CHANGE CURVE ESTIMATION BY WAVELETS

Yazhen Wang *

To be published in *Journal of the American Statistical Association*

*Yazhen Wang is Assistant Professor of Statistics, University of Missouri-Columbia, Columbia, MO 65211. Research is supported in part by NSF Grant DMS-9404142 and Grants from Research Board at the University of Missouri and Research Council at the University of Missouri-Columbia. The author would like to thank two anonymous referees, an Associate Editor and the Editor for their stimulating comments that led to significant improvement of the paper.

Abstract

The recently developed theory of wavelets has a remarkable property to "zoom in" on very short lived frequency phenomena, such as transients in signals and singularities in functions, and hence provides an ideal tool to study localized changes. This paper proposes a wavelet method to estimate jump and sharp cusp curves of a function in the plane. The method is first to compute wavelet transformation of data and then to estimate jump and sharp cusp curves by wavelet transformation across fine scales. Asymptotic theory is established and simulations are carried out to lend some credence to the asymptotic theory. The wavelet estimate is nearly optimal and can be computed by fast algorithms. The method is applied to a real image.

KEY WORDS: Asymptotic theory; Boundary estimation; Edge estimation; Image processing; Jump curve; Sharp cusp curve; Multidimensional change-point; Wavelet transformation.

1. INTRODUCTION

Suppose function $f(x,y)$, $(x,y) \in [0,1]^2$, is observed from model

$$z_{i_1 i_2} = f(x_{i_1}, y_{i_2}) + \varepsilon_{i_1 i_2}, \quad i_1, i_2 = 1, \cdots, n = 2^J, \tag{1}$$

where $x_{i_1} = i_1/n$, $y_{i_2} = i_2/n$, and $\varepsilon_{i_1 i_2}$ are i.i.d. normal errors with mean zero and variance σ^2. The unknown function f has sharp cusps (defined below) along some curves in $(0,1)^2$, and our goal is to estimate these sharp cusp curves.

We say f has α sharp cusps $(0 \leq \alpha < 1)$ along a continuous curve θ in $(0,1)^2$, if for each $(x,y) \in \theta$, there exists a $\beta_0 \in [0, 2\pi)$ such that for $\beta \in (\beta_0 - \delta, \beta_0 + \delta)$, as $h \to 0$,

$$|f(x + h\,cos(\beta), y + h\,sin(\beta)) - f(x,y)| \geq K|h|^\alpha,$$

where K and δ are positive constants depending only on θ and f. $\alpha = 0$ corresponds to discontinuities along θ.

Estimation of jump and sharp cusp curves is a change curve estimation problem, often referred to as a multidimensional change-point, boundary estimation or edge estimation problem. Edge estimation plays an important role in signal and image processing and a large variety of edge estimation methods have been proposed in different contexts (Blake and Zisserman 1987; Marr 1982; Pratt 1991). Boundaries (edges) are local features of a function (an image), and thus nonparametric approaches are most suitable for boundary estimation problems. Several nonparametric methods were proposed to estimate boundaries for various models. Korostelev and Tsyabkov (and their co-authors) established asymptotic minimax theory for the problem of estimating a boundary curve and the support of a density (Korostelev 1991; Korostelev, Simar, and Tsyabkov 1995; Korostelev and Tsyabkov 1992, 1993a, b; Mammen and Tsyabkov 1995; Tsyabkov 1994). They showed that estimates constructed by piecewise polynomial approximation and sieve maximum likelihood methods can achieve the optimal convergence rate. For nonparametric regression with general random errors, Müller and Song (1994a) proposed a maximin method based on directional limits and multivariate kernel estimates to estimate a jump curve, and established asymptotics for their estimate. Puri and Ruymgaart (1994) considered nonparametric regression models with dependent random errors, and used a kernel method to reduce data to nearly black objects and then recovered a boundary curve. Methods based on penalizing and regularization techniques were developed by Shiau, Wahba and Johnson (1986) and Titterington (1991). Carlstein and Krishnamoorthy (1992, 1994) and Rudemo, Skovgaard and Stryhn (1990) studied a general problem of estimating a boundary curve across which distributions of observed random variables change.

The theory of wavelets is a recent breakthrough in applied mathematics. Because of the interdisciplinary origin, wavelets appeal to scientists and engineers of many different backgrounds as a fairly simple mathematical tool with a great variety of applications. Wavelets permit decomposition of functions into localized oscillating components and thus provide an ideal tool to study localized changes of a function. Fast algorithms are available to numerically carry out wavelet analysis. Several methods based on wavelets are proposed to detect singularities and edges (Antoine, Murenzi and Piette 1992; Froment and Mallat 1992; Mallat and Zhong 1992a, b, c; Mallat and Hwang 1992). However, these methods are proposed mostly in a deterministic setting and are generally unsuitable for noisy data. We proposed wavelet methods to estimate jumps and sharp cusps. Wang (1995 1997b) studied the one dimensional problem for direct or indirect noisy data. In this paper we consider estimation of jump and sharp cusp curves in the plane and establish asymptotics for the wavelet method. The two dimensional problem is much more complicated and difficult. We consider directions and curves and deal with random fields, and totally different techniques are employed to prove asymptotic results.

It is very common in edge estimation problems that estimates achieving (nearly) optimal rates of convergence are often complicated from the computational point of view, while practical data processing algorithms are frequently suboptimal. In contrast, the wavelet estimate is nearly optimal on the one hand and can be computed by fast algorithms on the other.

This paper is organized as follows. Section 2 reviews wavelets. Section 3 describes the methods and establishes asymptotics. Simulated examples and a real image are considered in Section 4. Section 5 presents conclusions and discussion. The technical proofs are collected in the Appendix.

2. WAVELETS

Let (ϕ, ψ) be Daubechies' wavelets (Daubechies 1992, chap. 6), and define wavelets in two dimensions as follows:

$$\Psi^{h,\beta}(x,y) = \phi(x\,cos\beta - y\,sin\beta)\,\psi(x\,sin\beta + y\,cos\beta),$$

$$\Psi^{v,\beta}(x,y) = \psi(x\,cos\beta - y\,sin\beta)\,\phi(x\,sin\beta + y\,cos\beta),$$

$$\Psi^{d,\beta}(x,y) = \psi(x\,cos\beta - y\,sin\beta)\,\psi(x\,sin\beta + y\,cos\beta),$$

where h, v and d stand for "horizontal," "vertical" and "diagonal," respectively, and $\beta \in [0, \pi/4]$ is an orientation turning parameter (Antoine, Murenzi and Piette 1992; Daubechies 1992; and Mallat and Zhong 1993a, b, c).

Let $\Psi_s^{\gamma,\beta}(x,y) = s^{-1}\Psi^{\gamma,\beta}(x/s, y/s)$ for $\gamma = h, v, d$. The wavelet transformation of $f(x,y)$ is defined to be

$$T^{\gamma,\beta} f(s,x,y) = \int \Psi_s^{\gamma,\beta}(x-u, y-v)\, f(u,v)\, du dv,$$

and

$$Tf(s,x,y) = \max_{0 \le \beta \le \pi/4} \{|T^{h,\beta} f(s,x,y)|,\ |T^{v,\beta} f(s,x,y)|,\ |T^{d,\beta} f(s,x,y)|\}.$$

For compactly supported wavelets, the value of $Tf(s,x,y)$ depends upon the value of f in a neighborhood of (x,y) of size proportional to s^2. At fine scales s, $Tf(s,x,y)$ provides localized information such as Lipschitz regularity on $f(x,y)$ as follows.

We say $f(x,y)$ is Lipschitz a at (x_0, y_0) if there exists a positive constant K such that for all β and as $h \to 0$,

$$|f(x_0 + h\,cos\beta, y_0 + h\,sin\beta) - f(x_0, y_0)| \le K|h|^a.$$

Mathematically global and local Lipschitz regularities can be characterized by the asymptotic decay of wavelet transformation at small scales (Mallat and Zhong 1992a, theorem 2; Mallat and Huang 1992, theorem 6). For example, if f is differentiable, $Tf(s,x,y)$ is of order s^2; if f has α sharp cusps along curve θ, then the minimum of $Tf(s,x,y)$ along θ converges to zero at a rate no faster than $s^{\alpha+1}$ as s tends to zero.

3. CHANGE CURVE ESTIMATION

3.1 Threshold Method

Given data from model (1), we define two processes $Z_n(x,y)$, $(x,y) \in [0,1]^2$, and $W_n(x,y)$, $(x,y) \in [0,1]^2$, via $Z_n(0,0) = W_n(0,0) = 0$ and

$$Z_n(x,y) = n^{-2} \sum_{i_1 \le nx} \sum_{i_2 \le ny} z_{i_1 i_2}, \qquad W_n(x,y) = n^{-2} \sum_{i_1 \le nx} \sum_{i_2 \le ny} \varepsilon_{i_1 i_2}. \tag{2}$$

The wavelet transformations of W_n and Z_n are defined as follows,

$$\begin{aligned} T^{\gamma,\beta} W_n(s,x,y) &= \int \Psi_s^{\gamma,\beta}(x-u, y-v) W_n(du, dv) \\ &= n^{-2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \Psi_s^{\gamma,\beta}(x - i_1/n, y - i_2/n)\, \varepsilon_{i_1 i_2}, \end{aligned}$$

$$TW_n(s,x,y) = \max_{0 \le \beta \le \pi/4} \{ |T^{h,\beta} W_n(s,x,y)|,\ |T^{v,\beta} W_n(s,x,y)|,\ |T^{d,\beta} W_n(s,x,y)| \},$$

$$T^{\gamma,\beta} Z_n(s,x,y) = \int \Psi_s^{\gamma,\beta}(x-u, y-v) Z_n(du, dv),$$

and

$$TZ_n(s,x,y) = \max_{0 \le \beta \le \pi/4} \{ |T^{h,\beta} Z_n(s,x,y)|,\ |T^{v,\beta} Z_n(s,x,y)|,\ |T^{d,\beta} Z_n(s,x,y)| \}.$$

By (1), (2), the above definitions of wavelet transformation and Donoho (1993, sec. 7; 1995, sec. 8) we have

$$T^{\gamma,\beta} Z_n(s,x,y) = T^{\gamma,\beta} f(s,x,y)\,\{1+o(1)\} + T^{\gamma,\beta} W_n(s,x,y). \tag{3}$$

Since $T^{\gamma,\beta} W_n(s,x,y)$ is a normal random variable with zero mean and variance of order $1/n$ (Lemma A.1 in the Appendix), and $Tf(s,x,y)$ is of order s^2 and $s^{\alpha+1}$, respectively, for the two situations that f is differentiable at (x,y) and f has α sharp cusp at (x,y), (3) implies that, at a very fine scale s, $TZ_n(s,x,y)$ is dominated by $TW_n(s,x,y)$, while at a coarse scale s, $Tf(s,x,y)$ controls $TZ_n(s,x,y)$.

We select scales s_n such that $s_n \;\; (n/|log\, n|^{\eta})^{1/(\alpha+1)}$ is bounded away below from zero and above from infinity as $n \to \infty$, with η a constant greater than 1, that is,

$$s_n \asymp (|log\, n|^{\eta}/n)^{1/(\alpha+1)}. \tag{4}$$

s_n is selected such that $TW_n(s_n,x,y)$ is smaller than $Tf(s_n,x,y)$ for nearby (x,y) where f has sharp cusps, and is much larger than $Tf(s_n,x,y)$ for (x,y) where f is differentiable. Define

$$C_n = 2\,\sigma\, n^{-1}\,(log\, n)^{1/2}. \tag{5}$$

With probability tending to one, $TW_n(s_n,x,y)$ is bounded by C_n (Lemma A.6 in the Appendix), and for (x,y) where f is differentiable, $TW_n(s_n,x,y)$ dominates $TZ_n(s_n,x,y)$, so $TZ_n(s_n,x,y)$ is asymptotically bounded by C_n at (x,y) where f has no sharp cusps; while for (x,y) near a sharp cusp curve, $TZ_n(s_n,x,y)$ is dominated by $Tf(s_n,x,y)$, and thus significantly larger than C_n. Hence, $TZ_n(s_n,x,y)$ exceeds

C_n only nearby sharp cusp curves, and C_n is served as a threshold in estimation of the sharp cusp curves.

Suppose f has sharp cusps along curves $\theta_1, \cdots, \theta_q$, and is continuously differentiable elsewhere. Denote by Θ the class of smooth candidate curves which contains the true curves $\theta_1, \cdots, \theta_q$. We find all curves in Θ along which the minima of $TZ_n(s_n, x, y)$ exceed threshold C_n, that is,

$$\left\{\hat{\theta}\right\} = \left\{\vartheta \in \Theta : \min_{(x,y)\in\vartheta} TZ_n(s_n, x, y) \geq C_n\right\}, \tag{6}$$

and estimate $\{\theta_i\}_{1\leq i\leq q}$ by $\left\{\hat{\theta}\right\}$.

Denote by $d(x, y; A)$ the distance between a point (x, y) and a closed set $A \subset [0, 1]^2$, that is, the distance between (x, y) and its nearest neighbor in A, and define the Hausdorff distance between two closed sets A, $B \subset [0, 1]^2$ by

$$L(A, B) = max\left\{\max_{(x,y)\in A} d(x, y; B),\ \max_{(x,y)\in B} d(x, y; A)\right\}.$$

Treating each of $\{\theta_i\}_{1\leq i\leq q}$ and $\left\{\hat{\theta}\right\}$ as a group of closed subsets of $[0, 1]^2$, we measure the distance between $\{\theta_i\}_{1\leq i\leq q}$ and $\left\{\hat{\theta}\right\}$ by the Hausdorff distance $L\left(\bigcup\left\{\hat{\theta}\right\}, \bigcup_{i=1}^q \theta_i\right)$ between unions $\bigcup\left\{\hat{\theta}\right\}$ and $\bigcup_{i=1}^q \theta_i$. The following theorem establishes a convergence rate for the distance.

Theorem 1 *Suppose the following assumptions are satisfied.*

(1). $\theta_1, \cdots, \theta_q$ are continuous curves in $(0, 1)^2$ with finite length, and satisfy the Lipschitz condition that the arc length of θ_i between any two points on θ_i is of order of their geometric distance.

(2). f has α sharp cusps along curves $\theta_1, \cdots, \theta_q$, and is continuously differentiable on $[0,1]^2 \setminus \cup_{i=1}^{q} \theta_i$.

(3). ϕ and ψ are continuously differentiable.

Then

$$L\left(\bigcup \{\hat{\theta}\},\ \bigcup_{i=1}^{q} \theta_i\right) = O_p(s_n).$$

Assumption (3) in Theorem 1 is required for technical reasons such as Taylor expansion and integration by parts and satisfied by commonly used wavelets. For example, Daubechies wavelets with two or more vanishing moments are continuously differentiable (Daubechies 1992, chap. 7).

As suggested in Donoho and Johnstone 1994), we estimate σ in C_n by the median of $|n\, T^{\gamma,\beta} Z_n(s,x,y)|$ at fine scale s, divided by 0.6745. The consistency of the estimate is justified as follows. $T^{\gamma,\beta} W_n(s,x,y)$ follows a normal distribution with mean zero and standard deviation close to σ/n (Lemma A.1 in the Appendix), (3) implies that $T^{\gamma,\beta} Z_n(s,x,y)$ is a normal random variable with mean of order $T f(s,x,y)$ and standard deviation near σ/n. Since f is smooth on $[0,1]^2$ except for curves $\theta_1, \cdots, \theta_q$, then for (x,y) away from θ_i, $T f(s,x,y)$ is of order s^2. Therefore, for all (x,y) except for a small portion of points near the change curves, $n\, T^{\gamma,\beta} Z_n(s,x,y)$ follows a normal distribution with negligible mean of order $n\, s^2$ and standard deviation approaching σ. Since median is hardly influenced by a small fraction of extreme values, σ can be consistently estimated by the median of $|n\, T^{\gamma,\beta} Z_n(s,x,y)|$, divided by 0.6745.

3.2 Maximin Method for One Change Curve

Suppose f has only one sharp cusp curve θ. Discussion in previous section implies that $TZ_n(s_n, x, y)$ exceeds threshold C_n only at (x, y) near θ, and hence $TZ_n(s_n, x, y)$ at (x, y) near θ are significantly larger than the others, so we can construct an alternative estimate of θ by the maximin method developed in Müller and Song (1994a) as follows: find the curve that maximizes the minimum of $TZ_n(s_n, x, y)$ along curves in Θ, that is,

$$\hat{\theta}^* = arg \max_{\vartheta \in \Theta} \min_{(x,y) \in \vartheta} TZ_n(s_n, x, y). \tag{7}$$

In the following theorem we show that under the Hausdorff distance $\hat{\theta}^*$ also has a convergence rate s_n.

Theorem 2 *Suppose $q = 1$ and the three assumptions in Theorem 1 are satisfied. Then*

$$L(\hat{\theta}^*, \theta) = O_p(s_n).$$

For the problem of estimating a jump curve, Korostelev, Tsyabkov, and their co-authors established minimax risks and proved that estimates constructed by piecewise polynomial approximation and sieve maximum likelihood methods can be tuned to achieve the optimal rate (Korostelev and Tsyabkov 1992, 1993a; Tsyabkov 1994). For example, for estimating a jump curve satisfying a cone condition (which is equivalent to a piecewise Lipschitz boundary assumption) based on observations from nonparametric regression with random errors whose moment generating functions are bounded by that of a normal random variable, the minimax risk (defined by Hausdorff distance) converges to zero at a rate $n^{-1}\,(log\,n)^{1/2}$. For nonparametric regression with general random errors, Müller and Song (1994a) proposed a maximin estimate based on directional limits of multivariate kernel estimates and proved that the Hausdorff distance between their estimate and the true jump curve almost surely converges to zero at a rate $n^{-1+\rho}\,(log\,n)^{1/2}$, with $\rho > 0$ close to the ratio of 4 and the number of finite moments of random errors. These methods are proposed for different models and results were obtained under different conditions and criteria. However, in the case of nonparametric regression with normal random errors, they all have convergence rates pointing at n^{-1} multiplying by a $log\,n$ term. For the jump case, both threshold estimate $\hat{\theta}$ and maximin estimate $\hat{\theta}^*$ have convergence rate $n^{-1}\,(log\,n)^{\eta}$, so it is reasonable to believe that $\hat{\theta}$ and $\hat{\theta}^*$ catch the best possible convergence rate except for a $log\,n$ term.

4. EXAMPLES

4.1 Simulated Examples

The data $\{z_{ij}\}$ form a square matrix which is called a matrix data set. In image processing a matrix data set represents a digital image. For the discrete data we compute only the wavelet transformation $TZ_n(2^{-j}, k_1\,2^{-j}, k_2\,2^{-j})$, $k_1, k_2 = 1, \cdots, 2^j - 1$, $j = 1, \cdots, J$. They are called discrete wavelet transformation and can be computed by Mallat's pyramidal algorithm (Mallat 1989) which essentially applies many one dimensional discrete wavelet transformation (Daubechies 1988, 1992; Mallat 1989;

Nason and Silverman 1994). The computational complexity of the algorithm is of order of data size.

To test the methods, three simulated examples were carried out and illustrated in Figures 1, 2 and 3. f was taken to be a hyperboloid with jumps along a cosine type curve in Example 1, a paraboloid with jumps along a circle in Example 2, and a cosine surface with sharp cusps along the two diagonal lines in Example 3, and $n = 128$ in all three examples.

By Mallat's pyramid algorithm, we used least asymmetric Daubechies wavelet with three vanishing moments and reflecting boundary (Daubechies 1992, chap. 8) to compute the discrete wavelet transformations of data at scale levels 2^{-j}, $j = 1, \cdots, 7$, and at positions $(x, y) = (k_1 2^{-j}, k_2 2^{-j})$, $k_1, k_2 = 1, \cdots, 2^j - 1$, $j = 1, \cdots, 7$, which were also called empirical wavelet coefficients. At the highest level 7, only a few percent of the empirical wavelet coefficients were somewhat influenced by the presence of change curves, so we estimated σ by the median of the absolute values of the empirical wavelet coefficients at level 7, divided by 0.6745, and then computed C_n by (5).

By checking TZ_n at these 7 levels, we found location points $(i/n, j/n)$ at some levels whose corresponding $TZ_n(s, x, y)$ exceeded threshold C_n and were significantly larger than the others. Thus jumps and sharp cusps must be located in these squares around the location points. Since at high resolution levels the squares were very small with areas proportional to $s^2 = 2^{-2j}$, the jumps and sharp cusps were located with sufficient accuracy. In Figures 1, 2 and 3, (a) and (b) were the true surfaces and noisy surfaces, respectively. The (c) figures illustrated the grid location points at which TZ_n at levels $j = 6$ or 7 were significantly large and exceeded C_n. The location points were around a cosine type curve in Example 1, a circle in Example 2 and two diagonal lines in Example 3. We smoothed the points to obtain smoothed curves as follows: first divided the points into ten segments and then fit a line to each segment by minimizing the sum of squared distances from the points to the line. These smoothed curves along with the true curves were plotted in (d) figures. Visual comparison showed that they estimated the true curves very well.

For a fixed sample size in the simulated examples, as the signal to noise ratios decreased, the estimate became less and less precise. Fixed $n = 128$ in Example 2, as σ increased from 0.5 to 1.0 in steps of 0.1, the jump curve was estimated by wavelet coefficients at levels from 5 to 7. The case of $\sigma = 0.7$ was illustrated in Figure 2. The jump curve was hardly visible from the plot of the noisy data in Figure 2 (b). For $\sigma > 1$, the jumps were totally buried in the noise and the estimate was often hardly recognized to be a circle.

4.2 Edge Detection in Image Compression

Image compression is to code image with a minimum number of computer bits for transmission or storage. For an image like a TV picture, compression is done by taking advantage of the structural properties of images, and the way they are observed by a human observer. The details are as follows.

The image information can be divided into three categories. The first corresponds to edge contours. Edge contours are vital visual clues and particularly important for image understanding, which is well illustrated by our ability to recognize objects from a drawing that only outlines edges. Edges can not be modified substantially without producing perceivable changes in an image. The second category regroups the different textures which are defined for coding purposes as the image variations that do not have an information content by themselves but as a group. Important distortions can be introduced without affecting their visualization. The third category is the image "noise" which is defined as the information that can be removed without affecting the image visualization. This suggests removing noise and coding the texture information to be less precise than the edge information, and one can obtain high compression while introducing distortions that do not affect the image visualization (Canny 1986; Cohen and Froment 1992; Froment and Mallat 1992). So it is very crucial for image coding to identify the "important" feature: Edge contours. As a digitized image is represented by a bivariate function, edge contours are sharp variation curves along which the function has jumps and sharp cusps.

Now we apply the method to Lena image. Figure 4 (a) is Lena image with 256×256 pixels. We computed the wavelet transformation of the image and estimated the edge contours whose image was illustrated in Figure 4 (b). Coding the edges needed only a small fraction of the space required to store the original image.

5. CONCLUSIONS AND DISCUSSION

A wavelet method is proposed to estimate jump and sharp cusp curves. Asymptotic theory for the method is established and simulations are carried out to check the performance of the method for finite sample size. The method is applied to a real image. It is shown that the wavelet estimate is nearly optimal and can be computed by fast algorithms.

For estimating jump and sharp cusp curves, only the data around the curves matter. By appropriate interpolation or extrapolation, the method can be applied to a data set of any size, which may be neither square nor uniformly sampled.

This paper also leaves some issues for further investigations. For example, because simulations and arguments similar to the proofs in the appendix indicate that threshold C_n bounds all discrete wavelet transformation of normal noise, the procedure used in section 4 requires to check discrete wavelet coefficients at all levels with threshold C_n. We may give more theoretical justification for the procedure. Also we may consider non-normal random errors.

APPENDIX: PROOFS

Proofs of Lemmas

Denote by K generic positive constants which may depend on (ϕ, ψ, σ) and whose value may change from line to line. $n\, T^{\gamma,\beta} W_n(s, x, y)/\sigma$ follows a normal distribution with mean zero. Its variance and covariance are studied in three subsequent lemmas.

Lemma A.1 *For* $(\beta, x, y) \in [0, \pi/4] \times [0,1]^2$,

$$\left| E \left\{ n\, T^{\gamma,\beta} W_n(s, x, y)/\sigma \right\}^2 - 1 \right| \le K\, (n\, s)^{-1}.$$

Proof. Let $D_{ij} = ((i-1)/n, i/n] \times ((j-1)/n, j/n]$. As Ψ has unit L_2 norm, we have

$$
\begin{aligned}
& E\left\{n\, T^{\gamma,\beta} W_n(s,x,y)/\sigma\right\}^2 - 1 \\
&= n^{-2} \sum_{1\le i,j\le n} |\Psi_s^{\gamma,\beta}(x-x_i, y-y_j)|^2 - \int |\Psi_s^{\gamma,\beta}(x-u, y-v)|^2 du dv \\
&= s^{-2} \sum_{1\le i,j\le n} \int_{D_{ij}} \left\{ \left|\Psi^{\gamma,\beta}(x/s - x_i/s, y/s - y_j/s)\right|^2 \right. \\
&\qquad \left. - \left|\Psi^{\gamma,\beta}(x/s - u/s, y/s - v/s)\right|^2 \right\} du dv.
\end{aligned}
$$

Because Ψ has a compact support, of the n^2 integrals in the above sum, only $K\,(n\,s)^2$ integrals have nonzero integrands. On the other hand, since ϕ and ψ are continuously differentiable and have compact supports, Ψ satisfies Lipschitz condition and thus the absolute values of the nonzero integrands on D_{ij} are bounded by $K\,(n\,s)^{-1}$. Hence,

$$
\left| E\left\{n\, T^{\gamma,\beta} W_n(s,x,y)/\sigma\right\}^2 - 1 \right| \le K\, s^{-2}\, (n\,s)^2\, (n\,s)^{-1}\, n^{-2} \le K\,(n\,s)^{-1}. \qquad \square
$$

Lemma A.2

$$
\begin{aligned}
& Var\left\{n\, T^{\gamma,\beta} W_n(s,x,y)/\sigma - n\, T^{\gamma,\beta'} W_n(s,x',y')/\sigma\right\} \\
&\qquad \le K\left(|\beta-\beta'|^2 + |x-x'|^2/s^2 + |y-y'|^2/s^2\right).
\end{aligned}
$$

The proof is similar to that of Lemma A.1 and thus omitted.

Lemma A.3

$$
(1). \qquad Cov\left\{n\, T^{\gamma,\beta} W_n(s,x,y)/\sigma, n\, T^{\gamma,\beta} W_n(s,x',y')/\sigma\right\} = 0,
$$

if $|(x-x')\, cos\beta - (y-y')\, sin\beta| > K\, s$ *or* $|(x-x')\, sin\beta + (y-y')\, cos\beta| > K\, s$.

$$
(2). \qquad Cov\left\{n\, T^{\gamma,\beta} W_n(s,x,y)/\sigma, n\, T^{\gamma,\beta} W_n(s,x',y')/\sigma\right\} =
$$

$$1 - C_1^\gamma \left\{\frac{(x-x')\,cos\beta - (y-y')\,sin\beta}{s}\right\}^2 - C_2^\gamma \left\{\frac{(x-x')\,sin\beta + (y-y')\,cos\beta}{s}\right\}^2$$
$$+\,o\left(\left\{\frac{x-x'}{s}\right\}^2 + \left\{\frac{y-y'}{s}\right\}^2 + \frac{1}{n\,s}\right),$$

where

$$C_1^\gamma = \frac{1}{2}\int \left(\frac{\partial \Psi^{\gamma,0}(x,y)}{\partial x}\right)^2\, dxdy, \qquad C_2^\gamma = \frac{1}{2}\int \left(\frac{\partial \Psi^{\gamma,0}(x,y)}{\partial y}\right)^2\, dxdy.$$

Proof. Denote by $(u(\beta), v(\beta))$ the rotation of a point (u, v) counterclockwise by angle β, that is,

$$u(\beta) = u\,cos\beta - v\,sin\beta, \qquad v(\beta) = u\,sin\beta + v\,cos\beta.$$

In the following $(x(\beta), y(\beta))$, $(x'(\beta), y'(\beta))$, and $(x_i(\beta), y_j(\beta))$ stand for the rotations of (x, y), (x', y'), and (x_i, y_j), respectively. Then

$$\begin{aligned}
&Cov\left\{n\,T^{\gamma,\beta} W_n(s,x,y)/\sigma,\, n\,T^{\gamma,\beta} W_n(s,x',y')/\sigma\right\}\\
&\quad = n^{-2} \sum_{1\le i,j\le n} \Psi_s^{\gamma,\beta}(x-x_i, y-y_j)\; \Psi_s^{\gamma,\beta}(x'-x_i, y'-y_j)\\
&\quad = (n\,s)^{-2} \sum_{1\le i,j\le n} \Psi^{\gamma,0}\left(\frac{x(\beta)-x_i(\beta)}{s}, \frac{y(\beta)-y_j(\beta)}{s}\right)\\
&\qquad\qquad \Psi^{\gamma,0}\left(\frac{x'(\beta)-x_i(\beta)}{s}, \frac{y'(\beta)-y_j(\beta)}{s}\right).
\end{aligned} \tag{A.1}$$

(A.1) implies that the covariance is equal to zero, if $\{x(\beta) - x'(\beta)\}/s$ or $\{y(\beta) - y'(\beta)\}/s$ exceed the diameter of the compact support of $\Psi^{\gamma,0}$. This proves (1).

Because of similarity, we prove (2) only for $\gamma = h$. The right hand side of (A.1) can be approximated, with error of order $(n\,s)^{-1}$, by

$$\int_{-\infty}^{\infty} \phi(a)\phi\left(a + \frac{x'(\beta)-x(\beta)}{s}\right)\, da \int_{-\infty}^{\infty} \psi(b)\psi\left(b + \frac{y'(\beta)-y(\beta)}{s}\right)\, db. \tag{A.2}$$

From (A.2) we can easily prove (2) by the fact that wavelets are differentiable and compactly supported. $\square$

Now we derive a sharp bound on $T^{\gamma,\beta} W_n(s_n, x, y)$. The idea is to discretize index parameter space $[0, \beta] \times [0, 1]^2$ and approximate the process by its values at these discrete points. First we prove the following two lemmas which derive bounds on its maximum over the discrete points and on the approximation error.

Lemma A.4 *For some $\zeta > 1/4$, take integers $m_n \sim n\,(log\,n)^{-\zeta}$ and $q_n \sim log\,n$, and let $\beta_\ell = \ell\,\pi/(4\,q_n)$, $\ell \in \{1, \cdots, q_n\}$, $x_{a_1} = a_1/m_n$, $y_{a_2} = a_2/m_n$, $\mathbf{a} = (a_1, a_2) \in \{1, \cdots, m_n\}^2$. Set $u_n = 2\,\{log\,n - log\,log\,log\,n\}^{1/2}$. Then*

$$P\left\{\max\left(\left|n\,T^{\gamma,\beta_\ell} W_n(s_n, x_{a_1}, y_{a_2})/\sigma\right| : \ell \in \{1, \cdots, q_n\}, \mathbf{a} \in \{1, \cdots, m_n\}^2\right) \leq u_n\right\}$$
$$\geq 1 - K\,(log\,n)^{1/2 - 2\,\zeta}\,(log\,log\,n)^2.$$

Proof. Let

$$V_{\mathbf{a}}(\beta_\ell) = n\,T^{\gamma,\beta_\ell} W_n(s_n, x_{a_1}, y_{a_2})/\sigma.$$

If we have that for each fixed β,

$$P\left\{\max\left(|V_{\mathbf{a}}(\beta)| : \mathbf{a} \in \{1, \cdots, m_n\}^2\right) \leq u_n\right\}$$
$$\geq 1 - K\,(log\,n)^{-1/2 - 2\,\zeta}\,(log\,log\,n)^2, \tag{A.3}$$

then

$$P\left(\max_{1 \leq \ell \leq q_n} \max_{\mathbf{a} \in \{1, \cdots, m_n\}^2} |V_{\mathbf{a}}(\beta_\ell)| \leq u_n\right) \geq 1 - K\,q_n\,(log\,n)^{-1/2 - 2\,\zeta}\,(log\,log\,n)^2$$
$$\geq 1 - K\,(log\,n)^{1/2 - 2\,\zeta}\,(log\,log\,n)^2.$$

So the proof is reduced to show (A.3). For simplicity, we assume $\beta = 0$ and write $V_{\mathbf{a}}$ for $V_{\mathbf{a}}(\beta)$.

$V_{\mathbf{a}}$, $\mathbf{a} \in \{1, \cdots, m_n\}^2$, is a normal sequence with mean zero, variance and covariance derived in Lemmas A.1 and A.3. Denote by $\hat{V}_{\mathbf{a}}$, $\mathbf{a} \in \{1, \cdots, m_n\}^2$, an independent normal sequence with mean zero and the same variance as that of $V_{\mathbf{a}}$. Applying inequality (4.2.1) in theorem 4.2.1 of Leadbetter, Lindgren and Rootzén (1983, p. 81), we obtain that for any u,

$$\begin{aligned} & P(\max_{\mathbf{a}\in\{1,\cdots,m_n\}^2} \hat{V}_{\mathbf{a}} \le u) - P(\max_{\mathbf{a}\in\{1,\cdots,m_n\}^2} V_{\mathbf{a}} \le u) \\ & \quad \le (2\,\pi)^{-1} \sum_{\mathbf{a}\neq\mathbf{a}'} \max\{0, -Cov(V_{\mathbf{a}}, V_{\mathbf{a}'})\}\ \{1 - |Cov(V_{\mathbf{a}}, V_{\mathbf{a}'})|\}^{-1/2} \\ & \qquad\quad exp\Big\{-u^2/\{1 + |Cov(V_{\mathbf{a}}, V_{\mathbf{a}'})|\}\Big\}. \end{aligned} \tag{A.4}$$

By Lemma A.3 we have that $Cov(V_{\mathbf{a}}, V_{\mathbf{a}'}) = 0$ for $\mathbf{a} = (a_1, a_2)$ and $\mathbf{a}' = (a_1', a_2')$ satisfying $|a_1 - a_1'| > K\, m_n\, s_n$ or $|a_2 - a_2'| > K\, m_n\, s_n$, and

$$Cov(V_{\mathbf{a}}, V_{\mathbf{a}'}) \sim 1 - (m_n\, s_n)^{-2}\ \Big\{C_1^\gamma\,(a_1 - a_1')^2 + C_2^\gamma\,(a_2 - a_2')^2\Big\}.$$

Since $Cov(V_{\mathbf{a}}, V_{\mathbf{a}'})$ asymptotically decreases from 1 to zero as $\mathbf{a}$ and $\mathbf{a}'$ move apart, so for some $0 < \delta < 1/(2\,\alpha + 1)$, there exists a constant $K_1 > 0$ such that for $\mathbf{a} = (a_1, a_2)$ and $\mathbf{a}' = (a_1', a_2')$ satisfying $|a_1 - a_1'| \le K_1\, m_n\, s_n$ and $|a_2 - a_2'| \le K_1\, m_n\, s_n$, $Cov(V_{\mathbf{a}}, V_{\mathbf{a}'}) > 0$, and for other $\mathbf{a}$ and $\mathbf{a}'$, $|Cov(V_{\mathbf{a}}, V_{\mathbf{a}'})| \le \delta$. Hence, in the sum of the right hand side of (A.4), there are at most $m_n^2\,(2\,K_1\, m_n\, s_n)^2$ nonzero terms and each nonzero term is bounded by

$$\delta\ (1 - \delta)^{-1/2}\ exp\Big\{-u^2/(1 + \delta)\Big\}.$$

Using these facts and (A.4) we get

$$
\begin{aligned}
& P(\max_{\mathbf{a}} \hat{V}_{\mathbf{a}} \leq u_n) - P(\max_{\mathbf{a}} V_{\mathbf{a}} \leq u_n) \\
& \leq K\, m_n^2\, (m_n\, s_n)^2\, \delta\, (1-\delta)^{-1/2}\, exp\Big\{-u_n^2/(1+\delta)\Big\} \\
& \leq K\, n^{4\,\delta/(1+\delta)-2/(1+\alpha)}\, (log\, n)^{2\eta/(1+\alpha)-4\,\zeta}\, (log\, log\, n)^{4/(1+\delta)}. \qquad (A.5)
\end{aligned}
$$

On the other hand, $\hat{V}_{\mathbf{a}}$, $\mathbf{a} \in \{1, \cdots, m_n\}^2$, are independent normal random variables with mean zero and variance tending to one, so simple calculations lead to

$$
\begin{aligned}
& P(\max_{\mathbf{a}} \hat{V}_{\mathbf{a}} \leq u_n) = \prod_{\mathbf{a}} P(\hat{V}_{\mathbf{a}} \leq u_n) \\
& \sim \prod_{\mathbf{a}} \Big(1 - \int_{u_n}^{\infty} (2\,\pi)^{-1/2}\, exp\{-x^2/2\}\, dx\Big) \\
& \sim \Big\{1 - u_n^{-1}\, exp(-u_n^2/2)\Big\}^{m_n^2} \\
& \sim 1 - K\, (log\, n)^{-1/2-2\,\zeta}\, (log\, log\, n)^2. \qquad (A.6)
\end{aligned}
$$

Combing (A.5) and (A.6) we have

$$
P(\max_{\mathbf{a}} V_{\mathbf{a}} \leq u_n) \geq 1 - K\, (log\, n)^{-1/2-2\,\zeta}\, (log\, log\, n)^2.
$$

Finally (A.3) is a consequence of above inequality and the fact that for any u,

$$
P(\max_{\mathbf{a}} |V_{\mathbf{a}}| \geq u) \geq 2\, P(\max_{\mathbf{a}} V_{\mathbf{a}} \geq u). \qquad \square
$$

Given $(\beta, x, y) \in [0, \pi/4] \times [0,1]^2$, there exist $\ell \in \{1, \cdots, q_n\}$ and $(a_1, a_2) \in \{1, \cdots, m_n\}^2$ such that

$$
\beta_{\ell-1} < \beta \leq \beta_\ell, \qquad x_{a_1-1} < x \leq x_{a_1}, \qquad y_{a_2-1} < y \leq y_{a_2},
$$

and set $(\beta^*, x^*, y^*) = (\beta_\ell, x_{a_1}, y_{a_2})$. Then

$$
|\beta - \beta^*| \leq 1/q_n, \qquad |x - x^*| \leq 1/m_n, \qquad |y - y^*| \leq 1/m_n. \qquad (A.7)
$$

Lemma A.5

$$P\left(\max\left\{\left|n\,T^{\gamma,\beta}W_n(s_n,x,y)/\sigma - n\,T^{\gamma,\beta^*}W_n(s_n,x^*,y^*)/\sigma\right| :\right.\right.$$
$$\left.\left.(\beta,x,y)\in[0,\pi/4]\times[0,1]^2\right\}\le K\,(log\,n)^{-1/2}\right)\ge 1-K_1\,n^{-1}\,(log\,n)^2.$$

Proof. Set $\boldsymbol{\xi}=(\xi_1,\xi_2,\xi_3)=(\beta,x/s_n,y/s_n)$, $\boldsymbol{\xi}^*=(\xi_1^*,\xi_2^*,\xi_3^*)=(\beta^*,x^*/s_n,y^*/s_n)$, $\Xi_n=[0,\pi/4]\times[0,1/s_n]^2$, and let

$$\begin{aligned}X_n(\boldsymbol{\xi}) &= n\,T^{\gamma,\beta}W_n(s_n,x,y)/\sigma\\ &= (\sigma\,n\,s_n)^{-1}\sum_{1\le i_1,i_2\le n}\varepsilon_{i_1i_2}\,\Psi^{\gamma,0}\left(\left\{\xi_2-i_1\,(n\,s_n)^{-1}\right\}\,cos\,\xi_1-\left\{\xi_3-i_2\,(ns_n)^{-1}\right\}\,sin\,\xi_1,\right.\\ &\qquad\left.\left\{\xi_2-i_1\,(n\,s_n)^{-1}\right\}\,sin\,\xi_1+\left\{\xi_3-i_2\,(ns_n)^{-1}\right\}\,cos\,\xi_1\right).\end{aligned}$$

Then $X_n(\boldsymbol{\xi})$, $\boldsymbol{\xi}\in\Xi_n$, is a Gaussian random field with mean zero. Define

$$\mu_n(\boldsymbol{\xi},\delta)=E\left(\sup\left\{|X_n(\boldsymbol{\xi}')-X_n(\boldsymbol{\xi}'')|:|\boldsymbol{\xi}'-\boldsymbol{\xi}|\le\delta,|\boldsymbol{\xi}''-\boldsymbol{\xi}|\le\delta\right\}\right),$$

$$\text{and}\qquad \mu_n(\delta)=\sup_{\boldsymbol{\xi}\in\Xi_n}\mu_n(\boldsymbol{\xi},\delta).$$

First we use the theory of empirical processes to show that for $\delta_n=K\,(log\,n)^{-1}$,

$$\mu_n(\delta_n)\le K\,(log\,n)^{-1/2}. \tag{A.8}$$

Denote by $N_n(\delta)$ the minimal number of balls of radius δ needed to cover Ξ_n. Then $N_n(\delta)\le 2\,s_n^{-2}\,\delta^{-3}$. Let $H(x)=exp(x^2)/5$ and define the Orlicz norm of a random variable U by

$$\|U\|_H=\inf\{a>0:E\{H(U/a)\}\le 1\}.$$

Simple computations show that for a zero-mean normal random variable U, $\|U\|_H^2\le 4\,Var(U)$. So Lemma A.2 implies that

$$\|X_n(\boldsymbol{\xi}')-X_n(\boldsymbol{\xi}'')\|_H\le K\,|\boldsymbol{\xi}'-\boldsymbol{\xi}''|.$$

Applying the maximal inequality (Pisier 1983, eq. 1.2 in theorem 1.1 on p. 126; Pollard 1990, eq. 3.5 or eq. 3.6 on pp.12-13) to $X_n(\boldsymbol{\xi})$ we obtain

$$\begin{aligned}
&\mu_n(t,\delta_n) \le K \int_0^{2\,\delta_n} \{log\, N_n(\delta)\}^{1/2}\, d\delta \\
&\le K \int_0^{2\,\delta_n} \{|log\, s_n| + log(1/\delta)\}^{1/2}\, d\delta \\
&\le K\, \delta_n \left\{|log\, s_n|^{1/2} + |log\, \delta_n|^{1/2}\right\} \\
&\le K\,(log\, n)^{-1/2},
\end{aligned}$$

and thus $\mu_n(\delta_n) \le K(log\, n)^{-1/2}$, which proves (A.8).

By Lemma A.2 and (A.7) we get

$$\begin{aligned}
&Var\{X_n(\boldsymbol{\xi}) - X_n(\boldsymbol{\xi}^*)\} \le K\,(|\beta-\beta^*|^2 + |x-x^*|^2/s_n^2 + |y-y^*|^2/s_n^2) \\
&\le K\,\left(q_n^{-2} + 2\,(m_n\, s_n)^{-2}\right) \\
&\le K\,(log\, n)^{-2}. \qquad \text{(A.9)}
\end{aligned}$$

Now applying inequality (5.10) on p.119 of Adler (1990) to $X_n(\boldsymbol{\xi}) - X_n(\boldsymbol{\xi}^*)$ and using (A.9) and (A.8), we have

$$\begin{aligned}
&P\left(\max_{\boldsymbol{\xi}\in\Xi_n}\{X_n(\,\boldsymbol{\xi}) - X_n(\boldsymbol{\xi}^*)\} > M\,(log\, n)^{-1/2}\right) \\
&\le 2\,N_n(\delta_n)\, exp\left(-\frac{1}{2}\left\{M\,(log\, n)^{-1/2} - \mu_n(\delta_n)\right\}^2 \Big/ \left\{K\,(log\, n)^{-2}\right\}\right) \\
&\le 2\,N_n(\delta_n)\, exp\left(-K\,M^2\, log\, n + 2\,K\,M\,(log\, n)^{3/2}\,\mu_n(\delta_n)\right) \\
&\le K_1\, s_n^{-2}\,\delta_n^{-3}\, exp\left\{-K_2\,M^2\, log\, n + 2\,K_3\,M\, log\, n\right\} \\
&\le K_1\, n^{-2/(\alpha+1)}\,(log\, n)^{3-2\,\eta/(\alpha+1)}\, exp\left\{-K_2\,M^2\, log\, n + 2\,K_3\,M\, log\, n\right\} \\
&\le K_1\, n^{2-K_2\,M^2+2\,K_3\,M}\,(log\, n)^2. \qquad \text{(A.10)}
\end{aligned}$$

Taking $M \geq 2K_3/K_2 + (3/K_2)^{1/2}$ we have $2 - K_2 M^2 + 2K_3 M \leq -1$. Hence (A.10) implies that the probability in the left hand side of (A.10) is bounded by $K_1 n^{-1} (log\, n)^2$. Finally the lemma is proved by using the fact that for any u,

$$P\left(\max_{\boldsymbol{\xi}\in\Xi_n} |X_n(\boldsymbol{\xi}) - X_n(\boldsymbol{\xi}^*)| > u\right) \leq 2P\left(\max_{\boldsymbol{\xi}\in\Xi_n}\{X_n(\boldsymbol{\xi}) - X_n(\boldsymbol{\xi}^*)\} > u\right). \qquad \square$$

Now we are in the position to prove the main lemma.

Lemma A.6 *As $n \to \infty$, with probability tending to one,*

$$\max_{0\leq x,y\leq 1} TW_n(s_n, x, y) \leq 2\,\sigma\, n^{-1} (log\, n)^{1/2}.$$

Proof. Since

$$\left|n\, T^{\gamma,\beta} W_n(s_n, x, y)/\sigma\right| \leq \left|n\, T^{\gamma,\beta^*} W_n(s_n, x^*, y^*)/\sigma\right| +$$
$$\left|n\, T^{\gamma,\beta} W_n(s_n, x, y)/\sigma - n\, T^{\gamma,\beta^*} W_n(s_n, x^*, y^*)/\sigma\right|,$$

By Lemmas A.4 and A.5 we have that with probability tending to one, for $(\beta, x, y) \in [0, \beta/4] \times [0, 1]^2$,

$$\left|n\, T^{\gamma,\beta} W_n(s_n, x, y)/\sigma\right| \leq 2\,(log\, n - log\, log\, log\, n)^{1/2} + K\,(log\, n)^{-1/2}$$
$$\leq 2\,(log\, n)^{1/2} - (log\, n)^{-1/2}\, log\, log\, log\, n + K\,(log\, n)^{-1/2}$$
$$\leq 2\,(log\, n)^{1/2} - (1/2)\,(log\, n)^{-1/2}\, log\, log\, log\, n. \tag{A.11}$$

Finally the lemma is a consequence of above inequality and

$$\max_{0\leq x,y\leq 1} TW_n(s, x, y) = max\left\{\max_{\beta,x,y} |T^{\gamma,\beta} W_n(s, x, y)| : \gamma = h, v, d\right\}. \qquad \square$$

Proofs of Theorems 1 and 2

Due to the similarity, we give only arguments for the proof of Theorem 1.

Proof of Theorem 1. Denote by $supp(\psi)$ and $supp(\phi)$ the respective compact supports of ψ and ϕ. Set $supp(\Psi) = [supp(\psi) \cup supp(\phi)]^2$, and $\bigcup_{i=1}^q \{\theta_i\} + s_n \, supp(\Psi) = \{(x,y) + s_n \, (u,v) : (x,y) \in \theta_i, i = 1, \ldots, q, \ (u,v) \in supp(\Psi)\}$ be the union of sausages formed by moving the set $s_n \, supp(\Psi)$ along curves θ_i, $i = 1, \cdots, q$.

Since $f(x)$ has α sharp cusps along θ_i and is differentiable elsewhere, from the discussion near the end of Section 2 we obtain that for all (s_n, x, y) with $(x,y) \notin \bigcup_{i=1}^q \{\theta_i\} + s_n \, supp(\Psi)$,

$$T f(s_n, x, y) \leq K s_n^2,$$

and

$$\min_{(x,y) \in \theta_i} T f(s_n, x, y) \geq K s_n^{\alpha+1}.$$

The above two inequalities together with (A.11) in the proof of Lemma A.6 imply that, with probability tending to one, for all (s_n, x, y) with $(x,y) \notin \bigcup_{i=1}^q \{\theta_i\} + s_n \, usupp(\Psi)$,

$$\begin{aligned} & T Z_n(s_n, x, y) \leq T f(s_n, x, y) + T W_n(s_n, x, y) \\ & \leq K \, s_n^2 + C_n - (1/2) \, \sigma \, n^{-1} \, (log \, n)^{-1/2} \, log \, log \, log \, n \\ & \leq C_n, \end{aligned}$$

and thus for any ϑ with $\vartheta \not\subset \bigcup_{i=1}^q \{\theta_i\} + s_n \, supp(\Psi)$,

$$\min_{(x,y) \in \vartheta} T Z_n(s_n, x, y) \leq C_n; \tag{A.12}$$

and

$$\min_{(x,y)\in\theta_i} TZ_n(s_n,x,y) \geq \min_{(x,y)\in\theta_i} Tf(s_n,x,y) - \max_{x,y} TW_n(s_n,x,y)$$
$$\geq K\, s_n^{\alpha+1} - 2\,\sigma\, n^{-1}\,(log\, n)^{1/2}$$
$$\geq K\, n^{-1}\,(log\, n)^{\eta}\,\{1-(log\, n)^{1/2-\eta}\}. \tag{A.13}$$

Since $\eta > 1/2$, as $n \to \infty$, the bound in the right hand side of (A.13) exceeds threshold C_n, which bounds (A.12). From the definition of $\left\{\hat{\theta}\right\}$ we have that, with probability tending to one, $\bigcup\left\{\hat{\theta}\right\} \subset \bigcup_{i=1}^{q}\{\theta_i\} + s_n\, supp(\Psi)$. Therefore, $L(\bigcup\left\{\hat{\theta}\right\}, \bigcup_{i=1}^{q}\theta_i) = O_p(s_n)$. $\square$

REFERENCES

Adler, R. (1981), *The Geometry of Random Fields*, New York: John Wiley.

Adler, R. (1990), *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes* (Ser. 12), Hayward, California: Institute of Mathematical Statistics.

Antoniadis, A., Gregoire, G. and McKeague, I. W. (1994), "Wavelet methods for curve estimation," *Journal of the American Statistical Association*, 89, 1340-1353.

Antoine, J. P., Murenzi, R., and Piette, B. (1992), "Image Analysis With 2D Continuous Wavelet Transform: Detection of Position, Orientation and Visual Contrast of Simple Objects," in *Wavelets and Applications*, ed. Y. Meyer, New York: Springer-Verlag, pp. 144-159.

Bickel, P. and Rosenblatt, M. (1973), "Two-Dimensional Random Fields," in *Multivariate Analysis* III, ed. P. R. Krishnaiah, New York: Academic Press, pp. 3-15.

Blake, A. and Zisserman, A. (1987), *Visual Reconstruction*, Cambridge: MIT Press.

Canny, J. (1986), "A Computational Approach to Edge Detection," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 8, 679-698.

Carlstein, E. and Krishnamoorthy, C. (1992), "Boundary Estimation," *Journal of the American Statistical Association*, 87, 430-438.

Carlstein, E. and Krishnamoorthy, C. (1994), "Practical Considerations in Boundary Estimation: Model-Robustness, Efficient Computation, and Bootstrapping," in *Change-Point Problems* (Ser. 23), eds. E. Carlstein, H. G. Müller, and D. Siegmund, Hayward, California: Institute of Mathematical Statistics, pp. 177-193.

Carlstein, E., Müller, H. G., and Siegmund, D. (1994), *Change-Point Problems* (Ser. 23), Hayward, California: Institute of Mathematical Statistics.

Cohen, A. and Froment, J. (1992), "Image Compression and Multiscale Approximation," in *Wavelets and Applications*, ed. Y. Meyer, New York: Springer-Verlag, pp. 181-206.

Daubechies, I. (1988), "Orthonormal Bases of Compactly Supported Wavelets," *Communications on Pure and Applied Mathematics*, 41, 909-996.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, Philadelphia: Society for Industrial and Applied Mathematics.

DeVore, R. A., Jawerth, B., and Lucier, B. J. (1992), "Image Compression Through

Wavelet Transform Coding," *IEEE Transaction on Information Theory*, 38, 719-746.

Devroye, L. and Wise, G. L. (1980), "Detection of Abnormal Behavior via Nonparametric Estimation of the Support," *SIAM Journal on Applied Mathematics*, 38, 480-488.

Donoho, D. L. (1993), "Smooth Wavelet Decompositions With Blocky Coefficient Kernels," in *Recent Advances in Wavelet Analysis*, eds. L. L. Schumaker and G. Webb, New York: Academic Press, pp. 259-308.

Donoho, D. L. (1995), "De-Noising by Soft Thresholding," *IEEE Transaction on Information Theory*, 41, 613-627.

Donoho, D. L. and Johnstone, I. M. (1994), "Ideal Spatial Adaptation by Wavelet Shrinkage," *Biometrika*, 81, 425-455.

Donoho, D. L. Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1995), "Wavelet Shrinkage: Asymptopia ?" *Journal of the Royal Statistical Society*, Ser. B, 57, 301-369.

Froment, J. and Mallat. S. G. (1992), "Second Generation Compact Image Coding With Wavelets," in *Wavelet and Their Applications*, eds. M. B. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer, and L. Raphael, Boston: Jones and Bartlett Publishes, pp. 655-677.

Goel, P. and Vidakovic, B. (1995), "Wavelet Transformations as Diversity Enhancers," Technical report, Duke University.

Hartigan, J. A. (1987), "Estimation of a Convex Density Contour in Two Dimensions," *Journal of the American Statistical Association*, 82, 267-270.

Korostelev, A. P. (1991), "Minimax Reconstruction of Two-Dimensional Image," *Probability Theory and its Applications*, 36, 153-159.

Korostelev, A. P., Simar, L., and Tsybakov, A. B. (1995), "Efficient Estimation of Monotone Boundaries," *The Annals of Statistics*, 23, 476-489.

Korostelev, A. P. and Tsybakov, A. B. (1992), "Asymptotically Minimax Image Reconstruction Problems," in *Topics Nonparametric Estimation*, ed. R. Z. Khasminshii, Providence: American Mathematical Society, pp. 45-86.

Korostelev, A. P. and Tsybakov, A. B. (1993a), *Minimax Theory of Image Reconstruction*, Lecture Notes in Statistics, 82, New York: Springer-Verlag.

Korostelev, A. P. and Tsybakov, A. B. (1993b), "Estimation of Support of a Probability Density and Estimation of Support of Functionals," *Problems of Information Transmission*, 29, 3-18.

Leadbetter, Lindgren and Rootzén (1983), "Extremes and Related Properties of Random Sequences and Processes," New York: Springer-Verlag.

Mallat, S. G. and Hwang, W. L. (1992), "Singularity Detection and Processing With Wavelets," *IEEE Transaction on Information Theory*, 38, 617-643.

Mallat, S. G. and Zhong, S. (1992a), "Characterization of Signals From Multiscale Edge," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 14, 710-732.

Mallat, S. G. and Zhong, S. (1992b), "Wavelet Transformation Maxima and Multiscale Edges," in *Wavelets: A Tutorial in Theory and Applications*, ed. C. K. Chui, Boston: Academic Press, pp. 66-104.

Mallat, S. G. and Zhong, S. (1992c), "Wavelet Maxima Representation," in *Wavelets and Applications*, ed. Y. Meyer, New York: Springer-Verlag, pp. 207-284.

Mammen, E. and Tsybakov, A. B. (1995), "Asymptotical Minimax Recovery of Sets With Smooth Boundaries," *The Annals of Statistics*, 23, 502-524.

Marr, D. (1982), *Vision*, San Francisco: Freeman.

Müller, H. G. and Song, K. (1994a), "Maximin Estimation of Multidimensional Boundaries," *Journal of Multivariate Analysis*, 50, 265-281.

Müller, H. G. and Song, K. (1994b), "Cube Splitting in Multidimensional Edge Estimation," in *Change-Point Problems* (Ser. 23), eds. E. Carlstein, H. G. Müller, and D. Siegmund, Hayward, California: Institute of Mathematical Statistics, pp. 210-223.

Nason, G. P. and Silverman, B. W. (1994), "The Discrete Wavelet Transformation in S," *Journal of Computational and Graphical Statistics*, 3, 163-191.

Nason, G. P. (1996), "Wavelet Shrinkage Using Cross-Validation," *Journal of the Royal Statistical Society*, Series B, 58, 463-479.

Pollard, D. (1990), *Empirical Processes: Theory and Applications*, NSF-CBMS Regional Conference Series in Probability and Statistics, 2, Hayward, California: Institute of Mathematical Statistics.

Pratt, W. K. (1991), *Digital Image Processing* (2nd ed.), New York: Wiley.

Puri, M. L. and Ruymgaart, F. H. (1994), "Change Curves in the Presence of Dependent Noise," In *Change-Point Problems* (Ser. 23), eds. E. Carlstein, H. G. Müller, and D. Siegmund, Hayward, California: Institute of Mathematical Statistics, pp. 242-254.

Ruddemo, M. and Stryhn, H. (1994), "Boundary Estimation for Star-Shaped Objects," in *Change-Point Problems* (Ser. 23), E. Carlstein, H. G. Müller, and D. Siegmund, Hayward, California: Institute of Mathematical Statistics, pp. 276-283.

Shiau, J. J., Wahba, G., and Johnson, D. R. (1986), "Partial Spline Models for the Inclusion of Tropopause and Fronral Boundary Information in Otherwise Smooth Two- and Three-Dimensional Objective Analysis," *Journal of Atmospheric and Oceanic Technology*, 3, 714-725.

Titterington, D. M. (1991), "Choosing the Regularization Parameter in Image Reconstruction, " In *Spatial Statistical Imaging*, ed. A. Possolo, Hayward, California: Institute of Mathematical Statistics, pp. 392-402.

Tsybakov, A. B. (1994), "Multidimensional Change-Point Problems and Boundary Estimation," in *Change-Point Problems* (Ser. 23), eds. E. Carlstein, H. G. Müller, and D. Siegmund, Hayward, California: Institute of Mathematical Statistics, pp. 317-329.

Vidakovic, B. (1994), "Nonlinear Wavelet Shrinkage With Bayes Rules and Bayes Factors," Technical report, Duke University.

Wang, Y. (1995), "Jump and Sharp Cusp Detection by Wavelets," *Biometrika*, 82, 385-397.

Wang, Y. (1996), "Function Estimation via Wavelet Shrinkage for Long-Memory Data," *The Annals of Statistics*, 24, 466-484.

Wang, Y. (1997a), "Fractal Function Estimation via Wavelet Shrinkage," *Journal of Royal Statistical Society B*, 59, 603-613.

Wang, Y. (1997b), "Change-Point Analysis via Wavelets for Indirect Data," *Statistica Sinca*, to be published.

Legend for Figures

Figure 1. The data are simulated from the model $y_{ij} = f(i/n, j/n) + \varepsilon_{ij}$, $i, j = 1, \cdots, 128$, $f(x_1, x_2) = (2x_1-1)^2-(2x_2-1)^2+2\cdot 1\{1.4(x_1+x_2)-2.1 > cos(\pi x_1-\pi x_2)\}$, $\varepsilon_{ij} \sim N(0, \sigma^2)$, $\sigma = 0.5$, $n = 128$; (a) the true surface; (b) the noisy surface; (c) the location points (dots) at which wavelet coefficients at level $j = 7$ exceed the threshold, and the smoothed curve (dot line); (d) the smoothed curve (dot line) and the true curve (solid line).

Figure 2. The data are simulated from the model $y_{ij} = f(i/n, j/n) + \varepsilon_{ij}$, $i, j = 1, \cdots, 128$, $f(x_1, x_2) = 1-(2x_1-1)^2-(2x_2-1)^2-2\cdot 1\{(2x_1-1)^2+(2x_2-1)^2 < 0.36\}$, $\varepsilon_{ij} \sim N(0, \sigma^2)$, $\sigma = 0.7$, $n = 128$; (a) the true surface; (b) the noisy surface; (c) the location points (dots) at which wavelet coefficients at level $j = 6$ exceed the threshold, and the smoothed curve (dot line); (d) the smoothed curve (dot line) and the true curve (solid line).

Figure 3. The data are simulated from the model $y_{ij} = f(i/n, j/n) + \varepsilon_{ij}$, $i, j = 1, \cdots, 128$, $f(x_1, x_2) = 1-cos\{2\pi(x_1+x_2)\}-2\,|\,|2x_1-1|-|2x_2-1|\,|^{1/5}$, $\varepsilon_{ij} \sim N(0, \sigma^2)$, $\sigma = 0.3$, $n = 128$; (a) the true surface; (b) the noisy surface; (c) the location points (dots) at which wavelet coefficients at level $j = 7$ exceed the threshold, and the smoothed curve (dot line); (d) the smoothed curve (dot line) and the true curve (solid line).

Figure 4. (a) Lena image with 256×256 pixels; (b) edge contours.