JOURNAL OF MULTIVARIATE ANALYSIS 50, 265–281 (1994)

Maximin Estimation of Multidimensional Boundaries

HANS-GEORG MÜLLER

University of California

AND

KAI-SHENG SONG

Purdue University

We consider the problem of estimating the location and size of a discontinuity in an otherwise smooth multidimensional regression function. The boundary or location of the discontinuity is assumed to be a closed curve respective surface, and we aim to estimate this closed set. Our approach utilizes the uniform convergence of multivariate kernel estimators for directional limits. Differences of such limits converge to zero under smoothness assumptions, and to the jump size along the discontinuity. This leads to the proposal of a maximin estimator, which selects the boundary for which the minimal estimated directional difference among all points belonging to this boundary is maximized. It is shown that this estimated boundary is almost surely enclosed in a sequence of shrinking neighborhoods around the true boundary, and corresponding rates of convergence are obtained. © 1994 Academic Press, Inc.

1. INTRODUCTION

The problem of estimating a multidimensional boundary has recently found increasing interest, as it is closely related to the problem of detecting edges in image analysis and has also applications in ecology and general statistical modelling. The connection between edge detection and the estimation of a multivariate boundary or change-point was recently explored by Carlstein and Krishnamoorthy (1992), in a recent monograph by Korostelev and Tsybakov (1993), and by Rudemo and Stryhn (1993a, 1993b). An earlier approach for boundary detection was described in

Received March 23, 1992; revised November 10, 1993.

AMS 1991 subject classifications: 62G07, 62H11.

Key words and phrases: change-point, change curve, directional limits, discontinuity, edge, image analysis, jump size, kernel estimator, nonparametric regression, rate of convergence, strong consistency.

0047-259X/94 $6.00

Brodskii and Darkhovskii (1986). Optimal rates in a minimax sense were developed in Tsybakov (1989), and a good review of the current knowledge regarding the boundary estimation problem is given in Tsybakov (1993). This area is currently under rapid development, and these references are only a selection of significant work. For further relevant references consult Korstelev and Tsybakov (1993), Rudemo and Stryhn (1993a), and Tsybakov (1993).

We consider the multidimensional boundary estimation problem here in the context of a multivariate fixed design regression setting and propose a new class of estimators for the boundary, which alternatively could be described as a multidimensional change-point, edge, or change curve. Strong consistency with rates of convergence for these estimators will be established in a suitable metric. Since we aim to estimate a set in d-dimensional Euclidean space, choice of an appropriate metric on sets in $\mathbb{R}^d$ is of relevance in order to obtain a reasonable measure for the performance of the proposed estimators.

We assume that the boundary Γ_0 induces a partition of the compact subspace $[0, 1]^d$ of $\mathbb{R}^d$, within which measurements are available, into two sets C_1, C_2, such that $\Gamma_0 = \partial C_1 \backslash \partial([0, 1]^d) = \partial C_2 \backslash \partial([0, 1]^d)$, i.e., Γ_0 is a (possibly closed) curve in the two-dimensional case and a (possibley closed) surface in the case that $d > 2$. The boundary property of Γ_0 is a consequence of the following assumptions on the regression function g: $[0, 1]^d \to \mathbb{R}$: This function is assumed to be "smooth" everywhere, except at the boundary Γ_0, where it has a discontinuity. For each point $y \in \Gamma_0$, the maximum of the differences of directional limits, $|\lim_{x \to y, x \in C_1} g(x) - \lim_{x \to y, x \in C_2} g(x)|$, is bounded away from zero.

We obtain this by assuming

$$g(x) = h(x) + \Delta_n f(x) 1_{C_2}(x), \tag{1.1}$$

where h, f are smooth on $[0, 1]^d$, and $\inf_{y \in \Gamma_0} |f(y)| = 1$. In this model, Δ_n is seen to be a lower bound for the jump size across the discontinuity Γ.

The available data are assumed to be noisy measurements of g, taken on a regular grid in $[0, 1]^d$. This corresponds to the situation in image analysis for $d = 2$, where the pixels are located on a regular grid and the grey level of each pixel would correspond to a continuous, noisy measurement of g at the respective point. Since often a boundary respective edge or change curve is of primary interest when analyzing such data, the corresponding statistical problem of estimating a boundary has many important applications. Model (1.1) will be appropriate whenever discontinuous and continuous changes occur simultaneously in different regions. If several boundaries exist simultaneously, defining a partition of $[0, 1]^d$ into more than two subsets or "cells" $C_1, \ldots, C_m$, the proposed method could be modified to estimate the boundaries of several "cells" as well.

Besides estimating the boundary Γ_0, it is often also of interest to obtain good global estimates for the regression function g itself, the "surface" or "image." Ignoring the boundary and applying simple smoothing procedures will result in a "blurred" image where the originally sharp boundary is "smoothed out," producing inconsistent estimates in a neighborhood of this boundary. Alternatively, a two-step procedure can be devised, in analogy to a proposal in Müller (1992) for the one-dimensional case, where in a first step the boundary is estimated and in a second step a nonparametric regression estimate adapted to this boundary is obtained. Such estimators can preserve the discontinuity and thus lead to globally consistent estimators of the regression function. Related global discontinuity preserving curve estimators in the one-dimensional case were considered by McDonald and Owen (1986) and Hall and Titterington (1992).

Sometimes it might also be of interest to divide the total space $[0, 1]^d$ into several "cells" as mentioned above, to estimate the dividing boundaries, and then to fit a multivariate parametric regression model within each cell. If the number of cells is assumed to grow with the number of observations, this leads ultimately to a nonparametric multivariate regression procedure which could be described as a "honeycomb" method.

The paper is organized as follows: The model, the notation, assumptions, and precise formulation of the problem are introduced in Section 2. The proposed maximin estimators which are based on differences of kernel estimators for directional limits are discussed in Section 3. The main result on strong consistency of these boundary estimators as well as the rate of convergence in the Hausdorff metric is presented in Theorem 4.1 and Corollary 4.1 in Section 4. Section 5 contains a result on the uniform a.s. convergence of kernel estimators for directional limits (Theorem 5.1) and auxiliary results.

2. Multidimensional Boundaries

We assume that noisy measurements $y_{i,n}$ are obtained of a regression surface g: $\mathbb{R}^d \to \mathbb{R}$, $d \geq 1$, which has a discontinuity, the location of which is the boundary we intend to estimate. These measurements are made at non-random locations $x_{i,n}$ which lie on a regular grid in $[0, 1]^d$. More precisely, let $n = \prod_{l=1}^{d} n_l$, $i = (i_1, ..., i_d)$, $1 \leq i_l \leq n_l$, $1 \leq l \leq d$, be a multi-index and assume

(A1) $y_{i,n} = g(x_{i,n}) + \varepsilon_{i,n}$, where the $\varepsilon_{i,n}$ are all i.i.d. random variables with $E\varepsilon_{i,n} = 0$, and

(A2) $x_{i,n} = ((i_1 - 1/2)/n_1, (i_2 - 1/2)/n_2, ..., (i_d - 1/2)/n_d) \in [0, 1]^d$, $\min_{1 \leq l \leq d} n_l \geq cn^{1/d}$ for a constant $c > 0$.

In the following denote the index set of all i such that $x_{i,n}$ is a grid point by I; I has n elements.

The function g is assumed to be smooth everywhere except at a boundary $\Gamma_0 \subset [0, 1]^d$. We assume that Γ_0 has the following properties:

(A3) Γ_0 is closed and there exist two arcwise connected sets C_1, C_2 such that $C_1 \cup C_2 \cup \Gamma_0 = [0, 1]^d$, $C_i \cap \Gamma_0 = \varnothing$, $i = 1, 2$, $C_1 \cap C_2 = \varnothing$.

(A4) There exists a δ with $0 < \delta < 1$ such that for any given sequence $\alpha_n \to 0$, $\alpha_n > 0$, for sufficiently large n the following holds. For each $y \in \Gamma_0$, there exists an orthonormal mapping $T_y(\cdot): \mathbb{R}^d \to \mathbb{R}^d$ with the properties:

(i) $T_y(z) = R(y + z)$, $z \in \mathbb{R}^d$, where R is a rotation in $\mathbb{R}^d$.

(ii) Considering the cubes

$$D_{1n} = [\delta\alpha_n, 2\alpha_n + \delta\alpha_n] \times [-\alpha_n, \alpha_n]^{d-1},$$

$$D_{2n} = [-\delta\alpha_n - 2\alpha_n, -\delta\alpha_n] \times [-\alpha_n, \alpha_n]^{d-1},$$

it holds that $T_y(D_{1n}) \subset C_1$, $T_y(D_{2n}) \subset C_2$.

A consequence is that for each $y \in \Gamma_0$, the sequences $y_{1n} = T_y\{(\delta\alpha_n, 0, ..., 0)^T\} \in C_1$, $y_{2n} = T_y\{(-\delta\alpha_n, 0, ..., 0)^T\} \in C_2$, satisfy

$$y_{1n} \to y \quad \text{and} \quad y_{2n} \to y \quad \text{as} \quad n \to \infty, \tag{2.1}$$

so that Γ_0 is in the closure of both C_1 and C_2.

Note that condition (A4) is always satisfied by twice differentiable boundaries with uniformly bounded curvature. This can be seen geometrically, since then for each $y \in \Gamma_0$ there exist two cones with their tips meeting at y, centered around the same axis through y, such that except for y one of the two cones belongs to C_1 and the other one to C_2. The angle of aperture for these cones can be chosen to be the same for all $y \in \Gamma_0$ due to the uniform curvature of Γ_0. We can then inscribe cubes with properties as required in (ii) into the cones, where the uniform angle of aperture of the cones determines the corresponding δ. This shows the close relation of (A4) with the "double cone condition" described in Korostelev and Tsybakov (1993).

Note that assumption (A4) also covers non-differentiable boundaries like rectangles or other polygons in $\mathbb{R}^d$. Condition (A4) is a "separability" requirement which ensures that we can estimate the jump across the boundary Γ_0 by employing smoothers with supports $T_y(D_{1n}) \subset C_1$ and $T_y(D_{2n}) \subset C_2$ contained in the regions where the regression function is smooth in order to obtain estimators for the limits $\lim_{x \to y, x \in C_1} g(x)$, $\lim_{x \to y, x \in C_2} g(x)$. Furthermore, also owing to the continuity of g in subregions C_1, C_2 as required in (A5) below, these limits do not depend on the particular sequence x chosen in the respective subregion.

The regression function g is supposed to be smooth within subregions C_1, C_2 but to have a discontinuity across Γ_0; the jumpsize Δ_n may depend on n. These assumptions on g are reflected in the following model:

(A5) $g(x) = h(x) + \Delta_n f(x)\, 1_{C_2}(x),\ x \in [0, 1]^d.$

where $h,\ f \in \mathscr{C}^k([0, 1]^d)$, $\inf_{y \in \Gamma_0} |f(y)| = 1$, Δ_n is either constant or monotone decreasing and $k \geq 1$ is a given integer; it is required (see (K7) below) that $n^{1/d}\Delta_n \to \infty$ as $n \to \infty$.

The jumpsize when crossing the boundary at any given point $y \in \Gamma_0$ is then $|\Delta_n f(y)| \geq \Delta_n$; it is not necessarily constant on Γ_0 and may decrease as $n \to \infty$.

Furthermore, for the strong consistency result to be derived, the following requirement on the moments of the errors is needed:

(A6) The error variables $\varepsilon_{i,n}$ in model (A1) form a triangular scheme of i.i.d. random variables such that $E\varepsilon_{i,n} = 0$, $E\varepsilon_{i,n}^2 = \sigma^2$, and $E|\varepsilon_{i,n}|^s < \infty$ for some $s > 4$.

The estimators for the multidimensional boundary or change curve Γ_0 to be introduced in the next section will be shown to be strongly consistent under model assumptions (A1)–(A6) in Theorem 4.1. Note that the model assumptions (A1)–(A6) apply to a large range of multidimensional boundary problems which are relevant in applications. For instance, Γ_0 could be a simple Jordan curve in $\mathbb{R}^2$ and h, f could be chosen as constant functions with constants $c_1, c_2, c_1 \neq c_2$, compare Rudemo and Stryhn (1993b) for a discussion of this case. In image analysis, this case corresponds to the situation of two different constant grey levels separated by the boundary Γ_0.

3. The Maximin Method

Let $\varphi = (\varphi_1, ..., \varphi_{d-1}) \in [0, 2\pi]^{d-1}$. For any $x \in [0, 1]^d$ and $\omega \in \mathbb{R}$, $\omega > 0$, define

$$z(x, \omega, \varphi) = (x_1 + \omega \cos \varphi_1, ..., x_{d-1} + \omega \cos \varphi_{d-1}, x_d + \omega \sin \varphi_{d-1}) \in \mathbb{R}^d.$$

For any point $x \in [0, 1]^d$, define the directional limit

$$g(x, \varphi) = \lim_{\omega \downarrow 0} g(z(x, \omega, \varphi)), \tag{3.1}$$

and the absolute difference of directional limits taken in opposite directions,

$$\Delta(x, \varphi) = |g(x, \varphi) - g(x, \bar{\varphi})|,$$

where

$$\bar{\varphi}_j = \pi + \varphi_j, \qquad 1 \leqslant j \leqslant d-1.$$

According to (A3)–(A5), these quantities are well-defined for all $x \in [0, 1]^d$ and $\varphi \in [0, 2\pi]^{d-1}$. The maximum difference of directional limits associated with a point $x \in [0, 1]^d$ is then

$$\Delta(x) = \sup_{\varphi} \Delta(x, \varphi). \tag{3.2}$$

Note that the mapping $T_y(\cdot)$ which is defined in (A4) corresponds to a rotation in $\mathbb{R}^d$, coupled with a translation. This implies that for each $y \in \Gamma_0$, there exists $\varphi = \varphi(y) \in [0, 2\pi]^{d-1}$ such that

$$z(y, \omega, \varphi(y)) \to y \qquad z(y, \omega, \bar{\varphi}(y)) \to y, \qquad \text{as} \quad \omega \to 0, \tag{3.3}$$

where $\bar{\varphi}_j(y) = \pi + \varphi_j(y)$, $1 \leqslant j \leqslant d-1$, and, without loss of generality,

$$z(y, \omega, \varphi(y)) \in C_1, \qquad z(y, \omega, \bar{\varphi}(y)) \in C_2.$$

It follows immediately from (3.3) and (A5) that

$$\inf_{y \in \Gamma_0} \Delta(y) \geqslant \inf_{y \in \Gamma_0} \Delta(y, \varphi(y)) = \Delta_n > 0, \tag{3.4}$$

whereas the continuity of g implies

$$\Delta(x) = 0 \qquad \text{for all} \quad x \in [0, 1]^d \setminus \Gamma_0. \tag{3.5}$$

Assume that $\hat{\Delta}(x)$ are estimates of $\Delta(x)$. Then we define the maximin estimator

$$\hat{\Gamma} = \arg\sup_{\Gamma \in \Xi} \left\{ \inf_{y \in \Gamma} \hat{\Delta}(y) \right\} \tag{3.6}$$

for Γ_0, where Ξ is a sufficiently rich class of candidate boundaries, containing Γ_0, and being such that for each $\Gamma \in \Xi$ it holds that $\Gamma \subset [\xi, 1-\xi]^d$ for an arbitrarily small fixed $\xi > 0$ (compare the comment at the end of this section).

This estimator corresponds to the boundary in Ξ which maximizes the minimal maximum difference of directional limits achieved for the points belonging to this boundary. Intuitively, the smaller $\inf_{y \in \Gamma} \hat{\Delta}(y)$ is for a boundary Γ, the more likely it is that Γ ventures away from Γ_0 into the "continuous" part of g where $\Delta(x) = 0$ according to (3.5), so that an estimated boundary on which $\inf_{y \in \Gamma} \hat{\Delta}(y)$ is maximized should be close to Γ_0. This heuristic idea will be made more precise in Section 4, where we will show that $\hat{\Gamma}$ (3.6) almost surely stays within a sequence of neighborhoods of Γ_0 which converges to Γ_0 as $n \to \infty$.

We still need to specify the estimators $\hat{\Delta}(x)$ of $\Delta(x)$ for $x \in [0, 1]^d$ which are essential for the boundary estimates $\hat{\Gamma}$ (3.6). We use a kernel method to construct such estimators, but other asymptotically equivalent smoothing methods, in particular locally weighted least squares estimators, could be used as well; the results would remain essentially unchanged for locally weighted least squares estimators.

Let $(K_\gamma), \gamma \in \mathbb{R}$, be a family of kernel functions with the following properties:

(K1) support $(K_\gamma) = [\gamma, 2+\gamma]$,

$$\text{(K2)} \quad \int K_\gamma(x)\, x^i dx = \begin{cases} 1, & i=0 \\ 0, & i=1, ..., k-1 \\ \neq 0, & i=k. \end{cases}$$

For each γ, K_γ is integrable and

(K3) $\sup |K_\gamma(\cdot)| < \infty$.

(K4) For each γ, there exists some ζ, $0 < \zeta \leqslant 1$, such that for all bounded funcions $\xi_n(v)$ satisfying $\xi_n = \sup|\xi_n(v)| \to 0$ as $n \to \infty$, it holds that

$$\int |K_\gamma(v + \xi_n(v)) - K_\gamma(v)|\, dv = O(\xi_n^\zeta).$$

Note that (K1), (K2) relate to a kernel function of order k with support $[\gamma, 2+\gamma]$; for $\gamma = -1$, these are usual kernels with symmetric support. For γ with $-1 < \gamma \leqslant 0$, $-2 \leqslant \gamma < -1$, these kernels correspond to boundary kernels (see, for instance, Müller, 1991); for $\gamma < -2$ or $\gamma > 0$ they correspond to "extrapolation kernels." The boundedness condition (K3) is satisfied for all practically relevant kernel functions, and condition (K4) is satisfied with $\zeta = 1$ for kernel functions which are Lipschitz continuous on $\mathbb{R}$. It is easy to see that one sufficient for (K4) is Lipschitz continuity on $\mathbb{R}$ except at a finite number of discontinuities where one-sided limits exist. Note that therefore kernels with discontinuities at the endpoints of their support like the rectangular kernel satisfy (K4).

Set now $\gamma = \delta$ with δ as defined in (A4). We consider a sequence of bandwidths $b = b(n)$ satisfying

(K5) $b \to 0,\ nb^d/\log n \to \infty \quad \text{as} \quad n \to \infty$.

For any vector $x \in \mathbb{R}^d$, let $x_{(j)}$ be its jth component, and let x/b denote $(x_{(1)}/b, ..., x_{(d)}/b)$. Furthermore, let A_φ be the rotation with the center at 0 corresponding to an angle $\varphi \in [0, 2\pi]^{d-1}$, i.e., if $x \in \mathbb{R}^d$ has the polar coordinates (r_x, φ_x), $r_x \in \mathbb{R}$, $\varphi_x \in [0, 2\pi]^{d-1}$, then $A_\varphi x$ has the polar coordinates $(r_x, \varphi + \varphi_x)$. Note that according to (A2), there exists a partition of $[0, 1]^d$ into n rectangles $A_{i,n}$ of equal size such that $x_{i,n} \in A_{i,n}$, where the $x_{i,n}$ are the gridpoints defined in (A1), (A2).

For $v \in \mathbb{R}^d$, define

$$\tilde{K}(v) = \sum_{j=2}^{d} K_{-1}(v_{(j)}) K_\delta(v_{(1)}). \tag{3.7}$$

Then for $x \in [0, 1]^d$, $\varphi \in [0, 2\pi]^{d-1}$, the *directional limit estimator* is

$$\hat{g}(x, \varphi) = b^{-d} \sum_{i=1}^{n} y_{i,n} \int_{A_{i,n}} \tilde{K}\left(\frac{A_\varphi^{-1}(u-x)}{b}\right) du. \tag{3.8}$$

Observe that $A_\varphi^{-1} = A_{-\varphi}$. The effective support of the transformed kernel is then

$$S_{x,\varphi} = \tilde{A}_{\varphi,x}\left\{[x_{(1)} + \delta b, x_{(1)} + (\delta+2)b] \times \prod_{j=2}^{d} [x_{(j)} - b, x_{(j)} + b]\right\}, \tag{3.9}$$

where $\tilde{A}_{\varphi,x}$ is the rotation corresponding to an angle φ with center at x. This is a cube of volume $(2b)^d$. This cube is obtained from a cube with midpoint x and sidelength $2b$ with sides parallel to the coordinate axes, by a translation by $b + b\delta$ in the direction of the first coordinate $x_{(1)}$ such that the distance between x and the the shifted cube is $b\delta$, and a subsequent rotation $\tilde{A}_{\varphi,x}$ centered around x. Note that the distance between x and this rotated cube is also $b\delta$. Therefore, estimator (3.8) is a (weighted) average of data obtained in a cubic area lying in the direction φ away from x, where the distance $b\delta$ between this cubic area and x goes to 0 as $n \to \infty$.

The moments of the kernel function need to satisfy (K2) which implies that applying this kernel estimator corresponds to extrapolating the function observed on the cube (3.9) onto x. Therefore the directional limit estimator (3.8) corresponds to an approximation of $g(x)$ in the direction φ, and as $n \to \infty$ we can expect that it converges to the corresponding directional limit $g(x, \varphi)$ (3.1).

Having constructed an estimator for directional limits, we obtain $\hat{\Gamma}$ (3.6) by setting

$$\hat{\Delta}(x) = \sup_{\varphi \in [0, 2\pi]^{d-1}} |\hat{g}(x, \varphi) - \hat{g}(x, \bar{\varphi})|, \tag{3.10}$$

where $\bar{\varphi} = \pi + \varphi$, in accordance with (3.2).

It remains to provide explicit constructions of kernels K_γ satisfying (K1)–(K4). For this purpose we consider ultraspherical polynomials $\bar{P}_{j\mu}$ on $[-1, 1]$, i.e., polynomials which are orthonormal with respect to the scalar product $(f, g) = \int f(x)\, g(x)\, \bar{G}_\mu(x)\, dx$, where the weight function is given by $\bar{G}_\mu(x) = (1+x)^\mu (1-x)^\mu$ for some $\mu \geqslant 0$, see Szegö (1975). For $\mu = 0$, these polynomials are the normalized Legendre polynomials on $[-1, 1]$. The normalized ultraspherical polynomials on an interval $[c, d]$, $c < d$, are then obtained by means of the transformation formula

$$P_{j\mu}(x) = \left(\frac{2}{d-c}\right)^{\mu+1/2} \bar{P}_{j\mu}\left(2\frac{x-c}{d-c} - 1\right).$$

and the corresponding weight function is $G_\mu(x) = (x-c)^\mu (d-x)^\mu$.

Define now the corresponding polynomials and weight function on $[\gamma, 2+\gamma]$, i.e.,

$$P_{j\mu} = (\gamma, x) = \bar{P}_{j\mu}(x - \gamma - 1), \tag{3.11}$$

$$G_\mu(\gamma, x) = (2 + \gamma - x)^\mu (x - \gamma)^\mu. \tag{3.12}$$

We set $K_\gamma(x) = P(\gamma, x)\, G_\mu(\gamma, x)$, where $K_\gamma(\cdot)$ is supposed to be a kernel function with support $[\gamma, 2+\gamma]$ satisfying (K1)–(K4), and $P(\gamma, \cdot)$ is assumed to be a polynomial of degree $(k-1)$.

According to (K2), we find

$$\int_\gamma^{2+\gamma} P(\gamma, x)\, P_{j\mu}(\gamma, x)\, G_\mu(\gamma, x)\, dx = P_{j\mu}(\gamma, 0), \qquad 0 \leqslant j < k,$$

and therefore

$$P(\gamma, x) = \sum_{j=0}^{k-1} P_{j\mu}(\gamma, 0)\, P_{j\mu}(\gamma, x). \tag{3.13}$$

This motivates the following result.

LEMMA 3.1. *The family of kernels*

$$K_\gamma = \left\{ \sum_{j=0}^{k-1} P_{j\mu}(\gamma, 0)\, P_{j\mu}(\gamma, x) \right\} G_\mu(\gamma, x), \tag{3.14}$$

where $P_{j\mu}$, G_μ *are defined in* (3.11), (3.12), *satisfies conditions* (K1)–(K4), *where* (K4) *holds with* $\zeta = 1$.

Proof. Property (K1) is satisfied since $P_{j\mu}(\gamma, \cdot)$ and $G_\mu(\gamma, \cdot)$ both have support $[\gamma, 2+\gamma]$, and (K3) is obvious. For (K4), we use the fact that $K_\gamma(\cdot)$ (3.14) is Lipschitz continuous on $\mathbb{R}$ for $\mu \geqslant 1$ and is Lipschitz continuous on $\mathbb{R}$ except at the two points γ and $2+\gamma$ for $\mu=0$, where a discontinuity occurs. As discussed after (K4), in either case (K4) holds with $\zeta=1$. It remains to show (K2). Let

$$x^i=\sum_{j=0}^{\infty}\lambda_{ji}P_{j\mu}(\gamma,x)=\sum_{j=0}^{i}\lambda_{ji}P_{j\mu}(\gamma,x)$$

be the unique representation of the monomial x^i in the system $P_{j\mu}(\gamma,\cdot)$, where

$$\lambda_{ji}=\int P_{j\mu}(\gamma,x)\,x^iG_\mu(\gamma,x)\,dx,\qquad \text{for}\quad 0\leqslant j\leqslant i,\quad \lambda_{ji}=0\qquad \text{for}\quad j>i.$$

Then

$$\begin{aligned}\int K_\gamma(x)\,x^i dx&=\int\sum_{j=0}^{k-1}P_{j\mu}(\gamma,0)\,P_{j\mu}(\gamma,x)\,x^iG_\mu(\gamma,x)\,dx\\&=\sum_{j=0}^{k-1}\int P_{j\mu}(\gamma,x)\,x^iG_\mu(\gamma,x)\,dx\cdot P_{j\mu}(\gamma,0)\\&=\sum_{j=0}^{k-1}\lambda_{ji}P_{j\mu}(\gamma,0)=\sum_{j=0}^{i}\lambda_{ji}P_{j\mu}(\gamma,0)=x^i\Big|_{x=0}=\begin{cases}1, & i=0\\0, & 0<i\leqslant k-1\end{cases},\end{aligned}$$

which implies (K2). ∎

According to (3.11)–(3.13), this family of kernels can be written as

$$K_\gamma(x)=\left\{(2+\gamma-x)^\mu(x-\gamma)^\mu\sum_{j=0}^{k-1}\bar{P}_{j\mu}(-1-\gamma)\,\bar{P}_{j\mu}(x-\gamma-1)\right\}1_{[\gamma,\,2+\gamma]}. \tag{3.15}$$

Explicit formulas for the $\bar{P}_{j\mu}$ are available from Szegő (1975). For example, $\bar{P}_{00}(x)=2^{-1/2}$, $\bar{P}_{10}(x)=(3/2)^{1/2}x$, $\bar{P}_{01}(x)=3^{1/2}/2$, and $\bar{P}_{11}(x)=(15^{1/2}/2)x$. This implies

$$K_\gamma(x)=\tfrac{1}{2}(1-3(1+\gamma)(x-(1+\gamma)))\,1_{[\gamma,\,2+\gamma]}\qquad \text{for}\quad \mu=0,\,k=2;$$

$$K_\gamma(x)=\tfrac{3}{4}(2+\gamma-x)(x-\gamma)(1-5(1+\gamma)(x-\gamma-1))\,1_{[\gamma,\,2+\gamma]}\qquad \text{for}\ \mu=1,\,k=2.$$

The special cases of a symmetrically supported kernel, i.e., for $\gamma = -1$, correspond to $K_{-1}(x) = \frac{1}{2} 1_{[-1,1]}$ for $\mu = 0$, $k = 2$, i.e., the rectangular kernel, and to $K_{-1}(x) = \frac{3}{4}(1 - x^2) 1_{[-1,1]}$ for $\mu = 1$, $k = 2$, i.e., the parabolic or Bartlett–Priestley–Epanechnikov kernel.

Note that the corresponding kernels $\tilde{K}$ for the estimation of directional limits are obtained from univariate kernels K_γ (3.15) by means of (3.7). Further, the condition that $\Gamma \subset [\xi, 1-\xi]^d$ for all $\Gamma \in \Xi$ stated after (3.6) can now be relaxed to $\Gamma \subset [3d^{1/2}b, 1 - 3d^{1/2}b]^d$, $\Gamma \in \Xi$. Since $b \to 0$, this is a weaker requirement. This condition conveniently allows us to ignore boundary effects when estimating directional limits.

4. Strong Consistency of Boundary Estimators

We consider here consistency properties of the maximin estimators $\hat{\Gamma}$ (3.6) for the boundary Γ_0. The proof relies on uniform convergence results for the directional limit estimators $\hat{g}(x, \varphi)$ (3.8) which will be discussed and proved in the following section.

An additional technical requirement on the sequence of bandwidths $b = b(n)$ is

(K6) For one arbitrary fixed number r which satisfies $4 < r < s$ and where s is as in (A6), it holds that

$$\liminf_{n \to \infty} n^{-2/r} (nb^d \log n)^{1/2} > 0.$$

A further restriction connects the sequence of bandwidths b with the sequence of jump sizes Δ_n as defined in (A5):

(K7) $\{b^k + n^{-1/d} + (\log n/nb^d)^{1/2}\}/\Delta_n \to 0 \quad$ as $\quad n \to \infty$.

For any $\rho > 0$ define a ρ-neighborhood $U_\rho(\Gamma_0)$ of the boundary Γ_0 as

$$U_\rho(\Gamma_0) = \bigcup_{y \in \Gamma_0} S(y; \rho),$$

where for any $x \in \mathbb{R}^d$, $S(x; \rho) = \{z \in \mathbb{R}^d : \|x - z\| \leqslant \rho\}$ is the ball with radius ρ around x, $\|\cdot\|$ denoting the Euclidean norm in $\mathbb{R}^d$. Note that for the Hausdorff distance between $U_\rho(\Gamma_0)$ and Γ_0, observing that $\Gamma_0 \subset U_\rho(\Gamma_0)$,

$$d(U_\rho(\Gamma_0), \Gamma_0) = \sup_{x \in U_\rho(\Gamma_0)} \inf_{y \in \Gamma_0} \|x - y\| \leqslant \rho \tag{4.1}$$

The following separability property is useful.

LEMMA 4.1. *For any* $x \in [0, 1]^d \backslash U_\rho(\Gamma_0)$, $S(x; \rho/2) \cap U_{\rho/3}(\Gamma_0) = \varnothing$.

Proof. Assume $z \in S(x; \rho/2) \cap U_{\rho/3}(\Gamma_0)$. Then $\|z - x\| \leqslant \rho/2$, and there exists $y_0 \in \Gamma_0$ with $\|z - y_0\| \leqslant \rho/3$, which leads to the contradiction $x \in U_\rho(\Gamma_0)$. ∎

Denote by $U_b(\Gamma_0) = U_{b(n)}(\Gamma_0)$ a sequence of $b(n)$-neighborhoods of Γ_0. Since according to (K5), $b \to 0$ as $n \to \infty$, $U_b(\Gamma_0) \to \Gamma_0$ as $n \to \infty$. Our central result is

THEOREM 4.1. *Under* (A1)–(A6), (K1)–(K7), *setting* $\beta_n = 6d^{1/2}b(n)$,

$$P(\hat{\Gamma} \in U_{\beta_n}(\Gamma_0) \text{ for sufficiently large } n) = 1. \tag{4.2}$$

Furthermore

$$d(\hat{\Gamma}, \Gamma_0) = O(b(n)) \qquad a.s., \tag{4.3}$$

where $d(A, B)$ *denotes the Hausdorff distance between sets* A, B.

The proof is given at the end of Section 5. Note that (4.3) is an immediate consequence of (4.2) in view of (4.1). The special case $d = 1$ was investigated in Müller (1992).

COROLLARY 4.1. *Under* (A1)–(A6), (K1)–(K7), *assuming that* $s > 2d$ *in* (A6), *one obtains for fixed jump sizes* $\Delta_n = \Delta$ *and any given* η *with* $0 < \eta < s - 4$ *the rate of convergence*

$$d(\hat{\Gamma}, \Gamma_0) = O\{(n^{-1 + 4/(s-\eta)}(\log n)^{-1})^{1/d}\} \ a.s. \tag{4.4}$$

In order to see how this follows from (4.3), consider the restrictions imposed by (K5)–(K7). For fixed Δ_n, the main restriction is (K6). The minimal b compatible with (K6) is of the type given on the r.h.s. of (4.4), using $r = s - \eta$.

Note that $s > 4$ in (A6), which ensures that all restrictions can be satisfied simultaneously, so that in (4.3) the rate is at least $o(1)$ a.s. According to (4.4), under these minimal assumptions, the rate is seen to be actually $O\{(n^{-\eta'}(\log n)^{-1})^{1/d}\}$ for some $\eta' > 0$. Observe that by assuming large enough s in (A6), rates $\{n^{-1+\alpha}(\log n)^{-1}\}^{1/d}$ are achieved for any given $\alpha > 0$ by choosing small values for η.

Considering the case where $\Delta_n \to 0$, we observe that in view of (K5)–(K7) we obtain the same a.s. rate as in (4.4) provided that $(n^{1/d}\Delta_n)^{-1} \to 0$, $\log n/(n^{2/(s-\eta)}\Delta_n) \to 0$, and $(n^{1-4/(s-\eta)} \log n \, \Delta_n^{d/k})^{-1} \to 0$ as $n \to \infty$. If for instance $\Delta_n = n^{-\alpha}$ for some $\alpha > 0$, this would be satisfied if $\alpha < \min\{d^{-1}, 2/(s - \eta), (k/d)(1 - 4/(s - \eta))\}$.

5. Uniform Convergence of Directional Limit Estimators

The results given in this section are needed for the proof of Theorem 4.1 and are also of interest in their own right. Considering kernel estimators $\hat{g}(x, \varphi)$ (3.8) of directional limits $g(x, \varphi)$ (3.1), we decompose the difference

$$\hat{g}(x, \varphi) - g(x, \varphi) = \hat{g}(x, \varphi) - E\hat{g}(x, \varphi) + E\hat{g}(x, \varphi) - g(x, \varphi)$$

into a stochastic and a bias part.

Thereom 5.1. *Let $A \subset [0, 1]^d \times [0, 2\pi]^{d-1}$ and $B \subset [0, 1]^d$ be compact sets such that for any $(x, \varphi) \in A$ it holds that the effective support $S_{x, \varphi}$ (3.9) of the kernel employed in $\hat{g}(x, \varphi)$ satisfies $S_{x, \varphi} \subset B$. Furthermore, assume $g \in \mathscr{C}^k(B)$ and* (A1), (A2), (A6), (K1)–(K6). *Then*

$$\sup_{(x, \varphi) \in A} |\hat{g}(x, \varphi) - g(x, \varphi)| = O\{b^k + n^{-1/d} + (\log n/nb^d)^{1/2}\} \qquad a.s.$$

The proof follows immediately from the subsequent two lemmas for the bias and stochastic parts which hold under the assumptions of Theorem 5.1.

Lemma 5.1. $\sup_{(x, \varphi) \in A} |E\hat{g}(x, \varphi) - g(x, \varphi)| = O(b^k + n^{-1/d})$.

Proof. Let the rotation A_φ be as defined before (3.7) and observe that by a Riemann sum approximation argument, for given $(x, \varphi) \in A$, using that $g \in \mathscr{C}^k(B)$,

$$E\hat{g}(x, \varphi) = b^{-d} \int g(u)\, \tilde{K}(b^{-1}A_\varphi^{-1}(u - x))\, du + O(n^{-1/d})$$

$$= \int g(x + bA_\varphi v)\, \tilde{K}(v)\, dv + O(n^{-1/d}),$$

where for $x \in \mathbb{R}^d$, $bx = (bx_{(1)}, ..., bx_{(d)})$ and $\tilde{K}$ is as in (3.7). A multivariate Taylor expansion of $g(\cdot, \varphi)$ around $g(x, \varphi)$, noting that $\operatorname{supp}\{g(x + bA_\varphi v)\, \tilde{K}(v)\} \subset B$ due to (K1), then yields

$$\int g(x + bA_\varphi v)\, \tilde{K}(v)\, dv$$

$$= g(x, \varphi) + \sum_{l=1}^{k-1} b^l \sum_{|\alpha| = l} \frac{1}{\alpha!} D^\alpha g(x, \varphi) \int (A_\varphi v)^\alpha \tilde{K}(v)\, dv$$

$$+ b^k \sum_{|\alpha| = k} \frac{1}{\alpha!} D^\alpha g(\xi, \varphi) \int (A_\varphi v)^\alpha \tilde{K}(v)\, dv,$$

where $\alpha = (\alpha_1, ..., \alpha_d)$ is a multi-index, $\alpha! = \alpha_1!...\alpha_d!$, $x^\alpha = x_{(1)}^{\alpha_1}...\alpha_{(d)}^{\alpha_d}$, $D^\alpha g(x, \varphi) = (\partial^{\alpha_1}...\partial^{\alpha_d}/\partial^{\alpha_1}v_{(1)}...\partial^{\alpha_d}v_{(d)})\, g(v, \varphi)|_{v=x}$ is the mixed partial derivative of order α of $g(x, \varphi)$, and ξ is an intermediate value between x and $x + bA_\varphi v$, $\xi \in B$. We used $\int \tilde{K}(v)\, dv = 1$, which is a consequence of the moment conditions (K2). These conditions also imply that $\int \tilde{K}(v)\, v^\alpha dv = 0$ for α with $0 < |\alpha| < k$, so that we conclude $E\hat{g}(x, \varphi) - g(x, \varphi) = O(n^{-1/d} + b^k)$, for the chosen point $(x, \varphi) \in A$. It is now easy to see that these remainder terms are uniform over compact sets. ∎

LEMMA 5.2. $\sup_{(x,\varphi)\in A} |\hat{g}(x, \varphi) - E\hat{g}(x, \varphi)| = O(\{\log n/nb^d\}^{1/2})$ *a.s.*

Proof. Let $\sigma_n = \{\log n/nb^d\}^{1/2}$. Consider a partition (D_j) of A such that $d(D_j) = \sup_{x,y\in D_j} \|x - y\| \leqslant n^{-3/\zeta}$, where ζ is as in (K4). For each j, fix a point $\tau_j \in D_j$, and define the function $\tau: A \to \mathbb{R}$ by $\tau(t) = \arg\min_{\tau_j} \|\tau_j - t\|$ for $t \in A$. For r as in (K6), define the truncated errors $\tilde{\varepsilon}_i = \varepsilon_i 1_{\{|\varepsilon_i| \leqslant (in)^{1/r}\}}$. Then, defining

$$W_i(t) = b^{-d} \int_{A_{i,n}} \tilde{K}\left(\frac{A_\varphi^{-1}(u - x)}{b}\right) du,$$

for $t = (x, \varphi)$, we have

$$\begin{aligned}\sup_{(x,\varphi)\in A} |\hat{g}(x, \varphi) - E\hat{g}(x, \varphi)| &\sup_{t\in A} \sum |W_i(t)\, \varepsilon_i| \\ &\leqslant \sup_{t\in A} \sum |\{W_i(t) - W_i(\tau(t))\}\, \varepsilon_i| \\ &\quad + \sup_{t\in A} \left|\sum W_i(\tau(t))(\varepsilon_i - \tilde{\varepsilon}_i)\right| \\ &\quad + \sup_{t\in A} \left|\sum W_i(\tau(t))\, \tilde{\varepsilon}_i\right| = I + II + III.\end{aligned}$$

Now, observe that by the strong law of large numbers,

$$\begin{aligned} I &\leqslant \sup_{t\in A} \left\{\sum [W_i(t) - W_i(\tau(t))]^2\right\}^{1/2} \left\{\sum \varepsilon_i^2\right\}^{1/2} \\ &= O\left(n^{1/2} \left\{\sup_{t\in A} \sum [W_i(t) - W_i(\tau(t))]^2\right\}^{1/2}\right) \quad \text{a.s.} \end{aligned} \tag{5.1}$$

In the following, $\|x\|$ is the Euclidean norm for a vector $x \in \mathbb{R}^d$ and $\|A\|$ is the matrix norm $\|A\| = \sup_{\|x\|=1} \|Ax\|$ for a matrix A.

By (K3), (K4), substituting $v = b^{-1} A_{\varphi}^{-1}(u - x)$ and letting $t = (x, \varphi)$, $t' = (x', \varphi')$,

$$\sup_{t \in A} \sum [W_i(t) - W_i(\tau(t))]^2$$

$$\leqslant c' \max_j \sup_{t, t' \in D_j} \frac{1}{nb^{2d}} \sum \int_{A_{i,n}} \left| \tilde{K}\left(\frac{A_{\varphi}^{-1}(u-x)}{b}\right) - \tilde{K}\left(\frac{A_{\varphi'}^{-1}(u-x')}{b}\right) \right| du$$

$$= c' \max_j \sup_{t, t' \in D_j} \frac{1}{nb^d} \int \left| \tilde{K}(v) - \tilde{K}\left(A_{\varphi'}^{-1} A_{\varphi} v - \frac{A_{\varphi'}^{-1}(x'-x)}{b}\right) \right| dv$$

$$= O(n^{-1}\{b^{-d}\|A_{\varphi} - A_{\varphi'}\| + b^{-2d}\|x' - x\|\}^{\zeta}) = O([n^4 b^{2d}]^{-1})$$

as $\zeta \leqslant 1$, and

$$\int |\tilde{K}(v) - \tilde{K}(v + \xi_n)| \, dv$$

$$\leqslant C \int \left\{ \sum_{j=2}^{d} |K_{-1}(v_{(j)} + \xi_{n(j)})| + |K_{\delta}(v_{(1)}) - K_{\delta}(v_{(1)} + \xi_{n(1)})| \right\} dv$$

$$= O\left\{ \sum_{j=1}^{d} |\xi_{n(j)}|^{\zeta} \right\} = O(\|\xi_n\|^{\zeta}).$$

Relations (5.1) and (5.2) imply

$$I = O(\sigma_n). \tag{5.3}$$

As for II, by the same arguments as in Lemma 5.2 of Müller and Stadtmüller (1987), one shows that

$$\sup_{t \in A} \left| \sum W_i(\tau(t))(\varepsilon_i - \tilde{\varepsilon}_i) \right| = O\,(n^{2/r} \sup_{\substack{t \in A \\ 1 \leqslant i \leqslant n}} |W_i(t)|) \qquad \text{a.s.,}$$

and with $\sup_{t \in A,\, 1 \leqslant i \leqslant n} |W_i(t)| \leqslant c/(nb^d)$ then one obtains from (K6) that

$$\mathrm{II} = O(\sigma_n) \qquad \text{a.s.} \tag{5.4}$$

Observing that due to (K1), the number of nonzero weights in the estimator (3.8) is $O(nb^d)$, one finds that

$$\sup_{t \in A} \left(\sum W_i^2(t) \log n \right)^{1/2} = O(\sigma_n).$$

Under this condition, term III can be handled as in the univariate case, see Müller and Stadtmüller (1987), to yield

$$\mathrm{III} = O(\sigma_n). \tag{5.5}$$

The result follows immediately from (5.3), (5.4), and (5.5). ∎

Proof of Theorm 4.1. Set $\gamma_n = b^k + n^{-1/d} + (\log n/nb^d)^{1/2}$. Choosing $A_1 = \{[3d^{1/2}b, 1-3d^{1/2}b]^d \backslash U_{\beta_n}(\Gamma_0)\} \times [0, 2\pi]^{d-1}$, $B_1 = [0, 1]^d \backslash U_{\beta_n/3}(\Gamma_0)$, Theorem 5.1 implies

$$\sup_{(x,\varphi)\in A_1} |\hat{g}(x, \varphi) - g(x, \varphi)| = O(\gamma_n) \qquad \text{a.s.} \tag{5.6}$$

Note that according to (3.9) and (A4) it follows from Pythagoras' Theorem for the effective support $S_{x,\varphi}$ (3.9) of the kernel employed in $\hat{g}(x, \varphi)$ that $S_{x,\varphi} \subset S(x, 3d^{1/2}b)$, where $S(x, \rho)$ is defined after (K7). Together with Lemma 4.1 this implies that $S_{x,\varphi} \cap U_{\beta_n/3}(\Gamma_0) = \varnothing$, so that either $S_{x,\varphi} \subset C_1 \backslash U_{\beta_n/3}(\Gamma_0)$ or $S_{x,\varphi} \subset C_2 \backslash U_{\beta_n/3}(\Gamma_0)$(see (A3)), i.e., $S_{x,\varphi} \subset B_1$. Obviously, $g \in \mathscr{C}^k(B_1)$.

Furthermore, choosing $A_2 = \{(y, \varphi(y)),\ y \in \Gamma_0\}$, $\bar{A}_2 = \{(y, \bar{\varphi}(y)),\ y \in \Gamma_0\}$, where $\varphi(y)$, $\bar{\varphi}(y)$ are the angles defined by the rotation part of the orthonormal mapping T_y in (A4), as in (3.3), and $B_2 = C_1$, $\bar{B}_2 = C_2$, as suggested by (3.3), Theorem 5.1 implies

$$\begin{aligned} &\sup_{y\in\Gamma_0} |\hat{g}(y, \varphi) - g(y, \varphi)| = O(\gamma_n) \qquad \text{a.s.,} \\ &\sup_{y\in\Gamma_0} |\hat{g}(y, \bar{\varphi}) - g(y, \bar{\varphi})| = O(\gamma_n) \qquad \text{a.s.,} \end{aligned} \tag{5.7}$$

as $g \in \mathscr{C}^k(C_1)$, $g \in \mathscr{C}^k(C_2)$, defining g on Γ_0 by the corresponding limits

$$g(y, \varphi(y)) = \lim_{n\to\infty} g(y_{1n}), \qquad y_{1n} \in C_1,$$

$$g(y, \bar{\varphi}(y)) = \lim_{n\to\infty} g(y_{2n}), \qquad y_{2n} \in C_2, \quad \text{according to (2.1).}$$

From (5.6), (5.7), we obtain in conjuction with (3.4), (3.5)

$$\sup_{x\in A_1\backslash U_{\beta_n}(\Gamma_0)} \hat{\Delta}(x) = O(\gamma_n) \qquad \text{a.s.} \tag{5.8}$$

and

$$\begin{aligned} \inf_{y\in\Gamma_0} \hat{\Delta}(y) &\geqslant \inf_{y\in\Gamma_0} \sup_{\varphi(y)} |\hat{g}(y, \varphi(y)) - \hat{g}(y, \bar{\varphi}(y))| \\ &= \inf_{y\in\Gamma_0} \sup_{\varphi(y)} |g(y, \varphi(y)) - g(y, \bar{\varphi}(y))| + O(\gamma_n) \qquad \text{a.s.} \\ &\geqslant \Delta_n + O(\gamma_n) \qquad \text{a.s.,} \end{aligned} \tag{5.9}$$

where $\hat{\Delta}(\cdot)$ is defined in (3.10), and the jump size Δ_n is defined in (A5).

It follows from (5.8), (5.9), and (K7) that for any $\Gamma \in \Xi$ with $\Gamma \not\subseteq U_{\beta_n}(\Gamma_0)$ it holds that for sufficiently large n,

$$\inf_{y \in \Gamma_0} \hat{\Delta}(y) > \inf_{x \in \Gamma} \hat{\Delta}(x) \qquad \text{a.s.},$$

and according to the definition (3.6) of $\hat{\Gamma}$, this implies (4.2). ▮

Acknowledgments

This research was partially supported by NSF Grants DMS 90-02423 and DMS 93-05484.

References

[1] Brodskii, B., and Dharkhovskii, B. (1986). The a posteriori method of detecting the disruption of a random field. In *Detection of Changes in Random Processes* (L. Telksnys, Ed.), pp. 32–38. Optimization Software, New York.

[2] Carlstein, E., and Krishnamoorthy, C. (1992). Boundary estimation. *J. Amer. Statist. Assoc.* **87** 430–438.

[3] Hall, P., and Titterington, D. M. (1992). Edge and peak preserving smoothing *Technometrics* **34** 429–440.

[4] Korostelev, A. P., and Tsybakov, A. B. (1993). *Minimax Theory of Image Reconstruction.* Lecture Notes in Statistics, Vol. 82, Springer, New York.

[5] McDonald, J. A., and Owen, A. B. (1986). Smoothing with split linear fits. *Technometrics* **28** 195–208.

[6] Müller, H. G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika* **78** 321–330.

[7] Müller, H. G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* **20** 737–761.

[8] Müller, H. G. and Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15** 610–625.

[9] Rudemo, M. and Stryhn, H. (1993a). Boundary estimation for star-shaped objects. In *Change-Point Problems* (E. Carlstein, H. G. Müller, and D. Siegmund, Eds.), IMS Lecture Notes and Monographs Series, in press.

[10] Rudemo, M., and Stryhn, H. (1993b). Approximating the distribution of maximum likelihood contour estimators in two-region images. *Scand. J. Statist.*, in press.

[11] Szegő, G. (1975). *Orthogonal Polynomials.* Amer. Math. Soc. Providence, RI.

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

[13] Tsybakov, A. B. (1993). Multidimensional change-point problems and boundary estimation. In *Change-Point Problems* (E. Carlstein, H. G. Müller, and D. Siegmund, Eds.), IMS Lecture Notes and Monographs Series, in press.