ADAPTATION TO LOWEST DENSITY REGIONS WITH APPLICATION TO SUPPORT RECOVERY

By Tim Patschkowski and Angelika Rohde

Ruhr-Universität Bochum

A scheme for locally adaptive bandwidth selection is proposed which sensitively shrinks the bandwidth of a kernel estimator at lowest density regions such as the support boundary which are unknown to the statistician. In case of a Hölder continuous density, this locally minimax-optimal bandwidth is shown to be smaller than the usual rate, even in case of homogeneous smoothness. Some new type of risk bound with respect to a density-dependent standardized loss of this estimator is established. This bound is fully non-asymptotic and allows to deduce convergence rates at lowest density regions that can be substantially faster than $n^{-1/2}$. It is complemented by a weighted minimax lower bound which splits into two regimes depending on the value of the density. The new estimator adapts into the second regime, and it is shown that simultaneous adaptation into the fastest regime is not possible in principle as long as the Hölder exponent is unknown. Consequences on plug-in rules for support recovery are worked out in detail. In contrast to those with classical density estimators, the plug-in rules based on the new construction are minimax-optimal, up to some logarithmic factor.

1. Introduction.

Adaptation in the classical context of nonparametric function estimation in Gaussian white noise has been extensively studied in the statistical literature. Since Nussbaum (1996) has established asymptotic equivalence in Le Cam's sense for the nonparametric models of density estimation and Gaussian white noise, a rigorous framework is provided which allows to carry over specific statistical results established for the Gaussian white noise model to the model of density estimation, at least in dimension one. Density estimation is as one of the most fundamental problems in statistics subject to a variety of recent studies, see e.g. Efromovich (2008), Gach, Nickl and Spokoiny (2013), Lepski (2013), Birgé (2014) and Liu and Wong (2014). It has become clear that under the conditions for the asymptotic equivalence to hold, minimax rates of convergence in density estimation with respect to pointwise or mean integrated squared error loss coincide with the optimal convergence rates obtained in the context of nonparametric regression, and the procedures are typically identical on the level of ideas. A main requisite on the density for Nussbaum's (1996) asymptotic equivalence is the assumption that it is compactly supported and uniformly bounded away from zero on its support. If this assumption is violated, the density estimation experiment may produce statistical features which do not have any analog in the regression context. For instance, minimax estimation of non-compactly supported

*Supported by the DFG Priority Program SPP 1324, RO 3766/2-1.

Keywords and phrases: Anisotropic density estimation, bandwidth selection, adaptation to lowest density regions, density dependent minimax optimality, support estimation

densities under L_p-loss bears striking differences to the compact case, see Juditsky and Lambert-Lacroix (2004), Reynaud-Bouret, Rivoirard and Tuleau-Malot (2011) and Goldenshluger and Lepski (2011, 2013). The minimax rates reflect an interplay of the regularity parameters and the parameter of the loss function, an effect which is caused by the tail behavior of the densities under consideration. In this article we recover such an exclusive effect even for compactly supported densities. It turns out that minimax estimation in regions where the density is small is possible with higher accuracy although fewer observations are available, leading to rates which can be substantially faster than $n^{-1/2}$. Even more, this accuracy can be achieved to a large extent without a priori knowledge of these regions by a kernel density estimator with an adaptively selected bandwidth. As discovered by Butucea (2001), the exact constant of normalization for pointwise adaptive univariate density estimation on Sobolev classes depends increasingly on the density at the point of estimation itself. The crucial observation is that the classical bias variance trade-off does not reflect the dependence of the kernel estimator's variance on the density, which brings the idea of an estimated variance in the bandwidth selection rule into play. Although Butucea's interesting result requires the point of estimation to be fixed, it suggests that a potential gain in the rate might be possible at lowest density regions. In this paper we investigate the problem of adaptation to lowest density regions under anisotropic Hölder constraints. A bandwidth selection rule is introduced which provably attains fast pointwise rates of convergence at lowest density regions. On this way, new weighted lower risk bounds over anisotropic Hölder classes are established, which split into two regimes depending on the value of the density. We show that the new estimator uniformly improves the global minimax rate of convergence, adapts to the second regime and finally that adaptation into the fastest regime is not possible in principle if the density's regularity is unknown. We identify the best possible adaptive rate of convergence

$$n^{-\frac{\bar{\beta}}{\bar{\beta}+d}}$$

up to a logarithmic factor, where $\bar{\beta}$ is the unnormalized harmonic mean of the d-dimensional Hölder exponent.
This breakpoint determines the attainable speed of convergence of plug-in estimators for functionals of the density where the quality of estimation at the boundary is crucial. We exemplarily demonstrate it for the problem of support recovery. In order to line up with the related results of Cuevas and Fraiman (1997) about plug-in rules for support estimation and Rigollet and Vert (2009) on minimax analysis of plug-in level-set estimators, we measure the performance of the plug-in support estimator with respect to the global measure of symmetric difference of sets under the margin condition (Polonik (1995), see also Mammen and Tsybakov (1999) and Tsybakov (2004)). In contrast to level set estimation however, plug-in rules for the support functional possess sub-optimal convergence rates when the classical kernel density estimator with minimax-optimal global bandwidth choice is used. We derive the optimal minimax rate for support recovery

$$n^{-\frac{\gamma\beta}{\beta+d}},$$

where γ denotes the margin exponent, d the dimension and β the isotropic Hölder exponent. Our result demonstrates that support recovery is possible with higher accuracy than level set estimation as already conjectured by Tsybakov (1997). We finally show that the performance of the plug-in support estimator resulting from our new density estimator turns out to be minimax-optimal up to a logarithmic factor.

The article is organized as follows. Section 2 contains the basic notations. In Section 3 the adaptive density estimator is introduced, new weighted lower pointwise risk bounds are derived and the optimality performance of the estimator is proved. Section 4 addresses the important problem of density support estimation as an example of a functional which substantially benefits from the new density estimator. The proofs are deferred to Section 5 and the supplemental article [Patschkowski and Rohde (2015)].

2. Preliminaries and notation.

All our estimation procedures are based on a sample of n real-valued d-dimensional random vectors $X_i = (X_{i,1}, \dots, X_{i,d})$, $i = 1, \dots, n$ ($d \geq 1$ and if not stated otherwise $n \geq 2$), that are independent and identically distributed according to some unknown probability measure $\mathbb{P}$ on $\mathbb{R}^d$ with continuous Lebesgue density p. $\mathbb{E}_p^{\otimes n}$ denotes the expectation with respect to the n-fold product measure $\mathbb{P}^{\otimes n}$. Let

$$\hat{p}_{n,h}(t) = \hat{p}_{n,h}(t, X_1, \dots, X_n) := \frac{1}{n} \sum_{i=1}^{n} K_h(t - X_i),$$

denote the kernel density estimator with d-dimensional bandwidth $h = (h_1, \dots, h_d)$ at point $t \in \mathbb{R}^d$, where

$$K_h(x) := \left(\prod_{i=1}^{d} h_i \right)^{-1} K\left(\frac{x_1}{h_1}, \dots, \frac{x_d}{h_d} \right)$$

describes a rescaled kernel supported on $\prod_{i=1}^{d} [-h_i, h_i]$. The kernel function K is assumed to be compactly supported on $[-1, 1]^d$ and to be of product structure, i.e. $K(x_1, \dots, x_d) = \prod_{i=1}^{d} K_i(x_i)$. Additionally, $K_{i,h_i}(x) := h_i^{-1} K_i(x/h_i), i = 1, \dots, d$. The components K_i are assumed to integrate to one and to be continuous on its support with $K_i(0) > 0$. If not stated otherwise, they are symmetric and nonnegative, implying that the kernel is of first order. Recall that K is said to be of kth order, $k = (k_1, ..., k_d) \in \mathbb{N}^d$, if the functions $x \mapsto x_i^{j_i} K_i(x_i)$, $j_i \in \mathbb{N}$ with $1 \leq j_i \leq k_i$, $i = 1, \dots, d$, satisfy

$$\int x_i^{j_i} K_i(x_i) d\lambda(x_i) = 0,$$

where λ^d denotes the Lebesgue measure on $\mathbb{R}^d$ throughout the article. The Lebesgue measure on $\mathbb{R}$ is denoted by λ. For any function $f : \mathbb{R}^d \to \mathbb{R}$ and $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ we define the univariate functions

$$\begin{aligned} f_{i,x} : \mathbb{R} &\longrightarrow \mathbb{R} \\ y &\longmapsto f(x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_d). \end{aligned} \tag{2.1}$$

and denote by $P_{y,l}^{(f_{i,x})}$ the Taylor polynomial

$$P_{y,l}^{(f_{i,x})}(\cdot) := \sum_{k=0}^{l} \frac{f_{i,x}^{(k)}(y)}{k!}(\cdot - y)^k \tag{2.2}$$

of $f_{i,x}$ at the point $y \in \mathbb{R}$ of degree l (whenever it exists). Let $\mathscr{H}_d(\beta, L)$ be the anisotropic Hölder class with regularity parameters (β, L), i.e. any function f belonging to this class fulfills for all $y, y' \in \mathbb{R}$ the inequality

$$\sup_{x\in\mathbb{R}^d} |f_{i,x}(y) - f_{i,x}(y')| \;\le\; L\,|y - y'|^{\beta_i}$$

for those $i \in \{1, \ldots, d\}$ with $\beta_i \le 1$, and in case $\beta_i > 1$ admits derivates with respect to its i-th coordinate up to the order $\lfloor \beta_i \rfloor := \max\{n \in \mathbb{N} : n < \beta_i\}$, such that the approximation by the Taylor polynomial satisfies

$$\sup_{x\in\mathbb{R}^d} \left| f_{i,x}(y) - P_{y',\lfloor\beta_i\rfloor}^{(f_{i,x})}(y) \right| \;\le\; L\,|y - y'|^{\beta_i} \qquad \text{for all} \;\; y, y' \in \mathbb{R}.$$

For adaptation issues, it is assumed that $\beta = (\beta_1, \ldots, \beta_d) \in \prod_{i=1}^d [\beta_{i,l}^*, \beta_{i,u}^*]$ and $L \in [L_l^*, L_u^*]$ for some positive constants $\beta_{i,l}^* < \beta_{i,u}^*$, $i = 1, ..., d$, and $L_l^* < L_u^*$. For short, we simply write β^* and L^* for the couples (β_l^*, β_u^*) and (L_l^*, L_u^*), and finally $\mathcal{R}(\beta^*, L^*)$ for the rectangle $\prod_{i=1}^d [\beta_{i,l}^*, \beta_{i,u}^*] \times [L_l^*, L_u^*]$. It turns out that all rates of convergence emerging in an anisotropic setting involve the unnormalized harmonic mean of the smoothness parameters

$$\bar{\beta} := \left(\sum_{i=1}^d \frac{1}{\beta_i} \right)^{-1}.$$

To focus on rates only and for ease of notation we denote by c positive constants that may change from line to line. All relevant constants will be numbered consecutively. Dependencies of the constants on the functional classes' parameters are always indicated and it should be kept in mind that the constants can potentially depend on the chosen kernel, the loss function and the dimension as well. Furthermore, $\mathscr{P}_d(\beta, L)$ denotes the set of all probability densities in $\mathscr{H}_d(\beta, L)$. It is well-known that any function $f \in \mathscr{P}_d(\beta, L)$ is uniformly bounded by a constant

$$c_1(\beta, L) = \sup\{\|p\|_{\sup} : p \in \mathscr{P}_d(\beta, L)\} \tag{2.3}$$

depending on the regularity parameters only.

3. New lower risk bounds, adaptation to lowest density regions.

The fully nonparametric problem of estimating a density p at some given point $t = (t_1, \ldots, t_d)$ has quite a long history in the statistical literature and has been extensively studied. Considering different estimators, a very natural question is whether there is an estimator that is optimal and how optimality can be exactly described. A common concept of optimality is stated in a minimax framework.

An estimator $T_n(t) = T_n(t, X_1, \ldots, X_n)$ is called minimax-optimal over the class $\mathscr{P}_d(\beta, L)$ if its risk matches the minimax risk

$$\inf_{T_n(t)} \sup_{p \in \mathscr{P}_d(\beta,L)} \mathbb{E}_p^{\otimes n} \left| T_n(t) - p(t) \right|^r$$

for some $r \geq 1$, where the infimum is taken over all estimators. However, the minimax approach is often rated as quite pessimistic as it aims at finding an estimator which performs best in the worst situation. Different in spirit is the oracle approach. Within a prespecified class $\mathscr{T}$ of estimators, it aims at finding for any individual density the estimator $\hat{T}_n \in \mathscr{T}$ which is optimal, leading to oracle inequalities of the form

$$\mathbb{E}_p^{\otimes n} \left| \hat{T}_n(t) - p(t) \right|^r \;\leq\; c \inf_{T_n \in \mathscr{T}} \mathbb{E}_p^{\otimes n} \left| T_n(t) - p(t) \right|^r \;+\; R_n$$

with a remainder term R_n depending on the class $\mathscr{T}$, the underlying density p and the sample size only. Besides having the drawback that there is no notion of optimality judging about the adequateness of the estimator's class, an equally severe problem may be caused by the fact that the remainder term is uniform in $\mathscr{T}$ and thus a worst case remainder. The latter is responsible for the fact that our fast convergence rates cannot be deduced from the oracle inequality in Goldenshluger and Lepski (2013), the order for their remainder being unimprovable, however. In this article, we introduce the notion of best possible p-dependent minimax speed of convergence $\psi^n_{p(t),\beta,L}$ within the function class $\mathscr{P}_d(\beta, L)$ and aim at constructing an estimator $T_n(t)$ bounding the risk

$$\sup_{p \in \mathscr{P}_d(\beta,L)} \sup_{\substack{t \in \mathbb{R}^d: \\ p(t) > 0}} \mathbb{E}_p^{\otimes n} \left(\frac{|T_n(t) - p(t)|}{\psi^n_{p(t),\beta,L}} \right)^r$$

uniformly over a range of parameters (β, L). Firstly, this requires a suitable definition of the quantity $\psi^n_{p(t),\beta,L}$.

3.1. *New weighted lower risk bound.* As we want to work out the explicit dependence on the value of the density, it seems suitable to fix an arbitrary constant $\varepsilon \in (0,1)$, and to pick out maximal not necessarily disjoint subsets U_δ of $\mathscr{P}_d(\beta, L)$ with the following properties: $\cup U_\delta = \{p \in \mathscr{P}_d(\beta, L) : p(t) > 0\}$, and pairwise ratios $p(t)/q(t)$, $p, q \in U_\delta$, are bounded away from zero by ε and from infinity by $1/\varepsilon$. This motivates the construction of the subsequent theorem.

Theorem 3.1 (New weighted lower risk bound). *For any $\beta = (\beta_1, \ldots, \beta_d)$ with $0 < \beta_i \leq 2$, $i = 1, \ldots, d$, $L > 0$ and $r \geq 1$, there exist constants $c_2(\beta, L, r) > 0$ and $n_0(\beta, L) \in \mathbb{N}$, such that for every $t \in \mathbb{R}$ the pointwise minimax risk over Hölder-smooth densities is bounded from below by*

$$\inf_{0 < \delta \leq c_1(\beta,L)} \inf_{T_n(t)} \sup_{\substack{p \in \mathscr{P}_d(\beta,L): \\ \delta/2 \leq p(t) \leq \delta}} \mathbb{E}_p^{\otimes n} \left(\frac{|T_n(t) - p(t)|}{\psi^n_{p(t),\beta}} \right)^r \;\geq\; c_2(\beta, L, r)$$

for all $n \geq n_0(\beta, L)$, where $\psi^n_{x,\beta} := x \wedge (x/n)^{\frac{\bar{\beta}}{2\bar{\beta}+1}}$ and $c_1(\beta, L)$ defined in (2.3).

REMARK 3.2. *(i) The lower bound of the above theorem is attained by the oracle estimator*

$$T_n(t) \;:=\; \hat{p}_{n,h_{n,\delta}}(t) \cdot \mathbb{1}\big\{\delta \geq n^{-\bar{\beta}/(\bar{\beta}+1)}\big\} \tag{3.1}$$

with $h_{n,\delta,i} = (\delta/n)^{\frac{1}{2\bar{\beta}+1}\frac{1}{\beta_i}}$. *Hence,* $\psi^n_{p(t),\beta}$ *cannot be improved in principle. We refer to it in the sequel as* p*-dependent speed of convergence within the functional class* $\mathscr{P}_d(\beta, L)$.

(ii) Note that for the classical minimax rate $n^{-\bar{\beta}/(2\bar{\beta}+1)}$,

$$\lim_{n\to\infty} \inf_{0<\delta\leq c_1(\beta,L)} \inf_{T_n(t)} \sup_{\substack{p\in\mathscr{P}_d(\beta,L):\\ \delta/2\leq p(t)\leq\delta}} \mathbb{E}_p^{\otimes n}\left(\frac{|T_n(t)-p(t)|}{n^{-\bar{\beta}/(2\bar{\beta}+1)}}\right)^r \;=\; 0$$

as a direct consequence of the subsequently formulated Theorem 3.3. The p*-dependent speed of convergence* $\psi^n_{p(t),\beta}$ *is of substantially smaller order than the classical one along a shrinking neighborhood of lowest density regions.*

Note that the exponent $\bar{\beta}/(2\bar{\beta}+1)$ implicitly depends on the dimension d and coincides in case of isotropic smoothness with the well-known exponent $\beta/(2\beta+d)$. It splits into two regimes which are listed and specified in the following table.

	Regime	Rate $\psi^n_{x,\beta}$
(i)	$x \leq n^{-\frac{\bar{\beta}}{\bar{\beta}+1}}$	x
(ii)	$n^{-\frac{\bar{\beta}}{\bar{\beta}+1}} < x \leq c_1(\beta, L)$	$\left(\frac{x}{n}\right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}}$

The worst p-dependent speed of convergence within $\mathscr{P}_d(\beta, L)$, namely

$$\sup_{0<x\leq c_1(\beta,L)} \psi^n_{x,\beta},$$

reveals the classical minimax rate $n^{-\bar{\beta}/(2\bar{\beta}+1)}$. The fastest rate in regime (ii) is of the order

$$n^{-\bar{\beta}/(\bar{\beta}+1)} \quad \text{for} \quad x = n^{-\bar{\beta}/(\bar{\beta}+1)},$$

which is substantially smaller than the classical minimax risk bound. Figure 1 visualizes the split-up into the regimes and relates the new p-dependent rate of Theorem 3.1 to the classical minimax rate for different sample sizes from $n = 50$ to $n = 800$.

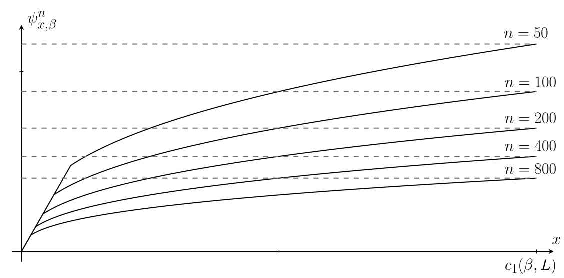

Fig 1. *New lower bound (solid line), Classical lower bound (dashed line)*

It becomes apparent from the proof that the lower bound actually even holds for the subset of (β, L)-regular densities with compact support. At first glance however, the new lower bound is of theoretical value only, because the value of a density at some point to be estimated is unknown. The question is whether it is possible to improve the local rate of convergence of an estimator without prior knowledge in regions where fewer observations are available, that is, to which extent it is possible to adapt to lowest density regions.

3.2. *Adaptation to lowest density regions.* Adaptation is an important challenge in nonparametric estimation. Lepski (1990) introduced a sequential multiple testing procedure for bandwidth selection of kernel estimators in the Gaussian white noise model. It has been widely used and refined for a variety of adaptation issues over the last two decades. For recent references, see Giné and Nickl (2010), Chichignoud (2012), Goldenshluger and Lepski (2011, 2013), Chichignoud and Lederer (2014), Jirak, Meister and Reiß (2014), Dattner, Reiß and Trabs (2014), and Bertin, Lacour and Rivoirard (2014) among many others. Our subsequently constructed estimator is based on the anisotropic bandwidth selection procedure of Kerkyacharian, Lepski and Picard (2001), which has been developed in the Gaussian white noise model, but incorporates the new approach of adaptation to lowest density regions. Although Goldenshluger and Lepski (2013) pursue a similar goal via some kind of empirical risk minimization, their oracle inequality provides no faster rates than $n^{-1/2}$ times the average of the density over the unit cube around the point under consideration. They deduce from it adaptive minimax rates of convergence with respect to the L_p-risk over anisotropic Nikol'skii classes for density estimation on $\mathbb{R}^d$. As concerns adaptation to lowest density regions such as the unknown support boundary, this oracle inequality is not sufficient as no faster rates than $n^{-1/2}$ can be deduced from it, and it is not clear whether these faster rates are attainable for their estimator in principle. Besides having the drawback that there is no notion of optimality judging about the adequateness of the estimator's class, an equally severe problem of the oracle approach may be caused by the fact that the remainder term is uniform in the estimator's class and thus a worst case remainder. The latter is responsible for the fact that our fast convergence rates cannot be deduced from the oracle inequality in Goldenshluger and Lepski (2013), the order for their remainder being unimprovable, however. It raises the question whether this imposes a fundamental limit on the possible range of adaptation. We shall demonstrate in what follows that it is even possible to attain substantially faster rates, indeed that adaptation to the whole second regime of Theorem 3.1 is an achievable goal, and that this describes precisely the full range where adaptation to lowest density regions is possible as long as the density's regularity is unknown. Our procedure uses kernel density estimators $\hat{p}_{n,h}(t)$ with multivariate bandwidths $h = (h_1, \ldots, h_d)$, which are able to deal with different degrees of smoothness in different coordinate directions. Note that optimal bandwidths for estimation of Hölder-continuous densities are typically derived by a bias-variance trade-off balancing the bias bound

$$\left|p(t) - \mathbb{E}_p^{\otimes n} \hat{p}_{n,h}(t)\right| \ \leq \ c(\beta, L) \cdot \sum_{i=1}^{d} h_i^{\beta_i}, \tag{3.2}$$

see (5.3) in Section 5 for details, against the rough variance bound

$$\operatorname{Var}(\hat{p}_{n,h}(t)) \ \leq \ \frac{c_1(\beta, L) \|K\|_2^2}{n \prod_{i=1}^d h_i}, \tag{3.3}$$

where $\|\cdot\|_2$ is the Euclidean norm (on $L_2(\lambda^d)$). This bound leads to suboptimal rates of convergence whenever the density is small since it is not able to capture small values of p in a small neighborhood around t in contrast to the sharp convolution bound

$$\operatorname{Var}(\hat{p}_{n,h}(t)) \ \leq \ \frac{1}{n}((K_h)^2 * p)(t) \ =: \ \sigma_t^2(h). \tag{3.4}$$

Balancing (3.2) and (3.4) leads to smaller bandwidths at lowest density regions as compared to bandwidths resulting from the classical bias-variance trade-off between (3.2) and (3.3). The convolution bound (3.4) is unknown and it is natural to replace it by its unbiased empirical version

$$\tilde{\sigma}_t^2(h) \ := \ \frac{1}{n^2 \prod_{i=1}^d h_i^2} \sum_{i=1}^{n} K^2\left(\frac{t - X_i}{h}\right).$$

However, $\tilde{\sigma}_t^2(h)$ concentrates extremely poorly around its mean if the bandwidth h is small, which is just the important situation at lowest density regions. Precisely, Bernstein's inequality provides the bound

$$\mathbb{P}^{\otimes n}\left(\left|\frac{\tilde{\sigma}_t^2(h)}{\sigma_t^2(h)} - 1\right| \geq \eta\right) \ \leq \ 2\exp\left(-\frac{3\eta^2}{2(3+2\eta)\|K\|_{\sup}^2}\ \sigma_t^2(h) \cdot n^2 \prod_{i=1}^d h_i^2\right), \tag{3.5}$$

which suggests to study the following truncated versions instead

$$\begin{aligned} \sigma_{t,\mathrm{trunc}}^2(h) &:= \max\left\{\frac{\log^2 n}{n^2 \prod_{i=1}^d h_i^2},\ \sigma_t^2(h)\right\}, \\ \tilde{\sigma}_{t,\mathrm{trunc}}^2(h) &:= \max\left\{\frac{\log^2 n}{n^2 \prod_{i=1}^d h_i^2},\ \tilde{\sigma}_t^2(h)\right\}. \end{aligned} \tag{3.6}$$

Without the logarithmic term, the truncation level ensures tightness of the family of random variables $\tilde{\sigma}^2_{t,\mathrm{trunc}}(h)/\sigma^2_{t,\mathrm{trunc}}(h)$, because the exponent in (3.5) remains a non-degenerate function in η. The logarithmic term is introduced in order to guarantee sufficient concentration of $\sup_h |1-\tilde{\sigma}^2_{t,\mathrm{trunc}}(h)/\sigma^2_{t,\mathrm{trunc}}(h)|$.

Construction of the adaptive estimator. Our estimation procedure is developed in the anisotropic setting, in which neither the variance bound nor the bias bound provides an immediate monotone behavior in the bandwidth. Unlike in the univariate or isotropic multivariate case, Lepski's (1990) idea of mimicking the bias-variance trade-off fails. Consequently, our estimation scheme imitates the anisotropic procedure of Kerkyacharian, Lepski and Picard (2001) and Klutchnikoff (2005), developed in the Gaussian white noise model, with the following changes. Firstly, their threshold given by the variance bound in the Gaussian white noise setting is replaced essentially with the truncated estimate in (3.6), which is sensitive to small values of the density. Moreover, it is crucial in the anisotropic setting that our procedure uses an ordering of bandwidths according to these estimated variances instead of an ordering according to the product of the bandwidth's components. The bandwidth selection scheme chooses a bandwidth in the set

$$\mathcal{H} := \left\{ h=(h_1,\ldots,h_d) \in \prod_{i=1}^d (0,h_{\max,i}] \ : \ \prod_{i=1}^d h_i \geq \frac{\log^2 n}{n} \right\},$$

where for simplicity we set $(h_{\max,1},\ldots,h_{\max,d}) = (1,\ldots,1)$. Let furthermore

$$\mathcal{J} := \left\{ j=(j_1,\ldots,j_d) \in \mathbb{N}_0^d \ : \ \sum_{i=1}^d j_i \leq \left\lfloor \log_2\left(\frac{n}{\log^2 n}\right)\right\rfloor \right\}$$

be a set of indices and denote by

$$\mathcal{G} := \left\{ (2^{-j_1},\ldots,2^{-j_d}) \ : \ j\in\mathcal{J} \right\} \ \subset \ \mathcal{H}$$

the corresponding dyadic grid of bandwidths, that serves as a discretization for the multiple testing problem in Lepski's selection rule. For ease of notation, we abbreviate dependences on the bandwidth $(2^{-j_1},\ldots,2^{-j_d})$ by the multiindex j. Next, with $j\wedge m$ denoting the minimum by component, the set of admissible bandwidths is defined as

$$\begin{aligned} \mathcal{A} = \mathcal{A}(t) := \Big\{ j\in\mathcal{J} : |\hat{p}_{n,j\wedge m}(t) - \hat{p}_{n,m}(t)| \leq c_{14}\sqrt{\hat{\sigma}_t^2(m)\log n} \\ \text{for all } m\in\mathcal{J} \text{ with } \hat{\sigma}_t^2(m) \geq \hat{\sigma}_t^2(j) \Big\}, \end{aligned} \tag{3.7}$$

with a properly chosen constant $c_{14} = c_{14}(\beta^*, L^*)$ satisfying the constraint (5.17) appearing in the proof of Theorem 3.3. Here, both the threshold and the ordering of bandwidths are defined via the truncated variance estimator

$$\hat{\sigma}_t^2(h) := \min\left\{ \tilde{\sigma}^2_{t,\mathrm{trunc}}(h), \frac{\|K\|_2^2\, c_1}{n\prod_{i=1}^d h_i} \right\}$$

$$= \min\left\{ \max\left[\frac{\log^2 n}{n^2 \prod_{i=1}^d h_i^2},\ \frac{1}{n^2 \prod_{i=1}^d h_i^2} \sum_{i=1}^n K^2\left(\frac{t-X_i}{h}\right)\right],\ \frac{\|K\|_2^2\, c_1}{n \prod_{i=1}^d h_i} \right\}, \tag{3.8}$$

where $c_1 = c_1(\beta^*, L^*)$ is an upper bound on $c_1(\beta, L)$ in the range of adaptation. The threshold in (3.7) could be modified by a further logarithmic factor to avoid the dependence of the constants on the range of adaptation. Recall again that this refined estimated threshold is crucial for our estimation scheme. The procedure selects the bandwidth among all admissible bandwidths with

$$\hat{j} = \hat{j}(t) \in \underset{j \in \mathcal{A}}{\arg\min}\ \hat{\sigma}_t^2(j). \tag{3.9}$$

Finally,

$$\hat{p}_n := \hat{p}_{n,\hat{j}} \wedge c_1$$

defines the adaptive estimator. In case of isotropic Hölder smoothness it is sufficient to restrict the grid to bandwidths with equal components, and we even simplify the method by replacing the ordering by estimated variances in condition (3.8) "for all $m \in \mathcal{J}$ with $\hat{\sigma}_t^2(m) \geq \hat{\sigma}_t^2(j)$" by the classical order "for all $m \in \mathcal{J}$ with $m \geq j$" as the componentwise ordering is the same for all components.

Performance of the adaptive estimator. Clearly, the truncation in the threshold imposes serious limitations to which extent adaptation to lowest densities regions is possible. However, a careful analysis of the ratio

$$\sup_h \left| \frac{\tilde{\sigma}^2_{t,\mathrm{trunc}}(h)}{\sigma^2_{t,\mathrm{trunc}}(h)} - 1 \right|$$

rather than the difference $\sup_h |\tilde{\sigma}^2_{t,\mathrm{trunc}}(h) - \sigma^2_{t,\mathrm{trunc}}(h)|$ allows to prove indeed that adaptation is possible in the whole second regime.

THEOREM 3.3 (New upper bound). *For any rectangle $\mathcal{R}(\beta^*, L^*)$ with $[\beta^*_{i,l}, \beta^*_{i,u}] \subset (0, 2]$, $[L^*_l, L^*_u] \subset (0, \infty)$ and $r \geq 1$, there exists a constant $c_3(\beta^*, L^*, r) > 0$, such that the new density estimator $\hat{p}_n$ with adaptively chosen bandwidth according to (3.9) satisfies*

$$\sup_{(\beta,L)\in\mathcal{R}(\beta^*,L^*)}\ \sup_{p\in\mathscr{P}_d(\beta,L)}\ \sup_{t\in\mathbb{R}^d} \mathbb{E}_p^{\otimes n}\left(\frac{|\hat{p}_n(t)-p(t)|}{\tilde{\psi}^n_{p(t),\beta}}\right)^r \ \leq\ c_3(\beta^*, L^*, r),$$

where

$$\tilde{\psi}^n_{x,\beta}\ :=\ \left[n^{-\frac{\bar{\beta}}{\bar{\beta}+1}} \vee \left(x/n\right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}}\right] (\log n)^{3/2}.$$

The p-dependent speed of convergence $\tilde{\psi}^n_{p(t),\beta}$ (except the logarithmic factor) is plotted in Figure 2, which shows the superiority of the new estimator in low density regions. It also depicts that the new estimator is able to adapt to regime

(ii) up to a logarithmic factor, and that it improves the rate of convergence significantly in both regimes as compared to the classical minimax rate. Besides, although not emphasized before, $\hat{p}_n$ is fully adaptive to the smoothness in terms of Hölder regularity.

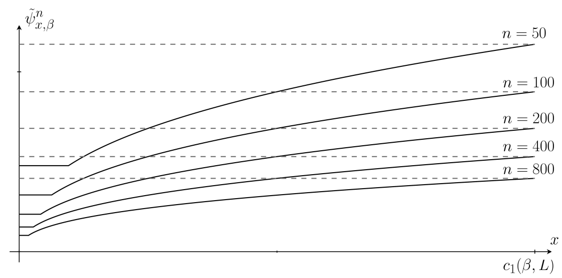

FIG 2. *New upper bound without logarithmic factor (solid line), Classical upper bound (dashed line)*

As ψ and $\tilde{\psi}$ coincide (up to a logarithmic factor) in regime (ii) but differ in regime (i), the question arises whether the breakpoint

$$n^{-\bar{\beta}/(\bar{\beta}+1)}$$

describes the fundamental bound on the range of adaptation to lowest density regions. The following result shows that this is indeed the case as long as the density's regularity is unknown.

THEOREM 3.4. *For any $\beta_2 < \beta_1 \leq 2$ and any sequence $(\rho(n))$ converging to infinity with*

$$\rho(n) = O\Big(n^{\frac{\beta_1-\beta_2}{(2\beta_1+1)(\beta_2+1)}} (\log n)^{-3/2}\Big),$$

there exist $L_1, L_2 > 0$ and densities $p_n \in \mathscr{P}_1(\beta_1, L_1)$ with

$$\frac{n^{-\beta_1/(\beta_1+1)}}{p_n(t)} = o(1)$$

as $n \to \infty$, such that for every estimator $T_n(t)$ satisfying

$$\mathbb{E}^{\otimes n}_{p_n} |T_n(t) - p_n(t)| \;\leq\; c_3(\beta_1^*, L_1^*, r) \left(\frac{p_n(t)}{n}\right)^{\frac{\beta_1}{2\beta_1+1}} (\log n)^{3/2}, \tag{3.10}$$

there exist $n_0(\beta_1, \beta_2, L_1, L_2)$ and a constant $c > 0$ both independent of t, with

$$\sup_{\substack{q \in \mathscr{P}_d(\beta_2, L_2): \\ q(t) \leq c(n)\cdot n^{-\frac{\beta_2}{\beta_2+1}}}} \frac{\mathbb{E}^{\otimes n}_q |T_n(t) - q(t)|}{n^{-\frac{\beta_2}{\beta_2+1}}} \geq c$$

for all $n \geq n_0(\beta_1, \beta_2, L_1, L_2)$ and any sequence $(c(n))$ with $c(n) \geq \rho(n)^{-1}$.

The following consideration provides a heuristic reason why adaptation to regime (i) is not possible in principle. Consider the univariate and Lipschitz continuous triangular density $p : \mathbb{R} \to \mathbb{R}, x \mapsto (1 - |x|)\mathbb{1}\{|x| \le 1\}$. If $\delta_n < n^{-\beta/(\beta+1)} = n^{-1/2}$, the expected number of observations in $\{p \le \delta_n\}$ is less than one. Without the knowledge of the regularity, it is intuitively clear that it is impossible to predict whether local averaging is preferable to just estimating by zero.

3.2.1. *Adaptation to lowest density regions when β is known.* If the Hölder-exponent $\beta \in (0, 2]$ is known to the statistician, the form of the oracle estimator (3.1) suggests that some further improvement in regime (i) might be possible by considering the truncated estimator

$$\hat{p}_n(\cdot) \cdot \mathbb{1}\Big\{\hat{p}_n(\cdot) \ge n^{-\frac{\bar{\beta}}{\bar{\beta}+1}} (\log n)^{\zeta_1}\Big\} \tag{3.11}$$

for some suitable constant $\zeta_1 > 0$. In fact, elementary algebra shows that this threshold does not affect the performance in regime (ii) (up to a logarithmic term). For isotropic Hölder smoothness, we prove in the supplemental article [Patschkowski and Rohde (2015)] that the estimator (3.11) indeed attains the p-dependent speed of convergence

$$\vartheta^n_{p(t),\beta} = \psi^n_{p(t),\beta} \vee n^{-\zeta_2}$$

up to logarithmic terms, with $\psi^n_{x,\beta}$ as defined in Theorem 3.1. Here, the constant ζ_2 can be made arbitrarily large by enlarging c_{14} and ζ_1. That is, if the Hölder exponent is known, adaptation to regime (i) is possible to a large extent.

3.2.2. *Extension to $\beta > 2$.* As concerns an extension of Theorem 3.1 and Theorem 3.3 to arbitrary $\beta > 2$, Lemma 5.1 (ii) demonstrates that the variance of the kernel density estimator never falls below the reference speed of convergence $\tilde{\psi}^n_{p(t),\beta}$. However, it can be substantially larger, resulting in a lower speed of convergence as compared to the reference speed of convergence. Therefore, it seems necessary to introduce a p-dependent speed of convergence which does not incorporate the value of the density $p(t)$ only but also information on the derivatives. An exception of outstanding importance are points t close to the support boundary, because not only $p(t)$ itself but also all derivatives are necessarily small. Theorem A.1, which is deferred to the supplemental article [Patschkowski and Rohde (2015)], reveals that our procedure then even reaches the fast adaptive speed of convergence at the support boundary for *every* $\beta > 0$. In fact, as $\beta \to \infty$, adaptive rates arbitrarily close to n^{-1} can be attained.

4. Application to support recovery. The phenomenon of faster rates of convergence in regions where the density is small may have strong consequences on plug-in rules for certain functionals of the density. As an application of the results of Section 3, we investigate the support plug-in functional. Support estimation has a long history in the statistical literature. Geffroy (1964) and Rényi and Sulanke (1963, 1964) are cited as pioneering reference most commonly, followed by further contributions of Chevalier (1976), Devroye and Wise (1980), Grenander (1981), Hall (1982), Groeneboom (1988), Tsybakov (1989, 1991, 1997), Cuevas (1990), Korostelev and Tsybakov (1993), Härdle, Park and Tsybakov (1995), Mammen and Tsybakov (1995), Cuevas and Fraiman (1997), Gayraud (1997), Hall, Nussbaum and Stern (1997), Baíllo, Cuevas and Justel (2000), Cuevas and Rodríguez-Casal (2004), Klemelä (2004), and Biau, Cadre and Pelletier (2008), Biau, Cadre, Mason and Pelletier (2009), Brunel (2013), and Cholaquidis, Cuevas and Fraiman (2014) as a by far non-exhaustive list of contributions. In order to demonstrate the substantial improvement in the rates of convergence for the plug-in support estimator based on the new density estimator, we first establish minimax lower bounds for support estimation under the margin condition which have not been provided in the literature so far. Theorem 4.4 and Theorem 4.5 then reveal that the minimax rates for the support estimation problem are substantially faster than for the level set estimation problem, as already conjectured in Tsybakov (1997). In fact, in the level set estimation framework, when β and L are given, the classical choice of a bandwidth of order $n^{-1/(2\beta+d)}$ in case of isotropic Hölder smoothness leads directly to a minimax-optimal plug-in level set estimator as long as the offset is suitably chosen (Rigollet and Vert 2009). In contrast, this bandwidth produces suboptimal rates in the support estimation problem, no matter how the offset is chosen. At first sight, this makes the plug-in rule as a by-product of density estimation inappropriate. We shall demonstrate subsequently, however, that our new density estimator avoids this problem. In order to line up with the results of Cuevas and Fraiman (1997) and Rigollet and Vert (2009), we work essentially under the same type of conditions. The distance between two subsets A and B of $\mathbb{R}^d$ is measured by

$$d_\Delta(A, B) := \lambda^d(A \Delta B),$$

where Δ denotes the symmetric difference of sets

$$A \Delta B := (A \setminus B) \cup (B \setminus A).$$

Subsequently, $\bar{A}$ denotes the topological closure of a set $A \subset \mathbb{R}^d$. We impose the following condition, which characterizes the complexity of the problem. It was introduced by Polonik (1995), see also Mammen and Tsybakov (1999), Tsybakov (2004) and Cuevas and Fraiman (1997), where the latter authors referred to it as sharpness order.

Definition 4.1 (Margin condition). *A density $p : \mathbb{R}^d \to \mathbb{R}$ is said to satisfy the κ-margin condition with exponent $\gamma > 0$, if*

$$\lambda^d \left(\overline{\{x \in \mathbb{R}^d \mid 0 < p(x) \le \varepsilon\}} \right) \le \kappa_2 \cdot \varepsilon^\gamma$$

for all $0 < \varepsilon \le \kappa_1$, where $\kappa = (\kappa_1, \kappa_2) \in (0, \infty)^2$.

In particular, $\lambda^d(\partial \Gamma_p) = 0$ for every density which satisfies the margin condition, where $\partial \Gamma_p$ denotes the boundary of the support Γ_p. To highlight the line of ideas, we restrict the application to the important special case of isotropic smoothness.

Let $\mathscr{H}_d^{iso}(\beta, L)$ denote the isotropic Hölder class with one-dimensional parameters β and L, which is for $0 < \beta \leq 1$ defined by

$$\mathscr{H}_d^{iso}(\beta, L) := \Big\{ f : \mathbb{R}^d \to \mathbb{R} \,:\, |f(x) - f(y)| \,\leq\, L\, \|x-y\|_2^\beta \ \text{ for all } \ x, y \in \mathbb{R}^d \Big\}.$$

For $\beta > 1$ it is defined as the set of all functions $f : \mathbb{R}^d \to \mathbb{R}$ that are $\lfloor \beta \rfloor$ times continuously differentiable such that the following property is satisfied

$$\Big| f(x) - P_{y, \lfloor \beta \rfloor}^{(f)}(x) \Big| \,\leq\, L\, \|x-y\|_2^\beta \quad \text{for all } x, y \in \mathbb{R}^d, \tag{4.1}$$

where

$$P_{y,l}^{(f)}(x) := \sum_{|k| \leq l} \frac{D^k f(y)}{k_1! \cdots k_d!} (x_1 - y_1)^{k_1} \cdots (x_d - y_d)^{k_d}$$

with $|k| := \sum_{i=1}^d k_i$ and the partial differential operator

$$D^k := \frac{\partial^{|k|}}{\partial x_1^{k_1} \ldots \partial x_d^{k_d}}$$

denotes the multivariate Taylor polynomial of f at the point $y \in \mathbb{R}^d$ up to the l-th order, see also (2.2) for the coinciding definition in one dimension. Correspondingly, $\mathscr{P}_d^{iso}(\beta, L)$ denotes the set of probability densities contained in $\mathscr{H}_d^{iso}(\beta, L)$. The following lemma demonstrates, that not every combination of margin exponent and Hölder continuity is possible.

LEMMA 4.2. *There exists a compactly supported density in $\mathscr{P}_d^{iso}(\beta, L)$ satisfying the margin condition to the exponent γ if and only if $\gamma\beta \leq 1$.*

4.1. *Lower risk bounds for support recovery.* For any subset $A \subset \mathbb{R}^d$ and $\varepsilon > 0$ the closed outer parallel set of A at distance $\varepsilon > 0$ is given by

$$A^\varepsilon := \Big\{ x \in \mathbb{R}^d : \inf_{y \in A} \|x - y\|_2 \leq \varepsilon \Big\}$$

and the closed inner ε-parallel set by $A^{-\varepsilon} := \overline{\left((A^c)^\varepsilon \right)^c}$. Here, $\|\cdot\|_2$ denotes the Euclidean norm (on $\mathbb{R}^d$). A support satisfying

$$0 \,<\, \liminf_{\varepsilon \to 0} \frac{\lambda^d(\Gamma_p \setminus \Gamma_p^{-\varepsilon})}{\lambda^d(\Gamma_p^\varepsilon \setminus \Gamma_p)} \,\leq\, \limsup_{\varepsilon \to 0} \frac{\lambda^d(\Gamma_p \setminus \Gamma_p^{-\varepsilon})}{\lambda^d(\Gamma_p^\varepsilon \setminus \Gamma_p)} \,<\, \infty$$

is referred to as boundary regular support. Note that a support is always boundary regular if its Minkowski surface measure is well defined (in the sense that outer and inner Minkowski content exist and coincide). The minimax lower bound is formulated under the assumption of Γ_p fulfilling the following complexity condition (to the exponent $\mu = \gamma\beta$), which even slightly weakens the assumption of boundary regularity under the margin condition.

Definition 4.3 (Complexity condition). *A set A is said to satisfy the ξ-complexity condition to the exponent $\mu > 0$ if for all $0 < \varepsilon \le \xi_1$ there exists a disjoint decomposition $A = A_{1,\varepsilon} \cup A_{2,\varepsilon}$ such that*

$$\frac{\lambda^d(A^{\varepsilon}_{1,\varepsilon} \setminus A_{1,\varepsilon}) \vee \lambda^d(A_{2,\varepsilon})}{\varepsilon^{\mu}} \le \xi_2,$$

where $\xi = (\xi_1, \xi_2) \in (0,\infty)^2$.

Note that a boundary regular support of a (β, L)-Hölder-smooth density satisfying the margin condition to the exponent γ fulfills the complexity condition to the exponent $\mu \ge \gamma\beta$ for the canonical decomposition $\Gamma_p = \Gamma_p \cup \emptyset$. Let us finally relate the margin condition 4.1 to the two-sided margin condition

$$\lambda^d\{x \in \mathbb{R}^d : 0 < |p(x) - \lambda| \le \varepsilon\} \le c\,\varepsilon^{\gamma},$$

which is imposed in the context of density level set estimation for some level $\lambda > 0$, c.f. Rigollet and Vert (2009). If $\Gamma_{p,\lambda} = \{x \in \mathbb{R}^d : p(x) > \lambda\}$ denotes the λ-level set at level $\lambda > 0$, the two-sided (κ, γ)-margin condition provides the bound

$$\lambda^d\left(\Gamma^{\varepsilon}_{p,\lambda} \setminus \Gamma_{p,\lambda}\right) \le \kappa_2 (c\varepsilon^{\beta\wedge 1})^{\gamma} \tag{4.2}$$

for all $\varepsilon \le \kappa_1$, where $c = L$ for $\beta \le 1$ and $c = \sup_{x\in\mathbb{R}^d} \|\nabla p(x)\|_2$ for $\beta > 1$. In contrast, the margin condition at $\lambda = 0$ provides no bound on $\lambda^d(\Gamma^{\varepsilon}_p \setminus \Gamma_p)$. The complexity condition is a mild assumption which guarantees such type of bound. For $\beta \le 1$, the relation (4.2) for $\lambda = 0$ implies the complexity condition to the exponent $\mu = \gamma\beta$. Note that the typical situation is indeed

$$\lambda^d(\Gamma^{\varepsilon}_p \setminus \Gamma_p)/\varepsilon = \mathcal{O}(1) \quad \text{and} \quad \varepsilon/\lambda^d(\Gamma^{\varepsilon}_p \setminus \Gamma_p) = \mathcal{O}(1)$$

as $\varepsilon \to 0$. For instance, this holds true for any finite union of convex sets in $\mathbb{R}^d$ as a consequence of the isoperimetric inequality (Theorem III.2.2, Chavel 2001) and Theorem 3.1 (Bhattacharya and Rango Rao 1976). If it exists, the limit

$$\lim_{\varepsilon \searrow 0} \frac{\lambda^d(\Gamma^{\varepsilon}_p \setminus \Gamma_p)}{\varepsilon}$$

corresponds to the surface measure of the boundary if the latter is sufficiently regular. Due to the relation $\gamma\beta \le 1$ by Lemma 4.2 and the decomposition into suitable subsets, the complexity condition relaxes this regularity condition on the surface area substantially. The subset of $\mathscr{P}^{iso}_d(\beta, L)$ consisting of densities satisfying the κ-margin condition to the exponent γ with support fulfilling the ξ-complexity condition to the exponent $\mu = \gamma\beta$ is denoted by $\mathscr{P}^{iso}_d(\beta, L, \gamma, \kappa, \xi)$.

Theorem 4.4 (Minimax lower bound). *For any $\beta > 0$ and any margin exponent $\gamma > 0$ with $\gamma\beta \le 1$, there exist $c_4(\beta, L) > 0$, $n_0(\beta, L, \gamma) \in \mathbb{N}$ and parameters $\kappa, \xi \in (0,\infty)^2$, such that the minimax risk with respect to the measure of symmetric difference of sets is bounded from below by*

$$\inf_{\hat\Gamma_n} \sup_{p \in \mathscr{P}^{iso}_d(\beta,L,\gamma,\kappa,\xi)} \mathbb{E}^{\otimes n}_p \left[d_{\Delta}(\hat\Gamma_n, \Gamma_p) \right] \ge c_4(\beta, L) \cdot n^{-\frac{\gamma\beta}{\beta+d}}$$

for all $n \ge n_0(\beta, L, \gamma)$.

4.2. *Minimax-optimal plug-in rule.* We use the plug-in support estimator with the kernel density estimator of Section 3. This density estimator improves the rate of convergence in particular at the support boundary. For the isotropic procedure, the index set $\mathcal{J}$ is restricted to bandwidths coinciding in all components. We even simplify the ordering by estimated variances in condition (3.8) "for all $m \in \mathcal{J}$ with $\hat{\sigma}_t^2(m) \geq \hat{\sigma}_t^2(j)$" by the classical order "for all $m \in \mathcal{J}$ with $m \geq j$" as Lemma 5.2 shows that the relevant orderings are equivalent up to multiplicative constants for $0 < \beta \leq 2$. Furthermore, under isotropic smoothness it is natural to use a rotation invariant kernel, i.e. $K(x) = \tilde{K}(\|x\|_2)$ with $\tilde{K}$ supported on $[0,1]$ and continuous on its support with $\tilde{K}(0) > 0$. The following theorem shows that the corresponding plug-in rule

$$\hat{\Gamma}_n = \overline{\{x \in \mathbb{R}^d : \hat{p}_n(x) > \alpha_n\}}$$

with offset level

$$\alpha_n := c_5(\beta, L)\left(\frac{(\log n)^{3/2}}{n}\right)^{\frac{\beta}{\beta+d}}\sqrt{\log n} \tag{4.3}$$

and constant $c_5(\beta, L)$ specified in the proof of the following theorem, is able to recover the support with minimax optimal rate, up to a logarithmic factor.

THEOREM 4.5 (Uniform upper bound). *For any $\beta \leq 2$, $\gamma > 0$ with $\gamma\beta \leq 1$ and $\kappa, \xi \in (0,\infty)^2$, there exist a constant $c_6 = c_6(\beta, L, \gamma, \kappa, \xi) > 0$ and $n_0 \in \mathbb{N}$, such that*

$$\sup_{p \in \mathscr{P}_d^{iso}(\beta, L, \gamma, \kappa, \xi)} \mathbb{E}_p^{\otimes n}\left[d_\Delta\left(\Gamma_p, \hat{\Gamma}_n\right)\right] \leq c_6 \cdot n^{-\frac{\gamma\beta}{\beta+d}} (\log n)^{2\gamma}$$

for all $n \geq n_0$.

As the rate already indicates, it is getting apparent from the proof that this result can be established only if the minimax optimal density estimator actually adapts up to the fastest rate in regime (ii).

REMARK 4.6. *The results show the simultaneous optimality of the adaptive density estimator of Section 3 in the plug-in rule for support estimation. Correspondingly, they are restricted to $\beta \leq 2$. Whether the rate $n^{-\gamma\beta/(\beta+d)}$ is minimax optimal for $\beta > 2$ provided $\gamma\beta \leq 1$, and whether it can be attained by a plug-in rule in principle, remains open for the moment.*

Let us finally point out two consequences. We have shown that the optimal minimax rates for support estimation are significantly faster than the corresponding rates for level set estimation

$$n^{-\frac{\gamma\beta}{2\beta+d}}$$

under the margin condition (Rigollet and Vert 2009). Although any level set of a fixed density satisfying the margin condition to the exponent γ fulfills the complexity condition to the exponent $\mu = \gamma\beta$ as long as $\beta \leq 1$, the hypotheses in the proof of the lower bounds of Rigollet and Vert (2009) do even satisfy this condition for some fixed ξ, uniformly in n, as well. Hence, their optimal minimax rates of convergence remain the same under our condition. On an intuitive level, this phenomenon can be nicely motivated by comparing the Hellinger distance $H(\mathbb{P},\mathbb{Q})$ between the probability measure $\mathbb{P}$ with Lebesgue density p and $\mathbb{Q}$ whose Lebesgue density $q = p + \tilde{p}$ is a perturbation of p with a small function $\tilde{p}$ around the level $\alpha \geq 0$, see Tsybakov (1997), Extension (E4). If $\alpha > 0$, then simple Taylor expansion of $\sqrt{p+\tilde{p}}$ yields $H^2(\mathbb{P},\mathbb{Q}) \sim \int \tilde{p}^2 d\lambda^d$, whereas $H^2(\mathbb{P},\mathbb{Q}) \sim \int \tilde{p}\, d\lambda^d$ in case $\alpha = 0$. Thus, perturbations at the boundary ($\alpha = 0$) can be detected with the higher accuracy resulting in faster attainable rates for support estimation than for level set estimation. Moreover, the rates for plug-in support estimators already established in the literature by Cuevas and Fraiman (1997) turn out to be always suboptimal in case of Hölder continuous densities of boundary regular support. To be precise, Cuevas and Fraiman (1997) establish in Theorem 1 (c) a convergence rate under the margin condition given in terms of $\rho_n = n^\rho$ and the offset level $\alpha_n = n^{-\alpha}$ (in their notation), which are assumed to satisfy $0 < \alpha < \rho$ and their condition (R2), namely

$$\rho_n \int |\hat{p}_n - p| \, d\lambda^d \;=\; o_{\mathbb{P}}(1) \;\text{ and }\; \rho_n \alpha_n^{1+\gamma} \;=\; o(1) \;\text{ as }\; n \to \infty.$$

As a consequence, $\rho_n = o(n^{\beta/(2\beta+d)})$ for typical candidates $p \in \mathscr{P}_d^{iso}(\beta, L)$, i.e. densities p which are locally not smoother than (β, L)-regular. Under the margin condition to the exponent $\gamma > 0$, this limits their rate of convergence $n^{-\rho+\alpha}$ to

$$d_\Delta(\Gamma_p, \hat{\Gamma}_n) \;=\; o_{\mathbb{P}}\left(n^{-\frac{\beta}{2\beta+d}\frac{\gamma}{1+\gamma}}\right),$$

which is substantially slower than the above established minimax rate. The crucial point is that even with the improved density estimator of Section 3, the above mentioned condition on ρ_n in (R2) cannot be improved, because any estimator can possess the improved performance at lowest density regions only. For this reason, the L_1-speed of convergence of a density estimator is not an adequate quantity to characterize the performance of the corresponding plug-in support estimator.

5. Lemma 5.1 - 5.7, Proofs of Theorem 3.3 and Theorem 3.4.

Due to space constraints, all remaining proofs are deferred to the supplemental article [Patschkowski and Rohde (2015)]. In the proof of Theorem 3.3, we frequently make use of the bandwidth

$$\bar{h}_i := c_{11}(\beta, L) \cdot \max\left\{ \left(\frac{\log n}{n}\right)^{\frac{\bar{\beta}}{\bar{\beta}+1}\frac{1}{\beta_i}}, \left(\frac{p(t)\log n}{n}\right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}\frac{1}{\beta_i}} \right\} \tag{5.1}$$

for $i = 1, \ldots, d$, with constant $c_{11}(\beta, L)$ of Lemma 5.1, which can be thought of as an optimal adaptive bandwidth. The truncation in the definition of $\bar{h}$ results from the necessary truncation in $\sigma^2_{t,\mathrm{trunc}}$. With the exponents

$$\bar{j}_i = \bar{j}_i(t) := \left\lfloor \log_2\left(\frac{1}{\bar{h}_i}\right)\right\rfloor + 1, \quad i = 1, \ldots, n \tag{5.2}$$

the bandwidth $2^{-\bar{j}_i}$ is an approximation of $\bar{h}_i$ by the next smaller bandwidth on the grid $\mathcal{G}$ such that $\bar{h}_i/2 \leq 2^{-\bar{j}_i} \leq \bar{h}_i$ for all $i = 1, \ldots, d$.

Before turning to the proof of Theorem 3.3, we collect some technical ingredients. First, recall the classical upper bound on the bias of a kernel density estimator. With the notation provided in Section 2, and K of order $\max_i \beta_i$ at least, we obtain

$$\begin{aligned} b_t(h) := p(t) - \mathbb{E}_p^{\otimes n} \hat{p}_{n,h}(t) &= \int K(x)\Big(p(t+hx) - p(t)\Big) d\lambda^d(x) \\ &= \sum_{i=1}^d \int K(x)\Big(p([t,t+hx]_{i-1}) - p([t,t+hx]_i)\Big) d\lambda^d(x), \end{aligned}$$

using the notation $[x,y]_0 = y$, $[x,y]_d = x$, $[x,y]_i = (x_1, \ldots, x_i, y_{i+1}, \ldots, y_d)$, $i = 1, \ldots, d-1$ for two vectors $x, y \in \mathbb{R}^d$ and denoting by $hx = (h_1x_1, \ldots, h_dx_d)$ the componentwise product. Taylor expansions for those components i with $\beta_i \geq 1$ lead to

$$\begin{aligned} p([t,t+hx]_{i-1}) - p([t,t+hx]_i) &= \sum_{k=1}^{\lfloor \beta_i \rfloor} p_{i,[t,t+hx]_i}^{(k)}(t_i)\, \frac{(h_ix_i)^k}{k!} \\ &\quad + \Big(p([t,t+hx]_{i-1}) - P_{t_i,\lfloor \beta_i \rfloor}^{(p_{i,[t,t+hx]_i})}(t_i + h_ix_i)\Big). \end{aligned}$$

Hence,

$$|b_t(h)| \leq L \sum_{i=1}^d c_{12,i}(\beta)\, h_i^{\beta_i} =: B_t(h) \tag{5.3}$$

with constants $c_{12,i}(\beta) := \int |x_i|^{\beta_i} |K(x)|\, d\lambda^d(x) < \infty$.

With a slight abuse of notation, dependencies on some bandwidth $h = 2^{-j}$ are subsequently expressed in terms of the corresponding grid exponent $j = (j_1, \ldots, j_d)$, i.e. $B_t(h)$ equals $B_t(j)$, etc. For any multiindex j, we use the abbreviation

$$|j| := \sum_{i=1}^d j_i.$$

The following lemmata are crucial ingredients for the proof of Theorem 3.3.

Lemma 5.1. *(i) For any (β, L) with $0 < \beta_i \leq 2$, $p \in \mathscr{P}_d(\beta, L)$, and for any bandwidth $h = (h_1, \ldots, h_d)$ with $h_i \leq c_{11}(\beta, L)\, p(t)^{1/\beta_i}$, $i = 1, \ldots, d$ with*

$$c_{11}(\beta, L) := \min_{i=1,\ldots,d} \left(\frac{2dL}{\|K\|_2^2} \int |x_i|^{\beta_i} K^2(x) d\lambda^d(x) \right)^{-1/\beta_i},$$

the following inequality chain holds true

$$\frac{1}{2} \frac{\|K\|_2^2}{n \prod_{i=1}^d h_i} p(t) \;\leq\; \frac{1}{n}((K_h)^2 * p)(t) \;\leq\; \frac{3}{2} \frac{\|K\|_2^2}{n \prod_{i=1}^d h_i} p(t).$$

(ii) For any constant $c_{27} > 0$, there exists a constant $c_{23}(\beta, L) = c_{23}(\beta, L, c_{27}) > 0$, such that for any (β, L), $0 < \beta_i < \infty$, $i = 1, \ldots, d$, and $p \in \mathscr{P}_d(\beta, L)$,

$$\frac{c_{23}(\beta, L)}{n \prod_{i=1}^d h_i} p(t) \;\leq\; \frac{1}{n}((K_h)^2 * p)(t)$$

for every bandwidth $h = (h_1, \ldots, h_d)$ with $h_i \leq c_{27}\, p(t)^{1/\beta_i}$, $i = 1, \ldots, d$.

(iii) For any density p with isotropic Hölder smoothness (β, L), $0 < \beta < \infty$ and bandwidth h, we have

$$\frac{1}{n}((K_h)^2 * p)(t) \;\leq\; \frac{L\|K\|_2^2}{nh^d} \Big(h + \inf_{y \in \Gamma_p^c} \|t - y\|_2 \Big)^{\beta},$$

where K is a rotation invariant kernel supported on the closed Euclidean unit ball.

LEMMA 5.2. *There exists some constant $c_{13}(\beta, L) > 0$, such that for any $p \in \mathscr{P}_d(\beta, L)$, $0 < \beta_i \leq 2$, $i = 1, \ldots, d$, and $t \in \mathbb{R}^d$ the inequality*

$$\sigma^2_{t,\mathrm{trunc}}(j \wedge m) \;\leq\; c_{13}(\beta, L) \left(\sigma^2_{t,\mathrm{trunc}}(j) \vee \sigma^2_{t,\mathrm{trunc}}(m) \right)$$

holds true for all (non-random) indices $j = (j_1, \ldots, j_d)$ and $m = (m_1, \ldots, m_d)$ with $j \geq \bar{j}$ componentwise. If additionally $m \geq j$ componentwise, then

$$\sigma^2_{t,\mathrm{trunc}}(j) \leq c_{13}(\beta, L)\, \sigma^2_{t,\mathrm{trunc}}(m).$$

The next lemma carefully analyzes the ratio of the truncated quantities $\sigma^2_{t,\mathrm{trunc}}$ and $\tilde{\sigma}^2_{t,\mathrm{trunc}}$.

LEMMA 5.3. *For the quantities $\sigma^2_{t,\mathrm{trunc}}(h)$ and $\tilde{\sigma}^2_{t,\mathrm{trunc}}(h)$ defined in (3.6) and any $\eta \geq 0$ holds*

$$\mathbb{P}^{\otimes n} \left(\left| \frac{\tilde{\sigma}^2_{t,\mathrm{trunc}}(h)}{\sigma^2_{t,\mathrm{trunc}}(h)} - 1 \right| \geq \eta \right) \;\leq\; 2 \exp\left(- \frac{3\eta^2}{2(3 + 2\eta)\, \|K\|^2_{\sup}} \log^2 n \right).$$

LEMMA 5.4. *For any (β, L) with $0 < \beta_i \leq 2$, $i = 1, \ldots, d$, there exist constants $c_{15}(\beta, L)$ and $c_{21}(\beta, L) > 0$ such that for the multiindex $\bar{j}$ as defined in (5.2) and the bias upper bound B_t as given in (5.3),*

$$B_t(\bar{j}) \;\leq\; c_{15}(\beta, L) \sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j}) \log n} \tag{5.4}$$

$$\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j})} \;\leq\; c_{21}(\beta, L) \left\{ \left(\frac{\log n}{n} \right)^{\frac{\bar{\beta}}{\bar{\beta}+1}} \vee \left(\frac{p(t) \log n}{n} \right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}} \right\}. \tag{5.5}$$

LEMMA 5.5. *For any (non-random) index $j = (j_1, \ldots, j_d)$, the tail probabilities of the random variable*

$$Y := \frac{\hat{p}_{n,j}(t) - \mathbb{E}_p^{\otimes n}\hat{p}_{n,j}(t)}{\sqrt{\sigma^2_{t,\mathrm{trunc}}(j)\log n}},$$

are bounded by

$$\mathbb{P}^{\otimes n}(|Y| \geq \eta) \;\leq\; 2\exp\Big(-\frac{\log n}{4}\cdot(\eta^2\wedge\eta)\Big)$$

for any $\eta \geq 0$, any $t \in \mathbb{R}^d$ and $n \geq n_0$ with n_0 depending on $\|K\|_{\sup}$ only.

LEMMA 5.6. *Let Z be some non-negative random variable satisfying*

$$\mathbb{P}(Z \geq \eta) \;\leq\; 2\exp\Big(-A\eta\Big).$$

for some $A > 0$. Then

$$(\mathbb{E}Z^m)^{1/m} \leq c_{28}\frac{m}{A}$$

for any $m \in \mathbb{N}$, where the constant c_{28} does not depend on A and m.

LEMMA 5.7 (Klutchnikoff 2005). *For all $k, l \in \mathcal{J}$, the absolute value of the difference of bias terms is bounded by*

$$|b_t(k\wedge l) - b_t(l)| \;\leq\; 2B_t(k)$$

for all $t \in \mathbb{R}^d$.

PROOF OF THEOREM 3.3. Recall the notation of Section 3 and denote $\hat{p}_{n,\hat{j}} = \hat{p}_n$. In a first step, the risk

$$\mathbb{E}_p^{\otimes n}|\hat{p}_{n,\hat{j}}(t) - p(t)|^r$$

is decomposed as follows:

$$\begin{aligned}\mathbb{E}_p^{\otimes n}|\hat{p}_{n,\hat{j}}(t) - p(t)|^r &= \mathbb{E}_p^{\otimes n}\Big[|\hat{p}_{n,\hat{j}}(t) - p(t)|^r\cdot\mathbb{1}\{\hat{\sigma}_t^2(\hat{j}) \leq \hat{\sigma}_t^2(\bar{j})\}\Big]\\ &\quad + \mathbb{E}_p^{\otimes n}\Big[|\hat{p}_{n,\hat{j}}(t) - p(t)|^r\cdot\mathbb{1}\{\hat{\sigma}_t^2(\hat{j}) > \hat{\sigma}_t^2(\bar{j})\}\Big]\\ &=: R^+ + R^-. \qquad (5.6)\end{aligned}$$

We start with R^+, which is decomposed again as follows

$$\begin{aligned}R^+ \leq 3^{r-1}\bigg(&\mathbb{E}_p^{\otimes n}\Big[|\hat{p}_{n,\hat{j}}(t) - \hat{p}_{n,\hat{j}\wedge\bar{j}}(t)|^r\cdot\mathbb{1}\{\hat{\sigma}_t^2(\hat{j}) \leq \hat{\sigma}_t^2(\bar{j})\}\Big]\\ &+ \mathbb{E}_p^{\otimes n}\Big[|\hat{p}_{n,\hat{j}\wedge\bar{j}}(t) - \hat{p}_{n,\bar{j}}(t)|^r\cdot\mathbb{1}\{\hat{\sigma}_t^2(\hat{j}) \leq \hat{\sigma}_t^2(\bar{j})\}\Big]\\ &+ \mathbb{E}_p^{\otimes n}\Big[|\hat{p}_{n,\bar{j}}(t) - p(t)|^r\cdot\mathbb{1}\{\hat{\sigma}_t^2(\hat{j}) \leq \hat{\sigma}_t^2(\bar{j})\}\Big]\bigg)\end{aligned}$$

$$=: 3^{r-1}(S_1 + S_2 + S_3), \tag{5.7}$$

where we used the inequality $(x+y+z)^r \leq 3^{r-1}(x^r + y^r + z^r)$ for all $x, y, z \geq 0$. This decomposition bears the advantage that only kernel density estimators with well-ordered bandwidths are compared. We focus on the estimation of S_1, S_2 and S_3 and start with S_2 using the selection scheme's construction. Clearly, $\hat{j} \in \mathcal{A}$ as defined in (3.7). As a consequence the following inequality holds true

$$\begin{aligned}
S_2 \; &\leq c_{14}^r \, \mathbb{E}_p^{\otimes n} \left[\left(\hat{\sigma}_t^2(\bar{j}) \log n \right)^{r/2} \cdot \mathbb{1} \left\{ \left| \frac{\tilde{\sigma}_{t,\mathrm{trunc}}^2(\bar{j})}{\sigma_{t,\mathrm{trunc}}^2(\bar{j})} - 1 \right| < \; 1 \right\} \right] \\
&\quad + \; c_{14}^r \, \mathbb{E}_p^{\otimes n} \left[\left(\hat{\sigma}_t^2(\bar{j}) \log n \right)^{r/2} \cdot \mathbb{1} \left\{ \left| \frac{\tilde{\sigma}_{t,\mathrm{trunc}}^2(\bar{j})}{\sigma_{t,\mathrm{trunc}}^2(\bar{j})} - 1 \right| \geq \; 1 \right\} \right] \\
&\leq 2^{r/2} c_{14}^r \left(\min \left\{ \sigma_{t,\mathrm{trunc}}^2(\bar{j}), \frac{\|K\|_2^2 \, c_1}{n 2^{-|\bar{j}|}} \right\} \log n \right)^{r/2} \\
&\quad + \; c_{14}^r \left(\frac{\|K\|_2^2 \, c_1}{n 2^{-|\bar{j}|}} \log n \right)^{r/2} \mathbb{P}^{\otimes n} \left(\left| \frac{\tilde{\sigma}_{t,\mathrm{trunc}}^2(\bar{j})}{\sigma_{t,\mathrm{trunc}}^2(\bar{j})} - 1 \right| \geq \; 1 \right),
\end{aligned}$$

where we used the condition in the indicator function in the first summand to bound the estimated truncated variance $\tilde{\sigma}_{t,\mathrm{trunc}}^2$ from above by $2\sigma_{t,\mathrm{trunc}}^2$, and additionally the upper truncation level in the second summand. By the deviation inequality of Lemma 5.3, we can further estimate S_2 by

$$\begin{aligned}
S_2 \leq 2^{r/2} c_{14}^r \left(\sigma_{t,\mathrm{trunc}}^2(\bar{j}) \log n \right)^{r/2} \\
+ c_{14}^r \left(\frac{\|K\|_2^2 c_1}{n 2^{-|\bar{j}|}} \log n \right)^{r/2} \cdot 2 \exp \left(- \frac{3}{10 \|K\|_{\sup}^2} \log^2 n \right).
\end{aligned}$$

The second term is always of smaller order than the first term because $2^{-|\bar{j}|} \leq 1$, and therefore for $n \geq 2$,

$$\left(\frac{\|K\|_2^2 c_1}{n 2^{-|\bar{j}|}} \log n \right)^{r/2} \cdot 2 \exp \left(- \frac{3}{10 \|K\|_{\sup}^2} \log^2 n \right) \leq c \left(\frac{\log^3 n}{n^2 \left(2^{-|\bar{j}|} \right)^2} \right)^{r/2}$$

for some constant c depending on c_1, r and the kernel K only. Finally,

$$S_2 \leq c(\beta, L) \left(\sigma_{t,\mathrm{trunc}}^2(\bar{j}) \log n \right)^{r/2}.$$

We will now turn to S_3, the third term in (5.7). We split the risk into bias and stochastic error. It holds

$$S_3 \leq \mathbb{E}_p^{\otimes n} \left(|\hat{p}_{n,\bar{j}}(t) - \mathbb{E}_p^{\otimes n} \hat{p}_{n,\bar{j}}(t)| + B_t(\bar{j}) \right)^r \tag{5.8}$$

and by Lemma 5.4

$$B_t(\bar{j}) \leq c_{15}(\beta, L) \sqrt{\sigma_{t,\mathrm{trunc}}^2(\bar{j}) \log n}. \tag{5.9}$$

Denoting by

$$Z_k := \frac{\hat{p}_{n,k}(t) - \mathbb{E}_p^{\otimes n}\hat{p}_{n,k}(t)}{\sqrt{\sigma^2_{t,\mathrm{trunc}}(k)\log n}} \qquad \text{for } k \in \mathcal{J}, \tag{5.10}$$

the decomposition (5.8), the bias variance relation (5.9) and the inequality $(x+y)^r \le 2^{r-1}(x^r+y^r)$, $x,y \ge 0$ together with Lemma 5.6 yields

$$\begin{aligned} S_3 &\le \left(\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n\right)^{r/2} \cdot \mathbb{E}_p^{\otimes n}\left(|Z_{\bar{j}}| + c_{15}(\beta,L)\right)^r \\ &\le \left(\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n\right)^{r/2} \cdot 2^{r-1}\mathbb{E}_p^{\otimes n}\left(|Z_{\bar{j}}|^r + c_{15}(\beta,L)^r\right) \\ &\le c(\beta,L)\left(\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n\right)^{r/2}. \end{aligned}$$

It remains to show an analogous result for S_1, the first term in (5.7). Clearly,

$$\begin{aligned} S_1 \le \sum_{j\in\mathcal{J}} \mathbb{E}_p^{\otimes n}\Big[\Big(&\left|\hat{p}_{n,j}(t) - \mathbb{E}_p^{\otimes n}\hat{p}_{n,j}(t)\right| + \left|\hat{p}_{n,j\wedge\bar{j}}(t) - \mathbb{E}_p^{\otimes n}\hat{p}_{n,j\wedge\bar{j}}(t)\right| \\ &+ \left|b_t(j\wedge\bar{j}) - b_t(j)\right|\Big)^r \cdot \mathbb{1}\{\hat{\sigma}_t^2(j) \le \hat{\sigma}_t^2(\bar{j}),\, \hat{j} = j\}\Big]. \end{aligned} \tag{5.11}$$

By Lemma 5.7 and Lemma 5.4,

$$|b_t(j\wedge\bar{j}) - b_t(j)| \le 2B_t(\bar{j}) \le 2c_{15}(\beta,L)\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n}.$$

On account of this inequality and in view of (5.11), it suffices to bound the expectations in the following expression

$$\begin{aligned} S_1 \ &\le 3^{r-1}\left(\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n\right)^{r/2} \\ &\cdot \Bigg\{ \sum_{j\in\mathcal{J}} \mathbb{E}_p^{\otimes n}\left[\left(\frac{|\hat{p}_{n,j}(t) - \mathbb{E}_p^{\otimes n}\hat{p}_{n,j}(t)|}{\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n}}\right)^r \mathbb{1}\{\hat{\sigma}_t^2(j) \le \hat{\sigma}_t^2(\bar{j}),\, \hat{j} = j\}\right] \\ &\quad + \sum_{j\in\mathcal{J}} \mathbb{E}_p^{\otimes n}\left[\left(\frac{|\hat{p}_{n,j\wedge\bar{j}}(t) - \mathbb{E}_p^{\otimes n}\hat{p}_{n,j\wedge\bar{j}}(t)|}{\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n}}\right)^r \mathbb{1}\{\hat{\sigma}_t^2(j) \le \hat{\sigma}_t^2(\bar{j}),\, \hat{j} = j\}\right] \\ &\quad + \sum_{j\in\mathcal{J}} 2^r c_{15}(\beta,L)^r \cdot \mathbb{P}^{\otimes n}(\hat{j} = j)\Bigg\}. \end{aligned} \tag{5.12}$$

Denoting

$$A_{j,\bar{j}} := \left\{\left|\frac{\tilde{\sigma}^2_{t,\mathrm{trunc}}(j)}{\sigma^2_{t,\mathrm{trunc}}(j)} - 1\right| < \ \frac{1}{2} \text{ and } \left|\frac{\tilde{\sigma}^2_{t,\mathrm{trunc}}(\bar{j})}{\sigma^2_{t,\mathrm{trunc}}(\bar{j})} - 1\right| < \ \frac{1}{2}\right\}, \tag{5.13}$$

it follows

$$\sum_{j\in\mathcal{J}} \mathbb{E}_p^{\otimes n}\left[\left(\frac{|\hat{p}_{n,j}(t) - \mathbb{E}_p^{\otimes n}\hat{p}_{n,j}(t)|}{\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n}}\right)^r \mathbb{1}\{\hat{\sigma}_t^2(j) \le \hat{\sigma}_t^2(\bar{j}),\, \hat{j} = j\}\right]$$

$$
\begin{aligned}
&= \sum_{j\in\mathcal{J}} \mathbb{E}_p^{\otimes n}\left[\left(\frac{|\hat{p}_{n,j}(t)-\mathbb{E}_p^{\otimes n}\hat{p}_{n,j}(t)|}{\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n}}\right)^r \mathbb{1}\{\hat{\sigma}_t^2(j)\le\hat{\sigma}_t^2(\bar{j}),\, \hat{j}=j\}\cdot\mathbb{1}_{A_{j,\bar{j}}}\right]\\
&\quad+\sum_{j\in\mathcal{J}} \mathbb{E}_p^{\otimes n}\left[\left(\frac{|\hat{p}_{n,j}(t)-\mathbb{E}_p^{\otimes n}\hat{p}_{n,j}(t)|}{\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n}}\right)^r \mathbb{1}\{\hat{\sigma}_t^2(j)\le\hat{\sigma}_t^2(\bar{j}),\, \hat{j}=j\}\cdot\mathbb{1}_{A^c_{j,\bar{j}}}\right]\\
&=: S_{1,1}+S_{1,2}.
\end{aligned}
$$

Applying Lemma 5.6 and Hölder's inequality for any $p>1$,

$$
\begin{aligned}
S_{1,1} &\le \left(\frac{3(1\vee c_1\|K\|_2^2)}{c_1\|K\|_2^2}\right)^{r/2}\sum_{j\in\mathcal{J}}\mathbb{E}_p^{\otimes n}\left[|Z_j|^r\cdot\mathbb{1}\{\hat{j}=j\}\right]\\
&\le \left(\frac{3(1\vee c_1\|K\|_2^2)}{c_1\|K\|_2^2}\right)^{r/2}\left(1+\sum_{j\in\mathcal{J}}\mathbb{E}_p^{\otimes n}\left[|Z_j|^r\,\mathbb{1}\{|Z_j|\ge 1\}\mathbb{1}\{\hat{j}=j\}\right]\right)\\
&\le \left(\frac{3(1\vee c_1\|K\|_2^2)}{c_1\|K\|_2^2}\right)^{r/2}\left(1+\sum_{j\in\mathcal{J}}\mathbb{E}_p^{\otimes n}\Big[|Z_j|^{rp}\,\mathbb{1}\{|Z_j|\ge 1\}\Big]^{1/p}\cdot\mathbb{P}(\hat{j}=j)^{\frac{p-1}{p}}\right)\\
&\le \left(\frac{3(1\vee c_1\|K\|_2^2)}{c_1\|K\|_2^2}\right)^{r/2}\left(1+c_{28}^r\left(\frac{8rp}{\log n}\right)^r\sum_{j\in\mathcal{J}}\mathbb{P}(\hat{j}=j)^{\frac{p-1}{p}}\right)\\
&\le \left(\frac{3(1\vee c_1\|K\|_2^2)}{c_1\|K\|_2^2}\right)^{r/2}\left(1+c_{28}^r\left(\frac{8rp}{\log n}\right)^r\left(\sum_{j\in\mathcal{J}}\mathbb{P}(\hat{j}=j)\right)^{\frac{p-1}{p}}\cdot|\mathcal{J}|^{\frac{1}{p}}\right).
\end{aligned}
$$

By the constraint $2^{-|j|}\ge\log^2 n/n$ for any $j\in\mathcal{J}$, there exists some constant $c>0$ such that $|\mathcal{J}|\le c\,(\log n)^d$. Setting finally $p=d\log n$, yields $S_{1,1}\le c(\beta^*,L^*)$. As concerns $S_{1,2}$, by the Cauchy-Schwarz inequality,

$$
\begin{aligned}
S_{1,2} &\le \sum_{j\in\mathcal{J}}\left(\frac{\sigma^2_{t,\mathrm{trunc}}(j)}{\sigma^2_{t,\mathrm{trunc}}(\bar{j})}\right)^{r/2}\mathbb{E}_p^{\otimes n}\left[|Z_j|^r\,\mathbb{1}\{\hat{j}=j\}\,\mathbb{1}_{A^c_{j,\bar{j}}}\right]\\
&\le \sum_{j\in\mathcal{J}}\left(\frac{\sigma^2_{t,\mathrm{trunc}}(j)}{\sigma^2_{t,\mathrm{trunc}}(\bar{j})}\right)^{r/2}\mathbb{E}_p^{\otimes n}\left[|Z_j|^{2r}\,\mathbb{1}\{\hat{j}=j\}\right]^{1/2}\\
&\qquad\cdot\left\{\mathbb{P}^{\otimes n}\left(\left|\frac{\tilde{\sigma}^2_{t,\mathrm{trunc}}(j)}{\sigma^2_{t,\mathrm{trunc}}(j)}-1\right|\ge\frac{1}{2}\right)+\mathbb{P}^{\otimes n}\left(\left|\frac{\tilde{\sigma}^2_{t,\mathrm{trunc}}(\bar{j})}{\sigma^2_{t,\mathrm{trunc}}(\bar{j})}-1\right|\ge\frac{1}{2}\right)\right\}^{1/2}.
\end{aligned}
$$

Via the lower and upper truncation levels in the definition of $\sigma^2_{t,\mathrm{trunc}}$,

$$
\frac{\sigma^2_{t,\mathrm{trunc}}(k)}{\sigma^2_{t,\mathrm{trunc}}(l)}\;\le\;\frac{(1\vee c_1\|K\|_2^2)n^2}{\log^4 n}\qquad\text{for any } k,l\in\mathcal{J}, \tag{5.14}
$$

and the remaining expectation $\sum_{j\in\mathcal{J}} \mathbb{E}_p^{\otimes n}[|Z_j|^{2r}\,\mathbb{1}\{\hat{j}=j\}]$ can be bounded by Lemma 5.6 as above. Finally, the probabilities compensate (5.14) by Lemma 5.3. As concerns the expectation in (5.12), we proceed analogously using

$$\sigma^2_{t,\mathrm{trunc}}(j\wedge\bar{j}) \le c_{13}(\beta,L)\left(\sigma^2_{t,\mathrm{trunc}}(\bar{j})\vee\sigma^2_{t,\mathrm{trunc}}(j)\right)$$

by Lemma 5.2 and $\sigma^2_{t,\mathrm{trunc}}(j)\le c(\beta,L)\,\sigma^2_{t,\mathrm{trunc}}(\bar{j})$ on $A_{j,\bar{j}}\cap\{\hat{\sigma}^2_t(j)\le\hat{\sigma}^2_t(\bar{j})\}$. Combining the results for S_1, S_2 and S_3 proves that R^+ as defined in (5.6) is estimated by

$$R^+ \;\le\; c(\beta,L)\left(\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n\right)^{r/2}.$$

To deduce a similar inequality for R^-, it remains to investigate the probability

$$\mathbb{P}^{\otimes n}\Big(\hat{\sigma}^2_t(\hat{j})>\hat{\sigma}^2_t(\bar{j})\Big),$$

since $\hat{p}_n$ and p are both upper bounded by c_1. If $\hat{\sigma}^2_t(\hat{j})>\hat{\sigma}^2_t(\bar{j})$, then $\bar{j}$ cannot be an admissible exponent, see (3.7), because $\hat{j}$ had not been chosen in the minimization problem (3.9) otherwise. Hence, by definition there exists a multiindex $m\in\mathcal{J}$ with $\hat{\sigma}^2_t(m)\ge\hat{\sigma}^2_t(\bar{j})$ such that

$$|\hat{p}_{n,\bar{j}\wedge m}(t)-\hat{p}_{n,m}(t)|>c_{14}\sqrt{\hat{\sigma}^2_t(m)\log n}.$$

Subsuming, we get

$$\begin{aligned}&\mathbb{P}^{\otimes n}\Big(\hat{\sigma}^2_t(\hat{j})>\hat{\sigma}^2_t(\bar{j})\Big)\\&\qquad\le\sum_{m\in\mathcal{J}}\mathbb{P}^{\otimes n}\left(|\hat{p}_{n,\bar{j}\wedge m}(t)-\hat{p}_{n,m}(t)|>c_{14}\sqrt{\hat{\sigma}^2_t(m)\log n}\;,\;\;\hat{\sigma}^2_t(m)\ge\hat{\sigma}^2_t(\bar{j})\right),\end{aligned}$$

and we divide the absolute value of the difference of the kernel density estimators as in (5.11) into the difference of biases $|b_t(\bar{j}\wedge m)-b_t(m)|$ and two stochastic terms $|\hat{p}_{n,\bar{j}\wedge m}(t)-\mathbb{E}_p^{\otimes n}\hat{p}_{n,\bar{j}\wedge m}(t)|$ and $|\hat{p}_{n,m}(t)-\mathbb{E}_p^{\otimes n}\hat{p}_{n,m}(t)|$. As before,

$$|b_t(\bar{j}\wedge m)-b_t(m)|\le 2B_t(\bar{j})\le 2c_{15}(\beta,L)\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n}$$

by Lemma 5.7 and Lemma 5.4, leading to the inequality

$$\begin{aligned}&\mathbb{P}^{\otimes n}\Big(\hat{\sigma}^2_t(\hat{j})>\hat{\sigma}^2_t(\bar{j})\Big)\\&\quad\le\sum_{m\in\mathcal{J}}\mathbb{P}^{\otimes n}\Bigg(|\hat{p}_{n,\bar{j}\wedge m}(t)-\mathbb{E}_p^{\otimes n}\hat{p}_{n,\bar{j}\wedge m}(t)|+|\hat{p}_{n,m}(t)-\mathbb{E}_p^{\otimes n}\hat{p}_{n,m}(t)|\\&\qquad\qquad>c_{14}\sqrt{\hat{\sigma}^2_t(m)\log n}-2c_{15}(\beta,L)\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n}\;,\;\hat{\sigma}^2_t(m)\ge\hat{\sigma}^2_t(\bar{j})\Bigg)\\&\quad\le\sum_{m\in\mathcal{J}}\Big(\mathbb{P}^{\otimes n}(B_{1,m})+\mathbb{P}^{\otimes n}(B_{2,m})\Big)\end{aligned}$$

with

$$\begin{aligned}B_{1,m} := \Bigg\{ & \Big|\hat{p}_{n,\bar{j}\wedge m}(t) - \mathbb{E}_p^{\otimes n}\hat{p}_{n,\bar{j}\wedge m}(t)\Big| \\ & > \frac{1}{2}\left(c_{14}\sqrt{\hat{\sigma}_t^2(m)\log n} - 2c_{15}(\beta,L)\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n}\right), \hat{\sigma}_t^2(m) \geq \hat{\sigma}_t^2(\bar{j})\Bigg\}\\ B_{2,m} := \Bigg\{ & \Big|\hat{p}_{n,m}(t) - \mathbb{E}_p^{\otimes n}\hat{p}_{n,m}(t)\Big| \\ & > \frac{1}{2}\left(c_{14}\sqrt{\hat{\sigma}_t^2(m)\log n} - 2c_{15}(\beta,L)\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n}\right), \hat{\sigma}_t^2(m) \geq \hat{\sigma}_t^2(\bar{j})\Bigg\}.\end{aligned}$$

To start with the second probability, we intersect event $B_{2,m}$ with $A_{m,\bar{j}}$ as defined in (5.13). Obviously,

$$\mathbb{P}^{\otimes n}(B_{2,m}) \leq \mathbb{P}^{\otimes n}(B_{2,m} \cap A_{m,\bar{j}}) + \mathbb{P}^{\otimes n}(A^c_{m,\bar{j}}).$$

The definition of c_{14} and Lemma 5.5 allow to bound the probability

$$\begin{aligned}\mathbb{P}^{\otimes n}(B_{2,m} \cap A_{m,\bar{j}}) \ &\leq\ \mathbb{P}^{\otimes n}\left(\frac{|\hat{p}_{n,m}(t) - \mathbb{E}_p^{\otimes n}\hat{p}_{n,m}(t)|}{\sqrt{\sigma^2_{t,\mathrm{trunc}}(m)\log n}}\ >\ c_{16}(\beta,L)\right)\\ &\leq\ 2\exp\left(-\frac{c_{16}(\beta,L)^2 \wedge c_{16}(\beta,L)}{4}\log n\right) \qquad (5.15)\end{aligned}$$

with

$$c_{16}(\beta,L) := \left(\frac{c_{14}}{2} - c_{15}(\beta,L)\sqrt{2\,\frac{1\vee c_1\|K\|_2^2}{c_1\|K\|_2^2}}\right)\cdot\sqrt{\frac{1}{2}\,\frac{c_1\|K\|_2^2}{1\vee c_1\|K\|_2^2}}. \qquad (5.16)$$

At this point, we specify a lower bound on c_{14}. Precisely, c_{14} has to be chosen large enough to guarantee that

$$\frac{c_{16}(\beta,L)^2 \wedge c_{16}(\beta,L)}{4} \geq \frac{r\bar{\beta}}{\bar{\beta}+1} + 1 \qquad (5.17)$$

for any β in the range of adaptation. Finally, by means of Lemma 5.3,

$$\begin{aligned}\mathbb{P}^{\otimes n}(A^c_{m,\bar{j}}) &\leq \mathbb{P}^{\otimes n}\left(\left|\frac{\tilde{\sigma}^2_{t,\mathrm{trunc}}(\bar{j})}{\sigma^2_{t,\mathrm{trunc}}(\bar{j})} - 1\right| \geq \frac{1}{2}\right) + \mathbb{P}^{\otimes n}\left(\left|\frac{\tilde{\sigma}^2_{t,\mathrm{trunc}}(m)}{\sigma^2_{t,\mathrm{trunc}}(m)} - 1\right| \geq \frac{1}{2}\right)\\ &\leq 4\exp\left(-\frac{3}{32\|K\|^2_{\sup}}\log^2 n\right), \qquad (5.18)\end{aligned}$$

which is of smaller order than the bound in (5.15). Altogether, with this restriction on c_{14},

$$\mathbb{P}^{\otimes n}(B_{2,m}) \leq c(\beta,L)\left(\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n\right)^{r/2}.$$

By Lemma 5.2, the probability $\mathbb{P}^{\otimes n}(B_{1,m})$ can be bounded in the same way using additionally

$$\sigma^2_{t,\mathrm{trunc}}(\bar{j} \wedge m) \leq c_{13}(\beta, L)\Big(\sigma^2_{t,\mathrm{trunc}}(\bar{j}) \vee \sigma^2_{t,\mathrm{trunc}}(m)\Big) = c(\beta, L)\,\sigma^2_{t,\mathrm{trunc}}(m),$$

because $\sigma^2_{t,\mathrm{trunc}}(\bar{j}) \leq c(\beta, L)\,\sigma^2_{t,\mathrm{trunc}}(m)$ on the event $A_{m,\bar{j}} \cap \{\hat{\sigma}^2_t(m) \geq \hat{\sigma}^2_t(\bar{j})\}$. Summarizing,

$$\mathbb{P}^{\otimes n}\Big(\hat{\sigma}^2_t(\hat{j}) > \hat{\sigma}^2_t(\bar{j})\Big) \;\leq\; c(\beta, L)\left(\sigma^2_{t,\mathrm{trunc}}(\bar{j}) \log n\right)^{r/2}. \tag{5.19}$$

Finally, by Lemma 5.4,

$$\Big(\mathbb{E}_p^{\otimes n}|\hat{p}_{n,\hat{j}}(t) - p(t)|^r\Big)^{1/r} \leq c(\beta, L)\left\{\left(\frac{\log n}{n}\right)^{\frac{\bar{\beta}}{\bar{\beta}+1}} \vee \left(\frac{p(t)\log n}{n}\right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}}\right\}\sqrt{\log n}.$$

This completes the proof of Theorem 3.3. □

Proof of Theorem 3.4. Before we construct the densities p_n and q_n, we first specify their amplitudes Δ_n and δ_n in t, respectively. Let

$$\begin{aligned}
\Delta_n &:= n^{-\frac{\beta_1}{\beta_1+1}} \cdot \varrho(n) \\
\delta_n &:= 4c_3(\beta_1^*, L_1^*, r)\left(\frac{\Delta_n}{n}\right)^{\frac{\beta_1}{2\beta_1+1}}(\log n)^{3/2} \\
&= 4c_3(\beta_1^*, L_1^*, r)\Delta_n \cdot \varrho(n)^{-\frac{\beta_1+1}{2\beta_1+1}}(\log n)^{3/2},
\end{aligned} \tag{5.20}$$

for

$$\varrho(n) := n^{\frac{\beta_1-\beta_2}{(\beta_1+1)(\beta_2+1)}}$$

converging to infinity. Note first that with this choice of $\varrho(n)$ it holds that $\Delta_n = n^{-\beta_2/(\beta_2+1)}$ and hence tends to zero as n goes to infinity. The amplitude δ_n is smaller than Δ_n for sufficiently large n and hence also tends to zero. Furthermore, it holds

$$\delta_n = 4c_3(\beta_1^*, L_1^*, r) \cdot n^{-\frac{\beta_2}{\beta_2+1}} \cdot n^{\frac{\beta_2-\beta_1}{(2\beta_1+1)(\beta_2+1)}} \cdot (\log n)^{3/2} = o\left(n^{-\frac{\beta_2}{\beta_2+1}}\right)$$

Denote by $K(\,\cdot\,;\beta_i), i = 1, 2$ the univariate, symmetric and non-negative functions to the Hölder exponent β_i, respectively, as defined in the supplemental article [Patschkowski and Rohde (2015)], Section A.4, normalized by appropriate choices of $c_{17}(\beta_i)$ such that both functions integrate to one. Let $\tilde{L}_i = \tilde{L}_i(\beta_i), i = 1, 2$ be such that $K(\,\cdot\,;\beta_i) \in \mathscr{P}_1(\beta_i, \tilde{L}_i)$. Note that $K(\,\cdot\,;h,\beta_i) := h^{\beta_i}K(\cdot/h;\beta_i)$ has the same Hölder regularity as K (as opposed to $K_h(\,\cdot\,;\beta_i) := h^{-1}K(\cdot/h;\beta_i)$, which has the same Hölder parameter β_i but not necessarily the same $\tilde{L}_i$).

To ensure that $p_n(t) = \Delta_n$ we use the scaled version $K(\cdot - t; g_{1,n}, \beta_1)$ for some bandwidth $g_{1,n}$ defined below, preserving the Hölder regularity. In order to re-establish integrability to one, a second part is added alongside. The density q_n is then defined as p_n with a perturbation added and subtracted around t, i.e.

$$\begin{aligned} p_n(x) &= K(x-t; g_{1,n}, \beta_1) + K(x - t - g_{1,n} - g_{2,n}; g_{2,n}, \beta_1) \;\; \in \mathscr{P}_1(\beta_1, L_1) \\ q_n(x) &= p_n(x) - K(x-t; h_n, \beta_2) + K(x - t - 2h_n; h_n, \beta_2) \;\;\; \in \mathscr{P}_1(\beta_2, L_2), \end{aligned}$$

with

$$g_{1,n} := \left(\frac{\Delta_n}{K(0;\beta_1)}\right)^{\frac{1}{\beta_1}}, \quad g_{2,n} := \left(1 - g_{1,n}^{\beta_1+1}\right)^{\frac{1}{\beta_1+1}}, \quad h_n := \left(\frac{\Delta_n - \delta_n}{K(0;\beta_2)}\right)^{\frac{1}{\beta_2}}.$$

and suitable constants L_1 and L_2 independent of n. The construction of the hypotheses is depicted in Figure 3. Recall that the particular construction of $K(\,\cdot\,; h, \beta)$ does not change the Hölder parameters and note that the classes $\cup_{L>0}\mathscr{C}_c \cap \mathscr{P}_1(\beta, L)$, $0 < \beta \le 2$, are nested ($\mathscr{C}_c$ denotes the set of continuous functions from $\mathbb{R}$ to $\mathbb{R}$ of compact support). The bandwidth $g_{1,n}$ tends to zero and hence $g_{2,n}$ converges to one. In particular, $g_{2,n}$ is positive for sufficiently large n. In turn, h_n ensures that $q_n(t) = \delta_n$. Note furthermore that $\Delta_n > \Delta_n - \delta_n$ and $K(0;\beta_1) < K(0;\beta_2)$ since the constant $c_{17}(\beta)$ is monotonously increasing in β and $\beta_2 < \beta_1$. Thus, h_n is smaller than $g_{1,n}$ and consequently q_n is non-negative for sufficiently large n.

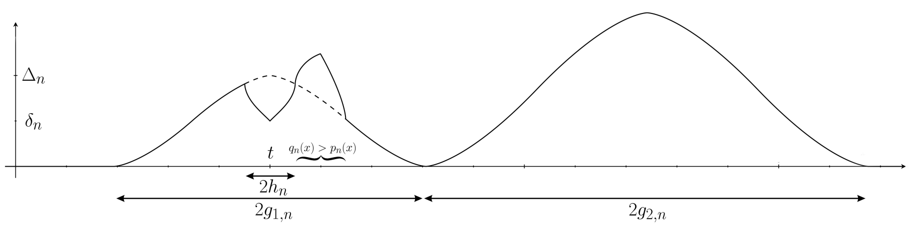

FIG 3. *Construction of p_n (dashed line) and q_n (solid line)*

Let $T_n(t)$ be an arbitrary estimator with property (3.10). Note first that we can pass on to the consideration of the estimator

$$\tilde{T}_n(t) := T_n(t) \cdot \mathbb{1}\left\{T_n(t) \le 2\Delta_n\right\},$$

since it both improves the quality of estimation of $p_n(t)$ and $q_n(t)$: Obviously,

$$\begin{aligned} \mathbb{E}_{p_n}^{\otimes n}|\tilde{T}_n(t) - p_n(t)| &= \mathbb{E}_{p_n}^{\otimes n}\left[p_n(t) \cdot \mathbb{1}\left\{T_n(t) - p_n(t) > p_n(t)\right\}\right] \\ &\quad + \mathbb{E}_{p_n}^{\otimes n}\left[\,|T_n(t) - p_n(t)| \cdot \mathbb{1}\left\{T_n(t) - p_n(t) \le p_n(t)\right\}\right] \\ &\le \mathbb{E}_{p_n}^{\otimes n}|T_n(t) - p_n(t)| \end{aligned}$$

and because of $q_n(t) \le p_n(t)$ also

$$\mathbb{E}_{q_n}^{\otimes n}|\tilde{T}_n(t) - q_n(t)| \le \mathbb{E}_{q_n}^{\otimes n}|T_n(t) - q_n(t)|.$$

As in the proof of the constrained risk inequality in Cai, Low and Zhao (2007), by reverse triangle inequality holds

$$\mathbb{E}_{q_n}^{\otimes n}|\tilde{T}_n(t) - q_n(t)| \geq (\Delta_n - \delta_n) - \mathbb{E}_{q_n}^{\otimes n}|\tilde{T}_n(t) - p_n(t)|.$$

In contrast to their proof, we need the decomposition

$$\begin{aligned} &\mathbb{E}_{q_n}^{\otimes n}|\tilde{T}_n(t) - q_n(t)| \\ &\qquad \geq (\Delta_n - \delta_n) - \mathbb{E}_{q_n}^{\otimes n}\Big[|T_n(t) - p_n(t)|\mathbb{1}_{B_n}\Big] - \mathbb{E}_{q_n}^{\otimes n}\Big[|\tilde{T}_n(t) - p_n(t)|\mathbb{1}_{B_n^c}\Big] \\ &\qquad =: (\Delta_n - \delta_n) - S_1 - S_2, \end{aligned} \tag{5.21}$$

where

$$B_n := \left\{ x = (x_1, \ldots, x_n) \in \mathbb{R}^n : \prod_{i=1}^n \frac{q_n(x_i)}{p_n(x_i)} \leq \frac{\Delta_n}{\delta_n} \right\}.$$

By definition of Δ_n and δ_n in (5.20) and the risk bound (3.10) the first two summands in (5.21) can be further estimated by

$$\begin{aligned} (\Delta_n - \delta_n) - S_1 &\geq (\Delta_n - \delta_n) - \mathbb{E}_{p_n}^{\otimes n}|T_n(t) - p_n(t)| \cdot \frac{\Delta_n}{\delta_n} \\ &\geq (\Delta_n - \delta_n)\left(1 - \frac{c_3(\beta_1^*, L_1^*, r)\left(\frac{\Delta_n}{n}\right)^{\frac{\beta_1}{2\beta_1+1}}(\log n)^{3/2}\frac{\Delta_n}{\delta_n}}{\Delta_n - \delta_n}\right) \\ &= \delta_n\left(\frac{\varrho(n)^{\frac{\beta_1+1}{2\beta_1+1}}(\log n)^{-3/2}}{4c_3(\beta_1^*, L_1^*, r)} - 1\right) \\ &\qquad \cdot \left(1 - \frac{c_3(\beta_1^*, L_1^*, r)\left(\frac{\Delta_n}{n}\right)^{\frac{\beta_1}{2\beta_1+1}}(\log n)^{3/2}\frac{\Delta_n}{\delta_n}}{\Delta_n\left(1 - 4c_3(\beta_1^*, L_1^*, r)\cdot\varrho(n)^{-\frac{\beta_1+1}{2\beta_1+1}}(\log n)^{3/2}\right)}\right), \end{aligned}$$

which is lower bounded by

$$\begin{aligned} (\Delta_n - \delta_n) - S_1 &\geq \delta_n \frac{\varrho(n)^{\frac{\beta_1+1}{2\beta_1+1}}(\log n)^{-3/2}}{8c_3(\beta_1^*, L_1^*, r)}\left(1 - \frac{2c_3(\beta_1^*, L_1^*, r)\left(\frac{\Delta_n}{n}\right)^{\frac{\beta_1}{2\beta_1+1}}(\log n)^{3/2}}{\delta_n}\right) \\ &= \delta_n \frac{\varrho(n)^{\frac{\beta_1+1}{2\beta_1+1}}(\log n)^{-3/2}}{16c_3(\beta_1^*, L_1^*, r)} \end{aligned}$$

for sufficiently large n. Furthermore,

$$S_2 \;\leq\; 2\Delta_n \cdot \mathbb{Q}_n^{\otimes n}(B_n^c) = \delta_n \frac{\varrho(n)^{\frac{\beta_1+1}{2\beta_1+1}}(\log n)^{-3/2}}{2c_3(\beta_1^*, L_1^*, r)} \cdot \mathbb{Q}_n^{\otimes n}(B_n^c),$$

and it remains to show that $\mathbb{Q}_n^{\otimes n}(B_n^c)$ tends to zero. By Markov's inequality,

$$
\begin{aligned}
\mathbb{Q}_n^{\otimes n}(B_n^c) = \mathbb{Q}_n^{\otimes n}\left(\prod_{i=1}^n \frac{q_n(X_i)}{p_n(X_i)} > \frac{\Delta_n}{\delta_n}\right)
&\leq \frac{\delta_n}{\Delta_n}\left(\mathbb{E}_{q_n}\frac{q_n(X_1)}{p_n(X_1)}\right)^n \\
&\leq \frac{\delta_n}{\Delta_n}\left(1 + \int \frac{q_n(x)}{p_n(x)} q_n(x) \mathbb{1}\left\{q_n(x) > p_n(x)\right\} dx\right)^n \\
&\leq \frac{\delta_n}{\Delta_n}\left(1 + \frac{(2\Delta_n - \delta_n)^2}{K(3h_n; g_{1,n}, \beta_1)} \cdot 2h_n\right)^n \\
&\leq \frac{\delta_n}{\Delta_n}\left(1 + \frac{4\Delta_n^2}{g_{1,n}^{\beta_1} K(3h_n/g_{1,n}; \beta_1)} \cdot 2h_n\right)^n \\
&\leq \frac{\delta_n}{\Delta_n}\left(1 + c(\beta_1, \beta_2)\Delta_n^{\frac{\beta_2+1}{\beta_2}}\right)^n
\end{aligned}
$$

for sufficiently large n, where the last inequality is due to

$$
h_n/g_{1,n} = c(\beta_1, \beta_2)\Delta_n^{\frac{\beta_1-\beta_2}{\beta_1\beta_2}} \longrightarrow 0\,,
$$

i.e. $K(3h_n/g_{1,n}; \beta_1)$ stays uniformly bounded away from zero. Finally,

$$
\begin{aligned}
\mathbb{Q}_n^{\otimes n}(B_n^c) &\leq \frac{\delta_n}{\Delta_n} \exp\left(n \log\left(1 + c(\beta_1, \beta_2)\Delta_n^{\frac{\beta_2+1}{\beta_2}}\right)\right) \\
&\leq \frac{\delta_n}{\Delta_n} \exp\left(n \cdot c(\beta_1, \beta_2)\Delta_n^{\frac{\beta_2+1}{\beta_2}}\right)
\end{aligned}
$$

and

$$
n\Delta_n^{\frac{\beta_2+1}{\beta_2}} = 1,
$$

such that $\mathbb{Q}_n^{\otimes n}(B_n^c) \leq c(\beta_1, \beta_2) \cdot \delta_n/\Delta_n \longrightarrow 0.$ □

Acknowledgement. We are very grateful to two anonymous referees and an associate editor for three constructive and detailed reports which led to a substantial improvement of our presentation and stimulated further interesting research.

SUPPLEMENTARY MATERIAL

Supplement A: Supplement to "Adaptation to lowest density regions with application to support recovery"
(doi: COMPLETED BY THE TYPESETTER; .pdf). Supplement A is organized as follows. Section A.1 contains the proofs of Lemma 5.1 – Lemma 5.6, which are central ingredients for the proof of Theorem 3.3. Section A.2 is concerned with the remaining proofs of Section 3. Section A.3 contains the proofs of Section 4. Section A.4 introduces a specific construction of a kernel function with prescribed Hölder regularity, which is frequently used throughout the article.

References.

[1] Baíllo, A., Cuevas, A. and Justel, A. (2000). Set estimation and nonparametric detection. *Canad. J. Statist.* **28** 765–782. MR1821433

[2] Bertin, K., Lacour, C. and Rivoirard, V. (2014). Adaptive pointwise estimation of conditional density function. arXiv:1312.7402

[3] Bhattacharya, R. N. and Rango Rao, R. (1976). *Normal Approximations and Asymptotic Expansions*, Wiley, New York. MR0436272

[4] Biau, G., Cadre, B. and Pelletier, B. (2008). Exact rates in density support estimation. *J. Multivariate Analysis* **99** 2185–2207. MR2463383

[5] Biau, G., Cadre, B., Mason, D.M. and Pelletier, B. (2009). Asymptotic normality in density support estimation. *Electron. J. Prob.* **14** 2617–2635. MR2570013

[6] Birgé, L. (2014). Model selection for density estimation with $\mathbb{L}_2$-loss. *Probab. Theory Relat. Fields* **158** 533–574. MR2256570

[7] Brunel, V.-E. (2013). Adaptive estimation of convex and polytopal density support. *Probab. Theory Relat. Fields*, to appear. arXiv:1309.6602

[8] Butucea, C. (2001). Exact adaptive pointwise estimation on Sobolev classes of densities. *ESAIM: Probability and Statistics* **5** 1–31. MR1845320

[9] Cai, T.T., Low, M.G. and Zhao, L.H. (2007). Trade-offs between global and local risks in nonparametric function estimation. *Bernoulli* **13** 1–19. MR2307391

[10] Chavel, I. (2001). *Isoperimetric Inequalities - Differential Geometric and Analytic Perspectives*, Cambridge University Press. MR1849187

[11] Chevalier, J. (1976). Estimation du support et du contour du support dune loi de probabilité. *Annales de l'Institut Henri Poincaré (B) Probabilités & Statistiques* **12** 339–364. MR0451491

[12] Chichignoud, M. and Lederer, J. (2014). A robust, fully adaptive M-estimator for pointwise estimation in heteroscedastic regression. *Bernoulli* **20** 1560–1599. MR3217454

[13] Chichignoud, M. (2012). Minimax and minimax adaptive estimation in multiplicative regression : locally bayesian approach. *Probab. Theory Relat. Fields* **153** 543–586. MR2948686

[14] Cholaquidis, A., Cuevas, A. and Fraiman, R. (2014). On Poincaré cone property. *Ann. Statist.* **42** 1–412. MR3189486

[15] Cuevas, A. (1990). On pattern analysis in the non-convex case. *Kybernetes* **19** 26–33. MR1084947

[16] Cuevas, A. and Fraiman, R. (1997). A plug-in approach to support estimation. *Ann. Statist.* **25** 2300–2312. MR1604449

[17] Cuevas, A. and Rodríguez-Casal, A. (2004). On boundary estimation. *Adv. in Appl. Probab.* **36** 340–354. MR2058139

[18] Dattner, I., Reiss, M. and Trabs, M. (2014). Adaptive quantile estimation in deconvolution with unknown error distribution. *Bernoulli*, to appear.

[19] Devroye, L. and Wise, G. L. (1980). Detection of abnormal behavior via nonparametric estimation of the support. *SIAM J. Appl. Math.* **38** 480–488. MR0579432

[20] Efromovich, S. (2008). Adaptive estimation of and oracle inequalities for probability densities and characteristic functions. *Ann. Statist.* **36** 1127–1155. MR2418652

[21] Gach, F., Nickl, R. and Spokoiny, V. (2013). Spatially adaptive density estimation by localised Haar projections. *Ann. Inst. Henri Poincar Probab. Stat.* **49** 900–914. MR3112439

[22] Gayraud, G. (1997). Estimation of functionals of density support. *Math. Methods Statist.* **6** 26–46. MR1456645

[23] Geffroy, J. (1964). Sur un probléme d'estimation géométrique. *Publ. Inst. Statist. Univ. Paris* **13** 191–210. MR0202237

[24] Giné, E. and Nickl, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. MR2604707

[25] Goldenshluger, A. and Lepski, O. V. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. MR2850214

[26] Goldenshluger, A. and Lepski, O. V. (2013). On adaptive minimax estimation on $\mathbb{R}^d$. *Prob. Theory Relat. Fields* **159** 479–543. MR3230001

[27] Grenander, U. (1981). *Abstract Inference*, Wiley, New York. MR0599175

[28] Groeneboom, P., (1988). Limit theorems for convex hulls. *Probab. Theory Relat. Fields* **79** 327–368. MR0959514

[29] Hall, P. (1982). On estimating the endpoint of a distribution. *Ann. Statist.* **10** 556–568. MR0653530

[30] Hall, P., Nussbaum, M. and Stern, S.E. (1997). On the estimation of a support curve of indeterminate sharpness. *J. Multivariate Analysis* **62** 204–232. MR1473874

[31] Härdle, W., Park, B. U. and Tsybakov, A. B. (2001). Estimation of non-sharp support boundaries. *J. Multivariate Analysis* **55** 205–218. MR1370400

[32] Jirak, M., Meister, A. and Reiss, M. (2014). Adaptive function estimation in nonparametric regression with one-sided errors. *Ann. Statist.* **42** 1970–2002.

[33] Juditsky, A. and Lambert-Lacroix, S. (2004). On minimax density estimation on $\mathbb{R}$. *Bernoulli* **10**, 187–220. MR2046772

[34] Kerkyacharian, G., Lepski, O. and Picard, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Relat. Fields* **121** 137–170. MR1863916

[35] Klemelä, J. (2004). Complexity penalized support estimation. *J. Multivariate Analysis* **88** 274–297. MR2025614

[36] Klutchnikoff, N. (2005). *Sur l'estimation adaptative de fonctions anisotropes.*, Ph.D.Thesis. Université Aix-Marseille I.

[37] Korostelev, A. P. and Tsybakov, A. B. (1993). *Minimax Theory of Image Reconstruction*, Springer, New York. MR1226450

[38] Lepski, O. V. (1990). One problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 459–470. MR1091202

[39] Lepski, O. V. (2013). Multivariate density estimation under sup-norm loss: oracle approach, adaptation and independence structure. *Ann. Statist.* **41** 1005–1034. MR3099129

[40] Lepski, O. V. (2014). Adaptive estimation over anisotropic functional classes via oracle approach. arXiv:1405.4504

[41] Liu, L. and Wong, W. H. (2014). Multivariate density estimation based on adaptive partitioning: Convergence rate, variable selection and spatial adaptation. arXiv:1401.2597

[42] Mammen, E. and Tsybakov, A. B. (1995). Asymptotical minimax recovery of sets with smooth boundaries. *Ann. Statist.* **23** 502–524. MR1332579

[43] Mammen, E. and Tsybakov, A. B. (1999). Smooth discrimination analysis. *Ann. Statist.* **27** 1808–1829. MR1765618

[44] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430. MR1425959

[45] Patschkowski, T. and Rohde, A. (2015). Supplement to "Adaptation to lowest density regions with application to support recovery".

[46] Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters - an excess mass approach. *Ann. Statist.* **23** 855–881. MR1345204

[47] Rényi, A. and Sulanke, R. (1963). Über die konvexe Hülle von n zufällig gewählten Punkten. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **2** 75–84. MR0156262

[48] Rényi, A. and Sulanke, R. (1964). Über die konvexe Hülle von n zufällig gewählten Punkten II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **3** 138–147. MR0169139

[49] Reynaud-Bouret, P., Rivoirard, V. and Tuleau-Malot, C. (2011). Adaptive density estimation: A curse of support? *J. Statist. Planning and Inference* **141** 115–139. MR2719482

[50] Rigollet, P. and Vert, R. (2009). Optimal rates for plug-in estimators of density level sets. *Bernoulli* **15** 1154–1178. MR2597587

[51] Rohde, A. (2008). Adaptive goodness-of-fit tests based on signed ranks. *Ann. Statist.* **36** 1346–1374. MR2418660

[52] Rohde, A. (2011). Optimal calibration for multiple testing against local inhomogeneity in higher dimension. *Probab. Theory Relat. Fields* **149** 515–559. MR2776625

[53] Tsybakov, A. B. (1989). Optimal estimation accuracy of nonsmooth images. *Problems of Information Transmission* **25** 180–191. MR1021196

[54] Tsybakov, A. B. (1991). Nonparametric techniques in image estimation. In *Nonparametric Functional Estimation and Related Topics. Proceedings of the NATO Advanced Study Institute, Spetses, Greece* (G. Roussas, ed.) 669–678. Kluwer, Dordrecht. MR1154358

[55] Tsybakov, A. B. (1997). On nonparametric estimation of density level sets. *Ann. Statist.* **25** 948–969. MR1447735

[56] Tsybakov, A. B. (2004). Optimal aggregation of classifiers in statistical learning. *Ann. Statist.* **32** 135–166. MR2051002

[57] Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*, Springer, New York. MR2724359

Fakultät für Mathematik
Ruhr-Universität Bochum
44780 Bochum
Germany
E-mail: tim.patschkowski@ruhr-uni-bochum.de
angelika.rohde@ruhr-uni-bochum.de

SUPPLEMENT TO "ADAPTATION TO LOWEST DENSITY REGIONS WITH APPLICATION TO SUPPORT RECOVERY"

By Tim Patschkowski and Angelika Rohde

Ruhr-Universität Bochum

Supplement A is organized as follows. Section A.1 contains the proofs of Lemma 5.1 – 5.7, which are central ingredients for the proof of Theorem 3.3. Section A.2 is concerned with the remaining proofs of Section 3. Section A.3 contains the proofs of Section 4. Section A.4 introduces a specific construction of a kernel function with prescribed Hölder regularity, which is frequently used throughout the article.

For any $A \subset \mathbb{R}^d$ define

$$d(A,t) := \inf_{y \in A} \|t - y\|_2 \tag{0.1}$$

where $\|\cdot\|_2$ denotes the Euclidean norm (on $\mathbb{R}^d$).

A.1. Proofs of Lemma 5.1 – 5.6.

Proof of Lemma 5.1.

(i) Recall the decomposition

$$p(t+hx) = p(t) + \sum_{i=1}^{d} \Big(p([t,t+hx]_{i-1}) - p([t,t+hx]_i) \Big).$$

It holds for $\beta_i \leq 1$,

$$\Big| p([t,t+hx]_{i-1}) - p([t,t+hx]_i) \Big| \leq L |h_i x_i|^{\beta_i} \tag{0.2}$$

and for $1 < \beta_i \leq 2$

$$\Big| p([t,t+hx]_{i-1}) - p([t,t+hx]_i) + p'_{i,[t,t+hx]_i}(t_i) \cdot h_i x_i \Big| \leq L |h_i x_i|^{\beta_i}.$$

In both cases

$$\begin{aligned} \frac{1}{n}((K_h)^2 * p)(t) &= \frac{1}{n \prod_{i=1}^d h_i} \int K^2(x)\, p(t+hx) d\lambda^d(x) \\ &= \frac{\|K\|_2^2\, p(t)}{n \prod_{i=1}^d h_i} + \frac{1}{n \prod_{i=1}^d h_i} \sum_{i=1}^d S_i \end{aligned} \tag{0.3}$$

with

$$S_i := \int K^2(x)\Big(p([t,t+hx]_{i-1}) - p([t,t+hx]_i)\Big) d\lambda^d(x).$$

For $\beta_i \leq 1$ and

$$h_i \leq \left(\frac{2dL}{\|K\|_2^2}\int |x_i|^{\beta_i} K^2(x)\, d\lambda^d(x)\right)^{-1/\beta_i} p(t)^{1/\beta_i},$$

inequality (0.2) implies

$$\begin{aligned} |S_i| &\leq L\int |x_i|^{\beta_i} K^2(x)\, d\lambda^d(x)\cdot h_i^{\beta_i} \\ &\leq \frac{\|K\|_2^2\, p(t)}{2d}. \end{aligned} \tag{0.4}$$

For $1 < \beta_i \leq 2$, in case of a product kernel (anisotropic smoothness) the i-th factor K_i^2 is of first order again as it remains symmetric. The same holds true for a rotation invariant kernel (isotropic smoothness), because the function $K_{i,x}$ (as defined in (2.1)) is symmetric for every $x \in \mathbb{R}^d$. Hence, the quantity

$$\begin{aligned} |S_i| = \bigg| &\int K^2(x)\left(p([t,t+hx]_{i-1}) - p([t,t+hx]_i) + p'_{i,[t,t+hx]_i}(t_i)\cdot h_i x_i\right) d\lambda^d(x) \\ &- \int K^2(x)\, p'_{i,[t,t+hx]_i}(t_i)\cdot h_i x_i\, d\lambda^d(x)\bigg| \end{aligned}$$

is bounded from above by

$$\begin{aligned} |S_i| &\leq L\int |x_i|^{\beta_i} K^2(x) d\lambda^d(x)\cdot h_i^{\beta_i} \\ &\leq \frac{\|K\|_2^2\, p(t)}{2d}, \end{aligned} \tag{0.5}$$

which proves together with (0.3) the claim.

(ii) We prove the statement for $d = 1$. For higher dimension and product kernel, the result follows by telescoping and Fubini's Theorem. Denote by $\mathcal{H}_1(\beta, L; I)$ the Hölder class of functions from some interval $I \subset \mathbb{R}$ to $\mathbb{R}$ with parameters (β, L). With the previously introduced notation, $\mathcal{H}_1(\beta, L) = \mathcal{H}_1(\beta, L; \mathbb{R})$. The result has been shown in Rohde (2008) for $f \in \mathcal{H}_1(\beta, L; [0,1])$ and $t = \mathrm{argmax}_{x\in[0,1]} |f(x)|$ and is now generalized for arbitrary t. Since the kernel K is continuous on its support with $K(0) > 0$, there exists an

$$\varepsilon \in \left(0, \left(\frac{1}{2L}\right)^{1/\beta} \wedge 1\right],$$

such that $K(x) \geq K(0)/2$ for all $x \in [-\varepsilon, \varepsilon]$. It is sufficient to prove the following statement: For any $f \in \{g \in \mathcal{H}_1(\beta, L) : \|g\|_{\sup} \leq E\}$ for some $E > 0$, and $c_{27} > 0$, there exists a constant $c_{25}(\beta, L) > 0$, such that for every $h \le c_{27}\,|f(t)|^{1/\beta}$, there exists an interval

$$I_t(f,h) \subset J_t(f,h) := [t - \varepsilon h,\, t + \varepsilon h]$$

with

$$\lambda(I_t(f,h)) \ge c_{25}(\beta, L)\, h \tag{0.6}$$

and the property

$$\frac{1}{2}|f(t)| \;\le\; |f(x)| \qquad \text{for every } x \in I_t(f,h). \tag{0.7}$$

This in turn implies

$$\begin{aligned}
\frac{1}{n}((K_h)^2 * |f|)(t) &\ge \frac{1}{nh^2}\int_{J_t(f,h)} K^2\left(\frac{t-x}{h}\right)|f(x)|d\lambda(x)\\
&\ge \frac{K^2(0)}{4}\cdot\frac{1}{nh^2}\int_{J_t(f,h)}|f(x)|d\lambda(x)\\
&\ge \frac{K^2(0)}{4}\cdot\frac{1}{nh^2}\int_{I_t(f,h)}|f(x)|d\lambda(x)\\
&\ge \frac{K^2(0)}{8}\cdot\frac{|f(t)|}{nh^2}\cdot\lambda(I_t(f,h))\\
&\ge \frac{K^2(0)\,c_{25}(\beta,L)}{8}\cdot\frac{|f(t)|}{nh}.
\end{aligned}$$

It remains to prove the existence of such an interval $I_t(f,h)$ with properties (0.6) and (0.7). For $\beta \le 1$, choose $I_t(f,h) = J_t(f,(c_{27}^{-1}\wedge 1)h) \subset J_t(f,h)$. For $\beta > 1$, we consider the rescaled function

$$u_f(x) := \frac{f\left(t + (c_{27}^{-1}\wedge 1)\varepsilon h x\right)}{f(t)}, \quad x \in [-1,1],$$

which is contained in $\mathcal{H}_1(\beta, L; [-1,1])$ with $\|u_f\|_{\sup} \ge 1$. Taylor expansion around any point $y \in [-1,1]$ provides the approximation

$$u_f(x) = P^{(u_f)}_{y,\lfloor\beta\rfloor}(x) + R_{u_f}(x,y)$$

with a remainder term $|R_{u_f}(x,y)| \le L\,|x-y|^\beta$. Hence,

$$\begin{aligned}
\left|(x-y)\,u_f'(y) + \ldots + \frac{(x-y)^{\lfloor\beta\rfloor}}{\lfloor\beta\rfloor!}\,u_f^{(\lfloor\beta\rfloor)}(y)\right| &\le 2\|u_f\|_{\sup} + L\,|x-y|^\beta\\
&\le 2\|u_f\|_{\sup} + 2^\beta L.
\end{aligned} \tag{0.8}$$

For any polynomial $P(x) = \sum_{k=1}^D a_k x^k$ of degree D the norms

$$\|P\|_{(1)} = \sup_{x\in[-1,1]}|P(x)| \qquad \text{and} \qquad \|P\|_{(2)} = \max_{0\le k\le D}|a_k|$$

are two norms on the $(D+1)$-dimensional space of polynomials on $[-1,1]$ of degree D, and these norms are equivalent. Consequently, there exists a constant $C_{D,[-1,1]}$ depending on the degree D and on the interval $[-1,1]$, such that $\|P\|_{(2)} \leq C_{D,[-1,1]}\|P\|_{(1)}$. In particular, there exists a constant $C = C_{\beta,[-1,1]} \geq 1$ such that u_f' is in view of (0.8) uniformly bounded by $C(2\|u_f\|_{\sup} + 2^\beta L)$, and hence, by the mean value theorem,

$$|u_f(x) - u_f(y)| \leq \|u_f'\|_{\sup} \cdot |x-y| \leq C(2\|u_f\|_{\sup} + 2^\beta L) \cdot |x-y|$$

for all $x, y \in [-1,1]$. Denoting $x_0 := \operatorname{argmax}_{x\in[-1,1]} |u_f(x)|$, then

$$|u_f(x) - u_f(x_0)| \leq \frac{1}{2}|u_f(x_0)| \tag{0.9}$$

for

$$x \in \left[x_0 - \frac{\|u_f\|_{\sup}}{4C\|u_f\|_{\sup} + 2^{\beta+1}CL},\, x_0 + \frac{\|u_f\|_{\sup}}{4C\|u_f\|_{\sup} + 2^{\beta+1}CL}\right] \cap [-1,1].$$

Due to $\|u_f\|_{\sup} \geq 1$, inequality (0.9) holds also true on

$$\left[x_0 - \frac{1}{4C + 2^{\beta+1}CL},\, x_0 + \frac{1}{4C + 2^{\beta+1}CL}\right] \cap [-1,1].$$

Since $C \geq 1$, we assume without loss of generality that $[x_0,\, x_0 + 1/(4C + 2^{\beta+1}CL)]$ is fully contained in $[-1,1]$. By the triangle inequality and (0.9),

$$|f(z)| \geq \frac{1}{2}\left|f(t + (c_{27}^{-1} \wedge 1)\varepsilon h x_0)\right|$$

for all $z \in I_t(f,h)$, where

$$I_t(f,h) := \left[t + (c_{27}^{-1} \wedge 1)\varepsilon h x_0,\, t + (c_{27}^{-1} \wedge 1)\varepsilon h\left(x_0 + \frac{1}{4C + 2^{\beta+1}CL}\right)\right] \subset J_t(f,h).$$

Because $|u_f(x_0)| \geq |u_f(0)|$ and consequently $|f(t + \varepsilon h x_0)| \geq |f(t)|$, the result follows.

The result of Rohde (2008) is established for isotropic smoothness and rotation invariant kernel in Rohde (2011). Our result analogously extends to isotropic smoothness and rotation invariant kernel following the previous steps.

(iii) It holds

$$\begin{aligned}\frac{1}{n}((K_h)^2 * p)(t) &= \frac{1}{nh^{2d}} \int K^2\left(\frac{t-x}{h}\right) p(x) d\lambda^d(x) \\ &\leq \frac{\|K\|_2^2}{nh^d} \cdot \sup_{x \in \mathcal{B}_h(t)} p(x),\end{aligned}$$

where $\mathcal{B}_h(t)$ denotes the closed Euclidean ball with radius h around t. We have for any x with $\|x-t\|_2 \le h$

$$d(\Gamma_p^c, x) \le h + d(\Gamma_p^c, t).$$

Since $P^{(p)}_{y,\lfloor\beta\rfloor} = 0$ for any $y \in (\Gamma_p)^c$, and therefore

$$p(x) = p(x) - P^{(p)}_{y,\lfloor\beta\rfloor}(x), \qquad \text{for any } x \in \Gamma_p$$

it follows that

$$\sup_{x\in\mathcal{B}_h(t)} p(x) \le L\, d(\Gamma_p^c, x)^\beta \le L\left(h + d(\Gamma_p^c, t)\right)^\beta.$$

□

Proof of Lemma 5.2. We define

$$J_1 := \{i \in \{1,\dots,d\} : m_i > j_i\} := \{i_1,\dots,i_s\}$$
$$J_2 := \{1,\dots,d\} \setminus J_1.$$

With $x_1 := (x_i, i \in J_1)$ and $x_2 := (x_i, i \in J_2)$, $h_k = (2^{-k_1},\dots,2^{-k_d})$ and $K^2_{i,h_{k,i}}(\cdot) = h_{k,i}^{-1} K_i^2(\cdot/h_{k,i})$ for $k = j, m$, and

$$M_{J_2}^{t_2}(x_1) := \begin{cases} \displaystyle\int \prod_{i\in J_2} K^2_{i,h_{m,i}}(t_i - x_i)\, p(x)\, d\lambda^{d-s}(x_2) & \text{, if } s < d \\ p(x) & \text{, if } s = d, \end{cases}$$

we have the representation

$$\sigma_t^2(j\wedge m) = \frac{1}{n\prod_{i=1}^d (h_{j,i}\vee h_{m,i})} \int \prod_{i\in J_1} K^2_{i,h_{j,i}}(t_i - x_i)\, M_{J_2}^{t_2}(x_1)\, d\lambda^s(x_1), \tag{0.10}$$
$$\sigma_t^2(m) = \frac{1}{n\prod_{i=1}^d h_{m,i}} \int \prod_{i\in J_1} K^2_{i,h_{m,i}}(t_i - x_i)\, M_{J_2}^{t_2}(x_1)\, d\lambda^s(x_1).$$

Note that $M_{J_2}^{t_2}(\cdot) \in \mathcal{H}_s(\beta_{J_1}, L\, c_{J_2})$, where

$$c_{J_2} := \begin{cases} \|\prod_{i\in J_2} K_i\|_{\sup} & \text{, if } J_2 \ne \emptyset \\ 1 & \text{, if } J_2 = \emptyset \end{cases}$$

and $\beta_{J_1} = (\beta_i)_{i\in J_1}$. If h_j satisfies

$$h_{j,i} \le c_{11}(\beta_{J_1}, L\, c_{J_2}) \cdot M_{J_2}^{t_2}(t_1)^{\frac{1}{\beta_i}} \quad \text{for all } i \in J_1,$$

then Lemma 5.1 (i) yields

$$\frac{1}{\prod_{i\in J_1} h_{j,i}} \int \prod_{i\in J_1} K^2_{i,h_{j,i}}(t_i - x_i)\, M_{J_2}^{t_2}(x_1)\, d\lambda^s(x_1) \tag{0.11}$$

$$\leq \frac{3}{2} \cdot \frac{\| \prod_{i \in J_1} K_i \|_2^2}{\prod_{i \in J_1} h_{j,i}} M_{J_2}^{t_2}(t_1)$$
$$\leq \frac{3}{2} \cdot \frac{\| \prod_{i \in J_1} K_i \|_2^2}{\prod_{i \in J_1} h_{m,i}} M_{J_2}^{t_2}(t_1)$$
$$\leq 3 \cdot \frac{1}{\prod_{i \in J_1} h_{m,i}} \int \prod_{i \in J_1} K_{i,h_{m,i}}^2 (t_i - x_i) \, M_{J_2}^{t_2}(x_1) \, d\lambda^s(x_1),$$

which is equivalent to

$$\sigma_t^2(j \wedge m) \leq 3\sigma_t^2(m).$$

Due to the monotonicity of the truncation level in the product of the bandwidth's components, this implies

$$\sigma_{t,\mathrm{trunc}}^2(j \wedge m) \leq 3\sigma_{t,\mathrm{trunc}}^2(m).$$

If there exists an index $l \in J_1$ with

$$h_{j,l} > c_{11}(\beta_{J_1}, L\, c_{J_2}) \cdot M_{J_2}^{t_2}(t_1)^{\frac{1}{\beta_l}},$$

we obtain that

$$M_{J_2}^{t_2}(t_1) \leq c_6(\beta, L) \left\{ \left(\frac{\log n}{n} \right)^{\frac{\bar{\beta}}{\bar{\beta}+1}} \vee \left(\frac{p(t) \log n}{n} \right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}} \right\} \tag{0.12}$$

with

$$c_6(\beta, L) := \max_{J \subset \{1,\ldots,d\}} \max_{l \in J} \left(\frac{c_{11}(\beta, L)}{c_{11}(\beta_J, L\, c_{J_2})} \right)^{\beta_l},$$

since $j \geq \bar{j}$ componentwise. The maximum in (0.12) is attained by its right hand side term if and only if $p(t) \geq (\log n / n)^{\bar{\beta}/(\bar{\beta}+1)}$ in which case

$$\left(\frac{p(t) \log n}{n} \right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}} \leq p(t), \tag{0.13}$$

whence

$$M_{J_2}^{t_2}(t_1) \leq c_6(\beta, L) \left\{ \left(\frac{\log n}{n} \right)^{\frac{\bar{\beta}}{\bar{\beta}+1}} \vee p(t) \right\}.$$

With (0.10), we obtain by the same arguments as in (0.3), (0.4) and (0.5) applied to $M_{J_2}^{t_2}(\cdot) \in \mathcal{H}_s(\beta_{J_1}, L\, c_{J_2})$ as well as $j \geq \bar{j}$ componentwise,

$$\sigma_t^2(j \wedge m) \leq \frac{c(\beta, L)}{n \prod_{i=1}^d (h_{j,i} \vee h_{m,i})} \left\{ \left(\frac{\log n}{n} \right)^{\frac{\bar{\beta}}{\bar{\beta}+1}} \vee p(t) \right\}$$

$$\leq c(\beta,L)\begin{cases}\dfrac{\log^2 n}{n^2\prod_{i=1}^d \bar{h}_i\prod_{i=1}^d h_{j,i}} & \text{, if } \ p(t)\leq\left(\frac{\log n}{n}\right)^{\frac{\bar\beta}{\bar\beta+1}}\\ \dfrac{p(t)}{n\prod_{i=1}^d h_{j,i}} & \text{, if } \ p(t)>\left(\frac{\log n}{n}\right)^{\frac{\bar\beta}{\bar\beta+1}}\end{cases}$$

$$\leq c(\beta,L)\begin{cases}\dfrac{\log^2 n}{n^2\prod_{i=1}^d h_{j,i}^2} & \text{, if } \ p(t)\leq\left(\frac{\log n}{n}\right)^{\frac{\bar\beta}{\bar\beta+1}}\\ \dfrac{p(t)}{n\prod_{i=1}^d h_{j,i}} & \text{, if } \ p(t)>\left(\frac{\log n}{n}\right)^{\frac{\bar\beta}{\bar\beta+1}}.\end{cases}$$

Since $h_{j,i}\leq\bar{h}_{i,j}\leq c_{11}(\beta,L)\,p(t)^{1/\beta_i}$, $i=1,\dots,d$ whenever $p(t)>(\log n/n)^{\bar\beta/(\bar\beta+1)}$ due to (0.13), Lemma 5.1(i) can be applied and yields

$$\sigma_t^2(j\wedge m)\leq c(\beta,L)\,\sigma^2_{t,\mathrm{trunc}}(j).$$

Finally, by monotonicity of the truncation level in the product of the bandwidth's components,

$$\sigma^2_{t,\mathrm{trunc}}(j\wedge m)\leq c(\beta,L)\,\sigma^2_{t,\mathrm{trunc}}(j),$$

which proves the first statement of Lemma 5.2. As concerns the second claim, assume now that $m\geq j$ and $j\geq\bar{j}$ componentwise. We distinguish between the two cases of a truncated and non-truncated reference bandwidth $\bar{h}$, i.e.

$$\bar{h}_i=c_{11}(\beta,L)\left(\frac{\log n}{n}\right)^{\frac{\bar\beta}{\bar\beta+1}\frac{1}{\beta_i}}\quad\text{ and }\quad\bar{h}_i=c_{11}(\beta,L)\left(\frac{p(t)\log n}{n}\right)^{\frac{\bar\beta}{2\bar\beta+1}\frac{1}{\beta_i}}$$

for all $i=1,\dots,d$, respectively. If $\bar{h}$ is non-truncated, that is

$$\left(\frac{p(t)\log n}{n}\right)^{\frac{\bar\beta}{2\bar\beta+1}}\ \geq\left(\frac{\log n}{n}\right)^{\frac{\bar\beta}{\bar\beta+1}},$$

we obtain $p(t)\geq(\log n/n)^{\bar\beta/(\bar\beta+1)}$ and therefore $\bar{h}_i\leq c_{11}(\beta,L)\,p(t)^{1/\beta_i}$ for all $i=1,\dots,d$. Consequently, by Lemma 5.1 (i),

$$\frac{1}{n}((K_{\bar{h}})^2*p)(t)\ \leq\ 3\frac{1}{n}((K_{h_j})^2*p)(t)\ \leq\ 9\frac{1}{n}((K_{h_m})^2*p)(t).$$

By the monotonicity of the truncation level the claim follows for non-truncated $\bar{h}$. If $\bar{h}$ is truncated, that is

$$\left(\frac{p(t)\log n}{n}\right)^{\frac{\bar\beta}{2\bar\beta+1}}\ <\left(\frac{\log n}{n}\right)^{\frac{\bar\beta}{\bar\beta+1}},$$

we have $p(t)\leq(\log n/n)^{\bar\beta/(\bar\beta+1)}$. Thus, following the steps in (0.3), (0.4) and (0.5), for any $h\leq\bar{h}$ componentwise,

$$\frac{1}{n}((K_h)^2*p)(t)\leq\frac{\|K\|_2^2p(t)}{n\prod_{i=1}^d h_i}+\frac{c(\beta,L)}{n\prod_{i=1}^d h_i}\sum_{i=1}^d h_i^{\beta_i}\leq c(\beta,L)\frac{\log^2 n}{n^2\prod_{i=1}^d h_i^2}$$

and therefore
$$\frac{\log^2 n}{n^2\prod_{i=1}^d h_i^2} \le \sigma^2_{t,\mathrm{trunc}}(h) \le c(\beta,L)\frac{\log^2 n}{n^2\prod_{i=1}^d h_i^2},$$
where the left and right hand side are monotone in the product of bandwidth components. $\square$

Proof of Lemma 5.3. The proof is based on Bernstein's inequality. First,
$$\mathbb{P}^{\otimes n}\Big(|\tilde\sigma^2_{t,\mathrm{trunc}}(h) - \sigma^2_{t,\mathrm{trunc}}(h)| \ge \eta\,\sigma^2_{t,\mathrm{trunc}}(h)\Big)$$
$$\le\ \mathbb{P}^{\otimes n}\Big(|\tilde\sigma^2_t(h) - \sigma^2_t(h)| \ge \eta\,\sigma^2_{t,\mathrm{trunc}}(h)\Big).$$

The random variable $\tilde\sigma^2_t(h) - \sigma^2_t(h)$ can be rewritten as a sum of centered and independent random variables
$$Z_k := \frac{1}{n^2\prod_{i=1}^d h_i^2}\left(K^2\left(\frac{t-X_k}{h}\right) - \mathbb{E}_p K^2\left(\frac{t-X_k}{h}\right)\right)$$
with the properties
$$|Z_k| \le \frac{2\|K\|^2_{\sup}}{n^2\prod_{i=1}^d h_i^2}$$
and
$$\sum_{k=1}^n \mathrm{Var}\left(Z_k\right) \le \frac{1}{n^3\prod_{i=1}^d h_i^4}\mathbb{E}_p K^4\left(\frac{t-X_1}{h}\right) \le \frac{\|K\|^2_{\sup}}{n^2\prod_{i=1}^d h_i^2}\sigma^2_{t,\mathrm{trunc}}(h).$$
Hence, Bernstein's inequality yields the following exponential tail bound
$$\mathbb{P}^{\otimes n}\left(|\tilde\sigma^2_t(h) - \sigma^2_t(h)| \ge \eta\,\sigma^2_{t,\mathrm{trunc}}(h)\right)$$
$$\le 2\exp\left(-\frac{1}{2}\frac{\eta^2\sigma^4_{t,\mathrm{trunc}}(h)}{\frac{\|K\|^2_{\sup}\sigma^2_{t,\mathrm{trunc}}(h)}{n^2\prod_{i=1}^d h_i^2}\left(1+\frac{2\eta}{3}\right)}\right) \tag{0.14}$$
$$\le 2\exp\left(-\frac{3\eta^2}{2(3+2\eta)\|K\|^2_{\sup}}\log^2 n\right).$$
$\square$

Proof of Lemma 5.4. The inequalities are proven separately and both the proofs distinguish between the cases $p(t) \le (\log n/n)^{\bar\beta/(\bar\beta+1)}$ and $p(t) > (\log n/n)^{\bar\beta/(\bar\beta+1)}$.

Proof of (5.4): Recall the definition of the reference bandwidth in (5.1), which for $p(t) \le (\log n/n)^{\bar\beta/(\bar\beta+1)}$ is equal to
$$\bar h_i = c_{11}(\beta,L)\left(\frac{\log n}{n}\right)^{\frac{\bar\beta}{\bar\beta+1}\frac{1}{\beta_i}}, \quad i=1,\dots,d. \tag{0.15}$$

The corresponding truncation level satisfies

$$\frac{\log^2 n}{n^2 \prod_{i=1}^d \bar{h}_i^2} = c_{11}(\beta, L)^{-2d} \left(\frac{\log n}{n}\right)^{\frac{2\bar{\beta}}{\bar{\beta}+1}}. \tag{0.16}$$

Consequently,

$$\begin{aligned} B_t(\bar{j}) &\le B_t(\bar{h}) \\ &= \left(L \sum_{i=1}^d c_{12,i}(\beta)\, c_{11}(\beta, L)^{\beta_i}\right) \left(\frac{\log n}{n}\right)^{\frac{\bar{\beta}}{\bar{\beta}+1}} \\ &= \left(L \sum_{i=1}^d c_{12,i}(\beta)\, c_{11}(\beta, L)^{\beta_i}\right) c_{11}(\beta, L)^d \left(\frac{\log^2 n}{n^2 \prod_{i=1}^d \bar{h}_i^2}\right)^{1/2} \\ &\le \left(L \sum_{i=1}^d c_{12,i}(\beta)\, c_{11}(\beta, L)^{\beta_i}\right) c_{11}(\beta, L)^d \sqrt{\sigma_{t,\mathrm{trunc}}^2(\bar{j})}. \end{aligned}$$

For $p(t) > (\log n/n)^{\bar{\beta}/(\bar{\beta}+1)}$, the reference bandwidth $\bar{h}$ is defined as

$$\bar{h}_i = c_{11}(\beta, L) \left(\frac{p(t) \log n}{n}\right)^{\frac{\bar{\beta}}{2\bar{\beta}+1} \frac{1}{\beta_i}}, \quad i = 1, \ldots, d, \tag{0.17}$$

and therefore

$$\begin{aligned} B_t(\bar{j}) &\le \left(L \sum_{i=1}^d c_{12,i}(\beta)\, c_{11}(\beta, L)^{\beta_i}\right) \left(\frac{p(t) \log n}{n}\right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}} \\ &= \left(L \sum_{i=1}^d c_{12,i}(\beta)\, c_{11}(\beta, L)^{\beta_i}\right) c_{11}(\beta, L)^{d/2} \cdot \left(\frac{p(t) \log n}{n \prod_{i=1}^d \bar{h}_i}\right)^{1/2} \\ &\le \left(L \sum_{i=1}^d c_{12,i}(\beta)\, c_{11}(\beta, L)^{\beta_i}\right) c_{11}(\beta, L)^{d/2} \cdot \left(\frac{p(t) \log n}{n 2^{-|\bar{j}|}}\right)^{1/2}. \end{aligned} \tag{0.18}$$

Since for $p(t) \ge (\log n/n)^{\bar{\beta}/(\bar{\beta}+1)}$,

$$2^{-\bar{j}_i} \le \bar{h}_i \le c_{11}(\beta, L)\, p(t)^{1/\beta_i} \quad \text{for all } i = 1, \ldots, d, \tag{0.19}$$

Lemma 5.1 (i) yields

$$\frac{p(t) \log n}{n 2^{-|\bar{j}|}} \le \frac{2}{\|K\|_2^2} \cdot \sigma_t^2(\bar{j}) \log n$$

and together with (0.18)

$$B_t(\bar{j}) \le \left(L \sum_{i=1}^d c_{12,i}(\beta)\, c_{11}(\beta, L)^{\beta_i}\right) \left(\frac{2\, c_{11}(\beta, L)^d}{\|K\|_2^2}\right)^{1/2} \sqrt{\sigma_{t,\mathrm{trunc}}^2(\bar{j}) \log n}.$$

Proof of (5.5): For $p(t) \leq (\log n/n)^{\bar\beta/(\bar\beta+1)}$, the reference bandwidth $\bar h$ is given by (0.15). Hence by (0.16),

$$\frac{\log^2 n}{n^2(2^{-|\bar j|})^2} \leq 2^{2d}\frac{\log^2 n}{n^2\prod_{i=1}^d \bar h_i^2} = \left(\frac{2}{c_{11}(\beta,L)}\right)^{2d}\left(\frac{\log n}{n}\right)^{\frac{2\bar\beta}{\bar\beta+1}}. \tag{0.20}$$

Furthermore, by (0.3), (0.4) and (0.5),

$$\begin{aligned}
\sigma_t^2(\bar j) &\leq \frac{\|K\|_2^2\, p(t)}{n2^{-|\bar j|}} + \frac{1}{n2^{-|\bar j|}}\sum_{i=1}^d L\int |x_i|^{\beta_i}K^2(x)d\lambda^d(x)\cdot\left(2^{-\bar j_i}\right)^{\beta_i}\\
&\leq c(\beta,L)\cdot\frac{\log n}{n\prod_{i=1}^d \bar h_i}\left(\frac{\log n}{n}\right)^{\frac{\bar\beta}{\bar\beta+1}}\\
&= c(\beta,L)\cdot\left(\frac{\log n}{n}\right)^{\frac{2\bar\beta}{\bar\beta+1}}
\end{aligned}$$

and finally for $p(t) \leq (\log n/n)^{\bar\beta/(\bar\beta+1)}$,

$$\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar j)} \;\leq\; c(\beta,L)\left(\frac{\log n}{n}\right)^{\frac{\bar\beta}{\bar\beta+1}}. \tag{0.21}$$

For $p(t) > (\log n/n)^{\bar\beta/(\bar\beta+1)}$, the reference bandwidth $\bar h$ is given by (0.17) and hence

$$\begin{aligned}
\frac{\log^2 n}{n^2\left(2^{-|\bar j|}\right)^2} &\leq 2^{2d}c_{11}(\beta,L)^{-d}\cdot\frac{\log^2 n}{n^2\prod_{i=1}^d \bar h_i}\left(\frac{p(t)\log n}{n}\right)^{-\frac{1}{2\bar\beta+1}}\\
&\leq 2^{2d}c_{11}(\beta,L)^{-d}\cdot\frac{\log n}{n\prod_{i=1}^d \bar h_i}\left(\frac{\log n}{n}\right)^{\frac{\bar\beta}{\bar\beta+1}}\\
&\leq 2^{2d}c_{11}(\beta,L)^{-d}\cdot\frac{p(t)\log n}{n\prod_{i=1}^d \bar h_i}\\
&= 2^{2d}c_{11}(\beta,L)^{-2d}\cdot\left(\frac{p(t)\log n}{n}\right)^{\frac{2\bar\beta}{2\bar\beta+1}}.
\end{aligned}$$

Furthermore, since $\bar j$ satisfies property (0.19), Lemma 5.1 (i) reveals

$$\sigma_t^2(\bar j) \leq \frac{3}{2}\,\frac{\|K\|_2^2\, p(t)}{n2^{-|\bar j|}} = 3\cdot 2^d\|K\|_2^2\cdot c_{11}(\beta,L)^{-d}\cdot\left(\frac{p(t)\log n}{n}\right)^{\frac{2\bar\beta}{2\bar\beta+1}},$$

such that together with (0.21)

$$\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar j)} \;\leq c(\beta,L)\cdot\begin{cases}\left(\frac{\log n}{n}\right)^{\frac{\bar\beta}{\bar\beta+1}} & \text{, if } \; p(t)\leq\left(\frac{\log n}{n}\right)^{\frac{\bar\beta}{\bar\beta+1}}\\[2ex] \left(\frac{p(t)\log n}{n}\right)^{\frac{\bar\beta}{2\bar\beta+1}} & \text{, if } \; p(t)>\left(\frac{\log n}{n}\right)^{\frac{\bar\beta}{\bar\beta+1}}\end{cases}$$

$$= c(\beta, L) \cdot \left\{ \left(\frac{\log n}{n} \right)^{\frac{\bar{\beta}}{\bar{\beta}+1}} \vee \left(\frac{p(t) \log n}{n} \right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}} \right\}.$$

$\square$

Proof of Lemma 5.5. Observe first that Y can be expressed as a sum of centered and independent random variables $Y = \sum_{i=1}^n Y_i$ with

$$Y_i := \frac{\frac{1}{n 2^{-|j|}} \left(K\left(\frac{t_1 - X_{i,1}}{2^{-j_1}}, \dots, \frac{t_d - X_{i,d}}{2^{-j_d}} \right) - \mathbb{E}_p K\left(\frac{t_1 - X_{i,1}}{2^{-j_1}}, \dots, \frac{t_d - X_{i,d}}{2^{-j_d}} \right) \right)}{\sqrt{\sigma^2_{t,\mathrm{trunc}}(j) \log n}}.$$

For $n \geq n_0$ with n_0 depending on $\|K\|_{\sup}$ only, it holds

$$\frac{1}{3}|Y_i| \leq \frac{2\|K\|_{\sup}}{3\sqrt{\log^3 n}} \leq \frac{1}{\log n} \quad \text{and} \quad \sum_{i=1}^n \mathrm{Var}(Y_i) \leq \frac{1}{\log n}.$$

Bernstein's inequality yields

$$\mathbb{P}^{\otimes n}(|Y| \geq \eta) \;\leq\; 2 \exp\left(-\frac{1}{2} \frac{\eta^2}{1+\eta} \log n \right),$$

leading to subgaussian and subexponential tail behavior for $\eta \leq 1$ and $\eta > 1$, respectively. $\square$

Proof of Lemma 5.6. Fubini's theorem and the classical moment bound for the exponential distribution reveal

$$\begin{aligned}
\mathbb{E} Z^m &= \int_0^\infty x^m p_Z(x) d\lambda(x) \\
&= \int_0^\infty \int_0^x m t^{m-1} d\lambda(t)\, p_Z(x) d\lambda(x) \\
&= \int_0^\infty m t^{m-1} \mathbb{P}(Z \geq t) d\lambda(t) \\
&\leq 2m \int_0^\infty t^{m-1} \exp(-At) d\lambda(t) \\
&\leq 2m \frac{(m-1)!}{A^m}.
\end{aligned}$$

$\square$

A.2. Remaining proofs of Section 3.

Proof of Theorem 3.1. The construction of the hypotheses requires functions $K_1 \in \mathscr{P}_d(\beta, L')$ and $K_2 \in \mathscr{P}_d(\beta, L - L')$, integrating to one and compactly supported within a rectangle, say $\prod_{i=1}^{d}[-g_{1,i}, g_{1,i}]$ and $\prod_{i=1}^{d}[-g_{2,i}, g_{2,i}]$, respectively, with $K_1(0) = \sqrt{3/4} \cdot c_1(\beta, L')$ and $L' < L$ chosen such that $c_1(\beta, L') > \sqrt{3/4} \cdot c_1(\beta, L)$. The auxiliary constant L' is introduced to permit the construction of perturbed hypotheses in $\mathscr{P}_d(\beta, L)$ with value larger than $3/4 \cdot c_1(\beta, L)$ at the point t. First observe that

$$
\begin{aligned}
&\inf_{0<\delta\le c_1(\beta,L)} \inf_{T_n(t)} \sup_{\substack{p\in\mathscr{P}_d(\beta,L)\\ \delta/2\le p(t)\le\delta}} \mathbb{E}_p^{\otimes n}\left(\frac{|T_n(t)-p(t)|}{\psi^n_{p(t),\beta}}\right)^r \\
&\quad = \inf_{0<\delta\le c_1(\beta,L)/K_2(0)} \inf_{T_n(t)} \sup_{\substack{p\in\mathscr{P}_d(\beta,L)\\ \delta/2\le p(t)/K_2(0)\le\delta}} \mathbb{E}_p^{\otimes n}\left(\frac{|T_n(t)-p(t)|}{\psi^n_{p(t),\beta}}\right)^r \\
&\quad = \min\Bigg\{ \inf_{\delta\le n^{-\bar\beta/(\bar\beta+1)}} \inf_{T_n(t)} \sup_{\substack{p\in\mathscr{P}_d(\beta,L)\\ \delta/2\le p(t)/K_2(0)\le\delta}} \mathbb{E}_p^{\otimes n}\left(\frac{|T_n(t)-p(t)|}{\psi^n_{p(t),\beta}}\right)^r, \\
&\qquad \inf_{n^{-\bar\beta/(\bar\beta+1)}<\delta\le c_1(\beta,L)/K_2(0)} \inf_{T_n(t)} \sup_{\substack{p\in\mathscr{P}_d(\beta,L)\\ \delta/2\le p(t)/K_2(0)\le\delta}} \mathbb{E}_p^{\otimes n}\left(\frac{|T_n(t)-p(t)|}{\psi^n_{p(t),\beta}}\right)^r \Bigg\}.
\end{aligned}
$$

The two situations

$$
(0.22) \qquad (i)\ \ \delta \le n^{-\frac{\bar\beta}{\bar\beta+1}}, \qquad (ii)\ \ n^{-\frac{\bar\beta}{\bar\beta+1}} < \delta \le c_1(\beta,L)/K_2(0)
$$

are analyzed separately. In case (i), for any $\kappa_1 > 0$ Markov's inequality yields

$$
\begin{aligned}
&\inf_{\delta\le n^{-\bar\beta/(\bar\beta+1)}} \inf_{T_n(t)} \sup_{\substack{p\in\mathscr{P}_d(\beta,L)\\ \delta/2\le p(t)/K_2(0)\le\delta}} \mathbb{E}_p^{\otimes n}\left(\frac{|T_n(t)-p(t)|}{\psi^n_{p(t),\beta}}\right)^r \\
&\qquad \ge \inf_{\delta\le n^{-\bar\beta/(\bar\beta+1)}} \inf_{T_n(t)} \sup_{\substack{p\in\mathscr{P}_d(\beta,L)\\ \delta/2\le p(t)/K_2(0)\le\delta}} \mathbb{E}_p^{\otimes n}\left(\frac{|T_n(t)-p(t)|}{\delta K_2(0)}\right)^r \\
&\qquad \ge \inf_{\delta\le n^{-\bar\beta/(\bar\beta+1)}} \inf_{T_n(t)} \sup_{\substack{p\in\mathscr{P}_d(\beta,L)\\ \delta/2\le p(t)/K_2(0)\le\delta}} \kappa_1^r \cdot \mathbb{P}^{\otimes n}\Big(|T_n(t)-p(t)| \ge \kappa_1 \cdot \delta K_2(0)\Big).
\end{aligned}
$$

Denote by h_n the multidimensional bandwidth with components

$$
h_{n,i} := \delta^{\frac{1}{\beta_i}}, \quad i = 1, \dots, d,
$$

chosen in a manner such that

$$
K_2(x; h_n) := \left(\prod_{i=1}^{d} h_{n,i}\right)^{\bar\beta} K_2\left(\frac{x}{h_n}\right)
$$

attains the value $\delta \cdot K_2(0)$ at the point 0. Setting $s_n := (t_1 + h_{n,1} + g_{1,1}, t_2, \ldots, t_d)$, we define the hypotheses

$$\begin{aligned} p_{0,n}(x) &:= K_1(x - s_n) + \frac{1}{2}\Big(K_2(x-t;h_n) - K_2(x-s_n;h_n)\Big) \\ p_{1,n}(x) &:= K_1(x - s_n) + K_2(x-t;h_n) - K_2(x-s_n;h_n). \end{aligned}$$

Both hypotheses $p_{0,n}$ and $p_{1,n}$ have anisotropic Hölder smoothness with parameters (β, L), since

$$\left(\prod_{i=1}^{d} h_{n,i}\right)^{\bar\beta} = h_{n,i}^{\beta_i} \quad \text{for all } i = 1, \ldots, d$$

and $L' + (L - L') = L$. Moreover, they integrate to one, are positive for sufficiently large $n \geq n_0(\beta, L)$ and attain the values $p_{0,n}(t) = \delta \cdot K_2(0)/2$ and $p_{1,n}(t) = \delta \cdot K_2(0)$. The absolute distance in t equals

$$|p_{0,n}(t) - p_{1,n}(t)| = \frac{\delta \cdot K_2(0)}{2}.$$

It remains to bound the distance between the associated product probability measures $\mathbb{P}_{0,n}^{\otimes n}$ and $\mathbb{P}_{1,n}^{\otimes n}$. The squared Hellinger distance is bounded from above by 2, so Bernoulli's inequality yields the upper bound

$$H^2(\mathbb{P}_{0,n}^{\otimes n}, \mathbb{P}_{1,n}^{\otimes n}) = 2\left(1 - \left(1 - \frac{H^2(\mathbb{P}_{0,n}, \mathbb{P}_{1,n})}{2}\right)^n\right) \leq n\, H^2(\mathbb{P}_{0,n}, \mathbb{P}_{1,n}),$$

which in turn is bounded by

$$\begin{aligned} & n \int \left(\sqrt{K_2(x-t;h_n)/2} - \sqrt{K_2(x-t;h_n)}\,\right)^2 d\lambda^d(x) \\ & \qquad + n \int \Big(\sqrt{K_1(x-s_n) - K_2(x-s_n;h_n)/2} \\ & \qquad\qquad - \sqrt{K_1(x-s_n) - K_2(x-s_n;h_n)}\,\Big)^2 d\lambda^d(x) \\ \leq\ & n \int K_2(x;h_n)\, d\lambda^d(x) \\ =\ & n\, \delta^{\frac{\bar\beta+1}{\bar\beta}}, \end{aligned}$$

where the inequality is due to

$$\left(\sqrt{x - y/2} - \sqrt{x-y}\right)^2 \leq \frac{y}{2} \quad \text{for all } 0 \leq y \leq x. \tag{0.23}$$

The last expression is bounded by 1 as $\delta \leq n^{-\frac{\bar\beta}{\bar\beta+1}}$. Finally, by Theorem 2.2 (Tsybakov 2009) (Hellinger version) with $\kappa_1 = 1/4$, we arrive for $n \geq n_0(\beta, L)$ at

$$\inf_{\delta \leq n^{-\bar\beta/(\bar\beta+1)}} \inf_{T_n(t)} \sup_{\substack{p \in \mathscr{P}_d(\beta,L) \\ \delta/2 \leq p(t)/K_2(0) \leq \delta}} \mathbb{E}_p^{\otimes n}\left(\frac{|T_n(t) - p(t)|}{\psi_{p(t),\beta}^n}\right)^r \geq \frac{4^{-r}}{2}\left(1 - \sqrt{\frac{3}{4}}\right) > 0.$$

In case (0.22) (ii) the hypotheses have to be chosen in a different way. To this aim, the interval

$$\left(n^{-\bar\beta/(\bar\beta+1)},\, c_1(\beta,L)/K_2(0)\right]$$

is decomposed again into

$$I_1 := \left(n^{-\bar\beta/(\bar\beta+1)},\, c_7(\beta,L)\right] \quad \text{and} \quad I_2 := \left(c_7(\beta,L),\, c_1(\beta,L)/K_2(0)\right]$$

with a constant $c_7(\beta,L)$ specified later. Since

$$\begin{aligned}
&\inf_{n^{-\bar\beta/(\bar\beta+1)}<\delta\le c_1(\beta,L)/K_2(0)} \inf_{T_n(t)} \sup_{\substack{p\in\mathscr{P}_d(\beta,L)\\ \delta/2\le p(t)/K_2(0)\le\delta}} \mathbb{E}_p^{\otimes n}\left(\frac{|T_n(t)-p(t)|}{\psi^n_{p(t),\beta}}\right)^r \\
&\qquad = \min_{i=1,2} \inf_{\delta\in I_i} \inf_{T_n(t)} \sup_{\substack{p\in\mathscr{P}_d(\beta,L)\\ \delta/2\le p(t)/K_2(0)\le\delta}} \mathbb{E}_p^{\otimes n}\left(\frac{|T_n(t)-p(t)|}{\psi^n_{p(t),\beta}}\right)^r,
\end{aligned}$$

it is sufficient to treat the infima over I_1 and I_2 separately. We start with I_2. Again, by Markov's inequality, for any $\kappa_2>0$,

$$\begin{aligned}
&\inf_{\delta\in I_2} \inf_{T_n(t)} \sup_{\substack{p\in\mathscr{P}_d(\beta,L)\\ \delta/2\le p(t)/K_2(0)\le\delta}} \mathbb{E}_p^{\otimes n}\left(\frac{|T_n(t)-p(t)|}{\psi^n_{p(t),\beta}}\right)^r \\
&\quad \ge \inf_{\delta\in I_2} \inf_{T_n(t)} \sup_{\substack{p\in\mathscr{P}_d(\beta,L)\\ \delta/2\le p(t)/K_2(0)\le\delta}} \mathbb{E}_p^{\otimes n}\left(\frac{|T_n(t)-p(t)|}{(\delta K_2(0)/n)^{\bar\beta/(2\bar\beta+1)}}\right)^r \\
&\quad \ge \inf_{\delta\in I_2} \inf_{T_n(t)} \sup_{\substack{p\in\mathscr{P}_d(\beta,L)\\ \delta/2\le p(t)/K_2(0)\le\delta}} \kappa_2^r\cdot\mathbb{P}^{\otimes n}\left(|T_n(t)-p(t)|\ge\kappa_2\,(\delta K_2(0)/n)^{\bar\beta/(2\bar\beta+1)}\right).
\end{aligned}$$

As before, we construct a density shifted to an appropriate center s_n' and perturbate it. This time, the centering point s_n' is chosen such that it fulfills the equation

$$K_1(t-s_n') = \frac{3}{4}\delta K_2(0) - \frac{1}{4}K_2(0)\left(\frac{\delta}{n}\right)^{\frac{\bar\beta}{2\bar\beta+1}}. \tag{0.24}$$

This point exists since the function K_1 is continuous and takes values between 0 and

$$\|K_1\|_{\sup} \ge K_1(0) = \sqrt{3/4}\cdot c_1(\beta,L') > 3/4\cdot c_1(\beta,L) \ge 3/4\cdot\delta K_2(0),$$

and $K_1(t-s_n')$ is larger than $\delta K_2(0)/2$ due to (0.22) (ii). Define

$$h_{n,i} := c_{8,i}(\beta,L)\left(\frac{\delta}{n}\right)^{\frac{\bar\beta}{2\bar\beta+1}\frac{1}{\beta_i}}, \quad i=1,\dots,d$$

with

$$c_{8,i}(\beta,L) := \left(\frac{2L}{\|K_2\|_2^2}\int |x_i|^{\beta_i} K_2^2(x)\,d\lambda^d(x)\right)^{-1/\beta_i}.$$

The hypotheses can now be formulated as

$$\begin{aligned}
p_{0,n}(x) &:= K_1(x - s'_n)\\
p_{1,n}(x) &:= K_1(x - s'_n) + K_2(x - t; h_n) - K_2(x - s'_n; h_n).
\end{aligned}$$

Note that

$$\limsup_{n\to\infty} \sup_{\delta \in I_1 \cup I_2} K_1(t - s'_n) < K_1(0)$$

and hence the perturbations' supports do not intersect for $n \geq n_1(\beta, L)$ for sufficiently large $n_1(\beta, L) \in \mathbb{N}$ (not depending on $\delta \in I_1 \cup I_2$). Again, both hypotheses are contained in $\mathscr{P}_d(\beta, L)$. Furthermore, as for $p_{0,n}$, the hypothesis $p_{1,n}$ is bounded from above by $\delta K_2(0)$ in t and bounded from below by $K_1(t - s'_n) \geq \delta K_2(0)/2$. The hypotheses' distance in t

$$|p_{0,n}(t) - p_{1,n}(t)| = K_2(0; h_n) = \left(\prod_{i=1}^d c_{8,i}(\beta, L)\right)^{\bar\beta} \left(\frac{\delta}{n}\right)^{\frac{\bar\beta}{2\bar\beta+1}} K_2(0)$$

determines the choice of

$$\kappa_2 = \frac{1}{2}\left(\prod_{i=1}^d c_{8,i}(\beta, L)\right)^{\bar\beta} K_2(0)^{\frac{\bar\beta+1}{2\bar\beta+1}}.$$

Furthermore, with $\mathcal{K}(.,.)$ denoting the Kullback-Leibler divergence,

$$\begin{aligned}
\mathcal{K}(\mathbb{P}_{1,n}^{\otimes n}, \mathbb{P}_{0,n}^{\otimes n}) &= n\,\mathcal{K}(\mathbb{P}_{1,n}, \mathbb{P}_{0,n})\\
&= n\int \log\left(\frac{p_{1,n}(x)}{p_{0,n}(x)}\right) p_{1,n}(x)\, d\lambda^d(x)\\
&\leq n\int \Bigg\{\frac{K_2(x - t; h_n) - K_2(x - s'_n; h_n)}{K_1(x - s'_n)} \qquad (0.25)\\
&\qquad \cdot \Big(K_1(x - s'_n) + K_2(x - t; h_n) - K_2(x - s'_n; h_n)\Big)\Bigg\}\, d\lambda^d(x)\\
&= n\int (K_2(x - t; h_n) - K_2(x - s'_n; h_n))\, d\lambda^d(x)\\
&\quad + n\int \frac{(K_2(x - t; h_n) - K_2(x - s'_n; h_n))^2}{K_1(x - s'_n)}\, d\lambda^d(x)\\
&= n\int \frac{(K_2(x - t; h_n) - K_2(x - s'_n; h_n))^2}{K_1(x - s'_n)}\, d\lambda^d(x)\\
&= n\int \frac{K_2^2(x - t; h_n)}{K_1(x - s'_n)}\, d\lambda^d(x) + n\int \frac{K_2^2(x - s'_n; h_n)}{K_1(x - s'_n)}\, d\lambda^d(x)\\
&\leq 2n\int \frac{K_2^2(x - t; h_n)}{K_1(x - s'_n)}\, d\lambda^d(x)\\
&\leq 2n\left(\min_{x \in \prod_{i=1}^d [-g_{2,i}h_{n,i},\, g_{2,i}h_{n,i}]} K_1(x + t - s'_n)\right)^{-1} \int K_2^2(x; h_n)\, d\lambda^d(x)
\end{aligned}$$

$$\leq 2n \left(\frac{c_7(\beta, L)}{4}\right)^{-1} \|K_2\|_2^2 \left(\prod_{i=1}^{d} h_{n,i}\right)^{2\bar{\beta}+1}, \tag{0.26}$$

for $n \geq n_2(\beta, L)$ for sufficiently large $n_2(\beta, L) \in \mathbb{N}$ (not depending on $\delta \in I_2$). Here, inequality (0.25) is due to the inequality $\log(1+x) \leq x$ for $x > -1$ and (0.26) holds true for $n \geq n_2(\beta, L)$ because $c_7(\beta, L)$ does not depend on n while h_n tends to zero. Next, the latter expression (0.26) is bounded from above by

$$\frac{8\|K_2\|_2^2}{c_7(\beta, L)} \left(\prod_{i=1}^{d} c_{8,i}(\beta, L)\right)^{\frac{2\bar{\beta}+1}{\bar{\beta}}} \delta \leq \frac{8c_1(\beta, L)\,\|K_2\|_2^2}{c_7(\beta, L} \left(\prod_{i=1}^{d} c_{8,i}(\beta, L)\right)^{\frac{2\bar{\beta}+1}{\bar{\beta}}} =: \alpha.$$

Combining all results, we obtain by Theorem 2.2 in Tsybakov (2009) (Kullback version) for $n \geq (n_1(\beta, L) \vee n_2(\beta, L))$

$$\inf_{\delta \in I_2} \inf_{T_n(t)} \sup_{\substack{p \in \mathscr{P}_d(\beta, L) \\ \delta/2 \leq p(t) \leq \delta}} \mathbb{E}_p^{\otimes n} \left(\frac{|T_n(t) - p(t)|}{\psi_{p(t),\beta}^n}\right)^r \geq \kappa_2^r \cdot \max\left\{\frac{\exp(-\alpha)}{4}, \frac{1 - \sqrt{\alpha/2}}{2}\right\}.$$

For the remaining infimum over I_1, we use for K_1 the specific choice of the product kernel as described in Section A.4 with factors specified in (0.54), rescaled and normed such that it integrates to one and has the prescribed Hölder regularity (β, L). The corresponding norming constant of the i'th factor in the resulting product kernel is denoted by $c_{9,i}(\beta, L)$, and at this point, we specify the choice

$$c_7(\beta, L) := \|K_1\|_{\sup}.$$

By symmetry of K_1 we may assume without loss of generality $s'_{n,i} \geq t_i$ for all $i = 1, \ldots, d$. The proof is conducted in complete analogy to the case I_2 except for the bound (0.26), which is too rough for the case under consideration now. Instead, define

$$h_{n,i} := c_{10,i}(\beta, L) \left(\frac{\delta}{n}\right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}\frac{1}{\beta_i}}, \quad i = 1, \ldots, d$$

with

$$c_{10,i}(\beta, L) := \left(\min_{j=1,\ldots,d} \frac{K_2(0)}{2} \left(\frac{c_{9,j}(\beta_j, L)^{\beta_j - 2}}{2\beta_j}\right)^{\beta_j}\right)^{\frac{1}{\beta_i}}.$$

Since $K_1(t - s'_n)$ as given in (0.24) is bounded from below by $\delta/2 \cdot K_2(0)$ for any $\delta \in I_1$ due to (0.22) (ii),

$$\begin{aligned} h_{n,i} &\leq \frac{c_{9,i}(\beta_i, L)^{\beta_i - 2}}{2\beta_i} \left(\frac{K_2(0)}{2}\right)^{\frac{1}{\beta_i}} \delta^{\frac{1}{\beta_i}} \\ &\leq \frac{c_{9,i}(\beta_i, L)^{\beta_i - 2}}{2\beta_i} \left(\frac{K_2(0)}{2}\right)^{\frac{1}{\beta_i}} \left(\frac{2}{K_2(0)} K_1(t_i - s'_{n,i})\right)^{\frac{1}{\beta_i}} \end{aligned}$$

$$= \frac{c_{9,i}(\beta_i, L)^{\beta_i-2}}{2\beta_i} K_1(t_i - s'_{n,i})^{\frac{1}{\beta_i}}.$$

By the mean value theorem, for any $0 \le \eta \le g_{2,i}\, h_{n,i}$,

$$\begin{aligned}
\left|\left(1 - \frac{t_i - s'_{n,i}}{g_{2,i}}\right)^{\beta_i} - \left(1 - \frac{t_i - s'_{n,i} + \eta}{g_{2,i}}\right)^{\beta_i}\right| &= \left|-\beta_i \left(1 - \frac{\xi_{n,i}}{g_{2,i}}\right)^{\beta_i - 1} \frac{\eta}{g_{2,i}}\right| \\
&\le \beta_i \left(1 - \frac{t_i - s'_{n,i}}{g_{2,i}}\right)^{\beta_i - 1} \frac{\eta}{g_{2,i}} \\
&= \beta_i c_{9,i}(\beta_i, L)^{1-\beta_i} \frac{K_1(t_i - s'_{n,i})}{K_1(t_i - s'_{n,i})^{1/\beta_i}} \frac{\eta}{g_{2,i}} \\
&\le \frac{1}{2c_{9,i}(\beta_i, L)} K_1(t_i - s'_{n,i}),
\end{aligned}$$

where $\xi_{n,i} \in [t_i - s'_{n,i},\, t_i - s'_{n,i} + g_{2,i}\, h_{n,i}]$ denotes some suitably chosen intermediate point. Consequently, for all $i = 1, \ldots, d$,

$$K_1(t_i - s'_{n,i} + \eta) \;\ge\; \frac{1}{2} K_1(t_i - s'_{n,i}) \qquad \text{for all } |\eta| \le g_{2,i}\, h_{n,i},$$

such that applied to the bound in (0.26),

$$\begin{aligned}
K(\mathbb{P}_{1,n}^{\otimes n}, \mathbb{P}_{0,n}^{\otimes n}) &\le \frac{2\|K_2\|_2^2\, \delta}{2^{-d}\, K_1(t - s'_n)} \left(\prod_{i=1}^d c_{8,i}(\beta, L)\right)^{\frac{2\bar\beta+1}{\bar\beta}} \\
&\le \frac{2\|K_2\|_2^2\, \delta}{2^{-d}\, \delta\, K_2(0)/2} \left(\prod_{i=1}^d c_{8,i}(\beta, L)\right)^{\frac{2\bar\beta+1}{\bar\beta}} \\
&= \frac{2^{d+2}\|K_2\|_2^2}{K_2(0)} \left(\prod_{i=1}^d c_{8,i}(\beta, L)\right)^{\frac{2\bar\beta+1}{\bar\beta}},
\end{aligned}$$

and we conclude as before. □

Proof of Remark 3.2.1. Let $\hat p_n$ be the adaptive estimator as defined in Subsection 3.2. In case of isotropic Hölder regularity with known exponent $\beta \in (0, 2]$ define the threshold

$$\tilde\alpha_n := n^{-\frac{\beta}{\beta+d}} (\log n)^{\zeta_1}$$

for some constant

$$\zeta_1 > \frac{2\beta + d}{2\beta + 2d}.$$

Recall that the construction of $\hat p_n$ makes use of an upper bound on the second parameter L^* of the Hölder class. Hence, since $\mathscr{P}_d(\beta, L) \subset \mathscr{P}_d(\beta, L')$ whenever

$0 < L < L'$, we may assume without loss of generality that both, β and L, are given. It has to be shown that the estimator $\hat{p}_n(t)\mathbb{1}\{\hat{p}_n(t) \geq \tilde{\alpha}_n\}$ attains the bound

$$\limsup_n \sup_{p \in \mathscr{P}_d(\beta,L)} \sup_{\substack{t \in \mathbb{R}^d: \\ p(t) > n^{-\beta/(\beta+d)}}} \frac{\left(\mathbb{E}_p^{\otimes n} |\hat{p}_n(t) \cdot \mathbb{1}\{\hat{p}_n(x) > \tilde{\alpha}_n\} - p(t)|^r\right)^{1/r}}{\vartheta_{p(t),\beta}^n (\log n)^{\frac{3}{2} \vee \zeta_1}} < \infty.$$

We distinguish between the two cases $p(t) > n^{-\beta/(\beta+d)}$ and $p(t) \leq n^{-\beta/(\beta+d)}$. For any $p(t) > n^{-\beta/(\beta+d)}$, elementary algebra reveals

$$\begin{aligned}
&(\mathbb{E}_p^{\otimes n} |\hat{p}_n(t)\mathbb{1}\{\hat{p}_n(t) \geq \tilde{\alpha}_n\} - p(t)|^r)^{1/r} \\
&\quad \leq (\mathbb{E}_p^{\otimes n} (|\hat{p}_n(t) - p(t)|^r \mathbb{1}\{\hat{p}_n(t) \geq \tilde{\alpha}_n\}))^{1/r} + (\mathbb{E}_p^{\otimes n} p(t)^r \mathbb{1}\{\hat{p}_n(t) < \tilde{\alpha}_n\})^{1/r} \\
&\quad \leq (\mathbb{E}_p^{\otimes n} (|\hat{p}_n(t) - p(t)|^r))^{1/r} + (\mathbb{E}_p^{\otimes n} |p(t) - \hat{p}_n(t)|^r \mathbb{1}\{\hat{p}_n(t) < \tilde{\alpha}_n\})^{1/r} \\
&\qquad + (\mathbb{E}_p^{\otimes n} |\hat{p}_n(t)|^r \mathbb{1}\{\hat{p}_n(t) < \tilde{\alpha}_n\})^{1/r} \\
&\quad \leq 2(\mathbb{E}_p^{\otimes n} (|\hat{p}_n(t) - p(t)|^r))^{1/r} + \tilde{\alpha}_n.
\end{aligned}$$

That is, the threshold does not affect the performance of the estimator in regime (ii) (up to a logarithmic term).
The case $p(t) \leq n^{-\beta/(\beta+d)}$ is more involved. We show the bound

$$\limsup_n \sup_{p \in \mathscr{P}_d(\beta,L)} \sup_{\substack{t \in \mathbb{R}^d: \\ p(t) \leq n^{-\beta/(\beta+d)}}} \frac{\left(\mathbb{E}_p^{\otimes n} |\hat{p}_n(t) \cdot \mathbb{1}\{\hat{p}_n(x) > \tilde{\alpha}_n\} - p(t)|^r\right)^{1/r}}{\vartheta_{p(t),\beta}^n (\log n)^{\frac{3}{2} \vee \zeta_1}} < \infty.$$

By Minkowski's inequality it suffices to prove

$$\limsup_n \sup_{p \in \mathscr{P}_d(\beta,L)} \sup_{\substack{t \in \mathbb{R}^d: \\ p(t) \leq n^{-\beta/(\beta+d)}}} \frac{\left(\mathbb{E}_p^{\otimes n} \hat{p}_n(t)^r \cdot \mathbb{1}\{\hat{p}_n(x) > \tilde{\alpha}_n\}\right)^{1/r}}{\vartheta_{p(t),\beta}^n (\log n)^{\frac{3}{2} \vee \zeta_1}} < \infty.$$

Recall from the construction of the density estimator in Subsection 3.2, that in case of isotropic Hölder smoothness we simplify the method by replacing the ordering by estimated variances in condition (3.8) "for all $m \in \mathcal{J}$ with $\hat{\sigma}_t^2(m) \geq \hat{\sigma}_t^2(j)$" by the classical order "for all $m \in \mathcal{J}$ with $m \geq j$" as the componentwise ordering is the same for all components. We decompose

$$\left(\mathbb{E}_p^{\otimes n} \left(\hat{p}_n(t)^r \cdot \mathbb{1}\{\hat{p}_n(t) > \tilde{\alpha}_n\}\right)\right)^{1/r} = S_1 + S_2$$

with

$$\begin{aligned}
S_1 &= \left(\mathbb{E}_p^{\otimes n} \left[\hat{p}_n(t)^r \cdot \mathbb{1}\{\hat{p}_n(t) > \tilde{\alpha}_n\}\mathbb{1}\{\hat{h} > \bar{h}\}\right]\right)^{1/r} \\
S_2 &= \left(\mathbb{E}_p^{\otimes n} \left[\hat{p}_n(t)^r \cdot \mathbb{1}\{\hat{p}_n(t) > \tilde{\alpha}_n\}\mathbb{1}\{\hat{h} \leq \bar{h}\}\right]\right)^{1/r},
\end{aligned}$$

with $\bar{h}$ as defined in (0.46). As concerns S_1,

$$S_1 \leq c_1 \left\{ \mathbb{P}^{\otimes n} \left(\hat{p}_n(t) - \hat{p}_{n,\bar{h}}(t) > \frac{1}{2}\tilde{\alpha}_n,\ \hat{h} > \bar{h} \right) + \mathbb{P}^{\otimes n} \left(\hat{p}_{n,\bar{h}}(t) > \frac{1}{2}\tilde{\alpha}_n \right) \right\}^{1/r}$$

since $\hat{p}_n \leq c_1$ by construction. Using the selection procedure in the first probability and the bound

$$\mathbb{E}_p^{\otimes n} \hat{p}_{n,\bar{h}}(t) \leq b_t(\bar{h}) + p(t) \leq c(\beta, L)\bar{h}^{\beta} + n^{-\frac{\beta}{\beta+d}},$$

which is bounded by $\tilde{\alpha}_n/4$ for all $n \geq n_0$ and suitable $n_0 \in \mathbb{N}$ not depending on p since $\zeta_1 > \beta/(\beta+d)$. Thus, the term S_1 is for $n \geq n_0$ bounded by

$$c_1 \left\{ \mathbb{P}^{\otimes n} \left(\sqrt{\hat{\sigma}^2_{t,\mathrm{trunc}}(\bar{h}) \log n} > \frac{1}{2}\tilde{\alpha}_n \right) + \mathbb{P}^{\otimes n} \left(\hat{p}_{n,\bar{h}}(t) - \mathbb{E}_p^{\otimes n} \hat{p}_{n,\bar{h}}(t) > \frac{1}{4}\tilde{\alpha}_n \right) \right\}^{1/r}.$$

Denoting

$$A_{\bar{h}} := \left\{ \left| \frac{\tilde{\sigma}^2_{t,\mathrm{trunc}}(\bar{h})}{\sigma^2_{t,\mathrm{trunc}}(\bar{h})} - 1 \right| < \frac{1}{2} \right\},$$

we obtain

$$\mathbb{P}^{\otimes n} \left(\sqrt{\hat{\sigma}^2_{t,\mathrm{trunc}}(\bar{h}) \log n} > \frac{1}{2}\tilde{\alpha}_n \right) \leq \mathbb{P}^{\otimes n}(A^c_{\bar{h}}) + \mathbb{P}^{\otimes n} \left(\sqrt{\frac{3}{2}\sigma^2_{t,\mathrm{trunc}}(\bar{h}) \log n} > \frac{1}{2}\tilde{\alpha}_n \right).$$

The second term in the last inequality vanishes by Lemma 5.4 and the choice of ζ_1, while

$$\mathbb{P}^{\otimes n}(A^c_{\bar{h}}) \leq 2 \exp \left(\frac{3}{32\|K\|_{\sup}} \log^2 n \right)$$

according to Lemma 5.3. As concerns S_2,

$$S_2 \leq c_1 \mathbb{P}^{\otimes n}(\hat{h} \leq \bar{h})^{1/r}.$$

If $\hat{h} > \bar{h}$, then $\bar{h}$ cannot be an admissible bandwidth, see (3.7), because $\hat{h}$ had not been chosen in the minimization problem (3.9) otherwise. Hence, by definition there exists a bandwidth $h \in \mathcal{G}$ with $h \leq \bar{h}$ such that

$$|\hat{p}_{n,\bar{h}}(t) - \hat{p}_{n,h}(t)| > c_{14}\sqrt{\hat{\sigma}^2_t(h) \log n}.$$

Following the lines for the bound on R^- in the proof of Theorem 3.3, we obtain

$$\mathbb{P}^{\otimes n}(\hat{h} \leq \bar{h}) \leq \sum_{h \in \mathcal{G}} \left(\mathbb{P}^{\otimes n}(B_{1,h}) + \mathbb{P}^{\otimes n}(B_{2,h} \cap A_{h,\bar{h}}) + \mathbb{P}^{\otimes n}(A^c_{h,\bar{h}}) \right),$$

where

$$B_{1,h} := \left\{ \left| \hat{p}_{n,\bar{h}}(t) - \mathbb{E}_p^{\otimes n} \hat{p}_{n,\bar{h}}(t) \right| \right.$$
$$\left. > \frac{1}{2} \left(c_{14} \sqrt{\hat{\sigma}_t^2(h) \log n} - 2c_{15}(\beta, L) \sqrt{\sigma_{t,\mathrm{trunc}}^2(\bar{h}) \log n} \right),\ h \leq \bar{h} \right\},$$

$$B_{2,h} := \left\{ \left| \hat{p}_{n,h}(t) - \mathbb{E}_p^{\otimes n} \hat{p}_{n,h}(t) \right| \right.$$
$$\left. > \frac{1}{2} \left(c_{14} \sqrt{\hat{\sigma}_t^2(h) \log n} - 2c_{15}(\beta, L) \sqrt{\sigma_{t,\mathrm{trunc}}^2(\bar{h}) \log n} \right),\ h \leq \bar{h} \right\},$$

$$A_{h,\bar{h}} := \left\{ \left| \frac{\tilde{\sigma}_{t,\mathrm{trunc}}^2(h)}{\sigma_{t,\mathrm{trunc}}^2(h)} - 1 \right| < \frac{1}{2} \text{ and } \left| \frac{\tilde{\sigma}_{t,\mathrm{trunc}}^2(\bar{h})}{\sigma_{t,\mathrm{trunc}}^2(\bar{h})} - 1 \right| < \frac{1}{2} \right\}.$$

The cardinality of $\mathcal{G}$ is of logarithmic size in n, while all probabilities can be bounded by n^{-c} with a constant c depending monotonously increasing on the constant c_{14} of the bandwidth selection rule. Alltogether,

$$\sup_{p \in \mathscr{P}_d(\beta, L)} \sup_{t \in \mathbb{R}^d} \mathbb{E}_p^{\otimes n} \left(\frac{|\hat{p}_n(t) - p(t)|}{\vartheta_{p(t),\beta}^n (\log n)^{\frac{3}{2} \vee \zeta_1}} \right)^r < \infty.$$

□

THEOREM A.1 (Fast adaptive convergence rate at the support boundary). *For any $[\beta_l^*, \beta_u^*] \subset (0, \infty)$, $[L_l^*, L_u^*] \subset (0, \infty)$ and $r \geq 1$, there exists a constant $c_3(\beta^*, L^*, r) > 0$, such that the new density estimator $\hat{p}_n$ based on a compactly supported kernel of order $\lceil \beta_u^* \rceil$ with adaptively chosen bandwidth according to (3.9) satisfies*

$$\sup_{(\beta,L) \in \mathcal{R}(\beta^*, L^*)} \sup_{p \in \mathscr{P}_d^{iso}(\beta, L)} \sup_{\substack{t \in \mathbb{R}^d: \\ d(\Gamma_p^c, t) \leq \left(\frac{\log n}{n} \right)^{\frac{1}{\beta + d}}}} \mathbb{E}_p^{\otimes n} \left(\frac{|\hat{p}_n(t) - p(t)|}{\tilde{\tau}_\beta^n} \right)^r \leq c_3(\beta^*, L^*, r),$$

where $\tilde{\tau}_\beta^n := n^{-\frac{\beta}{\beta+d}} (\log n)^{3/2}$.

Note that $\hat{p}_n$ requires no a priori information about $\partial \Gamma_p$. The result likewise extends to the anisotropic setting because the Euclidean norm on $\mathbb{R}^d$ (in the definition of $d(\Gamma_p^c, t)$) and the maximum norm $\|t - y\|_{\max} = \max_{i=1,\ldots,d} |t_i - y_i|$ are equivalent.

PROOF OF THEOREM A.1. The proof follows the same lines as the proof of Theorem 3.3 except for the following modifications. We write

$$\mathbb{E}_p^{\otimes n} |\hat{p}_{n,\hat{j}}(t) - p(t)|^r = \mathbb{E}_p^{\otimes n} \left[|\hat{p}_{n,\hat{j}}(t) - p(t)|^r \cdot \mathbb{1}\{\hat{j} \leq \bar{j}\} \right]$$

$$
\begin{aligned}
&+ \mathbb{E}_p^{\otimes n}\Big[|\hat{p}_{n,\hat{j}}(t) - p(t)|^r \cdot \mathbb{1}\{\hat{j} > \bar{j}\}\Big] \\
&=: \tilde{R}^+ + \tilde{R}^-,
\end{aligned}
$$

where this time $\tilde{R}^+$ is decomposed as

$$
\begin{aligned}
\tilde{R}^+ &\le 2^{r-1}\bigg(\mathbb{E}_p^{\otimes n}\Big[|\hat{p}_{n,\hat{j}}(t) - \hat{p}_{n,\bar{j}}(t)|^r \cdot \mathbb{1}\{\hat{j} \le \bar{j}\}\Big] \\
&\qquad + \mathbb{E}_p^{\otimes n}\Big[|\hat{p}_{n,\bar{j}}(t) - p(t)|^r \cdot \mathbb{1}\{\hat{j} \le \bar{j}\}\Big]\bigg) \\
&=: 2^{r-1}(\tilde{S}_1 + \tilde{S}_3). \qquad (0.27)
\end{aligned}
$$

Here, $\bar{j}$ corresponds to the reference bandwidth $\bar{h}$ defined in (0.46). We need to verify the bounds

$$B_t(\bar{j}) \le c(\beta, L)\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j}) \log n} \qquad (0.28)$$

$$\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j})} \le c(\beta, L)\left(\frac{\log n}{n}\right)^{\frac{\beta}{\beta+d}} \qquad (0.29)$$

for isotropic Hölder smoothness of arbitrary $\beta > 0$ and $d(\Gamma_p^c, t) \le (\log n/n)^{1/(\beta+d)}$. Since for any $x \in \mathbb{R}^d$ the corresponding value of p is bounded by $p(x) \le L\, d(\Gamma_p^c, x)^\beta$ and in particular $p(t) \le L\,(\log n/n)^{\beta/(\beta+d)}$, we obtain

$$\bar{h} \le c(\beta, L)\left(\frac{\log n}{n}\right)^{\frac{1}{\beta+d}}$$

The first bound (0.28) is a consequence of the classical upper bound on the bias for higher order kernels, whereas the second bound (0.29) follows by Lemma 5.1 (iii). The terms $\tilde{S}_1$ and $\tilde{S}_3$ in (0.27) then require no further arguments. As concerns $\tilde{R}^-$, it remains to investigate

$$\mathbb{P}^{\otimes n}\left(\hat{j} > \bar{j}\right) \le \sum_{m > \bar{j}} \mathbb{P}^{\otimes n}\left(|\hat{p}_{n,\bar{j}} - \hat{p}_{n,m}(t)| > c_{14}\sqrt{\hat{\sigma}_t^2(m)\log n}\right).$$

Note that only indices $m > \bar{j}$ are taken into account. In order to line up with the previously developed arguments, it is sufficient to prove

$$\sigma^2_{t,\mathrm{trunc}}(\bar{j}) \le c(\beta, L)\,\sigma^2_{t,\mathrm{trunc}}(m) \qquad (0.30)$$

for all $m > \bar{j}$. Since $p(t) \le L d(\Gamma_p^c, t)^\beta \le L(\log n/n)^{\beta/(\beta+d)}$, the reference bandwidth $\bar{h}$ satisfies

$$c_{11}(\beta, L)\left(\frac{\log n}{n}\right)^{\frac{1}{\beta+d}} \le \bar{h} \le c_{26}(\beta, L)\left(\frac{\log n}{n}\right)^{\frac{1}{\beta+d}}$$

for some constant $c_{26}(\beta, L) > c_{11}(\beta, L)$. By Lemma 5.1 (iii),

$$\sigma_t^2(\bar{h}) \leq \frac{L\|K\|_2^2}{n\bar{h}^d} (c_{26}(\beta, L) + 1)^\beta \left(\frac{\log n}{n}\right)^{\frac{\beta}{\beta+d}} \leq c(\beta, L) \frac{\log^2 n}{n^2 \bar{h}^{2d}},$$

which is for $h < \bar{h}$ smaller than

$$\frac{\log^2 n}{n^2 h^{2d}} \leq \sigma_{t,\mathrm{trunc}}^2(h),$$

that is, (0.30) is verified. The further proof can then be conducted as before for Theorem 3.3. $\square$

A.3. Proofs of Section 4.

Theorem (Theorem III.2.2, Chavel 2001) *Given any compact set $K \subset \mathbb{R}^d$, let D denote the closed d-disc of the same measure as K, i.e. $\lambda^d(D) = \lambda^d(K)$. Then*

$$\lambda^d(D^\varepsilon) \leq \lambda^d(K^\varepsilon)$$

for all $\varepsilon > 0$.

Proof of Lemma 4.2. We first show the necessity of $\gamma\beta \leq 1$. Let $p \in \mathscr{P}_d^{iso}(\beta, L)$, and let y be an arbitrary point in the open set Γ_p^c. Then p is constant zero in a neighborhood of y, i.e. all derivatives are zero in y and thus

$$p(x) \;=\; \left|p(x) - P_{y,\lfloor\beta\rfloor}^{(p)}(x)\right| \;\leq\; L\|x - y\|_2^\beta$$

for any $x \in \mathbb{R}^d$ and therefore

$$p(x) \leq L\,(d(\partial\Gamma_p, x))^\beta$$

with d as defined in (0.1). Consequently,

$$\begin{aligned}\lambda^d\left(\overline{\{x \in \mathbb{R}^d \,|\, 0 < p(x) \leq \varepsilon\}}\right) &\geq \lambda^d\left(\overline{\{x \in \Gamma_p \,|\, 0 < L\,(d(\partial\Gamma_p, x))^\beta \leq \varepsilon\}}\right)\\ &= \lambda^d\left(\overline{\{x \in \Gamma_p \,|\, 0 < d(\partial\Gamma_p, x) \leq (\varepsilon/L)^{1/\beta}\}}\right).\end{aligned}$$

It remains to prove that

$$\liminf_{\varepsilon \searrow 0} \frac{\lambda^d\left(\overline{\{x \in \Gamma_p \,|\, 0 < d(\partial\Gamma_p, x) \leq (\varepsilon/L)^{1/\beta}\}}\right)}{\varepsilon^{1/\beta}} > 0.$$

Let K_η denote the closed Euclidean ball with volume

$$\lambda^d\left(\Gamma_p^{-\eta}\right), \quad \eta > 0.$$

For $\delta = (\varepsilon/L)^{1/\beta}$,

$$
\begin{aligned}
\lambda^d\left(\Gamma_p \setminus \Gamma_p^{-\delta}\right) &\geq \lambda^d\left(\left(\Gamma_p^{-\delta}\right)^{\delta} \setminus \Gamma_p^{-\delta}\right) \\
&= \lambda^d\left(\left(\Gamma_p^{-\delta}\right)^{\delta}\right) - \lambda^d\left(\Gamma_p^{-\delta}\right) \\
&\geq \lambda^d\left((K_\delta)^\delta\right) - \lambda^d\left(\Gamma_p^{-\delta}\right) \qquad (0.31) \\
&= \lambda^d\left((K_\delta)^\delta \setminus K_\delta\right),
\end{aligned}
$$

where the isoperimetric inequality for the Minkowski area of compact sets (Theorem III.2.2, Chavel 2001) is applied in inequality (0.31). Since Γ_p has non-empty interior, because it is the support of a continuous Lebesgue density, there exists some $\delta_0 > 0$ such that

$$\lambda^d\left(\Gamma_p^{-\delta_0}\right) > 0.$$

Finally,

$$\lambda^d\left((K_\delta)^\delta \setminus K_\delta\right) \geq \lambda^d\left((K_{\delta_0})^\delta \setminus K_{\delta_0}\right)$$

for all $\delta \leq \delta_0$, while

$$\liminf_{\delta \searrow 0} \frac{\lambda^d\left((K_{\delta_0})^\delta \setminus K_{\delta_0}\right)}{\delta} > 0,$$

which implies $\gamma\beta \leq 1$.
Finally, we need to show that for any $\beta > 0$, $\gamma > 0$ with $\gamma\beta \leq 1$, there exists a compactly supported density $p \in \mathscr{P}_d^{iso}(\beta, L)$ which satisfies the margin condition to the exponent γ. For this aim, it remains to verify that the density $K(\cdot, \beta)$ of (0.55) in Section A.4 satisfies the margin condition to any exponent $\gamma \leq 1/\beta$, which is shown in the proof of Theorem 4.4. □

Let $K_g(\cdot) = g^{-d}K(\cdot/g)$ with $g = g_{\beta,L/2,d}$ be the specific kernel given by (0.55) in Section A.4, and define $K(\cdot\,; h, \beta) := h^\beta K(\cdot/h; \beta)$. The proof of Theorem 4.4 requires sharp estimates of the Lebesgue volume

$$\Lambda_d(K_g(\cdot\,; h, \beta), \varepsilon) \; := \; \lambda^d\left(\overline{\{x \in \mathbb{R}^d \,:\, 0 < K_g(x; h, \beta) \leq \varepsilon\}}\right) \qquad (0.32)$$

of complementary level sets of $K_g(\cdot\,; h, \beta)$, provided by the following lemma.

Lemma 0.1. *There exists a constant $c_{29}(\beta) > 0$, such that for any bandwidths g, h, and any*

$$\varepsilon \leq c_{29}(\beta) h^\beta g^{-d},$$

the volume of the complementary level set defined in (0.32) is upper bounded by

$$\Lambda_d(K_g(\cdot\,; h, \beta), \varepsilon) \; \leq \; c_{29}(\beta)^{-\frac{1}{\beta}}\, dV_d \cdot (gh)^{d-1} g^{\frac{\beta+d}{\beta}} \varepsilon^{\frac{1}{\beta}},$$

where V_d denotes the volume of the d-dimensional unit ball.

Proof. Since $K(\cdot,\beta)$ is uniformly bounded away from zero on $\mathcal{B}_\delta(0)$ for any $\delta<1$,

$$\varepsilon_0(\beta) := \inf_{z\in\mathcal{B}_{\frac{1}{2}}(0)} K(z,\beta) > 0.$$

Hence, since $w(x)$ (with w specified in (0.55)) is monotonously increasing in $\|x\|_2$,

$$\Big\{x\in\mathbb{R}^d \,:\, 0<K(x,\beta)\le\varepsilon\Big\} \subset \Big\{x\in\mathbb{R}^d \,:\, 0< c_{30}(\beta)w(1/2)(1-\|x\|_2)_+^\beta \le \varepsilon\Big\}$$

for $0<\varepsilon<\varepsilon_0$. Consequently, whenever $\varepsilon < h^\beta g^{-d}\varepsilon_0$,

$$\Big\{x\in\mathbb{R}^d \,: 0<K_g(x;h,\beta)\le\varepsilon\Big\} \subset \big\{x\in\mathbb{R}^d \,:\, 0<Q_{g,h}(\cdot,\beta)\le\varepsilon\big\}$$

with

$$Q_{g,h}(\cdot,\beta) := c_{30}(\beta)w(1/2)\frac{h^\beta}{g^d}\left(1-\left\|\frac{\cdot}{gh}\right\|_2\right)_+^\beta.$$

It remains to bound $\Lambda_d(Q_{g,h}(\cdot,\beta),\varepsilon)$ for any $\varepsilon<h^\beta g^{-d}\varepsilon_0$. Note first that both, the support and the level sets of $Q_{g,h}(\cdot,\beta)$ are concentric balls, and hence $\Lambda_d(Q_{g,h}(\cdot,\beta),\varepsilon)$ is for

$$\varepsilon < h^\beta g^{-d}\left\{\varepsilon_0 \wedge \left(\frac{1}{2}\right)^\beta c_{30}(\beta)w(1/2)\right\} \tag{0.33}$$

the volume of a spherical shell with inner radius larger than the radius $gh/2$. The volume of a d-dimensional spherical shell with outer radius R and inner radius $r\le R$ equals

$$\lambda^d(\mathcal{B}_R(0)) - \lambda^d(\mathcal{B}_r(0)) = V_d(R^d-r^d) = V_d(R-r)\sum_{j=0}^{d-1} r^j R^{d-1-j}, \tag{0.34}$$

which in turn is directly upper bounded by $dV_d\, R^{d-1}(R-r)$. Since $Q_{g,h}(\cdot,\beta)$ attains ε on the sphere with radius

$$gh - [c_{30}(\beta)w(1/2)]^{-\frac{1}{\beta}} g^{\frac{\beta+d}{\beta}} \varepsilon^{\frac{1}{\beta}}$$

for ε satisfying (0.33),

$$\Lambda_d(K_g(\cdot\,;h,\beta),\varepsilon) \;\le\; [w(1/2)c_{30}(\beta)]^{-\frac{1}{\beta}}\, dV_d\cdot(gh)^{d-1} g^{\frac{\beta+d}{\beta}}\varepsilon^{\frac{1}{\beta}}.$$

□

Proof of Theorem 4.4. Since we measure the risk with respect to the L_1-type distance d_Δ it does not suffice to reduce the problem to two hypotheses. Instead, we use Assouad's hypercube technique where the hypotheses constitute an m-dimensional hypercube and thereby reduce the problem of testing m problems to m

problems of testing two hypotheses. As before, we construct a Hölder smooth density with prescribed regularity (β, L) using the function $K_g(x;\beta)$ and a perturbation based on $K_g(x;h_n,\beta)$ with bandwidth

$$h_n := (2n)^{-\frac{1}{\beta+d}} . \tag{0.35}$$

Recall that $K_g(\cdot;h_n,\beta)$ implicitly depends on L via g. Furthermore, choose

$$m_n = \lfloor h_n^{\gamma\beta-d} \rfloor + 1. \tag{0.36}$$

Again, denote by $\mathcal{B}_r(x)$ the closed Euclidean ball with radius r around x. Now choose points $z_i = (z_{i,1}, \ldots, z_{i,d})$, $i = 1, \ldots, m_n$ in $\mathcal{B}_{g/2}(0)$ separated in each coordinate by at least $2gh_n$, which is possible for n large enough since the total support volume of all perturbations is of the order $m_n(gh_n)^d$ and tends to zero. These points are shifted outside the support of K_g and the new points are denoted by

$$\begin{aligned} z'_{i,1} &= z_{i,1} + 2g, \\ z'_{i,j} &= z_{i,j}, \quad j = 2, \ldots, d \end{aligned}$$

for $i = 1, \ldots, m_n$. Then, for $\omega = (\omega_1, \ldots, \omega_{m_n}) \in \Omega := \{0,1\}^{m_n}$ denote the hypotheses by

$$p_{\omega,n}(x) = K_g(x;\beta) + \sum_{k=1}^{m_n} \omega_k \left[K_g(x - z'_k; h_n, \beta) - K_g(x - z_k; h_n, \beta) \right],$$

their supports by $\Gamma_{p_{\omega,n}}$ and the corresponding probability measures by $\mathbb{P}_{\omega,n}$. Obviously, $p_{\omega,n}$ is for sufficiently large n a density again and is contained in $\mathscr{P}_d(\beta, L)$. We will now show that $p_{\omega,n}$ has the right margin exponent as well. For sufficiently large n it holds

$$\begin{aligned} \Lambda_d(p_{\omega,n}, \varepsilon) \leq{}& \Lambda_d(K_g, \varepsilon) \\ &+ m_n \Lambda_d(K_g(\cdot\,; h_n, \beta), \varepsilon)\ \mathbb{1}\left\{\varepsilon \leq c_{29}(\beta) h_n^\beta g^{-d}\right\} \\ &+ 2m_n V_d \cdot (gh_n)^d \mathbb{1}\left\{\varepsilon > c_{29}(\beta) h_n^\beta g^{-d}\right\}. \end{aligned} \tag{0.37}$$

Now, because $m_n \leq h_n^{\gamma\beta-d} + 1 \leq 2h_n^{\gamma\beta-d}$, Lemma 0.1 yields

$$\begin{aligned} & m_n \Lambda_d(K_g(\cdot\,; h_n, \beta), \varepsilon) \cdot \mathbb{1}\left\{\varepsilon \leq c_{29}(\beta) h_n^\beta g^{-d}\right\} \qquad (0.38)\\ &\quad \leq c(\beta, L) \cdot h_n^{\gamma\beta-d} h_n^{d-1} \varepsilon^{\frac{1}{\beta}} \cdot \mathbb{1}\left\{\varepsilon \leq c_{29}(\beta) h_n^\beta g^{-d}\right\} \\ &\quad = c(\beta, L) \cdot h_n^{\gamma\beta-1} \varepsilon^{\frac{1}{\beta}} \cdot \mathbb{1}\left\{\varepsilon \leq c_{29}(\beta) h_n^\beta g^{-d}\right\} \\ &\quad \leq c(\beta, L) \cdot \varepsilon^{\gamma-\frac{1}{\beta}}\ \varepsilon^{\frac{1}{\beta}} \cdot \mathbb{1}\left\{\varepsilon \leq c_{29}(\beta) h_n^\beta g^{-d}\right\} \\ &\quad = c(\beta, L) \cdot \varepsilon^{\gamma} \cdot \mathbb{1}\left\{\varepsilon \leq c_{29}(\beta) h_n^\beta g^{-d}\right\}, \qquad (0.39) \end{aligned}$$

where the last inequality is due to the property $\gamma\beta \leq 1$. Furthermore, since $m_n h_n^d \leq 2h_n^{\gamma\beta}$, we can derive a similar bound for the last term in (0.37)

$$2m_n V_d \cdot (gh_n)^d \cdot \mathbb{1}\left\{\varepsilon > c_{29}(\beta) h_n^\beta g^{-d}\right\}$$

$$\leq c(\beta, L, \gamma) \cdot \varepsilon^{\gamma} \cdot \mathbb{1}\left\{\varepsilon > c_{29}(\beta) h_n^{\beta} g^{-d}\right\}. \tag{0.40}$$

Clearly, for $\varepsilon \leq c_{29}(\beta) h_n^{\beta} g^{-d} \wedge 1$, Lemma 0.1 also yields

$$\Lambda_d(K_g, \varepsilon) \leq c(\beta, L) \cdot \varepsilon^{\frac{1}{\beta}} \leq c(\beta, L) \cdot \varepsilon^{\gamma} \tag{0.41}$$

using the property $\gamma\beta \leq 1$ again. In summary, inequality (0.37) simplifies with (0.38), (0.40) and (0.41) for $\varepsilon \leq c_{29}(\beta) h_n^{\beta} g^{-d} \wedge 1$ to

$$\Lambda_d(p_{\omega,n}, \varepsilon) \;\leq\; c(\beta, L, \gamma) \cdot \varepsilon^{\gamma},$$

i.e. there exist constants $\kappa_1 = \kappa_1(\beta, L)$ and $\kappa_2 = \kappa_2(\beta, L, \gamma)$ such that $p_{\omega,n}$ fulfills the κ-margin condition.
It remains to show that $p_{\omega,n}$ also satisfies the complexity condition to the exponent $\mu = \gamma\beta$. To check this condition, two different types of decompositions are considered, depending on whether $\varepsilon \leq gh_n$ or $\varepsilon > gh_n$. For $\varepsilon \leq gh_n$ we consider the canonical disjoint decomposition $\Gamma_{p_{\omega,n}} = \Gamma_{p_{\omega,n}} \cup \emptyset =: A_{1,\varepsilon} \cup A_{2,\varepsilon}$. Clearly, by formula (0.34),

$$\begin{aligned}
\lambda^d\left(\Gamma^{\varepsilon}_{p_{\omega,n}} \setminus \Gamma_{p_{\omega,n}}\right) &\leq V_d \cdot \Big((g+\varepsilon)^d - g^d + m_n(gh_n+\varepsilon)^d - m_n(gh_n)^d\Big) \\
&\leq dV_d \cdot \Big((g+\varepsilon)^{d-1}\varepsilon + m_n(gh_n+\varepsilon)^{d-1}\varepsilon\Big) \\
&\leq dV_d \cdot \Big((2g)^{d-1}\varepsilon + m_n(2gh_n)^{d-1}\varepsilon\Big) \\
&= dV_d \cdot (2g)^{d-1}\varepsilon \cdot (1 + m_n h_n^{d-1}) \\
&\leq dV_d \cdot (2g)^{d-1}\varepsilon \cdot 3h_n^{\gamma\beta-1} \\
&\leq c(\beta, L, \gamma, d) \cdot \varepsilon^{\gamma\beta} \qquad (0.42)
\end{aligned}$$

where inequality (0.42) follows from $\gamma\beta \leq 1$. For $\varepsilon > gh_n$ let ξ_1 be an arbitrary constant and choose the following decomposition for the complexity condition

$$\Gamma_{p_{\omega,n}} = \mathcal{B}_g(0) \;\cup \bigcup_{k\,:\,\omega_k=1} \mathcal{B}_{gh_n}(z'_k) =: A_{1,\varepsilon} \cup A_{2,\varepsilon} \tag{0.43}$$

Then for all $\varepsilon \leq \xi_1$ similar calculations as before yield

$$\begin{aligned}
\lambda^d\left(\mathcal{B}_g(0)^{\varepsilon} \setminus \mathcal{B}_g(0)\right) &\leq dV_d \cdot (g+\varepsilon)^{d-1}\varepsilon \\
&\leq dV_d \cdot (g+\xi_1)^{d-1}\xi_1^{1-\gamma\beta}\varepsilon^{\gamma\beta},
\end{aligned}$$

where the last inequality again follows from $\gamma\beta \leq 1$. To check the complexity condition it remains to upper bound the Lebesgue volume of the second part in the decomposition (0.43). For $gh_n < \varepsilon \leq \xi_1$,

$$\begin{aligned}
\lambda^d\left(\bigcup_{k\,:\,\omega_k=1} \mathcal{B}_{gh_n}(z'_k)\right) &\leq m_n V_d \cdot (gh_n)^d \\
&\leq 2V_d g^d h_n^{\gamma\beta}
\end{aligned}$$

$$\leq 2V_d g^{d-\gamma\beta}\varepsilon^{\gamma\beta},$$

i.e. there exist constants $\xi_1 = \xi_1(\beta, L, \gamma)$ and $\xi_2 = \xi_2(\beta, L, \gamma)$ such that $\Gamma_{p_{\omega_n}}$ satisfies the ξ-complexity condition to the exponent $\mu = \gamma\beta$. The further proof accomplishes two tasks. Firstly, the minimax risk will be reduced to the form

$$\inf_{\hat{\omega}} \sup_{\omega\in\Omega} \mathbb{E}^{\otimes n}_{\omega,n} \rho(\hat{\omega}, \omega),$$

where ρ is the Hamming distance and expectation is taken with respect to $\mathbb{P}^{\otimes n}_{\omega,n}$. Afterwards, Assouad's lemma can be applied and it remains to bound a suitable distance of all neighboring probability measures with Hamming distance one. For any support estimator $\hat{\Gamma}_n$ we evaluate

$$\begin{aligned}\mathbb{E}^{\otimes n}_{\omega,n}\left[d_\Delta(\hat{\Gamma}_n, \Gamma_{p_{\omega,n}})\right] &= \mathbb{E}^{\otimes n}_{\omega,n}\left[\lambda^d\left(\left(\hat{\Gamma}_n \setminus \Gamma_{p_{\omega,n}}\right) \cup \left(\Gamma_{p_{\omega,n}} \setminus \hat{\Gamma}_n\right)\right)\right] \\ &= \mathbb{E}^{\otimes n}_{\omega,n}\left[\int \left|\mathbb{1}_{\hat{\Gamma}_n}(x) - \mathbb{1}_{\Gamma_{p_{\omega,n}}}(x)\right| d\lambda^d(x)\right]\end{aligned}$$

and due to the non-negativity of the integrand this expression can be estimated from below by

$$\mathbb{E}^{\otimes n}_{\omega,n}\left[d_\Delta(\hat{\Gamma}_n, \Gamma_{p_{\omega,n}})\right] \geq \sum_{i=1}^{m_n} \mathbb{E}^{\otimes n}_{\omega,n}\left[\int_{\mathcal{B}_{gh_n}(z_i')} \left|\mathbb{1}_{\hat{\Gamma}_n}(x) - \mathbb{1}_{\Gamma_{p_{\omega,n}}}(x)\right| d\lambda^d(x)\right],$$

which in turn simplifies to

$$\mathbb{E}^{\otimes n}_{\omega,n}\left[d_\Delta(\hat{\Gamma}_n, \Gamma_{p_{\omega,n}})\right] \geq \sum_{i=1}^{m_n} \mathbb{E}^{\otimes n}_{\omega,n}\left[\int_{\mathcal{B}_{gh_n}(z_i')} \left|\mathbb{1}_{\hat{\Gamma}_n}(x) - \omega_i\right| d\lambda^d(x)\right].$$

Introducing $\bar{\omega} = (\bar{\omega}_1, \ldots, \bar{\omega}_{m_n})$ with

$$\bar{\omega}_i := \operatorname*{arg\,min}_{\omega_i\in\{0,1\}} \int_{\mathcal{B}_{gh_n}(z_i')} \left|\mathbb{1}_{\hat{\Gamma}_n}(x) - \omega_i\right| d\lambda^d(x)$$

depending on $\hat{\Gamma}_n$, we obtain

$$\begin{aligned}\mathbb{E}^{\otimes n}_{\omega,n}\left[d_\Delta(\hat{\Gamma}_n, \Gamma_{p_{\omega,n}})\right] &\geq \sum_{i=1}^{m_n}\left(\frac{1}{2}\mathbb{E}^{\otimes n}_{\omega,n}\left[\int_{\mathcal{B}_{gh_n}(z_i')} \left|\mathbb{1}_{\hat{\Gamma}_n}(x) - \omega_i\right| d\lambda^d(x)\right]\right. \\ &\qquad \left. + \frac{1}{2}\mathbb{E}^{\otimes n}_{\omega,n}\left[\int_{\mathcal{B}_{gh_n}(z_i')} \left|\mathbb{1}_{\hat{\Gamma}_n}(x) - \bar{\omega}_i\right| d\lambda^d(x)\right]\right) \\ &\geq \frac{1}{2}\sum_{i=1}^{m_n} \mathbb{E}^{\otimes n}_{\omega,n}\left[\int_{\mathcal{B}_{gh_n}(z_i')} |\omega_i - \bar{\omega}_i|\, d\lambda^d(x)\right] \\ &\geq \frac{1}{2}V_d(gh_n)^d\, \mathbb{E}^{\otimes n}_{\omega,n}\left[\sum_{i=1}^{m_n} |\omega_i - \bar{\omega}_i|\right]\end{aligned}$$

$$= \frac{1}{2} V_d (g h_n)^d \, \mathbb{E}_{\omega,n}^{\otimes n} \rho(\bar{\omega}, \omega)$$

and consequently

$$\inf_{\hat{\Gamma}_n} \sup_{p \in \mathscr{P}_d(\beta,L,\gamma,\kappa,\xi)} \mathbb{E}_p^{\otimes n} \left[d_\Delta(\hat{\Gamma}_n, \Gamma_p) \right] \;\geq\; c(\beta, L) h_n^d \cdot \inf_{\hat{\omega}} \max_{\omega \in \Omega} \mathbb{E}_{\omega,n}^{\otimes n} \rho(\omega, \hat{\omega}), \tag{0.44}$$

where the infimum runs over all measurable $\hat{\omega} = \hat{\omega}(X_1, \ldots, X_n)$ with values in $\{0,1\}^{m_n}$. Now we use the Hellinger version of Assouad's lemma, cf. Theorem 2.12 (iii) in Tsybakov (2009), to bound the expression on the right-hand side of (0.44). For this purpose, the squared Hellinger distance between two arbitrary probability measures $\mathbb{P}_{\omega,n}$ and $\mathbb{P}_{\omega',n}$ with $\omega, \omega' \in \Omega$ and $\rho(\omega, \omega') = 1$ has to be bounded and we use inequality (0.23) for this purpose. Of course, ω and ω' coincide except for one component, say j. Again, by Bernoulli's inequality

$$\begin{aligned}
H^2(\mathbb{P}_{\omega,n}^{\otimes n}, \mathbb{P}_{\omega',n}^{\otimes n}) &\leq n \int \left(\sqrt{p_{\omega,n}(x)} - \sqrt{p_{\omega',n}(x)} \right)^2 d\lambda(x) \\
&= n \int \left(\sqrt{K_g(x;\beta)} - \sqrt{K_g(x;\beta) - K_g(x - z_j; h_n, \beta)} \right)^2 d\lambda^d(x) \\
&\quad + n \int K_g(x - z_j'; h_n, \beta) d\lambda^d(x) \\
&\leq 2n \int K_g(x; h_n, \beta) d\lambda^d(x) \\
&= 2n h_n^{\beta+d}.
\end{aligned}$$

By the choice of h_n in (0.35), this distance is bounded by one which yields together with (0.36) and inequality (0.44), see Tsybakov (2009),

$$\begin{aligned}
\inf_{\hat{\Gamma}_n} \sup_{p \in \mathscr{P}_d(\beta,L,\gamma,\kappa,\xi)} \mathbb{E}_p^{\otimes n} \left[d_\Delta(\hat{\Gamma}_n, \Gamma_p) \right] \;&\geq\; c(\beta, L)\, h_n^d \cdot \frac{m_n}{2} \left(1 - \sqrt{\frac{3}{4}} \right) \\
&\geq\; c(\beta, L)\, n^{-\frac{\gamma\beta}{\beta+d}}.
\end{aligned}$$

□

For the proof of Theorem 4.5 we need the following lemma, which is based on the work of Tsybakov (2004) and has been formulated for the problem of level set estimation by Rigollet and Vert (2009), Proposition A.1. The proof likewise holds for the support estimation problem and is transferred without any major modifications. However, we additionally need to verify that the bound holds uniformly over the class of densities satisfying the κ-margin condition to the exponent γ.

LEMMA 0.2. *For any density p which satisfies the κ-margin condition with exponent $\gamma > 0$, there exists a constant $c_{18}(\kappa, \gamma)$ such that the Lebesgue volume of a measurable subset G of Γ_p is bounded by*

$$\lambda^d(G) \leq c_{18}(\kappa, \gamma) \left(\int_G p(x)\, d\lambda^d(x) \right)^{\frac{\gamma}{1+\gamma}}.$$

PROOF. First note that for any p satisfying the κ-margin condition to the exponent $\gamma > 0$,

$$\begin{aligned}
\lambda^d(\Gamma_p) &= \lambda^d(\Gamma_p \cap \{p \le \kappa_1\}) \ + \ \lambda^d(\Gamma_p \cap \{p > \kappa_1\}) \\
&\le \kappa_2 \cdot \kappa_1^\gamma \ + \ \frac{1}{\kappa_1} \int_{\mathbb{R}^d} p \mathbb{1}_{p>\kappa_1} d\lambda^d \\
&\le \ \kappa_2 \cdot \kappa_1^\gamma \ + \ \frac{1}{\kappa_1} \\
&=: \ c_{19}(\kappa, \gamma).
\end{aligned}$$

Let G be a measurable subset of Γ_p. Then,

$$\begin{aligned}
\lambda^d(G \cap \{p > \varepsilon\}) &= \int_G \mathbb{1}\{p(x) > \varepsilon\} d\lambda^d(x) \\
&\le \int_G \mathbb{1}\{p(x) > \varepsilon\} \frac{p(x)}{\varepsilon} d\lambda^d(x) \\
&\le \frac{1}{\varepsilon} \int_G p(x) d\lambda^d(x)
\end{aligned}$$

and consequently for all $0 < \varepsilon \le \kappa_1$,

$$\begin{aligned}
\int_G p(x) d\lambda^d(x) &\ge \varepsilon \, \lambda^d(G \cap \{p > \varepsilon\}) \\
&= \varepsilon \Big(\lambda^d(G) - \lambda^d(G \cap \{p \le \varepsilon\}) \Big) \\
&\ge \varepsilon \Big(\lambda^d(G) - \lambda^d \left(\overline{\{0 < p \le \varepsilon\}} \right) \Big) \\
&\ge \varepsilon \lambda^d(G) - c_{20}(\kappa, \gamma) \varepsilon^{\gamma+1} \qquad (0.45)
\end{aligned}$$

with

$$c_{20}(\kappa, \gamma) \ = \ \kappa_2 \vee \frac{c_{19}(\kappa, \gamma)}{\kappa_1^\gamma (\gamma + 1)}.$$

The right hand side of (0.45) is maximized by the specific choice of

$$\varepsilon = \left(\frac{\lambda^d(G)}{c_{20}(\kappa, \gamma) \cdot (\gamma + 1)} \right)^{1/\gamma} \ \le \ \left(\frac{\lambda^d(\Gamma_p)}{c_{20}(\kappa, \gamma) \cdot (\gamma + 1)} \right)^{1/\gamma} \ \le \ \kappa_1.$$

Plugging this specific ε in (0.45) yields

$$\int_G p(x) d\lambda^d(x) \ \ge \ \Big(\lambda^d(G) \Big)^{\frac{\gamma+1}{\gamma}} \big(c_{20}(\kappa, \gamma) \big)^{-1/\gamma} \Big((\gamma + 1)^{-1/\gamma} - (\gamma + 1)^{-(\gamma+1)/\gamma} \Big).$$

□

We now turn to the proof of the upper bound on the support estimator's risk, which can be proved for c_5 satisfying

$$c_5(\beta, L) \ \ge \ 2 \left\{ c_1(\beta, L) \ \vee \ \left(c_{21}(\beta, L) \Big(2^{3/2} (1 + \sqrt{3/2} c_{14} + c_{15}(\beta, L)) \Big) \right)^{\frac{2\beta+d}{\beta+d}} \right\}.$$

Proof of Theorem 4.5. We prove an upper bound on the risk with respect to the symmetric difference of sets for some $p \in \mathscr{P}_d(\beta, L, \gamma, \kappa, \xi)$. All constants hold uniformly in p over this class. For notational convenience, we write $\hat{p}_n(x) = \hat{p}_{n,\hat{j}}(x)$ and denote by $\bar{j} = \bar{j}(x)$ the exponent corresponding to the reference bandwidth

$$\bar{h} = \bar{h}(x) := c_{11}(\beta, L) \cdot \max\left\{ \left(\frac{\log n}{n}\right)^{\frac{1}{\beta+d}}, \left(\frac{p(x)\log n}{n}\right)^{\frac{1}{2\beta+d}} \right\}. \tag{0.46}$$

with $\bar{h}(x)/2 \leq 2^{-\bar{j}(x)} \leq \bar{h}(x)$. We decompose the error into the two different kinds of errors

$$\mathbb{E}_p^{\otimes n}\, d_\Delta\left(\Gamma_p, \hat{\Gamma}_n\right) = E_1 + E_2$$

with

$$\begin{aligned} E_1 &:= \mathbb{E}_p^{\otimes n}\, \lambda^d \left(\hat{\Gamma}_n \setminus \Gamma_p\right) \\ E_2 &:= \mathbb{E}_p^{\otimes n}\, \lambda^d \left(\Gamma_p \setminus \hat{\Gamma}_n\right) \end{aligned} \tag{0.47}$$

and start with E_1. We split E_1 again

$$E_1 = E_{1,1} + E_{1,2}$$

with

$$\begin{aligned} E_{1,1} &:= \mathbb{E}_p^{\otimes n}\left[\lambda^d\Big(x \in \mathbb{R}^d \ :\ \hat{p}_{n,\hat{j}}(x) \geq \alpha_n,\ p(x) = 0,\ \hat{j} \geq \bar{j}\Big)\right] \\ E_{1,2} &:= \mathbb{E}_p^{\otimes n}\left[\lambda^d\Big(x \in \mathbb{R}^d \ :\ \hat{p}_{n,\hat{j}}(x) \geq \alpha_n,\ p(x) = 0,\ \hat{j} < \bar{j}\Big)\right]. \end{aligned} \tag{0.48}$$

We start with $E_{1,1}$. Since

$$2^{-\bar{j}(x)} \leq \bar{h}(x) = c_{11}(\beta, L)\left(\frac{\log n}{n}\right)^{\frac{1}{\beta+d}} =: \delta_n$$

uniformly for all x with $p(x) = 0$, it follows

$$\left\{x \in \mathbb{R}^d \,:\, \hat{p}_{n,\hat{j}}(x) \geq \alpha_n,\ p(x) = 0,\ \hat{j} \geq \bar{j}\right\} \ \subset\ \bigcup_{i=1}^n \left(\mathcal{B}_{\delta_n}(X_i) \setminus \Gamma_p\right)$$

with $\mathcal{B}_r(x)$ again denoting the Euclidean ball with radius r around x. The support Γ_p is assumed to satisfy the complexity condition 4.3 to the exponent $\mu = \gamma\beta$. Note that $\delta_n \leq \xi_1$ for sufficiently large $n \geq n_0(\xi_1)$. We denote by Γ_{1,δ_n} and Γ_{2,δ_n} the related disjoint decomposition of Γ_p. Then,

$$
\begin{aligned}
E_{1,1} \;\leq\;\; & \mathbb{E}_p^{\otimes n}\, \lambda^d \left(\bigcup_{i\,:\,X_i \in \Gamma_{1,\delta_n}} \left(\mathcal{B}_{\delta_n}(X_i) \setminus \Gamma_p\right) \right) \\
& + \mathbb{E}_p^{\otimes n}\, \lambda^d \left(\bigcup_{i\,:\,X_i \in \Gamma_{2,\delta_n} \cap \{p \leq \alpha_n\}} \left(\mathcal{B}_{\delta_n}(X_i) \setminus \Gamma_p\right) \right) \\
& + \mathbb{E}_p^{\otimes n}\, \lambda^d \left(\bigcup_{i\,:\,X_i \in \Gamma_{2,\delta_n} \cap \{p > \alpha_n\}} \left(\mathcal{B}_{\delta_n}(X_i) \setminus \Gamma_p\right) \right).
\end{aligned} \tag{0.49}
$$

The expectation of the first Lebesgue volume is immediately controlled by the complexity condition

$$
\mathbb{E}_p^{\otimes n}\, \lambda^d \left(\bigcup_{i\,:\,X_i \in \Gamma_{1,\delta_n}} \left(\mathcal{B}_{\delta_n}(X_i) \setminus \Gamma_p\right) \right) \;\leq\; \lambda^d \left(\Gamma_{1,\delta_n}^{\delta_n} \setminus \Gamma_{1,\delta_n} \right) \;\leq\; \xi_2\, \delta_n^{\gamma\beta}.
$$

The expectation of the second Lebesgue volume is also controlled by the complexity condition

$$
\begin{aligned}
& \mathbb{E}_p^{\otimes n}\, \lambda^d \left(\bigcup_{i\,:\,X_i \in \Gamma_{2,\delta_n} \cap \{p \leq \alpha_n\}} \left(\mathcal{B}_{\delta_n}(X_i) \setminus \Gamma_p\right) \right) \\
& \qquad \leq\; \sum_{i=1}^{n} \mathbb{E}_p \Big[V_d\, \delta_n^d \cdot \mathbb{1}\Big\{ X_i \in \Gamma_{2,\delta_n} \cap \{p \leq \alpha_n\} \Big\} \Big] \\
& \qquad =\; V_d\, n\, \delta_n^d \cdot \mathbb{P}\left(X_1 \in \Gamma_{2,\delta_n} \cap \{p \leq \alpha_n\} \right) \\
& \qquad \leq\; V_d\, n\, \delta_n^d\, \alpha_n \cdot \lambda^d \left(\Gamma_{2,\delta_n} \right) \\
& \qquad \leq\; c(\beta, L)\, \xi_2\, \delta_n^{\gamma\beta} (\log n)^2,
\end{aligned}
$$

where V_d again denotes the volume of the d-dimensional unit ball. Considering the third expectation in (0.49), let z be some point with $p(z) > \alpha_n$ and y any point in the open set Γ_p^c. Then p is constant zero in a neighborhood of y, i.e. all derivatives are zero in y and thus

$$
\alpha_n \;<\; p(z) \;=\; \left| p(z) - P_{y,\lfloor \beta \rfloor}^{(p)}(z) \right| \;\leq\; L \|z - y\|_2^{\beta}
$$

for all $y \in \Gamma_p^c$. If $\|z - y\|_2$ was smaller than δ_n, this inequality contradicts the choice of the offset level (4.3) for $n \geq n_1$ with n_1 depending on β and L only. Hence, the subset of Γ_p where p exceeds the offset level is contained in $\Gamma_p^{-\delta_n}$ and consequently $\mathcal{B}_{\delta_n}(z) \subset \Gamma_p$. Therefore, the third expectation vanishes and finally

$$
E_{1,1} \;\leq\; c(\beta, L, \xi_2) \left(\frac{\log n}{n} \right)^{\frac{\gamma\beta}{\beta + d}} (\log n)^2.
$$

Regarding $E_{1,2}$ in (0.48), only the points $x \in \mathbb{R}^d$ that belong to $(\bigcup_{i=1}^n \mathcal{B}_{\delta_n}(X_i))^c \setminus \Gamma_p$ have to be considered. Otherwise, the point is contained in $\bigcup_{i=1}^n (B_{\delta_n}(X_i) \setminus \Gamma_p)$ and we proceed as before in (0.49). Note, that for $x \in (\bigcup_{i=1}^n B_{\delta_n}(X_i))^c \setminus \Gamma_p$ both $\hat{p}_{n,\bar{j}}(x)$ and $\tilde{\sigma}_x^2(\bar{j})$ vanish. According to Lepski's selection rule, see (3.7), and Lemma 5.4,

$$\begin{aligned}
\hat{p}_{n,\hat{j}}(x) &= |\hat{p}_{n,\hat{j}}(x) - \hat{p}_{n,\bar{j}}(x)| \\
&\leq c_{14}\sqrt{\hat{\sigma}_x^2(\bar{j}) \log n} \\
&\leq c_{14}\frac{\sqrt{\log^3 n}}{n(2^{-\bar{j}})^d} \\
&\leq \frac{2^d c_{14}}{c_{11}(\beta, L)^d}\left(\frac{\log n}{n}\right)^{\frac{\beta}{\beta+d}} \sqrt{\log n}.
\end{aligned}$$

For sufficiently large $n \geq n_2$ with n_2 depending on β and L (precisely β^* and L^*) only, $\hat{p}_{n,\hat{j}}(x)$ cannot exceed α_n for $x \in (\bigcup_{i=1}^n B_{\delta_n}(X_i))^c \setminus \Gamma_p$ and hence the expectation $E_{1,2}$ provides the same bound

$$E_{1,2} \;\leq\; c(\beta, L, \xi_2)\left(\frac{\log n}{n}\right)^{\frac{\gamma\beta}{\beta+d}} (\log n)^2.$$

The second part of the proof is partially based on the proof for density level sets of Rigollet and Vert (2009). The second type of error E_2 in (0.47) has to be estimated. Since $\Gamma_p \setminus \hat{\Gamma}_n$ is a subset of the support Γ_p, Jensen's inequality and Lemma 0.2 provide

$$E_2 \;\leq\; c_{18}(\kappa, \gamma)\left(\mathbb{E}_p^{\otimes n}\left[\int_{\Gamma_p \setminus \hat{\Gamma}_n} p(x) d\lambda^d(x)\right]\right)^{\frac{\gamma}{1+\gamma}}.$$

Furthermore, the support can be decomposed as follows

$$\Gamma_p = \bigcup_{q \geq 0} \chi_q,$$

where

$$\begin{aligned}
\chi_0 &:= \overline{\{x \in \mathbb{R}^d : 0 < p(x) \leq 2\alpha_n\}} \\
\chi_q &:= \left\{x \in \mathbb{R}^d : 2^q\alpha_n < p(x) \leq 2^{q+1}\alpha_n\right\}, \quad q \geq 1.
\end{aligned}$$

Then,

$$E_2 \;\leq\; c_{18}(\kappa, \gamma)\left(\sum_{q \geq 0} E_{2,q}\right)^{\frac{\gamma}{1+\gamma}}$$

with

$$E_{2,q} := \mathbb{E}_p^{\otimes n}\left[\int_{\chi_q} p(x) \cdot \mathbb{1}\left\{\hat{p}_{n,\hat{j}}(x) < \alpha_n\right\} d\lambda^d(x)\right].$$

For $x \in \chi_0$ we estimate $p(x)$ from above and use the margin condition such that

$$E_{2,0} \ \leq \ c(\beta, L, \gamma, \kappa) \cdot \alpha_n^{1+\gamma}.$$

For $q \geq 1$, we distinguish between the error of stopping too late and stopping too early, leading to the following decomposition

$$\begin{aligned} E_{2,q} &= \int_{\chi_q} p(x)\Big(\mathbb{P}^{\otimes n}(A_{x,1}) + \mathbb{P}^{\otimes n}(A_{x,2})\Big) d\lambda^d(x) \\ &\leq \int_{\chi_q} p(x)\Big(\mathbb{P}^{\otimes n}(A_{x,1} \cap B_x) + \mathbb{P}^{\otimes n}(B_x^c) + \mathbb{P}^{\otimes n}(A_{x,2})\Big) d\lambda^d(x) \end{aligned} \tag{0.50}$$

with

$$\begin{aligned} A_{x,1} &:= \Big\{\hat{p}_{n,\hat{j}}(x) < \alpha_n \ \cap \ \{\hat{j} \leq \bar{j}\}\Big\} \\ A_{x,2} &:= \Big\{\hat{p}_{n,\hat{j}}(x) < \alpha_n \ \cap \ \{\hat{j} > \bar{j}\}\Big\} \\ B_{x,j} \ &:= \left\{\left|\frac{\tilde{\sigma}^2_{x,\mathrm{trunc}}(j)}{\sigma^2_{x,\mathrm{trunc}}(j)} - 1\right| \leq \frac{1}{2}\right\}. \end{aligned}$$

We start with the first probability in (0.50)

$$\begin{aligned} &\mathbb{P}^{\otimes n}(A_{x,1} \cap B_{x,\bar{j}}) \cdot \mathbb{1}\{x \in \chi_q\} \\ &\quad \leq \mathbb{P}^{\otimes n}\left(|\hat{p}_{n,\hat{j}}(x) - p(x)| > (2^q - 1)\,\alpha_n \ \cap \ \{\hat{j} \leq \bar{j}\} \ \cap \ B_{x,\bar{j}}\right) \cdot \mathbb{1}\{x \in \chi_q\}, \\ &\quad \leq \mathbb{P}^{\otimes n}\left(|\hat{p}_{n,\bar{j}}(x) - p(x)| + |\hat{p}_{n,\hat{j}}(x) - \hat{p}_{n,\bar{j}}(x)| > 2^{q-1}\alpha_n \cap \{\hat{j} \leq \bar{j}\} \cap B_{x,\bar{j}}\right) \cdot \mathbb{1}\{x \in \chi_q\} \end{aligned}$$

and the construction of Lepski's method controls the second term $|\hat{p}_{n,\hat{j}}(x) - \hat{p}_{n,\bar{j}}(x)|$, see (3.7). This yields, together with a bias variance decomposition

$$\begin{aligned} &\mathbb{P}^{\otimes n}(A_{x,1} \cap B_{x,\bar{j}}) \cdot \mathbb{1}\{x \in \chi_q\} \\ &\quad \leq \mathbb{P}^{\otimes n}\Bigg(|\hat{p}_{n,\bar{j}}(x) - \mathbb{E}_p^{\otimes n}\hat{p}_{n,\bar{j}}(x)| > 2^{q-1}\alpha_n \\ &\qquad\qquad - c_{14}\sqrt{\hat{\sigma}_x^2(\bar{j})\log n} - |b_x(\bar{j})| \ \cap \ B_{x,\bar{j}}\Bigg) \cdot \mathbb{1}\{x \in \chi_q\} \\ &\quad \leq \mathbb{P}^{\otimes n}\Bigg(|\hat{p}_{n,\bar{j}}(x) - \mathbb{E}_p^{\otimes n}\hat{p}_{n,\bar{j}}(x)| > 2^{q-1}\alpha_n \\ &\qquad\qquad - \left(\sqrt{3/2}\,c_{14} + c_{15}(\beta, L)\right)\sqrt{\sigma^2_{x,\mathrm{trunc}}(\bar{j})\log n}\Bigg) \cdot \mathbb{1}\{x \in \chi_q\}, \end{aligned}$$

where we used the definition of $B_{x,\bar{j}}$ and Lemma 5.4 in the second step. The lemma also yields for $x \in \chi_q$, $q \geq 1$

$$\sqrt{\sigma^2_{x,\mathrm{trunc}}(\bar{j})\log n} \leq c_{21}(\beta, L) \cdot \left\{\left(\frac{\log n}{n}\right)^{\frac{\beta}{\beta+d}} \vee \left(\frac{p(x)\log n}{n}\right)^{\frac{\beta}{2\beta+d}}\right\}\sqrt{\log n}$$

$$
\begin{aligned}
&\le c_{21}(\beta,L)\cdot\left\{\left(\frac{\log n}{n}\right)^{\frac{\beta}{\beta+d}}\vee\left(\frac{2^{q+1}\alpha_n\log n}{n}\right)^{\frac{\beta}{2\beta+d}}\right\}\sqrt{\log n}\\
&\le 2^{(q+1)/2}\,c_{21}(\beta,L)\,c_5(\beta,L)^{-\frac{\beta+d}{2\beta+d}}\,\alpha_n,
\end{aligned}
$$

such that

$$
\begin{aligned}
&\mathbb{P}^{\otimes n}(A_{x,1}\cap B_{x,\bar{j}})\cdot\mathbb{1}\{x\in\chi_q\}\\
&\qquad\le\mathbb{P}^{\otimes n}\Bigg(\frac{|\hat{p}_{n,\bar{j}}(x)-\mathbb{E}_p^{\otimes n}\hat{p}_{n,\bar{j}}(x)|}{\sqrt{\sigma^2_{x,\mathrm{trunc}}(\bar{j})\log n}}\\
&\qquad\qquad\qquad> 2^{q-1}2^{-(q+1)/2}\frac{c_5(\beta,L)^{\frac{\beta+d}{2\beta+d}}}{c_{21}(\beta,L)}-\left(\sqrt{3/2}c_{14}+c_{15}(\beta,L)\right)\Bigg)\\
&\qquad\le\mathbb{P}^{\otimes n}\Bigg(\frac{|\hat{p}_{n,\bar{j}}(x)-\mathbb{E}_p^{\otimes n}\hat{p}_{n,\bar{j}}(x)|}{\sqrt{\sigma^2_{x,\mathrm{trunc}}(\bar{j})\log n}}\\
&\qquad\qquad\qquad> 2^{q/2-3/2}\left[\frac{c_5(\beta,L)^{\frac{\beta+d}{2\beta+d}}}{c_{21}(\beta,L)}-2^{3/2}\left(\sqrt{3/2}c_{14}+c_{15}(\beta,L)\right)\right]\Bigg)\\
&\qquad\le\mathbb{P}^{\otimes n}\left(\frac{|\hat{p}_{n,\bar{j}}(x)-\mathbb{E}_p^{\otimes n}\hat{p}_{n,\bar{j}}(x)|}{\sqrt{\sigma^2_{x,\mathrm{trunc}}(\bar{j})\log n}}>2^{q/2}\right)
\end{aligned}
$$

by definition of $c_5(\beta,L)$, and thus

$$
\mathbb{P}^{\otimes n}(A_{x,1}\cap B_{x,\bar{j}})\cdot\mathbb{1}\{x\in\chi_q\}\;\le\;2\exp\left(-\frac{\log n}{4}\,2^{q/2}\right) \tag{0.51}
$$

by Lemma 5.5. Furthermore, Lemma 5.3 can be used to bound the probability of B_x^c by

$$
\mathbb{P}^{\otimes n}(B^c_{x,\bar{j}})\;\le\;2\exp\left(-\frac{3}{32\|K\|^2_{\sup}}\log^2 n\right). \tag{0.52}
$$

A sufficiently tight bound on the probability of $A_{x,2}$ in inequality (0.50) is required. By definition of $A_{x,2}$,

$$
\begin{aligned}
\mathbb{P}^{\otimes n}(A_{x,2})&\le\mathbb{P}^{\otimes n}(\hat{j}>\bar{j})\\
&\le\sum_{m\in\mathcal{J}\,:\,m>\bar{j}}\mathbb{P}^{\otimes n}\left(|\hat{p}_{n,\bar{j}}(t)-\hat{p}_{n,m}(t)|>c_4\sqrt{\hat{\sigma}^2_t(m)\log n}\right),
\end{aligned}
$$

and we divide the absolute value of the difference of the kernel density estimators into the difference of bias terms $|b_t(\bar{j})-b_t(m)|$ and two stochastic terms $|\hat{p}_{n,\bar{j}}(t)-\mathbb{E}_p^{\otimes n}\hat{p}_{n,\bar{j}}(t)|$ and $|\hat{p}_{n,m}(t)-\mathbb{E}_p^{\otimes n}\hat{p}_{n,m}(t)|$. By Lemma 5.4,

$$
|b_t(\bar{j})-b_t(m)|\;\le\;2B_t(\bar{j})\;\le\;2c_{20}(\beta,L)\sqrt{\sigma^2_{x,\mathrm{trunc}}(\bar{j})\log n}.
$$

Furthermore, $x \in \chi_q$ for some $q \geq 1$ and therefore $p(x) \geq 2\alpha_n$. Consequently, for $n \geq n_2$ with n_2 depending on β and L only,

$$p(x) \geq \left(\frac{\log n}{n}\right)^{\frac{\beta}{\beta+d}},$$

such that $\bar{h}$ as defined in (0.46) satisfies

$$\bar{h} \leq c_{11}(\beta, L)\, p(x)^{\frac{1}{\beta}}.$$

Lemma 5.1 (i) then yields for $m > \bar{j}$

$$\sigma^2_{x,\mathrm{trunc}}(\bar{j}) \leq 3\,\sigma^2_{x,\mathrm{trunc}}(m).$$

Thus, we can follow the arguments in the proof of Theorem 3.3 line by line straightforwardly, and arrive as in (5.15) and (5.18) at

$$\mathbb{P}^{\otimes n}(A_{x,2}) \;\leq\; \exp\Big(- \tilde{c}_{16}(\beta, L) \log n + \log\big(|\mathcal{J}|\big)\Big) \tag{0.53}$$

with a constant $\tilde{c}_{16}(\beta, L)$ that again, as $c_{16}(\beta, L)$, is monotonously increasing in c_{14}. Via $\tilde{c}_{16}(\beta, L)$ this requires some further restriction on the lower bound on c_{14}, namely such that

$$\tilde{c}_{16}(\beta, L) \geq 2\Big(1 + \frac{1}{\beta}\Big)\frac{\beta}{\beta+d},$$

which implies in particular $\tilde{c}_{16}(\beta, L) > 2(1+\gamma)\beta/(\beta+d)$. Plugging now the bounds (0.51), (0.52) and (0.53) into (0.50) and applying the margin condition, we arrive at

$$\begin{aligned}
E_{2,q} &\leq \int_{\chi_q} p(x)\Big(\mathbb{P}^{\otimes n}(A_{x,1} \cap B_x) + \mathbb{P}^{\otimes n}(B_x^c) + \mathbb{P}^{\otimes n}(A_{x,2})\Big) d\lambda^d(x) \\
&\leq (2^{q+1}\alpha_n)\Big\{ 2\exp\left(-\frac{\log n}{4}\, 2^{(q+1)/2}\right) + 2\exp\left(-\frac{3}{32\|K\|^2_{\sup}}\log^2 n\right) \\
&\qquad\qquad + 2\exp\Big(- \tilde{c}_{16}(\beta, L)\log n + \log\big(|\mathcal{J}|\big)\Big)\Big\} \int_{\chi_q} d\lambda^d(x) \\
&\leq \kappa_2 (2^{q+1}\alpha_n)^{1+\gamma}\Big\{ 2\exp\left(-\frac{\log n}{4}\, 2^{(q+1)/2}\right) + 2\exp\left(-\frac{3}{32\|K\|^2_{\sup}}\log^2 n\right) \\
&\qquad\qquad + 2\exp\Big(- \tilde{c}_{16}(\beta, L)\log n + \log\big(|\mathcal{J})|\big)\Big\} \\
&\leq \alpha_n^{1+\gamma}\cdot 2\kappa_2 \Big\{ \exp\left(-\frac{\log n}{4}\, 2^{(q+1)/2} + (q+1)(1+\gamma)\log 2\right) \\
&\qquad\qquad + \exp\left(-\frac{3}{32\|K\|^2_{\sup}}\log^2 n + (q+1)(1+\gamma)\log 2\right) \\
&\qquad\qquad + \exp\Big(- \tilde{c}_{16}(\beta, L)\log n + (q+1)(1+\gamma)\log 2 + \log\big(|\mathcal{J}|\big)\Big)\Big\}
\end{aligned}$$

$$=: \alpha_n^{1+\gamma}\Delta_{n,q}(\beta, L, \gamma, \kappa).$$

Since densities $p \in \mathscr{P}_d(\beta, L, \gamma, \kappa, \xi)$ are uniformly bounded by $c_1(\beta, L)$, χ_q is empty as soon as

$$q > q_{\max} = \frac{\log\big(c_1(\beta, L)/\alpha_n\big)}{\log 2},$$

whence

$$E_2 \leq c(\beta, L, \gamma, \kappa) \cdot \alpha_n^{\gamma} \cdot \left(\sum_{q=0}^{q_{\max}} \Delta_{n,q}(\beta, L, \gamma, \kappa)\right)^{\frac{\gamma}{\gamma+1}} \leq c(\beta, L, \gamma, \kappa) \cdot \alpha_n^{\gamma}.$$

□

A.4. (β, L)-regular kernels. The proofs of Theorem 3.1, Theorem 3.4 and Theorem 4.4 make use of the following specific constructions of functions with prescribed Hölder regularity (β, L). The first construction, which is appealing due to its simplicity, is taken from Rigollet and Vert (2009). Note that it works for $\beta \leq 2$ only because the second derivative is not continuous. Define the function $K : \mathbb{R}^d \to \mathbb{R}$ by

$$K(x;\beta) := c_{17}(\beta) \begin{cases} (1-\|x\|_2)_+^{\beta} & \text{, if } \beta \leq 1 \\ \begin{cases} 2^{1-\beta} - \|x\|_2^{\beta} & \text{, if } \|x\|_2 \leq \frac{1}{2} \\ (1-\|x\|_2)_+^{\beta} & \text{, if } \frac{1}{2} < \|x\|_2 \end{cases} & \text{, if } 1 < \beta \leq 2. \end{cases} \tag{0.54}$$

with the normalizing constant $c_{17}(\beta)$ ensuring that K integrates to one, and $f_+ = \max\{f, 0\}$ the positive part for a real-valued function f.
The second construction is a pointwise convex combination

$$K(x;\beta) := c_{30}(\beta)\Big(w(x)\,(1-\|x\|_2)_+^{\beta} + (1-w(x))\,u(x)\Big) \tag{0.55}$$

with

$$u(x) = \begin{cases} \exp\Big(-\frac{1}{1-\|x\|_2^2}\Big) & \text{if } \|x\|_2 < 1 \\ 0 & \text{if } \|x\|_2 \geq 1 \end{cases}$$

$$w(x) = \begin{cases} \exp\Big(1-\frac{1}{\|x\|_2^2}\Big) & \text{if } \|x\|_2 > 0 \\ 0 & \text{if } \|x\|_2 = 0 \end{cases}$$

and the normalizing constant $c_{30}(\beta)$. The idea behind is that $(1-\|x\|_2)_+^{\beta}$ is dominating when $\|x\|_2$ is close to one, while the choice of u and of the weight w guarantees that $K(\cdot, \beta)$ remains $\lfloor\beta\rfloor$-times differentiable at zero. Note that u has Hölder regularity to the exponent β for every $\beta > 0$, see Tsybakov (2009), Section 2.5.

Lemma 0.3. *For any $\beta > 0$, the kernel $K(\cdot, \beta)$ in (0.55) is Hölder continuous to the exponent β.*

For both constructions, (0.54) and (0.55), the dependence of $K(\cdot,\beta)$ on β is omitted when there is no ambiguity. The function $K(\cdot,\beta)$ is supported on the closed Euclidean unit ball $\mathcal{B}_1(0)$, integrates to one (by definition of $c_{17}(\beta)$ and $c_{30}(\beta)$, respectively) and has Hölder regularity $(\beta,\tilde{L})$ for a constant $\tilde{L} = \tilde{L}(\beta)$. Recall that $K(x;h,\beta) := h^\beta K(x/h;\beta)$ has the same Hölder regularity as K, but does not necessarily integrate to one, whereas $K_h(x;\beta) := h^{-d}K(x/h;\beta)$ is the rescaled kernel having the same Hölder parameter β but not necessarily the same parameter $\tilde{L}$ and is still integrating to one. With the choice

$$g = g_{\beta,L,d} := 1 \vee \left(\frac{\tilde{L}}{L}\right)^{\frac{1}{\beta+d}}, \quad i = 1,\dots,d$$

the function $K_g(x;\beta)$ is supported on $\mathcal{B}_g(0)$ and is contained in $\mathscr{P}_d(\beta,L)$.

In case of anisotropic smoothness we frequently use the product kernel $K = \prod_{i=1}^d K_i$ with factors $K_i = K_{g_{\beta_i,L,1}}$.

Proof of Lemma 0.3. Rewriting

$$K(x;\beta) = c_{30}(\beta)\left(w(x)d(x) + u(x)\right),$$

with $d(x) = (1-\|x\|_2)_+^\beta - u(x)$, it remains to prove that $w\cdot d$ has the Hölder exponent β. For $\beta \le 1$, this is an easy consequence of

$$|w(x)d(x) - w(y)d(y)| \le |w(x)|\cdot|d(x)-d(y)| + |d(y)|\cdot|w(x)-w(y)|$$

because both w and d are uniformly bounded on $\mathcal{B}_1(0)$, while being Hölder continuous to the exponent β.

We now treat the case $\beta > 1$.

Case 1: ($\|x\|_2 \ge 1$ and $\|y\|_2 \ge 1$) Condition (4.1) obviously holds.

Case 2: ($\|x\|_2 \ge 1$ and $\|y\|_2 < 1$) The expression $(w\cdot d)(x)$ equals zero. Using the definition of the derivative via the difference quotient, $w\cdot d$ is shown to be $\lfloor\beta\rfloor$-times continuously differentiable at $y=0$ with value zero. Hence, its Taylor polynomial $P_{0,\lfloor\beta\rfloor}^{(wd)}$ vanishes. If $y \ne 0$, both d and w are $\lfloor\beta\rfloor$-times differentiable in y and

$$\left|(w\cdot d)(x) - P_{y,\lfloor\beta\rfloor}^{(wd)}(x)\right| = \left|P_{y,\lfloor\beta\rfloor}^{(wd)}(x)\right|$$

with the Taylor polynomial

$$P_{y,\lfloor\beta\rfloor}^{(wd)}(x) = \sum_{|n|\le\lfloor\beta\rfloor} \gamma_n(y)(x_1-y_1)^{n_1}\cdots(x_d-y_d)^{n_d} \tag{0.56}$$

and the coefficients

$$\gamma_n(y) = \sum_{k+l=n} \frac{1}{|k|!\,|l|!}\binom{|k|}{k_1,\dots,k_d}\binom{|l|}{l_1,\dots,l_d} D^k w(y) D^l d(y),$$

where
$$\binom{|k|}{k_1,\ldots,k_d} = \frac{|k|!}{k_1!\cdots k_d!}$$
denotes the multinomial coefficient. For $l \in \mathbb{N}^d$ with $1 \le |l| \le \lfloor \beta \rfloor$,

$$D^l d(y) = \beta(\beta-1)\cdots(\beta-|l|+1)\,(1-\|y\|_2)^{\beta-|l|}\prod_{i=1}^{d}(-\operatorname{sgn} y_i)^{l_i}\,\mathbb{1}\{y_i \neq 0\} - D^l u(y)$$

where $D^l u(\cdot)$ is of the form

$$D^l u(y) = u(y)\cdot\sum_{k=1}^{|l|}\frac{Q_k(y)}{(1-\|y\|_2^2)^{k+1}}$$

with polynomials Q_k, $k=1,\ldots,|l|$ in at most d variables of degree at most $|l|$. Using the representation of $1/(1-\|y\|_2^2)$ as geometric series as well as the series expansion of the logarithm, we get

$$\frac{\sqrt{u(y)}}{(1-\|y\|_2)_+^m} = \exp\left(-\frac{1}{2}\frac{1}{1-\|y\|_2^2} - m\log\Big((1-\|y\|_2)_+\Big)\right) \to 0 \tag{0.57}$$

for any $m \in \mathbb{N}$, as $\|y\|_2$ approaches one from below. Hence, there exists a constant $c_{31}(\beta) \in (0,1)$, such that $\sqrt{u(y)} \le (1-\|y\|_2)_+^{\beta-m}$ for all $1 \le m \le \lfloor \beta \rfloor$ and for all y with $\|y\|_2 \ge c_{31}(\beta)$. For any y with $\|y\|_2 < c_{31}(\beta)$,

$$\sqrt{u(y)} \le \frac{1}{\sqrt{\exp(1)}} \le \frac{1}{(1-c_{31}(\beta))^\beta\sqrt{\exp(1)}}\,(1-\|y\|_2)_+^{\beta-m}$$

for any $1 \le m \le \lfloor \beta \rfloor$. Summarizing,

$$\begin{aligned}|D^l u(y)| &\le \left(\frac{1}{(1-c_{31}(\beta))^\beta\sqrt{\exp(1)}} \vee 1\right)(1-\|y\|_2)_+^{\beta-|l|}\left(\sqrt{u(y)}\cdot\sum_{k=1}^{|l|}\frac{|Q_k(y)|}{(1-\|y\|_2^2)_+^{k+1}}\right)\\ &\le C\left(\frac{1}{(1-c_{31}(\beta))^\beta\sqrt{\exp(1)}} \vee 1\right)(1-\|y\|_2)_+^{\beta-|l|},\end{aligned}$$

with a constant

$$C = \sup_{y:\|y\|_2<1}\left(\sqrt{u(y)}\cdot\sum_{k=1}^{|l|}\frac{|Q_k(y)|}{(1-\|y\|_2^2)_+^{k+1}}\right),$$

which is due to (0.57) and the uniform boundedness of the polynomials Q_k, $k=1,\ldots,|l|$, on the closed unit ball. Finally,

$$c_d = \sup_{y:0<\|y\|_2<1}\frac{|D^l d(y)|}{(1-\|y\|_2)_+^{\beta-|l|}} < \infty.$$

Furthermore,

$$c_w = \max_{k:1\le|k|\le\lfloor\beta\rfloor} \sup_{y:\|y\|<1} |D^k w(y)| < \infty,$$

because the partial derivatives are products of $w(y)$ and rational functions with a pole in zero only. Summarizing,

$$\begin{aligned}|\gamma_n(y)| &\le \sum_{k+l=n} \frac{1}{|k|!\,|l|!}\binom{|k|}{k_1,\dots,k_d}\binom{|l|}{l_1,\dots,l_d} c_d c_w\,(1-\|y\|_2)_+^{\beta-|l|}\\ &\le (1-\|y\|_2)_+^{\beta-|n|}\,c_d c_w \sum_{k+l=n} \frac{1}{|k|!\,|l|!}\binom{|k|}{k_1,\dots,k_d}\binom{|l|}{l_1,\dots,l_d}\\ &\le \|x-y\|_2^{\beta-|n|} c_d c_w \sum_{k+l=n} \frac{1}{|k|!\,|l|!}\binom{|k|}{k_1,\dots,k_d}\binom{|l|}{l_1,\dots,l_d}\end{aligned}$$

as $\|x\|_2 \ge 1$. Together with (0.56), this proves

$$\sup_{x,y:\|x\|_2\ge1,\|y\|_2<1} \frac{\left|(w\cdot d)(x) - P^{(wd)}_{y,\lfloor\beta\rfloor}(x)\right|}{\|x-y\|_2^\beta} = \sup_{x,y:\|x\|_2\ge1,\|y\|_2<1} \frac{|P^{(wd)}_{y,\lfloor\beta\rfloor}(x)|}{\|x-y\|_2^\beta} < \infty.$$

Case 3: ($\|x\|_2 < 1$ and $\|y\|_2 \ge 1$) The Taylor polynomial $P^{(wd)}_{y,\lfloor\beta\rfloor}(x)$ vanishes, such that

$$\left|(w\cdot d)(x) - P^{(wd)}_{y,\lfloor\beta\rfloor}(x)\right| = |(w\cdot d)(x)| \le |d(x)| \le (1-\|x\|_2)_+^\beta + u(x).$$

By the same arguments as in (0.57) and below, there exists a constant $c_{32}(\beta) \in (0,1)$, such that $u(x) \le (1-\|x\|_2)_+^\beta$ and hence

$$\left|(w\cdot d)(x) - P^{(wd)}_{y,\lfloor\beta\rfloor}(x)\right| \le 2\,(1-\|x\|_2)_+^\beta$$

for all x with $\|x\|_2 \ge c_{32}(\beta)$. For x with $\|x\|_2 < c_{32}(\beta)$, note first that $w\cdot d$ is by triangle inequality uniformly bounded from above by $1+\exp(-1)$, implying

$$\left|(w\cdot d)(x) - P^{(wd)}_{y,\lfloor\beta\rfloor}(x)\right| < \frac{1+\exp(-1)}{(1-c_{32}(\beta))^\beta}(1-\|x\|_2)_+^\beta$$

for all x with $\|x\|_2 < c_{31}(\beta)$. Summarizing,

$$\left|(w\cdot d)(x) - P^{(wd)}_{y,\lfloor\beta\rfloor}(x)\right| \le c\,(1-\|x\|_2)_+^\beta \le c\,(\|y\|_2 - \|x\|_2)^\beta \le c\,\|x-y\|_2^\beta$$

for

$$c = \max\left\{2, \frac{1+\exp(-1)}{(1-c_{31}(\beta))^\beta}\right\}.$$

Case 4: ($\|x\|_2 < 1$ and $\|y\|_2 < 1$) If $y = 0$, then $P^{(wd)}_{y,\lfloor\beta\rfloor}$ vanishes and it remains to note that

$$\sup_{x:0<\|x\|_2<1} \frac{|(wd)(x)|}{\|x\|_2^\beta} < \infty.$$

From now on we assume $y \neq 0$. By the triangle inequality

$$\left|(w \cdot d)(x) - P^{(wd)}_{y,\lfloor\beta\rfloor}(x)\right| \tag{0.58}$$

$$\leq \|w\|_{\sup} \left|d(x) - P^{(d)}_{y,\lfloor\beta\rfloor}(x)\right| + \|P^{(d)}_{y,\lfloor\beta\rfloor}(\cdot)\|_{\sup} \left|w(x) - P^{(w)}_{y,\lfloor\beta\rfloor}(x)\right|$$

$$+ \left|P^{(w)}_{y,\lfloor\beta\rfloor}(x) P^{(d)}_{y,\lfloor\beta\rfloor}(x) - P^{(wd)}_{y,\lfloor\beta\rfloor}(x)\right| \tag{0.59}$$

with the Taylor polynomial $P^{(wd)}_{y,\lfloor\beta\rfloor}$ as defined in (0.56). As concerns the product

$$P^{(w)}_{y,\lfloor\beta\rfloor}(x) \cdot P^{(d)}_{y,\lfloor\beta\rfloor}(x) = \sum_{|n| \leq 2\lfloor\beta\rfloor} \tilde{\gamma}_n(y)(x_1 - y_1)^{n_1} \cdots (x_d - y_d)^{n_d},$$

we have $\gamma_n(y) = \tilde{\gamma}_n(y)$ for $0 \leq |n| \leq \lfloor\beta\rfloor$. For $\lfloor\beta\rfloor < |n| \leq 2\lfloor\beta\rfloor$, the coefficient $\tilde{\gamma}_n$ is a linear combination of products $D^j w(y) D^{j'} d(y)$ with $0 \leq j, j' \leq \lfloor\beta\rfloor$, which are uniformly bounded in $y \in \mathbb{R}^d \setminus \{0\}$. For d, this can be easily checked via the explicit form of the derivatives (note that d is not differentiable at zero), while w is $\lfloor\beta\rfloor$-times continuously differentiable and of compact support. Hence, (0.59) is bounded by

$$\sum_{\lfloor\beta\rfloor < |n| \leq 2\lfloor\beta\rfloor} \sup_{0 < \|z\|_2 < 1} |\tilde{\gamma}_n(z)| \max_{i=1,\ldots,d} |x_i - y_i|^{|n|}$$

$$\leq \sum_{\lfloor\beta\rfloor < |n| \leq 2\lfloor\beta\rfloor} \sup_{0 < \|z\|_2 < 1} |\tilde{\gamma}_n(z)| \, 2^{|n|} \left(\frac{\|x - y\|_2}{2}\right)^{|n|}$$

$$\leq \left(\frac{\|x - y\|_2}{2}\right)^{\beta} \sum_{\lfloor\beta\rfloor < |n| \leq 2\lfloor\beta\rfloor} \sup_{0 < \|z\|_2 < 1} |\tilde{\gamma}_n(z)| \, 2^{2\lfloor\beta\rfloor}$$

$$= \|x - y\|_2^{\beta} \sum_{\lfloor\beta\rfloor < |n| \leq 2\lfloor\beta\rfloor} \sup_{0 < \|z\|_2 < 1} |\tilde{\gamma}_n(z)| \, 2^{2\lfloor\beta\rfloor - \beta}.$$

Therefore, (0.58) is bounded by $\|x - y\|_2^{\beta}$ times a constant, the latter uniformly bounded in $x, y \in \mathcal{B}_1(0)$.

Combining Case 1-4 implies that $wd \in \mathscr{H}^{iso}_d(\beta, \tilde{L})$ for some constant $\tilde{L} > 0$. □

Fakultät für Mathematik
Ruhr-Universität Bochum
44780 Bochum
Germany
E-mail: tim.patschkowski@ruhr-uni-bochum.de
angelika.rohde@ruhr-uni-bochum.de