Robust Jump Detection in Regression Surface *

by Tim Garlipp and Christine H. Müller

Carl von Ossietzky University of Oldenburg

November 2, 2004

Abstract

We use the difference of two asymmetric M-kernel estimators to detect jumps in two-dimensional regression functions. The method extends and corrects the Rotational Difference Kernel Estimator method proposed by Qiu (1997). For regression functions with only one explicit jump curve and additive noise, we show consistency for the jump location and height. In a simulation study, the consistency is also demonstrated for the case that 30% of the observations are replaced by outliers. In this case, the robust M-kernel estimators are superior to the classical kernel-estimators.

Keywords: M-kernel estimation, consistency, edge detection, jump regression function, robustness against outliers.

AMS Subject classification: 62 G 20, 62 G 35, 62 G 08, 62 G 05, 62 H 35

*Research supported by the grant Mu 1031/4-1/2 of the Deutsche Forschungsgemeinschaft.

1 Introduction

Two-dimensional regression functions with discontinuities are used in various fields of application. For example, meteorological or geological departments have observations from fixed gauging stations, which are used to estimate the border of air layers or mine surfaces. Another application is the detection of edges in image analysis.

There are global an local methods for detecting discontinuities. Mainly in the theory of multivariate boundary estimation, which is closely related to edge detection, global approaches are used. For boundary estimation, observations are assumed to follow a regression function which is smooth except on the boundary of a region. Referred to images, estimating this boundary is detecting the edge of an object in the image. (See, e.g., Korostelev and Tsybakov, 1993). For a global approach see, for example, Carlstein and Krishnamoorthy (1992). A recent extension of their method is found in Ferger (2004).

Most considerations about estimation of jump locations in regression surfaces concern edge detection in image analysis where diverse local methods are known. For example, the so-called *filter*-methods, which like many others use the fact that the derivative of the image function becomes very large in the vicinity of edges. Other methods use statistical tests based on the representation of the image as a Markov field. See Davis (1975) for an overview about some of the "classical" methods, or Peli and Malah (1982) for a comparison. For some newer techniques see, for example, Müller and Song (1994), Qiu and Yandell (1997), or Hou and Koh (2003).

For one-dimensional jump detection, Qiu et al. (1991), Müller (1992), and Wu and Chu (1993) introduced similar estimators based on the difference of two one-sided kernel estimates (DKE - Difference Kernel Estimators). In smooth regions of the regression function, an estimator using only observations on the left side will be similar to an estimator using only observations on the right side. In contrast, near jump points, the difference of these two estimates will be close to the jump height (see Figure 1).

As generalisation of the one-dimensional DKE, Qiu introduced in 1997 the Rotational Difference Kernel Estimators (RDKE), on which the method discussed in this paper is based. The most important difference between the two-dimensional and the one-dimensional case is that the distinction between "left" and "right" side has now to be done along a direction. According to this direction the difference may strongly vary. If the considered point lies on an edge, differences calculated along that edge will be close to the jump height, whereas for other directions, the difference may even vanish (see Figure 2).

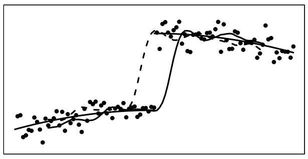


Figure 1: left sided (solid) and right sided (dashed) one-dimensional kernel estimations

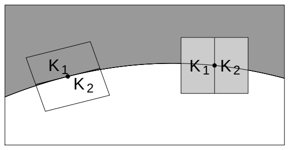


Figure 2: difference of two asymmetric two-dimensional kernel estimations

To cope with this problem, Qiu introduced rotated kernel functions. Based on two asymmetric two-dimensional kernel functions

$$K_1^*(x_1, x_2) = 0 \quad \text{for} \quad (x_1, x_2) \notin \left[-\tfrac{1}{2}, \tfrac{1}{2}\right] \times [-1, 0]$$

$$K_2^*(x_1, x_2) = 0 \quad \text{for} \quad (x_1, x_2) \notin \left[-\tfrac{1}{2}, \tfrac{1}{2}\right] \times [0, 1]$$

he defined rotated kernel functions $K_j(\theta, x_1, x_2) = K_j^*(r_1(\theta, x_1, x_2), r_2(\theta, x_1, x_2))$, where r_1 and r_2 describe the coordinates of the pixel $x^\top = (x_1, x_2)^\top$ rotated around an angle θ. With these kernel functions, asymmetric kernel estimators are defined as

$$m_n^{(j)}(\theta, x_1, x_2) = \frac{1}{n h_{1n} h_{2n}} \sum_{i=1}^{n} K_j\left(\theta, \frac{x_{1i} - x_1}{h_{1n}}, \frac{x_{2i} - x_2}{h_{2n}}\right) Z_i, \tag{1}$$

where h_{1n} and h_{2n} are the bandwidths. Then, $m_n^{(2)}(\theta, x_1, x_2) - m_n^{(1)}(\theta, x_1, x_2)$ is the estimated difference between the weighted means of the observations located on the different sides of (x_1, x_2) along the direction described by θ.

In practical applications, kernel estimators have the disadvantage of not being robust. For example, outliers among the observations may have a strong influence on the estimation. For robust regression estimation, Härdle and Gasser (1984) proposed the M-kernel estimators which are a generalization of the kernel estimators. In Section 2, a robust estimator for two-dimensional jump regression functions combining the concepts of the M-kernel estimators with those of the RDKE method of Qiu is introduced. Thereby we correct an essential mistake in the original definitions for the RDKE introduced by Qiu, which is even restated in a recent modification of the RDKE (Qiu, 2002): The scaling by the bandwidths h_{1n} and h_{2n} is fundamental in the theory of kernel estimators (see e.g. Eubank, 1988) and according to the conceptional idea, the rotation should apply to the support of the scaled kernels. But Qiu applies the rotation after the scaling (see (1)), i.e., the support of the kernels is first rotated and then scaled. This leads to the effect that the supports are deformed if the two bandwidths are unequal (see Figure 3). Since this involves a change of the direction of the line that divides the two kernels, the proofs given by Qiu do not hold without substantial changes. Therefore, we introduce new kernel functions, which do not show this disadvantage (see Figure 4).

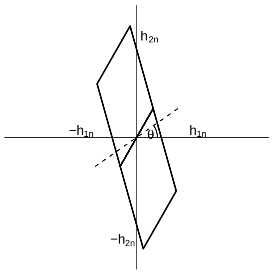


Figure 3: Domains of the two asymmetric kernel functions as defined by Qiu (1997)

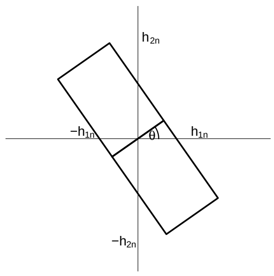


Figure 4: Domains of the two asymmetric kernel functions used in this paper

As in Qiu (1997), we introduce estimators for the jump curve ϕ and jump height C for regression functions of the form

$$m(x) \quad = \quad \widetilde{m}(x) + C(x_1)\mathbb{1}_A(x) \text{ with } A := \left\{x \in [0,1]^2 : x_2 > \phi(x_1)\right\},$$

where $\widetilde{m}$ is the (Lipschitz-) continuous part of the regression function. In Section 3, the uniform consistency of these estimators is presented for the case that the regression function is corrupted by an additive noise. The proofs are given in the appendix. Since, with the identity as score function, the M-kernel estimator becomes the ordinary kernel estimator, our proofs also include Qiu's RDKE with the now corrected succession of scaling and rotating. Moreover, Qiu uses a condition on the (asymptotical) number of observations in a given area, for example, the domains of the kernels, which is not proven. Since the design points are not required to be equidistant, this condition does not trivially hold true but needs further assumptions, which are included in the assumptions given in Section 3.

Section 4 provides a simulation study which shows the consistency also in the case where 30% of the observations are replaced by outliers. This simulation study demonstrates clearly the superiority of the RDKE based on robust M-kernel estimators to the RDKE based on the classical kernel estimators.

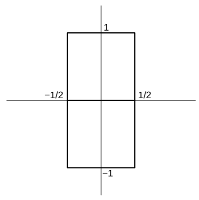
2 Definitions and assumptions

We consider n observations $Z_i = m(x_i) + \epsilon_i \in \mathbb{R}$ at design points $x_i = (x_{1i}, x_{2i})^\top \in [0,1]^2$ $(1 \leq i \leq n)$. The residuals $\epsilon_i \in \mathbb{R}$ are i.i.d. with symmetric density $f : \mathbb{R} \to \mathbb{R}$ and $E\,|\epsilon_i|^p < M < \infty$ for a $p > 4$. The regression function is supposed to be of the form $m(x) = \widetilde{m}(x) + C(x_1)\mathbb{1}_A(x)$, with $A := \{x = (x_1, x_2) \in [0,1]^2 : x_2 > \phi(x_1)\}$. Let the smooth part of the regression function $\widetilde{m} : [0,1]^2 \to \mathbb{R}$ and the jump curve $C : [0,1] \to \mathbb{R} \setminus \{0\}$ be Lipschitz-continous with Lipschitz-constant $C_{\widetilde{m}}$ and C_C respectively and w.l.o.g. let $C(x_1) > 0$. Further, let the jump height $\phi : [0,1] \to (0,1)$ be two times differentiable. The aim is to estimate the functions ϕ and C describing the jump of the regression function.

Let $K_1(x)$ and $K_2(x)$ be two one-sided, continuous kernel functions which fulfill the following conditions:

(A1) $K_1(x) = 0 \quad \text{for} \quad x \notin \left[-\frac{1}{2}, \frac{1}{2}\right] \times [-1, 0] \quad \text{and}$

$K_2(x) = 0 \quad \text{for} \quad x \notin \left[-\frac{1}{2}, \frac{1}{2}\right] \times [0, 1]$



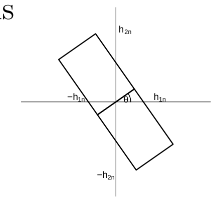
(A2) $\int\limits_{[-1,1]^2} K_j(x)dx = 1, \quad j \in \{1,2\}$

(A3) $K_j(x) \geq 0, \quad j \in \{1,2\}$.

The compact supports and the continuity imply Lipschitz-continuity of the kernel functions and their powers. Let C_{K^p} be the (common) Lipschitz-constant of the p-th power of K_1 and K_2.

With $a_{1\theta} := (\cos\theta, -\sin\theta)^\top$, $a_{2\theta} := (\sin\theta, \cos\theta)^\top$, $A_\theta := (a_{1\theta}, a_{2\theta})^\top = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, and $H_n := \begin{pmatrix} h_{1n} & 0 \\ 0 & h_{2n} \end{pmatrix}$ we define for $j \in \{1,2\}$ the rotated kernels

$$K_n^{(j)}(\theta, x) \;:=\; \frac{1}{h_{1n}h_{2n}} K_j(H_n^{-1} A_\theta x),$$



where h_{1n} and h_{2n} are the bandwidths.

For simplifying the notation let $\overline{h_n} := 2 \cdot \sqrt{h_{1n}^2 + h_{2n}^2}$ and w.l.o.g. let $h_{jn} < \frac{1}{2}$ for all $n \in \mathbb{N}$, $j \in \{1,2\}$. Since $\min_{x_1 \in [0,1]} \phi(x_1) > 0$, $\max_{x_1 \in [0,1]} \phi(x_1) < 1$, and $\overline{h_n} \to 0$, we can further assume w.l.o.g. that $0 < \overline{h_n} < \phi(x_1) < 1 - \overline{h_n} < 1$ for all $n \in \mathbb{N}$ and $x_1 \in [0,1]$. The rotated asymmetric M-kernel estimators $m_n^{(j)}(\theta, x)$ are defined as zeros of the objective functions $H_n^{(j)}(z;\theta,x)$ with

$$H_n^{(j)}(z;\theta,x) \;:=\; \sum_{i=1}^n \alpha_{i,n}^{(j)}(\theta,x)\psi(Z_i - z),$$

so that

$$m_n^{(j)}(\theta,x) \;\in\; \{z \in \mathbb{R} : H_n^{(j)}(z;\theta,x) = 0\},$$

where $\psi : \mathbb{R} \to \mathbb{R}$ is a score function and $\alpha_{i,n}^{(j)}(\theta,x)$ are rotated asymmetric Gasser-Müller weights

$$\alpha_{i,n}^{(j)}(\theta,x) \;:=\; \int_{\Delta_i} K_n^{(j)}(\theta, u - x)\, du.$$

Then, as in Qiu (1997), we define

$$M_n(\theta,x) \;:=\; m_n^{(2)}(\theta,x) - m_n^{(1)}(\theta,x).$$

It seems plausible that $|M_n(\theta,x)|$ will be close to zero independently of θ for $x \in \left[\overline{h_n}\,, 1 - \overline{h_n}\right]^2$ lying in a smooth region of the regression function, i.e. x_2 has an arbitrary distance to $\phi(x_1)$. But for x lying on the jump location curve, i.e. $x_2 = \phi(x_1)$, $|M_n(\theta,x)|$ will be close to the jump height $C(x_1)$, if the direction described by θ corresponds with the direction of the tangent of ϕ in x_1. Therefore, the jump height at a point $x \in \left[\overline{h_n}\,, 1 - \overline{h_n}\right]$ can be estimated by

$$\widetilde{C}_n(x) \;=\; |M_n(\theta_n(x), x)|$$

where $\theta_n(x)$ is the maximizing angle

$$\theta_n(x) \;\in\; \underset{\theta\in\left[-\frac{\pi}{2},\frac{\pi}{2}\right]}{\operatorname{argmax}}\; |M_n(\theta,x)|.$$

Note, that $\widetilde{C}_n(x)$ can be used as heuristic criterion for jump detection in more general models by regarding every point x as jump point, if $\widetilde{C}_n(x)$ is larger than a certain threshold value. In our model, the explicit form of the jump location curve enables us to estimate this curve by maximizing the estimated jump height for every x_1 in x_2-direction, i.e.

$$\phi_n(x_1) \;\in\; \underset{x_2\in\left[\overline{h_n}\,,1-\overline{h_n}\right]}{\operatorname{argmax}}\; \widetilde{C}_n\left((x_1,x_2)^\top\right).$$

Finally, the jump height $C(x_1)$ is estimated by inserting the estimated jump location $\phi_n(x_1)$ into $\widetilde{C}_n(x)$,

$$C_n(x_1) \;=\; \widetilde{C}_n\left((x_1,\phi_n(x_1))^\top\right).$$

In the following section the consistency of ϕ_n and C_n is shown.

3 Consistency results

Although the design points do not have to be equidistant, they must fulfill at least asymptotically a certain regularity:
Let $\Lambda = \{\Delta_i, 1 \le i \le n\}$ be a partition of $[0,1] \times [0,1]$ with

(B1) $x_i \in \Delta_i$ for $i = 1, \ldots, n$ and $\bigcup_{i=1}^{n} \Delta_i = [0,1] \times [0,1]$, $\Delta_i \cap \Delta_j = \emptyset$ for $i \neq j$,

(B2) $D_n := \max_{1 \le i \le n} d_i = O\left(n^{-1/2}\right)$ and $\max_{1 \le i \le n} |S(\Delta_i) - 1/n| = O\left(n^{-1-\lambda}\right)$ with $\lambda > 0$, where $d_i = \sup_{x,\widetilde{x} \in \Delta_i}\{||x - \widetilde{x}||\}$ is the diameter and $S(\Delta_i)$ is the area of Δ_i.

Obviously equidistant design points fulfill these conditions for arbitrary n.

Moreover, let $\psi(z) : \mathbb{R} \to \mathbb{R}$ be a score function which fulfill

(C1) ψ is monotone, antisymmetric about the origin, and Lipschitz-continous with Lipschitz-constant C_ψ,

(C2) $E\psi(Z_i - z) < \infty$ for any fixed z,

(C3) ψ is differentiable with $\psi'(0) > 0$.

Note, that (C3) implies that ψ is strictly monotone in an environment of 0.

Further, let h_{1n} and h_{2n} denote the – not necessarily equal – bandwidths which fulfill

(D1) $h_{jn} \to 0$ for $n \to \infty$, $j \in \{1,2\}$,

(D2) there exists a $\widetilde{p}$ with $p > \widetilde{p} > 4$ and $\alpha > 0$ with

$$\frac{h_{1n}^2 + h_{2n}^2}{n^{\frac{1}{4} - \frac{1}{\widetilde{p}} - \alpha} h_{1n}^2 h_{2n}^2} \to 0 \qquad \text{for } n \to \infty,$$

(D3) $0 < q := \lim_{n \to \infty} \frac{h_{1n}}{h_{2n}} < \infty$.

Condition (D1) is classical in the theory of kernel estimators, ensuring that the estimation becomes more and more local. (D2) replaces the classical condition $nh_n \to \infty$, what means, that the estimate is based on more and more observations although the windows becomes smaller. The last condition is needed to avoid that the support of the kernels become arbitrary narrow. Note for further use, that (D2) implies $\sqrt{n} h_{1n}^4 \to \infty$ and $\sqrt{n} h_{2n}^4 \to \infty$.

Consistency at the borders cannot be expected, but on $\left[\,\overline{h_n}\,, 1 - \overline{h_n}\,\right]$ the following two theorems show even uniform consistency.

Theorem 1 *For every $\epsilon > 0$, $\delta > 0$ there exists $N \in \mathbb{N}$ so that for all $n > N$ we have*

$$P\left(\sup_{x_1 \in \left[\,\overline{h_n}\,, 1-\overline{h_n}\,\right]} |C_n(x_1) - C(x_1)| > \delta\right) < \epsilon.$$

Theorem 2 *For every $\epsilon > 0$, $\delta > 0$ there exists $N \in \mathbb{N}$ so that for all $n > N$ we have*

$$P\left(\sup_{x_1 \in \left[\,\overline{h_n}\,, 1-\overline{h_n}\,\right]} \Big|\phi_n(x_1) - \phi(x_1)\Big| > \delta\right) < \epsilon.$$

The proofs of these theorems are based on some lemmas. At first, we need that the objective functions $H_n^{(j)}(z;\theta,x)$ $(j = 1,2)$ converge to certain limit-functions $h_{n,\theta,x}^{(j)}(z)$ which, in general, are mixtures of two functions. The ratio of these mixtures depends on the proportion of weights belonging to observations on each side of the jump curve. Therefore, for $j \in \{1,2\}$, we define the set of indices belonging to observations lying below the jump curve $J_n := \{1 \le i \le n : x_{2i} \le \phi(x_{1i})\}$. With the corresponding sums of weights

$$\lambda_{n,\theta}^{(j)}(x) \;:=\; \textstyle\sum_{i \in J_n} \alpha_{i,n}^{(j)}(\theta, x),$$

we then define the limit functions

$$h_{n,\theta,x}^{(j)}(z) := \int \psi(\mu - z)\left[\lambda_{n,\theta}^{(j)}(x) f(\mu - \widetilde{m}(x)) + (1 - \lambda_{n,\theta}^{(j)}(x)) f(\mu - \widetilde{m}(x) - C(x_1))\right] d\mu.$$

Note, that $h_{n,\theta,x}^{(j)}(z)$ are no "real" limit functions, since they still depend on n. The convergence of $H_n^{(j)}(z;\theta,x)$ is shown by the following two lemmas:

Lemma 1 *For all compact subsets $\mathcal{Z} = [z_L, z_R] \subset \mathbb{R}$, $\epsilon > 0$, $\delta > 0$, and $j \in \{1,2\}$ there exists $N \in \mathbb{N}$ so that for all $n > N$ we have*

$$P\Bigg(\sup_{\substack{z \in \mathcal{Z},\, \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}] \\ x \in [\,\overline{h_n}\,, 1-\overline{h_n}\,]^2}} \Big|H_n^{(j)}(z;\theta,x) - EH_n^{(j)}(z;\theta,x)\Big| > \delta\Bigg) < \epsilon.$$

Lemma 2 *For all $\epsilon > 0$, $\delta > 0$, and $j \in \{1,2\}$ there exists $N \in \mathbb{N}$ so that for all $n > N$ we have*

$$P\Bigg(\sup_{\substack{z \in \mathbb{R},\, \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}] \\ x \in [\,\overline{h_n}\,, 1-\overline{h_n}\,]^2}} \Big|EH_n^{(j)}(z;\theta,x) - h_{n,\theta,x}^{(j)}(z)\Big| > \delta\Bigg) < \epsilon.$$

Independently of x and θ, all zeros of $H_n^{(j)}(z;\theta,x)$ lie between the lower and the upper part of the regression function. Moreover, for x lying below the jump curve, both limit functions are zero for $z=\widetilde{m}(x)$ and for those x lying above the jump curve they vanish for $z=\widetilde{m}(x)+C(x_1)$.

Lemma 3 *With* $L_n := \left[-\frac{\pi}{2},\frac{\pi}{2}\right]\times\left[\overline{h_n},1-\overline{h_n}\right]^2$, $L_{j,n}^0 := \{(\theta,x)\in L_n$ *with* $\lambda_{n,\theta}^{(j)}(x)=0\}$ *and* $L_{j,n}^1 := \{(\theta,x)\in L_n$ *with* $\lambda_{n,\theta}^{(j)}(x)=1\}$ *we have that for all* $\epsilon>0$, $\delta>0$, *and* $j\in\{1,2\}$ *there exists* $N\in\mathbb{N}$ *so that for all* $n>N$ *we have*

(i) $P\left(m_n^{(j)}(\theta,x)\in\left(\widetilde{m}(x)-\delta,\widetilde{m}(x)+C(x_1)+\delta\right)\ \forall\,(\theta,x)\in L_n\right)>1-\epsilon.$

(ii) $P\left(m_n^{(j)}(\theta,x)\in\left(\widetilde{m}(x)-\delta,\widetilde{m}(x)+\delta\right)\ \forall\,(\theta,x)\in L_{j,n}^1\right)>1-\epsilon$ *and*

$P\left(m_n^{(j)}(\theta,x)\in\left(\widetilde{m}(x)+C(x_1)-\delta,\widetilde{m}(x)+C(x_1)+\delta\right)\ \forall(\theta,x)\in L_{j,n}^0\right)>1-\epsilon.$

This lemma already implies that the estimated jump height cannot be larger than the true jump height with high probability.

To show that it even cannot be too small, we need convergence of $h_{n,\theta,x}^{(j)}(z)$ on the jump curve to a "real" limit function $h_{\theta,x}^{(j)}(z)$ which is independent of n. Therefore, with q from Condition (D3), $Q:=\begin{pmatrix}\sqrt{q} & 0\\ 0 & \frac{1}{\sqrt{q}}\end{pmatrix}$, and $j\in\{1,2\}$ let

$$
\begin{aligned}
J_n^{(j)}(\theta,x) &:= \left\{i\in\{1,\dots,n\}:\alpha_{i,n}^{(j)}(\theta,x)\neq 0\right\}\\
T &:= \left\{x\in\mathbb{R}^2:\exists\theta\in\left[-\tfrac{\pi}{2},\tfrac{\pi}{2}\right],\ j\in\{1,2\}\text{ with }K_j(A_\theta x)\neq 0\right\}\\
A_{x_1} &:= \{u\in T: u_2\le\phi'(x_1)u_1\}\\
\lambda_\theta^{(j)}(x) &:= \begin{cases} 1 & x_2<\phi(x_1)\\ \int\limits_{A_{x_1}} K_j(Q^{-1}A_\theta u)du & x_2=\phi(x_1)\\ 0 & x_2>\phi(x_1)\end{cases}\\
h_{\theta,x}^{(j)}(z) &:= \int\psi(\mu-z)\Big[\lambda_\theta^{(j)}(x)f(\mu-\widetilde{m}(x))\\
&\qquad\qquad +(1-\lambda_\theta^{(j)}(x))f(\mu-\widetilde{m}(x)-C(x_1))\Big]\,d\mu.
\end{aligned}
$$

Lemma 4 *For all* $\delta>0$ *and* $j\in\{1,2\}$ *there exists* $N\in\mathbb{N}$ *with*

$$\sup_{\substack{z\in\mathbb{R},\,\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x_1\in[0,1]}}\left|h_{n,\theta,\widetilde{x}}^{(j)}(z)-h_{\theta,\widetilde{x}}^{(j)}(z)\right|\le\delta$$

for all $n>N$ *with* $\widetilde{x}:=(x_1,\phi(x_1))^\top$.

4 Comparisons and applications

To demonstrate the convergence of the robust Rotational Difference M-Kernel Estimator (RDMKE) also in cases where the assumptions for the distribution of the error term are violated, we applied the RDMKE with two different score functions ψ to images with different sample sizes which are blurred by 30% outliers.

We use $\phi(x_1) = 0.3 \cdot \sin(2\pi x_1) + 0.5$ and $C(x_1) = 0.5 \cdot (\log(10 \cdot x_1 + 1) + 2)$ as jump curve and height (Figure 5) and $\widetilde{m}(x) = \varphi_{0.2}(x_1) \cdot \varphi_{0.2}(x_2)$ as smooth part of the image, where $\varphi_\sigma(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2/2\sigma^2}$. The resulting original image $m(x)$ is shown in Figure 6.

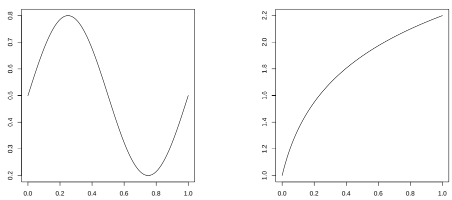


Figure 5: Jump curve $\phi(x_1)$ (left) and jump height $C(x_1)$ (right)

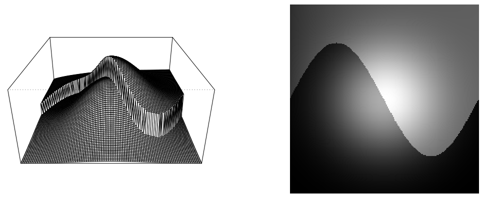


Figure 6: original image $m(x)$

To show the convergence, observations Z_i coming from the true image m overlaid by a "30% salt and pepper noise" are generated for sample sizes $n = 30^2$, $n = 50^2$, $n = 100^2$, and $n = 200^2$ at equidistant design points $x_i = (x_{1i}, x_{2i}) \in [0,1] \times [0,1]$. That means, every observation Z_i has with probability 0.7 the true value of the image at the corresponding design point x_i and with probability 0.3 a uniformly distributed value within the range of the true image, denoted by $\mathcal{D} = \left[0, \max_{x \in [0,1]^2} m(x)\right]$. The according generated images are shown in Figures 7.

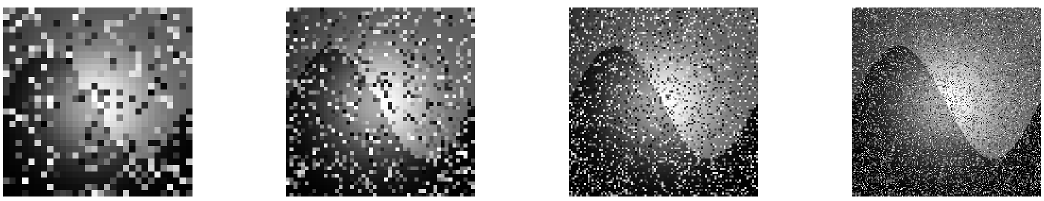


Figure 7: Noisy images for $n = 30^2$, $n = 50^2$, $n = 100^2$, and $n = 200^2$

To each of these images we applied the RDMKE with the identity as first, unbounded score function ψ and the negative derivative of the density of the standard normal distribution as second, bounded score function. In the former case the M-kernel estimator becomes an ordinary non robust kernel estimator. In the latter case the maximization is computed by the Newton-Raphson method with the median as starting value.

The computation can be strongly simplified since, in the case of equidistant design points, results hardly change by using Nadaraya-Watson weights instead of Gasser-Müller weights (see, for example, Eubank, 1988). This means that the piecewise integrals $\alpha_{i,n}^{(j)}(\theta, x)$ can be substituted by the Nadaraya-Watson weights $K_n^{(j)}(\theta, x_i - x) / \sum_{l=1}^{n} K_n^{(j)}(\theta, x_l - x)$. Note that Condition (A2), which is usual for Gasser-Müller kernels, is mainly needed for having the weights summing to one (compare Lemma 5.v in the appendix) what obviously is also fulfilled by the Nadaraya-Watson weights. Thus, this condition is not necessary anymore.

The kernels used were the products of two one-dimensional Gaussian kernels $K_1(x) = K_{\left[-\frac{1}{2},\frac{1}{2}\right]}(x_1) \cdot K_{[-1,0]}(x_2)$ and $K_2(x) = K_{\left[-\frac{1}{2},\frac{1}{2}\right]}(x_1) \cdot K_{[0,1]}(x_2)$ with $\sigma = \frac{1}{2}$, where $K_{[a,b]}(t) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2\sigma^2}} \cdot \mathbb{1}_{[a,b]}(t)$. The bandwidths were $h_{1n} = h_{2n} = \frac{1}{5} n^{-\frac{1}{4}}$.

At first, for every pixel x_i with sufficient distance to the margins, the maximal jump height $\widetilde{C}_n(x_i) = \max_{\theta \in \left[-\frac{\pi}{2},\frac{\pi}{2}\right]} |M_n(\theta, x_i)|$ was determined by calculating $M_n(\theta, x_i)$ for 1000 equidistant angles $\theta_j \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Then, for every x_{1i}, we got $\phi_n(x_{1i})$ as $\phi_n(x_{1i}) \in \operatorname{argmax}_{1 \le j \le n} \widetilde{C}_n\left((x_{1i}, x_{2j})^\top\right)$ and $C_n(x_{1i})$ by inserting the jump curve, $C_n(x_{1i}) = \widetilde{C}_n\left(\left(x_{1i}, \phi_n\left(x_{1i}\right)\right)^\top\right)$.

The plotted results for $\phi_n(x_{1i})$ and $C_n(x_{1i})$ are shown in Figures 8–11, while Tables 1 and 2 show the means of the absolute and quadratic distances of the estimates to the true values.

	absolute distance		quadratic distance	
$\sqrt{n}$	unbounded ψ	bounded ψ	unbounded ψ	bounded ψ
30	0.1205	0.1341	0.0441	0.0439
50	0.1333	0.0704	0.0488	0.0225
100	0.0303	0.0086	0.0086	0.0002
200	0.0147	0.0034	0.0042	0.0000

Table 1: Mean of the absolute and quadratic distances of the estimates $\phi_n(x_{1i})$ to the true values $\phi(x_{1i})$

Note that, for both score functions, the estimated jump curve ϕ_n converges very clearly to the true jump curve ϕ, while the robust version is much better already for $n = 100^2$ than the unrobust version is for $n = 200^2$ (see Table 1 and Figures 8 and 9). Concerning the jump height C, only the robust version converges (see Table 2 and Figure 11), while the unrobust version seems to underestimate systematically the

$\sqrt{n}$	absolute distance unbounded ψ	absolute distance bounded ψ	quadratic distance unbounded ψ	quadratic distance bounded ψ
30	0.6924	1.9941	0.6181	4.2532
50	0.2864	1.1455	0.1285	2.0515
100	0.2713	0.2847	0.1062	0.1266
200	0.4807	0.1546	0.2509	0.0322

Table 2: Mean of the absolute and quadratic distances of the estimates $C_n(x_{1i})$ to the true values $C(x_{1i})$

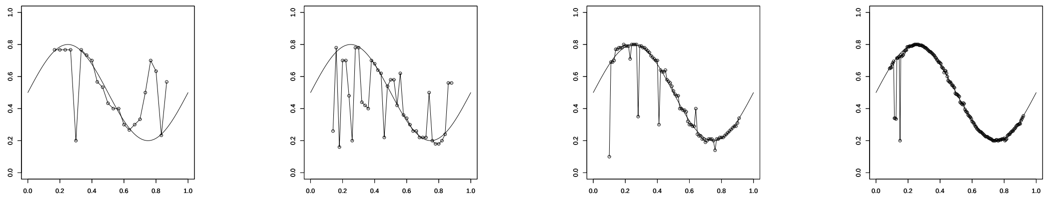


Figure 8: $\phi_n(x_{1i})$ for $n = 30^2$, $n = 50^2$, $n = 100^2$, $n = 200^2$, and unbounded score function

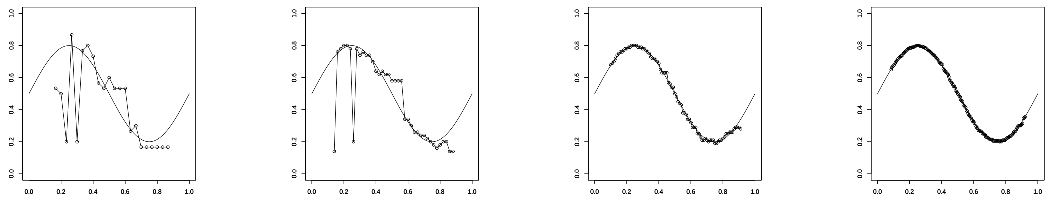


Figure 9: $\phi_n(x_{1i})$ for $n = 30^2$, $n = 50^2$, $n = 100^2$, $n = 200^2$, and bounded score function

true jump height for large sample sizes, although the jump curve is estimated well (see Figures 8 and 10). This can be explained by the special form of the noise used which is not additive but independent of the true pixel value. Therefore, the mean of the disturbed pixels is equal at both sides of the jump curve and consequently, the means of the observations within the windows get closer to each other even if the windows are on different sides of the jump curve. This way, with an amount of 30% disturbed pixels, the original pixels have only an influence of 70% to the estimation, so that the estimated jump height will be reduced to 70% of the true jump height. The robust estimator is less influenced by the outliers so that the jump height can be correctly estimated after all. The effect that both estimators overestimate the jump height for small sample sizes is a result of the maximization process: For x on the true jump curve and θ describing the direction of this curve in x, the jump height can already be estimated by $|M_n(\theta, x)|$ (see Theorem 1). Therefore it is obvious, that C_n which maximizes $|M_n(\theta, x)|$ over θ and x_2 converges from above.

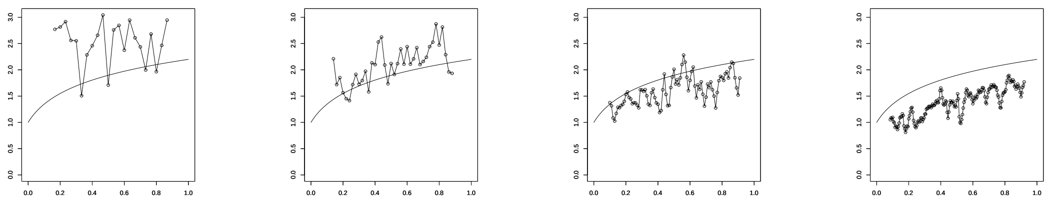


Figure 10: $C_n(x_{1i})$ for $n = 30^2$, $n = 50^2$, $n = 100^2$, $n = 200^2$, and unbounded score function

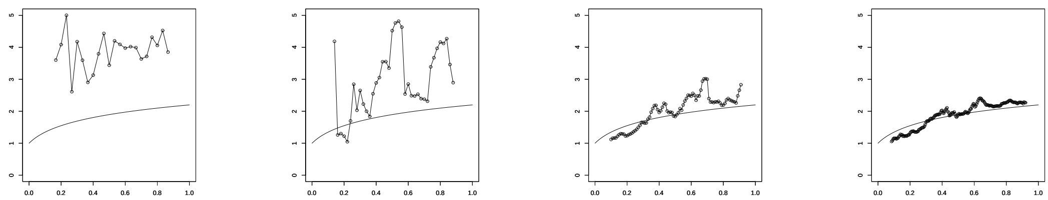


Figure 11: $C_n(x_{1i})$ for $n = 30^2$, $n = 50^2$, $n = 100^2$, $n = 200^2$, and bounded score function

Appendix: Proofs

The following lemma gives some properties of the design points and the kernels and weights.

Lemma 5 *For $j \in \{1,2\}$ we have*

(i) $||x_i - x|| = O\left(\overline{h_n}\right)$ *and* $|x_{1i} - x_1| = O\left(\overline{h_n}\right)$ *for all* $\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, $x \in [0,1]^2$, $n \in \mathbb{N}$, *and* $i \in J_n^{(j)}(\theta, x)$,

(ii) $S(\Delta_i) = O\left(\frac{1}{n}\right)$ *and* $\frac{1}{S(\Delta_i)} = O(n)$,

(iii) $\sup_{\substack{\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}] \\ x \in [0,1]^2}} \# J_n^{(j)}(\theta, x) = O\left(n\left(h_{1n}^2 + h_{2n}^2\right)\right)$.

(iv) $T \subset [-2,2]^2$,

(v) $\sum_{i \in J_n^{(j)}(\theta,x)} \alpha_{i,n}^{(j)}(\theta, x) = 1$ *for all* $\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, $x \in \left[\overline{h_n}, 1 - \overline{h_n}\right]^2$ *and* $n \in \mathbb{N}$,

(vi) $\sup_{\substack{\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}] \\ x \in [0,1]^2}} \alpha_{i,n}^{(j)}(\theta, x) = O\left(\frac{1}{n h_{1n} h_{2n}}\right)$,

(vii) $\left|\left(K_n^{(j)}(\theta, x)\right)^p - \left(K_n^{(j)}(\theta, \widetilde{x})\right)^p\right| \leq C_{K^p} \frac{1}{h_{1n}^{p+1} h_{2n}^{p+1}} \cdot ||x - \widetilde{x}||$ *and*

$\left|K_n^{(j)}(\theta, x) - K_n^{(j)}(\widetilde{\theta}, x)\right| \leq C_K \frac{\overline{h_n}}{h_{1n}^2 h_{2n}^2} \cdot (|x_1| + |x_2|) \cdot |\theta - \widetilde{\theta}|$

for all $\theta, \widetilde{\theta} \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, $x, \widetilde{x} \in \mathbb{R}^2$ *and* $p, n \in \mathbb{N}$,

(viii) $\sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x\in[\overline{h_n},1-\overline{h_n}]^2}} \left|\sum_{i=1}^n \left(\alpha_{i,n}^{(j)}(\theta,x)\right)^2 - \frac{1}{n}\int \left(K_n^{(j)}(\theta,u)\right)^2 du\right| = O\left(\frac{h_{1n}^2+h_{2n}^2}{nh_{1n}^2h_{2n}^2}\right)$

and $\sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x\in[\overline{h_n},1-\overline{h_n}]^2}} \left|\sum_{i=1}^n \left(\alpha_{i,n}^{(j)}(\theta,x)\right)^2\right| = O\left(\frac{h_{1n}^2+h_{2n}^2}{nh_{1n}^2h_{2n}^2}\right)$,

Proof

(i,ii,vi) Follow from Condition (B1) and (B2).

(iii) With

$$\begin{aligned} A_n^{(j)}(\theta,x) &:= \left\{u\in[0,1]^2 : K_n^{(j)}(\theta,u-x)\neq 0\right\} \quad \text{and} \\ \widetilde{A}_n^{(j)}(\theta,x) &:= U_{D_n}\left(A_n^{(j)}(\theta,x)\right) \end{aligned}$$

we have

$$\widetilde{A}_n^{(j)}(\theta,x) \subset \left[x_1-\overline{h_n}-D_n,\ x_1+\overline{h_n}+D_n\right]\times\left[x_2-\overline{h_n}-D_n,\ x_2+\overline{h_n}+D_n\right]$$

for all $\theta\in\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ and $x\in[0,1]^2$. For $i\in J_n^{(j)}(\theta,x)$ we have $\Delta_i\cap A_n^{(j)}(\theta,x)\neq\emptyset$ and therefore $\bigcup_{i\in J_n^{(j)}(\theta,x)}\Delta_i\subset\widetilde{A}_n^{(j)}(\theta,x)$ for all $\theta\in\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ and $x\in[0,1]^2$. From $\#J_n^{(j)}(\theta,x)=\sum_{i\in J_n^{(j)}(\theta,x)}\frac{S(\Delta_i)}{S(\Delta_i)} = O(n)\cdot S\left(\bigcup_{i\in J_n^{(j)}(\theta,x)}\Delta_i\right)$ the claim follows.

(iv) Follows from Condition (A1).

(v) Follows from Condition (A2).

(vii) Follows from the Lipschitz-continuity of K_j.

(viii) With (ii) and (vii) we have

$$\begin{aligned} &\sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x\in[\overline{h_n},1-\overline{h_n}]^2}} \left|\sum_{i=1}^n \left(\alpha_{i,n}^{(j)}(\theta,x)\right)^2 - \frac{1}{n}\int \left(K_n^{(j)}(\theta,u)\right)^2 du\right| \\ =\ &\sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x\in[\overline{h_n},1-\overline{h_n}]^2}} \left|\sum_{i\in J_n^{(j)}(\theta,x)} \left(\int_{\Delta_i} K_n^{(j)}(\theta,u-x)du\right)^2 - \frac{1}{n}\int_{\Delta_i}\left(K_n^{(j)}(\theta,u-x)\right)^2 du\right| \end{aligned}$$

$$
\begin{aligned}
&\leq \sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}] \\ x\in[\overline{h_n},1-\overline{h_n}]^2}} \sum_{i\in J_n^{(j)}(\theta,x)} S(\Delta_i)^2 \sup_{\xi_i,\zeta_i\in\Delta_i} \left| \left(K_n^{(j)}(\theta,\xi_i - x)\right)^2 - \left(K_n^{(j)}(\theta,\zeta_i - x)\right)^2 \right| \\
&\qquad + S(\Delta_i) \left| S(\Delta_i) - \frac{1}{n} \right| \sup_{\zeta_i\in\Delta_i} \left| \left(K_n^{(j)}(\theta,\zeta_i - x)\right)^2 \right| \\
&\leq \sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}] \\ x\in[\overline{h_n},1-\overline{h_n}]^2}} \sum_{i\in J_n^{(j)}(\theta,x)} \left(\frac{C_1^2}{n^2} C_{K^2} \frac{1}{h_{1n}^3 h_{2n}^3} \sup_{\xi_i,\zeta_i\in\Delta_i} ||\xi_i - \zeta_i|| + \frac{C_1}{n}\frac{C_2}{n} \frac{\max_x K_j^2(x)}{h_{1n}^2 h_{2n}^2} \right) \\
&= O\left(\frac{h_{1n}^2 + h_{2n}^2}{n h_{1n}^2 h_{2n}^2} \right).
\end{aligned}
$$

The second claim follows with $\int \left(K_n^{(j)}(\theta,u)\right)^2 du = O\left(\frac{1}{h_{1n}h_{2n}}\right)$.

Proof of Lemma 1 With $\widetilde{p}$ from Condition (D2) and $\overline{\epsilon_i} := \epsilon_i \mathbb{1}_{\left\{|\epsilon_i|\leq i^{1/\widetilde{p}}\right\}}$, $\overline{Z_i} := m(x_i) + \overline{\epsilon_i}$, we have for every $\delta > 0$

$$
\begin{aligned}
& P\left(\sup_{\substack{z\in\mathcal{Z},\,\theta\in[-\frac{\pi}{2},\frac{\pi}{2}] \\ x\in[\overline{h_n},1-\overline{h_n}]^2}} \left| H_n^{(j)}(z;\theta,x) - EH_n^{(j)}(z;\theta,x) \right| > \delta \right) \\
\leq\ & P\Big(S_{1,n}^{(j)} + S_{2,n}^{(j)} + S_{3,n}^{(j)} > \delta\Big) \leq P\Big(S_{1,n}^{(j)} > \delta/3\Big) + P\Big(S_{2,n}^{(j)} > \delta/3\Big) + P\Big(S_{3,n}^{(j)} > \delta/3\Big)
\end{aligned}
$$

with

$$
\begin{aligned}
S_{1,n}^{(j)} &:= \sup_{\substack{z\in\mathcal{Z},\,\theta\in[-\frac{\pi}{2},\frac{\pi}{2}] \\ x\in[\overline{h_n},1-\overline{h_n}]^2}} \left| \sum_{i=1}^n \alpha_{i,n}^{(j)}(\theta,x) \Big(\psi\left(Z_i - z\right) - \psi\left(\overline{Z_i} - z\right) \Big) \right| \\
S_{2,n}^{(j)} &:= \sup_{\substack{z\in\mathcal{Z},\,\theta\in[-\frac{\pi}{2},\frac{\pi}{2}] \\ x\in[\overline{h_n},1-\overline{h_n}]^2}} \left| \sum_{i=1}^n \alpha_{i,n}^{(j)}(\theta,x) \Big(\psi\left(\overline{Z_i} - z\right) - E\left(\psi\left(\overline{Z_i} - z\right)\right) \Big) \right| \\
S_{3,n}^{(j)} &:= \sup_{\substack{z\in\mathcal{Z},\,\theta\in[-\frac{\pi}{2},\frac{\pi}{2}] \\ x\in[\overline{h_n},1-\overline{h_n}]^2}} \left| \sum_{i=1}^n \alpha_{i,n}^{(j)}(\theta,x) \Big(E\left(\psi\left(\overline{Z_i} - z\right)\right) - E\left(\psi\left(Z_i - z\right)\right) \Big) \right|.
\end{aligned}
$$

We now show the convergence of $S_{1,n}^{(j)}$, $S_{2,n}^{(j)}$, and $S_{3,n}^{(j)}$.

$\boldsymbol{S_{1,n}^{(j)}}$ **:**With Lemma 5.vi we have

$$
\begin{aligned}
S_{1,n}^{(j)} &\leq \sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}] \\ x\in[\overline{h_n},1-\overline{h_n}]^2}} \sum_{i=1}^n \alpha_{i,n}^{(j)}(\theta,x) C_\psi \left|\epsilon_i - \overline{\epsilon_i}\right| \\
&\leq O\left(\frac{1}{n h_{1n} h_{2n}}\right) \lim_{m\to\infty} \sum_{i=1}^m \left|\epsilon_i - \overline{\epsilon_i}\right|
\end{aligned}
$$

With Markov's inequality we get $P\left(|\epsilon_i - \overline{\epsilon_i}| > 0\right) = P\left(|\epsilon_i| > i^{1/\widetilde{p}}\right) < \frac{E|\epsilon_i|^p}{\left(i^{1/\widetilde{p}}\right)^p} < \frac{M}{i^{p/\widetilde{p}}}$ and since $p > \widetilde{p}$, it follows $\sum_{i=1}^{\infty} P\left(|\epsilon_i - \overline{\epsilon_i}| > 0\right) < \infty$.

Therefore with the Borel-Cantelli Lemma we get

$$\begin{aligned} P\left(\lim_{n\to\infty} S_{1,n}^{(j)} = 0\right) &\geq P\left(\lim_{n\to\infty} \sum_{i=1}^{n} |\epsilon_i - \overline{\epsilon_i}| < \infty\right) \\ &\geq P\left(\bigcup_{n=1}^{\infty} \bigcap_{i\geq n} \{|\epsilon_i - \overline{\epsilon_i}| = 0\}\right) = 1. \end{aligned}$$

$\boldsymbol{S_{2,n}^{(j)}}$ **:**We split $S_{2,n}^{(j)}$ into three parts, $S_{2,n}^{(j)} \leq T_{1,n}^{(j)} + T_{2,n}^{(j)} + T_{3,n}^{(j)}$, with

$$\begin{aligned} T_{1,n}^{(j)} &= \sup_{\substack{z\in\mathcal{Z},\,\theta\in[-\frac{\pi}{2},\frac{\pi}{2}] \\ x\in[\overline{h_n},1-\overline{h_n}]^2}} \left| \sum_{i=1}^{n} \left[\alpha_{i,n}^{(j)}(\theta,x)\psi\left(\overline{Z_i} - z\right) - \alpha_{i,n}^{(j)}\big(t(\theta),r(x)\big)\psi\left(\overline{Z_i} - u(z)\right) \right] \right| \\ T_{2,n}^{(j)} &= \sup_{\substack{z\in\mathcal{Z},\,\theta\in[-\frac{\pi}{2},\frac{\pi}{2}] \\ x\in[\overline{h_n},1-\overline{h_n}]^2}} \left| \sum_{i=1}^{n} \alpha_{i,n}^{(j)}\big(t(\theta),r(x)\big) \left[\psi\left(\overline{Z_i} - u(z)\right) - E\left(\psi\left(\overline{Z_i} - u(z)\right)\right) \right] \right| \\ T_{3,n}^{(j)} &= \sup_{\substack{z\in\mathcal{Z},\,\theta\in[-\frac{\pi}{2},\frac{\pi}{2}] \\ x\in[\overline{h_n},1-\overline{h_n}]^2}} \left| \sum_{i=1}^{n} \left[\alpha_{i,n}^{(j)}\big(t(\theta),r(x)\big) E\left(\psi\left(\overline{Z_i} - u(z)\right)\right) - \alpha_{i,n}^{(j)}(\theta,x) E\left(\psi\left(\overline{Z_i} - z\right)\right) \right] \right|, \end{aligned}$$

where, for α from Condition (D2),

$$\begin{aligned} t(\theta) \in T_n &:= \left\{ \tfrac{k\pi}{n^{1/4}} - \tfrac{\pi}{2}; k \in \left\{0,\ldots,\lfloor n^{1/4}\rfloor\right\} \right\} &&\text{with} \quad |t(\theta) - \theta| \leq \tfrac{\pi}{2n^{1/4}} \\ r(x) \in R_n &:= \left\{ \left(\tfrac{k}{n^{1/4}}, \tfrac{l}{n^{1/4}}\right)^{\top}; k,l \in \left\{0,\ldots,\lfloor n^{1/4}\rfloor\right\} \right\} &&\text{with} \quad ||r(x) - x|| \leq \tfrac{1}{\sqrt{2}n^{1/4}} \\ u(z) \in U_n &:= \left\{ z_L + \tfrac{k}{n^{\alpha}}\left(z_R - z_L\right); k \in \{0,\ldots,\lfloor n^{\alpha}\rfloor\} \right\} &&\text{with} \quad |u(z) - z| \leq \tfrac{z_R - z_L}{2n^{\alpha}} \end{aligned}$$

for all θ, x, and z.

With the Lipschitz-continuity of ψ and with $\psi(0) = 0$ we have

$$\begin{aligned} \sup_{z\in\mathcal{Z}} \left|\psi\left(\overline{Z_i} - u(z)\right)\right| &= \sup_{z\in\mathcal{Z}} \left|\psi(\overline{Z_i} - u(z)) - \psi(0)\right| \leq \sup_{z\in\mathcal{Z}} C_{\psi} \left|m(x_i) + \overline{\epsilon_i} - u(z)\right| \\ &\leq C_{\psi} \left(\left| \max_{x} \{m(x)\} + i^{1/\widetilde{p}} \right| + |z_L| + |z_R| \right) = O\left(n^{1/\widetilde{p}}\right) \qquad (2) \end{aligned}$$

From Lemma 5.vii we get $\left|\alpha_{i,n}^{(j)}(\theta,x)-\alpha_{i,n}^{(j)}(t(\theta),r(x))\right| = O\left(\frac{1}{n}\frac{1}{n^{1/4}h_{1n}^2h_{2n}^2}\right)$ and since $\left|\alpha_{i,n}^{(j)}(\theta,x)-\alpha_{i,n}^{(j)}(t(\theta),r(x))\right| = 0$ for $i \notin J_n^{(j)}(\theta,x)\cup J_n^{(j)}(t(\theta),r(x))$ it follows

$$
\begin{aligned}
T_{1,n}^{(j)} &\leq \sup_{z,\theta,x}\left|\sum_{i=1}^n \alpha_{i,n}^{(j)}(\theta,x)\Big(\psi\left(\overline{Z_i}-z\right)-\psi\left(\overline{Z_i}-u(z)\right)\Big)\right| + \\
&\quad \sup_{z,\theta,x}\left|\sum_{i=1}^n \Big(\alpha_{i,n}^{(j)}(\theta,x)-\alpha_{i,n}^{(j)}\big(t(\theta),r(x)\big)\Big)\psi\left(\overline{Z_i}-u(z)\right)\right| \\
&\leq \sup_{z,\theta,x}\sum_{i=1}^n \alpha_{i,n}^{(j)}(\theta,x)C_\psi\left|z-u(z)\right| + \\
&\quad \sup_{z,\theta,x}\sum_{i\in(J_n^{(j)}(\theta,x)\cup J_n^{(j)}(t(\theta),r(x)))}\left|\alpha_{i,n}^{(j)}(\theta,x)-\alpha_{i,n}^{(j)}(t(\theta),r(x))\right|\cdot\left|\psi\left(\overline{Z_i}-u(z)\right)\right| \\
&\leq C_\psi\frac{z_R-z_L}{2n^\alpha}+\left(\#J_n^{(j)}(\theta,x)+\#J_n^{(j)}(t(\theta),r(x))\right)\cdot O\left(\frac{1}{n}\frac{1}{n^{1/4}h_{1n}^2h_{2n}^2}\right)\cdot O\left(n^{1/\widetilde{p}}\right) \\
&= O\left(\frac{1}{n^\alpha}\right)+O\left(\frac{h_{1n}^2+h_{2n}^2}{n^{\frac{1}{4}-\frac{1}{\widetilde{p}}}h_{1n}^2h_{2n}^2}\right)\to 0.
\end{aligned}
$$

Since also $\sup_{z\in\mathcal{Z}}\left|E\psi\left(\overline{Z_i}-u(z)\right)\right| = O\left(n^{1/\widetilde{p}}\right)$ the same result holds for $T_{3,n}^{(j)}$.

Since ψ^2 is Lipschitz-continous on $\mathcal{Z}$ we have as in (2) $\sup_{z\in\mathcal{Z}}\left|\psi^2\left(\overline{Z_i}-z\right)\right| = O\left(n^{1/\widetilde{p}}\right)$ and therefore $\sup_{z\in\mathcal{Z}} E\psi^2\left(\overline{Z_i}-z\right) = O\left(n^{1/\widetilde{p}}\right)$.

Now we have with Lemma 5.viii

$$
\begin{aligned}
&\sup_{\substack{z\in\mathcal{Z},\,\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x\in[\overline{h_n},1-\overline{h_n}]^2}} \operatorname{Var}\left(\sum_{i=1}^n \alpha_{i,n}^{(j)}(\theta,x)\psi\left(\overline{Z_i}-z\right)\right) \\
\leq\ &\sup_{\substack{z\in\mathcal{Z},\,\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x\in[\overline{h_n},1-\overline{h_n}]^2}} \sum_{i=1}^n\left(\alpha_{i,n}^{(j)}(\theta,x)\right)^2\left(E\psi^2\left(\overline{Z_i}-z\right)\right) \leq O\left(\frac{h_{1n}^2+h_{2n}^2}{nh_{1n}^2h_{2n}^2}\right)\cdot O\left(n^{1/\widetilde{p}}\right)
\end{aligned}
$$

and therefore,

$$
\begin{aligned}
&P\left(T_{2,n}^{(j)}>\delta/9\right) \\
=\ &P\left(\exists t\in T_n, r\in R_n, u\in U_n : \left|\sum_{i=1}^n \alpha_{i,n}^{(j)}(t,r)\Big(\psi\left(\overline{Z_i}-u\right)-E\left(\psi\left(\overline{Z_i}-u\right)\right)\Big)\right|>\delta/9\right)
\end{aligned}
$$

$$\leq \sum_{t,r,u} P\left(\left|\sum_{i=1}^{n} \alpha_{i,n}^{(j)}(t,r)\Big(\psi\left(\overline{Z_i}-u\right) - E\left(\psi\left(\overline{Z_i}-u\right)\right)\Big)\right| > \delta/9\right)$$

$$\leq \sum_{t,r,u} \frac{\mathrm{Var}\left(\sum_{i=1}^{n} \alpha_{i,n}^{(j)}(t,r)\psi\left(\overline{Z_i}-u\right)\right)}{(\delta/9)^2}$$

$$\leq (n^{1/4}+1)\cdot(n^{1/4}+1)^2\cdot(n^{\alpha}+1)\cdot O\left(\frac{h_{1n}^2+h_{2n}^2}{nh_{1n}^2h_{2n}^2}\right)\cdot O\left(n^{1/\widetilde{p}}\right)$$

$$= O\left(\frac{h_{1n}^2+h_{2n}^2}{n^{\frac{1}{4}-\frac{1}{\widetilde{p}}-\alpha}h_{1n}^2h_{2n}^2}\right) \to 0.$$

Finally, there exists $N \in \mathbb{N}$, so that for all $n > N$ we have $P\big(S_{2,n}^{(j)} > \delta/3\big) \leq P\big(T_{1,n}^{(j)} > \delta/9\big) + P\big(T_{2,n}^{(j)} > \delta/9\big) + P\big(T_{3,n}^{(j)} > \delta/9\big) < \frac{\epsilon}{3}$.

$\boldsymbol{S_{3,n}^{(j)}}$ **:**Since $|\epsilon_i| > i^{\frac{1}{\widetilde{p}}} \Leftrightarrow |\epsilon_i|^{p-1} > i^{\frac{p-1}{\widetilde{p}}} \Leftrightarrow |\epsilon_i|^{p-1}\cdot i^{-\frac{p-1}{\widetilde{p}}} > 1$ it follows that

$$E|\epsilon_i - \overline{\epsilon}_i| = \int_{|\epsilon_i|>i^{\frac{1}{\widetilde{p}}}} |\epsilon_i| dP(\epsilon_i) \leq \int_{|\epsilon_i|>i^{\frac{1}{\widetilde{p}}}} i^{-\frac{p-1}{\widetilde{p}}}|\epsilon_i|^p dP(\epsilon_i) < i^{-\frac{p-1}{p}}\cdot E|\epsilon_i|^p$$

and

$$\sum_{i=1}^{n} E|\epsilon_i - \overline{\epsilon}_i| < \sum_{i=1}^{n} n^{-\frac{p-1}{p}} E\left|\epsilon_1\right|^p = n^{\frac{1}{p}} E\left|\epsilon_1\right|^p < n^{1/4}M = O\left(n^{1/4}\right).$$

Therefore, we have

$$S_{3,n}^{(j)} \leq O\left(\tfrac{1}{nh_{1n}h_{2n}}\right)\sup_{z\in\mathcal{Z}}\textstyle\sum_{i=1}^{n} E\left|\psi\left(\overline{Z_i}-z\right)-\psi\left(Z_i-z\right)\right|$$

$$\leq O\left(\tfrac{1}{nh_{1n}h_{2n}}\right)C_{\psi}\textstyle\sum_{i=1}^{n} E|\epsilon_i-\overline{\epsilon_i}| = O\left(\tfrac{1}{n^{3/4}h_{1n}h_{2n}}\right)\to 0$$

and the claim follows.

Proof of Lemma 2 With Lemma 5.i and 5.v we have for all $z \in \mathbb{R}$, $\theta \in \left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ and $x \in \left[\,\overline{h_n}\,, 1-\overline{h_n}\,\right]$

$$\left|EH_n^{(j)}(z;\theta,x) - h_{n,\theta,x}^{(j)}(z)\right|$$

$$= \Big|\textstyle\sum_{i\in J_n}\alpha_{i,n}^{(j)}(\theta,x)\int\psi(\mu-z)f(\mu-m(x_i))d\mu +$$

$$\sum_{\substack{i=1\\ x_{2i}>\phi(x_{1i})}}^{n}\alpha_{i,n}^{(j)}(\theta,x)\int\psi(\mu-z)f(\mu-m(x_i))d\mu$$

$$-\lambda_{n,\theta}^{(j)}(x)\int\psi(\mu-z)f(\mu-\widetilde{m}(x))d\mu$$

$$-(1-\lambda_{n,\theta}^{(j)}(x))\int\psi(\mu-z)f(\mu-\widetilde{m}(x)-C(x_1))d\mu\Big|$$

$$
\begin{aligned}
&= \Big|\textstyle\sum_{i\in J_n} \alpha_{i,n}^{(j)}(\theta,x)\int\Big[\psi(\mu+\widetilde{m}(x_i)-z)-\psi(\mu+\widetilde{m}(x)-z)\Big]f(\mu)d\mu \\
&\quad + \textstyle\sum_{\substack{i=1\\ x_{2i}>\phi(x_{1i})}}^{n} \alpha_{i,n}^{(j)}(\theta,x)\int\Big[\psi(\mu+\widetilde{m}(x_i)+C(x_{1i})-z) \\
&\qquad\qquad -\psi(\mu+\widetilde{m}(x)+C(x_1)-z)\Big]f(\mu)d\mu\Big| \\
&\le \textstyle\sum_{\substack{i\in J_n^{(j)}(\theta,x)\\ i\in J_n}} \alpha_{i,n}^{(j)}(\theta,x)\int C_\psi C_{\widetilde{m}}||x_i-x||f(\mu)d\mu \\
&\quad + \textstyle\sum_{\substack{i\in J_n^{(j)}(\theta,x)\\ x_{2i}>\phi(x_{1i})}} \alpha_{i,n}^{(j)}(\theta,x)\int C_\psi\left[C_{\widetilde{m}}||x_i-x||+C_C|x_{1i}-x_1|\right]f(\mu)d\mu \\
&\le \widetilde{C}\cdot\textstyle\sum_{i=1}^n \alpha_{i,n}^{(j)}(\theta,x)\cdot\int f(\mu)d\mu\cdot O\left(\overline{h_n}\right) \\
&= O\left(\overline{h_n}\right)\to 0
\end{aligned}
$$

with $\widetilde{C}=C_\psi(C_{\widetilde{m}}+C_C)$ independent of θ and x. Therefore the claim follows.

Proof of Lemma 3 First, let $h(z):=\int\psi(\mu-z)f(\mu)d\mu$ and $\epsilon>0$ and $\delta>0$ be arbitrary. Since ψ is monotone increasing, h is monotone decreasing. The symmetry of f and the strict monotony in 0 and the antisymmetry of ψ imply $h(0)=0$ and the existence of a $\delta'>0$ with $h(z)<-\delta'$ for $z\ge\delta$ and $h(z)>\delta'$ for $z\le-\delta$. Since

$$
h_{n,\theta,x}^{(j)}(z) \;=\; \lambda_{n,\theta}^{(j)}(x)h(z-\widetilde{m}(x))+\big(1-\lambda_{n,\theta}^{(j)}(x)\big)h(z-\widetilde{m}(x)-C(x_1)) \tag{3}
$$

it follows, that

$$
\begin{aligned}
h_{n,\theta,x}^{(j)}(\widetilde{m}(x)-\delta) &= \lambda_{n,\theta}^{(j)}(x)h(-\delta)+\big(1-\lambda_{n,\theta}^{(j)}(x)\big)h(-\delta-C(x_1)) > \delta' \\
h_{n,\theta,x}^{(j)}(\widetilde{m}(x)+C(x_1)+\delta) &= \lambda_{n,\theta}^{(j)}(x)h(C(x_1)+\delta)+\big(1-\lambda_{n,\theta}^{(j)}(x)\big)h(\delta) < -\delta'
\end{aligned}
$$

for all $n\in\mathbb{N}$, $(\theta,x)\in L_n$.

Let $\mathcal{M}_n^{(j)}(\theta,x) := \left\{z\in\mathbb{R}: H_n^{(j)}(z;\theta,x)=0\right\} = \left(H_n^{(j)}\right)^{-1}(0;\theta,x)$, that means $m_n^{(j)}(\theta,x)\in\mathcal{M}_n^{(j)}(\theta,x)$. With K_j also the weights $\alpha_{i,n}^{(j)}(\theta,x)$ are nonnegative and therefore the monotony of ψ implies that $H_n^{(j)}(z;\theta,x)$ is monotone decreasing in z. Since ψ is even strictly monotone in an environment of 0 there exist real numbers $M_{1,n;\theta,x}^{(j)}$ and $M_{2,n;\theta,x}^{(j)}$ with $H_n^{(j)}(z;\theta,x)>0$ for $z\le M_{1,n;\theta,x}^{(j)}$ and $H_n^{(j)}(z;\theta,x)<0$ for $z\ge M_{2,n;\theta,x}^{(j)}$. Since $H_n^{(j)}(z;\theta,x)$ is continuous this implies the existence of a $z_{0,n;\theta,x}^{(j)}\in\mathbb{R}$ with $H_n^{(j)}\left(z_{0,n;\theta,x}^{(j)};\theta,x\right)=0$.

Now, define $a^{(j)}_{n;\theta,x} := \sup\{z \in \mathbb{R} : H^{(j)}_n(z;\theta,x) > 0\}$ and $bn := \inf\{z \in \mathbb{R} : H^{(j)}_n(z;\theta,x) < 0\}$. Obviously, these fulfill $a^{(j)}_{n;\theta,x} \leq z^{(j)}_{0,n;\theta,x} \leq b^{(j)}_{n;\theta,x}$ and $\left(a^{(j)}_{n;\theta,x}, b^{(j)}_{n;\theta,x}\right) \subset \mathcal{M}^{(j)}_n(\theta,x) \subset \left[a^{(j)}_{n;\theta,x}, b^{(j)}_{n;\theta,x}\right]$. But since $\mathcal{M}^{(j)}_n(\theta,x)$ is the inverse image of the compact set $\{0\}$ and $H^{(j)}_n(z;\theta,x)$ is continuous, it follows that $\mathcal{M}^{(j)}_n(\theta,x) = \left[a^{(j)}_{n;\theta,x}, b^{(j)}_{n;\theta,x}\right]$.

Since the set $\mathcal{Z} := \left[\min_{x\in[0,1]^2}\{\widetilde{m}(x)-\delta\}, \max_{x\in[0,1]^2}\{\widetilde{m}(x)+C(x_1)+\delta\}\right]$ is compact, with Lemma 1 and 2 there exists $N \in \mathbb{N}$ with

$$
\begin{aligned}
& P\left(\exists(\theta,x) \in L_n \,:\, m^{(j)}_n(\theta,x) \notin \left(\widetilde{m}(x)-\delta, \widetilde{m}(x)+C(x_1)+\delta\right)\right) \\
\leq\; & P\left(\exists(\theta,x) \in L_n \,:\, a^{(j)}_{n;\theta,x} \leq \widetilde{m}(x)-\delta\right) + \\
& P\left(\exists(\theta,x) \in L_n \,:\, b^{(j)}_{n;\theta,x} \geq \widetilde{m}(x)+C(x_1)+\delta\right) \\
\leq\; & P\left(\exists(\theta,x) \in L_n \,:\, H^{(j)}_n(\widetilde{m}(x)-\delta;\theta,x) \leq 0\right) + \\
& P\left(\exists(\theta,x) \in L_n \,:\, H^{(j)}_n(\widetilde{m}(x)+C(x_1)+\delta;\theta,x) \geq 0\right) \\
\leq\; & P\left(\exists(\theta,x) \in L_n \,:\, H^{(j)}_n(\widetilde{m}(x)-\delta;\theta,x) - h^{(j)}_{n,\theta,x}(\widetilde{m}(x)-\delta) < -\delta'\right) + \\
& P\left(\exists(\theta,x) \in L_n \,:\, H^{(j)}_n(\widetilde{m}(x)+C(x_1)+\delta;\theta,x) - h^{(j)}_{n,\theta,x}(\widetilde{m}(x)+C(x_1)+\delta) > \delta'\right) \\
\leq\; & 2 \cdot P\left(\sup_{\substack{z\in\mathcal{Z},\,\theta\in[-\frac{\pi}{2},\frac{\pi}{2}] \\ x\in[\overline{h_n},1-\overline{h_n}]^2}} \left|H^{(j)}_n(z;\theta,x) - h^{(j)}_{n,\theta,x}(z)\right| > \delta'\right) \\
<\; & \epsilon
\end{aligned}
$$

for all $n > N$.

We get (ii) in the same way, taking note of the fact that in (3) for $\lambda^{(j)}_{n,\theta}(x) = 1$ we have $h^{(j)}_{n,\theta,x}(z) = h(z-\widetilde{m}(x))$ and consequently $h^{(j)}_{n,\theta,x}(\widetilde{m}(x)-\delta) = h(-\delta) > \delta'$ and $h^{(j)}_{n,\theta,x}(\widetilde{m}(x)+\delta) = h(\delta) < -\delta'$ for all $n \in \mathbb{N}$, $(\theta,x) \in L^1_{j,n}$, and $j = \{1,2\}$. Analogously, for $\lambda^{(j)}_{n,\theta}(x) = 0$ we have $h^{(j)}_{n,\theta,x}(z) = h(z-\widetilde{m}(x)-C(x_1))$, so that $h^{(j)}_{n,\theta,x}(\widetilde{m}(x)+C(x_1)-\delta) = h(-\delta) > \delta'$ and $h^{(j)}_{n,\theta,x}(\widetilde{m}(x)+C(x_1)+\delta) = h(\delta) < -\delta'$ for all $n \in \mathbb{N}$, $(\theta,x) \in L^0_{j,n}$, and $j = \{1,2\}$.

Before we can show the convergence of $h^{(j)}_{n,\theta,x}(z)$ we need the following asymptotic behavior of $\lambda^{(j)}_{n,\theta}(x)$:

Lemma 6 *For $j \in \{1,2\}$ we have*

$$\sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x\in[0,1]^2}} \left|\lambda^{(j)}_{n,\theta}(x) - \int\limits_{A^C} K^{(j)}_n(\theta, u-x)du\right| \;=\; O\left(\frac{1}{\sqrt{n}h_{1n}h_{2n}}\right).$$

Proof Let $A_n := \bigcup_{i\in J_n} \Delta_i$, $A^{(1)}_n := \{u \in [0,1]^2 : \forall \widetilde{u} \in A : ||u - \widetilde{u}|| > D_n\}$, and $A^{(2)}_n := \{u \in [0,1]^2 : \exists \widetilde{u} \in A^C : ||u - \widetilde{u}|| \leq D_n\}$. Obviously $A^{(1)}_n \subset A^C \subset A^{(2)}_n$. As well we have $A^{(1)}_n \subset A_n \subset A^{(2)}_n$, since for any $u \in [0,1]^2$ there exists $1 \leq i \leq n$ with $u \in \Delta_i$, what means $||u - x_i|| < D_n$. For $u \in A^{(1)}_n$ we have $x_i \in A^C$ for this i, what means $i \in J_n$ and therefore $u \in A_n$. For $u \in A_n$ we have $i \in J_n$ respectively and therefore $x_i \in A^C$, what means $u \in A^{(2)}_n$. Consequently, we have $S\left(A_n \triangle A^C\right) \leq S\big(A^{(2)}_n \setminus A^{(1)}_n\big)$.

Further, for $u \in A^{(2)}_n$ and suitable $\widetilde{u} \in A^C$ we have

$$u_2 \leq \widetilde{u}_2 + D_n \leq \phi(\widetilde{u}_1) + D_n \leq \phi(u_1) + \max_{\xi\in[0,1]} |\phi'(\xi)| \cdot |\widetilde{u}_1 - u_1| + D_n \leq \phi(u_1) + C \cdot D_n.$$

Since $A^{(1)}_n \subset A^C$ we have $u_2 \leq \phi(u_1)$ for $u \in A^{(1)}_n$, and therefore

$$S\left(A^{(2)}_n \setminus A^{(1)}_n\right) \leq \int_0^1 (\phi(u_1) + C \cdot D_n) - \phi(u_1)du_1 = C \cdot D_n = O\left(\frac{1}{\sqrt{n}}\right).$$

This implies

$$\begin{aligned}
&\sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x\in[0,1]^2}} \left|\lambda^{(j)}_{n,\theta}(x) - \int\limits_{A^C} K^{(j)}_n(\theta, u-x)du\right| \\
= \;& \sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x\in[0,1]^2}} \left|\frac{1}{h_{1n}h_{2n}} \int\limits_{A_n} K_j(H_n^{-1}A_\theta(u-x))du - \frac{1}{h_{1n}h_{2n}} \int\limits_{A^C} K_j(H_n^{-1}A_\theta(u-x))du\right| \\
\leq \;& \frac{1}{h_{1n}h_{2n}} \cdot S\left(A_n \triangle A^C\right) \cdot \max_{x\in[0,1]^2} |K_j(x)| = O\left(\frac{1}{\sqrt{n}h_{1n}h_{2n}}\right).
\end{aligned}$$

Proof of Lemma 4 First, we show the convergence of $\lambda^{(j)}_{n,\theta}(\widetilde{x})$ to $\lambda^{(j)}_{\theta}(\widetilde{x})$. With Lemma 6 we have

$$\begin{aligned}
&\sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x_1\in[0,1]}} \left|\lambda^{(j)}_{n,\theta}(\widetilde{x}) - \lambda^{(j)}_{\theta}(\widetilde{x})\right| \\
= \;& \sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x_1\in[0,1]}} \left|\frac{1}{h_{1n}h_{2n}} \int\limits_{A^C} K_j(H_n^{-1}A_\theta(u-\widetilde{x}))du - \int\limits_{A_{x_1}} K_j(Q^{-1}A_\theta u)du\right| \\
& + O\left(\frac{1}{\sqrt{n}h_{1n}h_{2n}}\right)
\end{aligned}$$

$$= \sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x_1\in[0,1]}} \left| \int\limits_{H_n^{-1}A_\theta(A^C-\widetilde{x})\cap T} K_j(\widetilde{u})d\widetilde{u} - \int\limits_{Q^{-1}A_\theta(A_{x_1})\cap T} K_j(\widetilde{u})d\widetilde{u} \right| + O\left(\tfrac{1}{\sqrt{n}h_{1n}h_{2n}}\right)$$

$$\rightarrow 0,$$

since $\sup_{\theta,x_1} S\Big(\big(H_n^{-1}A_\theta\left(A^C-\widetilde{x}\right)\cap T\big)\triangle\big(Q^{-1}A_\theta(A_{x_1})\cap T\big)\Big)\rightarrow 0$, what can be shown as follows:

Let $T_1(\theta,x_1) := \phi'(x_1)\cos(-\theta)-\sin(-\theta)$ and $T_2(\theta,x_1) := \phi'(x_1)\sin(-\theta)+\cos(-\theta)$. Since $T_1(\theta,x_1)^2+T_2(\theta,x_1)^2=\phi'(x_1)^2+1\geq 1$, we have for all θ, x_1 that $|T_1(\theta,x_1)|\geq t_0$ or $|T_2(\theta,x_1)|\geq t_0$ where $t_0=\sqrt{\frac{1}{2}}$. Further, there exists C_T independent of θ and x_1 with $|T_1(\theta,x_1)|\leq C_T$ and $|T_2(\theta,x_1)|\leq C_T$ for all θ and x_1.

With $A_\theta^{-1}=A_{-\theta}$ and with the Taylor-expansion of ϕ in x_1 we have on the one hand $u\in H_n^{-1}A_\theta\left(A^C-\widetilde{x}\right)\Leftrightarrow A_\theta^{-1}H_nu+\widetilde{x}\in A^C\Leftrightarrow a_{2-\theta}^\top H_nu+\phi(x_1)\leq\phi(a_{1-\theta}^\top H_nu+x_1)$ $\Leftrightarrow T_2(\theta,x_1)u_2\leq\frac{h_{1n}}{h_{2n}}T_1(\theta,x_1)u_1+O\left(\frac{h_{1n}^2}{h_{2n}}+h_{1n}+h_{2n}\right)$ and on the other hand

$u\in Q^{-1}A_\theta(A_{x_1})\Leftrightarrow A_\theta^{-1}Qu\in A_{x_1}\Leftrightarrow a_{2-\theta}^\top Qu\leq\phi'(x_1)a_{1-\theta}^\top Qu$
$\Leftrightarrow T_2(\theta,x_1)u_2\leq q\,T_1(\theta,x_1)u_1$.

For $u\in\big(H_n^{-1}A_\theta\left(A^C-\widetilde{x}\right)\setminus Q^{-1}A_\theta(A_{x_1})\big)\cap T$ we have consequently

$$q\,T_1(\theta,x_1)u_1 < T_2(\theta,x_1)u_2\leq\frac{h_{1n}}{h_{2n}}T_1(\theta,x_1)u_1+O(h_{1n}+h_{2n}). \tag{4}$$

If now $|T_2(\theta,x_1)|\geq t_0$ holds, this implies

$$q\frac{T_1(\theta,x_1)}{T_2(\theta,x_1)}u_1 \;<\; u_2 \;\leq\; \frac{h_{1n}}{h_{2n}}\frac{T_1(\theta,x_1)}{T_2(\theta,x_1)}u_1+\frac{O(h_{1n}+h_{2n})}{T_2(\theta,x_1)} \quad \text{for } T_2(\theta,x_1)>0$$

$$q\frac{T_1(\theta,x_1)}{T_2(\theta,x_1)}u_1 \;>\; u_2 \;\geq\; \frac{h_{1n}}{h_{2n}}\frac{T_1(\theta,x_1)}{T_2(\theta,x_1)}u_1+\frac{O(h_{1n}+h_{2n})}{T_2(\theta,x_1)} \quad \text{for } T_2(\theta,x_1)<0$$

and with Lemma 5.iv we get

$$\sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x_1\in[0,1]\\ |T_2(\theta,x_1)|\geq t_0}} S\Big(\big(H_n^{-1}A_\theta\left(A^C-\widetilde{x}\right)\setminus Q^{-1}A_\theta(A_{x_1})\big)\cap T\Big)$$

$$\leq \sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x_1\in[0,1]\\ |T_2(\theta,x_1)|\geq t_0}} \int_{-2}^{2}\left|\tfrac{h_{1n}}{h_{2n}}-q\right|\tfrac{|T_1(\theta,x_1)|}{|T_2(\theta,x_1)|}u_1+\tfrac{O(h_{1n}+h_{2n})}{|T_2(\theta,x_1)|}du_1$$

$$\leq 4\left(\left|\tfrac{h_{1n}}{h_{2n}}-q\right|\tfrac{C_T}{t_0}+\tfrac{O(h_{1n}+h_{2n})}{t_0}\right)\rightarrow 0.$$

But if $|T_1(\theta, x_1)| \geq t_0$ holds, we similarly get the same result. Moreover, (4) holds for $u \in \left(Q^{-1}A_\theta(A_{x_1}) \setminus H_n^{-1}A_\theta\left(A^C - \widetilde{x}\right)\right) \cap T$ with changed signs, so that we also have $\sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x_1\in[0,1]}} S\Big(\left(Q^{-1}A_\theta(A_{x_1}) \setminus H_n^{-1}A_\theta\left(A^C - \widetilde{x}\right)\right) \cap T\Big) \to 0$ what means, that also

$$\sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x_1\in[0,1]}} S\left((H_n^{-1}A_\theta\left(A^C - \widetilde{x}\right) \cap T) \triangle (Q^{-1}A_\theta(A_{x_1} \cap T))\right) \to 0.$$

Now, we get

$$\begin{aligned}
&\sup_{\substack{z\in\mathbb{R},\,\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x_1\in[0,1]}} \left|h^{(j)}_{n,\theta,\widetilde{x}}(z) - h^{(j)}_{\theta,\widetilde{x}}(z)\right| \\
\leq\ &\sup_{\substack{z\in\mathbb{R},\,\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x_1\in[0,1]}} \left|\lambda^{(j)}_{n,\theta}(\widetilde{x}) - \lambda^{(j)}_{\theta}(\widetilde{x})\right| \\
&\qquad \cdot \left|\int \psi\Big(\mu - z + \widetilde{m}(\widetilde{x})\Big) f(\mu) d\mu - \int \psi\Big(\mu - z + \widetilde{m}(\widetilde{x}) + C(x_1)\Big) f(\mu) d\mu\right| \\
\leq\ &\sup_{\substack{\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]\\ x_1\in[0,1]}} \left|\lambda^{(j)}_{n,\theta}(\widetilde{x}) - \lambda^{(j)}_{\theta}(\widetilde{x})\right| \cdot C_\psi \cdot \max|C(x_1)| \cdot \left|\int f(\mu) d\mu\right| \to 0.
\end{aligned}$$

Proof of Theorem 1 For $x_1 \in \left[\,\overline{h_n}\,, 1 - \overline{h_n}\,\right]$ let first $\widetilde{\theta}(x_1) := -\arctan(\phi'(x_1))$.

Since $\arctan(\phi'(x_1)) \in \left(\frac{-\pi}{2}, \frac{\pi}{2}\right)$ and therefore $\cos\Big(\widetilde{\theta}(x_1)\Big) > 0$ we have for all $u \in T \cap (A_{x_1})^C$ (e.g. $u_2 > \phi'(x_1)u_1$)

$$\begin{aligned}
a^\top_{2\,\widetilde{\theta}(x_1)} u &> \left[\sin\left(\widetilde{\theta}(x_1)\right) + \phi'(x_1)\cos\left(\widetilde{\theta}(x_1)\right)\right] u_1 \\
&= \Big[\underbrace{\tan\left(\widetilde{\theta}(x_1)\right)}_{=-\phi'(x_1)} + \phi'(x_1)\Big] \cos\left(\widetilde{\theta}(x_1)\right) u_1 = 0.
\end{aligned}$$

With Condition (A1) that implies $K_1\left(Q^{-1}A_{\widetilde{\theta}(x_1)}u\right) = 0$ for all $u \in T \cap (A_{x_1})^C$. But for $u \notin T$ we also have $K_1\left(Q^{-1}A_{\widetilde{\theta}(x_1)}u\right) = 0$ and therefore, with $\widetilde{x} := (x_1, \phi(x_1))^\top$, we have

$$\lambda^{(1)}_{\widetilde{\theta}(x_1)}(\widetilde{x}) = \int_{A_{x_1}} K_1\left(Q^{-1}A_{\widetilde{\theta}(x_1)}u\right) du = \int_{\mathbb{R}^2} K_1\left(Q^{-1}A_{\widetilde{\theta}(x_1)}u\right) du = 1$$

what implies

$$h^{(1)}_{\widetilde{\theta}(x_1),\widetilde{x}}(z) = \int \psi(\mu - z) f(\mu - \widetilde{m}(\widetilde{x})) d\mu$$

for all $x_1 \in \left[\,\overline{h_n}\,, 1 - \overline{h_n}\,\right]$.

With the same arguments as in the proof of Lemma 3 we get

$$P\left(\exists\, x_1 \in \left[\,\overline{h_n}\,, 1-\,\overline{h_n}\,\right] \,:\, m_n^{(1)}\left(\widetilde{\theta}(x_1), \widetilde{x}\right) \notin \left(\widetilde{m}(\widetilde{x})-\delta, \widetilde{m}(\widetilde{x})+\delta\right)\right) < \epsilon$$

for all $n > N_1$ what implies

$$P\left(m_n^{(1)}\left(\widetilde{\theta}(x_1), \widetilde{x}\right) \in \left(\widetilde{m}(\widetilde{x})-\delta, \widetilde{m}(\widetilde{x})+\delta\right) \forall\, x_1 \in \left[\,\overline{h_n}\,, 1-\,\overline{h_n}\,\right]\right) > 1-\epsilon.$$

Analogously we get $\lambda^{(2)}_{\widetilde{\theta}(x_1)}(\widetilde{x}) = 0$ and

$$P\Big(m_n^{(2)}\left(\widetilde{\theta}(x_1), \widetilde{x}\right) \in \left(\widetilde{m}(\widetilde{x})+C(x_1)-\delta, \widetilde{m}(\widetilde{x})+C(x_1)+\delta\right) \qquad \forall\, x_1 \in \left[\,\overline{h_n}\,, 1-\,\overline{h_n}\,\right]\Big) \;>\; 1-\epsilon$$

for all $n > N_1$, what implies

$$P\left(\left|M_n\left(\widetilde{\theta}(x_1), \widetilde{x}\right)\right| \in \left(C(x_1)-2\delta, C(x_1)+2\delta\right) \forall\, x_1 \in \left[\,\overline{h_n}\,, 1-\,\overline{h_n}\,\right]\right) > 1-\epsilon.$$

Finally, we have

$$C_n(x_1) = \max_{x_2 \in \left[\,\overline{h_n}\,, 1-\overline{h_n}\,\right]} \; \max_{\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]} \left|M_n(\theta, (x_1, x_2)^\top)\right| \geq \left|M_n\left(\widetilde{\theta}(x_1), \widetilde{x}\right)\right|,$$

for all $n > N_1$ and $x_1 \in \left[\,\overline{h_n}\,, 1-\,\overline{h_n}\,\right]$ what implies

$$P\Big(C_n(x_1) > C(x_1) - 2\delta \;\forall\, x_1 \in \left[\,\overline{h_n}\,, 1-\,\overline{h_n}\,\right]\Big) > 1-\epsilon$$

for all $n > N_1$. But since also $\widetilde{x}_n := (x_1, \phi_n(x_1))^\top \in \left[\,\overline{h_n}\,, 1-\,\overline{h_n}\,\right]^2$, with Lemma 3.i there exists as well $N_2 \in \mathbb{N}$ with

$$\begin{aligned}
& P\Big(C_n(x_1) < C(x_1) + 2\delta \;\forall\, x_1 \in \left[\,\overline{h_n}\,, 1-\,\overline{h_n}\,\right]\Big) \\
\geq\;\; & P\left(m_n^{(1)}(\theta_n(\widetilde{x}_n), \widetilde{x}_n) \in \left(\widetilde{m}(\widetilde{x}_n)-\delta, \widetilde{m}(\widetilde{x}_n)+C(x_1)+\delta\right) \wedge \right. \\
& \left. m_n^{(2)}(\theta_n(\widetilde{x}_n), \widetilde{x}_n) \in \left(\widetilde{m}(\widetilde{x}_n)-\delta, \widetilde{m}(\widetilde{x}_n)+C(x_1)+\delta\right) \forall\, x_1 \in \left[\,\overline{h_n}\,, 1-\,\overline{h_n}\,\right]\right) \\
>\;\; & 1-2\epsilon
\end{aligned}$$

for all $n > N_2$. With $N := \max\{N_1, N_2\}$ the claim is proven.

Proof of Theorem 2 With $M_{\phi'} := \max_{x_1 \in [0,1]} |\phi'(x_1)|$, $q_n := \max\{h_{1n}, h_{2n}\}$, and $p_n := 2M_{\phi'}q_n + 2q_n + M_{\phi'}D_n + D_n$, let

$$\begin{aligned}
A_n^+ &:= \left\{(x_1, x_2) \in \left[\,\overline{h_n}\,, 1-\,\overline{h_n}\,\right]^2 : x_2 \geq \phi(x_1) + p_n\right\} \\
A_n^- &:= \left\{(x_1, x_2) \in \left[\,\overline{h_n}\,, 1-\,\overline{h_n}\,\right]^2 : x_2 \leq \phi(x_1) - p_n\right\}.
\end{aligned}$$

We show, that for all $\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, $n \in \mathbb{N}$, and $j \in \{1,2\}$ we have $\lambda^{(j)}_{n,\theta}(x) = 1$ for all $x \in A^-_n$ and $\lambda^{(j)}_{n,\theta}(x) = 0$ for all $x \in A^+_n$.

First, let $x \in A^-_n$ and $i \notin J_n$, what means $x_{2i} > \phi(x_{1i})$. Then, we have for all $u \in \Delta_i$ $u_2 > \phi(x_{1i}) - D_n \geq \phi(u_1) - M_{\phi'} D_n - D_n$. If $|u_1 - x_1| \leq 2q_n$, we therefore have $u_2 - x_2 \geq \phi(u_1) - M_{\phi'} D_n - D_n - (\phi(x_1) - p_n) = \phi'(\xi_u)(u_1 - x_1) + 2M_{\phi'} q_n + 2q_n \geq 2q_n$. That means we have $||u - x|| \geq 2q_n$ for all $u \in \Delta_i$ what implies $i \notin J^{(j)}_n(\theta, x)$. Consequently we have $J^{(j)}_n(\theta, x) \subset J_n$ for all $x \in A^-_n$ and $\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, what, with Lemma 5.v, immediately implies

$$\lambda^{(j)}_{n,\theta}(x) = \sum\nolimits_{i \in J_n} \alpha^{(j)}_{i,n}(\theta, x) = \sum\nolimits_{i \in J^{(j)}_n(\theta,x)} \alpha^{(j)}_{i,n}(\theta, x) = 1$$

for all $\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, $x \in A^-_n$ and $n \in \mathbb{N}$. $\lambda^{(j)}_{n,\theta}(x) = 0$ for $x \in A^+_n$ follows analogously.

With $L^0_{j,n}$ and $L^1_{j,n}$ from Lemma 3.ii we now have $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \times A^-_n \subset L^1_{j,n}$ and $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \times A^+_n \subset L^0_{j,n}$ for all $n \in \mathbb{N}$. Therefore, we get that for all $\epsilon > 0$, $\delta > 0$ there exists a $N_1 \in \mathbb{N}$ so that for all $n > N_1$

$$\begin{aligned} & P\left(\sup\nolimits_{x \in A^-_n} \widetilde{C}_n(x) > \delta\right) = P\left(\sup_{\substack{x \in A^-_n \\ \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]}} \left|m^{(2)}_n(\theta, x) - m^{(1)}_n(\theta, x)\right| > \delta\right) \\ \leq\ & P\left(\exists (\theta, x) \in L^1_{j,n} : m^{(1)}_n \notin \left(\widetilde{m}(x) - \delta, \widetilde{m}(x) + \delta\right) \vee m^{(2)}_n \notin \left(\widetilde{m}(x) - \delta, \widetilde{m}(x) + \delta\right)\right) \\ <\ & \epsilon/2 \end{aligned}$$

and in the same way $P\left(\sup_{x \in A^+_n} \widetilde{C}_n(x) > \delta\right) < \epsilon/2$.

Since $C_n(x_1) = \widetilde{C}_n\left((x_1, \phi_n(x_1))^\top\right)$ that implies

$$\begin{aligned} & P\left(\exists\, x_1 \in \left[\,\overline{h_n}\,, 1 - \overline{h_n}\,\right] \,:\, C_n(x_1) > \delta \,\wedge\, (x_1, \phi_n(x_1))^\top \in A^-_n \cup A^+_n\right) \qquad (5) \\ \leq\ & P\left(\sup\nolimits_{x \in A^-_n \cup A^+_n} \widetilde{C}_n(x) > \delta\right) < \epsilon \end{aligned}$$

for all $n > N_1$. Further, we have

$$\begin{aligned} & P\left(\exists\, x_1 \in \left[\,\overline{h_n}\,, 1 - \overline{h_n}\,\right] \,:\, C_n(x_1) > \delta \,\wedge\, (x_1, \phi_n(x_1))^\top \in A^-_n \cup A^+_n\right) \\ \geq\ & P\left(C_n(x_1) > \delta\, \forall x_1 \in \left[\,\overline{h_n}\,, 1 - \overline{h_n}\,\right]\right) + \\ & P\left(\exists\, x_1 \in \left[\,\overline{h_n}\,, 1 - \overline{h_n}\,\right] \,:\, (x_1, \phi_n(x_1))^\top \in A^-_n \cup A^+_n\right) - 1. \qquad (6) \end{aligned}$$

Now, w.l.o.g. let $\epsilon < \frac{1}{2}$ and $\delta < \min_{x_1 \in [0,1]} C(x_1)/2$. Then, with Theorem 1 there exists $N_2 \in \mathbb{N}$ with

$$\begin{aligned} & P\left(C_n(x_1) > \delta \,\forall\, x_1 \in \left[\,\overline{h_n}\,, 1 - \overline{h_n}\,\right]\right) \qquad (7) \\ \geq \;& 1 - P\left(\exists\, x_1 \in \left[\,\overline{h_n}\,, 1 - \overline{h_n}\,\right] \,:\, C_n(x_1) < C(x_1) - \delta\right) \\ \geq \;& 1 - P\left(\sup_{x_1 \in \left[\,\overline{h_n}\,, 1 - \overline{h_n}\,\right]} \left| C_n(x_1) - C(x_1) \right| > \delta\right) > 1 - \epsilon \end{aligned}$$

for all $n > N_2$. (5),(6), and (7) provide

$$\begin{aligned} & P\left(\exists\, x_1 \in \left[\,\overline{h_n}\,, 1 - \overline{h_n}\,\right] \,:\, (x_1, \phi_n(x_1))^\top \in A_n^- \cup A_n^+\right) \\ \leq \;& P\left(\exists\, x_1 \in \left[\,\overline{h_n}\,, 1 - \overline{h_n}\,\right] \,:\, C_n(x_1) > \delta \,\wedge\, (x_1, \phi_n(x_1))^\top \in A_n^- \cup A_n^+\right) \\ & - P\left(C_n(x_1) > \delta \,\forall\, x_1 \in \left[\,\overline{h_n}\,, 1 - \overline{h_n}\,\right]\right) + 1 \\ < \;& 2\epsilon \end{aligned}$$

for all $n > \max\{N_1, N_2\}$. Since $p_n \to 0$ there exists $N_3 \in N$ with $p_n < \delta$ for all $n > N_3$ and therefore we finally get

$$\begin{aligned} & P\left(\sup_{x_1 \in \left[\,\overline{h_n}\,, 1 - \overline{h_n}\,\right]} \left| \phi_n(x_1) - \phi(x_1) \right| > \delta\right) \\ < \;& P\left(\exists\, x_1 \in \left[\,\overline{h_n}\,, 1 - \overline{h_n}\,\right] \,:\, (x_1, \phi_n(x_1))^\top \in A_n^- \cup A_n^+\right) \\ < \;& 2\epsilon \end{aligned}$$

for all $n > N := \max\{N_1, N_2, N_3\}$.

References

[1] Carlstein, E., Krishnamoorthy, C. (1992), Boundary Estimation, *Journal of the American Statistical Association* **87**, 430–438.

[2] Davis, L. (1975), A Survey of Edge Detection Techniques, *Computer Graphics and Image Processing* **4**, 248–270.

[3] Eubank, R. L. (1988), *Spline Smoothing and Nonparametric Regression*, Dekker, New York.

[4] Ferger, D. (2004), Boundary estimation based on set-indexed empirical processes, *Nonparametric Statistics* **16**, 245–260.

[5] Härdle, W., Gasser, T. (1984), Robust Nonparametric Function Fitting, *Journal of the Royal Statistical Society, Series B* **46**, 42–51.

[6] Hou, Z., Koh, T.S. (2003), Robust edge detection, *Pattern Recognition* **36**, 2083–2091.

[7] Korostelev, A.P., Tsybakov, A.B. (1993), *Minimax theory of image reconstruction*, Lecture Notes in Statistics **82**, Springer-Verlag, New York.

[8] Müller, H.-G. (1992), Change-Points in Nonparametric Regression Analysis, *Annals of Statistics* **20**, 737–761.

[9] Müller, H.-G., Song, K.-S. (1994), Maximin estimation of multidimensional boundaries, *Journal of Multivariate Analysis* **50**, 265–281.

[10] Peli, T., Malah, D. (1982), A Study of Edge Detection Algorithms, *Computer Graphics and Image Processing* **20**, 1–21.

[11] Qiu, P. (1997), Nonparametric Estimation of Jump Surface, *Sankhya: The Indian Journal of Statistics, Series A* **59**, 268–294.

[12] Qiu, P. (2002), A nonparametric procedure to detect jumps in regression surfaces, *Journal of Computational and Graphical Statistics* **11**, 799–822.

[13] Qiu, P., Asano, C., Li, X. (1991), Estimation of Jump Regression Function, *Bulletin of Informatics and Cybernetics* **24**, 197–212.

[14] Qiu, P., Yandell, B. (1997), Jump Detection in Regression Surfaces, *Journal of Computational and Graphical Statistics* **6**, 332–354.

[15] Wu, J. S. Wu, Chu, C. K. (1993), Kernel Type Estimators of Jump Points and Values of a Regression Function, *Annals of Statistics* **21**, 1545–1566.

Tim Garlipp
garlipp@math.uni-oldenburg.de

Prof. Dr. Christine H. Müller
mueller@math.uni-oldenburg.de

Institut für Mathematik
Carl von Ossietzky Universität Oldenburg
Postfach 2503
D - 26111 Oldenburg
Germany