DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen"

Adaptation to lowest density regions with application to support recovery

T. Patschkowski, A. Rohde

Preprint 170

Edited by

AG Numerik/Optimierung
Fachbereich 12 - Mathematik und Informatik
Philipps-Universität Marburg
Hans-Meerwein-Str.
35032 Marburg

DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen”

Adaptation to lowest density regions with application to support recovery

T. Patschkowski, A. Rohde

Preprint 170

The consecutive numbering of the publications is determined by their chronological order.

The aim of this preprint series is to make new research rapidly available for scientific discussion. Therefore, the responsibility for the contents is solely due to the authors. The publications will be distributed by the authors.

ADAPTATION TO LOWEST DENSITY REGIONS WITH APPLICATION TO SUPPORT RECOVERY

By Tim Patschkowski and Angelika Rohde

Ruhr-Universität Bochum

In the context of pointwise density estimation, we introduce the notion of best possible individual minimax speed of convergence over Hölder classes $\mathscr{P}_d(\beta, L)$, leading to a new uniform risk criterion, weighted according to the inverse of this individual speed of convergence. While it bears similarities to the oracle approach, our concept keeps the classical notion of minimax optimality and is not restricted to a prespecified class of estimators. The individual speed of convergence is getting faster as the density approaches zero, being substantially faster than the classical minimax optimal rate. It splits into two regimes depending on the value of the density. An estimator $\hat{p}_n$ tailored to the new risk criterion is constructed such that

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

is bounded, uniformly over a range of parameters (β, L). It provably attains pointwise rates of convergence $\Psi_n(p(t), \beta, L)$ which can be substantially faster than $n^{-1/2}$. We demonstrate 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 principal if the density's regularity is unknown. Consequences on plug-in rules for support recovery based on the new estimator 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. As a by-product, we demonstrate that the rates on support estimation obtained in Cuevas and Fraiman (1997, Ann. Statist.) are always suboptimal in case of Hölder continuous densities.

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

*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, individual minimax optimality, support estimation

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 densities under L_p-loss bears striking differences to the compact case, see Juditsky and Lambert-Lacroix (2004), Renaud-Bouret et al. (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. The crucial fact is that the optimal bandwidth for estimating a density at some fixed point depends not on the local smoothness only but even on the value of the density itself in a neighborhood of the point under consideration. Under prespecified local smoothness, Samworth and Wand (2010) tailor the choice of a bandwidth to the problem of estimation of highest density regions by a plug-in rule. They do not work in a minimax framework but derive appealing asymptotic properties of their bandwidth estimate, essentially under some kind of margin condition, as introduced by Polonik (1995). Goldenshluger and Lepski (2013) develop a bandwidth estimator via some kind of empirical risk minimization, and they prove a pointwise oracle inequality with a remainder term of the order $n^{-1/2}$ 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 principal. 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. To overcome both the drawback of the oracle approach and the drawback of the minimax approach of considering always the worst case within a functional class, a notion of minimax optimal *individual* speed of convergence within the class is defined in the present article. A new type of risk criterion with respect to pointwise loss weighted according to the inverse of this best possible individual speed of convergence is introduced. While it bears similarities to the oracle approach, our criterion keeps the notion of optimality over the functional class, disregarding any preferred class of estimators. We exemplarily study the density estimation framework, where this minimax optimal individual speed of convergence depends not only on the parameters of the functional class under consideration, but also explicitly on the density to be estimated itself. A bandwidth selection rule tailored to the new risk criterion is introduced which provably attains pointwise rates of convergence which can be substantially faster than $n^{-1/2}$. On this way, new 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 principal if the density's regularity is unknown.

To demonstrate the great impact of the improved density estimator on plug-in properties of certain functionals of the density we exemplarily study the problem of density 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. At first glance, this makes the plug-in rule as a by-product of density estimation unfavorable. We derive optimal minimax rates for support recovery, which demonstrate that support recovery is possible with higher accuracy than level set estimation as already conjectured by Tsybakov (1997). We show finally 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.

To recapitulate the support estimation problem, let $X_1, ..., X_n$ be independent random variables, identically distributed according to some unknown probability measure on $\mathbb{R}^d$ with compactly supported continuous Lebesgue density p and support

$$\Gamma_p := \overline{\left\{x \in \mathbb{R}^d : p(x) > 0\right\}}.$$

Here and subsequently, $\bar{A}$ denotes the topological closure of a set $A \subset \mathbb{R}^d$. The classical problem of density support estimation based on the sample $X_1, \ldots, X_n$ has quite a long history in the statistical literature. It is related to cluster analysis, pattern recognition, image recovery and anomaly detection and still enjoys a growing variety of applications. Korostelev and Tsybakov (1993) nicely present a detailed overview and point out important connections to econometrics. Both concerned with the two-dimensional case, 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), Tsybakov (1989, 1991, 1997), Cuevas (1990), Korostelev and Tsybakov (1995), Mammen and Tsybakov (1995), Hall, Nussbaum and Stern (1997), Cuevas

and Fraiman (1997), Klemelä (2004), and Biau, Cadre and Pelletier (2008) to mention just a few. The approaches proposed so far in the literature are related to several shape constraints and regularity assumptions on the support or the density, such as uniform distribution on Γ_p, convexity or star-shapeness of Γ_p, smoothness conditions on the boundary or on the shape function among many others. See for instance Groeneboom (1988), Härdle, Park and Tsybakov (1995), Mammen and Tsybakov (1995), Gayraud (1997), Cuevas and Fraiman (1997), Baíllo, Cuevas and Justel (2000), Cuevas and Rodríguez-Casal (2004), Biau, Cadre, Mason and Pelletier (2009), Aaron (2013) and Brunel (2013) for some results on support estimation in the still growing literature. Recently, Cholaquidis, Cuevas and Fraiman (2014) introduced the so-called cone property as a generalized convexity property to the statistical literature and analyzed a support estimator tailored to this new condition.
But what can be done if we cannot impose such strong and restricting conditions on the support? Which approach is expedient if little is known about the support? Unlike support estimators that profit of these conditions, the plug-in approach is not motivated by any shape constraint. Cuevas and Fraiman (1997) were the first who studied plug-in estimators with off-set, i.e. estimators of the form

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

for some density estimator $\hat{p}_n$ and a sequence $\alpha_n \searrow 0$. They established consistency and upper bounds on the convergence rate under the sharpness or margin exponent condition and smoothness assumptions on the density p with respect to the pseudo-distance of symmetric difference of sets. Their results are not stated in a minimax framework, and optimality of their upper bounds has remained an open problem. As a by-product of our analysis, we demonstrate that the rates on support estimation obtained in Cuevas and Fraiman (1997) are always suboptimal in case of Hölder continuous densities.

The article is organized as follows. Section 2 introduces some basic notations. In Section 3 the notion of minimax optimal individual speed of convergence is introduced, new weighted lower pointwise risk bounds are derived for the density estimation framework, and an adaptive density estimator is proposed. Section 4 addresses the important problem of density support estimation as an example of a functional which benefits greatly from the new density estimator. Super-fast convergence rates at the boundary are deduced in Section 5. A discussion of the obtained results is given afterwards in Section 6 and the proofs are deferred to Section 7.

2. Preliminaries and notation.

All our estimation procedures are based on a sample of n real-valued d-dimensional random vectors $X_i = (X_{i,1}, \ldots, X_{i,d})$, $i = 1, \ldots, n$ (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

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

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

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

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. $K_h(x_1, \ldots, x_d) = \prod_{i=1}^{d} K_{i,h_i}(x_i)$. Its 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 non-negative, 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 $1 \leq j_i \leq k_i$, $i = 1, \ldots, d$, 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 $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ we define the univariate functions

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

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

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

for those $i \in \{1, \ldots, d\}$ with $\beta_i \leq 1$, 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

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

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

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

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

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

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 individual 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 $\psi^n_{p(t),\beta,L}$. 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 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 $\psi^n_{x,\beta} := x \wedge (x/n)^{\frac{\bar\beta}{2\bar\beta+1}}$.

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

$$T_n(t) \;:=\; \hat p_{n,h_{n,\delta}}(t) \cdot \mathbb{1}\big\{\delta \leq n^{-\bar\beta/(\bar\beta+1)}\big\}$$

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 principal. We refer to it in the sequel as individual 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 3.2. The individual 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)	$x \le n^{-\frac{\bar{\beta}}{\bar{\beta}+1}}$	x
(ii)	$n^{-\frac{\bar{\beta}}{\bar{\beta}+1}} < x \le c_1(\beta, L)$	$\left(\frac{x}{n}\right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}}$

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

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

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 individual rate of Theorem 3.1 to the classical minimax rate for different sample sizes from $n = 50$ to $n = 800$.

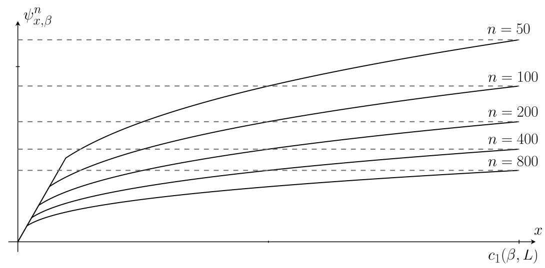

FIG 1. *New lower bound (solid line), Classical lower bound (dashed line)*

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.

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. It raises the question whether this imposes a fundamental limit on the possible range of adaptation. We shall demonstrate in what follows (see Section 5 in particular) 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

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

see (7.8) in Section 7 for details, against the rough variance bound

$$\operatorname{Var}(\hat{p}_{n,h}(t)) \;\leq\; \frac{c_1(\beta, L) \|K\|_2^2}{n \prod_{i=1}^{n} h_i}, \tag{3.1}$$

where $\| \cdot \|_2$ is the Euclidean norm (on $L_2(\lambda^d)$). In contrast to the classical bound (3.1), the sharp convolution bound

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

is able to capture small values of p in a small neighborhood around t and thus leads to suboptimal rates of convergence whenever the density is small. This convolution bound 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, $\tilde{\sigma}_t^2(h)$ is getting extremely unstable if the bandwidth h is small, which is just the important situation at lowest density regions. Precisely, Bernstein's inequality provides the bound

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

which suggests to study the following truncated versions instead

$$
\begin{aligned}
\sigma^2_{t,\mathrm{trunc}}(h) &:= \max\left\{\frac{\log^2 n}{n^2\prod_{i=1}^d h_i^2},\ \sigma_t^2(h)\right\}, \\
\tilde{\sigma}^2_{t,\mathrm{trunc}}(h) &:= \max\left\{\frac{\log^2 n}{n^2\prod_{i=1}^d h_i^2},\ \tilde{\sigma}_t^2(h)\right\}.
\end{aligned} \tag{3.2}
$$

Our approach imitates the bandwidth selection scheme developed in Kerkyacharian, Lepski and Picard (2001) with the following modifications. Firstly, their threshold given by the variance bound in the Gaussian white noise setting is replaced essentially with the truncated estimate in (3.2). 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 resulting estimator $\hat{p}_n$ is rigorously constructed in Appendix B. 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 3.2 (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 (B.3) 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 individual 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.

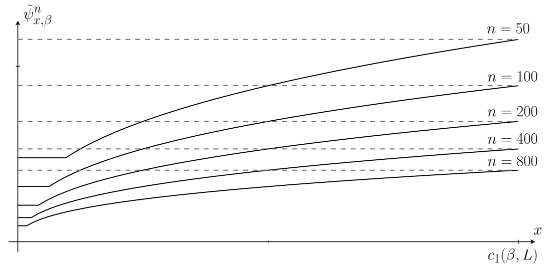

FIG 2. *New upper bound without logarithmic factor (solid line), Classical upper bound (dashed line)*

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 superefficiency result shows that this is indeed the case as long as the density's regularity is unknown.

THEOREM 3.3 (Superefficiency). *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 $p_n \in \mathscr{P}_1(\beta_1, L_1)$ and $q_n \in \mathscr{P}_1(\beta_2, L_2)$ with

$$\frac{n^{-\beta_1/(\beta_1+1)}}{p_n(t)} = o(1) \quad \text{and} \quad \frac{q_n(t)}{n^{-\beta_2/(\beta_2+1)}} = o(1) \tag{3.3}$$

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

$$\mathbb{E}_{p_n}^{\otimes n} |T_n(t) - p_n(t)| \;\leq\; c_3(\beta_1^*, L_1^*, r) \left(\frac{p_n(t)}{n}\right)^{\frac{\beta_1}{2\beta_1+1}} (\log n)^{3/2}, \tag{3.4}$$

there exists an $n_0(\beta_1, \beta_2, L_1, L_2)$ such that

$$\mathbb{E}_{q_n}^{\otimes n} |T_n(t) - q_n(t)| \;\geq\; q_n(t) \cdot \rho(n)$$

for all $n \geq n_0(\beta_1, \beta_2, L_1, L_2)$.

The following consideration provides a heuristic reason why adaptation to regime (i) is not possible in principal. Consider the univariate and Lipschitz continuous triangular density p. If $\delta_n < n^{-\beta/(\beta+1)} = n^{-1/2}$, the expected number of observations in $\{p \leq \delta_n\}$ 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.

REMARK. *While obtaining the classical adaptive rate of convergence in high density regions, the new estimator leads to a substantially improved rate of convergence in lowest density regions and hence demonstrates its superiority as compared to classical adaptive estimators. Indeed, the smaller $p(t)$ is, the better the rate of convergence gets. However, in regime (i) no further improvement can be achieved. Our estimator works without knowledge of the location of lowest density regions and adapts to regime (ii) up to a logarithmic factor and additionally to the unknown regularity. We have shown that simultaneous adaptation to the fastest regime (i) is an unachievable goal whenever the regularity is unknown.*

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 and demonstrate the substantial improvement in the rates of convergence. To this aim, 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 off-set 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 this problem does not arise when using the new density estimator as developed in the previous section instead. As a by-product, the rates of support estimation derived in Cuevas and Fraiman (1997) turn out to be always suboptimal for Hölder continuous densities. 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. That is, the distance between two subsets A and B of $\mathbb{R}^d$ is measured by the pseudo-distance

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

where Δ denotes the symmetric difference of sets

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

We impose the following condition, which controls the slope of the density at the support boundary and therefore characterizes the complexity of the problem and thus the attainable rates of convergence. 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*

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

for all $0 < \varepsilon \leq \kappa_1$*, 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. 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

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

and the closed inner ε-parallel set by $A^{-\varepsilon} := \left((A^c)^\varepsilon\right)^c$.

Definition 4.2 (Blowing-up condition). *The set* A *is said to satisfy the* η*-blowing-up condition with exponent* $BU > 0$ *if*

$$\frac{1}{\eta_2} \leq \frac{\lambda^d(A^\varepsilon \setminus A)}{\varepsilon^{BU}} \leq \eta_2$$

for all $0 < \varepsilon \leq \eta_1$*, where* $\eta = (\eta_1, \eta_2) \in (0, \infty)^2$*.*

The blowing-up condition controls the complexity of a set, i.e. the simpler the structure of a set, the larger is the blowing-up exponent BU. As a consequence of the isoperimetric inequality, any Euclidean ball in $\mathbb{R}^d$ and, more general, any finite union of convex sets in $\mathbb{R}^d$ fulfills the blowing-up condition to the exponent $BU = 1$, cf. Bhattacharya and Rango Rao (1976) and Cuevas and Fraiman (1997). 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

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

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

$$\left| f(x) - P_{y, \lfloor \beta \rfloor}^{(f)}(x) \right| \leq L \, \|x - y\|_2^\beta \qquad \text{for all } x, y \in \mathbb{R}^d,$$

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 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 $\mathscr{H}_d^{iso}(\beta, L)$. In order to be able to speak of optimality in a minimax framework, we first derive a lower bound for the support estimation problem, which has not been deduced in the literature before.

4.1. *Lower risk bounds for support recovery.* The following theorem provides a minimax lower bound for the support estimation problem with respect to the global measure of symmetric difference of sets d_Δ for Hölder continuous densities. Note that the blowing-up exponent BU of a density support cannot be less than $\gamma\beta$ whenever the density is Hölder-smooth to the exponent β, fulfills the margin condition to the exponent γ and has a boundary regular support, i.e.

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

The minimax lower bound is formulated under the assumption of Γ_p fulfilling the following complexity condition, which even slightly weakens the assumption of boundary regularity under the margin condition.

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

$$\frac{\lambda^d(A_{1,\varepsilon}^{\varepsilon} \setminus A_{1,\varepsilon}) \,\vee\, \lambda^d(A_{2,\varepsilon})}{\varepsilon^{\gamma\beta}} \,\leq\, \xi_2,$$

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

Recall that a boundary-regular support of a (β, L)-Hölder-smooth density satisfying the margin condition to the exponent γ fulfills the blowing-up condition with parameter $BU \geq \gamma\beta$. Hence, it also obviously fulfills the complexity condition for the canonical decomposition $A = A \cup \emptyset$. The complexity condition fills the gap between the margin condition 4.1 and the two-sided margin condition

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

which is imposed in the context of density level set estimation for some level $\lambda > 0$, c.f. Rigollet and Vert (2009). In fact, the two-sided margin condition is considerably weaker for $\lambda = 0$. In what follows, $\mathscr{P}_d^{iso}(\beta, L, \gamma, \kappa, \xi)$ denotes the subset of $\mathscr{P}_d^{iso}(\beta, L)$ consisting of densities satisfying the κ-margin condition to the exponent γ and the ξ-complexity condition.

Theorem 4.4 (Minimax lower bound). *For any $\beta \leq 2$ and any margin exponent $\gamma > 0$ with $\gamma\beta \leq 1$, 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*

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

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

This lower bound for the problem of support estimation is substantially smaller than lower bounds of the order

$$\left(\frac{\log n}{n}\right)^{\frac{\gamma\beta}{2\beta+d}}$$

for density level set estimation under the same type of 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.

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 a shrinking neighborhood of lowest density regions, in particular at the support boundary. For the isotropic procedure, the index set $\mathcal{J}$ is just restricted to bandwidths coinciding in all components, and we even simplify the ordering by estimated variances in condition (B.2) "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 7.2 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 $\tilde{K}(0) > 0$. Note that the plug-in support estimator appealingly emerges as a by-product of the density estimator of Section 3. 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

$$\alpha_n := c_5(\beta, L)\left(\frac{(\log n)^{3/2}}{n}\right)^{\frac{\beta}{\beta+d}} \sqrt{\log n} \tag{4.1}$$

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 $\beta \le 2$, $\gamma > 0$ with $\gamma\beta \le 1$ 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*

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

for all $n \ge n_0$.

This minimax upper bound points out that the performance of a density support estimator is specified by the rate of convergence of a density estimator in low density regions. 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). Theorem 4.4 and Theorem 4.5 characterize the minimax-optimal convergence rates for support estimation within a logarithmic factor. Let us point out two consequences. In view of Rigollet and Vert (2009), we have shown that the optimal minimax rates for support estimation are significantly faster than the corresponding rates of level set estimation under the same type of conditions. 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

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

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. Super-fast convergence rates at the support boundary.

So far, our results are restricted to Hölder smoothness to the exponent $\beta \le 2$. As concerns the analysis of our estimator $\hat{p}_n(t)$, the key property is the behavior of p in some $p(t)^{1/\beta}$ neighborhood around $p(t)$. Of course, the less deviation of p around $p(t)$ can be guaranteed, the better is the attainable gain of efficiency. Although the bias is not affected when using higher order kernels, the convolution $((K_h)^2 * p)(t)$ in the variance term may destroy the "low density effect", since K^2 is never of order larger than one. 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. Lemma A.1 (ii) – (iii) reveals that our procedure then even reaches the adaptive speed of convergence

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

(up to a logarithmic factor) under isotropic Hölder smoothness at the support boundary for *every* $\beta > 0$. That is, as $\beta \to \infty$, rates arbitrarily close to n^{-1} can be attained. Define

$$d(\partial\Gamma_p, t) := \inf_{y \in \Gamma_p^c} \|t - y\|_2.$$

THEOREM 5.1 (Super-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 (B.3) satisfies*

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

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 $d(\partial\Gamma_p, t)$) and the maximum norm $\|t - y\|_{\max} = \max_{i=1,\ldots,d} |t_i - y_i|$ are equivalent. As concerns an extension of the results in Section 3 to arbitrary $\beta > 2$, Lemma A.1 (ii) demonstrates that the variance of the kernel density estimator never falls below the reference speed of convergence $\tilde{\psi}_{p(t),\beta}^n$. 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 an individual speed of convergence which does not incorporate the value of the density $p(t)$ only but also information on the derivatives. We leave it open for future research.

6. Discussion.

A question of overriding interest is how fast the convergence of a density estimator can be and how an estimator has to be constructed to attain the optimal speed of convergence. The topic has gained a lot of interest and is still – besides nonparametric regression – one of the most studied problems in statistical literature. In this article, a notion of best possible *individual* minimax speed of convergence within the functional class $\mathscr{P}_d(\beta, L)$ has been developed. As compared to the classical minimax rate of convergence

$$n^{-\bar{\beta}/(2\bar{\beta}+1)}$$

over $\mathscr{P}_d(\beta, L)$, we have shown that each particular density p possesses an individual gain of efficiency for estimation at some point t as compared to the classical minimax rate, and

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

Concerning the logarithmic factor involved in the new rate, we did not put any effort in optimizing its exponent, which is not the smallest possible. In fact, our proof already yields a slightly better dependence in the logarithmic factor, but we preferred to state the bound in the clearest representation. We have indicated in Section 4 that this gain of efficiency is of major importance when the value of the density close to the support boundary is of particular interest. Presumably, similar improvements can be achieved for estimation of extreme quantiles by the plug-in approach, but it requires some additional standardization step as $\hat{p}_n$ does not necessarily integrate to one.
In view of confidence statements about the unknown density, the first question arises whether, up to some additional logarithmic payment, the upper bound in Theorem 3.2 can be established with the supremum over $t \in \mathbb{R}^d$ and expectation interchanged. Such a type of bound reminds of weighted uniform consistency results of the form

$$\sqrt{\frac{nh_n^d}{2|\log h_n^d|}}\left\|\Psi(\cdot)\Big(\hat{p}_n(\cdot)-\mathbb{E}^{\otimes n}\hat{p}_n(\cdot)\Big)\right\|_{\sup} = O_{\mathbb{P}}(1), \tag{6.1}$$

see Giné, Koltchinskii and Zinn (2004) for necessary and sufficient conditions on Ψ and p which guarantee (6.1). The classical case of norming is $\Psi = f^{-\alpha}$ for some $\alpha > 0$, and it is well known that for strictly positive densities of unbounded support only powers of α not exceeding $1/2$ can lead to finite a.s. limits of the expression in (6.1), see Stute (1984), Deheuvels (2000) and Giné and Guillou (2002). In our setting however, $\hat{p}_n$ is a kernel density estimator with randomly selected bandwidth, and the corresponding weight function incorporates an additional dependence on the sample size, leading to the expression

$$\left(\frac{n}{\log n}\right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}}\left\|\Psi_n(\cdot)\Big(\hat{p}_n(\cdot)-\mathbb{E}^{\otimes n}\hat{p}_n(\cdot)\Big)\right\|_{\sup}$$

as well as

$$\left(\frac{n}{\log n}\right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}}\left\|\Psi_n(\cdot)\Big(\hat{p}_n(\cdot)-p(\cdot)\Big)\right\|_{\sup}$$

with weight function

$$\Psi_n(t) = \left(p(t)\vee n^{-\frac{\bar{\beta}}{\bar{\beta}+1}}\right)^{-\frac{1}{2\bar{\beta}+1}}$$

(possibly up to an additional logarithmic factor). The weight function Ψ_n reflects the increased pointwise speed of convergence at lowest density regions as compared to the classical minimax rates on uniform consistency. Recall that the results of Section 3 do not require compactness of the density's support. How such a type of result can conduce as a first step to adaptation to lowest density regions on the level of statistical inference in terms of adaptive confidence bands, and to which extent adaptation to lowest density regions is possible there, is still an open problem and under current investigation by the authors. As an estimated version $\hat{\sigma}_t$ of σ_t is naturally involved in the expression in that case, studying this type of uniform bounds for self-normalizing processes (see de la Peña, Klass and Lai 2004) seems to be promising.

7. Proofs.

We will now turn to the proofs of the main results, stated in Section 3, Section 4 and Section 5.

7.1. *Proofs of Section 3 and Section 5.*

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

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

The two situations

$$
(i)\quad \delta \leq n^{-\frac{\bar\beta}{\bar\beta+1}}, \qquad (ii)\quad n^{-\frac{\bar\beta}{\bar\beta+1}} < \delta \leq c_1(\beta,L)/K_2(0) \tag{7.1}
$$

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

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

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

Denote by h_n the multidimensional bandwidth with components

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

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

$$
\begin{aligned}
n \int &\Big(\sqrt{K_2(x-t; h_n)/2} - \sqrt{K_2(x-t; h_n)} \Big)^2 \, d\lambda^d(x) \\
&+ n \int \Big(\sqrt{K_1(x - s_n) - K_2(x - s_n; h_n)/2} \\
&\qquad - \sqrt{K_1(x - s_n) - K_2(x - s_n; h_n)} \Big)^2 \, d\lambda^d(x)
\end{aligned}
$$

$$\begin{aligned} &\leq \; n\int K_2(x;h_n)\, d\lambda^d(x) \\ &= \; n\,\delta^{\frac{\bar\beta+1}{\bar\beta}}, \end{aligned}$$

where the inequality is due to

$$\left(\sqrt{x-y/2}-\sqrt{x-y}\right)^2 \leq \frac{y}{2} \qquad \text{for all } 0\leq y\leq x. \tag{7.2}$$

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

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

In case (7.1) (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

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

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 $\kappa_2 > 0$,

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

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

$$K_1(t - s_n') = \frac{3}{4}\delta K_2(0) - \frac{1}{4}K_2(0)\left(\frac{\delta}{n}\right)^{\frac{\bar\beta}{2\bar\beta+1}}. \tag{7.3}$$

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

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

and $K_1(t - s_n')$ is larger than $\delta K_2(0)/2$ due to (7.1) (ii). Define

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

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 $K(.,.)$ denoting the Kullback-Leibler divergence,

$$\begin{aligned}
K(\mathbb{P}_{1,n}^{\otimes n}, \mathbb{P}_{0,n}^{\otimes n}) &= n\, K(\mathbb{P}_{1,n}, \mathbb{P}_{0,n})\\
&= n\int \log\left(\frac{p_{1,n}(x)}{p_{0,n}(x)}\right) p_{1,n}(x)\, d\lambda^d(x)
\end{aligned}$$

$$
\begin{aligned}
(7.4) \qquad & \leq n \int \Bigg\{ \frac{K_2(x-t;h_n) - K_2(x-s_n';h_n)}{K_1(x-s_n')} \\
& \qquad \cdot \Big(K_1(x-s_n') + K_2(x-t;h_n) - K_2(x-s_n';h_n) \Big) \Bigg\} \, d\lambda^d(x) \\
& = n \int \left(K_2(x-t;h_n) - K_2(x-s_n';h_n) \right) d\lambda^d(x) \\
& \quad + n \int \frac{(K_2(x-t;h_n) - K_2(x-s_n';h_n))^2}{K_1(x-s_n')} \, d\lambda^d(x) \\
& = n \int \frac{(K_2(x-t;h_n) - K_2(x-s_n';h_n))^2}{K_1(x-s_n')} \, d\lambda^d(x) \\
& = n \int \frac{K_2^2(x-t;h_n)}{K_1(x-s_n')} \, d\lambda^d(x) + n \int \frac{K_2^2(x-s_n';h_n)}{K_1(x-s_n')} \, d\lambda^d(x) \\
& \leq 2n \int \frac{K_2^2(x-t;h_n)}{K_1(x-s_n')} \, d\lambda^d(x) \\
& \leq 2n \left(\min_{x \in \prod_{i=1}^d [-g_{2,i} h_{n,i},\, g_{2,i} h_{n,i}]} K_1(x+t-s_n') \right)^{-1} \int K_2^2(x;h_n) \, d\lambda^d(x) \\
(7.5) \qquad & \leq 2n \left(\frac{c_7(\beta,L)}{4} \right)^{-1} \|K_2\|_2^2 \left(\prod_{i=1}^d h_{n,i} \right)^{2\bar{\beta}+1},
\end{aligned}
$$

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 (7.4) is due to the inequality $\log(1+x) \leq x$ for $x > -1$ and (7.5) 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 (7.5) 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 Appendix C, 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 $i = 1, \dots, d$. The proof is conducted in complete analogy to the case I_2 except for the bound (7.5), which is too rough for the case under consideration now. Instead, define

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

with

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

Since $K_1(t - s'_n)$ as given in (7.3) is bounded from below by $\delta/2 \cdot K_2(0)$ for any $\delta \in I_1$ due to (7.1) (ii),

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

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

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

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 $i = 1, \dots, d$,

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

such that applied to the bound in (7.5),

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

$$
\begin{aligned}
&\le \frac{2\|K_2\|_2^2\,\delta}{2^{-d}\,\delta\,K_2(0)/2}\left(\prod_{i=1}^d c_{8,i}(\beta,L)\right)^{\frac{2\bar\beta+1}{\bar\beta}}\\
&= \frac{2^{d+2}\|K_2\|_2^2}{K_2(0)}\left(\prod_{i=1}^d c_{8,i}(\beta,L)\right)^{\frac{2\bar\beta+1}{\bar\beta}},
\end{aligned}
$$

and we conclude as before. □

In the proof of Theorem 3.2, we frequently make use of the bandwidth

$$
\bar h_i := c_{11}(\beta,L)\cdot\max\left\{\left(\frac{\log n}{n}\right)^{\frac{\bar\beta}{\bar\beta+1}\frac{1}{\beta_i}},\ \left(\frac{p(t)\log n}{n}\right)^{\frac{\bar\beta}{2\bar\beta+1}\frac{1}{\beta_i}}\right\} \tag{7.6}
$$

for $i=1,\ldots,d$, with constant $c_{11}(\beta,L)$ of Lemma A.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

$$
\bar j_i = \bar j_i(t) := \left\lfloor \log_2\left(\frac{1}{\bar h_i}\right)\right\rfloor + 1,\quad i=1,\ldots,n \tag{7.7}
$$

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 $\bar h_i/2\le 2^{-\bar j_i}\le \bar h_i$ for all $i=1,\ldots,d$.

Before turning to the proof of Theorem 3.2, 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 $\beta_i\ge 1$ lead to

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

Hence,

$$|b_t(h)| \leq L \sum_{i=1}^{d} c_{12,i}(\beta)\, h_i^{\beta_i} =: B_t(h) \tag{7.8}$$

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

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

Furthermore, the following lemma by Klutchnikoff (2005) is important to bound the difference of two bias terms in the anisotropic setting.

LEMMA 7.1 (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 $t \in \mathbb{R}^d$.

LEMMA 7.2. *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

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

PROOF. We define

$$\begin{aligned} J_1 &:= \{ i \in \{1, \ldots, d\} : m_i > j_i \} := \{ i_1, \ldots, i_s \} \\ J_2 &:= \{1, \ldots, d\} \setminus J_1. \end{aligned}$$

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

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

we have the representation

$$\sigma_t^2(j \wedge m) = \frac{1}{n \prod_{i=1}^d (h_{j,i} \vee h_{m,i})} \int \prod_{i \in J_1} K_{i,h_{j,i}}^2(t_i - x_i)\, M_{J_2}^{t_2}(x_1)\, d\lambda^s(x_1), \tag{7.9}$$

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

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

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

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

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

then Lemma A.1 (i) yields

$$\begin{aligned} &\frac{1}{\prod_{i \in J_1} h_{j,i}} \int \prod_{i \in J_1} K_{i,h_{j,i}}^2(t_i - x_i)\, M_{J_2}^{t_2}(x_1)\, d\lambda^s(x_1) \\ &\leq \frac{3}{2} \cdot \frac{\|\prod_{i \in J_1} K_i\|_2^2}{\prod_{i \in J_1} h_{j,i}} M_{J_2}^{t_2}(t_1) \\ &\leq \frac{3}{2} \cdot \frac{\|\prod_{i \in J_1} K_i\|_2^2}{\prod_{i \in J_1} h_{m,i}} M_{J_2}^{t_2}(t_1) \\ &\leq 3 \cdot \frac{1}{\prod_{i \in J_1} h_{m,i}} \int \prod_{i \in J_1} K_{i,h_{m,i}}^2(t_i - x_i)\, M_{J_2}^{t_2}(x_1)\, d\lambda^s(x_1), \end{aligned} \tag{7.10}$$

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

$$M_{J_2}^{t_2}(t_1) \leq c_6(\beta, L) \left\{ \left(\frac{\log n}{n} \right)^{\frac{\bar{\beta}}{\bar{\beta}+1}} \vee \left(\frac{p(t) \log n}{n} \right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}} \right\} \tag{7.11}$$

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 (7.11) 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

$$\left(\frac{p(t) \log n}{n} \right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}} \leq p(t), \tag{7.12}$$

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 (7.9), we obtain by the same arguments as in (A.2), (A.3) and (A.4) 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,

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

Since $h_{j,i} \leq \bar{h}_{i,j} \leq c_{11}(\beta, L)\, p(t)^{1/\beta_i}$, $i = 1, \ldots, d$ whenever $p(t) > (\log n/n)^{\bar{\beta}/(\bar{\beta}+1)}$ due to (7.12), Lemma A.1 (i) can be applied and yields

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

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

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

which proves the first statement of Lemma 7.2. 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 $i = 1, \ldots, d$, 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 $i = 1, \ldots, d$. Consequently, by Lemma A.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 (A.2), (A.3) and (A.4), for any $h \leq \bar{h}$ componentwise,

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

and therefore

$$\frac{\log^2 n}{n^2\prod_{i=1}^d h_i^2} \leq \sigma^2_{t,\mathrm{trunc}}(h) \leq c(\beta, L)\frac{\log^2 n}{n^2\prod_{i=1}^d h_i^2},$$

where the left and right hand side are monotone in the product of bandwidth components.

□

The following lemma carefully analyzes the ratio of the truncated quantities $\sigma^2_{t,\mathrm{trunc}}$ and $\tilde{\sigma}^2_{t,\mathrm{trunc}}$ and is crucial for the proof of Theorem 3.2.

LEMMA 7.3. *For the quantities $\sigma^2_{t,\mathrm{trunc}}(h)$ and $\tilde{\sigma}^2_{t,\mathrm{trunc}}(h)$ defined in (3.2) and any $\eta \geq 0$ holds*

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

PROOF. The proof is based on Bernstein's inequality. First,

$$\begin{aligned}&\mathbb{P}^{\otimes n}\Big(|\tilde{\sigma}^2_{t,\mathrm{trunc}}(h) - \sigma^2_{t,\mathrm{trunc}}(h)| \geq \eta\,\sigma^2_{t,\mathrm{trunc}}(h)\Big)\\ &\qquad\leq\; \mathbb{P}^{\otimes n}\Big(|\tilde{\sigma}^2_t(h) - \sigma^2_t(h)| \geq \eta\,\sigma^2_{t,\mathrm{trunc}}(h)\Big).\end{aligned}$$

The random variable $\tilde{\sigma}_t^2(h) - \sigma_t^2(h)$ 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

$$|Z_k| \leq \frac{2\|K\|_{\sup}^2}{n^2 \prod_{i=1}^d h_i^2}$$

and

$$\sum_{k=1}^n \operatorname{Var}(Z_k) \leq \frac{1}{n^3 \prod_{i=1}^d h_i^4} \mathbb{E}_p K^4 \left(\frac{t - X_1}{h} \right) \leq \frac{\|K\|_{\sup}^2}{n^2 \prod_{i=1}^d h_i^2} \sigma_{t,\mathrm{trunc}}^2(h).$$

Hence, Bernstein's inequality yields the following exponential tail bound

$$\begin{aligned}
&\mathbb{P}^{\otimes n} \left(|\tilde{\sigma}_t^2(h) - \sigma_t^2(h)| \geq \eta\, \sigma_{t,\mathrm{trunc}}^2(h) \right) \\
&\qquad \leq 2 \exp \left(-\frac{1}{2} \frac{\eta^2 \sigma_{t,\mathrm{trunc}}^4(h)}{\frac{\|K\|_{\sup}^2 \sigma_{t,\mathrm{trunc}}^2(h)}{n^2 \prod_{i=1}^d h_i^2} \left(1 + \frac{2\eta}{3}\right)} \right) \qquad (7.13)\\
&\qquad \leq 2 \exp \left(-\frac{3\eta^2}{2(3 + 2\eta)\|K\|_{\sup}^2} \log^2 n \right).
\end{aligned}$$

□

Obviously, a truncation of the variance estimator is necessary since inequality (7.13) is getting weak if the variance is very small, and sufficient concentration of the estimator around its mean cannot be guaranteed in this case.

Proof of Theorem 3.2 and Theorem 5.1. Recall the notation of Appendix B 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:

$$\begin{aligned}
\mathbb{E}_p^{\otimes n} |\hat{p}_{n,\hat{j}}(t) - p(t)|^r &= \mathbb{E}_p^{\otimes n} \left[|\hat{p}_{n,\hat{j}}(t) - p(t)|^r \cdot \mathbb{1}\{\hat{\sigma}_t^2(\hat{j}) \leq \hat{\sigma}_t^2(\bar{j})\} \right] \\
&\quad + \mathbb{E}_p^{\otimes n} \left[|\hat{p}_{n,\hat{j}}(t) - p(t)|^r \cdot \mathbb{1}\{\hat{\sigma}_t^2(\hat{j}) > \hat{\sigma}_t^2(\bar{j})\} \right] \\
&=: R^+ + R^-. \qquad (7.14)
\end{aligned}$$

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

$$\begin{aligned}R^+ \leq 3^{r-1}\Bigg(&\mathbb{E}_p^{\otimes n}\Big[|\hat{p}_{n,\hat{j}}(t) - \hat{p}_{n,\hat{j}\wedge\bar{j}}(t)|^r \cdot \mathbb{1}\{\hat{\sigma}_t^2(\hat{j}) \leq \hat{\sigma}_t^2(\bar{j})\}\Big]\\
&+ \mathbb{E}_p^{\otimes n}\Big[|\hat{p}_{n,\hat{j}\wedge\bar{j}}(t) - \hat{p}_{n,\bar{j}}(t)|^r \cdot \mathbb{1}\{\hat{\sigma}_t^2(\hat{j}) \leq \hat{\sigma}_t^2(\bar{j})\}\Big]\\
&+ \mathbb{E}_p^{\otimes n}\Big[|\hat{p}_{n,\bar{j}}(t) - p(t)|^r \cdot \mathbb{1}\{\hat{\sigma}_t^2(\hat{j}) \leq \hat{\sigma}_t^2(\bar{j})\}\Big]\Bigg)\\
=:&\, 3^{r-1}(S_1 + S_2 + S_3),\end{aligned} \tag{7.15}$$

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 (B.1). As a consequence the following inequality holds true

$$\begin{aligned}S_2 \;&\leq c_{14}^r\, \mathbb{E}_p^{\otimes n}\left[\left(\hat{\sigma}_t^2(\bar{j})\log n\right)^{r/2} \cdot \mathbb{1}\left\{\left|\frac{\tilde{\sigma}^2_{t,\mathrm{trunc}}(\bar{j})}{\sigma^2_{t,\mathrm{trunc}}(\bar{j})} - 1\right| < \;1\right\}\right]\\
&\quad+\; c_{14}^r\, \mathbb{E}_p^{\otimes n}\left[\left(\hat{\sigma}_t^2(\bar{j})\log n\right)^{r/2} \cdot \mathbb{1}\left\{\left|\frac{\tilde{\sigma}^2_{t,\mathrm{trunc}}(\bar{j})}{\sigma^2_{t,\mathrm{trunc}}(\bar{j})} - 1\right| \geq \;1\right\}\right]\\
&\leq 2^{r/2}c_{14}^r\left(\min\left\{\sigma^2_{t,\mathrm{trunc}}(\bar{j}), \frac{\|K\|_2^2\, c_1}{n2^{-|\bar{j}|}}\right\}\log n\right)^{r/2}\\
&\quad+\; c_{14}^r\left(\frac{\|K\|_2^2\, c_1}{n2^{-|\bar{j}|}}\log n\right)^{r/2}\mathbb{P}^{\otimes n}\left(\left|\frac{\tilde{\sigma}^2_{t,\mathrm{trunc}}(\bar{j})}{\sigma^2_{t,\mathrm{trunc}}(\bar{j})} - 1\right| \geq \;1\right),\end{aligned}$$

where we used the condition in the indicator function in the first summand to bound the estimated truncated variance $\tilde{\sigma}^2_{t,\mathrm{trunc}}$ from above by $2\sigma^2_{t,\mathrm{trunc}}$, and additionally the upper truncation level in the second summand. By the deviation inequality of Lemma 7.3, we can further estimate S_2 by

$$\begin{aligned}S_2 \leq\, &2^{r/2}c_{14}^r\left(\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n\right)^{r/2}\\
&+ c_{14}^r\left(\frac{\|K\|_2^2 c_1}{n2^{-|\bar{j}|}}\log n\right)^{r/2} \cdot 2\exp\left(-\frac{3}{10\|K\|^2_{\sup}}\log^2 n\right).\end{aligned}$$

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

$$\left(\frac{\|K\|_2^2 c_1}{n2^{-|\bar{j}|}}\log n\right)^{r/2} \cdot 2\exp\left(-\frac{3}{10\|K\|^2_{\sup}}\log^2 n\right) \leq c\left(\frac{\log^3 n}{n^2\left(2^{-|\bar{j}|}\right)^2}\right)^{r/2}$$

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

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

We will now turn to S_3, the third term in (7.15). We split the risk into bias and stochastic error. It holds

$$S_3 \leq \mathbb{E}_p^{\otimes n}\Big(|\hat{p}_{n,\bar{j}}(t) - \mathbb{E}_p\hat{p}_{n,\bar{j}}(t)| + B_t(\bar{j})\Big)^r \tag{7.16}$$

and by Lemma A.2

$$B_t(\bar{j}) \leq c_{15}(\beta, L)\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n}. \tag{7.17}$$

Denoting by

$$Z_k := \frac{\hat{p}_{n,k}(t) - \mathbb{E}_p\hat{p}_{n,k}(t)}{\sqrt{\sigma^2_{t,\mathrm{trunc}}(k)\log n}} \qquad \text{for } k \in \mathcal{J}, \tag{7.18}$$

the decomposition (7.16), the bias variance relation (7.17) and the inequality $(x + y)^r \leq 2^{r-1}(x^r + y^r)$, $x, y \geq 0$ together with Lemma A.4 yields

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

It remains to show an analogous result for S_1, the first term in (7.15). Clearly,

$$\begin{aligned} S_1 \leq \sum_{j\in\mathcal{J}} \mathbb{E}_p^{\otimes n}\Big[\Big(\big|\hat{p}_{n,j}(t) - \mathbb{E}_p^{\otimes n}\hat{p}_{n,j}(t)\big| + \big|\hat{p}_{n,j\wedge\bar{j}}(t) - \mathbb{E}_p^{\otimes n}\hat{p}_{n,j\wedge\bar{j}}(t)\big| \\ + \big|b_t(j\wedge\bar{j}) - b_t(j)\big|\Big)^r \cdot \mathbb{1}\{\hat{\sigma}_t^2(j) \leq \hat{\sigma}_t^2(\bar{j}),\, \hat{j} = j\}\Big]. \end{aligned} \tag{7.19}$$

By Lemma 7.1 and Lemma A.2,

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

On account of this inequality and in view of (7.19), it suffices to bound the expectations in the following expression

$$\begin{aligned} S_1 \;&\leq 3^{r-1}\left(\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n\right)^{r/2} \\ &\cdot\Bigg\{\sum_{j\in\mathcal{J}}\mathbb{E}_p^{\otimes n}\left[\left(\frac{|\hat{p}_{n,j}(t) - \mathbb{E}_p^{\otimes n}\hat{p}_{n,j}(t)|}{\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n}}\right)^r \mathbb{1}\{\hat{\sigma}_t^2(j) \leq \hat{\sigma}_t^2(\bar{j}),\, \hat{j} = j\}\right] \\ &\quad + \sum_{j\in\mathcal{J}}\mathbb{E}_p^{\otimes n}\left[\left(\frac{|\hat{p}_{n,j\wedge\bar{j}}(t) - \mathbb{E}_p^{\otimes n}\hat{p}_{n,j\wedge\bar{j}}(t)|}{\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n}}\right)^r \mathbb{1}\{\hat{\sigma}_t^2(j) \leq \hat{\sigma}_t^2(\bar{j}),\, \hat{j} = j\}\right] \\ &\quad + \sum_{j\in\mathcal{J}} 2^r c_{15}(\beta, L)^r \cdot \mathbb{P}^{\otimes n}(\hat{j} = j)\Bigg\}. \end{aligned} \tag{7.20}$$

Denoting

$$A_{j,\bar{j}} := \left\{ \left| \frac{\tilde{\sigma}^2_{t,\mathrm{trunc}}(j)}{\sigma^2_{t,\mathrm{trunc}}(j)} - 1 \right| < \frac{1}{2} \text{ and } \left| \frac{\tilde{\sigma}^2_{t,\mathrm{trunc}}(\bar{j})}{\sigma^2_{t,\mathrm{trunc}}(\bar{j})} - 1 \right| < \frac{1}{2} \right\}, \tag{7.21}$$

it follows

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

Applying Lemma A.4 and Hölder's inequality for any $p > 1$

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

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 $\sigma^2_{t,\mathrm{trunc}}$,

$$\frac{\sigma^2_{t,\mathrm{trunc}}(k)}{\sigma^2_{t,\mathrm{trunc}}(l)} \;\le\; \frac{(1\vee c_1\|K\|_2^2)n^2}{\log^4 n} \qquad \text{for any } k,l\in\mathcal{J}, \tag{7.22}$$

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 A.4 as above. Finally, the probabilities compensate (7.22) by Lemma 7.3. As concerns the expectation in (7.20), 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 7.2 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 (7.14) 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

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

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 (B.1), because $\hat{j}$ had not been chosen in the minimization problem (B.3) 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

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

and we divide the absolute value of the difference of the kernel density estimators as in (7.19) 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,

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

by Lemma 7.1 and Lemma A.2, leading to the inequality

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

with

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

To start with the second probability, we intersect event $B_{2,m}$ with $A_{m,\bar{j}}$ as defined in (7.21). 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 A.3 allow to bound the probability

$$\begin{aligned}
\mathbb{P}^{\otimes n}(B_{2,m} \cap A_{m,\bar{j}}) &\leq \mathbb{P}^{\otimes n}\left(\frac{|\hat{p}_{n,m}(t) - \mathbb{E}_p^{\otimes n}\hat{p}_{n,m}(t)|}{\sqrt{\sigma^2_{t,\mathrm{trunc}}(m)\log n}} > c_{16}(\beta, L)\right) \\
&\leq 2\exp\left(-\frac{c_{16}(\beta, L)^2 \wedge c_{16}(\beta, L)}{4}\log n\right) \qquad (7.23)
\end{aligned}$$

with

$$c_{16}(\beta, L) := \left(\frac{c_{14}}{2} - c_{15}(\beta, L)\sqrt{2\,\frac{1 \vee c_1\|K\|_2^2}{c_1\|K\|_2^2}}\right) \cdot \sqrt{\frac{1}{2}\,\frac{c_1\|K\|_2^2}{1 \vee c_1\|K\|_2^2}}. \qquad (7.24)$$

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

$$\frac{c_{16}(\beta,L)^2 \wedge c_{16}(\beta,L)}{4} \geq \frac{r\bar{\beta}}{\bar{\beta}+1} + 1 \tag{7.25}$$

for any β in the range of adaptation. Finally, by means of Lemma 7.3,

$$\begin{aligned}\mathbb{P}^{\otimes n}(A^c_{m,\bar{j}}) &\leq \mathbb{P}^{\otimes n}\left(\left|\frac{\tilde{\sigma}^2_{t,\mathrm{trunc}}(\bar{j})}{\sigma^2_{t,\mathrm{trunc}}(\bar{j})} - 1\right| \geq \frac{1}{2}\right) + \mathbb{P}^{\otimes n}\left(\left|\frac{\tilde{\sigma}^2_{t,\mathrm{trunc}}(m)}{\sigma^2_{t,\mathrm{trunc}}(m)} - 1\right| \geq \frac{1}{2}\right) \\ &\leq 4\exp\left(-\frac{3}{32\|K\|^2_{\sup}}\log^2 n\right), \qquad (7.26)\end{aligned}$$

which is of smaller order than the bound in (7.23). 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 7.2, 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,

$$\mathbb{P}^{\otimes n}\Big(\hat{\sigma}^2_t(\hat{j}) > \hat{\sigma}^2_t(\bar{j})\Big) \;\leq\; c(\beta,L)\left(\sigma^2_{t,\mathrm{trunc}}(\bar{j})\log n\right)^{r/2}. \tag{7.27}$$

Finally, by Lemma A.2,

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

This completes the proof of Theorem 3.2. The proof of Theorem 5.1 follows the same lines except for the following modifications. We write

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

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

$$\begin{aligned}\tilde{R}^+ &\leq 2^{r-1}\Big(\mathbb{E}^{\otimes n}_p\Big[|\hat{p}_{n,\hat{j}}(t) - \hat{p}_{n,\bar{j}}(t)|^r \cdot \mathbb{1}\{\hat{j} \leq \bar{j}\}\Big] \\ &\qquad + \mathbb{E}^{\otimes n}_p\Big[|\hat{p}_{n,\bar{j}}(t) - p(t)|^r \cdot \mathbb{1}\{\hat{j} \leq \bar{j}\}\Big]\Big) \\ &=: 2^{r-1}(\tilde{S}_1 + \tilde{S}_3). \qquad (7.28)\end{aligned}$$

Here, $\bar{j}$ corresponds to the reference bandwidth $\bar{h}$ defined in (7.47). We need to verify the bounds

$$B_t(\bar{j}) \leq c(\beta, L)\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j}) \log n} \tag{7.29}$$

$$\sqrt{\sigma^2_{t,\mathrm{trunc}}(\bar{j})} \leq c(\beta, L)\left(\frac{\log n}{n}\right)^{\frac{\beta}{\beta+d}} \tag{7.30}$$

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

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

The first bound (7.29) is a consequence of the classical upper bound on the bias for higher order kernels, whereas the second bound (7.30) follows by Lemma A.1 (iii).The terms $\tilde{S}_1$ and $\tilde{S}_3$ in (7.28) then require no further arguments. As concerns $\tilde{R}^-$, it remains to investigate

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

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

$$\sigma^2_{t,\mathrm{trunc}}(\bar{j}) \leq c(\beta, L)\,\sigma^2_{t,\mathrm{trunc}}(m) \tag{7.31}$$

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

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

for some constant $c_{26}(\beta, L) > c_{11}(\beta, L)$. By Lemma A.1 (iii),

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

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

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

that is, (7.31) is verified. The further proof can then be conducted as before for Theorem 3.2. $\square$

PROOF OF THEOREM 3.3. Before we construct the densities p_n and q_n, we first specify their amplitudes Δ_n and δ_n in t, respectively. Let

$$
\begin{aligned}
\Delta_n &:= n^{-\frac{\beta_1}{\beta_1+1}} \cdot \varrho(n) \\
\delta_n &:= 4c_3(\beta_1^*, L_1^*, r) \left(\frac{\Delta_n}{n}\right)^{\frac{\beta_1}{2\beta_1+1}} (\log n)^{3/2} \\
&= 4c_3(\beta_1^*, L_1^*, r)\Delta_n \cdot \varrho(n)^{-\frac{\beta_1+1}{2\beta_1+1}} (\log n)^{3/2},
\end{aligned}
\tag{7.32}
$$

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

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

and thus the conditions (3.3) are fulfilled.
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 (C.1), 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 $\cup_{L>0}\mathscr{C}_c \cap \mathscr{P}_1(\beta, L)$, $0 < \beta \leq 2$, 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.

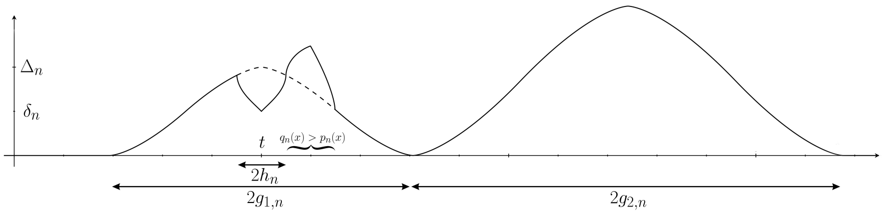

FIG 3. *Construction of p_n (dashed line) and q_n (solid line)*

Let $T_n(t)$ be an arbitrary estimator with property (3.4). 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

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

In contrast to their proof, we need the decomposition

$$\begin{aligned}
&\mathbb{E}_{q_n}^{\otimes n}|\tilde{T}_n(t) - q_n(t)| \\
&\qquad \ge (\Delta_n - \delta_n) - \mathbb{E}_{q_n}^{\otimes n}\Big[\,|T_n(t) - p_n(t)|\mathbb{1}_{B_n}\Big] - \mathbb{E}_{q_n}^{\otimes n}\Big[\,|\tilde{T}_n(t) - p_n(t)|\mathbb{1}_{B_n^c}\Big] \\
&\qquad =: (\Delta_n - \delta_n) - S_1 - S_2, \qquad (7.33)
\end{aligned}$$

where

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

By definition of Δ_n and δ_n in (7.32) and the risk bound (3.4) the first two summands in (7.33) 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,

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

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

$$\mathbb{Q}_n^{\otimes n}(B_n^c) \leq c(\beta_1,\beta_2)\frac{\delta_n}{\Delta_n} \longrightarrow 0\,.$$

□

7.2. *Proofs of Section 4.* Let $K_g(\cdot) = g^{-d}K(\cdot/g)$ with $g = g_{\beta,L/2,d}$ the specific kernel as described in Appendix C 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

$$\Lambda_d(K_g(\cdot\,;h,\beta),\varepsilon) \;:=\; \lambda^d\left(\overline{\{x\in\mathbb{R}^d \,:\, 0<K_g(x;h,\beta)\leq\varepsilon\}}\right) \tag{7.34}$$

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

Lemma 7.4. *For any bandwidth g and h and*

$$\varepsilon \leq c_{17}(\beta)\cdot 2^{-\beta}h^\beta g^{-d}$$

the volume of the complementary level sets defined in (7.34) is upper bounded by

$$\Lambda_d(K_g(\cdot\,;h,\beta),\varepsilon) \;\leq\; c_{17}(\beta)\,dV_d\cdot(gh)^{d-1}g^{\frac{\beta+d}{\beta}}\varepsilon^{\frac{1}{\beta}},$$

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

Proof. Note first, that both the support and the level sets of $K_g(\cdot\,;h,\beta)$ are concentric balls and hence $\Lambda_d(K_g(\cdot\,;h,\beta),\varepsilon)$ is for $\varepsilon \leq c_{17}(\beta)\cdot 2^{-\beta}h^\beta g^{-d}$ the volume of a spherical shell with inner radius larger than the radius $gh/2$, which is just the radius of the sphere where the definition of $K_g(\cdot;h,\beta)$ splits in case of $1<\beta\leq 2$. Both the cases $\beta\leq 1$ and $1<\beta\leq 2$ can hence be treated at once. The volume of a d-dimensional spherical shell with outer radius R and inner radius $r\leq R$

equals $V_d(R^d - r^d)$. An induction on the dimension d shows that this is in turn upper bounded by

$$\lambda^d(B_R(0)) - \lambda^d(B_r(0)) \;\le\; dV_d\, R^{d-1}(R-r). \tag{7.35}$$

For $d = 1$ this holds immediately. For $d = 2$ we have

$$R^2 - r^2 = (R+r)(R-r) \le 2R(R-r).$$

If the inequality holds for $d-1$, we obtain

$$\begin{aligned} R^d - r^d &= R^{d-1}(R-r) + r(R^{d-1} - r^{d-1}) \\ &\le R^{d-1}(R-r) + R(d-1)R^{d-2}(R-r) \\ &= dR^{d-1}(R-r). \end{aligned}$$

Since $K_g(x; h, \beta)$ attains ε on the sphere with radius $gh - c_{17}(\beta)^{-\frac{1}{\beta}} g^{\frac{\beta+d}{\beta}} \varepsilon^{\frac{1}{\beta}}$ for $\varepsilon \le c_{17}(\beta) \cdot 2^{-\beta} h^\beta g^{-d}$,

$$\Lambda_d(K_g(\cdot\,; h, \beta), \varepsilon) \;\le\; c_{17}(\beta)\, dV_d \cdot (gh)^{d-1} g^{\frac{\beta+d}{\beta}} \varepsilon^{\frac{1}{\beta}}.$$

□

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

$$h_n := (2n)^{-\frac{1}{\beta+d}}\,. \tag{7.36}$$

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

$$m_n = \lfloor h_n^{\gamma\beta-d} \rfloor + 1. \tag{7.37}$$

Now choose points $z_i = (z_{i,1}, \ldots, z_{i,d})$, $i = 1, \ldots, m_n$ in $B_{g/2}(0)$ 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

$$\begin{aligned}\Lambda_d(p_{\omega,n},\varepsilon) \leq \Lambda_d(K_g,\varepsilon)& \\ &+ m_n\Lambda_d(K_g(\cdot\,;h_n,\beta),\varepsilon)\ \mathbb{1}\left\{\varepsilon \leq c_{17}(\beta)\cdot 2^{-\beta}h_n^{\beta}\,g^{-d}\right\} \\ &+ 2m_nV_d\cdot(gh_n)^d\mathbb{1}\left\{\varepsilon > c_{17}(\beta)\cdot 2^{-\beta}h_n^{\beta}\,g^{-d}\right\}.\end{aligned} \tag{7.38}$$

Now, because $m_n \leq h_n^{\gamma\beta-d} + 1 \leq 2h_n^{\gamma\beta-d}$, Lemma 7.4 yields

$$m_n\Lambda_d(K_g(\cdot\,;h_n,\beta),\varepsilon)\cdot\mathbb{1}\left\{\varepsilon \leq c_{17}(\beta)\cdot 2^{-\beta}h_n^{\beta}g^{-d}\right\} \tag{7.39}$$

$$\begin{aligned}&\leq c(\beta,L)\cdot h_n^{\gamma\beta-d}h_n^{d-1}\varepsilon^{\frac{1}{\beta}}\cdot\mathbb{1}\left\{\varepsilon \leq c_{17}(\beta)\cdot 2^{-\beta}h_n^{\beta}\,g^{-d}\right\} \\ &= c(\beta,L)\cdot h_n^{\gamma\beta-1}\varepsilon^{\frac{1}{\beta}}\cdot\mathbb{1}\left\{\varepsilon \leq c_{17}(\beta)\cdot 2^{-\beta}h_n^{\beta}\,g^{-d}\right\} \\ &\leq c(\beta,L)\cdot\varepsilon^{\gamma-\frac{1}{\beta}}\ \varepsilon^{\frac{1}{\beta}}\cdot\mathbb{1}\left\{\varepsilon \leq c_{17}(\beta)\cdot 2^{-\beta}h_n^{\beta}\,g^{-d}\right\}\end{aligned}$$

$$= c(\beta,L)\cdot\varepsilon^{\gamma}\cdot\mathbb{1}\left\{\varepsilon \leq c_{17}(\beta)\cdot 2^{-\beta}h_n^{\beta}\,g^{-d}\right\}, \tag{7.40}$$

where the last inequality is due to the property $\gamma\beta \leq 1$. Furthermore, since $m_nh_n^d \leq 2h_n^{\gamma\beta}$, we can derive a similar bound for the last term in (7.38)

$$\begin{aligned}2m_nV_d\cdot(gh_n)^d\cdot\mathbb{1}\left\{\varepsilon > c_{17}(\beta)\cdot 2^{-\beta}h_n^{\beta}g^{-d}\right\}& \\ &\leq c(\beta,L,\gamma)\cdot\varepsilon^{\gamma}\cdot\mathbb{1}\left\{\varepsilon > c_{17}(\beta)\cdot 2^{-\beta}h_n^{\beta}\,g^{-d}\right\}.\end{aligned} \tag{7.41}$$

Clearly, for $\varepsilon \leq c_{17}(\beta)\cdot 2^{-\beta}g^{-d}\wedge 1$, Lemma 7.4 also yields

$$\Lambda_d(K_g,\varepsilon) \leq c(\beta,L)\cdot\varepsilon^{\frac{1}{\beta}} \leq c(\beta,L)\cdot\varepsilon^{\gamma} \tag{7.42}$$

using the property $\gamma\beta \leq 1$ again. In summary, inequality (7.38) simplifies with (7.39), (7.41) and (7.42) for $\varepsilon \leq c_{16}(\beta)\cdot 2^{-\beta}g^{-d}\wedge 1$ 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 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 (7.35),

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

$$
\begin{aligned}
&= dV_d \cdot (2g)^{d-1}\varepsilon \cdot (1+m_n) \\
&\le dV_d \cdot (2g)^{d-1}\varepsilon \cdot 3h_n^{\gamma\beta-1} \\
&\le c(\beta, L, \gamma, d) \cdot \varepsilon^{\gamma\beta} \qquad (7.43)
\end{aligned}
$$

where inequality (7.43) follows from $\gamma\beta \le 1$. For $\varepsilon > gh_n$ let ξ_1 be an arbitrary constant and choose the following decomposition for the complexity condition

$$
\Gamma_{p_{\omega,n}} = B_g(0) \;\cup \bigcup_{k\,:\,\omega_k=1} B_{gh_n}(z'_k) =: A_{1,\varepsilon} \cup A_{2,\varepsilon} \qquad (7.44)
$$

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

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

where the last inequality again follows from $\gamma\beta \le 1$. To prove the complexity condition it remains to upper bound the Lebesgue volume of the second part in the decomposition (7.44). For $gh_n < \varepsilon \le \xi_1$,

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

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. The further proof accomplishes two tasks. Firstly, the minimax risk will be reduced to the form

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

where ρ is the Hamming distance and expectation is taken with respect to $\mathbb{P}_{\omega,n}^{\otimes n}$. 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

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

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

$$
\mathbb{E}_{\omega,n}^{\otimes n}\left[d_\Delta(\hat\Gamma_n, \Gamma_{p_{\omega,n}})\right] \ge \sum_{i=1}^{m_n} \mathbb{E}_{\omega,n}^{\otimes n}\left[\int_{B_{gh_n}(z'_i)} \left|\mathbb{1}_{\hat\Gamma_n}(x) - \mathbb{1}_{\Gamma_{p_{\omega,n}}}(x)\right| d\lambda^d(x)\right],
$$

which in turn simplifies to

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

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

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

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

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

and consequently

$$\inf_{\hat{\Gamma}_n}\sup_{p\in\mathscr{P}_d(\beta,L,\gamma,\kappa,\xi)} \mathbb{E}_p^{\otimes n}\left[d_\Delta(\hat{\Gamma}_n,\Gamma_p)\right] \;\geq\; c(\beta,L)h_n^d\cdot\inf_{\hat{\omega}}\max_{\omega\in\Omega}\mathbb{E}_{\omega,n}^{\otimes n}\rho(\omega,\hat{\omega}), \tag{7.45}$$

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 (7.45). 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 (7.2) for this purpose. Of course, ω and ω' coincide except for one component, say j. Again, by Bernoulli's inequality

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

By the choice of h_n in (7.36), this distance is bounded by one which yields together with (7.37) and inequality (7.45), 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 7.5. *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 $\gamma > 0$,

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

Let G be a measurable subset of Γ_p. Then,

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

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

$$\begin{aligned} \int_G p(x) d\lambda^d(x) &\geq \varepsilon\, \lambda^d(G \cap \{p > \varepsilon\}) \\ &= \varepsilon \Big(\lambda^d(G) - \lambda^d(G \cap \{p \leq \varepsilon\}) \Big) \end{aligned}$$

$$\begin{aligned}&\geq \varepsilon\Big(\lambda^d(G) - \lambda^d\left(\overline{\{0 < p \leq \varepsilon\}}\right)\Big)\\ &\geq \varepsilon\lambda^d(G) - c_{20}(\kappa,\gamma)\varepsilon^{\gamma+1} \qquad (7.46)\end{aligned}$$

with

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

The right hand side of (7.46) is maximized by the specific choice of

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

Plugging this specific ε in (7.46) yields

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

□

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

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

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 $\hat{p}_n(x) = \hat{p}_{n,\hat{j}}(x)$, see Appendix B and denote by $\bar{j} = \bar{j}(x)$ the exponent corresponding to the reference bandwidth

$$\bar{h} = \bar{h}(x) := c_{11}(\beta,L)\cdot\max\left\{\left(\frac{\log n}{n}\right)^{\frac{1}{\beta+d}}, \left(\frac{p(x)\log n}{n}\right)^{\frac{1}{2\beta+d}}\right\}. \qquad (7.47)$$

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

$$\begin{aligned}E_1 &:= \mathbb{E}_p^{\otimes n}\,\lambda^d\left(\hat{\Gamma}_n \setminus \Gamma_p\right)\\ E_2 &:= \mathbb{E}_p^{\otimes n}\,\lambda^d\left(\Gamma_p \setminus \hat{\Gamma}_n\right)\end{aligned} \qquad (7.48)$$

and start with E_1. We split E_1 again

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

with

$$E_{1,1} := \mathbb{E}_p^{\otimes n}\left[\lambda^d\Big(x \in \mathbb{R}^d \ : \ \hat{p}_{n,\hat{j}}(x) \geq \alpha_n,\ p(x) = 0,\ \hat{j} \geq \bar{j}\Big)\right]$$
$$E_{1,2} := \mathbb{E}_p^{\otimes n}\left[\lambda^d\Big(x \in \mathbb{R}^d \ : \ \hat{p}_{n,\hat{j}}(x) \geq \alpha_n,\ p(x) = 0,\ \hat{j} < \bar{j}\Big)\right]. \tag{7.49}$$

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

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

The support Γ_p is assumed to satisfy the complexity condition 4.3. 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,

$$\begin{aligned} E_{1,1} \ \leq \quad & \mathbb{E}_p^{\otimes n}\,\lambda^d\left(\bigcup_{i\,:\,X_i \in \Gamma_{1,\delta_n}} \left(B_{\delta_n}(X_i) \setminus \Gamma_p\right)\right) \\ & + \mathbb{E}_p^{\otimes n}\,\lambda^d\left(\bigcup_{i\,:\,X_i \in \Gamma_{2,\delta_n} \cap \{p \leq \alpha_n\}} \left(B_{\delta_n}(X_i) \setminus \Gamma_p\right)\right) \\ & + \mathbb{E}_p^{\otimes n}\,\lambda^d\left(\bigcup_{i\,:\,X_i \in \Gamma_{2,\delta_n} \cap \{p > \alpha_n\}} \left(B_{\delta_n}(X_i) \setminus \Gamma_p\right)\right). \end{aligned} \tag{7.50}$$

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

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

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

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

$$
\begin{aligned}
&\leq \; V_d \, n \, \delta_n^d \, \alpha_n \cdot \lambda^d \left(\Gamma_{2,\delta_n}\right) \\
&\leq \; c(\beta, L) \, \xi_2 \, \delta_n^{\gamma\beta} (\log n)^2,
\end{aligned}
$$

where V_d again denotes the volume of the d-dimensional unit ball. Considering the third expectation in (7.50), 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 (4.1) 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 is contained in $\Gamma_p^{-\delta_n}$ and consequently $B_{\delta_n}(z) \subset \Gamma_p$. 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 (7.49), only the points $x \in \mathbb{R}^d$ that belong to $(\bigcup_{i=1}^n B_{\delta_n}(X_i))^c \setminus \Gamma_p$ 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 (7.50). 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 (B.1), and Lemma A.2,

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

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 (7.48) has to be estimated. Since $\Gamma_p \setminus \hat{\Gamma}_n$ is a subset of the support Γ_p, Jensen's inequality and Lemma 7.5 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

$$\begin{aligned}E_{2,q} &= \int_{\chi_q} p(x)\Big(\mathbb{P}^{\otimes n}(A_{x,1}) + \mathbb{P}^{\otimes n}(A_{x,2})\Big)d\lambda^d(x)\\ &\leq \int_{\chi_q} p(x)\Big(\mathbb{P}^{\otimes n}(A_{x,1}\cap B_x) + \mathbb{P}^{\otimes n}(B_x^c) + \mathbb{P}^{\otimes n}(A_{x,2})\Big)d\lambda^d(x) \qquad (7.51)\end{aligned}$$

with

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

We start with the first probability in (7.51)

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

and the construction of Lepski's method controls the second term $|\hat{p}_{n,\hat{\jmath}}(x)-\hat{p}_{n,\bar{\jmath}}(x)|$, see (B.1). This yields, together with a bias variance decomposition

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

where we used the definition of $B_{x,\bar{\jmath}}$ and Lemma A.2 in the second step. The lemma also yields for $x\in\chi_q$, $q\ge 1$

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

such that

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

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

$$\mathbb{P}^{\otimes n}(A_{x,1}\cap B_{x,\bar{\jmath}})\cdot \mathbb{1}\{x\in \chi_q\}\ \le\ 2\exp\left(-\frac{\log n}{4}\,2^{q/2}\right) \tag{7.52}$$

by Lemma A.3. Furthermore, Lemma 7.3 can be used to bound the probability of B_x^c by

$$\mathbb{P}^{\otimes n}(B^c_{x,\bar{j}}) \;\leq\; 2\exp\left(-\frac{3}{32\|K\|^2_{\sup}}\log^2 n\right). \tag{7.53}$$

A sufficiently tight bound on the probability of $A_{x,2}$ in inequality (7.51) is required. By definition of $A_{x,2}$,

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

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 A.2,

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

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 (7.47) satisfies

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

Lemma A.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 3.2 line by line straightforwardly, and arrive as in (7.23) and (7.26) at

$$\mathbb{P}^{\otimes n}(A_{x,2}) \;\leq\; \exp\Big(-\tilde{c}_{16}(\beta, L)\log n + \log\big(|\mathcal{J}|\big)\Big) \tag{7.54}$$

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 (7.52), (7.53) and (7.54) into (7.51) and applying the margin condition, we arrive at

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

$$
\begin{aligned}
&\leq (2^{q+1}\alpha_n)\Bigg\{2\exp\left(-\frac{\log n}{4}\,2^{(q+1)/2}\right) + 2\exp\left(-\frac{3}{32\|K\|_{\sup}^2}\log^2 n\right)\\
&\qquad + 2\exp\Big(-\tilde{c}_{16}(\beta,L)\log n + \log\big(|\mathcal{J}|\big)\Big)\Bigg\}\int_{\chi_q} d\lambda^d(x)\\
&\leq \kappa_2(2^{q+1}\alpha_n)^{1+\gamma}\Bigg\{2\exp\left(-\frac{\log n}{4}\,2^{(q+1)/2}\right) + 2\exp\left(-\frac{3}{32\|K\|_{\sup}^2}\log^2 n\right)\\
&\qquad + 2\exp\Big(-\tilde{c}_{16}(\beta,L)\log n + \log\big(|\mathcal{J}|\big)\Big)\Bigg\}\\
&\leq \alpha_n^{1+\gamma}\cdot 2\kappa_2\Bigg\{\exp\left(-\frac{\log n}{4}\,2^{(q+1)/2} + (q+1)(1+\gamma)\log 2\right)\\
&\qquad\qquad + \exp\left(-\frac{3}{32\|K\|_{\sup}^2}\log^2 n + (q+1)(1+\gamma)\log 2\right)\\
&\qquad\qquad + \exp\Big(-\tilde{c}_{16}(\beta,L)\log n + (q+1)(1+\gamma)\log 2 + \log\big(|\mathcal{J}|\big)\Big)\Bigg\}\\
&=: \alpha_n^{1+\gamma}\Delta_{n,q}(\beta,L,\gamma,\kappa).
\end{aligned}
$$

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

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

□

APPENDIX A: AUXILIARY LEMMATA

LEMMA A.1. *(i) For any (β,L) with $0<\beta_i\leq 2$ 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), $0<\beta_i<\infty$, $i=1,\ldots,d$,

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

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

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

where K is a rotation invariant kernel supported on $B_1(0)$.

Proof. (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 $\beta_i \leq 1$,

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

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

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

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 (A.1) implies

$$\begin{aligned}|S_i| &\leq L\int |x_i|^{\beta_i} K^2(x)\, d\lambda^d(x)\cdot h_i^{\beta_i}\\ &\leq \frac{\|K\|_2^2\, p(t)}{2d}.\end{aligned} \tag{A.3}$$

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 $x \in B_1(0)$. Hence, the quantity

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

is bounded from above by

$$\begin{aligned} |S_i| &\leq L \int |x_i|^{\beta_i} K^2(x) d\lambda^d(x) \cdot h_i^{\beta_i} \\ &\leq \frac{\|K\|_2^2 \, p(t)}{2d}, \end{aligned} \tag{A.4}$$

which proves together with (A.2) 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 $t = \operatorname{argmax}_{x \in [0,1]} |f(x)|$ 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 $f \in \{g \in \mathcal{H}_1(\beta, L) : \|g\|_{\sup} \leq D\}$ and $c_{27} > 0$, there exists a constant $c_{25}(\beta, L) > 0$, such that for every $h \leq c_{27} \, |f(t)|^{1/\beta}$, there exists an interval

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

with

$$\lambda(I_t(f,h)) \geq c_{25}(\beta, L) \, h \tag{A.5}$$

and the property

$$\frac{1}{2} |f(t)| \ \leq \ |f(x)| \qquad \text{for every } x \in I_t(f,h). \tag{A.6}$$

This in turn implies

$$\begin{aligned} \frac{1}{n} ((K_h)^2 * |f|)(t) &\geq \frac{1}{nh^2} \int_{J_t(f,h)} K^2 \left(\frac{t-x}{h} \right) |f(x)| d\lambda(x) \\ &\geq \frac{K^2(0)}{4} \cdot \frac{1}{nh^2} \int_{J_t(f,h)} |f(x)| d\lambda(x) \end{aligned}$$

$$\geq \frac{K^2(0)}{4} \cdot \frac{1}{nh^2} \int_{I_t(f,h)} |f(x)| d\lambda(x)$$
$$\geq \frac{K^2(0)}{8} \cdot \frac{|f(t)|}{nh^2} \cdot \lambda(I_t(f,h))$$
$$\geq \frac{K^2(0)\, c_{25}(\beta, L)}{8} \cdot \frac{|f(t)|}{nh}.$$

It remains to prove the existence of such an interval $I_t(f,h)$ with properties (A.5) and (A.6). For $\beta \leq 1$, 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 $\|u_f\|_{\sup} \geq 1$. Taylor expansion around any point $y \in [-1,1]$ provides the approximation

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

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

$$\text{(A.7)} \qquad \left|(x-y)\,u_f'(y) + \ldots + \frac{(x-y)^{\lfloor\beta\rfloor}}{\lfloor\beta\rfloor!}\,u_f^{(\lfloor\beta\rfloor)}(y)\right| \leq 2\|u_f\|_{\sup} + L\,|x-y|^\beta \leq 2\|u_f\|_{\sup} + 2^\beta L.$$

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

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

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 (A.7) 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

$$\text{(A.8)} \qquad |u_f(x) - u_f(x_0)| \leq \frac{1}{2}|u_f(x_0)|$$

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 (A.8) 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 (A.8),

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

for all

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

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

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

We have for any x with $\|x - t\|_2 \leq h$

$$d(\partial\Gamma_p, x) \leq h + d(\partial\Gamma_p, t).$$

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

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

it follows that

$$\sup_{x \in B_h(t)} p(x) \leq L\, d(\partial\Gamma_p, x)^\beta \leq L\left(h + d(\partial\Gamma_p, t)\right)^\beta.$$

□

LEMMA A.2. *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 (7.7) and the bias upper bound B_t as given in (7.8),*

$$B_t(\bar{j}) \;\leq\; c_{15}(\beta, L)\sqrt{\sigma_{t,\mathrm{trunc}}^2(\bar{j}) \log n} \tag{A.9}$$

$$\sqrt{\sigma_{t,\mathrm{trunc}}^2(\bar{j})} \;\leq\; c_{21}(\beta, L)\left\{\left(\frac{\log n}{n}\right)^{\frac{\bar{\beta}}{\bar{\beta}+1}} \vee \left(\frac{p(t)\log n}{n}\right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}}\right\}. \tag{A.10}$$

PROOF. 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 (A.9): Recall the definition of the reference bandwidth in (7.6), which for $p(t) \le (\log n/n)^{\bar\beta/(\bar\beta+1)}$ is equal to

$$\bar h_i = c_{11}(\beta, L)\left(\frac{\log n}{n}\right)^{\frac{\bar\beta}{\bar\beta+1}\frac{1}{\beta_i}}, \quad i = 1, \dots, d. \tag{A.11}$$

The corresponding truncation level satisfies

$$\frac{\log^2 n}{n^2 \prod_{i=1}^d \bar h_i^2} = c_{11}(\beta, L)^{-2d}\left(\frac{\log n}{n}\right)^{\frac{2\bar\beta}{\bar\beta+1}}. \tag{A.12}$$

Consequently,

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

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

$$\bar h_i = c_{11}(\beta, L)\left(\frac{p(t)\log n}{n}\right)^{\frac{\bar\beta}{2\bar\beta+1}\frac{1}{\beta_i}}, \quad i = 1, \dots, d, \tag{A.13}$$

and therefore

$$\begin{aligned}
B_t(\bar\jmath) &\le \left(L\sum_{i=1}^d c_{12,i}(\beta)\, c_{11}(\beta, L)^{\beta_i}\right)\left(\frac{p(t)\log n}{n}\right)^{\frac{\bar\beta}{2\bar\beta+1}}\\
&= \left(L\sum_{i=1}^d c_{12,i}(\beta)\, c_{11}(\beta, L)^{\beta_i}\right) c_{11}(\beta, L)^{d/2}\cdot\left(\frac{p(t)\log n}{n\prod_{i=1}^d \bar h_i}\right)^{1/2}\\
&\le \left(L\sum_{i=1}^d c_{12,i}(\beta)\, c_{11}(\beta, L)^{\beta_i}\right) c_{11}(\beta, L)^{d/2}\cdot\left(\frac{p(t)\log n}{n2^{-|\bar\jmath|}}\right)^{1/2}.
\end{aligned} \tag{A.14}$$

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

$$2^{-\bar\jmath_i} \le \bar h_i \le c_{11}(\beta, L)\, p(t)^{1/\beta_i} \quad \text{for all } i = 1, \dots, d, \tag{A.15}$$

Lemma A.1 (i) yields

$$\frac{p(t)\log n}{n2^{-|\bar{j}|}} \le \frac{2}{\|K\|_2^2} \cdot \sigma_t^2(\bar{j})\log n$$

and together with (A.14)

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

Proof of (A.10): For $p(t) \le (\log n/n)^{\bar{\beta}/(\bar{\beta}+1)}$, the reference bandwidth $\bar{h}$ is given by (A.11). Hence by (A.12),

$$\text{(A.16)} \qquad \frac{\log^2 n}{n^2(2^{-|\bar{j}|})^2} \le 2^{2d}\frac{\log^2 n}{n^2\prod_{i=1}^d \bar{h}_i^2} = \left(\frac{2}{c_{11}(\beta,L)}\right)^{2d}\left(\frac{\log n}{n}\right)^{\frac{2\bar{\beta}}{\bar{\beta}+1}}.$$

Furthermore, by (A.2), (A.3) and (A.4),

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

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

$$\text{(A.17)} \qquad \sqrt{\sigma_{t,\mathrm{trunc}}^2(\bar{j})} \;\le\; c(\beta,L)\left(\frac{\log n}{n}\right)^{\frac{\bar{\beta}}{\bar{\beta}+1}}.$$

For $p(t) > (\log n/n)^{\bar{\beta}/(\bar{\beta}+1)}$, the reference bandwidth $\bar{h}$ is given by (A.13) and hence

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

Furthermore, since $\bar{j}$ satisfies property (A.15), Lemma A.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 (A.17)

$$\begin{aligned}\sqrt{\sigma_{t,\mathrm{trunc}}^2(\bar{j})} \;&\leq c(\beta, L)\cdot \begin{cases} \left(\frac{\log n}{n}\right)^{\frac{\bar{\beta}}{\bar{\beta}+1}} & \text{, if } \; p(t) \leq \left(\frac{\log n}{n}\right)^{\frac{\bar{\beta}}{\bar{\beta}+1}} \\ \left(\frac{p(t)\log n}{n}\right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}} & \text{, if } \; p(t) > \left(\frac{\log n}{n}\right)^{\frac{\bar{\beta}}{\bar{\beta}+1}} \end{cases} \\ &= c(\beta, L)\cdot \left\{ \left(\frac{\log n}{n}\right)^{\frac{\bar{\beta}}{\bar{\beta}+1}} \vee \left(\frac{p(t)\log n}{n}\right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}} \right\}.\end{aligned}$$

□

Lemma A.3. *For any (non-random) index $j = (j_1, \ldots, j_d)$, the tail probabilities of the random variable*

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

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.

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

$$Y_i := \frac{\frac{1}{n2^{-|j|}}\left(K\left(\frac{t_1 - X_{i,1}}{2^{-j_1}}, \ldots, \frac{t_d - X_{i,d}}{2^{-j_d}}\right) - \mathbb{E}_p^{\otimes n} K\left(\frac{t_1 - X_{i,1}}{2^{-j_1}}, \ldots, \frac{t_d - X_{i,d}}{2^{-j_d}}\right)\right)}{\sqrt{\sigma_{t,\mathrm{trunc}}^2(j)\log n}}.$$

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. □

LEMMA A.4. *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.

PROOF. Fubini's theorem and the classical moment bound for the exponential distribution reveal

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

□

APPENDIX B: CONSTRUCTION OF THE DENSITY ESTIMATOR

Our estimation scheme is inspired by the anisotropic procedure of Kerkyacharian, Lepski and Picard (2001) and Klutchnikoff (2005), developed in the Gaussian white noise model, but uses the truncated variance estimator (B.2) as a threshold instead, which is sensitive to small values of the density. Unlike in the univariate or isotropic multivariate case, the problem of a missing notion of monotonicity becomes apparent in the anisotropic case. More precisely, neither the variance nor the bias provides an immediate monotone behavior in the bandwidth. Consequently, the idea of mimicking the bias-variance trade-off fails and the estimation scheme has to take this problem into account. Furthermore, in view of our goal of adaptation to lowest density regions, a truncated variance estimator, which is sensitive to small values of the density, is involved both as a threshold in the anisotropic Lepski method and in the ordering of the bandwidths. 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

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

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. It is well known that $n \prod h_i \to \infty$ is a necessary condition for consistency of the kernel density estimator $\hat{p}_{n,h}(t)$ and thus the bandwidth components should not be too small simultaneously. The logarithmic factor in the grid's lower limitation prevents the truncation level from getting too large and thus ensures that $\sigma^2_{t,\mathrm{trunc}}(h)$ does not exceed the order of the classical variance bound (3.1). 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

$$\begin{aligned} \mathcal{A} = \mathcal{A}(t) := \Big\{ j \in \mathcal{J} : |\hat{p}_{n,j\wedge m}(t) - \hat{p}_{n,m}(t)| \le c_{14} \sqrt{\hat{\sigma}_t^2(m) \log n} \\ \text{for all } m \in \mathcal{J} \text{ with } \hat{\sigma}_t^2(m) \ge \hat{\sigma}_t^2(j) \Big\}, \end{aligned} \tag{B.1}$$

with a properly chosen constant $c_{14} = c_{14}(\beta^*, L^*)$ satisfying the constraint (7.25) appearing in the proof of Theorem 3.2. Here, both the threshold and the ordering of bandwidths are defined via the truncated variance estimator

$$\begin{aligned} \hat{\sigma}_t^2(h) &:= \min \left\{ \tilde{\sigma}^2_{t,\mathrm{trunc}}(h), \frac{\|K\|_2^2 \, c_1}{n \prod_{i=1}^d h_i} \right\} \\ &= \min \left\{ \max \left[\frac{\log^2 n}{n^2 \prod_{i=1}^d h_i^2}, \; \frac{1}{n^2 \prod_{i=1}^d h_i^2} \sum_{i=1}^n K^2 \left(\frac{t - X_i}{h} \right) \right], \; \frac{\|K\|_2^2 \, c_1}{n \prod_{i=1}^d h_i} \right\} \end{aligned} \tag{B.2}$$

where $c_1 = c_1(\beta^*, L^*)$ is an upper bound on $c_1(\beta, L)$ in the range of adaptation. The threshold in (B.1) 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

$$\hat{j} = \hat{j}(t) \in \underset{j \in \mathcal{A}}{\arg\min}\; \hat{\sigma}_t^2(j). \tag{B.3}$$

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 (B.2) "for all $m \in \mathcal{J}$ with $\hat{\sigma}_t^2(m) \ge \hat{\sigma}_t^2(j)$" by the classical order "for all $m \in \mathcal{J}$ with $m \ge j$" as the componentwise ordering is the same for all components.

APPENDIX C: (β, L)-REGULAR KERNELS

The proofs of Theorem 3.1, Theorem 3.3 and Theorem 4.4 make use of the following specific construction of functions with prescribed Hölder regularity (β, L), which is taken from Rigollet and Vert (2009). Note that it works only for $\beta \leq 2$ because the second derivative is not continuous. Define the function $K : \mathbb{R}^d \to \mathbb{R}$ by

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

with a 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 dependence on β is omitted when there is no ambiguity. If $B_\varepsilon(x_0) := \{x \in \mathbb{R}^d : \|x - x_0\|_2 \leq \varepsilon\}$ denotes the closed Euclidean ball with radius ε around x_0, the function K is supported on $B_1(0)$, integrates to one 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

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

the function $K_g(x;\beta)$ is supported on $B_g(0)$ 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 $K_i = K_{g_{\beta_i,L,1}}$.

REFERENCES

[1] Aaron, C. (2013). Estimation of the support of the density and its boundary using Random Polyhedron. hal-00786393

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

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

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

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

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

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

[8] Brunel, V.-E. (2013). Adaptive estimation of convex and polytopal density support. arXiv:1309.6602

[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] 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

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

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

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

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

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

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

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

[18] Deheuvels, P. (2000). Uniform limit laws for kernel density estimators on possibly unbounded intervals. *Recent Advances in Reliability Theory: Methodology, Practice and Inference.* (N. Nimlios and M. Nikulin, eds.) Birkhäuser, Boston. **32** 477–492. MR1783500

[19] De La Peña, V.H., Klass, M.J. and Lai, T.Z. (2004). Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws. *Ann. Probab.* **32** 1902–1933. MR2565565

[20] 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

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

[22] 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

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

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

[25] Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Annales de l'Institut Henri Poincaré (B) Probabilités & Statistiques* **38** 907–921. MR1955344

[26] Giné, E., Koltchinskii, V. and Zinn, J. (2004). Weighted uniform consistency of kernel density estimators. *Ann. Probab.* **32** 2570–2605. MR2078551

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

[28] 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

[29] Goldenshluger, A. and Lepski, O. V. (2013). On adaptive minimax estimation on $\mathbb{R}^d$. *Prob. Theory Relat. Fields* **159** 479–543. MR3230001
[30] Grenander, U. (1981). *Abstract Inference*, Wiley, New York. MR0599175
[31] Groeneboom, P., (1988). Limit theorems for convex hulls. *Probab. Theory Relat. Fields* **79** 327–368. MR0959514
[32] Hall, P. (1982). On estimating the endpoint of a distribution. *Ann. Statist.* **10** 556–568. MR0653530
[33] 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
[34] Härdle, W., Park, B. U. and Tsybakov, A. B. (2001). Estimation of non-sharp support boundaries. *J. Multivariate Analysis* **55** 205–218. MR1370400
[35] Jirak, M., Meister, A. and Reiss, M. (2014). Adaptive function estimation in nonparametric regression with one-sided errors. *Ann. Statist.*, to appear.
[36] Juditsky, A. and Lambert-Lacroix, S. (2004). On minimax density estimation on $\mathbb{R}$. *Bernoulli* **10**, 187–220. MR2046772
[37] Kerkyacharian, G., Lepski, O. and Picard, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Relat. Fields* **121** 137–170. MR1863916
[38] Klemelä, J. (2004). Complexity penalized support estimation. *J. Multivariate Analysis* **88** 274–297. MR2025614
[39] Klutchnikoff, N. (2005). *Sur l'estimation adaptative de fonctions anisotropes.*, Ph.D.Thesis. Université Aix-Marseille I.
[40] Korostelev, A. P. and Tsybakov, A. B. (1993). *Minimax Theory of Image Reconstruction*, Springer, New York. MR1226450
[41] Lepski, O. V. (1990). One problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 459–470. MR1091202
[42] Lepski, O. V. (2013). Multivariate density estimation under sup-norm loss: oracle approach, adaptation and independence structure. *Ann. Statist.* **41** 1005–1034. MR3099129
[43] Lepski, O. V. (2014). Adaptive estimation over anisotropic functional classes via oracle approach. arXiv:1405.4504
[44] Liu, L. and Wong, W. H. (2014). Multivariate density estimation based on adaptive partitioning: Convergence rate, variable selection and spatial adaptation. arXiv:1401.2597
[45] Mammen, E. and Tsybakov, A. B. (1995). Asymptotical minimax recovery of sets with smooth boundaries. *Ann. Statist.* **23** 502–524. MR1332579
[46] Mammen, E. and Tsybakov, A. B. (1999). Smooth discrimination analysis. *Ann. Statist.* **27** 1808–1829. MR1765618
[47] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430. MR1425959
[48] Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters - an excess mass approach. *Ann. Statist.* **23** 855–881. MR1345204
[49] Rényi, A. and Sulanke, R. (1963). Über die konvexe Hülle von n zufällig gewählten Punkten. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **2** 75–84. MR0156262
[50] 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
[51] Rigollet, P. and Vert, R. (2009). Optimal rates for plug-in estimators of density level sets. *Bernoulli* **15** 1154–1178. MR2597587
[52] Rohde, A. (2008). Adaptive goodness-of-fit tests based on signed ranks. *Ann. Statist.* **36** 1346–1374. MR2418660

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

[54] Samworth, R. J. and Wand, M. P. (2010). Asymptotics and optimal bandwidth selection for highest density region estimation. *Ann. Statist.* **38** 1767–1792. MR2662359

[55] Stute, W. (1984). The oscillation behavior of empirical processes: The multivariate case. *Ann. Probab.* **12** 361–379. MR0735843

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

[57] 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

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

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

[60] 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

Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports

[1] R. Ramlau, G. Teschke, and M. Zhariy. A Compressive Landweber Iteration for Solving Ill-Posed Inverse Problems. Preprint 1, DFG-SPP 1324, September 2008.

[2] G. Plonka. The Easy Path Wavelet Transform: A New Adaptive Wavelet Transform for Sparse Representation of Two-dimensional Data. Preprint 2, DFG-SPP 1324, September 2008.

[3] E. Novak and H. Woźniakowski. Optimal Order of Convergence and (In-) Tractability of Multivariate Approximation of Smooth Functions. Preprint 3, DFG-SPP 1324, October 2008.

[4] M. Espig, L. Grasedyck, and W. Hackbusch. Black Box Low Tensor Rank Approximation Using Fibre-Crosses. Preprint 4, DFG-SPP 1324, October 2008.

[5] T. Bonesky, S. Dahlke, P. Maass, and T. Raasch. Adaptive Wavelet Methods and Sparsity Reconstruction for Inverse Heat Conduction Problems. Preprint 5, DFG-SPP 1324, January 2009.

[6] E. Novak and H. Woźniakowski. Approximation of Infinitely Differentiable Multivariate Functions Is Intractable. Preprint 6, DFG-SPP 1324, January 2009.

[7] J. Ma and G. Plonka. A Review of Curvelets and Recent Applications. Preprint 7, DFG-SPP 1324, February 2009.

[8] L. Denis, D. A. Lorenz, and D. Trede. Greedy Solution of Ill-Posed Problems: Error Bounds and Exact Inversion. Preprint 8, DFG-SPP 1324, April 2009.

[9] U. Friedrich. A Two Parameter Generalization of Lions' Nonoverlapping Domain Decomposition Method for Linear Elliptic PDEs. Preprint 9, DFG-SPP 1324, April 2009.

[10] K. Bredies and D. A. Lorenz. Minimization of Non-smooth, Non-convex Functionals by Iterative Thresholding. Preprint 10, DFG-SPP 1324, April 2009.

[11] K. Bredies and D. A. Lorenz. Regularization with Non-convex Separable Constraints. Preprint 11, DFG-SPP 1324, April 2009.

[12] M. Döhler, S. Kunis, and D. Potts. Nonequispaced Hyperbolic Cross Fast Fourier Transform. Preprint 12, DFG-SPP 1324, April 2009.

[13] C. Bender. Dual Pricing of Multi-Exercise Options under Volume Constraints. Preprint 13, DFG-SPP 1324, April 2009.

[14] T. Müller-Gronbach and K. Ritter. Variable Subspace Sampling and Multi-level Algorithms. Preprint 14, DFG-SPP 1324, May 2009.

[15] G. Plonka, S. Tenorth, and A. Iske. Optimally Sparse Image Representation by the Easy Path Wavelet Transform. Preprint 15, DFG-SPP 1324, May 2009.

[16] S. Dahlke, E. Novak, and W. Sickel. Optimal Approximation of Elliptic Problems by Linear and Nonlinear Mappings IV: Errors in L_2 and Other Norms. Preprint 16, DFG-SPP 1324, June 2009.

[17] B. Jin, T. Khan, P. Maass, and M. Pidcock. Function Spaces and Optimal Currents in Impedance Tomography. Preprint 17, DFG-SPP 1324, June 2009.

[18] G. Plonka and J. Ma. Curvelet-Wavelet Regularized Split Bregman Iteration for Compressed Sensing. Preprint 18, DFG-SPP 1324, June 2009.

[19] G. Teschke and C. Borries. Accelerated Projected Steepest Descent Method for Nonlinear Inverse Problems with Sparsity Constraints. Preprint 19, DFG-SPP 1324, July 2009.

[20] L. Grasedyck. Hierarchical Singular Value Decomposition of Tensors. Preprint 20, DFG-SPP 1324, July 2009.

[21] D. Rudolf. Error Bounds for Computing the Expectation by Markov Chain Monte Carlo. Preprint 21, DFG-SPP 1324, July 2009.

[22] M. Hansen and W. Sickel. Best m-term Approximation and Lizorkin-Triebel Spaces. Preprint 22, DFG-SPP 1324, August 2009.

[23] F.J. Hickernell, T. Müller-Gronbach, B. Niu, and K. Ritter. Multi-level Monte Carlo Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 23, DFG-SPP 1324, August 2009.

[24] S. Dereich and F. Heidenreich. A Multilevel Monte Carlo Algorithm for Lévy Driven Stochastic Differential Equations. Preprint 24, DFG-SPP 1324, August 2009.

[25] S. Dahlke, M. Fornasier, and T. Raasch. Multilevel Preconditioning for Adaptive Sparse Optimization. Preprint 25, DFG-SPP 1324, August 2009.

[26] S. Dereich. Multilevel Monte Carlo Algorithms for Lévy-driven SDEs with Gaussian Correction. Preprint 26, DFG-SPP 1324, August 2009.

[27] G. Plonka, S. Tenorth, and D. Roşca. A New Hybrid Method for Image Approximation using the Easy Path Wavelet Transform. Preprint 27, DFG-SPP 1324, October 2009.

[28] O. Koch and C. Lubich. Dynamical Low-rank Approximation of Tensors. Preprint 28, DFG-SPP 1324, November 2009.

[29] E. Faou, V. Gradinaru, and C. Lubich. Computing Semi-classical Quantum Dynamics with Hagedorn Wavepackets. Preprint 29, DFG-SPP 1324, November 2009.

[30] D. Conte and C. Lubich. An Error Analysis of the Multi-configuration Time-dependent Hartree Method of Quantum Dynamics. Preprint 30, DFG-SPP 1324, November 2009.

[31] C. E. Powell and E. Ullmann. Preconditioning Stochastic Galerkin Saddle Point Problems. Preprint 31, DFG-SPP 1324, November 2009.

[32] O. G. Ernst and E. Ullmann. Stochastic Galerkin Matrices. Preprint 32, DFG-SPP 1324, November 2009.

[33] F. Lindner and R. L. Schilling. Weak Order for the Discretization of the Stochastic Heat Equation Driven by Impulsive Noise. Preprint 33, DFG-SPP 1324, November 2009.

[34] L. Kämmerer and S. Kunis. On the Stability of the Hyperbolic Cross Discrete Fourier Transform. Preprint 34, DFG-SPP 1324, December 2009.

[35] P. Cerejeiras, M. Ferreira, U. Kähler, and G. Teschke. Inversion of the noisy Radon transform on $SO(3)$ by Gabor frames and sparse recovery principles. Preprint 35, DFG-SPP 1324, January 2010.

[36] T. Jahnke and T. Udrescu. Solving Chemical Master Equations by Adaptive Wavelet Compression. Preprint 36, DFG-SPP 1324, January 2010.

[37] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Irregular Shearlet Frames: Geometry and Approximation Properties. Preprint 37, DFG-SPP 1324, February 2010.

[38] G. Kutyniok and W.-Q Lim. Compactly Supported Shearlets are Optimally Sparse. Preprint 38, DFG-SPP 1324, February 2010.

[39] M. Hansen and W. Sickel. Best m-Term Approximation and Tensor Products of Sobolev and Besov Spaces – the Case of Non-compact Embeddings. Preprint 39, DFG-SPP 1324, March 2010.

[40] B. Niu, F.J. Hickernell, T. Müller-Gronbach, and K. Ritter. Deterministic Multi-level Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 40, DFG-SPP 1324, March 2010.

[41] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Construction of Compactly Supported Shearlet Frames. Preprint 41, DFG-SPP 1324, March 2010.

[42] C. Bender and J. Steiner. Error Criteria for Numerical Solutions of Backward SDEs. Preprint 42, DFG-SPP 1324, April 2010.

[43] L. Grasedyck. Polynomial Approximation in Hierarchical Tucker Format by Vector-Tensorization. Preprint 43, DFG-SPP 1324, April 2010.

[44] M. Hansen und W. Sickel. Best m-Term Approximation and Sobolev-Besov Spaces of Dominating Mixed Smoothness - the Case of Compact Embeddings. Preprint 44, DFG-SPP 1324, April 2010.

[45] P. Binev, W. Dahmen, and P. Lamby. Fast High-Dimensional Approximation with Sparse Occupancy Trees. Preprint 45, DFG-SPP 1324, May 2010.

[46] J. Ballani and L. Grasedyck. A Projection Method to Solve Linear Systems in Tensor Format. Preprint 46, DFG-SPP 1324, May 2010.

[47] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence Rates for Greedy Algorithms in Reduced Basis Methods. Preprint 47, DFG-SPP 1324, May 2010.

[48] S. Kestler and K. Urban. Adaptive Wavelet Methods on Unbounded Domains. Preprint 48, DFG-SPP 1324, June 2010.

[49] H. Yserentant. The Mixed Regularity of Electronic Wave Functions Multiplied by Explicit Correlation Factors. Preprint 49, DFG-SPP 1324, June 2010.

[50] H. Yserentant. On the Complexity of the Electronic Schrödinger Equation. Preprint 50, DFG-SPP 1324, June 2010.

[51] M. Guillemard and A. Iske. Curvature Analysis of Frequency Modulated Manifolds in Dimensionality Reduction. Preprint 51, DFG-SPP 1324, June 2010.

[52] E. Herrholz and G. Teschke. Compressive Sensing Principles and Iterative Sparse Recovery for Inverse and Ill-Posed Problems. Preprint 52, DFG-SPP 1324, July 2010.

[53] L. Kämmerer, S. Kunis, and D. Potts. Interpolation Lattices for Hyperbolic Cross Trigonometric Polynomials. Preprint 53, DFG-SPP 1324, July 2010.

[54] G. Kutyniok and W.-Q Lim. Shearlets on Bounded Domains. Preprint 54, DFG-SPP 1324, July 2010.

[55] A. Zeiser. Wavelet Approximation in Weighted Sobolev Spaces of Mixed Order with Applications to the Electronic Schrödinger Equation. Preprint 55, DFG-SPP 1324, July 2010.

[56] G. Kutyniok, J. Lemvig, and W.-Q Lim. Compactly Supported Shearlets. Preprint 56, DFG-SPP 1324, July 2010.

[57] A. Zeiser. On the Optimality of the Inexact Inverse Iteration Coupled with Adaptive Finite Element Methods. Preprint 57, DFG-SPP 1324, July 2010.

[58] S. Jokar. Sparse Recovery and Kronecker Products. Preprint 58, DFG-SPP 1324, August 2010.

[59] T. Aboiyar, E. H. Georgoulis, and A. Iske. Adaptive ADER Methods Using Kernel-Based Polyharmonic Spline WENO Reconstruction. Preprint 59, DFG-SPP 1324, August 2010.

[60] O. G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann. On the Convergence of Generalized Polynomial Chaos Expansions. Preprint 60, DFG-SPP 1324, August 2010.

[61] S. Holtz, T. Rohwedder, and R. Schneider. On Manifolds of Tensors of Fixed TT-Rank. Preprint 61, DFG-SPP 1324, September 2010.

[62] J. Ballani, L. Grasedyck, and M. Kluge. Black Box Approximation of Tensors in Hierarchical Tucker Format. Preprint 62, DFG-SPP 1324, October 2010.

[63] M. Hansen. On Tensor Products of Quasi-Banach Spaces. Preprint 63, DFG-SPP 1324, October 2010.

[64] S. Dahlke, G. Steidl, and G. Teschke. Shearlet Coorbit Spaces: Compactly Supported Analyzing Shearlets, Traces and Embeddings. Preprint 64, DFG-SPP 1324, October 2010.

[65] W. Hackbusch. Tensorisation of Vectors and their Efficient Convolution. Preprint 65, DFG-SPP 1324, November 2010.

[66] P. A. Cioica, S. Dahlke, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Spatial Besov Regularity for Stochastic Partial Differential Equations on Lipschitz Domains. Preprint 66, DFG-SPP 1324, November 2010.

[67] E. Novak and H. Woźniakowski. On the Power of Function Values for the Approximation Problem in Various Settings. Preprint 67, DFG-SPP 1324, November 2010.

[68] A. Hinrichs, E. Novak, and H. Woźniakowski. The Curse of Dimensionality for Monotone and Convex Functions of Many Variables. Preprint 68, DFG-SPP 1324, November 2010.

[69] G. Kutyniok and W.-Q Lim. Image Separation Using Shearlets. Preprint 69, DFG-SPP 1324, November 2010.

[70] B. Jin and P. Maass. An Analysis of Electrical Impedance Tomography with Applications to Tikhonov Regularization. Preprint 70, DFG-SPP 1324, December 2010.

[71] S. Holtz, T. Rohwedder, and R. Schneider. The Alternating Linear Scheme for Tensor Optimisation in the TT Format. Preprint 71, DFG-SPP 1324, December 2010.

[72] T. Müller-Gronbach and K. Ritter. A Local Refinement Strategy for Constructive Quantization of Scalar SDEs. Preprint 72, DFG-SPP 1324, December 2010.

[73] T. Rohwedder and R. Schneider. An Analysis for the DIIS Acceleration Method used in Quantum Chemistry Calculations. Preprint 73, DFG-SPP 1324, December 2010.

[74] C. Bender and J. Steiner. Least-Squares Monte Carlo for Backward SDEs. Preprint 74, DFG-SPP 1324, December 2010.

[75] C. Bender. Primal and Dual Pricing of Multiple Exercise Options in Continuous Time. Preprint 75, DFG-SPP 1324, December 2010.

[76] H. Harbrecht, M. Peters, and R. Schneider. On the Low-rank Approximation by the Pivoted Cholesky Decomposition. Preprint 76, DFG-SPP 1324, December 2010.

[77] P. A. Cioica, S. Dahlke, N. Döhring, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Adaptive Wavelet Methods for Elliptic Stochastic Partial Differential Equations. Preprint 77, DFG-SPP 1324, January 2011.

[78] G. Plonka, S. Tenorth, and A. Iske. Optimal Representation of Piecewise Hölder Smooth Bivariate Functions by the Easy Path Wavelet Transform. Preprint 78, DFG-SPP 1324, January 2011.

[79] A. Mugler and H.-J. Starkloff. On Elliptic Partial Differential Equations with Random Coefficients. Preprint 79, DFG-SPP 1324, January 2011.

[80] T. Müller-Gronbach, K. Ritter, and L. Yaroslavtseva. A Derandomization of the Euler Scheme for Scalar Stochastic Differential Equations. Preprint 80, DFG-SPP 1324, January 2011.

[81] W. Dahmen, C. Huang, C. Schwab, and G. Welper. Adaptive Petrov-Galerkin methods for first order transport equations. Preprint 81, DFG-SPP 1324, January 2011.

[82] K. Grella and C. Schwab. Sparse Tensor Spherical Harmonics Approximation in Radiative Transfer. Preprint 82, DFG-SPP 1324, January 2011.

[83] D.A. Lorenz, S. Schiffler, and D. Trede. Beyond Convergence Rates: Exact Inversion With Tikhonov Regularization With Sparsity Constraints. Preprint 83, DFG-SPP 1324, January 2011.

[84] S. Dereich, M. Scheutzow, and R. Schottstedt. Constructive quantization: Approximation by empirical measures. Preprint 84, DFG-SPP 1324, January 2011.

[85] S. Dahlke and W. Sickel. On Besov Regularity of Solutions to Nonlinear Elliptic Partial Differential Equations. Preprint 85, DFG-SPP 1324, January 2011.

[86] S. Dahlke, U. Friedrich, P. Maass, T. Raasch, and R.A. Ressel. An adaptive wavelet method for parameter identification problems in parabolic partial differential equations. Preprint 86, DFG-SPP 1324, January 2011.

[87] A. Cohen, W. Dahmen, and G. Welper. Adaptivity and Variational Stabilization for Convection-Diffusion Equations. Preprint 87, DFG-SPP 1324, January 2011.

[88] T. Jahnke. On Reduced Models for the Chemical Master Equation. Preprint 88, DFG-SPP 1324, January 2011.

[89] P. Binev, W. Dahmen, R. DeVore, P. Lamby, D. Savu, and R. Sharpley. Compressed Sensing and Electron Microscopy. Preprint 89, DFG-SPP 1324, March 2011.

[90] P. Binev, F. Blanco-Silva, D. Blom, W. Dahmen, P. Lamby, R. Sharpley, and T. Vogt. High Quality Image Formation by Nonlocal Means Applied to High-Angle Annular Dark Field Scanning Transmission Electron Microscopy (HAADF-STEM). Preprint 90, DFG-SPP 1324, March 2011.

[91] R. A. Ressel. A Parameter Identification Problem for a Nonlinear Parabolic Differential Equation. Preprint 91, DFG-SPP 1324, May 2011.

[92] G. Kutyniok. Data Separation by Sparse Representations. Preprint 92, DFG-SPP 1324, May 2011.

[93] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok. Introduction to Compressed Sensing. Preprint 93, DFG-SPP 1324, May 2011.

[94] H.-C. Kreusler and H. Yserentant. The Mixed Regularity of Electronic Wave Functions in Fractional Order and Weighted Sobolev Spaces. Preprint 94, DFG-SPP 1324, June 2011.

[95] E. Ullmann, H. C. Elman, and O. G. Ernst. Efficient Iterative Solvers for Stochastic Galerkin Discretizations of Log-Transformed Random Diffusion Problems. Preprint 95, DFG-SPP 1324, June 2011.

[96] S. Kunis and I. Melzer. On the Butterfly Sparse Fourier Transform. Preprint 96, DFG-SPP 1324, June 2011.

[97] T. Rohwedder. The Continuous Coupled Cluster Formulation for the Electronic Schrödinger Equation. Preprint 97, DFG-SPP 1324, June 2011.

[98] T. Rohwedder and R. Schneider. Error Estimates for the Coupled Cluster Method. Preprint 98, DFG-SPP 1324, June 2011.

[99] P. A. Cioica and S. Dahlke. Spatial Besov Regularity for Semilinear Stochastic Partial Differential Equations on Bounded Lipschitz Domains. Preprint 99, DFG-SPP 1324, July 2011.

[100] L. Grasedyck and W. Hackbusch. An Introduction to Hierarchical (H-) Rank and TT-Rank of Tensors with Examples. Preprint 100, DFG-SPP 1324, August 2011.

[101] N. Chegini, S. Dahlke, U. Friedrich, and R. Stevenson. Piecewise Tensor Product Wavelet Bases by Extensions and Approximation Rates. Preprint 101, DFG-SPP 1324, September 2011.

[102] S. Dahlke, P. Oswald, and T. Raasch. A Note on Quarkonial Systems and Multilevel Partition of Unity Methods. Preprint 102, DFG-SPP 1324, September 2011.

[103] A. Uschmajew. Local Convergence of the Alternating Least Squares Algorithm For Canonical Tensor Approximation. Preprint 103, DFG-SPP 1324, September 2011.

[104] S. Kvaal. Multiconfigurational time-dependent Hartree method for describing particle loss due to absorbing boundary conditions. Preprint 104, DFG-SPP 1324, September 2011.

[105] M. Guillemard and A. Iske. On Groupoid C*-Algebras, Persistent Homology and Time-Frequency Analysis. Preprint 105, DFG-SPP 1324, September 2011.

[106] A. Hinrichs, E. Novak, and H. Woźniakowski. Discontinuous information in the worst case and randomized settings. Preprint 106, DFG-SPP 1324, September 2011.

[107] M. Espig, W. Hackbusch, A. Litvinenko, H. Matthies, and E. Zander. Efficient Analysis of High Dimensional Data in Tensor Formats. Preprint 107, DFG-SPP 1324, September 2011.

[108] M. Espig, W. Hackbusch, S. Handschuh, and R. Schneider. Optimization Problems in Contracted Tensor Networks. Preprint 108, DFG-SPP 1324, October 2011.

[109] S. Dereich, T. Müller-Gronbach, and K. Ritter. On the Complexity of Computing Quadrature Formulas for SDEs. Preprint 109, DFG-SPP 1324, October 2011.

[110] D. Belomestny. Solving optimal stopping problems by empirical dual optimization and penalization. Preprint 110, DFG-SPP 1324, November 2011.

[111] D. Belomestny and J. Schoenmakers. Multilevel dual approach for pricing American style derivatives. Preprint 111, DFG-SPP 1324, November 2011.

[112] T. Rohwedder and A. Uschmajew. Local convergence of alternating schemes for optimization of convex problems in the TT format. Preprint 112, DFG-SPP 1324, December 2011.

[113] T. Görner, R. Hielscher, and S. Kunis. Efficient and accurate computation of spherical mean values at scattered center points. Preprint 113, DFG-SPP 1324, December 2011.

[114] Y. Dong, T. Görner, and S. Kunis. An iterative reconstruction scheme for photoacoustic imaging. Preprint 114, DFG-SPP 1324, December 2011.

[115] L. Kämmerer. Reconstructing hyperbolic cross trigonometric polynomials by sampling along generated sets. Preprint 115, DFG-SPP 1324, February 2012.

[116] H. Chen and R. Schneider. Numerical analysis of augmented plane waves methods for full-potential electronic structure calculations. Preprint 116, DFG-SPP 1324, February 2012.

[117] J. Ma, G. Plonka, and M.Y. Hussaini. Compressive Video Sampling with Approximate Message Passing Decoding. Preprint 117, DFG-SPP 1324, February 2012.

[118] D. Heinen and G. Plonka. Wavelet shrinkage on paths for scattered data denoising. Preprint 118, DFG-SPP 1324, February 2012.

[119] T. Jahnke and M. Kreim. Error bound for piecewise deterministic processes modeling stochastic reaction systems. Preprint 119, DFG-SPP 1324, March 2012.

[120] C. Bender and J. Steiner. A-posteriori estimates for backward SDEs. Preprint 120, DFG-SPP 1324, April 2012.

[121] M. Espig, W. Hackbusch, A. Litvinenkoy, H.G. Matthiesy, and P. Wähnert. Efficient low-rank approximation of the stochastic Galerkin matrix in tensor formats. Preprint 121, DFG-SPP 1324, May 2012.

[122] O. Bokanowski, J. Garcke, M. Griebel, and I. Klompmaker. An adaptive sparse grid semi-Lagrangian scheme for first order Hamilton-Jacobi Bellman equations. Preprint 122, DFG-SPP 1324, June 2012.

[123] A. Mugler and H.-J. Starkloff. On the convergence of the stochastic Galerkin method for random elliptic partial differential equations. Preprint 123, DFG-SPP 1324, June 2012.

[124] P.A. Cioica, S. Dahlke, N. Döhring, U. Friedrich, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R.L. Schilling. On the convergence analysis of Rothe's method. Preprint 124, DFG-SPP 1324, July 2012.

[125] P. Binev, A. Cohen, W. Dahmen, and R. DeVore. Classification Algorithms using Adaptive Partitioning. Preprint 125, DFG-SPP 1324, July 2012.

[126] C. Lubich, T. Rohwedder, R. Schneider, and B. Vandereycken. Dynamical approximation of hierarchical Tucker and Tensor-Train tensors. Preprint 126, DFG-SPP 1324, July 2012.

[127] M. Kovács, S. Larsson, and K. Urban. On Wavelet-Galerkin methods for semilinear parabolic equations with additive noise. Preprint 127, DFG-SPP 1324, August 2012.

[128] M. Bachmayr, H. Chen, and R. Schneider. Numerical analysis of Gaussian approximations in quantum chemistry. Preprint 128, DFG-SPP 1324, August 2012.

[129] D. Rudolf. Explicit error bounds for Markov chain Monte Carlo. Preprint 129, DFG-SPP 1324, August 2012.

[130] P.A. Cioica, K.-H. Kim, K. Lee, and F. Lindner. On the $L_q(L_p)$-regularity and Besov smoothness of stochastic parabolic equations on bounded Lipschitz domains. Preprint 130, DFG-SPP 1324, December 2012.

[131] M. Hansen. $n-$term Approximation Rates and Besov Regularity for Elliptic PDEs on Polyhedral Domains. Preprint 131, DFG-SPP 1324, December 2012.

[132] R. E. Bank and H. Yserentant. On the H^1 -stability of the L_2 -projection onto finite element spaces. Preprint 132, DFG-SPP 1324, December 2012.

[133] M. Gnewuch, S. Mayer, and K. Ritter. On Weighted Hilbert Spaces and Integration of Functions of Infinitely Many Variables. Preprint 133, DFG-SPP 1324, December 2012.

[134] D. Crisan, J. Diehl, P.K. Friz, and H. Oberhauser. Robust Filtering: Correlated Noise and Multidimensional Observation. Preprint 134, DFG-SPP 1324, January 2013.

[135] Wolfgang Dahmen, Christian Plesken, and Gerrit Welper. Double Greedy Algorithms: Reduced Basis Methods for Transport Dominated Problems. Preprint 135, DFG-SPP 1324, February 2013.

[136] Aicke Hinrichs, Erich Novak, Mario Ullrich, and Henryk Wozniakowski. The Curse of Dimensionality for Numerical Integration of Smooth Functions. Preprint 136, DFG-SPP 1324, February 2013.

[137] Markus Bachmayr, Wolfgang Dahmen, Ronald DeVore, and Lars Grasedyck. Approximation of High-Dimensional Rank One Tensors. Preprint 137, DFG-SPP 1324, March 2013.

[138] Markus Bachmayr and Wolfgang Dahmen. Adaptive Near-Optimal Rank Tensor Approximation for High-Dimensional Operator Equations. Preprint 138, DFG-SPP 1324, April 2013.

[139] Felix Lindner. Singular Behavior of the Solution to the Stochastic Heat Equation on a Polygonal Domain. Preprint 139, DFG-SPP 1324, May 2013.

[140] Stephan Dahlke, Dominik Lellek, Shiu Hong Lui, and Rob Stevenson. Adaptive Wavelet Schwarz Methods for the Navier-Stokes Equation. Preprint 140, DFG-SPP 1324, May 2013.

[141] Jonas Ballani and Lars Grasedyck. Tree Adaptive Approximation in the Hierarchical Tensor Format. Preprint 141, DFG-SPP 1324, June 2013.

[142] Harry Yserentant. A short theory of the Rayleigh-Ritz method. Preprint 142, DFG-SPP 1324, July 2013.

[143] M. Hefter and K. Ritter. On Embeddings of Weighted Tensor Product Hilbert Spaces. Preprint 143, DFG-SPP 1324, August 2013.

[144] M. Altmayer and A. Neuenkirch. Multilevel Monte Carlo Quadrature of Discontinuous Payoffs in the Generalized Heston Model using Malliavin Integration by Parts. Preprint 144, DFG-SPP 1324, August 2013.

[145] L. Kämmerer, D. Potts, and T. Volkmer. Approximation of multivariate functions by trigonometric polynomials based on rank-1 lattice sampling. Preprint 145, DFG-SPP 1324, September 2013.

[146] C. Bender, N. Schweizer, and J. Zhuo. A primal-dual algorithm for BSDEs. Preprint 146, DFG-SPP 1324, October 2013.

[147] D. Rudolf. Hit-and-run for numerical integration. Preprint 147, DFG-SPP 1324, October 2013.

[148] D. Rudolf and M. Ullrich. Positivity of hit-and-run and related algorithms. Preprint 148, DFG-SPP 1324, October 2013.

[149] L. Grasedyck, M. Kluge, and S. Krämer. Alternating Directions Fitting (ADF) of Hierarchical Low Rank Tensors. Preprint 149, DFG-SPP 1324, October 2013.

[150] F. Filbir, S. Kunis, and R. Seyfried. Effective discretization of direct reconstruction schemes for photoacoustic imaging in spherical geometries. Preprint 150, DFG-SPP 1324, November 2013.

[151] E. Novak, M. Ullrich, and H. Woźniakowski. Complexity of Oscillatory Integration for Univariate Sobolev Spaces. Preprint 151, DFG-SPP 1324, November 2013.

[152] A. Hinrichs, E. Novak, and M. Ullrich. A Tractability Result for the Clenshaw Curtis Smolyak Algorithm. Preprint 152, DFG-SPP 1324, November 2013.

[153] M. Hein, S. Setzer, L. Jost, and S. Rangapuram. The Total Variation on Hypergraphs - Learning on Hypergraphs Revisited. Preprint 153, DFG-SPP 1324, November 2013.

[154] M. Kovács, S. Larsson, and F. Lindgren. On the Backward Euler Approximation of the Stochastic Allen-Chan Equation. Preprint 154, DFG-SPP 1324, November 2013.

[155] S. Dahlke, M. Fornasier, U. Friedrich, and T. Raasch. Multilevel preconditioning for sparse optimization of functionals with nonconvex fidelity terms. Preprint 155, DFG-SPP 1324, December 2013.

[156] T. Müller-Gronbach, K. Ritter, and L. Yaroslavtseva. On the complexity of computing quadrature formulas for marginal distributions of SDEs. Preprint 156, DFG-SPP 1324, January 2014.

[157] M. Giles, T. Nagapetyan, and K. Ritter. Multi-Level Monte Carlo Approximation of Distribution Functions and Densities. Preprint 157, DFG-SPP 1324, February 2014.

[158] F. Dickmann and N. Schweizer. Faster Comparison of Stopping Times by Nested Conditional Monte Carlo . Preprint 158, DFG-SPP 1324, February 2014.

[159] L. Kämmerer, D. Potts, and T. Volkmer. Approximation of multivariate periodic functions by trigonometric polynomials based on sampling along rank-1 lattice with generating vector of Korobov form . Preprint 159, DFG-SPP 1324, February 2014.

[160] S. Dereich and T. Müller-Gronbach. Quadrature for Self-Affine Distributions on $\mathbb{R}^d$. Preprint 160, DFG-SPP 1324, March 2014.

[161] S. Dereich and S. Li. Multilevel Monte Carlo for Lévy-driven SDEs: central limit theorems for adaptive Euler schemes. Preprint 161, DFG-SPP 1324, March 2014.

[162] M. Kluge. Sampling Rules for Tensor Reconstruction in Hierarchical Tucker Format. Preprint 162, DFG-SPP 1324, April 2014.

[163] D. Rudolf and N. Schweizer. Error Bounds of MCMC for Functions with Unbounded Stationary Variance. Preprint 163, DFG-SPP 1324, April 2014.

[164] E. Novak and D. Rudolf. Tractability of the approximation of high-dimensional rank one tensors. Preprint 164, DFG-SPP 1324, April 2014.

[165] J. Dick and D. Rudolf. Discrepancy estimates for variance bounding Markov chain quasi-Monte Carlo. Preprint 165, DFG-SPP 1324, April 2014.

[166] L. Kämmerer, S. Kunis, I. Melzer, D. Potts, and T. Volkmer. Computational Methods for the Fourier Analysis of Sparse High-Dimensional Functions. Preprint 166, DFG-SPP 1324, April 2014.

[167] T. Müller-Gronbach and L. Yaroslavtseva. Deterministic quadrature formulas for SDEs based on simplified weak Ito-Taylor steps. Preprint 167, DFG-SPP 1324, June 2014.

[168] D. Belomestny, T. Nagapetyan, and V. Shiryaev. Multilevel Path Simulation for Weak Approximation Schemes: Myth or Reality. Preprint 168, DFG-SPP 1324, June 2014.

[169] S. Dahlke, N. Döhring, and S. Kinzel. A class of random functions in non-standard smoothness spaces. Preprint 169, DFG-SPP 1324, September 2014.

[170] T. Patschkowski and A. Rohde. Adaptation to lowest density regions with application to support recovery. Preprint 170, DFG-SPP 1324, September 2014.