SFB 823

Confidence bands for multivariate and time dependent inverse regression models

Katharina Proksch, Nicolai Bissantz, Holger Dette

Nr. 23/2012

Discussion Paper

Confidence bands for multivariate and time dependent inverse regression models

Katharina Proksch, Nicolai Bissantz, Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: holger.dette@rub.de

June 12, 2012

Abstract

Uniform asymptotic confidence bands for a multivariate regression function in an inverse regression model with a convolution-type operator are constructed. The results are derived using strong approximation methods and a limit theorem for the supremum of a stationary Gaussian field over an increasing system of sets. As a particular application asymptotic confidence bands for a time dependent regression function $f_t(x)$ ($x \in \mathbb{R}^d$, $t \in \mathbb{R}$) in a convolution-type inverse regression model are obtained. To the best knowledge of the authors the results presented in this paper are the first which provide uniform confidence bands for multivariate nonparametric function estimation in inverse problems.

Keywords and Phrases: Confidence bands, Inverse problems, Deconvolution, Rates of convergence, Multivariate regression, Nonparametric Regression, Uniform convergence, Time dependent regression function.

AMS Subject Classification: 62E20, 62G08, 42Q05

1 Introduction

1.1 Inverse regression models

In many applications it is impossible to observe a certain quantity of interest because only indirect observations are available for statistical inference. Problems of this type are called inverse problems and arise in many fields such as medical imaging, physics and biology. Mathematically the connection between the quantity of interest and the observable one can often be expressed in terms of a linear operator equation. Well known examples are Positron Emission Tomography, which involves the Radon Transform, (Cavalier (2000)), the heat equation (Mair and Ruymgaart (1996)), the Laplace Transform Saitoh (1997) and the reconstruction of astronomical and biological images from telescopic and microscopic imaging devices, which is closely connected to convolution-type operators (Adorf (1995), Bertero et al. (2009)).

Inverse problems have been studied intensively in a deterministic framework and in mathematical physics. See for example Engl et al. (1996) for an overview of existing methods in numerical analysis of inverse problems or Saitoh (1997) for techniques based on reproducing kernel Hilbert spaces. Recently, the investigation of inverse problems has also become of importance from a statistical point of view. Here, a particularly interesting and active field of research is the construction of statistical inference methods such as hypothesis tests or confidence regions.

In this paper we are interested in the convolution type inverse regression model

$$Y = (f * \psi)(x) + \varepsilon, \tag{1.1}$$

where ε is a random error, the operation $*$ denotes convolution, ψ is a given square integrable function and the object of interest is the function f itself. An important and interesting application of the inverse regression model (1.1) is the recovery of images from imaging devices such as astronomical telescopes or fluorescence microscopes in biology. In these cases, the observed, uncorrected image is always at least slightly blurry due to the physical characteristics of the propagation of light at surfaces of mirrors and lenses in the telescope. In this application the variable x represents the pixel of a CCD and we can only observe a blurred version of the true image modeled by the function f. In the corresponding mathematical model the observed image is (at least approximately) a convolution of the real image with the so-called point-spread-function ψ, i.e. an inverse problem with convolution operator.

The inference problem regarding the function f is called inverse problem with stochastic noise. In recent years, the problem of estimating the regression function f has become an important field of research, where the main focus is on a one dimensional predictor. Several authors propose Bayesian methods (Bertero et al. (2009); Kaipio and Somersalo (2010)) and construct estimators using tools from nonparametric curve estimation (Mair and Ruymgaart (1996); Cavalier (2008); Bissantz et al. (2007b)). Further inference methods, in particular the construction of confidence intervals and confidence bands, are much less developed. Birke et al. (2010) have constructed uniform confidence bands for the function f with a one-dimensional predictor.

The present work is motivated by the fact that in many applications one has to deal with an at least two-dimensional predictor. A typical example is image reconstruction since a picture is a two-dimensional object. Also in addition to the spatial dimensions, the data often show a

dynamical behavior, thus repeated measurements at different times can be used to extend the statistical inference. For example in astrophysics spectra of different objects like supernovae or variable stars underly changes in time on observable timescales. In this case the function f depends on a further parameter, say f_t and the reconstruction problem refers to a multivariate function even if the predictor is univariate.
The purpose of the present paper is the investigation of asymptotic properties of estimators for the function f in model (1.1) with a multivariate predictor. In particular we present a result on the weak convergence of the sup-norm of an appropriately centered estimate, which can be used to construct asymptotic confidence bands for the regression function f. In contrast to other authors (e.g. Cavalier and Tsybakov (2002)) we do not assume that the function ψ in model (1.1) is periodic, because in the reconstruction of astronomical or biological images from telescopes or microscopic imaging devices this assumption is often unrealistic.

1.2 Confidence bands

In a pioneering work, Bickel and Rosenblatt (1973b) extended results of Smirnov (1950) for a histogram estimate and constructed confidence bands for a density function of independent identical distributed (i.i.d) observations. Their method is based on the asymptotic distribution of the supremum of a centered kernel density estimator. Since then, their method has been further developed both in the context of density and regression estimation. For density estimation, Neumann (1998) derived bootstrap confidence bands, and Giné and Nickl (2010) derived adaptive asymptotic bands over generic sets. In a regression context, asymptotic confidence bands were constructed by Eubank and Speckman (1993) for the Nadaraya-Watson and by Xia (1998) for a local polynomial estimator. Bootstrap confidence bands for nonparametric regression were proposed by Hall (1993), Neumann and Polzehl (1998) and by Claeskens and van Keilegom (2003). For the statistical inverse problem of deconvolution density estimation, Bissantz et al. (2007a) constructed asymptotic and bootstrap confidence bands, where Lounici and Nickl (2011) obtained non-asymptotic confidence bands by using concentration inequalities. Recently, Birke et al. (2010) provided uniform asymptotic and bootstrap confidence bands for a spectral cut-off estimator in the one-dimensional indirect regression model with convolution operator.
All these results are limited to the estimation of univariate densities and regression functions, and are not applicable in cases, where the quantity of interest depends on a multivariate predictor. In such cases - to the best knowledge of the authors - confidence bands are not available. One reason for this gap is that a well-established way to construct asymptotic uniform confidence bands, which uses a pioneering result of Bickel and Rosenblatt (1973b) as the standard tool, cannot be extended in a straightforward manner to the multivariate case. There are substantial differences between the multivariate and one-dimensional case, and for multivariate inverse problems the mathematical construction of confidence bands requires different and/or extended methodology.

In the present paper we will consider the problem of constructing confidence bands for the regression function in an inverse regression model with a convolution-type operator with a multivariate predictor. The estimators and assumptions for our asymptotic theory are presented in Section 2, while Section 3 contains the main results of the paper. In Section 4 we consider the special case of time dependent regression functions with a univariate predictor, which originally motivated our investigations. The arguments of Section 5 and 6, which contain all technical details of the proofs, are based on results by Piterbarg (1996). These authors provided a limit theorem for the supremum

$$\sup_{t \in T_n} X(t)$$

of a stationary Gaussian field $\{X(t) \,|\, t \in \mathbb{R}^d\}$, where $\{T_n \subset \mathbb{R}^d\}_{n\in\mathbb{N}}$ is an increasing system of sets such that $\lambda^d(T_n) \to \infty$ as $n \to \infty$. This result generalized the multivariate extension in Bickel and Rosenblatt (1973a), who provided a limit theorem for the supremum $\sup_{t\in[0,T]^d} X(t)$, as $T \to \infty$.

2 Notation and assumptions

2.1 Model and notations

Suppose that $(2n+1)^d$ observations $(x_{\mathbf{k}}, Y_{\mathbf{k}}), \mathbf{k} = (k_1, \ldots, k_d) \in G_{\mathbf{n}} := \{-n, \ldots, n\}^d$ from the model

$$Y_{\mathbf{k}} = g(x_{\mathbf{k}}) + \varepsilon_{\mathbf{k}} := (f * \psi)(x_{\mathbf{k}}) + \varepsilon_{\mathbf{k}}, \tag{2.1}$$

are available, where the function $f : \mathbb{R}^d \to \mathbb{R}$ is unknown, $\psi : \mathbb{R}^d \to \mathbb{R}$ is a known function and $g := f * \psi$ denotes the convolution of f and ψ, that is

$$g(x) := (f * \psi)(x) := \int_{\mathbb{R}^d} f(s)\psi(x-s)\, ds. \tag{2.2}$$

The basic assumptions that guarantee the existence of the integral (2.2) and also assure $g \in L^2(\mathbb{R}^d)$ is that $f \in L^2(\mathbb{R}^d)$ and $\psi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, which will be assumed throughout this paper. In model (2.1) the predictors $x_{\mathbf{k}} := \mathbf{k} \cdot \frac{1}{na_n}$ are equally spaced fixed design points on a d-dimensional grid, with a sequence $(a_n)_{n\in\mathbb{N}}$ satisfying

$$na_n \to \infty \quad \text{and} \quad a_n \searrow 0 \quad \text{for} \quad n \to \infty.$$

The noise terms $\{\varepsilon_{\mathbf{k}} \,|\, \mathbf{k} \in G_{\mathbf{n}}\}$ are a field of centered i.i.d. random variables with variance $\sigma^2 := \mathbb{E}\varepsilon_{\mathbf{k}}^2 > 0$ and existing fourth moments. As a consequence of the convolution theorem and the formula for Fourier inversion we obtain the representation

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \frac{\mathcal{F}g(\xi)}{\mathcal{F}\psi(\xi)} \exp(i\xi^T x)\, d\xi. \tag{2.3}$$

An estimator for the regression function f can now easily be obtained by replacing the unknown quantity $\mathcal{F}g = \mathcal{F}(f * \psi)$ by an estimator $\mathcal{F}\hat{g}$. The random fluctuations in the estimator $\mathcal{F}\hat{g}$ cause instability of the ratio $\frac{\mathcal{F}\hat{g}(\xi)}{\mathcal{F}\psi(\xi)}$ if at least one of the components of ξ is large. As a consequence, the problem at hand is ill-posed and requires regularization. We address this issue by excluding large values of ξ_j for any $j = 1, \ldots, d$ from the domain of integration, i.e. we multiply the integrand in (2.3) with a sequence of Fourier transforms $\mathcal{F}\eta(h\cdot)$ of smooth functions with compact support $[-h^{-1}, h^{-1}]^d$. Here $h = h_n$ is a regularization parameter which corresponds to a bandwidth in nonparametric curve estimation and satisfies $h \to 0$ if $n \to \infty$. For the exact properties of the function η we refer to Assumption A below.
An estimator $\hat{f}_n$ for the function f in model (2.1) is now easily obtained as

$$\hat{f}_n(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \frac{\mathcal{F}\hat{g}(\xi)}{\mathcal{F}\psi(\xi)} \exp(i\xi^T x)\mathcal{F}\eta(h\xi)\, d\xi, \tag{2.4}$$

where

$$\mathcal{F}\hat{g}(\xi) = \frac{1}{(2\pi)^{\frac{d}{2}} n^d a_n^d} \sum_{\mathbf{k} \in G_n} Y_{\mathbf{k}} \exp(-i\xi^T x_{\mathbf{k}})$$

is the empirical analogue of Fourier transform of g. Note that with the definition of the kernel

$$K_n(x) = \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} \frac{\mathcal{F}\eta(\xi)}{\mathcal{F}\psi(\frac{\xi}{h})} \exp(i\xi^T x)\, d\xi, \tag{2.5}$$

the estimator (2.4) has the following representation

$$\hat{f}_n(x) = \frac{1}{(2\pi)^d n^d a_n^d h^d} \sum_{\mathbf{k} \in G_{\mathbf{n}}} Y_{\mathbf{k}} K_n\left((x - x_{\mathbf{k}})\frac{1}{h}\right). \tag{2.6}$$

Note that the kernel K_n can be expressed as a Fourier transform as follows

$$K_n = \overline{\mathcal{F}\Big(\frac{\mathcal{F}\eta}{\mathcal{F}\psi(\frac{\cdot}{h})}\Big)}.$$

The first step of the proof of our main result (see Theorem 1 in Section 3) will consist of a uniform approximation of $\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)$ by an appropriate stationary Gaussian field. In the second step, we apply results of Piterbarg (1996) and Bickel and Rosenblatt (1973a) to obtain the desired uniform convergence for the approximation process of the first step. Finally, these results are used to construct uniform confidence regions for $\mathbb{E}\hat{f}_n(x)$. Our approach is then based on undersmoothing: the choice of sufficiently small bandwidths assures the same limiting behaviour of $\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)$ and $\hat{f}_n(x) - f(x)$. This avoids the estimation of higher order derivatives, which often turns out to be difficult in applications. Thus, the limit theorem obtained in the second step will also provide uniform confidence regions for the function f itself. Whereas undersmoothing implies that the rate-optimal bandwidth cannot be used, there has also been some theoretical justification why this choice of the regularization parameter is useful for constructing confidence intervals (see Hall (1992)).

2.2 Assumptions

We now introduce the necessary assumptions which are required for the proofs of our main results in Section 3 . The first assumption refers to the type of (inverse) deconvolution problem describing the shape of the kernel function η in the spectral domain.

Assumption A. Let $\mathcal{F}\eta$ denote the Fourier transform of a function η such that

A1. $\operatorname{supp}\big(\mathcal{F}\eta\big) \subset [-1,1]^d$.

A2. $\mathcal{F}\eta \in \mathscr{D}(\mathbb{R}^d) = \{f : \mathbb{R}^d \to \mathbb{R} \,|\, f \in C^\infty(\mathbb{R}^d), \operatorname{supp}(f) \subset \mathbb{R}^d \text{ compact}\}$.

A3. There exists a constant $D > 0$, such that $\mathcal{F}\eta(\xi) = 1$ for all $\xi \in [-D, D]^d$ and $|\mathcal{F}\eta(\xi)| \leq 1$ for all $\xi \in \mathbb{R}^d$.

Remark 1.

1. The decay of the tails of the kernel K_n is given in terms of the smoothness of the integrand in (2.5). The choice of a smooth regularizing function $\mathcal{F}\eta$ has the advantage that the smoothness of $1/\mathcal{F}\psi$ carries over to $\mathcal{F}\eta(h\cdot)/\mathcal{F}\psi$.

2. Functions like $\mathcal{F}\eta$ are called bump functions. Their existence follows from the C^∞ Urysohn Lemma (see for example Folland (1984), Lemma 8.18).

3. Note that $\mathscr{D}(\mathbb{R}^d) \subset \mathscr{S}(\mathbb{R}^d)$, where $\mathscr{S}(\mathbb{R}^d)$ denotes the Schwartz space of smooth and rapidly decreasing functions. Since $\mathcal{F} : \mathscr{S}(\mathbb{R}^d) \to \mathscr{S}(\mathbb{R}^d)$ is a bijection (see for example Folland (1984), Corollary 8.28) we know that $\eta \in \mathscr{S}(\mathbb{R}^d)$ as well.

4. For the sake of transparency, we state the conditions and results with the same regularization parameter h for each direction. In practical applications this might not be the best strategy. The results presented in Section 3 and 4 also hold for different sequences of bandwidths $h_1, \ldots, h_d$ as long as the system of rectangles $\{[0, h_1^{-1}]\times, \ldots, [0, h_d^{-1}] \,|\, n \in \mathbb{N}\}$ is a blowing

up system of sets in the sense of Definition 14.1 in Piterbarg (1996). This is the case if the assumption

$$\sum_{p=1}^{d}\Big(\prod_{j=1,j\neq p}^{d}\frac{1}{h_j}\Big)\leq L_1\cdot\Big(\prod_{j=1}^{d}\frac{1}{h_j}\Big)^{L_2},$$

is satisfied for a constant L_1 that only depends on d and a constant $L_2<1$. This condition is not a restriction in our setting because it holds whenever $h_j\cdot n^{\gamma_j}\to C_j$ for constants $C_j,\gamma_j>0,\ \ j=1,\ldots,d$.

In general, two kinds of convolution problems are distinguished in the literature, because the decay of the Fourier transform of the convolution function ψ determines the degree of ill-posedness. In the case of an exponentially decreasing Fourier transform $\mathcal{F}\psi$ the problem is called severely ill-posed. In the present paper the class of moderately ill-posed problems is considered, where the Fourier transform of the convolution function decays at a polynomial rate (the precise condition will be specified in Assumption B below). Throughout this paper

$$\mathcal{W}^m(\mathbb{R}^d)=\{f\in L^2(\mathbb{R}^d)\,|\,\partial^{(\alpha)}f\in L^2(\mathbb{R}^d)\,\text{exists}\,\forall\,\alpha\in\mathbb{N}^d,\,|\alpha|\leq m\},$$

denotes the Sobolev space of order $m\in\mathbb{N}$, where $\partial^{(\alpha)}f$ is the weak derivative of f of order α. In the subsequent discussion we will also make use of the Sobolev space for general $m>0$, which is defined by

$$\mathcal{W}^m(\mathbb{R}^d)=\{f\in L^2(\mathbb{R}^d)\,|\,(1+|\xi|^2)^{\frac{m}{2}}\mathcal{F}f\in L^2(\mathbb{R}^d)\}.$$

Assumption B. We assume the existence of a function $\Psi:\mathbb{R}^d\to\mathbb{R}$ such that the kernel $K=\overline{\mathcal{F}\left(\Psi\cdot\mathcal{F}\eta\right)}$ satisfies

B1. $K\neq 0$ and there exist constants $\beta>d/2$, $M\in\mathbb{N}$, indices $0<\mu_1<\mu_2<\ldots<\mu_M$ and L^2-functions $f_1,\ldots,f_{M-1},f_M:\mathbb{R}^d\to\mathbb{R}$ with the property

$$\xi^\alpha f_p\in\mathcal{W}^m(\mathbb{R}^d)\qquad(p=1,\ldots,M-1)$$

for all multi-indices $\alpha\in\{0,\ldots,d\}^d$, $|\alpha|\leq d$ and all $m>\frac{d+|\alpha|}{2}$, such that

$$h^\beta K_n(x)-K(x)=\sum_{p=1}^{M-1}h^{\mu_p}\overline{\mathcal{F}f_p}(x)+h^{\mu_M}\overline{\mathcal{F}f_{n,M}}(x). \tag{2.7}$$

where f_M may depend on n, i.e. $f_M=f_{M,n}$ and $\|f_{M,n}\|_{L^1(\mathbb{R}^d)}=O(1)$.

B2. $\xi^\alpha\Psi\cdot\mathcal{F}\eta,\ \xi^\alpha\frac{h^\beta}{\mathcal{F}\psi(\frac{\cdot}{h})}\cdot\mathcal{F}\eta\in\mathcal{W}^m(\mathbb{R}^d)$ for some $m>\frac{d+|\alpha|}{2}$.

B3. $\log(n) \cdot h^{\mu_M}(a_n^{-\frac{d}{2}} h^{-\frac{d}{2}}) \cdot \|f_M\|_{L^1(\mathbb{R}^d)} = o(1)$ and $h^{\mu_1}(\log(n))^2 = o(1)$.

Remark 2. Assumption B1 implies $h^\beta K_n \to K$ in $L^2(\mathbb{R}^d)$ and also specifies the order of this convergence. It can be understood as follows. If the convergence of the difference $h^\beta K_n - K$ is fast enough, i.e.

$$\log(n) \cdot h^{\mu_1}(a_n h)^{-\frac{d}{2}} = o(1) \tag{2.8}$$

we have $M = 1$. On the other hand, in some relevant situations (see Example 1 (ii) below) the rate of convergence h^{μ_1} is given by h^2 for each d and (2.8) cannot hold for $d \geq 4$. Here, the expansion (2.7) provides a structure, such that our main results remain correct although the rate of convergence is not very fast. We can decompose the difference $h^\beta K_n - K$ in two parts, where one part depends on n only through the factors h^{μ_p} and the other part converges sufficiently fast (in some cases this term vanishes completely).

Example 1. This example illustrates the construction of the functions in the representation (2.7).

(i) Let $d = 2$ and $\psi(x) = \frac{1}{4}\exp(-|x_1|)\exp(-|x_2|)$, $x = (x_1, x_2)^T$, $\xi = (\xi_1, \xi_2)^T$. Then we have $\frac{h^4}{\mathcal{F}\psi(\xi)} = 2\pi\big(h^4 + h^2(\xi_1^2 + \xi_2^2) + \xi_1^2\xi_2^2\big)$, which implies $\beta = 4$, $M = 3$ and

$$h^4 \cdot K_n(x) = \int_{\mathbb{R}^2} \big(h^4 + h^2(\xi_1^2 + \xi_2^2) + \xi_1^2\xi_2^2\big)\mathcal{F}\eta(\xi)\exp(ix^T\xi)\,d\xi$$
$$K(x) = \int_{\mathbb{R}^2} \mathcal{F}\eta(\xi)\xi_1^2\xi_2^2\exp(ix^T\xi)\,d\xi.$$

With the definitions $f_1(\xi) = 2\pi(\xi_1^2 + \xi_2^2)\mathcal{F}\eta(\xi)$, $f_2(\xi) = 2\pi\mathcal{F}\eta(\xi)$ and $f_{n,3} \equiv 0$ we obtain

$$h^4 \cdot K_n(x) - K(x) = h^2 \cdot \overline{\mathcal{F}f_1(\xi)} + h^4 \cdot \overline{\mathcal{F}f_2(\xi)}.$$

In this example, the condition $\log(n)h^2/\sqrt{a_n^d h^d} = o(1)$ is satisfied. However, the following results are valid if the weaker condition of a decomposition of the form (2.7) holds. Furthermore, since the factors of $\mathcal{F}\eta$ in f_1 and f_2 are polynomials, we have $\mathcal{F}f_j(\xi) \in \mathscr{S}(\mathbb{R}^d)$, which implies $\xi^\alpha f_j \in \mathcal{W}^m(\mathbb{R}^d)$ for all α and all $m \in \mathbb{N}$.

(ii) If $|x| = \sqrt{x_1^2 + \ldots + x_d^2}$ and $\psi(x) = 2^{-\frac{d+1}{2}}e^{-|x|}$ we have

$$\mathcal{F}\psi(\xi) = \frac{1}{\sqrt{2\pi}}\Gamma\left(\frac{d+1}{2}\right)\frac{1}{(1+|\xi|^2)^{\frac{d+1}{2}}},$$

[see Folland (1984), Exercise 13]. If d is odd we use the identity

$$\left(h^2 + |\xi|^2\right)^{\frac{d+1}{2}} = \sum_{j=0}^{\frac{d+1}{2}} \binom{\frac{d+1}{2}}{j} h^{2j} |\xi|^{\frac{d+1}{2}-2j},$$

and an expansion of the form (2.7) is obvious from the definition of K_n in (2.5). If the dimension d is even the situation is more complicated. Consider for example the case $d = 4$, where

$$\frac{h^5}{\mathcal{F}\psi(\frac{\xi}{h})} = \rightarrow \frac{\sqrt{2\pi}}{\Gamma\left(\frac{5}{2}\right)} |\xi|^5 = \frac{\sqrt{2\pi}}{\Gamma\left(\frac{5}{2}\right)} \sqrt{\left(\xi_1^2 + \xi_2^2 + \xi_3^2 + \xi_4^2\right)^5} \quad \text{as} \quad n \to \infty.$$

It follows that the constant β and the functions Ψ, K_n and K from Assumption B are given by $\beta = d + 1 = 5$, $\Psi(\xi) = \frac{\sqrt{2\pi}}{\Gamma\left(\frac{5}{2}\right)} |\xi|^5$ and

$$\begin{aligned} h^\beta K_n(x) &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^d} \frac{\sqrt{2\pi}}{\Gamma\left(\frac{5}{2}\right)} \left(h^2 + |\xi|^2\right)^{\frac{d+1}{2}} \mathcal{F}\eta(\xi) \exp(i\xi^T x)\, d\xi, \\ K(x) &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^d} \frac{\sqrt{2\pi}}{\Gamma\left(\frac{5}{2}\right)} |\xi|^{d+1} \mathcal{F}\eta(\xi) \exp(i\xi^T x)\, d\xi, \end{aligned}$$

respectively. In order to show that Assumption B1 holds in this case we use Taylor's Theorem and obtain

$$\frac{h^5}{\mathcal{F}\psi(\frac{\xi}{h})} - \Psi(\xi) = \frac{\sqrt{2\pi}}{\Gamma\left(\frac{d+1}{2}\right)} \Big(h^2 \cdot \frac{5}{2} \cdot |\xi|^3 + h^4 \cdot \frac{5}{2} \cdot \frac{3}{2} \cdot (|\xi|^2 + \lambda_d h^2)^{\frac{1}{2}} \Big),$$

for some constant $\lambda_d \in [0, 1)$. Recalling the definition of K_n in (2.5) this gives

$$(h^\beta K_n - K)(x) = h^2\, \overline{\mathcal{F} f_1(\xi)} + h^4\, \overline{\mathcal{F} f_{2,n}(\xi)}$$

where the functions f_1 and $f_{2,n}$ are defined by

$$\begin{aligned} f_1(\xi) &= \frac{1}{(2\pi)^{\frac{3}{2}} \Gamma\left(\frac{5}{2}\right)} \cdot |\xi|^3 \cdot \frac{5}{2} \cdot \mathcal{F}\eta(\xi), \\ f_{2,n}(x) &= \frac{1}{(2\pi)^{\frac{3}{2}} \Gamma\left(\frac{5}{2}\right)} \frac{5}{2} \cdot \frac{3}{2} h^4 (|\xi|^2 + \lambda_d h^2)^{\frac{1}{2}} \cdot \mathcal{F}\eta(\xi), \end{aligned}$$

respectively. It can be shown by a straightforward calculation that $\xi^\alpha f_j \in \mathcal{W}^{6+|\alpha|}(\mathbb{R}^d)$ for all $\alpha \in \{0, \ldots, d\}^d$.

Remark 3. In the one-dimensional regression model (2.1), Birke et al. (2010) assume that the kernel K has exponentially decreasing tails in order to obtain asymptotic confidence bands, which, in combination with the other assumptions only allows for kernels that are Fourier transforms of C^∞-functions with square integrable derivatives. Our Assumption B is already satisfied if K is the Fourier transform of a once weakly differentiable function with square integrable weak derivative, such that all indices of ill-posedness β that satisfy $\beta > \frac{1}{2}$ are included if $d = 1$. Moreover, the assumptions regarding the bandwidths are less restrictive compared to Birke et al. (2010).

Our final assumptions refer to the smoothness of the function f and to the decay of the convolution $f * \psi$.

Assumption C. We assume that

C1. There exist constants $\gamma > 2$, $m > \gamma + \frac{d}{2}$ such that $f \in \mathcal{W}^m(\mathbb{R}^d)$.

C2. There exists a constant $\nu > 0$ such that

$$\int_{\mathbb{R}} |(f * \psi)(z)|^2 \, (1 + |z|^2)^\nu \, dz < \infty.$$

3 Asymptotic confidence regions

In this section we construct asymptotic confidence regions for the function f on the unit cube $[0,1]^d$. These results can easily be generalized to arbitrary rectangles $\times_{j=1}^d [a_j, b_j]$ for fixed constants $a_j < b_j$ $(j = 1, \dots, d)$ and the details are omitted for the sake of brevity. We investigate the limiting distribution of the supremum of the process $\{\tilde{Y}_n(x) \,|\, x \in [0,1]^d\}$, where

$$(3.1) \qquad \tilde{Y}_n(x) = \frac{(2\pi)^d h^\beta \sqrt{h^d n^d a_n^d}}{\sigma \|K\|_{L^2(\mathbb{R}^d)}} \left[\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x) \right] = \frac{(2\pi)^d h^\beta}{\sigma \|K\|_{L^2(\mathbb{R}^d)} \sqrt{h^d n^d a_n^d}} \sum_{\mathbf{k} \in G_n} K_n\Big((x - x_{\mathbf{k}}) \frac{1}{h} \Big) \varepsilon_{\mathbf{k}}.$$

and the kernel K_n is defined in (2.5). Note that

$$\sup_{x \in [0,1]^d} |\tilde{Y}_n(x)| = \sup_{x \in [0, h^{-1}]^d} |Y_n(x)|,$$

where the process

$$(3.2) \qquad Y_n(x) := \frac{(2\pi)^d h^\beta}{\sigma \|K\|_{L^2(\mathbb{R}^d)} \sqrt{h^d n^d a_n^d}} \sum_{\mathbf{k} \in G_n} K_n\Big(x - x_{\mathbf{k}} \frac{1}{h} \Big) \varepsilon_{\mathbf{k}}$$

can be approximated by a stationary Gaussian field uniformly with respect to $[0, h^{-1}]^d$. Thus the desired limiting distribution corresponds to the limiting distribution of the supremum of a stationary Gaussian process over a system of increasing smooth sets with sufficient similarity of their speed of increase, and is therefore of Gumbel-type. The precise result is given in the following Theorem.

Theorem 1. *Assume that for some fixed constant $\delta \in (0,1]$, $\delta < d$ and a constant $r > \frac{2d}{d-\delta}$ the r-th moment of the errors exists, i.e. $\mathbb{E}|\varepsilon_{\mathbf{k}}|^r < \infty$. If additionally Assumptions A and B are satisfied and $\frac{\log(n)}{n^\delta a_n^\delta h^d} = o(1)$, then we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{x\in[0,1]^d} \left(|\tilde{Y}_n(x)| - C_{n,3}\right) \cdot C_{n,3} < \kappa\right) = e^{-2e^{-\kappa}},$$

where

$$C_1 = \mathbf{det}\left(\left[\frac{(2\pi)^{2d}}{\|K\|_2^2}\int_{\mathbb{R}^d} |\Psi(v)\mathcal{F}\eta(v)|^2 v_i v_j \, dv\right], \, i,j = 1,\ldots,d\right)$$

$$C_{n,2} = \sqrt{\frac{C_1}{(2\pi)^{d+1}}\frac{1}{h^d}}$$

$$C_{n,3} = \sqrt{2\ln(C_{n,2})} + \frac{(d-1)\ln\big(2\ln(C_{n,2})\big)}{2\sqrt{2\ln(C_{n,2})}}.$$

The proof of this result is long and complicated and therefore deferred to Section 5 and 6. In the following we apply Theorem 1 to construct uniform confidence regions for the function f by choosing the bandwidth such that the bias decays to zero sufficiently fast. More precisely, if the condition

$$\log(n) \sup_{x\in[0,1]^d} \Big|f(x) - \mathbb{E}\hat{f}_n(x)\Big| = o\left(\left(h^\beta \sqrt{h^d n^d a_n^d}\right)^{-1}\right)$$

is satisfied, it follows directly that the random quantities $\sup_{x\in[0,1]^d} |\tilde{Y}_n(x)|$ and

$$\frac{(2\pi)^d h^\beta \sqrt{h^d n^d a_n^d}}{\|K\|_{L^2(\mathbb{R}^d)}\sigma} \sup_{x\in[0,1]^d} \Big|f(x) - \hat{f}_n(x)\Big|$$

have the same limiting behavior.

Corollary 1. *Assume that the conditions of Theorem 1, Assumption C and the condition*

$$\sqrt{h^d n^d a_n^d}\sqrt{\log(n)}\Big(\frac{1}{n^3 a_n^3 h^2} + \frac{a_n^\nu}{n} + a_n^{\nu+\frac{d}{2}} + h^{\gamma+\beta}\Big) = o(1) \quad \textit{for} \quad n \to \infty$$

are satisfied. Then we have for any $\kappa \in \mathbb{R}$

$$\lim_{n\to\infty} \mathbb{P}\left(\hat{f}_n(x) - \Phi_{n,\kappa} \leq f(x) \leq \hat{f}_n(x) + \Phi_{n,\kappa} \text{ for all } x \in [0,1]^d\right) = e^{-2e^{-\kappa}},$$

where the sequence $\Phi_{n,\kappa}$ *is defined by*

$$\Phi_{n,\kappa} = \frac{(\frac{\kappa}{C_{n,3}} + C_{n,3})\sigma\|K\|_{L^2(\mathbb{R}^d)}}{(2\pi)^d h^\beta \sqrt{h^d n^d a_n^d}}.$$

As a consequence of Corollary 1 an asymptotic uniform confidence region for the function f with confidence level $1-\alpha$ is given by

$$\{[\hat{f}_n(x) - \Phi_{n,-\ln(-0.5\ln(1-\alpha))}, \hat{f}_n(x) + \Phi_{n,-\ln(-0.5\ln(1-\alpha))}] \,|\, x \in [0,1]^d\}. \tag{3.3}$$

The corresponding $(1-\alpha)$-band has a width of $2\Phi_{n,-\ln(-0.5\ln(1-\alpha))}$. Here, the factor $\frac{1}{h^\beta}$ is due to the ill-posedness of the inverse problem (see Assumption B). It does not appear in corresponding results for the direct regression case. On the other hand the factor $a_n^{-\frac{d}{2}}$ arises from the design on the growing system of sets $\{[-a_n^{-1}, a_n^{-1}]^d \,|\, n \in \mathbb{N}, \}$. In the case of a regression on a fixed interval it does not appear as well. The width of the asymptotic point-wise confidence intervals in the multivariate indirect regression case as obtained in Bissantz and Birke (2009) is of order $\frac{1}{h^\beta\sqrt{Nh^d a_n^d}}$, where N is the total number of observations. Their point-wise confidence intervals are smaller than the uniform ones obtained in Corollary 1. The price for uniformity is an additional factor of logarithmic order, which is typical for results of this kind.
In applications the standard deviation is unknown but can be estimated easily from the data, because this does not require the estimation of the function f. In particular, (3.3) remains an asymptotic $(1-\alpha)$-confidence band, if σ is replaced by an estimator satisfying $\hat{\sigma} - \sigma = o_P(1/\log(n))$.

4 Time dependent regression functions

In this section we extend model (2.1) to include a time dependent regression function, that is

$$Y_{j,\mathbf{k},n} = \big(T_\psi f_{t_j}\big)(x_\mathbf{k}) + \varepsilon_\mathbf{k}, \quad \mathbf{k} \in G_\mathbf{n},\ j = -m, \ldots, m, \tag{4.1}$$

where $x_\mathbf{k} = \frac{\mathbf{k}}{na_n}$ and $t_j = \frac{j}{mb_m}$, $m = m(n)$, such that $m(n) \to \infty$ and $b_{m(n)} \searrow 0$ as $n \to \infty$.
We assume that ψ does not depend on the time and the operator T_ψ is defined by

$$\big(T_\psi f_t\big) = \int_{\mathbb{R}^d} f_t(y)\psi(\cdot - y)\, dy.$$

This assumption is reasonable in the context of imaging where the function ψ corresponds to the point spread function (Bertero et al. (2009)). If it is not satisfied, i.e. the convolution operator effects all coordinates, the problem can be modeled as in Section 2.
For a precise statement of the results we will add an index to the Fourier operator $\mathcal{F}$ which gives the dimension of the space under consideration. We will write $\mathcal{F}_{d+1}$ if the Fourier transform is taken over the whole space $\mathbb{R}^{d+1}$ and $\mathcal{F}_d$ to denote Fourier transformation with respect to the spatial dimensions. By the same considerations as given in Section 2 we obtain an estimator $\check{f}$ for the function f_t

$$\begin{aligned}\check{f}_n(x;t) &= \frac{1}{(2\pi)^{\frac{d+1}{2}}} \int_{\mathbb{R}^{d+1}} \frac{\mathcal{F}_{d+1}(\widehat{f*\psi})(\xi,\tau)}{(2\pi)^{\frac{d}{2}}\mathcal{F}_d\psi(\xi)} \mathcal{F}_d\check{\eta}(\xi h, \tau h_t)\exp(it\tau + ix^T\xi)\, d(\xi,\tau) \\ &= \frac{1}{(2\pi)^{d+\frac{1}{2}} n^d m a_n^d b_m} \sum_{(\mathbf{k},j)\in G^{d+1}_{(\mathbf{n},m)}} Y_{\mathbf{k},j} \check{K}_n\Big(\frac{x-x_{\mathbf{k}}}{h}, \frac{t-t_j}{h_t}\Big),\end{aligned}$$

where $G^{d+1}_{(n,m)}$ denotes the grid $\{-n,\dots,n\}^d \times \{-m,\dots,m\}$ and the kernel $\check{K}_n$ is given by

$$\check{K}_n(x;t) = \frac{1}{(2\pi)^{\frac{d+1}{2}}} \int_{\mathbb{R}^{d+1}} \frac{\exp\big(i\tau t + i\xi^T x\big)}{\mathcal{F}_d\psi(\frac{\xi}{h})} \mathcal{F}_{d+1}\check{\eta}(\xi,\tau)\, d(\xi,\tau). \tag{4.2}$$

Here the function $\check{\eta} : \mathbb{R}^{d+1} \to \mathbb{R}$ satisfies condition A and $h_t = h_t(n)$ is an additional sequence of bandwidths referring to the time domain. For the asymptotic analysis we require a modified version of Assumption B.

Assumption $\check{\mathbf{B}}$ Let Assumptions B1 (with corresponding kernel $\check{K}$) and B2 hold and additionally assume that

$\check{B}$3. $\log(n+m(n)) \cdot h^{\mu_M}(a_n^{-\frac{d}{2}} h^{-\frac{d}{2}} b_{m(n)}^{\frac{1}{2}} m(n)^{\frac{1}{2}}) = o(1)$ and for $p = 1,\dots,M-1$
$h^{\mu_p}(\log(n+m))^2 = o(1)$.

Theorem 2. *Define*

$$\check{Y}_n(x;t) := \frac{(2\pi)^{d+1} h^\beta \sqrt{h^d h_t n^d m b_m a_n^d}}{\sigma \|\check{K}\|_{L^2(\mathbb{R}^{d+1})}\|} \big[\check{f}_n(x;t) - \mathbb{E}\check{f}_n(x;t)\big]$$

and let the moment condition of Theorem 1 and Assumptions A and $\check{B}$ be satisfied. We further assume that the bandwidths h_t and h, and the sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_{m(n)})_{n\in\mathbb{N}}$ satisfy

$$\log(n+m)\Bigg(\sqrt{\frac{na_n}{mb_m}} \frac{1}{\sqrt{n^\delta h_t a_n^\delta h^d}} + \Big(\frac{mb_m}{na_n}\Big)^{\frac{d}{2}} \frac{1}{\sqrt{m^\delta h_t h^d}}\Bigg) = o(1) \quad \textit{for} \quad n \to \infty$$

$$h_t + h \leq L_1 \cdot h^{d(1-L_2)} h_t^{(1-L_2)}$$

for some constants $L_1 < \infty$ and $L_2 \in (0,1)$. Then we have for each $\kappa \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{x\in[0,1]^d} \Big(|\check{Y}_n(x;t)| - D_{n,3} \Big) \cdot D_{n,3} < \kappa \right) = e^{-2e^{-\kappa}},$$

where

$$D_1 = \mathbf{det}\left(\left[\frac{(2\pi)^{2(d+1)}}{\|\check{K}\|^2_{L^2(\mathbb{R}^{d+1})}} \int_{\mathbb{R}^{d+1}} |\Psi(v_1,\ldots,v_d)\mathcal{F}_{d+1}\check{\eta}(v)|^2 v_i v_j \, dv \right], \, i,j = 1,\ldots,d+1 \right),$$

$$D_{n,2} = \sqrt{\frac{D_1}{(2\pi)^{d+2}} \frac{1}{h^d h_t}} \quad \text{and}$$

$$D_{n,3} = \sqrt{2\ln(D_{n,2})} + \frac{(d-1)\ln\big(2\ln(D_{n,2})\big)}{2\sqrt{2\ln(D_{n,2})}}.$$

Corollary 2. *If the assumptions of Theorem 2 are satisfied, the limit kernel $\check{K}$ is defined by*

$$\check{K}(x,t) = \frac{1}{(2\pi)^{\frac{d+1}{2}}} \int_{\mathbb{R}^{d+1}} \Psi(\xi)\mathcal{F}_{d+1}\check{\eta}(\xi,\tau) \exp(i\xi^T x + i\tau t) \, d(\xi,\tau). \tag{4.3}$$

and the function $f_{(\cdot)}(\cdot) : \mathbb{R}^{d+1} \to \mathbb{R}^1$ satisfies Assumption C, then it follows that

$$\lim_{n\to\infty} \mathbb{P}\left(\check{f}_n(x;t) - \check{\Phi}_{n,\kappa} \leq f(x;t) \leq \check{f}_n(x;t) + \check{\Phi}_{n,\kappa} \text{ for all } (x,t) \in [0,1]^{d+1} \right) = e^{-2e^{-\kappa}},$$

where the constant $\check{\Phi}_{n,\kappa}$ is defined by

$$\check{\Phi}_{n,\kappa} = \frac{(\frac{\kappa}{D_{n,3}} + D_{n,3})\sigma \|\check{K}\|_{L^2(\mathbb{R}^{d+1})}}{h^\beta \sqrt{h^d n^d a_n^d m b_m h_t} (2\pi)^{d+1}}.$$

Asymptotic confidence bands for the function $f_t(x)$ at level $1-\alpha$ are hence given by

$$\{[\check{f}_n(x;t) - \check{\Phi}_{n,-\ln(-0.5\ln(1-\alpha))}, \check{f}_n(x;t) + \check{\Phi}_{n,-\ln(-0.5\ln(1-\alpha))}] \,|\, (x,t) \in [0,1]^{d+1}\}.$$

Acknowledgments: This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes"(SFB 823, Teilprojekte A1 and C4) of the German Research Foundation (DFG). The authors would like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise.

References

Adorf, H. (1995). Hubble space telescope image restauration in its fourth year. *Inverse Problems*, 11:639–653.

Bertero, M., Boccacci, P., Desiderà, G., and Vicidomini, G. (2009). Image deblurring with Poisson data: from cells to galaxies. *Inverse Problems*, 25(12):123006, 26.

Bickel, P. and Rosenblatt, M. (1973a). Two-dimensional random fields. In *Multivariate Analysis, III (Proc. Third Internat. Sympos., Wright State Univ., Dayton, Ohio, 1972)*, pages 3–15. Academic Press, New York.

Bickel, P. J. and Rosenblatt, M. (1973b). On some global measures of the deviations of density function estimates. *Annals of Statistics*, 1:1071–1095.

Birke, M., Bissantz, N., and Holzmann, H. (2010). Confidence bands for inverse regression models. *Inverse Problems*, 26(11):115020, 18.

Bissantz, N. and Birke, M. (2009). Asymptotic normality and confidence intervals for inverse regression models with convolution-type operators. *Journal of Multivariate Analysis*, 100:2364–2375.

Bissantz, N., Dümbgen, L., Holzmann, H., and Munk, A. (2007a). Nonparametric confidence bands in deconvolution density estimation. *Journal of the Royal Statistical Society, Series B*, 69:483–506.

Bissantz, N., Hohage, T., Munk, A., and Ruymgaart, F. (2007b). Convergence rates of general regularization methods for statistical inverse problems. *SIAM J. Num. Anal.*, 45:2610–2636.

Cavalier, L. (2000). Efficient estimation of a density in a problem of tomography. *Annals of Statistics*, 28:630–647.

Cavalier, L. (2008). Nonparametric statistical inverse problems. *Inverse Problems*, 24(3):034004, 19.

Cavalier, L. and Tsybakov, A. (2002). Sharp adaptation for inverse problems with random noise. *Probab. Theory Relat. Fields*, 123:323–354.

Claeskens, G. and van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Annals of Statistics*, 31:1852–1884.

Engl, H. W., Hanke, M., and Neubauer, A. (1996). *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht.

Eubank, R. L. and Speckman, P. L. (1993). Confidence bands in nonparametric regression. *Journal of the American Statistical Association*, 88:1287–1301.

Folland, G. B. (1984). *Real Analysis - Modern Techniques and their Applications.* Wiley, New York.

Giné, E. and Nickl, R. (2010). Confidence bands in density estimation. *Annals of Statistics*, 38:1122–1170.

Hall, P. (1992). On bootstrap confidence intervals in nonparametric regression. *Annals of Statistics*, 20:695–711.

Hall, P. (1993). On Edgeworth expansion and bootstrap confidence bands in nonparametric regression. *Journal of the Royal Statistical Society, Ser. B*, 55:291–304.

Kaipio, J. and Somersalo, E. (2010). *Statistical and Computational Inverse Problems.* Springer, Berlin.

Khoshnevisan, D. (2002). *Multiparameter Processes - An Introduction.* Springer, New York.

Lounici, K. and Nickl, R. (2011). Global uniform risk bounds for wavelet deconvolution estimators. *Journal of Multivariate Analysis*, 39:201–231.

Mair, B. A. and Ruymgaart, F. H. (1996). Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, 56:1424–1444.

Neumann, M. H. (1998). Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. *Annals of Statistics*, 26:2014–2048.

Neumann, M. H. and Polzehl, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *Journal of Nonparametric Statistics*, 9:307–333.

Owen, A. B. (2005). Multidimensional variation for quasi-Monte Carlo. In *Contemporary Multivariate Analysis and Design of Experiments*, volume 2 of *Ser. Biostat.*, pages 49–74. World Sci. Publ., Hackensack, NJ.

Paranjape, S. R. and Park, C. (1973). Laws of iterated logarithm of multiparameter Wiener processes. *Journal of Multivariate Analysis*, 3:132–136.

Piterbarg, V. I. (1996). *Asymptotic methods in the theory of Gaussian processes and fields*, volume 148 of *Translations of Mathematical Monographs.* American Mathematical Society, Providence, RI. Translated from the Russian by V. V. Piterbarg, Revised by the author.

Rio, E. (1993). Strong approximation for set-indexed partial sum processes via KMT constructions I. *Annals of Probability*, 21(2):759–790.

Saitoh, S. (1997). *Integral Transforms, Reproducing Kernels and their Applications*. Longman, Harlow.

Smirnov, N. V. (1950). On the construction of confidence regions for the density of distribution of random variables. *Doklady Akad. Nauk SSSR.*, 74:189–191.

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *Journal of the Royal Statistical Society, Ser. B*, 60:797–811.

5 Proofs of Theorem 1 and Corollary 1

5.1 Notation, preliminaries and remarks

First, we introduce some notation which is used extensively in the following proofs. Define for $a = (a_1, \dots, a_d), b = (b_1, \dots, b_d) \in \mathbb{R}^d$ the d-dimensional cube $[a, b] := \times_{j=1}^d [a_j, b_j]$. Let $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{Z}^d$, $\alpha = (\alpha_1, \dots, \alpha_d) \in \{0, 1\}^d$ be multi-indices, $\mathbf{0} := (0, \dots, 0)^T \in \mathbb{R}^d$ and $\mathbf{1} := (1, \dots, 1)^T \in \mathbb{R}^d$ and define $G_{\mathbf{k}} := \mathbb{Z}^d \cap [-\mathbf{k},\ \mathbf{k}]$. For $j \in \{1, \dots, d\}$ we denote by $G_{\mathbf{k}}^j$ the canonical projection of $G_{\mathbf{k}}$ onto $\mathbb{Z}^j$, i.e. $G_{\mathbf{k}}^j$ is a j-dimensional grid of integers with possibly different length in each direction. For $j \in \mathbb{N}$ let $G_{\mathbf{k}}^{j,+} := G_{\mathbf{k}}^j \cap \mathbb{N}^j$ denote the part of the grid $G_{\mathbf{k}}^j$ whose vectors contain only positive components and write $G_{\mathbf{k}}^+$ for $G_{\mathbf{k}}^{d,+}$. We further introduce the bijective map

$$E_d : \begin{cases} \{0,1\}^d & \to \quad \mathcal{P}(\{1, \dots, d\}) \\ (\alpha_1, \dots, \alpha_d) & \mapsto \quad v = \{v_1, \dots, v_{|\alpha|}\}; \quad \alpha_{v_j} = 1,\ j = 1, \dots, |\alpha| = \sum_{i=1}^d \alpha_i, \end{cases}$$

that maps each α to the set $v \subset \{1, \dots, d\}$ that contains the positions of its ones. For $\alpha \in \{0, 1\}^d$ and $\{v_1, \dots, v_{|\alpha|}\} = E_d(\alpha)$ let $(x)_\alpha := (x_{v_1}, x_{v_2}, \dots, x_{v_{|\alpha|}})$ denote the projection of $x \in \mathbb{R}^d$ onto the space spanned by the coordinate axes given by the positions of ones of the multi-index α. For $a, b \in \mathbb{R}^d$ let $(a)_\alpha : (b)_{\mathbf{1}-\alpha} = (a : b)_\alpha := (a_1^{\alpha_1} \cdot b_1^{1-\alpha_1}, \dots, a_d^{\alpha_d} \cdot b_d^{1-\alpha_d})$ denote the vector of the components of a and b specified by the index α. The following example illustrates these notations.

Example 2. For $d = 2$ we have $\{0, 1\}^2 = \{(1, 1), (1, 0), (0, 1), (0, 0)\}$ and the mapping E_2 is defined by

$$E_2\big((1,1)\big) = \{1,2\},\ E_2\big((1,0)\big) = \{1\},\ E_2\big((0,1)\big) = \{2\} \quad \text{and} \quad E_2\big((0,0)\big) = \emptyset.$$

For any $x = (x_1, x_2) \in \mathbb{R}^2$ we have

$$(x)_{(1,1)} = x,\ (x)_{(1,0)} = x_1 \quad \text{and} \quad (x)_{(0,1)} = x_2.$$

For $a = (a_1, a_2),\ b = (b_1, b_2) \in \mathbb{R}^2$ we have

$$(a:b)_{(1,1)} = (a_1, a_2) = a,\ (a:b)_{(1,0)} = (a_1, b_2),\ (a:b)_{(0,1)} = (b_1, a_2),\ (a:b)_{(0,0)} = (b_1, b_2) = b.$$

For the approximation of the integrals by Riemann sums we define for multi-indices $\tilde{\alpha}, \alpha \in \{0,1\}^d \backslash \{\mathbf{0}\}$

$$\Delta_\alpha(f; a, b) := \sum_{\tilde{\alpha} \in \{0,1\}^d,\, \tilde{\alpha} \leq \alpha} (-1)^{|\tilde{\alpha}|} f((a:b)_{\tilde{\alpha}}) = \sum_{\tilde{\alpha} \in \{0,1\}^d,\, \tilde{\alpha} \leq \alpha} (-1)^{d-|\tilde{\alpha}|} f\big((a)_{\mathbf{1}-\tilde{\alpha}} : (b)_{\tilde{\alpha}}\big), \tag{5.1}$$

where the symbol $\tilde{\alpha} \leq \alpha$ means $\tilde{\alpha}_j \leq \alpha_j$ for $j = 1, \ldots, d$. Note that for $\alpha = \mathbf{1} \in \mathbb{R}^d$ we obtain the special case of the d-fold alternating sum, i.e.

$$\Delta(f; a, b) := \Delta_{\mathbf{1}}(f; a, b) = \sum_{\alpha \in \{0,1\}^d} (-1)^{|\alpha|} f\left((a:b)_\alpha\right) = \sum_{\alpha \in \{0,1\}^d} (-1)^{d-|\alpha|} f((a:b)_{\mathbf{1}-\alpha}),$$

Note that $\Delta_\alpha(f; a, b)$ can be regarded as the increment of the function $f_\alpha((x)_\alpha) := f\big((x:b)_\alpha\big)$ over the interval $[(a)_\alpha, (b)_\alpha]$ which also gives rise to the alternative notation

$$\Delta_\alpha\left(f; a, b\right) = \Delta\left(f_\alpha, (a)_\alpha, (b)_\alpha\right). \tag{5.2}$$

5.2 Proof of Theorem 1

To prove the assertion of Theorem 1 we decompose the index set $G_{\mathbf{n}} = \{-n, \ldots, n\}^d$ of the sum in (3.1) into $2^d + 1$ parts: the respective intersections with the 2^d orthants of the origin in $\mathbb{R}^d$ and the marginal intersections with the coordinate axes. Our first auxiliary result shows that the contribution of the term representing the marginals is negligible (here and throughout the paper we use the convention $0^0 = 1$).

Lemma 1.

$$\sup_{x \in [0, h^{-1}]^d} \Big| \frac{h^\beta}{\sqrt{h^d n^d a_n^d}} \sum_{\alpha \in \{0,1\}^d \backslash \{\mathbf{1}\}} \sum_{(\mathbf{k}:\mathbf{0})_\alpha, \mathbf{k} \in G_n^+} K_n\Big(x - \frac{1}{h} x_{\mathbf{k}}\Big) \varepsilon_{\mathbf{k}} \Big| = o_P\left(\frac{1}{\log(n)}\right).$$

We obtain from its proof in Section 6 that Lemma 1 holds under the weaker condition $\frac{\log(n)}{\sqrt{n a_n h}} = o(1)$, which follows from the assumptions of Theorem 1. Next we consider the "positive" orthant G_n^+ and show in three steps that

$$\sup_{x \in [0, h^{-1}]^d} \Big| Y_n^{(+)}(x) - Y^{(+)}(x) \Big| \ = \ o_p(1), \tag{5.3}$$

where the processes $Y_n^{(+)}$ and $Y^{(+)}$ are defined by

$$Y_n^{(+)}(x) := \frac{(2\pi)^d h^\beta}{\sigma\|K\|_{L^2}\sqrt{h^d n^d a_n^d}} \sum_{\mathbf{k}\in G_n^+} K_n(x - \frac{1}{h}x_{\mathbf{k}})\varepsilon_{\mathbf{k}}, \tag{5.4}$$

$$Y^{(+)}(x) := \frac{(2\pi)^d}{\|K\|_{L^2}} \int_{\mathbb{R}_+^d} K(x-u)\, dB(u), \tag{5.5}$$

respectively, B is a standard Brownian sheet on $\mathbb{R}^d$ (see the proof of Lemma 2 for details) and K denotes the kernel defined in Assumption B. The final result is then derived using Theorem 14.1 in Piterbarg (1996). To be precise note that it can easily be shown that

$$\lim_{n\to\infty} n^d a_n^d h^d h^{2\beta} \cdot \text{Var}\left(\hat{f}_n(x)\right) = \frac{\sigma^2}{(2\pi)^{2d}} \int_{\mathbb{R}^d} |K(\frac{x}{h} - u)|^2\, du = \frac{\sigma^2\|K\|_{L_2}^2}{(2\pi)^{2d}}$$

(in particular the limit is independent of the variable x, which is typical for kernel estimates in homoscedastic regression models with equidistant design). We further obtain for the function $r(t) = (2\pi)^{2d}\|K\|_{L^2}^{-2} \int_{\mathbb{R}^d} K(v+t)K(v)dv$ that

$$\|r\|_{L^1} = \frac{(2\pi)^{2d}}{\|K\|_{L^2}^2} \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} K(v+t)K(v)\, dv \right| dt \le \frac{(2\pi)^{2d}\|K\|_{L^1}^2}{\|K\|_{L^2}^2} < \infty,$$

Therefore the conditions of Theorem 14.1 in Piterbarg (1996) are satisfied and the assertion of Theorem 1 follows.

The remaining proof of the uniform approximation (5.3) will be accomplished showing the following auxiliary results. For this purpose we introduce the process

$$Y_{n,1}^{(+)}(x) := \frac{(2\pi)^d h^\beta}{\sqrt{n^d a_n^d h^d}\|K\|_{L^2(\mathbb{R}^d)}} \sum_{\alpha\in\{0,1\}^d} (-1)^{|\alpha|} \sum_{\mathbf{j}\in G_n^{|\alpha|,+}} \Delta_\alpha\left(K_n \circ \tau_x, I_{\mathbf{j}}\right) B(\mathbf{j} : (\mathbf{n})_{\mathbf{1}-\alpha})$$

where the function τ_x is defined by $\tau_x(u) := x - \frac{u+1}{na_nh}$,

$$I_{\mathbf{j}} := \left[(\mathbf{j}-\mathbf{1}) : (\mathbf{n})_{1-\alpha}, \mathbf{j} : (\mathbf{n})_{1-\alpha}\right] \subset \mathbb{R}_+^d \tag{5.6}$$

and we use the notation (5.2).

Lemma 2. *There exists a Brownian sheet B on $\mathbb{R}^d$ such that*

$$\sup_{x\in[0,h^{-1}]^d} |Y_n^{(+)}(x) - Y_{n,1}^{(+)}(x)| = o\Big(\frac{1}{\sqrt{\log(n)}}\Big) \quad a.s.$$

We obtain from the proof in Section 6.1 that Lemma 2 holds under the condition $\frac{\log(n)}{n^{\frac{\delta}{2}} a_n^{\frac{\delta}{2}} h^{\frac{d}{2}}} = o(1)$, which follows from the assumptions of Theorem 1. The next step consists of the replacement of the kernel K_n in the process $Y_{n,1}$ by its limit.

Lemma 3.

$$\sup_{x\in[0,h^{-1}]}\Big|Y_{n,1}(x)-Y_{n,2}(x)\Big| = o_P\left(\frac{1}{\log(n)}\right),$$

where the process $Y_{n,2}$ is given by

$$Y_{n,2}(x) := \frac{(2\pi)^d}{\sqrt{n^d a_n^d h^d}\|K\|_{L^2}} \sum_{\alpha,\gamma\in\{0,1\}^d} (-1)^{|\alpha|} \sum_{\mathbf{j}\in G_n^{|\alpha|,+}} \Delta_\alpha\left(K\circ\tau_x, (-1)^\gamma I_{\mathbf{j}}\right) B((-1)^\gamma \mathbf{j} : (\mathbf{n})_{\mathbf{1}-\alpha}).$$

As described in Section 5.1 for fixed $\alpha\in\{0,1\}^d, j\in G_n^{|\alpha|,+}$ the quantity $\Delta_\alpha\left(K\circ\tau_x; I_{\mathbf{j}}\right)$ can be regarded as the increment of the function $(K_n\circ\tau_x)_\alpha((u)_\alpha) = K_n\circ\tau_x\big((u:n)_\alpha\big)$ on the cube $[(\mathbf{j}-\mathbf{1}),\mathbf{j}]$. This point of view is the basic step in the approximation by the Riemann-Stieltjes Integral of $B(((\cdot):\mathbf{n})_{\mathbf{1}-\alpha})$ with respect to the function $(K_n\circ\tau_x)_\alpha$ for each $\alpha\in\{0,1\}^d$.

Lemma 4.

$$\sup_{x\in[0,h^{-1}]^d}|Y_{n,2}^{(+)}(x) - Y_{n,3}^{(+)}(x)| = o_P\left(\frac{1}{\log(n)}\right),$$

where the process $Y_{n,3}^{(+)}$ is defined by

$$Y_{n,3}^{(+)}(x) \overset{\mathcal{D}}{=} \frac{(2\pi)^d}{\|K\|_{L^2}} \int_{[0,(a_nh)^{-1}]^d} K(x-u)\,dB(u). \tag{5.7}$$

We obtain from its proof in Section 6.2 that Lemma 4 holds under the condition $\frac{\log(n)}{nh^d} = o(1)$, which follows from the assumptions of Theorem 1. In the final step we show that the difference

$$Y^{(+)}(x) - Y_{n,3}^{(+)}(x) = \frac{(2\pi)^d}{\|K\|_{L^2}} \int_{\mathbb{R}_+^d} I_{\mathbb{R}_+^d\setminus[0,(a_nh)^{-1}]^d}(u)\,K(x-u)\,dB(u)$$

is asymptotically negligible.

Lemma 5. $\sup_{x\in[0,h^{-1}]^d}|Y_{n,3}(x)-Y(x)| = o_P((\log(n)^{-1})$.

5.3 Proof of Corollary 1

The assertion follows from the estimate

$$\sup_{[0,1]^d}\Big|f(x)-\mathbb{E}\hat{f}_n(x)\Big| = o\left(h^{-\beta}(h^d n^d a_n^d)^{-1/2}\right). \tag{5.8}$$

To prove (5.8) we use the representation (2.6) and obtain by a straightforward calculation

$$\begin{aligned}\mathbb{E}\hat{f}_n(x) &= \frac{1}{(2\pi)^d n^d a_n^d h^d} \sum_{\mathbf{k}\in G_\mathbf{n}} (f*\psi)(x_\mathbf{k}) \cdot K_n\left((x-x_\mathbf{k})\frac{1}{h}\right) \\ &= \frac{1}{(2\pi)^d h^d} \int_{[-\frac{1}{a_n},\frac{1}{a_n}]^d} (f*\psi)(z) \cdot K_n\left((x-z)\frac{1}{h}\right) dz \, + \, R_{n,1}(x) \\ &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} (f*\psi)(z\cdot h) \cdot K_n\left(\frac{x}{h}-z\right) dz \, + R_{n,1}(x) + R_{n,2}(x),\end{aligned}$$

where the term

$$(5.9)\quad R_{n,1}(x) \; = \; \frac{1}{(2\pi)^d h^d} \sum_{\mathbf{k}\in G_{\mathbf{n}-\mathbf{1}}} \int_{[x_\mathbf{k},x_{\mathbf{k}+\mathbf{1}}]} \Big\{ (f*\psi)(x_\mathbf{k}) K_n\Big(\frac{x-x_\mathbf{k}}{h}\Big) - (f*\psi)(z) K_n\Big(\frac{x-z}{h}\Big)\Big\}\, dz$$

denotes the "error" in the integral approximation and

$$R_{n,2}(x) := \frac{1}{(2\pi)^d h^d} \int_{\left([-\frac{1}{a_n},\frac{1}{a_n}]^d\right)^C} (f*\psi)(z) K_n\Big((x-z)\frac{1}{h}\Big)\, dz.$$

An application of the Plancherel identity (see for example Folland (1984), Theorem 8.29) gives (observing Assumption A1 and A3)

$$\begin{aligned}\mathbb{E}\hat{f}_n(x) &= \frac{1}{(2\pi)^{\frac{d}{2}} h^d} \int_{\mathbb{R}^d} \mathcal{F}f\big(h^{-1}\xi\big) \mathcal{F}\psi\big(h^{-1}\xi\big) \frac{\mathcal{F}\eta(\xi)}{\mathcal{F}\psi(\frac{\xi}{h})} \exp\big(ih^{-1}x^T\xi\big)\; d\xi + R_{n,1}(x) + R_{n,2}(x) \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} \mathcal{F}f(\xi)\cdot\mathcal{F}\eta(\xi h) \exp(ix^T\xi)\, d\xi + R_{n,1}(x) + R_{n,2}(x) \\ &= f(x) + R_{n,1}(x) + R_{n,2}(x) + R_{n,3}(x) + R_{n,4}(x)\end{aligned}$$

where

$$\begin{aligned}R_{n,3}(x) &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{([-\frac{D}{h},\frac{D}{h}]^d)^C} \mathcal{F}f(\xi) \exp(ix\xi)\, d\xi \\ R_{n,4}(x) &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{[-\frac{1}{h},\frac{1}{h}]^d \setminus [-\frac{D}{h},\frac{D}{h}]^d} \mathcal{F}f(\xi)\cdot\mathcal{F}\eta(\xi h) \exp(ix\xi)\, d\xi.\end{aligned}$$

We further obtain from Assumption C

$$\Big| \int_{\{\xi_j > \frac{D}{h}\}} \mathcal{F}f(\xi) \exp(-ix\xi)\, d\xi \Big| \le \frac{1}{D^\gamma} \int_{\{\xi_j > \frac{D}{h}\}} |\mathcal{F}f(\xi)| (h\xi_j)^\gamma\, d\xi = o(h^\gamma),$$

and finally $|R_{n,3}(x)| \le \sum_{j=1}^d \int_{\{\xi_j > \frac{D}{h}\}} |\mathcal{F}f(\xi)|\, d\xi = o(h^\gamma)$. With the same arguments it follows $R_{n,4}(x) = o(h^\gamma)$, since $|\mathcal{F}\eta(\xi h)| \le 1$ for all $\xi \in \mathbb{R}^d$. Define $\mathcal{A}_n = ([-\frac{1}{a_n},\frac{1}{a_n}]^d)^C$, then we obtain from the representation (2.7) the estimate

$$|R_{n,2}(x)| \leq \frac{1}{(2\pi)^d h^{\beta+d}} \Big(\int_{\mathcal{A}_n} |(f*\psi)(z)|^2\,dz\Big)^{\frac{1}{2}} \Big[\Big(\int_{\mathcal{A}_n} |K\Big((x-z)\frac{1}{h}\Big)|^2\,dz\Big)^{\frac{1}{2}} + \Big(\int_{\mathcal{A}_n} |(h^\beta K_n - K)\Big((x-z)\frac{1}{h}\Big)|^2\,dz\Big)^{\frac{1}{2}}\Big]$$

$$= \frac{1}{(2\pi)^d h^{\beta}} \Big(\int_{\mathcal{A}_n} |(f*\psi)(z)|^2\,dz\Big)^{\frac{1}{2}} \Big(O\big(h^d a_n^{\frac{d}{2}}\big) + O\big(h^{\mu_1+d} a_n^{\frac{d}{2}}\big)\Big) \;=\; O\Big(\frac{a_n^{\nu+\frac{d}{2}}}{h^\beta}\Big).$$

uniformly with respect to $x \in [0,1]^d$. Note that by Assumption C we have $f \in \mathcal{W}^{\lfloor m \rfloor}(\mathbb{R}^d)$ and since $m > 2 + \frac{d}{2}$ Sobolev's Embedding Theorem (Folland (1984), Theorem 8.54) implies the existence of a function $\tilde{f} \in C^2(\mathbb{R}^d)$ with $f = \tilde{f}$ almost everywhere. Observing that the convolution function ψ is integrable gives $\partial^\alpha (f*\psi) = (\partial^\alpha f)*\psi \in C(\mathbb{R}^d)$ for all $\alpha \in \{0,1,2\}$ with $|\alpha| \leq 2$ (see for example Folland (1984), Proposition 8.10), which justifies the application of Taylor's Theorem. Straightforward but tedious calculations give for the remaining term (5.9) $R_{n,1}(x) = O(\frac{1}{n^3 a_n^3 h^{\beta+2}}) + O(\frac{a_n^\nu}{n h^\beta})$ uniformly with respect to $x \in [0,1]^d$.

5.4 Proofs of Theorem 2 and Corollary 2

First we will show that the kernel $\check{K}_n$ satisfies conditions B1 and B2, with the kernel $\check{K}$ defined (4.3). If Assumption $\check{B}$ holds we have

$$\int_{\mathbb{R}^d} \Big(\frac{h^\beta}{\mathcal{F}_d \psi(\frac{\xi}{h})} - \Psi(\xi)\Big) \mathcal{F}_{d+1}\check{\eta}(\xi,\tau) \exp(ix^T\xi)\,d\xi$$
$$= \sum_{p=1}^{M-1} h^{\mu_p} \int_{\mathbb{R}^d} \Psi_p(\xi) \mathcal{F}_{d+1}\check{\eta}(\xi,\tau) \exp(ix^T\xi)\,d\xi + h^{\mu_M} \int_{\mathbb{R}^d} \Psi_{M,n}(\xi) \mathcal{F}_{d+1}\check{\eta}(\xi,\tau) \exp(ix^T\xi)\,d\xi,$$

which implies

$$h^\beta \check{K}_n(x,t) - \check{K}(x,t) = \sum_{p=1}^{M-1} h^{\mu_p} \int_{\mathbb{R}^{d+1}} \Psi_p(\xi) \mathcal{F}_{d+1}\check{\eta}(\xi,\tau) \exp(ix^T\xi + it\tau)\,d(\xi,\tau)$$
$$+ h^{\mu_M} \int_{\mathbb{R}^{d+1}} \Psi_{M,n}(\xi) \mathcal{F}_{d+1}\check{\eta}(\xi,\tau) \exp(ix^T\xi + it\tau)\,d(\xi,\tau).$$

A careful inspection of the proofs of Theorem 1 and Corollary 1 shows that the arguments can be transferred to the time-dependent case if the increase of n and $m(n)$ as well as the decrease of $a_n, b_{m(n)}$, h and h_t are balanced as given in the assumptions of the theorem. The details are omitted for the sake of brevity.

6 Proof of auxiliary results

6.1 Proof Lemma 2

Define $S_{\mathbf{k}} := \sum_{\mathbf{j}\in G^+_{\mathbf{k}}} \varepsilon_{\mathbf{j}}$, set $S_{\mathbf{j}} \equiv 0$ if $\min\{j_1,\ldots,j_d\} = 0$ and recall the definition of $Y_n^{(+)}$ and τ_x in (5.4) and before Lemma 2, respectively. In a first step we will replace the errors ε_k by increments given in terms of partial sums $S_{\mathbf{k}-\alpha}$ for $\alpha \in \{0,1\}^d$. To be precise, we use the representation

$$\varepsilon_{\mathbf{k}} = \sum_{\underline{\alpha}\in\{0,1\}^d} (-1)^{|\underline{\alpha}|} S_{(\mathbf{k}-\underline{\alpha})} = \sum_{\underline{\alpha}\in\{0,1\}^d} (-1)^{|\underline{\alpha}|} S_{\big((\mathbf{k}-\mathbf{1}):\mathbf{k}\big)_{\underline{\alpha}}}.$$

A straightforward calculation gives

$$\begin{aligned} Y_n^{(+)}(x) &:= \frac{h^\beta}{\sigma\|K\|_2\sqrt{n^d h^d a_n^d}} \sum_{\mathbf{k}\in G_n^+} K_n\circ\tau_x(\mathbf{k}-\mathbf{1}) \sum_{\alpha\in\{0,1\}^d} (-1)^{|\alpha|} S_{\mathbf{k}-\alpha} \\ &= \frac{h^\beta}{\sigma\|K\|_2\sqrt{n^d h^d a_n^d}} \sum_{\alpha\in\{0,1\}^d} (-1)^{|\alpha|} \sum_{\mathbf{k}\in G_n^+} K_n\circ\tau_x(\mathbf{k}-\mathbf{1}) S_{((\mathbf{k}-\mathbf{1}):\mathbf{k})\alpha} \\ &= \frac{h^\beta}{\sigma\|K\|_2\sqrt{n^d h^d a_n^d}} \Big(\sum_{\alpha\in\{0,1\}^d} (-1)^{|\alpha|} \sum_{\mathbf{k}\in G_n^+} (K_n\circ\tau_x(\mathbf{k}-\mathbf{1}) - K_n\circ\tau_x(((\mathbf{k}-\mathbf{1}):\mathbf{k})_\alpha)) S_{((\mathbf{k}-\mathbf{1}):\mathbf{k})\alpha} \\ &+ \sum_{\alpha\in\{0,1\}^d} (-1)^{|\alpha|} \sum_{\mathbf{k}\in G_n^+} K_n\circ\tau_x(((\mathbf{k}-\mathbf{1}):\mathbf{k})_\alpha) S_{((\mathbf{k}-\mathbf{1}):\mathbf{k})\alpha} \Big). \end{aligned}$$

Now we can make use of Proposition 6 and Proposition 3 of Owen (2005) to rewrite the sums, such that the increments given in terms of partial sums can be expressed by increments given in terms of the kernel K_n. We obtain

$$\begin{aligned} Y_n^{(+)}(x) = \frac{h^\beta}{\sigma\|K\|_2\sqrt{n^d h^d a_n^d}} \Big[& K_n\circ\tau_x(\mathbf{n}) S_{(\mathbf{n})} \\ & + \sum_{\alpha\in\{0,1\}^d} (-1)^{|\alpha|} \sum_{\mathbf{k}\in G_n^+} \sum_{\beta\in\{0,1\}^d\setminus\{\mathbf{0}\}} (-1)^{|\beta|} \Delta_\beta(K_n\circ\tau_x; \mathbf{k}-\mathbf{1},\ ((\mathbf{k}-\mathbf{1}):\mathbf{k})_\alpha) S_{((\mathbf{k}-\mathbf{1}):\mathbf{k})\alpha} \Big], \end{aligned}$$

The quantity $\Delta_\beta(K_n\circ\tau_x; \mathbf{k}-\mathbf{1},\ ((\mathbf{k}-\mathbf{1}):\mathbf{k})_\alpha)$ can only take values different from zero if $\alpha \le \mathbf{1}-\beta$. Note that for $\alpha \le \mathbf{1}-\beta$ the equality $(\mathbf{k})_\beta = ((\mathbf{k}-\mathbf{1}):\mathbf{k})_\alpha)_\beta$ holds which implies that in this case we also have $\big[(\mathbf{k}-\mathbf{1})_\beta, \big(((\mathbf{k}-\mathbf{1}):\mathbf{k})_\alpha\big)_\beta\big] = [(\mathbf{k}-\mathbf{1})_\beta, (\mathbf{k})_\beta]$. We further obtain

$$\begin{aligned} Y_n^{(+)}(x) = \frac{h^\beta}{\sigma\|K\|_2\sqrt{n^d h^d a_n^d}} \Big[& K_n\circ\tau_x(\mathbf{n}) S_{(\mathbf{n})} + \sum_{\beta\in\{0,1\}^d\setminus\{\mathbf{0}\}} (-1)^{|\beta|} \sum_{\mathbf{k}\in G_{\mathbf{n}}^+} \sum_{\tilde{\alpha}\in\{0,1\}^{d-|\beta|}} (-1)^{|\tilde{\alpha}|} \\ & \times \Delta_\beta\big(K_n\circ\tau_x; \mathbf{k}-\mathbf{1}\,,\ (\mathbf{k})_\beta : ((\mathbf{k}-\mathbf{1})_{\mathbf{1}-\beta} : (\mathbf{k})_{\mathbf{1}-\beta})_{\tilde{\alpha}}\big) S_{(\mathbf{k})_\beta : ((\mathbf{k}-\mathbf{1})_{\mathbf{1}-\beta} : (\mathbf{k})_{\mathbf{1}-\beta})_{\tilde{\alpha}}} \Big]. \end{aligned}$$

The alternating sum with respect to the index $\tilde{\alpha}$ can be written as an increment Δ as defined in (5.1) which then defines a telescope sum according to Owen (2005), Proposition 2. Taking into account that $S(\mathbf{k}) \equiv 0$ if $k_j = 0$ for at least one $j \in \{1, \ldots, d\}$ gives

$$Y_n^{(+)}(x) = \frac{h^\beta}{\sigma \|K\|_2 \sqrt{n^d h^d a_n^d}} \sum_{\beta \in \{0,1\}^d} (-1)^{|\beta|} \sum_{\mathbf{j} \in G_{\mathbf{n}}^{|\beta|,+}} \Delta_\beta \left(K_n \circ \tau_x; I_{\mathbf{j}}\right) \cdot S_{\mathbf{j}:(\mathbf{n})_{\mathbf{1}-\beta}}.$$

With the definitions $X(A) := \sum_{\mathbf{k} \in A \subset \mathbb{Z}^d} X_{\mathbf{k}}$ for any subset $A \in \mathbb{Z}^d$, we can rewrite these partial sums as set-indexed partial sums with index class $n \cdot \mathscr{S}$, where $\mathscr{S} := \{(0, \gamma] \mid 0 < \gamma_j \leq 1,\ 1 \leq j \leq d\}$ and $n \cdot \mathscr{S} := \{n \cdot S \,|\, S \in \mathscr{S}\}$. It follows directly that $\mathscr{S}$ is a sufficiently smooth VC-class of sets, which justifies the application of Theorem 1 in Rio (1993). Therefore there exists a version of a Brownian sheet on $[0, \infty)^d$, say B_1, such that

$$\sup_{\mathbf{k} \in G_{\mathbf{n}}^+} \left| \frac{S_{\mathbf{k}}}{\sigma} - B_1(\mathbf{k}) \right| = O\left((\log(n))^{\frac{1}{2}} n^{\frac{d-\delta}{2}} \right) \ a.s. \tag{6.1}$$

Recalling the definition of $I_{\mathbf{j}}$ in (5.6) we further obtain

$$Y_n^{(+)}(x) - Y_{n,1}^{(+)}(x) = \frac{h^\beta}{\|K\|_2 \sqrt{n^d h^d a_n^d}} \sum_{\beta \in \{0,1\}^d} (-1)^{|\beta|}$$
$$\sum_{\mathbf{j} \in G_{\mathbf{n}}^{|\beta|,+}} \Delta_\beta \left(K_n \circ \tau_x; I_{\mathbf{j}}\right) \cdot \left(\frac{1}{\sigma} S_{\mathbf{j}:(\mathbf{n})_{\mathbf{1}-\beta}} - B_1\big(\mathbf{j} : (\mathbf{n})_{\mathbf{1}-\beta}\big) \right).$$

The estimate (6.1) implies the existence of a constant $C \in \mathbb{R}_+$ such that

$$|Y_n^{(+)}(x) - Y_{n,1}^{(+)}(x)| \leq C \cdot \sqrt{\frac{\log(n)}{n^\delta h^\delta a_n^\delta}} h^\beta \Bigg[\sum_{\gamma \in \{0,1\}^d,\, |\gamma|=1} \int_{[0,(a_n h)^{-1}]^d} (u)_\gamma^{\frac{d-\delta}{2}} |\partial^{\mathbf{1}} K_n(x-u)| \, du$$
$$+ \sum_{\beta \in \{0,1\}^d \setminus \{\mathbf{0},\mathbf{1}\}} \int_{[0,(a_n h)^{-1}]^{|\beta|}} \left| \partial^\beta K_n \left(\left(x - (u : (a_n h)^{-1} \mathbf{1}) \right)_\beta \right) \right| (du)_\beta$$
$$+ |K_n(x - (a_n h)^{-1} \mathbf{1})| \Bigg] \quad \text{a.s.}$$

It follows from Assumption B that the function $u \mapsto (u)_\gamma^{\frac{|\alpha|}{2}} \partial^\alpha K(u)$ is integrable on $\mathbb{R}^d$ for all $\alpha \in \{0,1\}^d$ such that

$$\int_{[0,(a_n h)^{-1}]^d} (u)_\gamma^{\frac{d-\delta}{2}} |\partial^{\mathbf{1}} K_n(x-u)| \, du = O(h^{\frac{\delta-d}{2} - \beta})$$

and

$$\int_{[0,(a_nh)^{-1}]^{|\beta|}} \left|\partial^\beta K_n\left(\left(x-(u:(a_nh)^{-1}\mathbf{1})\right)_\beta\right)\right|(du)_\beta + |K_n(x-(a_nh)^{-1}\mathbf{1})| = O((a_nh)^{\frac{d}{2}}h^{-\beta}).$$

Note that for sufficiently large n such that $a_n < \frac{1}{2}$ we obtain $-\frac{1}{2a_nh} \geq x_j - (a_nh)^{-1} = \frac{a_n-1}{a_nh}$ uniformly with respect to j (note that $x_j \in [0, h^{-1}]$). Let $\tilde{B}$ be a continuous version of B_1. We set $\tilde{B}(t) \equiv 0$ if $t_j < 0$ for at least one index $j \in \{1,\ldots,d\}$ and let $\{\tilde{B}_\alpha \,|\, \alpha \in \{0,1\}^d\}$ be 2^d mutually independent copies of $\tilde{B}$. For $t \in \mathbb{R}^d$ define

$$B_\alpha(t) := \tilde{B}_\alpha((-1)^{\alpha_1}t_1, (-1)^{\alpha_2}t_2, \ldots, (-1)^{\alpha_d}t_d),$$

then the process $\{B(t) := \sum_{\alpha\in\{0,1\}^d} B_\alpha(t) \,|\, t \in \mathbb{R}^d\}$ is a Wiener field on $\mathbb{R}^d$.

6.2 Proof of Lemma 4

Note that $\partial^\alpha K$ exists and is integrable for each $\alpha \in \{0,1\}^d$. Consequently, the kernel K is of bounded variation on $[0,(a_nh)^{-1}]^d$ in the sense of Hardy Krause for each fixed n (see Owen (2005), Definition 2). Therefore an application of integration by parts for the Wiener integral (note that the kernel K has not necessarily a compact support) and rescaling of the Brownian sheet $Y_{n,3}^{(+)}$ yields

$$\begin{aligned}
Y_{n,3}^{(+)}(x) &\stackrel{\mathcal{D}}{=} \sum_{\alpha\in\{0,1\}^d\setminus\{\mathbf{0}\}} (-1)^{|\alpha|} \int_{[0,(a_nh)^{-1}]^{|\alpha|}} B\left(\left(u:(a_nh)^{-1}\mathbf{1}\right)_\alpha\right) dK\left(x-\left(u:(a_nh)^{-1}\mathbf{1}\right)_\alpha\right) \\
&\quad + \Delta\left(K(x-\cdot)\cdot B(\cdot), [0,(a_nh)^{-1}]^d\right) \\
&= \sum_{\alpha\in\{0,1\}^d\setminus\{\mathbf{0}\}} (-1)^{|\alpha|} \int_{[0,(a_nh)^{-1}]^{|\alpha|}} B\left(\left(u:(a_nh)^{-1}\mathbf{1}\right)_\alpha\right) \partial^\alpha K\left(x-\left(u:(a_nh)^{-1}\mathbf{1}\right)_\alpha\right)(du)_\alpha \\
&\quad + \Delta\left(K(x-\cdot)\cdot B(\cdot), [0,(a_nh)^{-1}]^d\right).
\end{aligned}$$

Recalling the definition of $Y_{n,2}^{(+)}(x)$ and identity (15) in Owen (2005) we can replace the increments by the corresponding integrals, that is

$$\begin{aligned}
Y_{n,2}^{(+)}(x) &\stackrel{\mathcal{D}}{=} \sum_{\alpha\in\{0,1\}^d\setminus\{\mathbf{0}\}} (-1)^{|\alpha|} \sum_{\mathbf{k}\in G_{\mathbf{n-1}}^{|\alpha|,+}} \int_{[(na_nh)^{-1}(\mathbf{k-1})_\alpha,(na_nh)^{-1}(\mathbf{k})_\alpha]} \partial^\alpha K\left(x-\left(u:(a_nh)^{-1}\mathbf{1}\right)_\alpha\right)(du)_\alpha \\
&\qquad \times B\left(\left((na_nh)^{-1}\mathbf{k}:(a_nh)^{-1}\mathbf{1}\right)_\alpha\right) + \Delta\left(K(x-\cdot)\cdot B(\cdot), [0,(a_nh)^{-1}]^d\right) \\
&= Y_{n,3}^{+}(x) + R_{n,SI}(x),
\end{aligned}$$

where the remainder $R_{n,SI}(x)$ is defined in an obvious manner. From the modulus of continuity for the Brownian Sheet (see Khoshnevisan (2002), Theorem 3.2.1) it follows that for $a, b \in \mathbb{R}^d$

$$\limsup_{\delta \to 0+} \sup_{s,t \in [a,b], \|s-t\|_\infty < \delta} \frac{|B(s) - B(t)|}{\sqrt{\delta \log(\frac{1}{\delta})}} \leq 24 \cdot d \|b\|_\infty^{d/2}, \tag{6.2}$$

which yields

$$\begin{aligned} |Y_{n,2}^{(+)}(x) - Y_{n,3}^{(+)}(x)| = |R_{n,SI}(x)| &\leq \sup_{\delta < \frac{1}{n}} \sup_{s,t \in [0,2]^d : \|s-t\|_\infty \leq \delta} |B(s) - B(t)| \\ &\times \sqrt{\frac{\log(n)}{n}} \Big[\int_{[0,(a-nh)^{-1}]^d} \big((u)^{\mathbf{1}}\big)^{\frac{1}{2}} |\partial^{\mathbf{1}} K(x-u)| \, du \, + \, O(h^{-\frac{d-1}{2}}) \Big] \end{aligned}$$

(note that the dominating term in $R_{n,SI}(x)$ is given by the summand where $|\alpha| = d$). With the same arguments as in the proof of Lemma 2 we finally obtain

$$|Y_{n,2}^{(+)}(x) - Y_{n,3}^{(+)}(x)| = O_P\left(\sqrt{\frac{\ln(na_n h)}{nh^d}}\right),$$

where we used the estimate (6.2) for the modulus of continuity of the Brownian sheet (note that this estimate is independent of x).

6.3 Proof of Lemma 5

Integration by parts gives

$$\begin{aligned} \Delta_{n,3} &:= |Y_{n,3}^{(+)}(x) - Y^{(+)}(x)| \\ &\leq \Big| \int_{[0,\infty)^d \setminus [0, \frac{1}{a_n h}]^d} B(u) \partial^{\mathbf{1}} K(x-u) \, du \Big| \\ &+ \Big| \sum_{\alpha \in \{0,1\}^d \setminus \{\mathbf{0},\mathbf{1}\}} (-1)^{|\alpha|} \int_{[0, \frac{1}{a_n h}]^{|\alpha|}} B\left(\left(u : (a_n h)^{-1} \mathbf{1}\right)_\alpha \right) \partial^\alpha K \left(x - \left(u : (a_n h)^{-1} \mathbf{1}\right)_\alpha \right) \, (du)_\alpha \Big| \\ &+ \left| \Delta \left(K(x - \cdot) B(\cdot); [0, (a_n h)^{-1}]^d \right) \right| := |\Delta_{n,3}^{(1)}(x)| + |\Delta_{n,3}^{(2)}(x)| + |\Delta_{n,3}^{(3)}(x)|, \end{aligned} \tag{6.3}$$

where the processes $\Delta_{n,3}^{(j)}(x)$, $j = 1, 2, 3$ are defined in an obvious manner. Let n be sufficiently large such that $\frac{1}{a_n h} \geq 1$ and $a_n < \frac{1}{2}$. Since $B(u) = 0$ if $t_j = 0$ for at least one index $j \in \{1, \dots, d\}$

we have

$$\begin{aligned}|\Delta^{(3)}_{n,3}(x)| &= \left|K\left(x-(a_nh)^{-1}\mathbf{1}\right)\cdot B\left((a_nh)^{-1}\mathbf{1}\right)\right| \\ &= \sqrt{2d\,(a_nh)^{-d}\ln(d\ln((a_nh)^{-1}))}\;\frac{|K\left(x-(a_nh)^{-1}\mathbf{1}\right)|\,|B\left((a_nh)^{-1}\mathbf{1}\right)|}{\sqrt{2d\,(a_nh)^{-d}\ln(d\ln((a_nh)^{-1}))}}.\end{aligned}$$

An application of the version of a law of the iterated logarithm given in Theorem 3 of Paranjape and Park (1973) yields the estimate

$$\begin{aligned}\sup_{x\in[0,h^{-1}]}|\Delta^{(3)}_{n,3}(x)| &= O(1)\cdot\sqrt{2d\,(a_nh)^{-d}\ln(d\ln((a_nh)^{-1}))}\sup_{x\in[0,h^{-1}]}\left|K\left(x-(a_nh)^{-1}\mathbf{1}\right)\right| \\ &\le O(1)\cdot\sqrt{2d\,(a_nh)^{-d}\ln(d\ln((a_nh)^{-1}))}\sup_{v\le\frac{a_n-1}{a_nh}}|K(v)| = o\left(\frac{1}{\log(n)}\right)\quad\text{a.s.}\end{aligned}$$

uniformly with respect to x.

To show that $\Delta^{(2)}_{n,3}(x)$ and $\Delta^{(1)}_{n,3}(x)$ are asymptotically negligible we also apply the LIL for the Brownian sheet. For each summand, say $\Delta^{(2)}_{n,3,\alpha}$, in $\Delta^{(2)}_{n,3}(x)$ ($|\alpha|<d$) we have

$$\begin{aligned}\Delta^{(2)}_{n,3,\alpha}(x) &:= \left|\int_{[0,\frac{1}{a_nh}]^{|\alpha|}} B\left(\left(u:(a_nh)^{-1}\mathbf{1}\right)_\alpha\right)\partial^\alpha K\left(x-\left(u:(a_nh)^{-1}\mathbf{1}\right)_\alpha\right)\,(du)_\alpha\right| \\ &= (a_nh)^{-|\alpha|}\left|\int_{[0,1]^{|\alpha|}} B\left(\left(u:\mathbf{1}\right)_\alpha(a_nh)^{-1}\right)\partial^\alpha K\left(x-\left(u:\mathbf{1}\right)_\alpha(a_nh)^{-1}\right)\,(du)_\alpha\right|.\end{aligned}$$

Scaling of the Brownian sheet yields

$$\begin{aligned}\Delta^{(2)}_{n,3,\alpha}(x) &\overset{\mathcal{D}}{=} (a_nh)^{-\frac{2|\alpha|+d}{2}}\left|\int_{[0,1]^{|\alpha|}} B\left(\left(u:\mathbf{1}\right)_\alpha\right)\partial^\alpha K\left(x-\left(u:\mathbf{1}\right)_\alpha(a_nh)^{-1}\right)\,(du)_\alpha\right| \\ &= O\left((a_nh)^{-\frac{d}{2}}\right)\left|\int_{[0,\frac{1}{a_nh}]^{|\alpha|}}\partial^\alpha K\left(x-\left(u:\mathbf{1}\right)_\alpha(a_nh)^{-1}\right)\,(du)_\alpha\right|\quad\text{a.s.}\end{aligned}$$

With the same arguments as in the proof of Lemma 2 we conclude that the leading contributions are given by the quantities $\Delta^{(2)}_{n,3,\alpha}(x)$, where $|\alpha|=d-1$. For $\alpha=(0,1,\dots,1)$ obtain

$$\sup_{x\in[0,h^{-1}]^d}\left|\Delta^{(2)}_{n,3,\alpha}(x)\right| = O_P\left((a_nh)^{-\frac{d}{2}}\right)\sup_{v\le-\frac{1}{2a_nh}}\int_{\mathbb{R}^{d-1}}\left|\partial^\alpha K\left(v,u_2,\dots,u_d\right)\right|\,d(u_2,\dots,u_d).$$

This gives $\sup_{x\in[0,h^{-1}]^d}\left|\Delta^{(2)}_{n,3,(0,1,\dots,1)}(x)\right| = o(\frac{1}{\log(n)})$. Applying the same argument to the other terms yields $\Delta^{(2)}_{n,3}(x)=o_P(\frac{1}{\log(n)})$ uniformly with respect to $x\in[0,1/h]^d$. Finally, a similar argument gives for the remaining term in (6.3) $\Delta^{(1)}_{n,3}(x)=o_P(\frac{1}{\log(n)})$, which completes the proof of Lemma 5.

6.4 Proof of Lemma 3

Note that we have

$$Y_{n,1}^{(+)}(x) \overset{\mathcal{D}}{=} \frac{h^{\beta}}{\sqrt{n^d a_n^d h^d}\|K\|_{L^2(\mathbb{R}^d)}} \sum_{\alpha \in \{0,1\}^d} (-1)^{|\alpha|} \sum_{\mathbf{j} \in G_n^{|\alpha|,+}} \Delta_\alpha(\overline{\mathcal{F} f_p} \circ \tau_x; I_{\mathbf{j}}) B(\mathbf{j} : (\mathbf{n})_{\mathbf{1}-\alpha}).$$

The representation (2.7) and the definition (5.6) yield

$$\begin{aligned}
&|Y_{n,1}^{(+)}(x) - Y_{n,2}^{+}(x)| \\
&= \sum_{p=1}^{M-1} \frac{h^{\mu_p}}{\sqrt{n^d a_n^d h^d}} \Big| \sum_{\alpha \in \{0,1\}^d} (-1)^{|\alpha|} \sum_{\mathbf{j} \in G_{\mathbf{n}}^{|\alpha|,+}} \Delta_\alpha(\overline{\mathcal{F} f_p} \circ \tau_x; I_{\mathbf{j}}) B(\mathbf{j} : (\mathbf{n})_{\mathbf{1}-\alpha}) \Big| + o_P\left(\frac{1}{\log(n)}\right).
\end{aligned}$$

For each fixed p we can now perform the approximation steps of the previous Lemmas and obtain

$$\begin{aligned}
\log(n) \sup_{x \in [0,h^{-1}]^d} \Big| &\frac{1}{\sqrt{n^d a_n^d h^d}} \sum_{\alpha \in \{0,1\}^d} (-1)^{|\alpha|} \sum_{\mathbf{j} \in G_{\mathbf{n}}^{|\alpha|,+}} \Delta_\alpha(\overline{\mathcal{F} f_p} \circ \tau_x; I_{\mathbf{j}}) B(\mathbf{j} : (\mathbf{n})_{\mathbf{1}-\alpha}) \\
&- \int_{\mathbb{R}_+^d} \overline{\mathcal{F} f_p}(x-u)\, dB(u) \Big| = o_P(1).
\end{aligned}$$

It can easily be shown that for all $p = 1, \ldots, M-1$

$$\lim_{n \to \infty} n^d a_n^d h^d \operatorname{Var}\Big(\frac{1}{n^d a_n^d h^d} \sum_{\mathbf{k} \in G_{\mathbf{n}}} Y_{\mathbf{k}} \overline{\mathcal{F} f_p}\Big((x - x_{\mathbf{k}}) \frac{1}{h} \Big) \Big) = \sigma^2 \|f_p\|_2^2,$$

where the limit does not depend on x. We finally obtain, repeating the approximation steps given in the previous Lemmas for each of the $2^d - 1$ remaining orthants

$$\begin{aligned}
\log(n) \sup_{x \in [0,h^{-1}]^d} \Big| &\frac{1}{\sqrt{n^d a_n^d h^d}} \sum_{\alpha, \gamma \in \{0,1\}^d} (-1)^{|\alpha|} \sum_{\mathbf{j} \in G_{\mathbf{n}}^{|\alpha|,+}} \Delta_\alpha(\overline{\mathcal{F} f_p} \circ \tau_x; (-1)^{\gamma} I_{\mathbf{j}}) B((-1)^{\gamma} \mathbf{j} : (\mathbf{n})_{\mathbf{1}-\alpha}) \\
&- \int_{\mathbb{R}^d} \overline{\mathcal{F} f_p}(x-u)\, dB(u) \Big| = o_P(1).
\end{aligned}$$

Note that

$$r(x-z) := \mathbb{E}\Big(\int_{\mathbb{R}^d} \overline{\mathcal{F} f_p}(x-u)\, dB(u) \int_{\mathbb{R}^d} \mathcal{F} f_p(z-u)\, dB(u) \Big) = \int_{\mathbb{R}^d} f_p(x-z+u) f_p(u)\, du$$

and $\|r\|_1 \leq \|f_p\|^2 < \infty$. The system of sets $\{[0,h^{-1}]^d \,|\, n \in \mathbb{N}\}$ is a blowing up system of sets in the sense of definition 14.1 in Piterbarg (1996). If we define

$$Z_p(x) = \frac{1}{\|f_p\|_{L^2(\mathbb{R}^d)}} \int_{\mathbb{R}^d} \overline{\mathcal{F} f_p}(x-u)\, dB(u),$$

then Theorem 1 in Bickel and Rosenblatt (1973a) gives the asymptotic independence of the scaled minimum and maximum of the process Z_p, which, with the observation that Z_p and $-Z_p$ have the same distribution and an application of Theorem 14.1 in Piterbarg (1996) yields that for $G \sim$ Gumbel$(\ln(2), 1)$

$$\sup_{x \in [0,h^{-1}]^d} \left(\left(|Z_p(x)| - \tilde{C}_{n,3} \right) \tilde{C}_{n,3} \right) \xrightarrow{\mathcal{D}} G \quad \text{for} \quad n \to \infty,$$

where the constants $\tilde{C}_1, \tilde{C}_{n,2}$ and $\tilde{C}_{n,3}$ are given by

$$\tilde{C}_1 = \mathbf{det}\left(\left[\frac{1}{\|f_p\|^2_{L^2(\mathbb{R}^d)}} \int_{\mathbb{R}^d} |f_p(v)|^2 v_i v_j \, dv \right],\, i,j = 1,\dots,d \right),$$

$$\tilde{C}_{n,2} = \sqrt{\frac{\tilde{C}_1}{(2\pi)^{d+1}} \frac{1}{h^d}}$$

$$\tilde{C}_{n,3} = \sqrt{2\ln(\tilde{C}_{n,2})} + \frac{(d-1)\ln\big(2\ln(\tilde{C}_{n,2})\big)}{2\sqrt{2\ln(\tilde{C}_{n,2})}}.$$

Since $h^{\mu_p} = o\left(\frac{1}{\log n}\right)$ we obtain $h^{\mu_p} \sup_{x \in [0,h^{-1}]} |Z_p(x)| = o_P((\log n)^{-1/2})$ for each $p = 1, \dots, M-1$, which justifies the replacement of $h^\beta K_n$ by K. Since the outer sum does not depend on n this gives the desired result.

6.5 Proof of Lemma 1

With the same arguments as in the proof of the previous Lemmas we can replace the errors by combinations of partial sums and perform the same approximation steps. In each replacement we obtain at most a $d-1$-fold sum which yields the desired result right away.