Confidence regions for images observed under the Radon transform

Nicolai Bissantz, Hajo Holzmann,
Katharina Proksch

Confidence regions for images observed under the Radon transform

Nicolai Bissantz[1], Hajo Holzmann[2] and Katharina Proksch[1]

[1]Fakultät für Mathematik
Ruhr-Universität Bochum, Germany

[2]Fachbereich Mathematik und Informatik
Philipps-Universität Marburg, Germany

November 14, 2013

Abstract

Recovering a function f from its integrals over hyperplanes (or line integrals in the two-dimensional case), that is, recovering f from the Radon transform Rf of f, is a basic problem with important applications in medical imaging such as computerized tomography (CT). In the presence of stochastic noise in the observed function Rf, we shall construct asymptotic uniform confidence regions for the function f of interest, which allows to draw conclusions regarding global features of f. Specifically, in a white noise model as well as a fixed-design regression model, we prove a Bickel-Rosenblatt-type theorem for the maximal deviation of a kernel-type estimator from its mean, and give uniform estimates for the bias for f in a Sobolev smoothness class. The finite sample properties of the proposed methods are investigated in a simulation study.

Key words: Confidence bands, Inverse problems, Nonparametric Regression, Radon Transform

AMS subject classification: Primary 62G15; Secondary 62G08, 65R10.

1 Introduction

Often, we would like to draw conclusions on the internal structure of a certain object but there is no possibility to take a direct look, say, by invasive means. Classical examples where this is

---

[2]Address for correspondence: Dr. Hajo Holzmann, Philipps-Universität Marburg, Fachbereich Mathematik und Informatik, Hans-Meerwein-Straße. D-35032 Marburg, Germany, Email: holzmann@mathematik.uni-marburg.de, Fon: + 49 6421 2825454

the case are the interior of the earth or features hidden inside a human body. In order to reveal parts of the respective inner structure under investigation one can often resort to tomography methods. Tomography is a collective term for noninvasive imaging methods that allow the reconstruction of the inner structure of the object of interest by cross-sections. In such an example only indirect observations (from the outside) are available for the reconstruction of the the invisible quantity (inside).

In general, problems of this nature are called inverse problems. Mathematically both the observed quantity and the quantity to be reconstructed are modelled as elements (functions) of suitable Hilbert spaces, $g \in Y$ and $f \in X$, respectively. Often, the connection between these functions can be modelled via some bounded linear operator $T : X \to Y$, i.e. $g = Tf$. Mathematically, the resulting inverse problem can be formulated as: Given $g \in Y$ find $f \in X$ such that $Tf = g$. Typically, the corresponding spaces are of infinite dimension which leads to ill-posed problems in the sense that $T(X) \neq \overline{T(X)}$ and hence, even if T is injective and bounded its inverse T^{-1} does not need to be bounded. In case of practical applications where one can never assume to observe the whole function g without any noise/ errors this causes severe problems in the reconstruction. Instead of the actual inverse T^{-1}, a regularized version thereof has to be used. In many examples of tomography the corresponding operator T is the Radon transform and the spaces under consideration are suitable Sobolev spaces.

An overview over inverse problems from a numerical, deterministic viewpoint and existing regularization methods can be found in the monograph Engl et al. (1996). From a statistical viewpoint, where errors are modelled as random quantities, the respective methods have to be revisited. Mair and Ruymgaart (1996); Kaipio and Somersalo (2005); Bissantz et al. (2007b) or Cavalier (2008) among others focus on the statistical modelling of inverse problems. Statistical inverse problems related to tomography have been studied in the last decades, where the main focus has been on positron emission tomograpy (PET) and compterized tomography (CT) in medical imaging which are both related to the Radon transform.

In PET we are given lines along which emissions have occured but the precise position on the lines is unknown and is reconstructed in order to obtain the emission distribution. This problem is related to nonparametric density estimation and is discussed in detail in, e.g., Johnstone and Silverman (1990); Korostelev and Tsybakov (1993); Cavalier (2000).

The example of CT leads to an inverse regression model. Here, from fixed positions thin beams of X-rays of known intensity are sent through the object of interest. The decrease of the intensity of the X-rays along several lines is measured from which, due to proportionality, the mass density of the object of interest can be reconstructed. Hence, the given data in this regression problem are integral values of the mass density along certain lines which are preselected and given by the design. Related problems have already been studied in Cavalier (1999); Kerkyacharian et al. (2010, 2012). The application related to the statistical models

regarded in this paper is also the problem of CT. An overview of the mathematical aspects of this particular method of medical imaging can be found in the monographs Natterer (1986) or Helgason (2011).

In this paper we will consider $T = R$ the Radon transform, which may be considered as a bounded operator between certain Sobolev spaces, see Natterer (1980). In our model we assume that the image Rf of f under R is observed with random noise. We then aim to construct uniform confidence regions for the function of interest f based on kernel-type estimators, which are closely related to the popular filtered backprojection algorithm.

To this end, we will first review kernel-type estimators for the function f from the literature and discuss their properties for flexible choices of the kernel. Mere estimation is usually only the first step in data analysis, further steps being statistical inference via goodness-of-fit tests or confidence regions. From the pointwise asymptotic distribution of an estimator and the resulting pointwise asymptotic confidence regions, statements regarding the function of interest at fixed points may be validated. For CT such a statement could be that the mass density of the object of interest at a fixed point does not exceed a certain threshold with high probability. In many cases, conclusions regarding global features of a curve, such as overall curvature or shape of the underlying function, are of particular interest. In the example of CT given above such a global statement could be that, with high probability, the mass density of an object of interest does not exceed a certain threshold at many points or a complete interval, rectangle or cube simultaneously. Here, the employment of confidence intervals is not sufficient without any additional considerations. A very common approach is based on the asymptotic distribution of the maximal deviation of the estimator and the function of interest which requires results from extreme value theory for stationary Gaussian processes. It was first introduced by Smirnov (1950) who constructed uniform confidence bands for the histogram estimate of a density and by Bickel and Rosenblatt (1973b) who derived the respective limit theorem for general kernel density estimates. Based on this approach simultaneous confidence regions have been constructed in many different problems of density estimation and nonparametric regression in the direct case Hall (1992); Eubank and Speckman (1993); Xia (1998); Claeskens and van Keilegom (2003); Giné and Nickl (2010), or Bissantz et al. (2007a); Birke et al. (2010); Lounici and Nickl (2011); Schmidt-Hieber et al. (2013) in density deconvolution and inverse regression. Neumann and Polzehl (1998) derived bootstrap-based uniform confidence bands without an application of such a limit theorem by directly linking the bootstrap statistic to the one based on the data.

However, all these results are for univariate problems. The construction of multivariate confidence regions has received much less attention in the literature, see Rosenblatt (1976); Konakov and Piterbarg (1984); Rio (1994). However, in many cases, multivariate problems arise naturally such as in the analysis of astronomical or biological images taken with telescopes or microscopes, respectively, that also involves deconvolution (Proksch et al. (2012)). Also

in the problem of reconstructing a function from noisy observations of its Radon transform discussed in this paper we have to deal with a two-dimensional problem at least.

The paper is structured as follows. In Section 2 the mathematical preliminaries will be discussed and the reconstruction problem will be formulated as an example of a multivariate inverse regression problem. Both an idealized Gaussian white noise model as well as a more commonly used discrete counterpart with fixed-design are discussed in Sections 2.2 and 4, respectively and kernel-based estimators will be proposed for all cases. Section 3.3 contains the limit theorems that will be used in order to construct the confidence bands and in Section 5 the finite sample properties of the estimator and the proposed methods will be investigated in a small simulation study. Finally, all proofs are given in Section 6. In the following, let $\|x\|_\infty = \max_{1 \le i \le N} |x_i|$ and $\|x\|$ be the usual Euclidean norm on $\mathbb{R}^N$.

2 The Radon transform and the white noise model

2.1 Radon transform

For $N \ge 2$, the N-dimensional Radon transform R is an integral operator that maps a real valued function f on $\mathbb{R}^N$ into the set of its integrals over the hyperplanes of $\mathbb{R}^N$. To be precise, for $f : \mathbb{R}^N \to \mathbb{R}$, $f \in L_1(\mathbb{R}^N)$ the Radon transform is defined by

$$Rf(s, u) = \int_{H(s,u)} f(v) \, dv,$$

where $u \in \mathbb{R}$, $s \in \mathbb{S}^{N-1}$, $\mathbb{S}^{N-1} = \{v \in \mathbb{R}^N \,|\, \|v\|_2 = 1\}$ is the unit sphere in $\mathbb{R}^N$, $H(s, u) := \{v \,|\, <v, s> = u\}$ and integration is with respect to the Lebesgue measure on the hyperplane. The function Rf is defined on the cylinder $\mathcal{Z} := \mathbb{S}^{N-1} \times \mathbb{R}$. Note that, for $f \in L^1(\mathbb{R}^N)$, the Radon transform exists for almost all $(s, u) \in \mathcal{Z}$ and the map $f \mapsto Rf$ is injective on $L^1(\mathbb{R}^N)$ (see, e.g., Helgason (2011), Proposition 3.4).

In spherical coordinates we have $s = s(\vartheta), \vartheta \in U \subset \mathbb{R}^{N-1}$ and $H(s, u) = H(s(\vartheta), u) = H(\vartheta, u)$. In the case $N = 2$ the parametrization of the hyperplane, i.e. line, H is illustrated in Figure 1.

2.2 Observations in the Gaussian white noise model

We shall derive our results first in an idealized white noise model, which we introduce below. The main advantage of such an idealized model is that unnecessary technicalities can be avoided, while the results obtained in the idealized setting also hold true under more realistic model assumptions. For further discussion and a discretized version of model (1) we refer to Section 4.

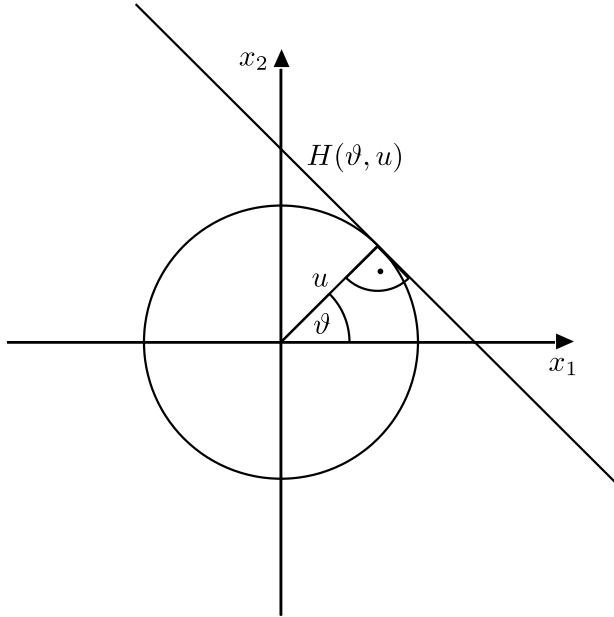To introduce the model we need to fix some notation first. Consider the σ-finite measure ν

Figure 1: Parametrization of the line $H = H(\vartheta, u)$ in the case $N = 2$.

on $\mathcal{B}(\mathcal{Z})$ defined by

$$A \mapsto \nu(A) = \int_{\mathbb{S}^{N-1}} \int_{\mathbb{R}} 1_A(s, u) \, du \, ds,$$

where ds is the common surface-measure on $\mathbb{S}^{N-1}$, such that $|A| = \int_A ds$ for $A \in \mathbb{S}^{N-1}$. We denote $\rho_N = |\mathbb{S}^{N-1}|$. Define

$$\left(\mathcal{B}(\mathcal{Z})\right)_\nu := \left\{ A \in \mathcal{B}(\mathcal{Z}) \,\middle|\, \nu(A) < \infty \right\}$$

and let W be Gaussian noise on $\mathcal{Z}$ based on ν, i.e. a Gaussian random set function such that for $A, B \in \left(\mathcal{B}(\mathcal{Z})\right)_\nu$, $A \cap B = \emptyset$ we have $W(A \cup B) = W(A) + W(B)$ a.s., and

$$W(A) \sim \mathcal{N}\left(0, \nu(A)\right),$$

and $W(A)$ and $W(B)$ are independent (see, e.g., Adler and Taylor (2007), Chapter 1.4.3 for details). Now, we consider the Gaussian white noise model

$$dY(s, u) = Rf(s, u) \, ds \, du + \epsilon dW(s, u), \qquad (s, u) \in \mathcal{Z}. \tag{1}$$

Here, $\epsilon > 0$ is a small parameter, representing the noise level of the observations. The meaning of equation (1) is that for every $A \in \left(\mathcal{B}(\mathcal{Z})\right)_\nu$ we observe the Gaussian set function Y on $\mathcal{Z}$

at A, i.e. $Y(A)$, where

$$Y(A) \sim \mathcal{N}\Big(\int_{\mathbb{S}^{N-1}} \int_{\mathbb{R}} Rf(s,u) I_{(s,u) \in A} \, du \, ds, \, \epsilon^2 \nu(A)\Big).$$

For $f \in L^2(\mathcal{Z})$ the process $W(f) = \int f(s,u) \, dW(s,u)$ is a centered Gaussian field with

$$\mathbb{E}\big[W(f) \cdot W(g)\big] = \int_{\mathbb{S}^{N-1}} \int_{\mathbb{R}} f(s,u) g(s,u) \, du \, ds = \langle f, g \rangle_{du \times ds}.$$

The goal is to recover f from the observation of Y, i.e. from indirect observations corrupted with random noise.

3 Estimation and uniform confidence regions

3.1 Derivation of the estimator

In this section we will derive suitable estimators for the regression function in the Gaussian white noise model (1). Our approach is based on an explicit inversion formula obtained by the application of results from Fourier analysis as it is given, e.g., in Natterer (1986) or Helgason (2011) which naturally leads to a nonparametric kernel-type estimator. The idea is similar to that of filtered backprojection, a reconstruction algorithm known from the numerical literature. This idea was also adopted by, e.g., Cavalier (1999, 2000) in a statistical framework. In the following we will give a heuristic derivation of the estimator. To this end let $\mathcal{F}f$ denote the Fourier transformation of a function $f : \mathbb{R}^N \to \mathbb{R}$, i.e. for $u, t \in \mathbb{R}^N$

$$\mathcal{F}f(t) = \int_{\mathbb{R}^N} f(u) \, e^{i<u,t>} \, du$$

so that

$$f(x) = \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} e^{-i<x,t>} \mathcal{F}_N f(t) \, dt. \tag{2}$$

In the following we will not only deal with Fourier transformation with respect to all N variables, especially the case of one-dimensional Fourier transformation of an N-dimensional function will also be important. In these cases we will point out the dimension by means of an index j such that $\mathcal{F}_j$ denotes j-dimensional Fourier transformation. Under mild Sobolev smoothness assumptions on f, the so-called Projection Theorem (cf. Natterer (1986), Theorem 1.1) $\mathcal{F}_N f(us) = \mathcal{F}_1 Rf(s,u)$, holds. Here, the second Fourier transform is taken with respect to the second variable only, i.e. $\mathcal{F}_1 Rf(s,u) = \mathcal{F}_1\big(Rf(s,\cdot)\big)(u)$. Thus one can derive explicit inversion formulae which will be the basis for the construction of our kernel-type estimators. With a smoothing parameter $\delta > 0$, satisfying $\delta \to 0$ as $\epsilon \to 0$ and $\delta \cdot \epsilon^{-2} \to \infty$, we will use kernels K_δ that are implicitly defined via their Fourier transforms $\mathcal{F}K_\delta$ which have

6

the property $\mathcal{F}K_\delta(t) \to \frac{1}{2(2\pi)^{N-1}}|t|^{1-N}$ as $\delta \to 0$. To see why this makes sense, introduce polar coordinates $t = u \cdot s$, $s \in \mathbb{S}^{N-1}$, and apply the projection theorem to obtain

$$
\begin{aligned}
f(x) &= \frac{1}{(2\pi)^N} \int_{\mathbb{S}^{N-1}} \int_{\mathbb{R}+} e^{-iu<x,s>} u^{N-1} \mathcal{F}_N f(us)\, du\, ds \\
&= \frac{1}{2(2\pi)^N} \int_{\mathbb{S}^{N-1}} \int_{\mathbb{R}} e^{-iu<x,s>} |u|^{N-1} \mathcal{F}_1 Rf(s,u)\, du\, ds,
\end{aligned}
\tag{3}
$$

where we also used that $\mathcal{F}_1 Rf(s,\cdot)$ is an even function. Since our data is on Rf, in a next step we derive a representation of the function f from (3) in which the Fourier transform $\mathcal{F}_1 Rf$ is replaced by Rf itself which could be achieved by an application of the Plancherel theorem if the function $u \mapsto |u|^{N-1}$ were square intagrable on $\mathbb{R}$. At this point we approximate $\frac{1}{2(2\pi)^{N-1}}|u|^{N-1}$ by some function $\mathcal{F}K_\delta$ of compact support such that $\frac{1}{2(2\pi)^{N-1}}|u|^{N-1} \approx \mathcal{F}K_\delta$ in an appropriate way for small values of δ (for details see Assumption 1 below). Then we can in fact write

$$
\begin{aligned}
f(x) &\approx \frac{1}{(2\pi)} \int_{\mathbb{S}^{N-1}} \int_{\mathbb{R}} e^{-iu<x,s>} \mathcal{F}K_\delta(u) \mathcal{F}_1 Rf(s,u)\, du\, ds \\
&= \int_{\mathbb{S}^{N-1}} \int_{\mathbb{R}} K_\delta(<x,s>-u) Rf(s,u)\, du\, ds =: (A_\delta f)(x).
\end{aligned}
\tag{4}
$$

Below we show that A_δ is actually a regularized inverse of Rf, also with respect to the sup-norm. In conclusion, to estimate f in model (1) we consider kernel-type estimators $\hat{f}(\cdot; \delta, \epsilon)$ of the form

$$
\hat{f}(x; \delta, \epsilon) = \int_{\mathbb{S}^{N-1}} \int_{\mathbb{R}} K_\delta\big(<s,x>-u\big)\, dY(s,u),
\tag{5}
$$

where evidently, $E\hat{f}(x; \delta, \epsilon) = A_\delta f$.

3.2 Kernel choice and bias estimates

We proceed by discussing possible choices for the kernel function K_δ in (5). For an overview of popular choices in the numerical literature see Natterer and Wübbeling (2007).
Cavalier (1999, 2000) proposes to use the kernel (Ram-Lak filter) $K_{\delta,1}$ with

$$
\mathcal{F}K_{\delta,1}(t) = \frac{1}{2}(2\pi)^{1-N}|t|^{N-1} I_\delta(t),
\tag{6}
$$

where $I_\delta = I_{[-\frac{1}{\delta}, \frac{1}{\delta}]}$ is the indicator function of the interval $[-\frac{1}{\delta}, \frac{1}{\delta}]$. It is a rather straightforward approximation of $\frac{1}{2}(2\pi)^{1-N}|t|^{N-1}$ that, due to the compact support results indeed in a smooth kernel. Nonetheless the rough edges also cause slow decay of the Fourier transform in the tails. To illustrate this, consider the case $N = 2$. Here we can give an explicit formula

for the resulting kernel $K_{\delta,1} = \frac{1}{\delta^2} K_1\left(\frac{\cdot}{\delta}\right)$ with

$$K_1(u) = \begin{cases} \frac{1}{(2\pi)^2}\left(\frac{u \cdot \sin(u) + \cos(u) - 1}{u^2}\right) & u \neq 0 \\ \frac{1}{2(2\pi)^2} & u = 0. \end{cases}$$

Note that we number some specific kernels by K_1, K_2, K_3, the element of scale family is then denoted by $K_{i,\delta}$. Obviously the kernel K_1 is smooth, $K_1 \in L^2(\mathbb{R}^N)$ but $K_1 \notin L^1(\mathbb{R}^N)$. The heavy tails result in poor practical performance of the estimator based on this kernel.

A natural alternative is to use smoothed versions of the indicator instead which will produce kernels with faster decay in the tails. Interestingly, representation (6) shows that there is a certain asymmetry regarding the dimension. While for odd dimensions N the absolute value in $|t|^{N-1}$ is redundant and, hence, smoothing of the indicator results in a smooth function $\mathcal{F}K$ and thus a fast decaying kernel K, the same no longer holds true for even dimensions. Here, the smoothness of $\mathcal{F}K$ is limited by the smoothness of $t \mapsto |t|^{N-1}$ in 0. Hence, a proper, continuous replacement for I_δ in (6) will already improve the decay properties as much as possible. In this case, for $N = 2$ this will lead to kernels K with rate of decay of $O\left(1/u^2\right)$ (see Example 1) at best, instead of $O\left(1/u\right)$. This asymmetry of the operator is further discussed in Natterer (1986), Chapter II.2.

Hoderlein et al. (2010) used generalised versions of $K_{\delta,1}$ by replacing the indicator by the function $L_{r,\delta}(t) := \left(1 - |\delta t|^r\right) I_{[0,\frac{1}{\delta}]}$ for an $r > 0$. This yields kernels of order r with the well-known properties from classical kernel smoothing problems, which, however, have slower order bias terms in the sup-norm.

Consider the following assumption on a function F.

Assumption 1. Let $F : \mathbb{R} \to \mathbb{R}$ be a symmetric function that satisfies **(i)** $\operatorname{supp}(F) \subset [-1, 1]$, **(ii)** $F(0) = 1$ and $0 \leq F(x) \leq 1$ for all $x \in \mathbb{R}$, and

(iiia) there exists $M > 0$ such that $|F(\delta\|\omega\|) - I_{[0,1]}(\delta\|\omega\|)| \leq \delta^M \|\omega\|^M$, or

(iiib) there is a $0 < D < 1$ such that $F(t) = 1$ for $t \in [-D, D]$.

Given such an F, define K_δ via its Fourier transform.

$$\mathcal{F}K_\delta(t) = \frac{1}{2 \cdot (2\pi)^{N-1}} |t|^{N-1} F(\delta|t|).$$

By symmetry of F,

$$K_\delta(u) = \frac{1}{(2\pi)^N} \int_{\mathbb{R}^+} t^{N-1} F(\delta t) \cos(tu)\, dt. \tag{7}$$

Example 1.

8

- Consider the filter $F = (1 - |t|^r) I_{[-1,1]}(t)$ as in Hoderlein et al. (2010). We have $F(\delta\|\omega\|) - I_{[0,1]}(\delta\|\omega\|) = \delta^r \|\omega\|^r$ and Assumption 1 **(iiia)** is satisfied for $M = r$. For $r = 2$ and $N = 2$ we obtain the explicit form

$$K_2(u) = \begin{cases} \frac{1}{(2\pi)^2}\left(\frac{-2u^2\cos(u)-u^2+6u\sin(u)+6\cos(u)-6}{u^4} \right) & u \neq 0 \\ \frac{1}{4(2\pi)^2} & u = 0. \end{cases}$$

- Setting $F(t) = (1 - t^2)^r I_{[-1,1]}$, Assumption 1 **(iiia)** holds with $M = 2$ and for $r = 2$ and $N = 2$ we obtain the expressions

$$K_3(u) = \begin{cases} \frac{1}{(2\pi)^2}\left(\frac{120\cos(u)-120+120u\sin(u)-48u^2\cos(u)-12u^2-8u^3\sin(u)-u^4}{u^6} \right) & u \neq 0 \\ \frac{1}{6(2\pi)^2} & u = 0. \end{cases} \quad (8)$$

- The full indicator $1_{[-1,1]}$, leading to K_1, evidently satisfies Assumption 1 **(iiib)**. In order to achieve lighter tails of the resulting kernel, the indicator on some $[-D, D]$, $0 < D < 1$, could be smoothly extended onto $\mathbb{R}$. Such a filter F would require numerical integration to evaluate the resulting kernel K.

- The cosine filter $F(t) = \cos(t\pi/2) I_{[-1,1]}(t)$ satisfies $1 - F(t) \leq (\pi/2)^3 t^2$, and hence satisfies Assumption 1 **(iiia)** with $M = 2$ (with an additional constant).

Consider the Sobolev space $\mathcal{W}^m(\mathbb{R}^N) := \left\{ f \in L^2(\mathbb{R}^N) \,|\, (1 + \|\cdot\|^2)^{\frac{m}{2}} \mathcal{F}f \in L^2(\mathbb{R}^N) \right\}$ with corresponding seminorm

$$\|f\|_m^2 = \int_{\mathbb{R}^N} (1 + \|\omega\|^2)^m |\mathcal{F}f|^2(\omega)\, d\omega, \qquad f \in \mathcal{W}^m(\mathbb{R}^N),$$

and given $L > 0$ set

$$\mathcal{W}^m(\mathbb{R}^N; L) = \{ f \in \mathcal{W}^m(\mathbb{R}^N) : \|f\|_m \leq L \}.$$

Lemma 1. *Suppose that $m > N/2$, and suppose that F satisfies Assumption 1* **(i)** *and* **(ii)**. *Then for the bias of $\hat{f}(x; \epsilon, \delta)$*

a. under Assumption 1 **(iiia)** *for $M \geq m - N/2$, given L and $0 < \eta < m - N/2$ there is a $C_1 = C_1(L, \eta, M, N, m) > 0$ such that*

$$\sup_{f \in \mathcal{W}^m(\mathbb{R}^N; L)} \sup_{x \in \mathbb{R}^N} |(A_\delta)f(x) - f(x)| \leq C_1\, \delta^{m-N/2-\eta}, \qquad \delta > 0.$$

b. under Assumption 1 **(iiib)**, *for $L, \delta > 0$ we have that*

$$\sup_{f \in \mathcal{W}^m(\mathbb{R}^N; L)} \sup_{x \in \mathbb{R}^N} |(A_\delta)f(x) - f(x)| \leq L \left(\frac{\rho_N}{(2m - N)(2\pi)^{2N}} \right)^{1/2} \left(1 + D^{-(2m-N)/2}\right) \delta^{m-N/2}.$$

Thus, kernels satisfying only Assumption 1 **(iiia)**, like the ones employed in Hoderlein et al. (2010) have suboptimal bias properties in the sup-norm. The reason is that they involve a remainder term in the bias which requires the L_1 norm of $\|\omega\|^n f(\omega)$ for the desired power n of δ. However, this is only garantueed to exist if $n < m - N/2$. As we shall see in the simulations, the finite sample properties of such kernels, especially the second class introduced in Example 1, may nevertheless be quite reasonable.

3.3 Asymptotic uniform confidence bands

In order to construct uniform confidence regions, we first require the following result on the asymptotic distribution of the maximal deviation of the estimator from its mean over the region of interest. We let

$$
\begin{aligned}
Y_\delta(x) &= \epsilon^{-1}\,\delta^{N-1/2}\left(\hat{f}(x;\delta,\epsilon) - E\hat{f}(x;\delta,\epsilon)\right) \\
&= \delta^{N-1/2}\int_{\mathbb{S}^{N-1}}\int_{\mathbb{R}}K_\delta\big(<s,x>-u\big)\,dW(s,u), \quad x \in \mathbb{R}^N,
\end{aligned}
\tag{9}
$$

which, because of the white noise structure, does no longer depend on ϵ nor f.

Theorem 2. *Suppose that F satisfies Assumption 1 **(i)** and **(ii)**. For a Jordan-measurable set $B \subset \mathbb{R}^N$ with $0 < \operatorname{Vol}B(<\infty)$ we set*

$$
M(\delta) = \sup_{x\in B}\,\epsilon^{-1}\,\delta^{N-1/2}\left|\hat{f}(x;\delta,\epsilon) - E\hat{f}(x;\delta,\epsilon)\right| = \sup_{x\in B}\left|Y_\delta(x)\right|.
$$

Then, as $\delta \to 0$,

$$
P\Big((2\log\delta^{-N})^{1/2}\Big(\frac{M(\delta)}{C_1^{1/2}} - D(\delta)\Big) \le z\Big) \to \exp\big(-2e^{-z}\big), \quad z \ge 0,
$$

where

$$
C_1 = \frac{1}{2(2\pi)^{(2N-1)}}\int_0^1 t^{2N-2}|F(t)|^2\,dt\,,
$$

and

$$
D(\delta) = (2\log\delta^{-N})^{1/2} + \Big(\frac{N-1}{2}\log\log\delta^{-1} + \log\big(\frac{(2N)^{(N-1)/2}C_2}{(2\pi)^{1/2}}\big)\Big)(2\log\delta^{-N})^{-1/2}, \tag{10}
$$

$$
C_2 = (2\pi)^{-N/2}\,\operatorname{Vol}B\left(\frac{\int_{\mathbb{R}^N}w_1^2\,\|w\|^{N-1}\,F(\|\omega\|)\,dw}{\int_{\mathbb{R}^N}\|w\|^{N-1}\,F(\|\omega\|)\,dw}\right)^{N/2}. \tag{11}
$$

Remarks

1. The Gaussian fields Y_δ in (9) are not of the convolution type as in Bickel and Rosenblatt (1973b), Rosenblatt (1976), or Bissantz et al. (2007a), thus, it is not obvious that the corresponding arguments apply.

2. We use the multivariate extreme value theory for Gaussian processes, and specifically Corollary 2 in Bickel and Rosenblatt (1973a).

3. The theorem in particular implies the rate of convergence $O_P\big(\epsilon\,\delta^{-N+1/2}\,(\log\delta^{-1})^{1/2}\big)$ for the maximal deviation $\sup_{x\in B}\big|\hat{f}(x;\delta,\epsilon) - E\hat{f}(x;\delta,\epsilon)\big|$, uniformly in f. If the kernel function F satisfies Assumption 1 (iiib) and $\delta(\epsilon)\sim\big(\epsilon^2\log(1/\epsilon)\big)^{1/(2m+N-1)}$, then we obtain the rate of convergence $O_P\big(\big(\epsilon^2\log(1/\epsilon)\big)^{(2m-N)/(2m+N-1)}\big)$ for $\sup_{x\in B}\big|\hat{f}(x;\delta(\epsilon),\epsilon) - f(x)\big|$, uniformly over $f\in\mathcal{W}^m(\mathbb{R}^N;L)$. This can be extended to

$$\sup_{f\in\mathcal{W}^m(\mathbb{R}^N;L)} E_f \sup_{x\in B}\big|\hat{f}(x;\delta(\epsilon),\epsilon) - f(x)\big| \le C\,\big(\epsilon^2\log(1/\epsilon)\big)^{(2m-N)/(2m+N-1)}.$$

In order to construct an asymptotic confidence set for the function f, we need to choose δ at a slightly faster rate than the optimal $\delta(\epsilon)$. For a given level $\alpha\in(0,1)$ we let

$$\mathcal{I}_\alpha(x;\delta,\epsilon) := \big[\hat{f}(x;\epsilon,\delta) - \Phi_{\alpha,\delta,\epsilon}\,,\ \hat{f}(x;\epsilon,\delta) + \Phi_{\alpha,\delta,\epsilon}\big],$$

$$\Phi_{\alpha,\delta,\epsilon} := \sqrt{C_1}\bigg(\frac{-\ln\big(-\frac{1}{2}\ln(1-\alpha)\big)}{\sqrt{2\ln(\delta^{-N})}} + D(\delta)\bigg)\varepsilon\delta^{\frac{1}{2}-N} \tag{12}$$

where the constants C_1 and $D(\delta)$ are specified in the theorem. Then we have the following corollary.

Corollary 3. *If the kernel function F satisfies Assumption 1* **(i)**, **(ii)** *and* **(iiib)**, *and if $\delta\to 0$ so that $\delta\big(\epsilon^2\log(1/\epsilon)\big)^{-1/(2m+N-1)}\to 0$, $\epsilon\to 0$, then for any $0<\alpha<1$ and $L>0$ we have that*

$$\liminf_{\epsilon\to 0}\ \inf_{f\in\mathcal{W}^m(\mathbb{R}^N;L)} P_f\big(f(x)\in\mathcal{I}_\alpha(x;\delta,\epsilon)\ \forall\ x\in B\big)\ge 1-\alpha.$$

For an application, we need to estimate the noise level ϵ. We shall discuss this in the discrete model in the following section.

3.4 Confidence sets for compactly-supported functions

The white noise model (1) is useful as an idealization of a more realistic discrete regression model. In such models, it is assumed that the function f has compact support, and that observations are only taken in a given set including the support of Rf.

Therefore, suppose that $\operatorname{supp} f\subset\{x\in\mathbb{R}^N\ :\ \|x\|<1\}=:B_1(0)$, so that $\operatorname{supp} Rf\subset\mathbb{S}^{N-1}\times[-1,1]$, and let

$$\mathcal{W}^m_c(\mathbb{R}^N;L) = \{f\in\mathcal{W}^m(\mathbb{R}^N)\ :\ \|f\|_m\le L,\ \operatorname{supp} f\subset B_1(0)\}.$$

Suppose that for such an f, observations are made according to the restricted white noise

model

$$dY(s, u) = Rf(s, u) \, ds \, du + \epsilon dW(s, u), \qquad s \in \mathbb{S}^{N-1}, \ u \in [-1, 1]. \tag{13}$$

As an estimator, we therefore take

$$\hat{f}_c(x; \delta, \epsilon) = \int_{\mathbb{S}^{N-1}} \int_{-1}^{1} K_\delta \big(< s, x > -u \big) \, dY(s, u), \tag{14}$$

In this case we impose an additional assumption on the decay of the kernel and its derivatives that guarantee that the influence of the noise simply weighted by the kernel is negligible.

Assumption 2. Let F be such that $K \in L^1(\mathbb{R})$ and such that for $u \neq 0$

$$|K^{(j)}(u)| \leq \frac{C_K}{u^{\alpha_N + j}}, \ j = 0, \ldots, N-1$$

for positive constants $C_K \in \mathbb{R}$, $\alpha_N > (N-1)/2$.

Note that Assumption 2 is satified if $F \in C^{2N-1}(\mathbb{R})$ by smoothness and integrability properties of the functions $t \mapsto t^j |t|^{N-1}$, $j = 1, \ldots, N$. While it is neither satisfied for K_1 nor for K_2 the assumption holds for the kernel K_3 or one constructed with a smoothed index function. Then we have the following result.

Corollary 4. *Suppose that the kernel function F satisfies Assumption 1* **(i)**, **(ii)** *and* **(iiib)**, *and additionally Assumption 2. Let $0 < \kappa < 1$ and suppose that $B \subset \{x \in \mathbb{R}^N : \|x\| \leq 1 - \kappa\}$ is Jordan measurable with $\mathrm{Vol}\, B > 0$. If $\delta \to 0$ so that $\delta \big(\epsilon^2 \log(1/\epsilon) \big)^{-1/(2m+N-1)} \to 0$, $\epsilon \to 0$, then for any $0 < \alpha < 1$ and $L > 0$ we have that*

$$\liminf_{\epsilon \to 0} \inf_{f \in \mathcal{W}_c^m(\mathbb{R}^N; L)} P_f \big(f(x) \in \mathcal{I}_{\alpha, c}(x; \delta, \epsilon) \ \forall \ x \in B \big) \geq 1 - \alpha,$$

where $\mathcal{I}_{\alpha, c}(x; \delta, \epsilon)$ is defined in (12) with $\hat{f}$ replaced by $\hat{f}_c$ in (14).

For the proof, we show that for the restricted noise process,

$$\tilde{Y}_\delta(x) = \delta^{N-1/2} \int_{\mathbb{S}^{N-1}} \int_{-1}^{1} K_\delta \big(< s, x > -u \big) \, dW(s, u).$$

we have that for some $\tilde{\eta} > 0$,

$$\sup_{\|x\| \leq 1 - \kappa} |Y_\delta(x) - \tilde{Y}_\delta(x)| = o_P(\tilde{\eta}).$$

Here, it is essential that a small boundary in $B_1(0)$ is excluded, the constant in o_P will depend on the value of κ. Due to the multi-dimensionality and the range of integration $\mathbb{S}^{N-1}$, the proof is somewhat more involved than in the standard univariate regression framework.
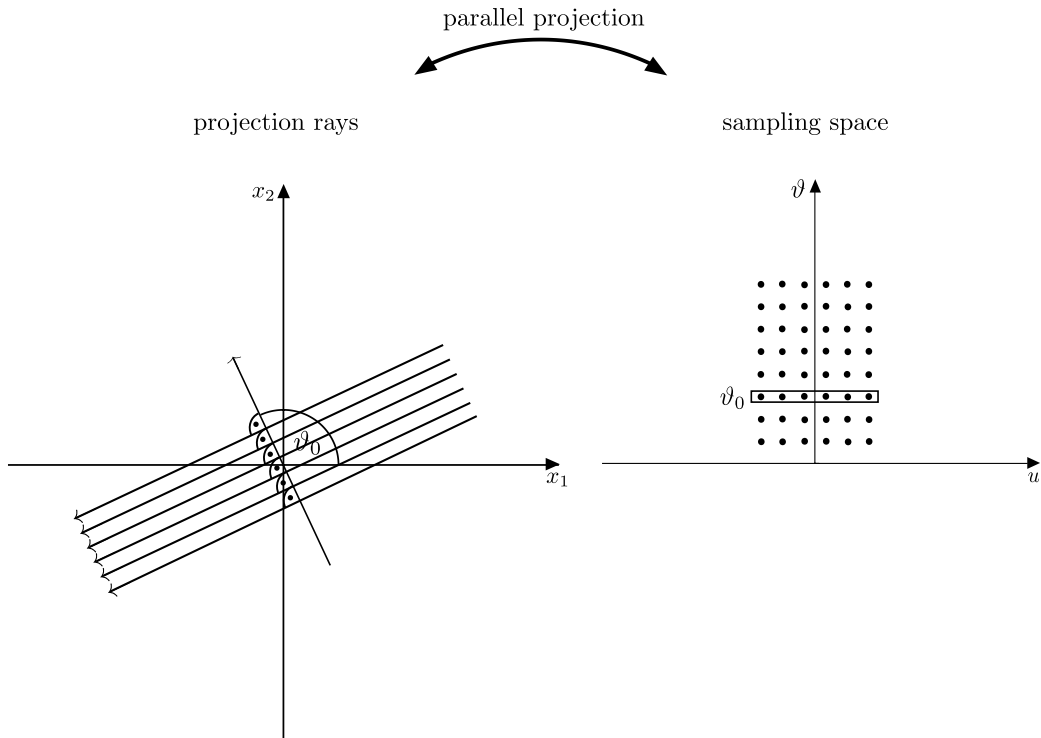
Figure 2: Schematic representation of the alignment of the projection rays in the parallel beam design (left) and the corresponding values in the sample space (right).

4   Discrete regression models

In this section, we restrict ourselves to the case $N = 2$. Here, we parametrize $s \in \mathbb{S}^1$ by the angle $\vartheta \in [0, 2\pi)$ according to $s = (\cos\vartheta, \sin\vartheta)$. Suppose that $\operatorname{supp} f \subset \{(x, y) \in \mathbb{R}^2 : x_1^2 + x_2^2 < 1\} =: B_1(0)$, and that observations are taken according to

$$Y_{(i,j)} = (Rf)(\vartheta_i, u_j) + \epsilon_{(i,j)}, \quad i = 1, \ldots, n_1, \, j = 1, \ldots, n_2, \tag{15}$$

where $\varepsilon_{(i,j)}$ are centered, i.i.d. random variables with finite variances $\mathbb{E}\varepsilon_{i,j}^2 = \sigma^2$ and existing moment larger than 2, and

$$\vartheta_i = 2\pi \, (i - 1/2)/n_1, \quad u_j = (j - 1/2)/n_2,$$

$i = 1, \ldots, n_1, \, j = 1, \ldots, n_2$, and n_1 and n_2 should be of the same order. As a physical model, it is the parallel-beam design: For each fixed angle $\vartheta_j \, j = 1, \ldots, n_1$, the set of all n_2 rays of different values for u, $u_1, \ldots, u_{n_2}$, are sent through the object, see Figure 2 for an illustration. Note that we only consider measurements at positive points u_j in order to avoid redundancies as simply changing the sign of the distance variable u only changes the orientation of the line

under consideration. A discrete analogue for the estimator (14) is then given by

$$\hat{f}_n(x_1, x_2; \delta) = \frac{2\pi}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} 2K_\delta\big(x_1 \cos(\vartheta_i) + x_2 \sin(\vartheta_i) - u_j\big) Y_{(i,j)}, \qquad (16)$$

where we abbreviate $n = (n_1, n_2)$. We start by estimating the additional discretization bias involved in the estimator.

Lemma 5. *Suppose that $m > 5/2$, and suppose that F satisfies Assumption 1* **(i)**, **(ii)** *and 1* **(iiib)** *as well as Assumption 2. Consider the estimator $\hat{f}_n(\cdot; \delta)$ in model (15). Then for $L, \delta > 0$ and constants C_1 and C_2 which are independent of the function f or the kernel K we have for the bias of $\hat{f}_n(x_1, x_2; \delta)$ that*

$$\sup_{f \in \mathcal{W}_c^m(\mathbb{R}^2; L)} \sup_{x \in B_1(0)} \big| E_f \hat{f}_n(x; \delta) - f(x) \big|$$

$$\leq L \frac{1 + D^{-(m-1)}}{\big(2(m-1)(2\pi)^3\big)^{1/2}} \delta^{m-1} + C_1 \Big(\frac{1}{n_1} + \frac{1}{n_2}\Big) \int_{\mathbb{R}^2} \|\omega\| |\mathcal{F}f(\omega)| \, d\omega$$

$$+ \frac{C_2}{\delta^3} \Bigg(\sum_{\substack{i,j \in \{0,1,2\} \\ i+j=2}} \frac{1}{n_1^i n_2^j} \Bigg) \Bigg(\max_{\substack{i,j \in \{0,1,2\} \\ i+j=2}} \sup_{(\vartheta,u) \in [0,2\pi] \times [0,1]} |D^{(i,j)} Rf(\vartheta, u)| \cdot \max_{j=0}^{2} \|K^{(j)}\|_1 \Bigg)$$

In order to construct uniform confidence sets, we let

$$\tilde{\mathcal{I}}_\alpha(x; \delta, n) := \big[\hat{f}_n(x; \delta) - \tilde{\Phi}_{\alpha,\delta,n} \,, \, \hat{f}_n(x; \delta) + \tilde{\Phi}_{\alpha,\delta,n} \big],$$

$$\tilde{\Phi}_{\alpha,\delta,n} := \sqrt{C_1} \left(\frac{-\ln\big(-\frac{1}{2}\ln(1-\alpha)\big)}{\sqrt{2\ln(\delta^{-2})}} + D(\delta) \right) \frac{(4\pi)^{1/2} \sigma}{\sqrt{n_1 n_2} \delta^{3/2}},$$

where

$$C_1 = \frac{1}{2(2\pi)^3} \int_0^1 t^2 |F(t)|^2 \, dt \,,$$

and

$$D(\delta) = (2\log \delta^{-2})^{1/2} + \Big(\frac{1}{2} \log\log \delta^{-1} + \log\big(\frac{2C_2}{(2\pi)^{1/2}}\big)\Big)(2\log \delta^{-2})^{-1/2},$$

$$C_2 = (2\pi)^{-1} \pi (1-\kappa)^2 \left(\frac{\int_{\mathbb{R}^2} w_1^2 \|w\| \, F(\|\omega\|) \, dw}{\int_{\mathbb{R}^2} \|w\| \, F(\|\omega\|) \, dw} \right)^{1/2}.$$

Corollary 6. *Suppose that $m > 5/2$, and suppose that F satisfies Assumption 1* **(i)**, **(ii)** *and 1* **(iiib)** *as well as Assumption 2. Further assume that $\mathbb{E}|\varepsilon_{i,j}|^r < \infty$ for some $r > 4$. Consider the estimator $\hat{f}_n(\cdot; \delta)$ in model (15), and suppose that $n_1 \asymp n_2$ as $n_1, n_2 \to \infty$. Suppose that $\delta \to 0$ such that $\delta\big(n_1^{-2} \log(n_1)\big)^{-1/(2m+1)} \to 0$ and $\log(n)/\sqrt{n_1 \delta^3} \to 0$, $n_1 \to \infty$, then for any*

$0 < \alpha, \kappa < 1$ and $L > 0$ we have that

$$\liminf_{n_1, n_2 \to \infty} \inf_{f \in \mathcal{W}_c^m(\mathbb{R}^2; L)} P_f\left(f(x) \in \tilde{\mathcal{I}}_\alpha(x; \delta, n) \ \forall \ \|x\| \leq 1 - \kappa\right) \geq 1 - \alpha.$$

The noise variance σ^2 can be estimated $\sqrt{n_1 n_2}$-consistently in two dimensions by difference-based estimators or by estimation of the squared residuals using smoothing, see Munk et al. (2005).

5 Simulations

In this section we illustrate the finite sample properties of both our estimator $\hat{f}_n$, defined in (16), and the proposed confidence sets by means of a small simulation study. To this end we use two different two-dimensional objects to generate data.

The first object is rather simple and consists of a sum of two (Gaussian-shaped) peaks. In more detail, the image is generated from the signal

$$f^0(x_1, x_2) = \left[e^{-8 \cdot (x^2 + (y - 0.3)^2)} + e^{-8 \cdot ((x - 0.2)^2 + (y + 0.3)^2)}\right] \cdot \mathbf{I}_{[(x^2 + y^2) \leq 1]}.$$

To achieve smoothness of the image at the boundary of the unit disc we have applied a smoothing filter with standard deviation ≈ 0.01 resulting in a smooth object.

Our second object has been constructed in a more complicated way and is motivated by the fact that a typical application of a Radon transform such as CT imaging requires the reconstruction of cross-sectional images of specific parts of a patient's body. Hence, a simple model of relevant strucures consists of a disc on which certain features of interest (such as organs, bones or spine, but also tumours etc.) are superimposed. The second object mimics such features by overlaying the disc-like main object with several additional, smaller ones. In more detail, it consists of a sum of several elliptical objects each of which is generated in two steps: firstly, a proto-image based on a scaled indicator function has been created which was smoothed by a Gaussian filter in a second step. Table 1 summarizes the relevant properties of the different objects and also contains the different standard deviations of the Gaussian filters for the case of an image of size 128×128 pixels. For other image sizes these standard deviations have been scaled accordingly.

We generate observations from the model (15), where the $\epsilon_{(i,j)}$ are taken as i.i.d. standard normally distributed. In the subsequent simulations we have always used $n_1 = n_2 = n_{x,y}$, where $n_{x,y}$ is the number of points used in the discretization of the true object along the x and y-axis. Moreover, we have used in all simulations described in the following the kernel $K_{\delta,3}$ with $r = 2$ in the estimator $\hat{f}_n$, which had shown best properties among the kernels discussed. We choose $\kappa = 0.2$ for the parameter which governs the exclusion of the boundary. Concerning the smoothing parameter $\delta > 0$, we first determined a suitable value for each

Component	Scaled indicator function	σ
Main object	$((x_1^2 + x_2^2) < 0.64) \cdot 1.$	5
Left eye	$(((x_1 - 0.25)^2 + (x_2 + 0.25)^2) < 0.04) \cdot 0.3$	3
Right eye	$(((x_1 + 0.25)^2 + (x_2 + 0.25)^2) < 0.04) \cdot 0.3$	3
Nose	$(((3x_1^2 + (x_2 - 0.1)^2) < 0.04) \cdot 0.2$	3
Mouth	$((x_1^2 + 5(x_2 - \sqrt{\max(0, 0.3 - 2x^2)})^2) < 0.09) \cdot 0.3$	2

Table 1: Properties of the components of the second object, where the column labeled σ gives the standard deviation of Gaussian smoothing in the generation of the true images (see text for details).

Image	Sample size	σ	δ	predicted	$(\hat{f}_n - E\hat{f}_n)$	$(\hat{f}_n - f_0)$
'two peaks'	128×128	0.01	0.01	0.051	0.048	0.049
'two peaks'	128×128	0.1	0.025	0.114	0.110	0.116
'two peaks'	256×256	0.01	0.007	0.046	0.044	0.043
'two peaks'	256×256	0.1	0.015	0.132	0.128	0.128
'face'	128×128	0.01	0.01	0.051	0.049	0.054
'face'	128×128	0.1	0.025	0.114	0.111	0.152
'face'	256×256	0.01	0.007	0.046	0.044	0.046
'face'	256×256	0.1	0.015	0.132	0.128	0.137

Table 2: Predicted and simulated widths of 80%-confidence bands. The column labeled $(\hat{f}_n - E\hat{f}_n)$ shows 80% quantiles of the supremum of the simulated distance between $\hat{f}_n$ and its mean and the column labeled $(\hat{f}_n - f_0)$ the respective quantile of the simulated distribution of $\sup|\hat{f}_n - f_0|$.

simulation scenario by applying the L_∞-motivated bandwidth selection method introduced in Bissantz et al. (2007a). This fixed smoothing parameter is then used in all runs for the respective scenario.

Figure 3 shows both objects under consideration and the corresponding sinograms in comparison to both one exemplary dataset with $\sigma = 0.1$ each and reconstructions from these datasets with bandwidth $\delta = 0.025$ and an image size of 128×128. Sample slices through the 90%-confidence surfaces for estimates with these parameters are shown in Figure 4. Finally, Fig. 5 illustrates (from left to right) the upper 90%-confidence surface, the estimate and the lower 90%-confidence surface for estimates of object 1 and 2, respectively.

Tables 2 and 3 summarise the results of the simulation study of the estimator, based on 500 simulation runs for each combination of object, image size and noise level. The results show a reasonably close proximity of the simulated width of confidence bands with their theoretical asymptotic counterparts.
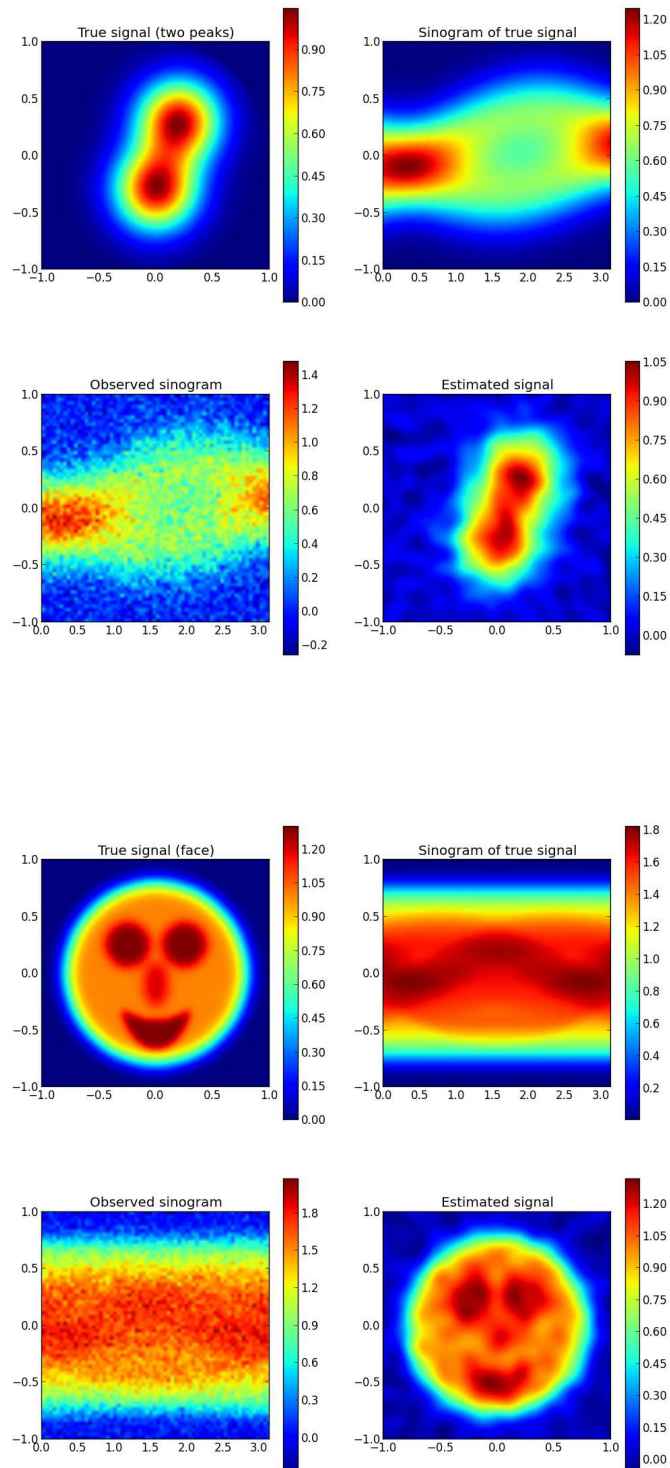
Figure 3: Sample reconstructions of the two objects. Top: 'two peaks', bottom: 'face'.
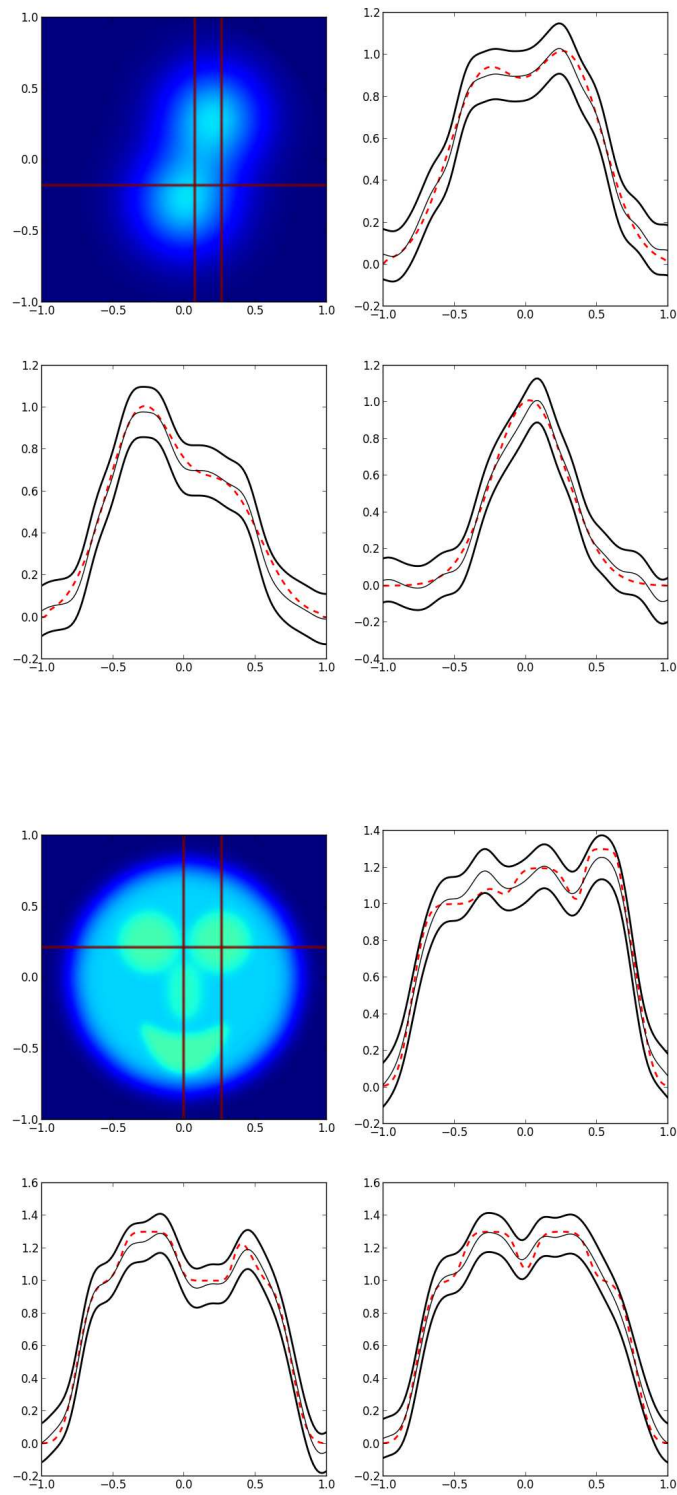
Figure 4: 1d-slices through confidence surfaces for the two objects. Top: 'two peaks', bottom: 'face'.
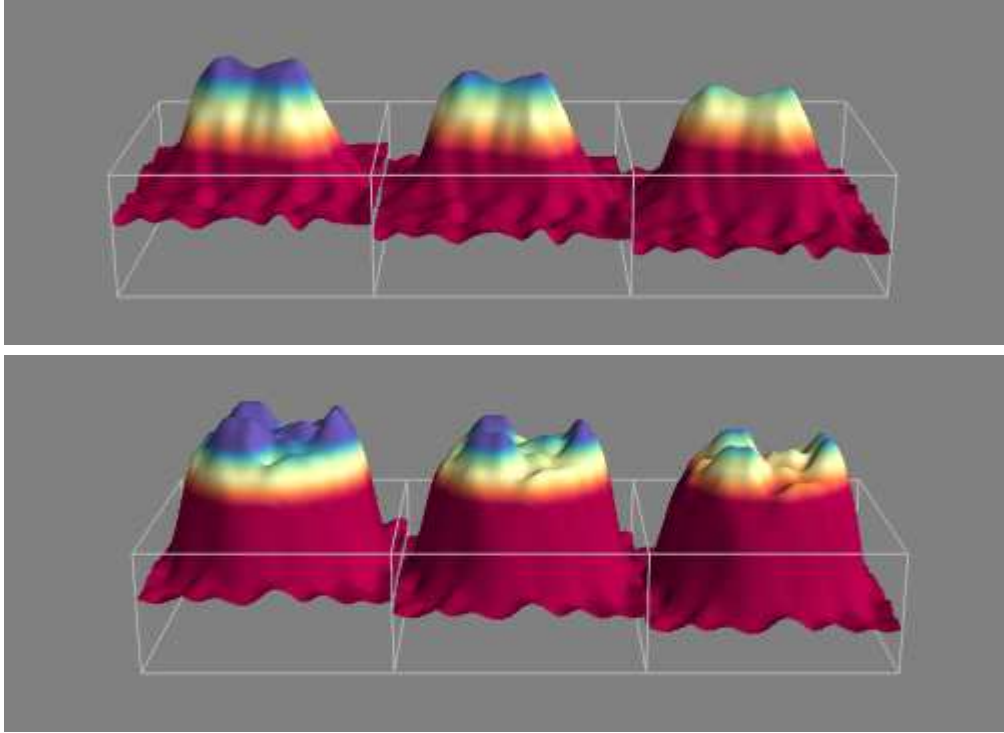
18

Figure 5: Confidence surfaces for object 'two peaks' (top) and 'face' (bottom).

Image	Sample size	σ	δ	predicted	$(\hat{f}_n - E\hat{f}_n)$	$(\hat{f}_n - f_0)$
'two peaks'	128×128	0.01	0.01	0.053	0.050	0.051
'two peaks'	128×128	0.1	0.025	0.120	0.116	0.122
'two peaks'	256×256	0.01	0.007	0.047	0.038	0.041
'two peaks'	256×256	0.1	0.015	0.138	0.135	0.136
'face'	128×128	0.01	0.01	0.053	0.051	0.057
'face'	128×128	0.1	0.025	0.120	0.118	0.161
'face'	256×256	0.01	0.007	0.047	0.046	0.048
'face'	256×256	0.1	0.015	0.138	0.134	0.143

Table 3: Predicted and simulated widths of 90%-confidence bands. Similar to table 2, the column labeled $(\hat{f}_n - E\hat{f}_n)$ shows 80% quantiles of the supremum of the simulated distance between $\hat{f}_n$ and its mean and the column labeled $(\hat{f}_n - f_0)$ the respective quantile of the simulated distribution of $\sup|\hat{f}_n - f_0|$.

6 Proofs

In the following remark we will list three results from Fourier analysis which will be frequently used throughout this section.

Remark 1.   (i) $\mathcal{F}f(x + \cdot)(\xi) = \exp(-ix\xi)\mathcal{F}f(\xi)$

(ii) $\mathcal{F}_1 Rf(s,t) = \mathcal{F}_N f(t \cdot s)$, $t \in \mathbb{R}$, $s \in \mathbb{S}^{N-1}$, where $\mathcal{F}_1 Rf(s,t)$ denotes one-dimensional Fourier transformation of $Rf(s, \cdot)$ for fixed $s \in \mathbb{S}^{N-1}$ and $\mathcal{F}_N f(t \cdot s)$ denotes N-dimensional Fourier transformation at the point $t \cdot s$, i.e. $\mathcal{F}_N f(t \cdot s) = (\mathcal{F}_N f)(t \cdot s)$. This identity is also known as Projection Theorem (cf. Natterer (1986), Theorem 1.1). Note that the constants in this identity differ from those given in Natterer (1986) which is due to the different choice of standardisation in the definition of the Fourier transform.

(iii) $(2\pi)^N \cdot \langle f, g \rangle = \langle \mathcal{F}f, \mathcal{F}g \rangle$, where the factor $(2\pi)^N$ is due to the choice of standardisation in the definition of the Fourier transform (Plancherel Theorem, see, e.g., Folland (1984), Theorem 8.29).

6.1 Proof of Lemma 1

Let $x \in \mathbb{R}^N$, $f \in \mathcal{W}^m(\mathbb{R}^N; L)$. Following (4) and (3) backwards yields

$$
\begin{aligned}
(A_\delta)f(x) =: & \int_{\mathbb{R}} \int_{\mathbb{S}^{N-1}} K_\delta\big(<s,x> -u\big) Rf(s,u)\, du \\
= & \frac{1}{2\pi} \int_{\mathbb{S}^{N-1}} \int_{\mathbb{R}} \mathcal{F}K_\delta(t) \mathcal{F}Rf(s,t) \exp(-it<s,x>)\, dt \\
= & \frac{1}{2(2\pi)^N} \int_{\mathbb{S}^{N-1}} \int_{\mathbb{R}} |t|^{N-1} F(\delta|t|) \mathcal{F}f(ts) \exp(-it<s,x>)\, du\, ds \\
= & \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} F(\delta\|\omega\|) \mathcal{F}f(\omega) \exp(-i<\omega,x>)\, d\omega.
\end{aligned}
$$

Therefore, we obtain

$$
\begin{aligned}
(A_\delta)f(x) - f(x) = & \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} \big(F(\delta\|\omega\|) - 1\big) \mathcal{F}f(\omega) \exp(-i<\omega,x>)\, d\omega \\
= & \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} \big(F(\delta\|\omega\|) - I_{[0,\frac{1}{\delta}]}(\|\omega\|)\big) \mathcal{F}f(\omega) \exp(-i<\omega,x>)\, d\omega \\
& - \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} I_{(\frac{1}{\delta},\infty)}(\|\omega\|) \mathcal{F}f(\omega) \exp(-i<\omega,x>)\, d\omega =: b_I(x) - b_{II}(x),
\end{aligned}
$$

where $b_I(x)$ and $b_{II}(x)$ are defined in an obvious manner.

An application of the Cauchy-Schwarz inequality gives

$$|b_{II}(x)| \le \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} I_{(1,\infty)}(\delta\|\omega\|)|\mathcal{F}f(\omega)| \left(\frac{1+\|\omega\|^2}{\|\omega\|^2}\right)^{\frac{m}{2}} d\omega$$

$$\le \frac{1}{(2\pi)^N} \left(\int_{\mathbb{R}^N} I_{(1,\infty)}(\delta\|\omega\|)\frac{1}{\|\omega\|^{2m}}\, d\omega \cdot \int_{\mathbb{R}^N} |\mathcal{F}f(\omega)|^2 \left(1+\|\omega\|^2\right)^m d\omega\right)^{\frac{1}{2}}.$$

Since $m > \frac{N}{2}$ we have

$$\int_{\mathbb{R}^N} I_{(1,\infty)}(\delta\|\omega\|)\frac{1}{\|\omega\|^{2m}}\, d\omega = \rho_N \int_{\frac{1}{\delta}}^\infty t^{N-1}\frac{1}{t^{2m}}\, dt = \frac{\rho_N}{2m-N}\delta^{2m-N},$$

therefore

$$\left|b_{II}(x)\right| = L \left(\frac{\rho_N}{(2m-N)\,(2\pi)^{2N}}\right)^{1/2} \delta^{m-N/2}.$$

Concerning $b_I(x)$, under Assumption 1**(iiib)** we obtain

$$|b_I(x)| \le \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} I_{(\frac{D}{\delta},\infty)}(\|\omega\|)|\mathcal{F}f(\omega)|\, d\omega,$$

which may be estimated as $b_{II}(x)$ to yield the second part of the lemma. Under Assumption 1**(iiia)** we have

$$\left|F(\delta\|\omega\|) - I_{[0,\frac{1}{\delta}]}(\|\omega\|)\right| \le (\delta\|\omega\|)^M \wedge 1 \le (\delta\|\omega\|)^{m-N/2-\eta}.$$

Since $f \in \mathcal{W}^m(\mathbb{R}^N; L)$, we have $\|\omega\|^{m-N/2-\eta}|\mathcal{F}f(\omega)| \in L_1(\mathbb{R}^N)$. Therefore

$$|b_I(x)| \le \frac{1}{(2\pi)^N}\delta^{m-N/2-\eta} \int_{\mathbb{R}^N} \|\omega\|^{m-N/2-\eta}|\mathcal{F}f(\omega)|\, d\omega \le C_1 \delta^{m-N/2-\eta}.$$

$\square$

6.2 Proof of Theorem 2

First observe that the Gaussian fields Y_δ are stationary since

$$Y_\delta(x+h) = \delta^{N-1/2} \int_{\mathbb{S}^{N-1}} \int_{\mathbb{R}} K_\delta\big(<s,x> + <s,h> -u\big)\, dW(s,u)$$

$$=^d Y_\delta(x), \qquad x \in \mathbb{R}^N,$$

where we used that integrals w.r.t. the Gaussian sheet are translation invariant in distribution. Next, we observe that the processes $(Y_\delta(x))_{x\in\mathbb{R}^N}$ scale as follows

$$(Y_\delta(x))_{x\in\mathbb{R}^N} \stackrel{d}{=} (Y_1(x/\delta))_{x\in\mathbb{R}^N}. \tag{17}$$

21

To show (17), we prove that

$$\mathrm{Cov}\big(Y_\delta(x), Y_\delta(y)\big) = \mathrm{Cov}\big(Y_1(x/\delta), Y_1(y/\delta)\big).$$

To this end, compute

$$
\begin{aligned}
\mathrm{Cov}\big(Y_\delta(x), Y_\delta(y)\big) &= \delta^{2N-1} \int_{\mathbb{S}^{N-1}} \int_{\mathbb{R}} K_\delta\big(<s,x>-u\big) K_\delta\big(<s,y>-u\big) \, ds \, du \\
&= \delta^{2N-1} \frac{1}{2\pi} \int_{\mathbb{S}^{N-1}} \int_{\mathbb{R}} e^{-it(<s,x>-<s,y>)} \, \mathcal{F}K_\delta(t)^2 dt \, ds \\
&= \delta^{2N-1} \frac{(2\pi)^{-(2N-1)}}{4} \int_{\mathbb{S}^{N-1}} \int_{\mathbb{R}} e^{-i<t\,s,x-y>} |t|^{2N-2} F(\delta|t|) dt \, ds \\
&= \delta^{2N-1} \frac{(2\pi)^{-(2N-1)}}{4} \rho_N \int_{\mathbb{R}^N} e^{-i<z,x-y>} \|z\|^{N-1} F(\delta\|z\|) dz \\
&= \frac{(2\pi)^{-(2N-1)}}{4} \rho_N \int_{\mathbb{R}^N} e^{-i<w/\delta,x-y>} \|w\|^{N-1} F(\|w\|) dw \\
&= \frac{(2\pi)^{-(2N-1)}}{4} \rho_N \int_{\mathbb{R}^N} e^{-i<w,x/\delta-y/\delta>} \|w\|^{N-1} F(\|w\|) dw \\
&= \mathrm{Cov}\big(Y_1(x/\delta), Y_1(y/\delta)\big),
\end{aligned}
$$

where the last step follows by going through the calculations backwards. In particular,

$$M(\delta) = \sup_{x\in B} |Y_\delta(x)| \overset{d}{=} \sup_{x\in B/\delta} |Y_1(x)|.$$

Therefore, we can analyze the asymptotic distribution of the right side, using corollary 2 in Bickel and Rosenblatt (1973a), similarly as Theorem 2 in Rosenblatt (1976). As calculated above, the covariance function of $Y_1(x)$ is

$$
\begin{aligned}
r(h) = \mathrm{Cov}\big(Y_1(x+h), Y_1(x)\big) &= \frac{(2\pi)^{-(2N-1)}}{4} \int_{\mathbb{S}^{N-1}} \int_{\mathbb{R}} e^{-i<t\,s,h>} |t|^{2N-2} F(t) dt \, ds \\
&= \frac{(2\pi)^{-(2N-1)}}{4} \rho_N \int_{\mathbb{R}^N} e^{-i<w,h>} \|w\|^{N-1} F(\|w\|) dw.
\end{aligned}
$$

For the variance, the first part gives

$$r(0) = \frac{(2\pi)^{-(2N-1)}}{4} \int_{-1}^{1} |t|^{2N-2} |F(t)|^2 \, dt = C_1.$$

For the covariance, we have for the vector of partial derivatives

$$\partial_h r(h) = \frac{(2\pi)^{-(2N-1)}}{4} \rho_N \int_{\mathbb{R}^N} (-i)w \, e^{-i<w,h>} \|w\|^{N-1} F(\|w\|) dw,$$

and thus $\partial_h r(0) = 0$ by symmetry. Similarly,

$$\partial_h \, \partial_h^T r(h) = -\frac{(2\pi)^{-(2N-1)}}{4} \, \rho_N \int_{\mathbb{R}^N} w w^T \, e^{-i<w,h>} \, \|w\|^{N-1} \, F(\|w\|) dw,$$

so that

$$-\partial_h \, \partial_h^T r(0) = \frac{(2\pi)^{-(2N-1)}}{4} \, \rho_N \int_{\mathbb{R}^N} w w^T \, \|w\|^{N-1} \, F(\|w\|) dw.$$

By symmetry again,

$$\int_{\mathbb{R}^N} w_i w_j \, \|w\|^{N-1} \, F(\|w\|) dw = 0, \quad i \neq j,$$

so that

$$r(h)/r(0) = \frac{1}{2} \frac{\int_{\mathbb{R}^N} w_1^2 \, \|w\|^{N-1} \, F(\|w\|) \, dw}{\int_{\mathbb{R}^N} \|w\|^{N-1} \, F(\|w\|) dw} \, \|h\|^2 + o(\|h\|^2), \qquad h \to 0.$$

Finally, $r(h)$ is the Fourier transform of an L_2-function (in fact a compactly supported function), thus it is itself in L_2. An application of corollary 2 in Bickel and Rosenblatt (1973a) finishes the proof in case $\mathrm{Vol}\, B = 1$.

For the general case, we let $s_0 = \left(\mathrm{Vol}\, B\right)^{1/N}$, so that $\mathrm{Vol}\,(B/s_0) = 1$. Consider $\tilde{Y}(x) = Y_1(s_0 x)$, $x \in \mathbb{R}^N$. Then

$$\tilde{M}(\delta) := \sup\left\{|\tilde{Y}(x)|, \ x \in B/(\delta s_0)\right\}$$
$$= \sup\left\{|Y_1(s_0 \, x)|, \ s_0 \, x \in B/\delta\right\}$$
$$= \sup\left\{|Y_1(x)|, \ x \in B/\delta\right\} = M(\delta).$$

Therefore, in order to treat the supremum of Y_1 over B, we can apply the case already proved to the process $\tilde{Y}$ where the supremum is taken with respect to $x \in B/s_0$. For its covariance, we have that

$$\tilde{r}(x) = \mathrm{Cov}\left(\tilde{Y}(x), \tilde{Y}(0)\right) = r(s_0 x),$$

therefore

$$\tilde{r}(h)/\tilde{r}(0) = \frac{s_0^N}{2} \frac{\int_{\mathbb{R}^N} w_1^2 \, \|w\|^{N-1} \, F(\|w\|) \, dw}{\int_{\mathbb{R}^N} \|w\|^{N-1} \, F(\|w\|) dw} \, \|h\|^2 + o(\|h\|^2), \qquad h \to 0,$$

and the conclusion follows since $s_0^N = \mathrm{Vol}\, B$. $\qquad\qquad\qquad\qquad \square$

6.3 Proof of Corollary 3

Let

$$\widetilde{M}(\delta, \epsilon) = \sup_{x \in B} \epsilon^{-1} \, \delta^{N-1/2} \left|\hat{f}(x; \delta, \epsilon) - f(x)\right|,$$

so that, for $z_\alpha = -\ln\left(-\frac{1}{2}\ln(1-\alpha)\right)$ we have

$$P_f\big(f(x) \in \mathcal{I}_\alpha(x; \delta, \epsilon) \ \forall \ x \in B\big) = P_f\Big((2\log\delta^{-N})^{1/2}\Big(\frac{\widetilde{M}(\delta,\epsilon)}{C_1^{1/2}} - D(\delta)\Big) \leq z_\alpha\Big).$$

Further, by Lemma 1 and the assumption on δ, we have uniformly in $f \in \mathcal{W}^m(\mathbb{R}^N; L)$,

$$\Big|(2\log\delta^{-N})^{\frac{1}{2}}\Big(\frac{M(\delta,\epsilon)}{C_1^{\frac{1}{2}}} - D(\delta)\Big) - (2\log\delta^{-N})^{\frac{1}{2}}\Big(\frac{\widetilde{M}(\delta,\epsilon)}{C_1^{\frac{1}{2}}} - D(\delta)\Big)\Big|$$

$$= \Big(\frac{2\log\delta^{-N}\delta^{2N-1}}{C_1\epsilon^2}\Big)^{\frac{1}{2}} \cdot \big|\|\hat{f} - \mathbb{E}\hat{f}\|_\infty - \|\hat{f} - f\|_\infty\big|$$

$$\leq \Big(\frac{2\log\delta^{-N}\delta^{2N-1}}{C_1\epsilon^2}\Big)^{\frac{1}{2}} \cdot \|\hat{f} - \mathbb{E}\hat{f} - \hat{f} + f\|_\infty$$

$$= \Big(\frac{2\log\delta^{-N}\delta^{2N-1}}{C_1\epsilon^2}\Big)^{\frac{1}{2}} \cdot \|f - \mathbb{E}\hat{f}\|_\infty = o(1),$$

Therefore,

$$P_f\Big((2\log\delta^{-N})^{1/2}\Big(\frac{\widetilde{M}(\delta,\epsilon)}{C_1^{1/2}} - D(\delta)\Big) \leq z_\alpha\Big) \geq P\Big((2\log\delta^{-N})^{1/2}\Big(\frac{M(\delta)}{C_1^{1/2}} - D(\delta)\Big) \leq z_\alpha + o(1)\Big),$$

where $o(1)$ is independent of f, and the conclusion follows from Theorem 2. $\qquad\square$

6.4 Proof of Corollary 4

We now give a detailed proof for the case $N = 2$ and briefly sketch its extension to the cases $N > 2$ afterwards. We have

$$|Y_\delta(x) - \tilde{Y}_\delta(x)| \leq \Big|\delta^{\frac{3}{2}} \int_{\mathbb{S}^1} \int_{(1,\infty)} K_\delta(<s,x> -u)\, dW(s,u)\Big|$$

$$+ \Big|\delta^{\frac{3}{2}} \int_{\mathbb{S}^1} \int_{(-\infty,-1)} K_\delta(<s,x> -u)\, dW(s,u)\Big| =: I + II,$$

where the summands I and II are defined in an obvious manner. Consider the term I. Let $\widetilde{W}$ be a Wiener sheet on $[0, 2\pi] \times [0, \infty)$, and let $s(\vartheta) = (\cos\vartheta, \sin\vartheta)$. Then

$$\delta^{\frac{3}{2}} \int_{\mathbb{S}^1} \int_{(1,\infty)} K_\delta(<s,x> -u)\, dW(s,u) \overset{d}{=} \delta^{\frac{3}{2}} \int_0^{2\pi} \int_{(1,\infty)} K_\delta\big(<s(\vartheta),x> -u\big)\, d\widetilde{W}(\vartheta,u).$$

Now, the partial integration formula for a continuously differentiable function $f : \mathbb{R}^2 \mapsto \mathbb{R}$ and a Wiener sheet on $[0,\infty)^2$, $z = (z_1, z_2)^T$, $a = (a_1, a_2)^T$, $b = (b_1, b_2)^T \in \mathbb{R}^2$ reads as follows:

$$
\int_{[a,b]} f(z)\, dW(z) = \int_{[a,b]} W(z) \frac{\partial^2}{\partial z} f(z)\, dz
$$

$$
- \int_{[a_1,b_1]} W(z_1, b_2) \frac{\partial}{\partial z_1} f(z_1, b_2)\, dz_1 + \int_{[a_1,b_1]} W(z_1, a_2) \frac{\partial}{\partial z_1} f(z_1, a_2)\, dz_1
$$

$$
- \int_{[a_2,b_2]} W(b_1, z_2) \frac{\partial}{\partial z_2} f(b_1, z_2)\, dz_2 + \int_{[a_2,b_2]} W(a_1, z_2) \frac{\partial}{\partial z_2} f(a_1, z_2)\, dz_2
$$

$$
+ W(b_1, b_2) f(b_1, b_2) - W(a_1, b_2) f(a_1, b_2) - W(b_1, a_2) f(b_1, a_2) + W(a_1, a_2) f(a_1, a_2).
$$

Therefore, integration by parts yields

$$
I = \delta^{\frac{3}{2}} \Big| \int_0^{2\pi} \int_{(1,\infty)} K_\delta^{(2)}(<s(\vartheta), x> -u)\widetilde{W}(\vartheta, u) <s^\perp(\vartheta), x>\, du\, d\vartheta
$$

$$
+ \int_0^{2\pi} K_\delta^{(1)}(<s, x> -1) <s^\perp(\vartheta), x> \widetilde{W}(\vartheta, 1)\, d\vartheta
$$

$$
- \int_1^\infty K_\delta^{(1)}(x_1 - u)\widetilde{W}(2\pi, u)\, du - \widetilde{W}(2\pi, 1)K_\delta(x_1 - 1) \Big| =: |I_1 + I_2 + I_3 + I_4|,
$$

since the other terms such as $W(0,1)K_\delta(x_1 - 1)$ vanish. First let us estimate I_1. Here we have to deal with a two parameter Wiener sheet, to which we can apply a two parameter version of the LIL, such as given in Csörgo and Révèsz (1981), Theorem 1.12.3. To this end define the set $D_u := \{(s,t) \,|\, s \cdot t \le 2\pi \cdot u,\ s \le 2\pi \cdot u,\ t \le 2\pi \cdot u\}$ and note that $(\vartheta, u) \in D_u$ if $(\vartheta, u) \in [0, 2\pi] \times (1, \infty)$, i.e., especially for all pairs (ϑ, u) in the domain of integration. We find

$$
|\widetilde{W}(\vartheta, u)| \overset{d}{=} |\sqrt{\delta}\, \widetilde{W}(\vartheta, u/\delta)| = \sqrt{8\pi\delta \frac{u}{\delta} \log\log(2\pi u/\delta)} \cdot \left| \frac{\widetilde{W}(\vartheta, u/\delta)}{\sqrt{8\pi \frac{u}{\delta} \log\log(2\pi u/\delta)}} \right|
$$

$$
\le \sqrt{8\pi u \log\log(2\pi u/\delta)} \cdot \sup_{u > \frac{1}{\delta}} \sup_{(s,t) \in D_u} \left| \frac{\widetilde{W}(s,t)}{\sqrt{8\pi u \log\log(2\pi u)}} \right| \tag{18}
$$

$$
\le C \cdot \sqrt{8\pi u \log\log(2\pi u/\delta)} \quad \text{a.s.,}
$$

for some random $C > 0$ by the version of the LIL cited above. We further observe

$$
\int_0^{2\pi} \int_1^\infty \left| \sqrt{u \log\log(2\pi u/\delta)} \cdot K_\delta^{(2)}(<s(\vartheta), x> -u) <s^\perp(\vartheta), x> \right| du\, d\vartheta
$$

$$
\le \int_0^{2\pi} \int_1^\infty \sqrt{u \log\log(2\pi u/\delta)} \cdot |K_\delta^{(2)}(<s(\vartheta), x> -u)|\, du\, d\vartheta \tag{19}
$$

$$
= \delta^{-3} \int_0^{2\pi} \int_{-\infty}^{\frac{\langle s(\vartheta), x\rangle - 1}{\delta}} |K^{(2)}(z)| \sqrt{(\langle s(\vartheta), x\rangle - \delta z) \log\log(2\pi\delta^{-1}(\langle s(\vartheta), x\rangle - \delta z))}\, dz\, d\vartheta,
$$

25

where $s^\perp(\vartheta) = (-\sin\vartheta, \cos\vartheta)$, by the definition (7) and a substitution. Since

$$< s(\vartheta), x > -\delta z \le 1 - \kappa - \delta z \le 1 - \delta z \le -\delta z(1/\kappa + 1) = \delta|z|(1/\kappa + 1)$$

we can estimate

$$\sqrt{((\langle s(\vartheta), x\rangle - \delta z)\log\log(2\pi\delta^{-1}(\langle s(\vartheta), x\rangle - \delta z))} \le \sqrt{\delta(1/\kappa + 1)|z|\log\log(2\pi|z|(1/\kappa + 1))} \tag{20}$$

and obtain from (18), (19) and (20) that a.s.,

$$\begin{aligned}
I_1 \le&\, C\,\delta^{3/2} \int_0^{2\pi} \int_1^\infty \sqrt{u\log\log(2\pi u/\delta)} \cdot |K_\delta^{(2)}(< s(\vartheta), x > -u)|\,du\,d\vartheta \\
\le&\, C\,2\pi(1/\kappa + 1)^{\frac{1}{2}}\delta^{3/2-3+1/2} \int_{-\infty}^{\frac{\langle s(\vartheta), x\rangle - 1}{\delta}} |K^{(2)}(z)|\sqrt{|z|\log\log(2\pi|z|(1/\kappa + 1))}\,dz \\
\le&\, C\,2\pi(1/\kappa + 1)^{\frac{1}{2}}\delta^{-1} \int_{-\infty}^{-\kappa/\delta} |K^{(2)}(z)|\sqrt{|z|\log\log(2\pi|z|(1/\kappa + 1))}\,dz,
\end{aligned}$$

where in the last step we used that $(\langle s(\vartheta), x\rangle - 1)/\delta < -\kappa/\delta$. By Assumption 2, $K^{(2)}(u)$ decays with a higher power than $|u|^{-5/2}$, so that for a sufficiently small constant $\check{\eta} > 0$,

$$\begin{aligned}
I_1 \le&\, 2\pi C(1/\kappa + 1)^{\frac{1}{2}}\delta^{-1} \int_{-\infty}^{-\kappa/\delta} |K^{(2)}(z)|\sqrt{|z|}\Big(\frac{\delta z}{\kappa}\Big)^{1+\check{\eta}}\sqrt{\log\log(2\pi|z|(1/\kappa + 1))}\,dz \\
\le&\, \tilde{C}\delta^{\check{\eta}}
\end{aligned}$$

for some finite (random) $\tilde{C}$. Concerning I_2, since

$$|K_\delta^{(1)}(< s, x > -1)| \le \Big(\delta^3\Big|\frac{< s, x > -1}{\delta}\Big|^{\alpha_2+1}\Big)^{-1}$$

as well as $\sup_{\vartheta\in[0,2\pi]}|\widetilde{W}(\vartheta, 1)| = O_P(1)$, we obtain $I_2 = o_P(\delta^{\alpha_2-3/2})$ and by assumption, $\alpha_2 - 3/2 > 0$. For I_4, the same result in proven in a similar manner. Finally we will give an estimate for the term I_3. Note that $(2\pi)^{-\frac{1}{2}}\widetilde{W}(2\pi, u) \overset{d}{=} \sqrt{\delta/2\pi}\,\widetilde{W}(2\pi, u/\delta)$ is the usual one parameter Wiener process. An application of the LIL yields

$$\sup_{u > \frac{1}{\delta}}\Big|\frac{(2\pi)^{-\frac{1}{2}}\widetilde{W}(2\pi, u)}{\sqrt{2u\log\log(u)}}\Big| \to 1 \quad \text{a.s. for } \delta \to 0.$$

This implies

$$I_3 = O_P(\delta^{\frac{3}{2}}\log\log(1/\delta)) \int_{(1,\infty)} |K_\delta^{(1)}(x_1 - u)|\,du.$$

By Assumption 2, $K^{(1)}(u)$ decays with a higher power than $|u|^{-3/2}$. Therefore, for a sufficiently small constant $\tilde{\eta} > 0$ we obtain

$$\delta^{\frac{3}{2}} \int_{(1,\infty)} |K_\delta^{(1)}(x_1 - u)|\, du = \delta^{-\frac{1}{2}} \int_{(-\infty,(x_1-1)/\delta)} |K^{(1)}(z)|\, dz$$

$$\leq \delta^{-\frac{1}{2}} \int_{(-\infty,-\kappa/\delta)} |K^{(1)}(z)| \left(\frac{\delta|z|}{\kappa}\right)^{\frac{1}{2}+\tilde{\eta}} dz = o\left(\delta^{\tilde{\eta}}\right)$$

and hence $I_3 = o_P\left(\delta^{\tilde{\eta}}\right)$. The combination of the results gives the estimate $I = o_P\left(\delta^{\eta}\right)$ for some small constant $\eta > 0$. All estimates hold uniformly with respect to the variable $x \in B_{1-\kappa}(0)$. For symmetry reasons we obtain the same rate of convergence for the term II as for I with the same arguments which completes the proof of the lemma in the case $N = 2$.

In the case $N > 2$, we introduce N-dimensional polar coordinates in order to parametrize the sphere $\mathbb{S}^{N-1}$

$$s(\vartheta) = \begin{pmatrix} \sin(\vartheta_1)\sin(\vartheta_2)\ldots\sin(\vartheta_{N-3})\sin(\vartheta_{N-2})\cos(\varphi) \\ \sin(\vartheta_1)\sin(\vartheta_2)\ldots\sin(\vartheta_{N-3})\sin(\vartheta_{N-2})\sin(\varphi) \\ \sin(\vartheta_1)\sin(\vartheta_2)\ldots\sin(\vartheta_{N-3})\cos(\vartheta_{N-2}) \\ \vdots \\ \sin(\vartheta_1)\cos(\vartheta_2)\sin(\vartheta_1) \end{pmatrix},$$

$\vartheta = (\vartheta_1,\ldots,\vartheta_{N-2},\varphi) \in (0,\pi)^{N-2} \times (0,2\pi)$ and $ds = \sin^{N-2}(\vartheta_1) \cdot \ldots \cdot \sin(\vartheta_{N-2})\, d\vartheta$. Let $\widehat{W}$ be a Wiener sheet on $[0,\pi]^{N-2} \times [0,2\pi] \times [0,\infty)$. We find

$$\delta^{\frac{N+1}{2}} \int_{\mathbb{S}^{N-1}} \int_{(1,\infty)} K_\delta(<s,x>-u)\, dW(s,u)$$

$$\overset{d}{=} \delta^{\frac{N+1}{2}} \int_0^{2\pi} \int_0^\pi \ldots \int_0^\pi \int_{(1,\infty)} K_\delta(<s,x>-u)\sqrt{\sin^{N-2}(\vartheta_1) \cdot \ldots \cdot \sin(\vartheta_{N-2})}\, d\widehat{W}(\vartheta,u).$$

Starting from the right hand side of the latter equation, again by integration by parts, we obtain integrals with respect to the variables (ϑ, u) instead of the integral with respect to the Wiener sheet $\widehat{W}$. The estimations are very similar to the ones in the two-dimensional case and are therefore omitted. For the N-dimensional integration by parts formula or N-dimensional versions of the LIL see, e.g., Proksch et al. (2012) or Paranjape and Park (1973), respectively. $\square$

6.5 Proof of Lemma 5

Since F is compactly supported it follows that K is smooth and an application of the Hausdorff-Young inequality shows that its derivatives of all orders are bounded and since

$K \in L^1(\mathbb{R})$ by assumption, also the derivatives $K^{(j)} \in L^1(\mathbb{R})$, $j \in \{1, 2\}$. Further, see Natterer (1980) Theorem 3.2, $Rf \in \mathcal{W}^{m+1/2}(\mathcal{Z})$, $m + \frac{1}{2} > 3$ and hence $Rf \in C^2(\mathcal{Z})$ a.s.. With these arguments we directly obtain the following estimation for the error in the integral approximation

$$\mathbb{E}\hat{f}_n(x; \delta) = \int_{\mathbb{S}^1} \int_{\mathbb{R}} K_\delta(<s, x> -u) Rf(s, u) \, du \, ds + O\left(\frac{1}{n_1} + \frac{1}{n_2} + \sum_{\substack{i,j \in \{0,1,2\} \\ i+j=2}} \frac{1}{\delta^3 n_1^i n_2^j} \right).$$

The constant given in the lemma can be derived by an application of Taylor's theorem and by following (3) and (4) backwards were we now regard K'_δ and K''_δ as well as K_δ and the respective derivatives of Rf. The calculations are straightforward but tedious and are therefore omitted. $\qquad \square$

6.6 Proof of Corollary 6

In order to prove the assertion of the corollary we need to show that under the given assumptions the quantity

$$Y_n(x) = \frac{1}{\sqrt{n_1 n_2 \sigma^2}} \left(\hat{f}_n(x_1, x_2, \delta) - \mathbb{E}\hat{f}_n(x_1, x_2, \delta) \right) = \frac{4\pi}{\sqrt{n_1 n_2 \sigma^2}} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} K_\delta(\langle s(\vartheta_i), x \rangle - u_j) \varepsilon_{i,j}$$

can be approximated by the Gaussian field $\tilde{Y}_\delta(x)$ with an error of smaller order than $1/\sqrt{\log(n)}$ uniformly with respect to the variable x. The lengthy details of this, i.e., the definition of suitable approximating processes and detailed calculations can be found in Proksch et al. (2012).

With an application of Corollary 4 we obtain the approximation of $\tilde{Y}_\delta$ by Y_δ and Lemma 5 yields the negligibility of the bias. An application of Theorem 2 completes the proof of this corollary.

$\qquad \square$

Acknowledgements

References

ADLER, R. J. and TAYLOR, J. E. (2007). *Random Fields and Geometry.* Springer, New York.

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

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** 115020.

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. (1999). Asymptotically efficient estimation in a problem related to tomography. *Math. Methods Statist.*, **7** 445–456.

CAVALIER, L. (2000). Efficient estimation of a density in a problem of tomography. *Ann. Statist.*, **28** 630–647.

CAVALIER, L. (2008). Nonparametric statistical inverse problems. *Inverse Problems*, **24** 034004, 19. URL http://dx.doi.org/10.1088/0266-5611/24/3/034004.

CLAESKENS, G. and VAN KEILEGOM, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Annals of Statistics*, **31** 1852–1884.

CSÖRGO, M. and RÉVÈSZ, P. (1981). *Strong Approximations in Probability and Statistics.* Academic Press, London.

ENGL, H. W., HANKE, M. and NEUBAUER, A. (1996). *Regularization of inverse problems*, vol. 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). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Annals of Statistics*, **20** 675–694.

HELGASON, S. (2011). *Integral geometry and Radon transforms*. Springer.

HODERLEIN, S., KLEMELÄ, J. and MAMMEN, E. (2010). Analyzing the random coefficient model nonparametrically. *Econometric Theory*, **26** 804–837.

JOHNSTONE, I. M. and SILVERMAN, B. W. (1990). Speed of estimation in positron emission tomography and related inverse problems. *Ann. Statist.*, **18** 251–280.

KAIPIO, J. and SOMERSALO, E. (2005). *Statistical and Computational Inverse Problems*. Springer, Berlin.

KERKYACHARIAN, G., KYRIAZIS, G., LE PENNEC, E., PETRUSHEV, P. and PICARD, D. (2010). Inversion of noisy radon transform by svd based needlets. *Appl. Comput. Harmon. Anal.*, **28** 24–45.

KERKYACHARIAN, G., PENNEC, E. L. and PICARD, D. (2012). Radon needlet thresholding. *Bernoulli*, **18** 391–433.

KONAKOV, V. D. and PITERBARG, V. I. (1984). On the convergence rate of maximal deviation distribution for kernel regression estimate. *Journal of Multivariate Analysis*, **15** 279–294.

KOROSTELEV, A. P. and TSYBAKOV, A. B. (1993). *Minimax theory of image reconstruction*, vol. 82 of *Lecture Notes in Statistics*. Springer-Verlag.

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.

MUNK, A., BISSANTZ, N., WAGNER, T. and FREITAG, G. (2005). On difference-based variance estimation in nonparametric regression when the covariate is high dimensional. *J. R. Stat. Soc., Ser. B, Stat. Methodol.*, **67** 19–41.

NATTERER, F. (1980). A sobolev space analysis of picture reconstruction. *SIAM J. Appl. Math.*, **39** 402–411.

NATTERER, F. (1986). *The mathematics of computerized tomography.* B. G. Teubner, Stuttgart; John Wiley & Sons, Ltd., Chichester.

NATTERER, F. and WÜBBELING, F. (2007). *Mathematical methods in image reconstruction. 1st paperback ed.* 1st ed. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM).

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

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

PROKSCH, K., BISSANTZ, N. and DETTE, H. (2012). Confidence bands for multivariate and time dependent inverse regression models. Preprint, arXiv:1206.2743v1 [math.ST].

RIO, E. (1994). Local invariance principles and their application to density estimation. *Probab. Theory Related Fields*, **98** 21–45.

ROSENBLATT, M. (1976). On the maximal deviation of k-dimensional density estimates. *Annals of Probability*, **6** 1009–1015.

SCHMIDT-HIEBER, J., MUNK, A. and DÜMBGEN, L. (2013). Multiscale methods for shape constraints in deconvolution: confidence statements for qualitative features. *Ann. Stat.*, **41** 1299–1328.

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.