

Optimal global rates of convergence for noiseless regression estimation problems with adaptively chosen design *

Michael Kohler[†]

Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany, email: kohler@mathematik.tu-darmstadt.de

September 20, 2013

Abstract

Given the values of a measurable function $m : \mathbb{R}^d \rightarrow \mathbb{R}$ at n arbitrarily chosen points in \mathbb{R}^d the problem of estimating m on whole \mathbb{R}^d , such that the L_1 error (with integration with respect to a fixed but unknown probability measure) of the estimate is small, is considered. Under the assumption that m is (p, C) -smooth (i.e., roughly speaking, m is p -times continuously differentiable) it is shown that the optimal minimax rate of convergence of the L_1 error is $n^{-p/d}$, where the upper bound is valid even if the support of the design measure is unbounded but the design measure satisfies some moment condition. Furthermore it is shown that this rate of convergence cannot be improved even if the function is not allowed to change with the size of the data.

AMS classification: Primary 62G08; secondary 62G05.

Key words and phrases: L_1 -error, minimax rate of convergence, noiseless regression, scattered data approximation.

1 Introduction

In this article the problem of estimating a measurable function $m : \mathbb{R}^d \rightarrow \mathbb{R}$ from n observations of the value of the function m at points $z_1, \dots, z_n \in \mathbb{R}^d$, which might be arbitrarily chosen, is considered. Any estimate of m uses in a first step a strategy to choose the points

$$z_1, z_2 = z_2((z_1, m(z_1))), \dots, z_n = z_n((z_1, m(z_1)), \dots, (z_{n-1}, m(z_{n-1}))) \quad (1)$$

and then uses the data

$$\mathcal{D}_n = \{(z_1, m(z_1)), \dots, (z_n, m(z_n))\} \quad (2)$$

to estimate m by $m_n(\cdot) = m_n(\cdot, \mathcal{D}_n) : \mathbb{R}^d \rightarrow \mathbb{R}$. In numerical analysis this problem is known under the name scattered data approximation (usually with non-adaptively chosen

*Running title: *Noiseless regression with adaptive design*

[†]Corresponding author: Tel. +49-6151-16-6846, Fax. +49-6151-16-6822

points $z_1, \dots, z_n \in \mathbb{R}^d$), see, e.g., Wendland (2010) and the literature cited therein. In this paper it is studied from a statistical point of view.

Motivated by a problem in density estimation, where m_n is used to generate additional data for the density estimate and where the error of the method crucially depends on the L_1 error of m_n (cf., Devroye, Felber and Kohler (2013) and Felber, Kohler and Krzyżak (2013)), the error of m_n is measured in this paper by the L_1 error computed with respect to a fixed but unknown probability measure μ , i.e., by

$$\int |m_n(x) - m(x)| \mu(dx). \quad (3)$$

In order to derive nontrivial rate of convergence results it is assumed in the sequel that the regression function is (p, C) -smooth according to the following definition.

Definition 1 *Let $p = k + \beta$ for some $k \in \mathbb{N}_0$ and $0 < \beta \leq 1$, and let $C > 0$. A function $m : \mathbb{R}^d \rightarrow \mathbb{R}$ is called (p, C) -smooth, if for every $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$ with $\sum_{j=1}^d \alpha_j = k$ the partial derivative $\frac{\partial^k m}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}$ exists and satisfies*

$$\left| \frac{\partial^k m}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}(x) - \frac{\partial^k m}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}(z) \right| \leq C \cdot \|x - z\|^\beta$$

for all $x, z \in \mathbb{R}^d$, where \mathbb{N}_0 is the set of non-negative integers.

In the sequel minimax rate of convergence results for the L_1 error (3) are derived. More precisely, for function classes $\mathcal{F}^{(p,C)}$ of (p, C) -smooth functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$ the behaviour of

$$\inf_{\tilde{m}_n} \sup_{m \in \mathcal{F}^{(p,C)}} \int |\tilde{m}_n(x) - m(x)| \mu(dx)$$

is analyzed, and estimates m_n are constructed such that

$$\int |m_n(x) - m(x)| \mu(dx) \approx \inf_{\tilde{m}_n} \sup_{m \in \mathcal{F}^{(p,C)}} \int |\tilde{m}_n(x) - m(x)| \mu(dx).$$

A related problem is nonparametric regression estimation, where the x -values of the data \mathcal{D}_n defined by (1) and (2) are given by an independent and identically distributed sample of μ and the corresponding function values are observed with additional errors with mean zero. This problem has been extensively studied in the literature. The most popular estimates include kernel regression estimate (cf., e.g., Nadaraya (1964, 1970), Watson (1964), Devroye and Wagner (1980), Stone (1977, 1982), Devroye and Krzyżak (1989) or Kohler, Krzyżak and Walk (2009)), partitioning regression estimate (cf., e.g., Györfi (1981), Beirlant and Györfi (1998) or Kohler, Krzyżak and Walk (2006)), nearest neighbor regression estimate (cf., e.g., Devroye (1982) or Devroye, Györfi, Krzyżak and Lugosi (1994)), least squares estimates (cf., e.g., Lugosi and Zeger (1995) or Kohler (2000)) and smoothing spline estimates (cf., e.g., Wahba (1990) or Kohler and Krzyżak (2001)). Minimax rates of convergence in this context have been derived in Stone (1980,1982,

1985, 1994), Antos, Györfi and Kohler (2000) and Kohler, Krzyżak and Walk (2006, 2009).

Kohler and Krzyżak (2013) have analyzed how the minimax rate of convergence results in Stone (1982) change in case that the function m can be observed without error. The main results there are that firstly for estimating (p, C) -smooth functions no estimate can achieve a rate better than $n^{-p/d}$. Secondly, a nearest neighbor estimate achieves this rate if $p \leq 1$. Thirdly, a nearest neighbor polynomial interpolation estimate achieves this rate for arbitrary $p \in \mathbb{N}$ in case $d = 1$ provided the distribution μ satisfies regularity assumptions (which are satisfied, e.g., in case of the uniform distribution). And fourthly it was shown that without regularity assumption on μ no estimate can achieve a rate of convergence better than n^{-1} . Throughout this paper it was assumed that the support of μ is bounded.

In this article it is investigated how these rate of convergence results change in case that the estimate is allowed to choose the design points, i.e., the points where the function values of m are observed, in an adaptive way as described by (1). Surprisingly, the minimax rate of convergence for estimation of (p, C) -smooth functions still remains $n^{-p/d}$, but this time it is achievable even in case $p/d > 1$ without regularity conditions on the design measure μ . Furthermore it is shown that this rate of convergence cannot be improved even if the function is not allowed to change with the size of the data, and that this rate of convergence can be achieved even if the support of the design measure is unbounded but the design measure satisfies some moment condition.

Throughout the paper the following notation is used: The sets of natural numbers, integers and real numbers are denoted by \mathbb{N} , \mathbb{Z} and \mathbb{R} , resp. For $z \in \mathbb{R}$ the smallest integer greater than or equal to z is denoted by $\lceil z \rceil$, and $\lfloor z \rfloor$ is the largest integer less than or equal to z . For $f : \mathbb{R}^d \rightarrow \mathbb{R}$

$$\|f\|_\infty = \sup_{x \in \mathbb{R}^d} |f(x)|$$

is its supremum norm, and the supremum norm of f on a set $A \subseteq \mathbb{R}^d$ is denoted by

$$\|f\|_{\infty, A} = \sup_{x \in A} |f(x)|$$

$\|x\|$ is the Euclidean norm of a vector $x \in \mathbb{R}^d$. The components of $x \in \mathbb{R}^d$ are denoted by $x^{(1)}, \dots, x^{(d)}$, i.e.,

$$x = (x^{(1)}, \dots, x^{(d)})^T.$$

The support of a probability measure μ defined on the Borel sets in \mathbb{R}^d is abbreviated by

$$\text{supp}(\mu) = \left\{ x \in \mathbb{R}^d : \mu(S_r(x)) > 0 \text{ for all } r > 0 \right\},$$

where $S_r(x)$ is the ball of radius r around x .

The outline of the paper is as follows: The main results are formulated in Section 2. The proofs are contained in Section 3.

2 Main results

In our first result we assume that the support of μ is bounded. In order to simplify the notation we assume w.l.o.g. that $\text{supp}(\mu) = [0, 1]^d$. We will use well-known results from spline theory to show that if we choose in this case the design points z_1, \dots, z_n equidistantly in $[0, 1]^d$, then a properly defined spline approximation of a (p, C) -smooth function achieves the rate of convergence $n^{-p/d}$.

In order to define the spline approximation, we introduce polynomial splines, i.e., sets of piecewise polynomials satisfying a global smoothness condition, and a corresponding B-spline basis consisting of basis functions with compact support as follows:

Choose $K \in \mathbb{N}$ and $M \in \mathbb{N}_0$, and set $u_k = k/K$ ($k \in \mathbb{Z}$). For $k \in \mathbb{Z}$ let $B_{k,M} : \mathbb{R} \rightarrow \mathbb{R}$ be the univariate B-spline of degree M with knot sequence $(u_l)_{l \in \mathbb{Z}}$ and support $\text{supp}(B_{k,M}) = [u_k, u_{k+M+1}]$. In case $M = 0$ this means that $B_{k,0}$ is the indicator function of the interval $[u_k, u_{k+1})$, and for $M = 1$ we have

$$B_{k,1}(x) = \begin{cases} \frac{x-u_k}{u_{k+1}-u_k} & , u_k \leq x \leq u_{k+1}, \\ \frac{u_{k+2}-x}{u_{k+2}-u_{k+1}} & , u_{k+1} < x \leq u_{k+2}, \\ 0 & , \text{else,} \end{cases}$$

(so-called hat-function). The general definition of $B_{k,M}$ can be found, e.g., in de Boor (1978), or in Section 14.1 of Görfi et al. (2002). These B-splines are basis functions of sets of univariate piecewise polynomials of degree M , where the piecewise polynomials are globally $(M-1)$ -times continuously differentiable and where the M -th derivative of the functions have jump points only at the knots u_l ($l \in \mathbb{Z}$).

For $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{Z}^d$ we define the tensor product B-spline $B_{\mathbf{k},M} : \mathbb{R}^d \rightarrow \mathbb{R}$ by

$$B_{\mathbf{k},M}(x^{(1)}, \dots, x^{(d)}) = B_{k_1,M}(x^{(1)}) \cdots B_{k_d,M}(x^{(d)}) \quad (x^{(1)}, \dots, x^{(d)} \in \mathbb{R}).$$

With these functions we define $\mathcal{S}_{K,M}$ as the set of all linear combinations of all those of the above tensor product B-splines, where the support has nonempty intersection with $[0, 1]^d$, i.e., we set

$$\mathcal{S}_{K,M} = \left\{ \sum_{\mathbf{k} \in \{-M, -M+1, \dots, K-1\}^d} a_{\mathbf{k}} \cdot B_{\mathbf{k},M} : a_{\mathbf{k}} \in \mathbb{R} \right\}.$$

It can be shown by using standard arguments from spline theory, that the functions in $\mathcal{S}_{K,M}$ are in each component $(M-1)$ -times continuously differentiable, that they are equal to a (multivariate) polynomial of degree less than or equal to M (in each component) on each rectangular

$$[u_{k_1}, u_{k_1+1}) \times \cdots \times [u_{k_d}, u_{k_d+1}) \quad (\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{Z}^d), \quad (4)$$

and that they vanish outside of the set

$$\left[-\frac{M}{K}, 1 + \frac{M}{K} \right]^d.$$

Next we define spline approximations using so-called quasi interpolands: For a function $m : [0, 1]^d \rightarrow \mathbb{R}$ we define a approximating spline by

$$(Qm)(x) = \sum_{\mathbf{k} \in \{-M, -M+1, \dots, K-1\}^d} Q_{\mathbf{k}} m \cdot B_{\mathbf{k}, M}$$

where

$$Q_{\mathbf{k}} m = \sum_{\mathbf{j} \in \{0, 1, \dots, M\}^d} a_{\mathbf{k}, \mathbf{j}} \cdot m(t_{k_1, j_1}, \dots, t_{k_d, j_d})$$

for some $a_{\mathbf{k}, \mathbf{j}} \in \mathbb{R}$ and some suitable chosen points $t_{k,j} \in \text{supp}(B_{k,M}) \cap [0, 1]$. It can be shown that if we set

$$t_{k,j} = \frac{k \cdot M + j}{K \cdot M} \quad (j \in \{0, \dots, M\}, k \in \{0, \dots, K-1\})$$

and

$$t_{k,j} = \frac{j}{K \cdot M} \quad (j \in \{0, \dots, M\}, k \in \{-M, -M+1, \dots, -1\}),$$

then there exists coefficients $a_{\mathbf{k}, \mathbf{j}}$ (which can be computed by solving a linear equation system), such that

$$|Q_{\mathbf{k}} f| \leq c_1 \cdot \|f\|_{\infty, [u_{k_1}, u_{k_1+M+1}] \times \dots \times [u_{k_d}, u_{k_d+M+1}]} \quad (5)$$

for any $\mathbf{k} \in \{-M, -M+1, \dots, K-1\}^d$, any $f : [0, 1]^d \rightarrow \mathbb{R}$ and some universal constant c_1 , and such that Q reproduces polynomials of degree M or less (in each component) on $[0, 1]^d$, i.e., for any multivariate polynomial $p : \mathbb{R}^d \rightarrow \mathbb{R}$ of degree M or less in each component we have

$$(Qp)(x) = p(x) \quad (x \in [0, 1]^d) \quad (6)$$

(cf., e.g., Theorem 14.4 and Theorem 15.2 in Györfi et al. (2002)).

Next we define our estimate m_n as a quasi interpoland. We fix the degree $M \in \mathbb{N}$ and set

$$K = K_n = \left\lfloor \frac{\lfloor n^{1/d} \rfloor - 1}{M} \right\rfloor.$$

Furthermore we choose z_1, \dots, z_n such that all of the $(M \cdot K + 1)^d$ points of the form

$$\left(\frac{j_1}{M \cdot K}, \dots, \frac{j_d}{M \cdot K} \right) \quad (j_1, \dots, j_d \in \{0, 1, \dots, M \cdot K\})$$

are contained in $\{z_1, \dots, z_n\}$, which is possible since $(M \cdot K + 1)^d \leq n$. Then we define

$$m_n(x) = (Qm)(x),$$

where Qm is the above defined quasi interpoland satisfying (5) and (6). The computation of Qm requires only function values of m at the points z_1, \dots, z_n and hence m_n is well defined.

Theorem 1 *Let $p = k + \beta$ for some $k \in \mathbb{N}_0$ and $0 < \beta \leq 1$, and let $C > 0$. Assume that $m : \mathbb{R}^d \rightarrow \mathbb{R}$ is (p, C) -smooth. Let m_n be the above defined estimate with degree $M \geq k$. Then for $n \geq \max\{2^d \cdot (M + 1)^d, 4^d\}$*

$$\|m_n - m\|_{\infty, [0,1]^d} \leq c_2 \cdot n^{-p/d} \quad (7)$$

for some constant $c_2 > 0$ which depends only on M, C, p and d . In particular, if $\text{supp}(\mu) \subseteq [0, 1]^d$, then we also have

$$\int_{\mathbb{R}^d} |m_n(x) - m(x)| \mu(dx) \leq c_2 \cdot n^{-p/d}. \quad (8)$$

Next we show, that if we want to estimate a (p, C) -smooth function $m : \mathbb{R}^d \rightarrow \mathbb{R}$ on whole \mathbb{R}^d , it is in case of the $L_1(\mu)$ -norm possible to achieve the same rate of convergence provided μ satisfies some moment condition, namely

$$\int \|x\|^\gamma \mu(dx) < \infty \quad \text{for some } \gamma > 1 + 3 \cdot p \quad (9)$$

We define our estimate in case of an unbounded support of μ as follows: Above we have introduced a quasi interpoland on the cube $[0, 1]^d$, which uses $(M \cdot K + 1)^d$ observations of function values of m at equidistant points in $[0, 1]^d$ and B-splines of degree M with knots $u_k = k/K$ ($k \in \mathbb{Z}$). Analogously we define a quasi interpoland

$$Q_{N,M,[\mathbf{a}, \mathbf{a}+\delta]^d} m$$

of the function $m : \mathbb{R}^d \rightarrow \mathbb{R}$ of degree M on the cube

$$[\mathbf{a}, \mathbf{a} + \delta] = [a^{(1)}, a^{(1)} + \delta] \times \dots \times [a^{(d)}, a^{(d)} + \delta]$$

for $\mathbf{a} = (a^{(1)}, \dots, a^{(d)})$ and $\delta > 0$. This quasi interpoland uses observations of m at

$$\left(\left\lfloor \frac{\lfloor N^{1/d} \rfloor - 1}{M} \right\rfloor \cdot M + 1 \right)^d$$

equidistant points in $[\mathbf{a}, \mathbf{a} + \delta]$ and B-splines of degree M with knots

$$u_k^{(i)} = a^{(i)} + k \cdot \frac{\delta}{\lfloor \frac{\lfloor N^{1/d} \rfloor - 1}{M} \rfloor} \quad (k \in \mathbb{Z})$$

for the univariate B-splines for the i -th component of \mathbb{R}^d ($i \in \{1, \dots, d\}$). Set

$$j_{\max}(n) = \left\lfloor n^{p/(d \cdot \gamma)} \right\rfloor$$

and define the estimate by zero outside of

$$[-j_{\max}(n) - 1, j_{\max}(n) + 1]^d.$$

For $c_3 > 0$ defined below we set

$$N_j = \frac{c_3 \cdot n}{(j+1)^{\frac{d}{p} \cdot (\gamma-1-2p)}} \quad (j \in \mathbb{N}_0).$$

Then we define

$$m_n(x) = \begin{cases} (Q_{N_0, M, [-1, 1]^d} m)(x) & \text{if } x \in [-1, 1]^d, \\ (Q_{N_j, M, [\mathbf{a}, \mathbf{a}+1]^d} m)(x) & \text{if } x \in [\mathbf{a}, \mathbf{a}+1]^d \subseteq [-j-1, j+1]^d \setminus (-j, j)^d \\ & \text{for some } \mathbf{a} \in \mathbb{Z}^d, 1 \leq j \leq j_{\max}(n), \\ 0 & \text{if } x \notin [-j_{\max}(n)-1, j_{\max}(n)+1]^d. \end{cases}$$

For $j \in \mathbb{N}$ the set $[-j-1, j+1]^d \setminus [-j, j]^d$ contains

$$(2j+2)^d - (2j)^d \leq c_4 \cdot j^{d-1}$$

sets of the form $[\mathbf{a}, \mathbf{a}+1]^d$, $\mathbf{a} \in \mathbb{Z}^d$. Since

$$\begin{aligned} N_0 + \sum_{j=1}^{j_{\max}(n)} c_4 \cdot j^{d-1} N_j &\leq n \cdot c_3 \cdot \left(1 + \sum_{j=1}^{\infty} \frac{c_4}{(j+1)^{\frac{d}{p} \cdot (\gamma-1-2p) - d + 1}} \right) \\ &= n \cdot c_3 \cdot \left(1 + \sum_{j=1}^{\infty} \frac{c_4}{(j+1)^{1 + \frac{d}{p} \cdot (\gamma-1-3p)}} \right) \end{aligned}$$

this estimate needs at most n observations of the function values of m provided we choose $c_3 > 0$ such that

$$c_3 \leq \frac{1}{1 + \sum_{j=1}^{\infty} \frac{c_4}{(j+1)^{1 + \frac{d}{p} \cdot (\gamma-1-3p)}}}.$$

For this estimate we can show the following result:

Theorem 2 *Let $p = k + \beta$ for some $k \in \mathbb{N}_0$ and $0 < \beta \leq 1$, and let $C > 0$. Assume that $m : \mathbb{R}^d \rightarrow \mathbb{R}$ is (p, C) -smooth and let μ be a probability measure satisfying (9). Let m_n be the above defined estimate with degree $M \geq k$. Then for sufficiently large n*

$$\int_{\mathbb{R}^d} |m_n(x) - m(x)| \mu(dx) \leq c_5 \cdot n^{-p/d} \quad (10)$$

for some constant $c_5 > 0$ which depends only on M, C, p and d .

Our next theorem shows that the above rate of convergence results are optimal in the minimax sense.

Theorem 3 *Let $p = k + \beta$ for some $k \in \mathbb{N}_0$ and $0 < \beta \leq 1$, let $C > 0$ and let $\mathcal{F}^{(p, C)}$ be the set of all (p, C) -smooth functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$. Then we have for any $n \in \mathbb{N}$*

$$\inf_{m_n} \sup_{m \in \mathcal{F}^{(p, C)}} \int_{[0, 1]^d} |m_n(x) - m(x)| dx \geq c_6 \cdot n^{-p/d}$$

for some constant $c_6 > 0$ which depends only on p , C and d . In other words: for any estimate m_n applied to a data set described by (1) and (2) and any data size n we can find a (p, C) -smooth function $m : [0, 1]^d \rightarrow \mathbb{R}$ such that

$$\int_{[0,1]^d} |m_n(x) - m(x)| dx \geq c_6 \cdot n^{-p/d}.$$

In the above result the worst function, which leads to a large error of an estimate m_n , changes with the number n of the observed function values. But in an application often it is possible to observe more and more function values of some fixed function. As our next result shows, the rate of convergence in Theorem 3 cannot be improved even if the function to be approximated is not allowed to change with n .

Theorem 4 *Let $p = k + \beta$ for some $k \in \mathbb{N}_0$ and $0 < \beta \leq 1$, and let $C > 0$. Let $(m_n)_{n \in \mathbb{N}}$ be an arbitrary sequence of estimates and let a_n be an arbitrary sequence of positive numbers converging monotonically towards zero. Then there exists a (p, C) -smooth function $m : [0, 1]^d \rightarrow \mathbb{R}$ such that*

$$\limsup_{n \rightarrow \infty} \frac{\int_{[0,1]^d} |m_n(x) - m(x)| dx}{a_n \cdot n^{-p/d}} = \infty.$$

3 Proofs

3.1 Proof of Theorem 1

In the proof of Theorem 1 we need the following lemma.

Lemma 1 *Let $p = k + \beta$ for some $k \in \mathbb{N}_0$ and $0 < \beta \leq 1$, and let $C > 0$. Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a (p, C) -smooth function, let $x_0 \in \mathbb{R}^d$ and let p_k be the Taylor polynomial of f of total degree k around x_0 , i.e.,*

$$p_k(x) = \sum_{\substack{j_1, \dots, j_d \in \{0, 1, \dots, k\}, \\ j_1 + \dots + j_d \leq k}} \frac{1}{j_1! \cdots j_d!} \cdot \frac{\partial^{j_1 + \dots + j_d} f}{\partial^{j_1} x^{(1)} \cdots \partial^{j_d} x^{(d)}}(x_0) \cdot (x^{(1)} - x_0^{(1)})^{j_1} \cdots (x^{(d)} - x_0^{(d)})^{j_d}.$$

Then we have for any $x \in \mathbb{R}^d$

$$|f(x) - p_k(x)| \leq c_7 \cdot C \cdot \|x - x_0\|^p$$

for some constant $c_7 \in \mathbb{R}$ depending only on k and on d .

Proof. The proof is a straightforward extension of the proof of Lemma 11.1 in Görfi et al. (2002). For the sake of completeness we present nevertheless a complete proof.

The definition of p_k and the integral form of the remainder of a Taylor series imply

$$f(x) - p_k(x)$$

$$\begin{aligned}
&= f(x) - p_{k-1}(x) - \sum_{\substack{j_1, \dots, j_d \in \{0, 1, \dots, k\}, \\ j_1 + \dots + j_d = k}} \frac{1}{j_1! \cdots j_d!} \cdot \frac{\partial^{j_1 + \dots + j_d} f}{\partial^{j_1} x^{(1)} \dots \partial^{j_d} x^{(d)}}(x_0) \\
&\quad \cdot (x^{(1)} - x_0^{(1)})^{j_1} \dots (x^{(d)} - x_0^{(d)})^{j_d} \\
&= \sum_{\substack{j_1, \dots, j_d \in \{0, 1, \dots, k\}, \\ j_1 + \dots + j_d = k}} \frac{k}{j_1! \cdots j_d!} \cdot \int_0^1 (1-t)^{k-1} \cdot \frac{\partial^{j_1 + \dots + j_d} f}{\partial^{j_1} x^{(1)} \dots \partial^{j_d} x^{(d)}}(x_0 + t \cdot (x - x_0)) dt \\
&\quad \cdot (x^{(1)} - x_0^{(1)})^{j_1} \dots (x^{(d)} - x_0^{(d)})^{j_d} \\
&- \sum_{\substack{j_1, \dots, j_d \in \{0, 1, \dots, k\}, \\ j_1 + \dots + j_d = k}} \frac{k}{j_1! \cdots j_d!} \cdot \int_0^1 (1-t)^{k-1} \frac{\partial^{j_1 + \dots + j_d} f}{\partial^{j_1} x^{(1)} \dots \partial^{j_d} x^{(d)}}(x_0) dt \\
&\quad \cdot (x^{(1)} - x_0^{(1)})^{j_1} \dots (x^{(d)} - x_0^{(d)})^{j_d} \\
&= \sum_{\substack{j_1, \dots, j_d \in \{0, 1, \dots, k\}, \\ j_1 + \dots + j_d = k}} \frac{k}{j_1! \cdots j_d!} \cdot (x^{(1)} - x_0^{(1)})^{j_1} \dots (x^{(d)} - x_0^{(d)})^{j_d} \\
&\quad \cdot \int_0^1 (1-t)^{k-1} \left(\frac{\partial^{j_1 + \dots + j_d} f}{\partial^{j_1} x^{(1)} \dots \partial^{j_d} x^{(d)}}(x_0 + t \cdot (x - x_0)) - \frac{\partial^{j_1 + \dots + j_d} f}{\partial^{j_1} x^{(1)} \dots \partial^{j_d} x^{(d)}}(x_0) \right) dt.
\end{aligned}$$

Using the triangle inequality and the (p, C) -smoothness of f we conclude

$$\begin{aligned}
&|f(x) - p_k(x)| \\
&\leq \sum_{\substack{j_1, \dots, j_d \in \{0, 1, \dots, k\}, \\ j_1 + \dots + j_d = k}} \frac{k}{j_1! \cdots j_d!} \cdot \|x - x_0\|^k \cdot \int_0^1 (1-t)^{k-1} \cdot C \cdot t^\beta \cdot \|x - x_0\|^\beta dt \\
&\leq c_7 \cdot C \cdot \|x - x_0\|^{k+\beta},
\end{aligned}$$

which completes the proof. \square

Proof of Theorem 1. The proof is an easy consequence of the properties (5) and (6) of Qm and of Lemma 1. Let $x \in [0, 1]^d$ be arbitrary. Let $\mathbf{k} \in \{0, 1, \dots, K\}^d$ be the index of the rectangular

$$[u_{k^{(1)}}, u_{k^{(1)}+1}] \times \cdots \times [u_{k^{(d)}}, u_{k^{(d)}+1}]$$

which contains x (in case that there exists several of these rectangles containing x , choose one of them). Let p_x be the Taylor polynomial of m of total degree k around x . Then $p_x(x) = m(x)$. Using this and (5) and (6) and the facts that only those B-splines with index \mathbf{j} satisfying $k^{(l)} - M \leq j^{(l)} \leq k^{(l)}$ for all $l \in \{1, \dots, d\}$ do not vanish at x and that the B-splines sum up to one (cf., e.g., Lemma 15.4 in Györfi et al. (2002)) we get

$$\begin{aligned}
&|m_n(x) - m(x)| \\
&= |(Qm)(x) - p_x(x)| \\
&= |(Qm)(x) - (Qp_x)(x)| \\
&= \left| \sum_{\mathbf{j} \in \{-M, -M+1, \dots, K-1\}^d} Q_{\mathbf{j}}(m - p_x) \cdot B_{\mathbf{j}, M}(x) \right|
\end{aligned}$$

$$\begin{aligned}
&\leq \sum_{\substack{\mathbf{j} \in \{-M, -M+1, \dots, K-1\}^d: \\ k^{(l)} - M \leq j^{(l)} \leq k^{(l)} \ (l \in \{1, \dots, d\})}} |Q_{\mathbf{j}}(m - p_x)| \\
&\leq (M+1)^d \cdot \|m - p_x\|_{\infty, [u_{k^{(1)}} - M, u_{k^{(1)}} + M + 1] \times \dots \times [u_{k^{(d)}} - M, u_{k^{(d)}} + M + 1]}.
\end{aligned}$$

By Lemma 1 we know that

$$\begin{aligned}
&\|m - p_x\|_{\infty, [u_{k^{(1)}} - M, u_{k^{(1)}} + M + 1] \times \dots \times [u_{k^{(d)}} - M, u_{k^{(d)}} + M + 1]} \\
&\leq c_7 \cdot C \cdot \left(\sqrt{d} \cdot \frac{M+1}{K} \right)^p \\
&= c_7 \cdot C \cdot (\sqrt{d} \cdot (M+1))^p \cdot \frac{1}{\left(\left\lfloor \frac{\lfloor n^{1/d} \rfloor - 1}{M} \right\rfloor \right)^p},
\end{aligned}$$

which implies (7). Inequality (8) is a immediate consequence of (7). \square

3.2 Proof of Theorem 2

The proof of Theorem 1 implies that in case of $N \geq 4^d \cdot (M+1)^d$ we have

$$\|Q_{N, M, [\mathbf{a}, \mathbf{a} + \delta]^d} m - m\|_{\infty, [\mathbf{a}, \mathbf{a} + \delta]^d} \leq c_8 \cdot \left(\frac{\delta}{\left\lfloor \frac{\lfloor N^{1/d} \rfloor - 1}{M} \right\rfloor} \right)^p,$$

hence $c_3 \cdot n \geq 4^d \cdot (M+1)^d$ implies

$$\|Q_{N_0, M, [-1, 1]^d} m - m\|_{\infty, [-1, 1]^d} \leq c_9 \cdot n^{-p/d}$$

and for any $j \in \mathbb{N}$ and $\mathbf{a} \in \mathbb{Z}^d$ we have

$$\|Q_{N_j, M, [\mathbf{a}, \mathbf{a} + 1]^d} m - m\|_{\infty, [\mathbf{a}, \mathbf{a} + 1]^d} \leq c_9 \cdot (j+1)^{\gamma-1-2p} \cdot n^{-p/d}$$

provided $c_3 \cdot n \geq 4^d \cdot (M+1)^d \cdot (j+1)^{(d/p) \cdot (\gamma-1-2p)}$. Since

$$\frac{(j_{\max}(n) + 1)^{(d/p) \cdot (\gamma-1-2p)}}{n} \rightarrow 0 \quad (n \rightarrow \infty),$$

we can assume w.l.o.g. that n satisfies the above two inequalities. We conclude

$$\begin{aligned}
&\int |m_n(x) - m(x)| \mu(dx) \\
&= \int_{[-1, 1]^d} |m_n(x) - m(x)| \mu(dx) + \sum_{j=1}^{j_{\max}(n)} \int_{[-j-1, j+1]^d \setminus [-j, j]^d} |m_n(x) - m(x)| \mu(dx) \\
&\quad + \int_{\mathbb{R}^d \setminus [-j_{\max}(n)-1, j_{\max}(n)+1]^d} |m_n(x) - m(x)| \mu(dx)
\end{aligned}$$

$$\begin{aligned}
&\leq c_9 \cdot n^{-p/d} + \sum_{j=1}^{j_{\max}(n)} c_9 \cdot (j+1)^{\gamma-1-2p} \cdot n^{-p/d} \cdot j^{-\gamma} \cdot \int_{[-j-1, j+1]^d \setminus [-j, j]^d} \|x\|^\gamma \mu(dx) \\
&\quad + \|m\|_\infty \cdot (j_{\max}(n) + 1)^{-\gamma} \cdot \int_{\mathbb{R}^d \setminus [-j_{\max}(n)-1, j_{\max}(n)+1]^d} \|x\|^\gamma \mu(dx) \\
&\leq c_{10} \cdot n^{-p/d} \cdot \sum_{j=0}^{\infty} (j+1)^{-1-2p} \cdot \left(1 + \int_{\mathbb{R}^d} \|x\|^\gamma \mu(dx)\right) \\
&\leq c_{11} \cdot n^{-p/d},
\end{aligned}$$

since by (9) we have

$$\int_{\mathbb{R}^d} \|x\|^\gamma \mu(dx) < \infty$$

and since $-1 - 2p < -1$ implies

$$\sum_{j=0}^{\infty} (j+1)^{-1-2p} < \infty.$$

□

3.3 Proof of Theorem 3

The proof is a modification of the proof of Theorem 3 in Kohler and Krzyżak (2013), which in turn is based on the proof of a lower bound presented in Stone (1982) (see also proof of Theorem 3.2 in Györfi et al. (2002)).

Set $M_n = \lceil (2 \cdot n)^{1/d} \rceil$ and let $\{A_{n,j}\}_{j=1, \dots, M_n^d}$ be a partition of $[0, 1]^d$ into cubes of side length $1/M_n$. Choose a $(p, 2^{\beta-1}C)$ -smooth function $g : \mathbb{R}^d \rightarrow [-1, 1]$ satisfying

$$\text{supp}(g) \subseteq \left[-\frac{1}{2}, \frac{1}{2}\right]^d \quad \text{and} \quad \int |g(x)| dx > 0.$$

For $j \in \{1, \dots, M_n^d\}$ let $a_{n,j}$ be the center of $A_{n,j}$ and set

$$g_{n,j}(x) = M_n^{-p} \cdot g(M_n \cdot (x - a_{n,j})).$$

We index the class of functions considered by $c_n = (c_{n,1}, \dots, c_{n,M_n^d}) \in \{-1, 1\}^{M_n^d}$ and define $m^{(c_n)} : \mathbb{R}^d \rightarrow [-1, 1]$ by

$$m^{(c_n)}(x) = \sum_{j=1}^{M_n^d} c_{n,j} \cdot g_{n,j}(x).$$

Let m_n be an arbitrary estimate of m . As in the proof of Theorem 3.2 in Györfi et al. (2002) we can see that $m^{(c_n)}$ is (p, C) -smooth, which implies

$$\sup_{m \in \mathcal{F}^{(p,C)}} \int |m_n(x) - m(x)| \mathbf{P}_X(dx)$$

$$\geq \sup_{c_n \in \{-1,1\}^{M_n^d}} \int_{[0,1]^d} |m_n(x) - m^{(c_n)}(x)| dx.$$

In order to bound the right-hand side of the inequality above we randomize c_n . Choose independent random variables $C_1, \dots, C_{M_n^d}$ satisfying

$$\mathbf{P}\{C_k = -1\} = \mathbf{P}\{C_k = 1\} = \frac{1}{2} \quad (k = 1, \dots, M_n^d),$$

which are, in case that z_1, \dots, z_n random, also independent from z_1, \dots, z_n , and set

$$C_n = (C_1, \dots, C_{M_n^d}).$$

Then

$$\begin{aligned} & \sup_{c_n \in \{-1,1\}^{M_n^d}} \int_{[0,1]^d} |m_n(x) - m^{(c_n)}(x)| dx \\ & \geq \mathbf{E} \int_{[0,1]^d} |m_n(x) - m^{(C_n)}(x)| dx \\ & = \sum_{j=1}^{M_n^d} \mathbf{E} \int_{A_{n,j}} |m_n(x) - C_{n,j} \cdot g_{n,j}(x)| dx \\ & \geq \sum_{j=1}^{M_n^d} \mathbf{E} \left\{ \int_{A_{n,j}} |m_n(x) - C_{n,j} \cdot g_{n,j}(x)| dx \cdot I_{\{z_1, \dots, z_n \notin A_{n,j}\}} \right\} \\ & = \sum_{j=1}^{M_n^d} \mathbf{E} \left\{ \mathbf{E} \left\{ \int_{A_{n,j}} |m_n(x) - C_{n,j} \cdot g_{n,j}(x)| dx \mid \mathcal{F}_{n,j} \right\} \cdot I_{\{z_1, \dots, z_n \notin A_{n,j}\}} \right\}, \end{aligned}$$

where $\mathcal{F}_{n,j}$ is the σ -field generated by $z_1, \dots, z_n, C_1, \dots, C_{j-1}, C_{j+1}, \dots, C_{M_n^d}$. (Here z_1, \dots, z_n are included in the σ -field $\mathcal{F}_{n,j}$ because it is in principle allowed that they are randomly chosen.) If z_1, \dots, z_n are not contained in $A_{n,j}$, then $m^{(C_n)}(z_1), \dots, m^{(C_n)}(z_n)$ and hence also $m_n(x)$ are independent of C_j , which implies

$$\begin{aligned} & \mathbf{E} \left\{ \int_{A_{n,j}} |m_n(x) - C_{n,j} \cdot g_{n,j}(x)| dx \mid \mathcal{F}_{n,j} \right\} \\ & = \frac{1}{2} \cdot \int_{A_{n,j}} |m_n(x) - g_{n,j}(x)| dx + \frac{1}{2} \cdot \int_{A_{n,j}} |m_n(x) + g_{n,j}(x)| dx \\ & = \frac{1}{2} \cdot \int_{A_{n,j}} (|g_{n,j}(x) - m_n(x)| + |g_{n,j}(x) + m_n(x)|) dx \\ & \geq \frac{1}{2} \cdot \int_{A_{n,j}} |(g_{n,j}(x) - m_n(x)) + (g_{n,j}(x) + m_n(x))| dx \\ & = \int_{A_{n,j}} |g_{n,j}(x)| dx, \end{aligned}$$

where the latter inequality follows from triangle inequality. From this we conclude

$$\begin{aligned}
& \sum_{j=1}^{M_n^d} \mathbf{E} \left\{ \mathbf{E} \left\{ \int_{A_{n,j}} |m_n(x) - C_{n,j} \cdot g_{n,j}(x)| dx \middle| \mathcal{F}_{n,j} \right\} \cdot I_{\{z_1, \dots, z_n \notin A_{n,j}\}} \right\} \\
& \geq \sum_{j=1}^{M_n^d} \mathbf{E} \left\{ \int_{A_{n,j}} |g_{n,j}(x)| dx \cdot I_{\{z_1, \dots, z_n \notin A_{n,j}\}} \right\} \\
& = M_n^{-p-d} \cdot \int |g(x)| dx \cdot \mathbf{E} \left\{ \sum_{j=1}^{M_n^d} I_{\{z_1, \dots, z_n \notin A_{n,j}\}} \right\} \\
& \geq M_n^{-p-d} \cdot \int |g(x)| dx \cdot (M_n^d - n),
\end{aligned}$$

where we have used that there are at most n sets of the M_n^d disjoint sets $A_{n,1}, \dots, A_{n,M_n^d}$, which contain at least one of the n points z_1, \dots, z_n . Using the definition of M_n we see that

$$\begin{aligned}
M_n^{-p-d} \cdot (M_n^d - n) & \geq M_n^{-p-d} \cdot n \geq \frac{1}{((2 \cdot n)^{1/p} + 1)^{p+d}} \cdot n \geq \frac{1}{2^{p+d} \cdot (2n)^{(p+d)/d}} \cdot n \\
& \geq \frac{1}{2^{p+d+1} \cdot 2^{p/d}} \cdot n^{-p/d},
\end{aligned}$$

which completes the proof. \square

3.4 Proof of Theorem 4

The proof is a modification of the proof of Theorem 3.3 in Györfi et al. (2002), which in turn is based on Antos, Györfi and Kohler (2000).

Let $\{I_j\}_j$ be a partition of $[0, 1]$ into intervals I_j of length 2^{-j} . The idea of the proof is to use simultaneously on each of the sets I_j^d a partition corresponding to the one in the proof of Theorem 3 for a suitable sample size n_j determined below.

Since a_n tends monotonically to zero, we can find a subsequence $(n_j)_j$ of $(n)_n$ such that

$$a_{n_j} \leq \frac{1}{2} \cdot 4^{-p-d} \cdot 2^{-j \cdot (p+d)}. \quad (11)$$

Set

$$M_j = \left\lceil (2 \cdot n_j)^{1/d} \right\rceil$$

and partition I_j^d into $S_j = M_j^d$ cubes $A_{j,1}, \dots, A_{j,S_j}$ of side length

$$\frac{2^{-j}}{M_j} \geq \frac{2^{-j}}{(2 \cdot n_j)^{1/d} + 1} \geq \frac{2^{-j}}{2 \cdot (2 \cdot n_j)^{1/d}} \geq \frac{1}{4} \cdot 2^{-j} \cdot n_j^{-1/d} =: p_j.$$

Choose a $(p, 2^{\beta-1}C)$ -smooth function $g : \mathbb{R}^d \rightarrow [-1, 1]$ satisfying

$$\text{supp}(g) \subseteq \left[-\frac{1}{2}, \frac{1}{2}\right]^d \quad \text{and} \quad \int |g(x)| dx > 0.$$

For $j \in \mathbb{N}$ and $k \in \{1, 2, \dots, S_j\}$ let $a_{j,k}$ be the center of $A_{j,k}$ and set

$$g_{j,k}(x) = p_j^p \cdot g\left(p_j^{-1} \cdot (x - a_{j,k})\right).$$

We index the class of functions considered by a vector

$$c = (c_{1,1}, c_{1,2}, \dots, c_{1,S_1}, c_{2,1}, c_{2,2}, \dots, c_{2,S_2}, \dots)$$

of $+1$ and -1 components and define $m^{(c)} : \mathbb{R}^d \rightarrow [-1, 1]$ by

$$m^{(c)}(x) = \sum_{j=1}^{\infty} \sum_{k=1}^{S_{j,k}} c_{j,k} \cdot g_{j,k}(x).$$

Let m_n be an arbitrary estimate of m . As in the proof of Theorem 3.2 in Györfi et al. (2002) we can see that $m^{(c)}$ is (p, C) -smooth, which implies

$$\begin{aligned} & \sup_{m \in \mathcal{F}(p, C)} \limsup_{n \rightarrow \infty} \frac{\int_{[0,1]^d} |m_n(x) - m(x)| dx}{a_n \cdot n^{-p/d}} \\ & \geq \sup_{c \in \{-1, 1\}^\infty} \limsup_{n \rightarrow \infty} \frac{\int_{[0,1]^d} |m_n(x) - m^{(c)}(x)| dx}{a_n \cdot n^{-p/d}}. \end{aligned}$$

Let $c \in \{-1, 1\}^\infty$ be arbitrary and set

$$\bar{c}_n = (\bar{c}_{n,1,1}, \dots, \bar{c}_{n,1,S_1}, \bar{c}_{n,2,1}, \dots, \bar{c}_{n,2,S_2}, \dots),$$

where

$$\bar{c}_{n,j,k} = \begin{cases} 1 & \text{if } \int_{A_{j,k}} |m_n(x) - 1 \cdot g_{j,k}(x)| dx \leq \int_{A_{j,k}} |m_n(x) - (-1) \cdot g_{j,k}(x)| dx \\ -1 & \text{else.} \end{cases}$$

Using the definition of $\bar{c}_{n,j,k}$ and the triangle inequality we see that

$$\begin{aligned} & \int_{A_{j,k}} |m_n(x) - c_{j,k} \cdot g_{j,k}(x)| dx \\ & \geq \frac{1}{2} \cdot \int_{A_{j,k}} |m_n(x) - c_{j,k} \cdot g_{j,k}(x)| dx + \frac{1}{2} \cdot \int_{A_{j,k}} |m_n(x) - \bar{c}_{n,j,k} \cdot g_{j,k}(x)| dx \\ & \geq \frac{1}{2} \cdot \int_{A_{j,k}} |(m_n(x) - c_{j,k} \cdot g_{j,k}(x)) + (\bar{c}_{n,j,k} \cdot g_{j,k}(x) - m_n(x))| dx \\ & = \frac{1}{2} \cdot |\bar{c}_{n,j,k} - c_{j,k}| \cdot p_j^{p+d} \cdot \int |g(x)| dx. \end{aligned}$$

From this we conclude

$$\int_{[0,1]^d} |m_n(x) - m^{(c)}(x)| dx$$

$$\begin{aligned}
&\geq \sum_{j=1}^{\infty} \sum_{k=1}^{S_j} \int_{A_{j,k}} |m_n(x) - c_{j,k} \cdot g_{j,k}(x)| dx \\
&\geq \int |g(x)| dx \cdot \sum_{j=1}^{\infty} \sum_{k=1}^{S_j} p_j^{p+d} \cdot I_{\{\bar{c}_{n,j,k} \neq c_{j,k}\}} \\
&\geq \int |g(x)| dx \cdot \sum_{j \in \mathbb{N}: S_j \geq 2n} \sum_{k=1}^{S_j} p_j^{p+d} \cdot I_{\{\bar{c}_{n,j,k} \neq c_{j,k}\}} \\
&=: \int |g(x)| dx \cdot R_n(c).
\end{aligned}$$

Hence

$$\sup_{m \in \mathcal{F}(p,C)} \limsup_{n \rightarrow \infty} \frac{\int_{[0,1]^d} |m_n(x) - m(x)| dx}{a_n \cdot n^{-p/d}} \geq \int |g(x)| dx \cdot \sup_{c \in \{-1,1\}^\infty} \limsup_{n \rightarrow \infty} \frac{R_n(c)}{a_n \cdot n^{-p/d}}.$$

Next we randomize c . Choose independent random variables $C_{1,1}, C_{1,2}, \dots, C_{1,S_1}, C_{2,1}, C_{2,2}, \dots, C_{2,S_2}, \dots$ satisfying

$$\mathbf{P}\{C_{j,k} = -1\} = \mathbf{P}\{C_{j,k} = 1\} = \frac{1}{2} \quad (j \in \mathbb{N}, k = 1, \dots, S_j),$$

which are, in case that z_1, \dots, z_n random, also independent from z_1, \dots, z_n , and set

$$C = (C_{1,1}, C_{1,2}, \dots, C_{1,S_1}, C_{2,1}, C_{2,2}, \dots, C_{2,S_2}, \dots).$$

In the sequel we derive a lower bound for $\mathbf{E}R_n(C)$. Let $\mathcal{F}_{j,k}$ be the σ -field generated by z_1, \dots, z_n , and $C_{i,l}$ ($i \in \mathbb{N}, l \in \{1, \dots, S_i\}, (i,l) \neq (j,k)$). If z_1, \dots, z_n are not contained in $A_{j,k}$, then $m^{(C_n)}(z_1), \dots, m^{(C_n)}(z_n)$ and hence also $m_n(x)$ and $\bar{c}_{n,j,k}$ are independent of $C_{j,k}$, which implies

$$\begin{aligned}
\mathbf{E}R_n(C) &= \mathbf{E} \left\{ \sum_{j \in \mathbb{N}: S_j \geq 2n} \sum_{k=1}^{S_j} p_j^{p+d} \cdot \mathbf{P}\{\bar{c}_{n,j,k} \neq C_{j,k} | \mathcal{F}_{j,k}\} \right\} \\
&\geq \mathbf{E} \left\{ \sum_{j \in \mathbb{N}: S_j \geq 2n} \sum_{k=1}^{S_j} p_j^{p+d} \cdot \mathbf{P}\{\bar{c}_{n,j,k} \neq C_{j,k} | \mathcal{F}_{j,k}\} \cdot I_{\{z_1, \dots, z_n \notin A_{j,k}\}} \right\} \\
&= \frac{1}{2} \cdot \mathbf{E} \left\{ \sum_{j \in \mathbb{N}: S_j \geq 2n} \sum_{k=1}^{S_j} p_j^{p+d} \cdot I_{\{z_1, \dots, z_n \notin A_{j,k}\}} \right\} \\
&\geq \frac{1}{2} \cdot \sum_{j \in \mathbb{N}: S_j \geq 2n} p_j^{p+d} \cdot (S_j - n),
\end{aligned}$$

where we have used that at most n of the S_j disjoint sets $A_{j,1}, \dots, A_{j,S_j}$ contain at least one of the n points z_1, \dots, z_n . Hence

$$\mathbf{E}R_n(C) \geq \frac{1}{4} \cdot \sum_{j \in \mathbb{N}: S_j \geq 2n} p_j^{p+d} \cdot S_j.$$

For $n = n_t$ we have

$$S_t = M_t^d \geq 2 \cdot n_t,$$

which together with inequality (11) implies

$$\mathbf{E}R_{n_t}(C) \geq \frac{1}{4} \cdot p_t^{p+d} \cdot S_t \geq \frac{1}{4} \cdot \left(\frac{1}{4} \cdot 2^{-t} \cdot n_t^{-1/d} \right)^{p+d} \cdot 2 \cdot n_t \geq a_{n_t} \cdot n_t^{-p/d}.$$

We conclude

$$\begin{aligned} \sup_{c \in \{-1,1\}^\infty} \limsup_{n \rightarrow \infty} \frac{R_n(c)}{a_n \cdot n^{-p/d}} &\geq \sup_{c \in \{-1,1\}^\infty} \limsup_{t \rightarrow \infty} \frac{R_{n_t}(c)}{\mathbf{E}R_{n_t}(C)} \\ &\geq \mathbf{E} \left\{ \limsup_{t \rightarrow \infty} \frac{R_{n_t}(C)}{\mathbf{E}R_{n_t}(C)} \right\}. \end{aligned}$$

Because of

$$\frac{R_{n_t}(c)}{\mathbf{E}R_{n_t}(C)} \leq \frac{\sum_{j \in \mathbb{N}: S_j \geq 2n_t} p_j^{p+d} \cdot S_j}{\frac{1}{4} \cdot \sum_{j \in \mathbb{N}: S_j \geq 2n_t} p_j^{p+d} \cdot S_j} \leq 4,$$

we can apply the Lemma of Fatou, which yields

$$\mathbf{E} \left\{ \limsup_{t \rightarrow \infty} \frac{R_{n_t}(C)}{\mathbf{E}R_{n_t}(C)} \right\} \geq \limsup_{t \rightarrow \infty} \mathbf{E} \left\{ \frac{R_{n_t}(C)}{\mathbf{E}R_{n_t}(C)} \right\} = 1.$$

Summarizing the above results we get

$$\limsup_{n \rightarrow \infty} \frac{\int_{[0,1]^d} |m_n(x) - m(x)| dx}{a_n \cdot n^{-p/d}} \geq \int |g(x)| dx.$$

By replacing a_n by \bar{a}_n satisfying

$$\frac{a_n}{\bar{a}_n} \rightarrow 0 \quad (n \rightarrow \infty)$$

we get the assertion. □

4 Acknowledgment

The author would like to thank the German Research Foundation (DFG) for funding this project within the Collaborative Research Center 805.

References

- [1] Antos, A., Györfi, L. and Kohler, M. (2000). Lower bounds on the rate of convergence of nonparametric regression estimates. *Journal of Statistical Planning and Inference*, **83**, pp. 91-100.

- [2] de Boor, C. (1978). *A practical guide to splines*. Springer.
- [3] Beirlant, J. and Györfi, L. (1998). On the asymptotic L_2 -error in partitioning regression estimation. *Journal of Statistical Planning and Inference*, **71**, pp. 93–107.
- [4] Cheney, W. and Kincaid, D. (2008). *Numerical Mathematics and Computing* Thomson Brooks/Cole, Belmont, CA.
- [5] Devroye, L. (1982). Necessary and sufficient conditions for the almost everywhere convergence of nearest neighbor regression function estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **61**, pp. 467–481.
- [6] Devroye, L., Felber, T., and Kohler, M. (2013). Estimation of a density using real and artificial data. *IEEE Transactions on Information Theory*, **59**, pp. 1917–1928.
- [7] Devroye, L., Györfi, L., Krzyżak, A. and Lugosi, G. (1994). On the strong universal consistency of nearest neighbor regression function estimates. *Annals of Statistics*, **22**, pp. 1371–1385.
- [8] Devroye, L. and Krzyżak, A. (1989). An equivalence theorem for L_1 convergence of the kernel regression estimate. *Journal of Statistical Planning and Inference*, **23**, pp. 71–82.
- [9] Devroye, L. and Wagner, T. J. (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Annals of Statistics*, **8**, pp. 231–239.
- [10] Felber, T., Kohler, M. and Krzyżak, A. (2013). Adaptive density estimation based on real and artificial data. Submitted for publication.
- [11] Györfi, L. (1981). Recent results on nonparametric regression estimate and multiple classification. *Problems of Control and Information Theory*, **10**, pp. 43–52.
- [12] Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer-Verlag, New York.
- [13] Kohler, M. (2000). Inequalities for uniform deviations of averages from expectations with applications to nonparametric regression. *Journal of Statistical Planning and Inference*, **89**, pp. 1–23.
- [14] Kohler, M. and Krzyżak, A. (2001). Nonparametric regression estimation using penalized least squares. *IEEE Transactions on Information Theory*, **47**, pp. 3054–3058.
- [15] Kohler, M. and Krzyżak, A. (2013). Optimal global rates of convergence for interpolation problems with random design. *Statistics and Probability Letters*, **83**, pp. 1871–1879.
- [16] Kohler, M., Krzyżak, A. and Walk, H. (2006). Rates of convergence for partitioning and nearest neighbor regression estimates with unbounded data. *Journal of Multivariate Analysis*, **97**, pp. 311–323.

- [17] Kohler, M., Krzyżak, A. and Walk, H. (2009). Optimal global rates of convergence in nonparametric regression with unbounded data. *Journal of Statistical Planning and Inference*, **139**, pp. 1286-1296.
- [18] Lugosi, G. and Zeger, K. (1995). Nonparametric estimation via empirical risk minimization. *IEEE Transactions on Information Theory*, **41**, pp. 677-687.
- [19] Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, **9**, pp. 141-142.
- [20] Nadaraya, E. A. (1970). Remarks on nonparametric estimates for density functions and regression curves. *Theory of Probability and its Applications*, **15**, pp. 134-137.
- [21] Stone, C. J. (1977). Consistent nonparametric regression. *Annals of Statistics*, **5**, pp. 595-645.
- [22] Stone, C. J. (1980). Optimal rates of convergence for nonparametric regression. *Annals of Statistics*, **8**, pp. 1348-1360.
- [23] Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Annals of Statistics*, **10**, pp. 1040-1053.
- [24] C. J. Stone (1985). Additive regression and other nonparametric models. *Annals of Statistics*, **13**, pp. 689-705.
- [25] C. J. Stone (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Annals of Statistics*, **22**, pp. 118-184.
- [26] Wahba, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.
- [27] Watson, G. S. (1964). Smooth regression analysis. *Sankhya Series A*, **26**, pp. 359-372.
- [28] Wendland, H. (2010). *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics Cambridge University Press, Cambridge.