Sankhyā: The Indian Journal of Statistics 1997, Volume 59, Series A, Pt. 2, pp.

NONPARAMETRIC ESTIMATION OF JUMP SURFACE

By PEIHUA QIU
The Ohio State University, Ohio

SUMMARY. In this paper, we discuss estimation of bivariate jump regression functions. An a.s. consistent estimator of the jump location curve is suggested. This estimator is based on difference of two one-sided kernel smoothers. A rotation transformation is also used. We consider an ideal case that the jump location curve has an explicit function form first and then generalize it to a more general case that the explicit function form does not exist. Comparing to some existing methods on this topic, mainly to the edge detection methods in image processing literature, our method uses less conditions on the design points and on the underlying regression function. So it is expected to find more applications.

1. **Introduction**

Geological departments need to estimate mine surface according to the mineral samples. Because of earth movement, the mine surface usually splits into several segments. It is very important, for mine exploring and for taking precautionary actions from accidents, to know the split locations of the mine surface. From a statistical view point, this is an estimation problem of 2-dimensional (2-D) regression function with discontinuities. We can find such problems in many other application fields such as meteorology, oceanography, image processing, etc.

First of all, let us make a simple review on statistical theory of discontinuous regression functions. Parametric regression models with some kind of change-points have been discussed for more than thirty years. (In literature, "change-point" has many different meanings. In this paper, we use "jump" or "discontinuous" to distinguish our problem from other kinds of change-point problems.) Readers are recommended to read Quandt (1958), Hinkley (1969, 1971), Shaban (1980) and Bhattacharya (1994). Recently nonparametric methods to fit jump regression models (JRMs, namely, models with discontinuous regression functions) have been developed in several related areas. Firstly, JRMs are regarded as a special kind of partial linear regressin models and are built with the so-called partial spline technique in smoothing spline area (e.g. Wahba, 1986; Shiau, 1987). Secondly, some special JRMs are regarded as random processes with change-points in their mean functions and are built under various conditions on the path, increment and other quantities of the processes (see e.g., Chen, 1988). Both of these two kinds of methods try to include the JRMs in other more general models. Doing so loses inevitably some special features of the jump regression functions (JRFs). McDonald and Owen (1986) might be the first to analyse the 1-dimensional (1-D) JRFs directly from their own features. They provided a so-called split linear smoothing algorithm for jump detection. Hall and Titterington (1992) proposed an alternative method to that of McDonald and Owen (1986). Yin (1988) suggested an algorithm to estimate the number, the locations and the magnitudes of jumps of 1-D JRFs. Almost at the same time, Müller (1992), Qiu (1991, 1994), Qiu *et al.* (1991) and Wu and Chu (1993) suggested similar estimators of the JRFs. These estimators were all based on difference between two one-sided kernel smoothers. One such smoother used the right-sided kernel while the other one used the left-sided kernel. Because

---

Paper received. October 1995; revised December 1996.
AMS 1991 *subject classifications.* Primary 62A25; secondary 62H05
Key Words and phrases. Jump regression function; difference kernel estimator; rotation transformation; uniformly a.s. consistent; image processing; edge detection.

of the weighted average nature of the kernel smoothers, the difference is close to zero when there is no jump at a given point. Otherwise, the difference is close to the jump magnitude. Qiu *et al.* (1991) called these estimators *Difference Kernel Estimators (DKEs)*. DKEs connected the JRMs with the kernel regression techniques. Doing so has at least two benefits. One is that it is possible for us to explore the properties of the DKEs by using the abundant kernel regression theory. Actually, Wu and Chu (1993) proved the asymptotic normality of a DKE; Müller (1992) obtained the global L^p consistency and Qiu *et al.* (1991) proved a.s. and L^2 consistencies of the DKEs in various cases. The second benefit is that we can make use of the existing kernel regression programs in SAS or other statistical softwares to do our computation. So these methods are not difficult to be used.

In this paper, we will discuss the estimation of the 2-D JRFs mentioned at the beginning of this section and we will suggest an a.s. consistent estimation method.

Obviously, there are many differences between 1-D and 2-D cases. In the former case, the jump locations are at most a series of points. But the jump location is a curve in the latter case. (We assume that there is one jump location curve in this paper.) In 1-D case, we can use the difference between averages of observations in a right-sided neighborhood and a left-sided neighborhood of a given design point to detect a jump. When we generalize this idea to a 2-D case, we can also make averages of observations on two different sides of a given point along a direction. But the difference of averages will change if we change the direction. In the case that this direction is not chosen properly, the difference can be very small even if the point in question is on the jump location curve. Furthermore, there is no coordinate system in hand in many applications. Or equivalently in some sense, the jump location curve can not be expressed by an explicit mathematical function under a predetermined coordinate system. This requires our method to be coordinate free or not to depend on the explicit function form of the jump location curve.

We suggest a so-called rotational difference kernel estimation method in this paper for the estimation of 2-D JRFs. For simplicity, we explain our idea below in a simple case that the jump location curve has an explicit mathematical function form. We leave the discussion about a more general case that the jump location curve does not have the explicit function form to Section 3.

Suppose that the regression model concerned is

$$Z_i = f(x_i, y_i) + \varepsilon_i, \quad i = 1, 2, \cdots, n \qquad \ldots (1.1)$$

where design points $(x_i, y_i) \in [0,1] \times [0,1], i = 1, 2, \cdots, n$, and $\{\varepsilon_i\}$ are i.i.d. random errors. The regression function $f(x, y)$ has the form

$$f(x, y) = g(x, y) + C(x) I_{y > \phi(x)}, \quad (x, y) \in [0,1] \times [0,1] \qquad \ldots (1.2)$$

where $g(x, y)$ is the continuous part, $\phi(x)$ denotes the jump location curve, $C(x)$ is the jump magnitude function and all of these functions are assumed to be continuous.

Let $K_1^*(x, y)$ and $K_2^*(x, y)$ be two kernel functions which satisfy

(i) $K_1^*(x, y) = 0$ when $(x, y) \notin [-1/2, 1/2] \times [-1, 0]$; $K_2^*(x, y) = 0$ when $(x, y) \notin [-1/2, 1/2] \times [0, 1]$.

(ii) $K_i^*(x, y) \geq 0, \quad i = 1, 2.$

(iii) $\int_{-1}^{1} \int_{-1}^{1} K_i^*(x, y)\, dx dy = 1, \quad i = 1, 2.$

Remark 1.1. Conditions (ii) and (iii) require that $K_i^*(x, y), i = 1, 2$, are the density kernels. This is an ordinary requirement of the kernel regression techniques (see e.g., Härdle, 1990). Condition (i) requires the kernels to be one-sided about the origin. $K_1^*(x, y)$ is lower-sided while $K_2^*(x, y)$ is upper-sided.

Remark 1.2. The non-zero-value domains of definition of $K_i^*(x, y), i = 1, 2$, in condition (i) can be generalized to $[-L/2, L/2] \times [-L, 0]$ and $[-L/2, L/2] \times [0, L]$ respectively where L is an arbitrarily fixed positive constant. We use $L = 1$ in this paper for simplicity. This, however, can be compensated by the selection of window sizes h_n and p_n which will be introduced later.

We now define rotational kernel functions as follows.

$$K_i(\theta, w, v) \stackrel{\triangle}{=} K_i^*(x, y) \qquad \ldots(1.3)$$

Variables (θ, w, v) and (x, y) are related by the following rotation transformation.

$$\begin{cases} x = \delta(w)\sqrt{w^2+v^2}\cos(\arctan(v/w)-\theta) \\ y = \delta(w)\sqrt{w^2+v^2}\sin(\arctan(v/w)-\theta) \end{cases} \qquad \ldots(1.4)$$

where $\theta \in [-\pi/2, \pi/2]$ is the rotation parameter, $(x, y) \in R^2, (w, v) \in R^2$ and $\delta(w) = 1$ or -1 when $w \geq 0$ or $w < 0$.

The non-zero-value domains of definition of the rotational kernel functions $K_i(\theta, w, v), i = 1, 2$, (defined by (1.3)-(1.4)) can be obtained by rotating the non-zero-value domains of definition of $K_i^*(x, y)$ by an angle θ. The values of $K_i(\theta, w, v)$ at (w, v) are equal to the values of $K_i^*(x, y)$ at (x, y) for $i = 1$ and 2 (c.f. Figure 1.1).

Put Figure 1.1 here

By convention, we use notation $K_i(\theta, x, y)$ instead of $K_i(\theta, w, v)$ in the rest part of this paper. Namely,

$$K_i(\theta, x, y) = K_i^*\left[\delta(x)\sqrt{x^2+y^2}\cos(\arctan(y/x)-\theta), \delta(x)\sqrt{x^2+y^2}\sin(\arctan(y/x)-\theta)\right].$$

We then define $M_n(\theta, x, y)$ as a difference of two kernel estimators as follows.

$$M_n(\theta, x, y) = \frac{1}{nh_np_n}\sum_{i=1}^{n} Z_i\left[K_2(\theta, \frac{x_i-x}{h_n}, \frac{y_i-y}{p_n}) - K_1(\theta, \frac{x_i-x}{h_n}, \frac{y_i-y}{p_n})\right]$$

where $(x, y) \in [b_n, 1-b_n]\times[b_n, 1-b_n]$ and $b_n = \sqrt{h_n^2/4 + p_n^2}$. The quantities h_n and p_n denote the window sizes. They are not necessary to be equal. Because of the weighted average nature of the kernel estimators, $M_n(\theta, x, y)$ is actually a difference between the weighted averages of the observations sited on two different sides of (x, y) along the direction of $(\cos\theta, \sin\theta)$ (c.f. the dotted rectangle in Figure 1.1). We can imagine that $|M_n(\theta, x_0, y_0)|$ will be very small no matter what the value of θ is if there is no jump at $(x_0, y_0) \in [b_n, 1-b_n] \times [b_n, 1-b_n]$. If (x_0, y_0) is on the jump location curve $(x, \phi(x))$, then $M_n(\theta, x_0, y_0)$ will be close to the jump magnitude $C(x_0)$ when $(\cos\theta, \sin\theta)$ is the tangent direction of $\phi(x)$ at x_0. Based on this observation, we can construct estimators of $\phi(x)$ and $C(x)$ as follows.

Let

$$|M_n(\theta^*(x, y), x, y)| = \max_{-\pi/2\leq\theta\leq\pi/2} |M_n(\theta, x, y)|$$

$$|M_n(\theta^*(x, \hat{\phi}(x)), x, \hat{\phi}(x))| = \max_{b_n\leq y\leq 1-b_n} |M_n(\theta^*(x, y), x, y)|$$

Then we use $\hat{\phi}(x)$ and $|M_n(\theta^*(x, \hat{\phi}(x)), x, \hat{\phi}(x))|$ as the estimators of $\phi(x)$ and $C(x)$ respectively and call them *Rotational Difference Kernel Estimators* (RDKEs). In Section 2, we will prove that they are a.s. consistent. Some simulation results will be presented in Section 4.

Estimation of the jump location curve of JRF is directly related to the edge detection problem in image processing. According to Haralick (1984), there are two kinds of edges in image processing. One is called *step edge*. The other one is called *roof edge*. The image intensity has jumps at step edges. The roof edges correspond to angles of the image intensity surface. So the step edge is similar to the jump location of the JRF while the roof edge corresponds to the jump location of the derivatives of the JRF. Jumps in derivatives will not be discussed in this paper. We can roughly divide edge detection methods into three categories (Sarker and Boyer, 1991). The first kind is the filtering methods. These methods are based on the fact that the derivative values of the image intensity are big or infinite at image edges. So there are many derivative-based criteria (called filters in image processing literature) in literature for edge detection such as the gradient, the Laplacian filters (page 85, vol. 2, Rosenfeld and Kak, 1982), the Laplacian of Gaussian filter (Marr and Hildreth, 1980) and some others (Canny, 1986; Qiu and Bhandarkar, 1996; Sarkar and Boyer, 1991). Design points (called pixels) in image processing are regularly spaced array of points. So the derivatives used in filters are usually replaced by the corresponding differences among observations within some neighborhoods. The second kind of methods models the image as a random field and try to detect changes of various statistical properties characterizing an edge (see Besag, 1986; Section 7.4, Cressie, 1990; Hansen and Elliot, 1982; etc.). The third kind of methods is based on surface fitting. We can fit a parametric model (usually polynomial model) in a neighborhood of a pixel. Then we can obtain estimators of the derivatives of the image intensity at this pixel and use these derivative estimates to do edge detection (Haralick, 1984). Other methods on edge detection are based on anisotropic diffusion (Perona and Malik, 1990; Saint-Marc *et al.*, 1991), residual analysis (Chen *et al.*, 1991) and global cost minimization using hill-climbing search (Tan *et al.*, 1989), simulated annealing (Tan *et al.*, 1991) and genetic algorithm (Bhandarkar *et al.*, 1994).

Comparing our RDKE method with those in image processing, there are three main differences. The first difference is that we establish some theory of the method. Statistical consistency is proved and convergence rates of the estimators are provided. The conditions imposed on the edge curves are mathematically explicitly expressed. The second difference is on the design points. Most edge detection methods in image processing make use of the "regularly spaced" pixels. (For example, filtering methods use differences among neighboring observations to replace derivatives.) But this "regularly spaced" condition on design points is hard to be satisfied in some other applications. On the other hand, it is not easy to generalize these methods to the cases with more general design points. A main difficulty is caused by the disorder of the design points. Our RDKE method do not need this "strict" condition on design points ("strict" for some other applications). All we need is the assumption (A) given in Section 2. This assumption requires the design points to have some homogeneity, but not necessarily to be regularly spaced. The third difference is on the use of derivatives. Many edge detection methods are based on the prerequisite condition that the image intensity has needed derivatives. This is reasonable for some problems. But it might be unreasonable for some others. Because of the assumed good smoothing properties of the image intensity, the gradient direction at a pixel (the direction along which the image intensity increases and/or decreases most rapidly) can be determined by the partial derivatives with respect to x and y (page 239, vol.1, Rosenfeld and Kak, 1982). So the edge detection problem is usually simplified to the estimation of the partial derivatives. And the partial derivative estimators are usually derived from the neighboring 2×2 observations (page 85, vol.2, Rosenfeld and Kak, 1982) or 3×3 or more observations (Haralick, 1984). RDKE does not need the derivative conditions. It only needs the Lipschitz (1) condition (c.f. conditions on $g(x, y)$ given in Section 2). That is because we use the difference of two directional kernel estimators (c.f. the construction of the RDKE) instead of the directional derivative estimators. Doing so does not lose anything but make our method be more efficient and flexible because of the "one-sided" nature of the kernels and the controllable window sizes. But a new problem appears at the same time. Because we have no derivative conditions, we have to search at a given design point for a direction with the biggest difference of two directional kernel estimators. We accomplish this by using a rotation transformation. But this step spends much computer time. As a conclusion, our method is more general and could be applied to more applications. But it requires more computing time. How can we lessen its computing time and keep its advantages at the same time ? This is what we will try to accomplish in our future research.

At the end of this section, we want to mention two facts. One is that it is straightforward to generalize the RDKE method to high dimensional cases. But like other multivariate methods, we will face the problem of "sparsity of data". At that time, some dimension reduction techniques should be useful such as the sliced inverse regression method (Li, 1991; Duan and Li, 1991), the projection pursuit regression method (Friedman and Stuetzle, 1981), etc. So far, it is unknown to us how to make the method be applicable in high dimensional cases. Another fact is that we have not considered the "boundary problem" in our method yet. Please read Gasser and Müller (1979), Rice (1984) and Qiu and Yandell (1997) for some boundary modification techniques.

2. Strong Consistency of the RDKE in an Ideal Case

In this section, we will prove the strong consistency of the RDKE in an ideal case that the jump location curve can be expressed by an explicit mathematical function as we mentioned in Section 1. Throughout this paper, we have the following assumption (A) on the design points $\{(x_i, y_i), i = 1, 2, \cdots, n\}$.

(A) There exists a partition $\Lambda = \{\Delta_i, i = 1, 2, \cdots, n\}$ of $[0,1] \times [0,1]$ such that

(1) $\bigcup_{i=1}^{n} \Delta_i = [0,1] \times [0,1]$; $\Delta_i \bigcap \Delta_j = \emptyset$, if $i \neq j$

(2) $(x_i, y_i) \in \Delta_i, \ \ i = 1, 2, \cdots, n$

(3) $\max_{1 \leq i \leq n} d_i = O(n^{-1/2})$, where d_i is the diameter of Δ_i for $i = 1, 2, \cdots, n$

(4) $\max_{1 \leq i \leq n} |S(\Delta_i) - 1/n| = O(n^{-1-\lambda})$, where $\lambda > 0$ is a constant and $S(\Delta_i)$ denotes the area of Δ_i.

REMARK 2.1. Assumption (A) requires that the design points have some homogeneity. This is a basic assumption of multivariate kernel regression methods (c.f. Chapter 6, Müller, 1988).

REMARK 2.2. The region $[0,1] \times [0,1]$ in assumption (A) will be replaced by a more general region $\Omega \subset R^2$ in Section 3.

For presentation convenience, the following notations will be used in the rest part of this paper.

(I) $||f(x)||_{[a,b]} \overset{\triangle}{=} \max_{a \leq x \leq b} |f(x)|$

(II) $||f(x,y)||_{[a,b]\times[c,d]} \overset{\triangle}{=} \max_{a \leq x \leq b, c \leq y \leq d} |f(x,y)|$

(III) $||f(\theta,x,y)||_{[-\pi/2,\pi/2]\times[a,b]\times[c,d]} \overset{\triangle}{=} \max_{-\pi/2 \leq \theta \leq \pi/2, a \leq x \leq b, c \leq y \leq d} |f(\theta,x,y)|$

First of all, we give an a.s. consistency result (Theorem 2.1) of a bivariate kernel regression estimator with a rotation parameter θ in the case that the regression function is continuous. We will use this conclusion when we establish the a.s. consistency of the RDKE afterwards. Proof of Theorem 2.1 will be given in Appendix A.

THEOREM 2.1. *Let ν be a positive number and $\{\beta_n\}$ be a series of numbers satisfying $\lim_{n\to\infty} \beta_n = \infty$. The design points $\{(x_i, y_i), i = 1, 2, \cdots, n\}$ satisfy the assumption (A). $f(x,y)$ satisfies the Lipschitz (1) condition. $E|\varepsilon_1|^p < \infty$ for $p \geq 2$. $K^*(x,y)$ is a non-negative kernel function which satisfies (i) $K^*(x,y) = 0$ when $(x,y) \notin [-1,1] \times [-1,1]$; (ii) $K^*(x,y)$ is Lipschitz (1) continuous and (iii) $\int_{-1}^{1}\int_{-1}^{1} K^*(x,y)\, dxdy = 1$. h_n and p_n are positive window sizes which satisfy, when n is large enough, the conditions that (1) $\frac{n^\nu}{\beta_n \log n}[h_n + p_n + \frac{1}{n^\lambda h_n p_n} + \frac{\sqrt{h_n^2+p_n^2}}{\sqrt{n} h_n^2 p_n^2}] = o(1)$; (2) $\frac{n^{2\nu}}{n h_n p_n \beta_n} = O(1)$ and (3) $\frac{n^{\nu+1/p-1}}{h_n p_n \beta_n \log n} = o(1)$. Then, for any constants $0 < a \leq b < 1$ and $0 < c \leq d < 1$, we have*

$$\lim_{n\to\infty} \frac{n^\nu}{\beta_n \log n} ||f_n(x,y) - f(x,y)||_{[a,b]\times[c,d]} = 0, \ \ a.s.$$

and

$$\lim_{n\to\infty} \frac{n^\nu}{\beta_n \log n} ||f_n(\theta,x,y) - f(x,y)||_{[-\pi/2,\pi/2]\times[a,b]\times[c,d]} = 0, \ \ a.s.$$

where

$$f_n(x,y) = \frac{1}{nh_np_n}\sum_{i=1}^{n} Z_iK^*(\frac{x_i-x}{h_n},\frac{y_i-y}{p_n})$$

$$f_n(\theta,x,y) = \frac{1}{nh_np_n}\sum_{i=1}^{n} Z_iK(\theta,\frac{x_i-x}{h_n},\frac{y_i-y}{p_n})$$

and

$$\begin{aligned} K(\theta,x,y) &= K^*\Big(\delta(x)\sqrt{x^2+y^2}\cos(\arctan(y/x)-\theta), \\ &\quad \delta(x)\sqrt{x^2+y^2}\sin(\arctan(y/x)-\theta)\Big). \end{aligned}$$

$f_n(x,y)$ in Theorem 2.1 is the Priestley-Chao kernel regression estimator. $f_n(\theta,x,y)$ uses the rotational kernel function $K(\theta,x,y)$ (c.f. the rotation transformation (1.4) and the definition of a rotational kernel function in Section 1). Theorem 2.1 says that the kernel estimator is uniformly a.s. consistent with respect to all x,y and θ.

COROLLARY 2.1. In Theorem 2.1, if we choose $h_n = p_n = O(n^{-1/8}), \nu = 1/8, \lambda \geq 3/8$, then

$$\lim_{n\to\infty}\frac{n^{1/8}}{\beta_n\log n}\|f_n(\theta,x,y)-f(x,y)\|_{[-\pi/2,\pi/2]\times[a,b]\times[c,d]} = 0, \quad a.s.$$

REMARK 2.3. If we use product kernel (namely, $K^*(x,y)$ has the form $K_1^*(x)K_2^*(y)$) or high order kernel, then the rates of convergence in Theorem 2.1 and Corollary 2.1 could be faster (c.f. Müller, 1988; Härdle, 1990, 1991). This remark is also true for the consistency results of RDKE discussed below.

Now let us return to the jump surface case. Under the jump regression model (1.1)-(1.2), we have the following consistency result.

THEOREM 2.2. *If the design points satisfy the assumption (A); the kernel functions $K_i^*(x,y)$ satisfy the conditions (i)-(iii) in Section 1 and (iv) $K_i^*(x,y)$ are Lipschitz (1) continuous for $i=1$ and 2; $E|\varepsilon_1|^p<\infty$ for $p\geq 2$; $g(x,y)$ is Lipschitz (1) continuous; h_n and p_n satisfy the conditions stated in Theorem 2.1; $\phi(x)$ has continuous second order derivative and it also satisfies $0<\phi(x)<1$ for any $x\in[0,1]$; $C(x)\neq 0$ for any $x\in[0,1]$, then for any $0<a\leq b<1$, we have*

$$\lim_{n\to\infty}\|\hat\phi(x)-\phi(x)\|_{[a,b]} = 0, \quad a.s.$$

and

$$\lim_{n\to\infty}\|M_n(\theta^*(x,\hat\phi(x)),x,\hat\phi(x))-C(x)\|_{[a,b]} = 0, \quad a.s.$$

PROOF. The dashed rectangle in Figure 2.1 denotes the non-zero-value domain of definition of $K_1(\arctan(\phi'(x_0)),(x-x_0)/h_n,(y-\phi(x_0))/p_n)$, where $a\leq x_0\leq b$.

<u>Put Figure 2.1 here</u>

It consists of two parts:

I = $\{(x,y)$: (x,y) belongs to the dashed rectangle and $y\geq\phi(x)$ $\}$
II = $\{(x,y)$: (x,y) belongs to the dashed rectangle and $y<\phi(x)$ $\}$

Then

$$\begin{aligned} & f_n^{(1)}(\arctan(\phi'(x_0)),x_0,\phi(x_0)) \\ \triangleq\ & \frac{1}{nh_np_n}\sum_{i=1}^{n} Z_iK_1(\arctan(\phi'(x_0)),\frac{x-x_0}{h_n},\frac{y-\phi(x_0)}{p_n}) \\ =\ & \frac{1}{nh_np_n}(\sum{}' + \sum{}'')Z_iK_1(\arctan(\phi'(x_0)),\frac{x-x_0}{h_n},\frac{y-\phi(x_0)}{p_n}) \end{aligned}$$

where $\sum'$ denotes summation of the terms with design points belong to I and $\sum''$ denotes summation of the remaining terms.

Let $\{Z_i^{(1)}\}$ and $\{Z_i^{(2)}\}$ be observations which design points belong to I and II respectively. Then

$$\begin{aligned} & f_n^{(1)}(\arctan(\phi'(x_0)), x_0, \phi(x_0)) \\ = \quad & \frac{1}{nh_np_n}\sum{}' (Z_i^{(1)} - C(x_i))K_1(\arctan(\phi'(x_0)), \frac{x-x_0}{h_n}, \frac{y-\phi(x_0)}{p_n}) \\ & + \frac{1}{nh_np_n}\sum{}'' Z_i^{(2)}K_1(\arctan(\phi'(x_0)), \frac{x-x_0}{h_n}, \frac{y-\phi(x_0)}{p_n}) \\ & + \frac{1}{nh_np_n}\sum{}' C(x_i)K_1(\arctan(\phi'(x_0)), \frac{x-x_0}{h_n}, \frac{y-\phi(x_0)}{p_n}). \end{aligned} \qquad \ldots(2.1)$$

According to Theorem 2.1, the summation of the first two terms of (2.1) has an a.s. limit $f_-(x_0, \phi(x_0)) \triangleq \lim_{(x,y)\to(x_0,\phi(x_0)), y<\phi(x)} f(x,y)$.

Since $\phi(x)$ has continuous second order derivative, it has the following Taylor expansion

$$\phi(x) = \phi(x_0) + \phi'(x_0)(x-x_0) + \frac{\phi''(x_0)}{2!}(x-x_0)^2 + o((x-x_0)^2).$$

By this expression, it is not difficult to check that the area of part I is $O(b_n^3)$. This along with a simple fact that $\lim_{n\to\infty} \frac{b_n^3}{h_n p_n} = 0$ gives us a conclusion that the third term of (2.1) converges to zero. Hence

$$\lim_{n\to\infty} f_n^{(1)}(\arctan(\phi'(x_0)), x_0, \phi(x_0)) = f_-(x_0, \phi(x_0)), \quad a.s.$$

By the same argument, we have

$$\begin{aligned} & \lim_{n\to\infty} f_n^{(2)}(\arctan(\phi'(x_0)), x_0, \phi(x_0)) \\ \triangleq \quad & \lim_{n\to\infty} \frac{1}{nh_np_n}\sum_{i=1}^{n} Z_iK_2(\arctan(\phi'(x_0)), \frac{x-x_0}{h_n}, \frac{y-\phi(x_0)}{p_n}) \\ = \quad & \lim_{(x,y)\to(x_0,\phi(x_0)), y>\phi(x)} f(x,y), \quad a.s. \end{aligned}$$

So

$$\lim_{n\to\infty} M_n\left(\arctan(\phi'(x_0)), x_0, \phi(x_0)\right) = C(x_0), \quad a.s. \qquad \ldots(2.2)$$

It is not difficult to check that the above convergence is uniformly true with respect to $x_0 \in [a,b]$. By (2.2), we have

$$\liminf_{n\to\infty} |M_n\left(\theta(x_0, \hat\phi(x_0)), x_0, \hat\phi(x_0)\right)| \geq C(x_0) > 0, \quad a.s.$$

By using the similar arguments to those in the proof of Theorem 1 in Qiu *et al.* (1991), we can prove that

$$\lim_{n\to\infty} M_n\left(\theta(x_0, \hat\phi(x_0)), x_0, \hat\phi(x_0)\right) = C(x_0), \quad a.s. \qquad \ldots(2.3)$$

and the convergence is also uniformly true with respect to $x_0 \in [a,b]$.

Let $\phi_{+b_n}(x), x \in [0,1]$, be a curve which has properties that $\phi_{+b_n}(x) \geq \phi(x), x \in [0,1]$, and the distance between $\phi_{+b_n}(x_0)$ and $\{(x,\phi(x)), x \in [0,1]\}$ is b_n for any $x_0 \in [0,1]$. (The distance used here has the same meaning as the distance used in Section 3.) The definition of $\phi_{-b_n}(x)$ is the same as that of $\phi_{+b_n}(x)$ except that $\phi_{-b_n}(x) \leq \phi(x)$ for any $x \in [0,1]$.

According to Theorem 2.1, it is easy to see that

$$\lim_{n\to\infty} ||M_n(\theta(x,y), x, y)||_{[a,b]\times[b_n, \phi_{-b_n}(x)]} = 0, \quad a.s. \qquad \ldots(2.4)$$

$$\lim_{n\to\infty} ||M_n(\theta(x,y), x, y)||_{[a,b]\times[\phi_{+b_n}(x), 1-b_n]} = 0, \quad a.s. \qquad \ldots(2.5)$$

Combining (2.3)-(2.5), we have the conclusion that when n is large enough

$$\phi_{-b_n}(x) \le \hat{\phi}(x) \le \phi_{+b_n}(x), \quad a.s.$$

and the inequalities are uniformly true with respect to $x \in [a,b]$. Hence

$$||\hat{\phi}(x) - \phi(x)||_{[a,b]} = O(b_n), \quad a.s.$$

The proof of the theorem is finished.

REMARK 2.4. From the above proof, we know that if we choose h_n and p_n as in the Corollary 2.1 and if $\lambda \ge 3/8$, then the rate of convergence of $||M_n(\theta^*(x,\hat{\phi}(x)), x, \hat{\phi}(x)) - C(x)||_{[a,b]}$ is $O(n^{-1/8}\log n)$ and the rate of convergence of $||\hat{\phi}(x) - \phi(x)||_{[a,b]}$ is $O(n^{-1/8})$.

REMARK 2.5. The condition $0 < \phi(x) < 1$ in Theorem 2.2 can be weakened to $0 \le \phi(x) \le 1$ where the equations can be reached at most finite times. $C(x)$ can also reach zero at finite points in $[0,1]$.

In the process of constructing the RDKE, we defined $\theta^*(x,y)$ with $(x,y) \in [b_n, 1-b_n] \times [b_n, 1-b_n]$. Intuitively, $\theta^*(x,\phi(x))$ should be close to the tangent angle of $\phi(x)$ at x. Its natural estimator $\theta^*(x,\hat{\phi}(x))$ should have this property too. The following Theorem 2.3 verifies our speculation.

THEOREM 2.3. *Under the conditions stated in Theorem 2.2, we have*

(i) $\lim_{n\to\infty} \theta^*(x,\phi(x)) = \arctan(\phi'(x)), \quad \forall x \in (0,1), \quad a.s.;$
$\lim_{n\to\infty} M_n(\theta^*(x,\phi(x)), x, \phi(x)) = C(x), \quad \forall x \in (0,1), a.s.$

(ii) $\lim_{n\to\infty} \theta^*(x,\hat{\phi}(x)) = \arctan(\phi'(x)), \quad \forall x \in (0,1), a.s.$

PROOF OF (i). Without loss of generality, we assume that $C(x) > 0$ for $x \in (0,1)$. Then by (2.2),

$$\liminf_{n\to\infty} |M_n(\theta^*(x,\phi(x)), x, \phi(x))| \ge \lim_{n\to\infty} |M_n(\arctan(\phi'(x)), x, \phi(x))| = C(x), \quad a.s. \quad \ldots(2.6)$$

Let Ω_0 be a sample subspace which satisfies $P(\Omega_0) = 1$ and (2.6) is true for any $\omega \in \Omega_0$. If $\theta^*(x,\phi(x))$ does not converge to $\arctan(\phi'(x))$ for some $\omega \in \Omega_0$ (when $n \to \infty$), then there exist some $0 < \theta_0 < \pi$ and a subseries of $\theta^*(x,\phi(x))$ (this subseries is also denoted as $\theta^*(x,\phi(x))$) such that $|\theta^*(x,\phi(x)) - \arctan(\phi'(x))| > \theta_0$. For this subseries, we have

$$\begin{aligned}
&M_n(\theta^*(x,\phi(x)), x, \phi(x)) \\
= \;& \frac{1}{nh_np_n}\sum_{i=1}^{n} \bar{Z}_i K_2(\theta^*(x,\phi(x)), (x_i - x)/h_n, (y_i - \phi(x))/p_n) \\
& - \frac{1}{nh_np_n}\sum_{i=1}^{n} \underline{Z}_i K_1(\theta^*(x,\phi(x)), (x_i - x)/h_n, (y_i - \phi(x))/p_n) \\
& - \frac{1}{nh_np_n}\sum_{i=1}^{n} C(x_i) I_{y_i \le \phi(x_i)} K_2(\theta^*(x,\phi(x)), (x_i - x)/h_n, (y_i - \phi(x))/p_n) \\
& - \frac{1}{nh_np_n}\sum_{i=1}^{n} C(x_i) I_{y_i > \phi(x_i)} K_1(\theta^*(x,\phi(x)), (x_i - x)/h_n, (y_i - \phi(x))/p_n)
\end{aligned} \quad \ldots(2.7)$$

where $\bar{Z}_i = Z_i + C(x_i) I_{y_i \le \phi(x_i)}$ and $\underline{Z}_i = Z_i - C(x_i) I_{y_i > \phi(x_i)}$.

It is not difficult to see that the first two terms of (2.7) converge to $C(x)$. For the fourth term, we have the following results.

Firstly,

$$\lim_{n\to\infty} \frac{1}{nh_np_n}\sum_{i=1}^{n} C(x_i) I_{y_i > \phi(x_i)} K_1(\theta^*(x,\phi(x)), (x_i - x)/h_n, (y_i - \phi(x))/p_n)$$

$$\begin{aligned}
\geq\ & \lim_{n\to\infty} \min\{\frac{1}{nh_np_n}\sum_{i=1}^{n} C(x_i)I_{y_i>\phi(x_i)}K_1(\arctan(\phi'(x))+\theta_0, \\
& (x_i-x)/h_n,(y_i-\phi(x))/p_n), \\
& \frac{1}{nh_np_n}\sum_{i=1}^{n} C(x_i)I_{y_i>\phi(x_i)}K_1(\arctan(\phi'(x))-\theta_0, \\
& (x_i-x)/h_n,(y_i-\phi(x))/p_n)\} \\
=\ & \min\{S_1C(x)\int\int_{-\theta_0\leq\arctan(y/x)\leq 0} K_1(x,y)\ dxdy, \\
& S_1C(x)\int\int_{\pi\leq\arctan(y/x)\leq\pi+\theta_0} K_1(x,y)\ dxdy\}
\end{aligned}$$

where S_1 is the area of the region

$$\{(-\theta_0\leq\arctan(y/x)\leq 0)\bigcap([-1/2,1/2]\times[-1,0])\}.$$

Secondly,

$$\begin{aligned}
& \lim_{n\to\infty}\frac{1}{nh_np_n}\sum_{i=1}^{n} C(x_i)I_{y_i>\phi(x_i)}K_1(\theta^*(x,\phi(x)),(x_i-x)/h_n,(y_i-\phi(x))/p_n) \\
\leq\ & \lim_{n\to\infty}\max\{\frac{1}{nh_np_n}\sum_{i=1}^{n} C(x_i)I_{y_i>\phi(x_i)}K_1(\pi/2,(x_i-x)/h_n,(y_i-\phi(x))/p_n), \\
& \frac{1}{nh_np_n}\sum_{i=1}^{n} C(x_i)I_{y_i>\phi(x_i)}K_1(-\pi/2,(x_i-x)/h_n,(y_i-\phi(x))/p_n)\} \\
=\ & \max\{S_2C(x)\int\int_{-(\pi/2-\arctan(\phi'(x)))\leq\arctan(y/x)\leq 0} K_1(x,y)\ dxdy, \\
& S_3C(x)\int\int_{\pi\leq\arctan(y/x)\leq 3\pi/2+\arctan(\phi'(x))} K_1(x,y)\ dxdy\}
\end{aligned}$$

where S_2 is the area of the region

$$\{(-(\pi/2-\arctan(\phi'(x)))\leq\arctan(y/x)\leq 0)\bigcap([-1/2,1/2]\times[-1,0])\}$$

and S_3 is the area of the region

$$\{(\pi\leq\arctan(y/x)\leq 3\pi/2+\arctan(\phi'(x)))\bigcap([-1/2,1/2]\times[-1,0])\}.$$

We can discuss the third term of (2.7) in the same way. As a result,

$$\liminf_{n\to\infty}|M_n\left(\theta^*(x,\phi(x)),x,\phi(x)\right)|<C(x)$$

and this contradicts with (2.6). So

$$\lim_{n\to\infty}\theta^*(x,\phi(x))=\arctan(\phi'(x)),\ \text{ for }\forall\ \omega\in\Omega_0.$$

By using this conclusion, we know that for $\forall\ \omega\in\Omega_0$ and $\forall\ 0<\theta_1<\pi/2$, there exists a positive integer N such that $|\theta^*(x,\phi(x))-\arctan(\phi'(x))|<\theta_1$, for $n>N$.

When n is large enough, it is not difficult to check that

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{nh_np_n}\sum_{i=1}^{n} C(x_i)I_{y_i>\phi(x_i)}K_1(\theta^*(x,\phi(x)),(x_i-x)/h_n,(y_i-\phi(x))/p_n) \\
\leq\ & \lim_{n\to\infty} \max\{\frac{1}{nh_np_n}\sum_{i=1}^{n} C(x_i)I_{y_i>\phi(x_i)}K_1(\arctan(\phi'(x))+\theta_1,(x_i-x)/h_n,(y_i-\phi(x))/p_n), \\
& \qquad \frac{1}{nh_np_n}\sum_{i=1}^{n} C(x_i)I_{y_i>\phi(x_i)}K_1(\arctan(\phi'(x))-\theta_1,(x_i-x)/h_n,(y_i-\phi(x))/p_n)\} \\
=\ & \max\{\frac{\tan(\theta_1)}{8}C(x)\int\int_{-\theta_1\leq\arctan(y/x)\leq 0} K_1(x,y)\ dxdy, \\
& \frac{\tan(\theta_1)}{8}C(x)\int\int_{\pi\leq\arctan(y/x)\leq\pi+\theta_1} K_1(x,y)\ dxdy\}.
\end{aligned}$$

Because of the arbitrariness of θ_1, we have

$$\lim_{n\to\infty} \frac{1}{nh_np_n}\sum_{i=1}^{n} C(x_i)I_{y_i>\phi(x_i)}K_1(\theta^*(x,\phi(x)),(x_i-x)/h_n,(y_i-\phi(x))/p_n) = 0.$$

By the same reason, the third term of (2.7) converges to zero. So

$$\lim_{n\to\infty} M_n(\theta^*(x,\phi(x)),x,\phi(x)) = C(x),\quad a.s.$$

By now, we have finished the proof of conclusion (i) of the theorem. The conclusion (ii) of the theorem could be proved in a similar way.

REMARK 2.6. It is not difficult to check that the convergences in the above theorem are uniformly true with respect to $x\in[a,b]$ where $0<a\leq b<1$.

Our RDKE method can be easily generalized to the case with more than one jump location curves. That is, the regression function $f(x,y)$ has the following form

$$f(x,y) = g(x,y) + \sum_{i=1}^{q} C_i(x)I_{y>\phi_i(x)},\quad (x,y)\in[0,1]\times[0,1]$$

where $g(x,y)$ is the continuous part, q is the number of jump location curves, $0<\phi_1(x)<\phi_2(x)<\cdots<\phi_q(x)<1$ are functions representing the jump location curves, $C_i(x), i=1,2,\cdots,q$, are related to the jump magnitude functions. Without loss of generality, we assume that $q=2$. Let

$$|M_n(\theta^*(x,\hat\phi_1(x)),x,\hat\phi_1(x))| = \max_{b_n\leq y\leq 1-b_n} |M_n(\theta^*(x,y),x,y)|$$

$$|M_n(\theta^*(x,\hat\phi_2(x)),x,\hat\phi_2(x))| = \max_{b_n\leq y\leq 1-b_n,|\hat\phi_1(x)-y|>b_n} |M_n(\theta^*(x,y),x,y)|$$

Let $\hat\phi_{(1)}(x)\leq\hat\phi_{(2)}(x)$, for $x\in[b_n,1-b_n]$, denote the order statistics of $\hat\phi_1(x)$ and $\hat\phi_2(x)$. Then we can use $\hat\phi_{(1)}(x),\hat\phi_{(2)}(x),|M_n(\theta^*(x,\hat\phi_{(1)}(x)),x,\hat\phi_{(1)}(x))|$ and $|M_n(\theta^*(x,\hat\phi_{(2)}(x)),x,\hat\phi_{(2)}(x))|$ as the estimators of $\phi_1(x),\phi_2(x),C_1(x)$ and $C_2(x)$ respectively and these estimators have all the consistency properties of the RDKE proved in the case with unique jump location curve.

3. RDKE in a More General Case

Firstly, let us discuss a simple case that the jump location curve is a circuit and it can be expressed by two functions $\phi_1(x) < \phi_2(x)$ on $[A, B]$ (c.f. Figure 3.1), where $0 \le A < B \le 1$ are two unknown constants. The corresponding jump magnitude functions are denoted by $C_1(x)$ and $C_2(x)$ which satisfy $|C_i(x)| \ge C_0 > 0$ for $i = 1, 2$ and $x \in [A, B]$, where C_0 is a known constant.

Put Figure 3.1 here

Let

$$\begin{aligned}\hat{A} &= \inf\Big\{x : |M_n(\theta^*(x, \hat{\phi}_{(1)}(x)), x, \hat{\phi}_{(1)}(x))| \\ &\quad +|M_n(\theta^*(x, \hat{\phi}_{(2)}(x)), x, \hat{\phi}_{(2)}(x))| \ge C_0, b_n \le x \le 1 - b_n\Big\} \\ \hat{B} &= \sup\Big\{x : |M_n(\theta^*(x, \hat{\phi}_{(1)}(x)), x, \hat{\phi}_{(1)}(x))| \\ &\quad +|M_n(\theta^*(x, \hat{\phi}_{(2)}(x)), x, \hat{\phi}_{(2)}(x))| \ge C_0, b_n \le x \le 1 - b_n\Big\}\end{aligned}$$

where $\hat{\phi}_{(1)}(x) \le \hat{\phi}_{(2)}(x)$ are the order statistics of $\hat{\phi}_1(x)$ and $\hat{\phi}_2(x)$ as we defined at the end of Section 2. Then it is not hard to check that $\hat{A}, \hat{B}, \hat{\phi}_{(1)}(x), \hat{\phi}_{(2)}(x)$, $|M_n(\theta^*(x, \hat{\phi}_{(1)}(x)), x, \hat{\phi}_{(1)}(x))|$ and $|M_n(\theta^*(x, \hat{\phi}_{(2)}(x)), x, \hat{\phi}_{(2)}(x))|$ are a.s. consistent estimators of $A, B, \phi_1(x), \phi_2(x), C_1(x)$ and $C_2(x)$ respectively.

REMARK 3.1. If C_0 is unknown, then we can use a series of constants $B_n = O(n^{-\nu}\beta_n \log n)$ to replace C_0 in the construction of $\hat{A}$ and $\hat{B}$ where ν and $\{\beta_n\}$ have the same meaning as those in Theorem 2.1. If the conditions in Corollary 2.1 are all satisfied, then we can use $B_n = O(n^{-1/8} \log n \log\log n)$. By using the fact that the rates of convergence of

$$\lim_{n\to\infty} ||M_n(\theta^*(x, \hat{\phi}_{(i)}(x)), x, \hat{\phi}_{(i)}(x))||_{(b_n, A-b_n)\bigcup(B+b_n, 1-b_n)} = 0$$

and

$$\lim_{n\to\infty} ||M_n(\theta^*(x, \hat{\phi}_{(i)}(x)), x, \hat{\phi}_{(i)}(x)) - C_i(x)||_{(A+b_n, B-b_n)} = 0, \quad i = 1, 2, \quad a.s.$$

are both faster than B_n, the consistency of the estimators can be proved without much difficulty.

Now let us discuss a more general case that the jump location curve ℓ has the following parametric form:

$$\begin{cases} x = x(t) \\ y = y(t), \quad t \in [0, T]. \end{cases}$$

The parameter t is the length of the curve from point $(x(0), y(0))$ to point $(x(t), y(t))$. As we know, almost all of the curves in applications can be expressed in this way.

In this paper, we assume that ℓ has no overlap. Namely, there are no $t_1 < t_2 \in [0, T]$ such that $(x(t_1), y(t_1)) = (x(t_2), y(t_2))$. Let us define the positive direction of ℓ by the parameter t from 0 to T. Then ℓ divide Ω, the domain of definition of the regression function $f(x, y)$, into two parts D_+ and D_-, where D_+ denotes the left-side part including the curve ℓ itself and D_- is the right-side part. Let $D = \{(x(t), y(t)) : t \in [0, T]\}$ be the image of ℓ. Then the regression function $f(x, y)$ can be expressed as follows.

$$f(x, y) = g(x, y) + C(x, y)I_{(x,y)\in D_+}, (x, y) \in \Omega,$$

where $g(x, y)$ and $C(x, y)$ are continuous functions, $C(x(t), y(t))$ with $t \in [0, T]$ is the corresponding jump magnitude function.

Suppose that G_1 and G_2 are two subsets of R^2. Then the Hausdorff distance between G_1 and G_2 is defined by

$$d(G_1, G_2) \triangleq \max\left\{\sup_{x\in G_1}\inf_{y\in G_2} ||x - y||, \sup_{x\in G_2}\inf_{y\in G_1} ||x - y||\right\}.$$

REMARK 3.2. If the Hausdorff distance between G_1 and G_2 is zero, then their closures are the same.

Denote $\Omega_{b_n} = \{(x,y) : d((x,y),\partial\Omega) \geq b_n, (x,y) \in \Omega\}$ where $\partial\Omega$ is the boundary set of Ω. For any $(x,y) \in \Omega_{b_n}$, $M_n(\theta^*(x,y),x,y)$ is defined as in Section 1. B_n is defined as in Remark 3.1. Let

$$\begin{aligned}\bar{D} &= \{(x,y) : |M_n(\theta^*(x,y),x,y)| \geq B_n, (x,y) \in \Omega_{b_n}\} \\ O_n(x,y) &= \{(x',y') : \sqrt{(x'-x)^2 + (y'-y)^2} \leq b_n, (x',y') \in \Omega_{b_n}\} \\ D_n &= \bigcup_{(x,y)\in D} O_n(x,y)\end{aligned}$$

Then we use $\bar{D}$ and $\{|M_n(\theta^*(x,y),x,y)| : (x,y) \in \bar{D}\}$ as the estimators of D and $\{|C(x,y)| : (x,y) \in D\}$ respectively.

THEOREM 3.1 *Suppose that the design points $\{(x_i,y_i), i = 1,2,\cdots,n\}$ satisfy the assumption (A); $E|\varepsilon_1|^p < \infty$ for $p \geq 2$; $g(x,y)$ is Lipschitz (1) continuous; $C(x,y)$ is continuous and $C(x,y) \neq 0$ at any $(x,y) \in D_+$; $x(t)$ and $y(t)$ have second order derivatives at any $t \in (0,T)$; $\nu, \{\beta_n\}, K_1^*(x,y), K_2^*(x,y), h_n$ and p_n are defined as in Theorem 2.2. Then*

$$\lim_{n\to\infty} (n^\nu/(\beta_n \log n)) d(\bar{D}, D) = 0, \quad a.s.$$

PROOF. According to Theorem 2.1,

$$\lim_{n\to\infty} ||M_n(\theta^*(x,y),x,y)||_{\Omega_{b_n}\setminus D_n} = 0, \quad a.s.$$

The rate of convergence of the above expression is $o(B_n)$. So $\bar{D} \subset D_n$, $a.s.$, when n is large enough. By the condition that $x(t)$ and $y(t)$ have second order derivatives, we can verify in the same way as that in the proof of Theorem 2.3 that

$$\lim_{n\to\infty} |M_n(\theta^*(x,y),x,y)| = |C(x,y)|, \text{for } \forall\ (x,y) \in D, \quad a.s.$$

and the convergence is uniformly true with respect to $(x,y) \in D$.

On the other hand,

$$\min_{(x,y)\in D} |C(x,y)| > 0.$$

This is a direct conclusion from the conditions that $C(x,y)$ is continuous and $C(c,y) \neq 0$ at any $(x,y) \in D_+$ and from the fact that D is a tight set. So when n is large enough,

$$D \subset \bar{D} \subset D_n, \quad a.s.$$

As a result, when n is large enough,

$$d(D,\bar{D}) \leq d(D,D_n), \quad a.s.$$

It is obvious that $d(D,D_n) \leq b_n$. By the condition (1) on h_n and p_n given in Theorem 2.1, we have $\lim_{n\to\infty} n^\nu b_n/(\beta_n \log n) = 0$. So

$$\lim_{n\to\infty} \frac{n^\nu}{\beta_n \log n} d(\bar{D}, D) = 0, \quad a.s.$$

REMARK 3.3. If the conditions in Corollary 2.1 are satisfied, then the rate of convergence of $\lim_{n\to\infty} d(\bar{D},D) = 0$, $a.s.$, reaches $O(n^{-1/8})$.

4. Some Simulation Results

In this section, we give some simulation results about the performance of RDKE in the ideal case discussed in Sections 1 and 2. The jump regression function used is $f(x,y) = \frac{1}{4}(1-x)y + [2+0.2\sin(2\pi x)]I_{y\geq 0.6\sin(\pi x)+0.2}$ for $(x,y) \in [0,1]\times[0,1]$, where $\phi(x) = 0.6\sin(\pi x)+0.2$ represents the jump location curve, $C(x) = 2+0.2\sin(2\pi x)$ is the jump magnitude function.

$\varepsilon_1 \sim N(0,1)$. The window sizes h_n and p_n are chosen to be equal to $\frac{1}{3}n^{-1/8}$. $K_2^*(x,y) = \frac{12}{11}(1-x^2) \cdot \frac{12}{11}(1-(y-0.5)^2)I_{[-1/2,1/2]\times[0,1]}$. It is a product of two Epanechnikov kernels (see page 45, Härdle, 1991). $K_1^*(x,y) = K_2^*(x,-y)$. Design points are regularly spaced in $[0,1]\times[0,1]$. Figure 4.1 shows the jump location curve (plot (a)), the jump magnitude function (plot(b)), the true regression surface (plot (c)) and a set of noisy data with $n = 400$ (plot (d)).

Put Figure 4.1 here

We then use the so-called scanning method to search for $\theta^*(x,y)$ and $\hat{\phi}(x)$. $[-\pi/2, \pi/2]$ is equally divided into 100 portions. So is the interval $[0,1]$ on y-axis. Figure 4.2 presents some simulation results at $x = 0.2, 0.4, 0.6$ and 0.8. We can see that the behavior of RDKE is acceptable in moderate sample size case ($n = 400$) and it is quite satisfactory when the sample size is large ($n = 10000$).

Put Figure 4.2 here

Appendix

LEMMA 1. *If $K^*(x,y)$ is Lipschitz (1) continuous, namely, there exists a constant $M > 0$ such that*

$$|K^*(x_1,y_1) - K^*(x_2,y_2)| \leq M\sqrt{(x_1-x_2)^2 + (y_1-y_2)^2},$$

for any $(x_i,y_i) \in [0,1]\times[0,1], i = 1,2$, then for any $\theta \in [-\pi/2,\pi/2]$, $K(\theta,x,y)$ which is defined by

$$K(\theta,x,y) = K^*\left(\delta(x)\sqrt{x^2+y^2}\cos(\arctan(y/x)-\theta), \delta(x)\sqrt{x^2+y^2}\sin(\arctan(y/x)-\theta)\right)$$

satisfies the Lipschitz (1) condition too with respect to x and y.

(The proof is trivial by using the fact that the rotation transformation is isometric.)

LEMMA 2. *If $K^*(x,y)$ is Lipschitz (1) continuous, then for any $(x,y) \in [0,1]\times[0,1]$, $K(\theta,x,y)$ satisfies the Lipschitz (1) condition with respect to θ.*

(The proof is trivial.)

LEMMA 3. *Suppose that the design points $\{(x_i,y_i), i = 1,2,\cdots,n\}$ satisfy the assumption (A); $K^*(x,y)$ is a non-negative kernel function which satisfies (i) $K^*(x,y) = 0$ when $(x,y) \notin [-1,1]\times[-1,1]$ (ii) $\int_{-1}^{1}\int_{-1}^{1} K^*(x,y)\,dxdy = 1$ and (iii) $K^*(x,y)$ is Lipschitz (1) continuous; the regression function $f(x,y)$ is Lipschitz (1) continuous, then for any $0 < a \leq b < 1$ and $0 < c \leq d < 1$, we have*

$$||Ef_n(x,y) - f(x,y)||_{[a,b]\times[c,d]} = O(h_n+p_n) + O(\frac{1}{n^\lambda h_n p_n}) + O(\frac{\sqrt{h_n^2+p_n^2}}{\sqrt{n}h_n^2p_n^2})$$

where

$$f_n(x,y) = \frac{1}{nh_np_n}\sum_{i=1}^{n} Z_iK^*(\frac{x_i-x}{h_n}, \frac{y_i-y}{p_n}).$$

PROOF.

$$\begin{aligned}
& |Ef_n(x,y) - f(x,y)| \\
= \quad & |\frac{1}{nh_np_n}\sum_{i=1}^{n}[f(x_i,y_i) - f(x,y)]K^*(\frac{x_i - x}{h_n}, \frac{y_i - y}{p_n}) + \\
& f(x,y)[\frac{1}{nh_np_n}\sum_{i=1}^{n}K^*(\frac{x_i - x}{h_n}, \frac{y_i - y}{p_n}) - 1]| \\
\leq \quad & \frac{C_f}{nh_np_n}\sum_{i=1}^{n}\sqrt{(x_i - x)^2 + (y_i - y)^2}K^*(\frac{x_i - x}{h_n}, \frac{y_i - y}{p_n}) + \\
& |f(x,y)||\frac{1}{nh_np_n}\sum_{i=1}^{n}K^*(\frac{x_i - x}{h_n}, \frac{y_i - y}{p_n}) - 1|
\end{aligned}$$

where C_f is a positive constant. It is not hard to see that

$$\begin{aligned}
& |\frac{1}{nh_np_n}\sum_{i=1}^{n}K^*(\frac{x_i - x}{h_n}, \frac{y_i - y}{p_n}) - 1| \\
\approx \quad & |\frac{1}{nh_np_n}\sum_{i=1}^{n}K^*(\frac{x_i - x}{h_n}, \frac{y_i - y}{p_n}) - \frac{1}{h_np_n}\sum_{i=1}^{n}\int\int_{\Delta_i}K^*(\frac{w - x}{h_n}, \frac{v - y}{p_n})\,dwdv| \\
\leq \quad & |\frac{1}{nh_np_n}\sum_{i=1}^{n}K^*(\frac{x_i - x}{h_n}, \frac{y_i - y}{p_n}) - \frac{1}{h_np_n}\sum_{i=1}^{n}\int\int_{\Delta_i}K^*(\frac{x_i - x}{h_n}, \frac{y_i - y}{p_n})\,dwdv| \\
+ \quad & |\frac{1}{h_np_n}\sum_{i=1}^{n}\int\int_{\Delta_i}[K^*(\frac{x_i - x}{h_n}, \frac{y_i - y}{p_n}) - K^*(\frac{w - x}{h_n}, \frac{v - y}{p_n})]\,dwdv| \\
\leq \quad & \frac{1}{h_np_n}\sum_{i=1}^{n}|\frac{1}{n} - S(\Delta_i)|K^*(\frac{x_i - x}{h_n}, \frac{y_i - y}{p_n}) \\
+ \quad & \frac{C_k}{h_np_n}\sum_{i=1}^{n}\int\int_{\Delta_i}\sqrt{(\frac{x_i - w}{h_n})^2 + (\frac{y_i - v}{p_n})^2}\,dwdv \\
= \quad & O(\frac{1}{n^{\lambda}h_np_n}) + O(\frac{\sqrt{h_n^2 + p_n^2}}{\sqrt{n}h_n^2p_n^2})
\end{aligned}$$

where "≈" means that a high order term has been neglected and C_k is a positive constant. Similarly, we can show that

$$\frac{C_f}{nh_np_n}\sum_{i=1}^{n}\sqrt{(x_i - x)^2 + (y_i - y)^2}K^*(\frac{x_i - x}{h_n}, \frac{y_i - y}{p_n}) = O(h_n + p_n)$$

Hence we have the conclusion of the Lemma.

A Proof of Theorem 2.1. First of all, from Lemma 3 and condition (1) on h_n and p_n, we have

$$\lim_{n\to\infty}\frac{n^{\nu}}{\beta_n\log n}||Ef_n(x,y) - f(x,y)||_{[a,b]\times[c,d]} = 0. \qquad \ldots(A.1)$$

Let

$$\begin{aligned} \bar{\varepsilon}_i &= \varepsilon_i I_{|\varepsilon_i| \le i^{1/p}}, \quad i = 1, 2, \cdots, n \\ \bar{g}_n(x,y) &= \frac{1}{nh_n p_n} \sum_{i=1}^{n} \bar{\varepsilon}_i K^*(\frac{x_i - x}{h_n}, \frac{y_i - y}{p_n}) \\ g_n^*(x,y) &= \frac{1}{nh_n p_n} \sum_{i=1}^{n} \varepsilon_i K^*(\frac{x_i - x}{h_n}, \frac{y_i - y}{p_n}). \end{aligned}$$

By using the same arguments as those in the proof of Theorem 3 in Cheng and Lin (1981), we can conclude that

$$P\left(\frac{n^\nu}{\beta_n \log n}[\bar{g}_n(x,y) - E\bar{g}_n(x,y)] > \varepsilon\right) = O\left(n^{-\varepsilon\beta_n^{1/2}} \exp(\frac{n^{2\nu}}{nh_n p_n \beta_n})\right) \qquad \ldots (A.2)$$

Define

$$A_n \triangleq \left\{(i/[n^\eta], j/[n^\zeta]), i = 1, 2, \cdots, [n^\eta], j = 1, 2, \cdots, [n^\zeta]\right\}$$

where η and ζ are two positive constants and $[x]$ denotes the integral part of x. Clearly, A_n has $[n^\eta] \cdot [n^\zeta]$ elements. For any $(x,y) \in [a,b] \times [c,d]$, there exists $(v(x), w(y)) \in A_n$ such that $|x - v(x)| \le 1/[n^\eta]$ and $|y - w(y)| \le 1/[n^\zeta]$.

Obviously,

$$\frac{n^\nu}{\beta_n \log n} ||\bar{g}_n(x,y) - E\bar{g}_n(x,y)||_{[a,b]\times[c,d]} \le S_{1n} + S_{2n} + S_{3n}$$

where

$$\begin{aligned} S_{1n} &= \frac{n^\nu}{\beta_n \log n} ||\bar{g}_n(x,y) - \bar{g}_n(v(x), w(y))||_{[a,b]\times[c,d]} \\ S_{2n} &= \frac{n^\nu}{\beta_n \log n} ||\bar{g}_n(v(x), w(y)) - E\bar{g}_n(v(x), w(y))||_{[a,b]\times[c,d]} \\ S_{3n} &= \frac{n^\nu}{\beta_n \log n} ||E\bar{g}_n(v(x), w(y)) - E\bar{g}_n(x,y)||_{[a,b]\times[c,d]} \end{aligned}$$

We can choose η and ζ large enough such that

$$\lim_{n\to\infty} S_{1n} \le \lim_{n\to\infty} \frac{n^\nu}{\beta_n \log n} \cdot \frac{1}{nh_n p_n} \cdot n \cdot n^{1/p} \sqrt{\frac{1}{[n^\eta]^2 h_n^2} + \frac{1}{[n^\zeta]^2 p_n^2}} = 0, \quad a.s.$$

By the same reason, $\lim_{n\to\infty} S_{3n} = 0$. By using (A.2) and the condition (2) on h_n and p_n, we have

$$P(S_{2n} > \varepsilon) \le Cn^{-\varepsilon\beta_n^{1/2} + \eta + \zeta}$$

where C is a constant. So $\sum_{n=1}^{\infty} P(S_{2n} > \varepsilon) < \infty$. By the Borel-Cantelli Lemma, $\lim_{n\to\infty} S_{2n} = 0$, $a.s.$ It is easy to see that

$$\begin{aligned} & ||f_n(x,y) - Ef_n(x,y)||_{[a,b]\times[c,d]} \\ = \; & ||g_n^*(x,y)||_{[a,b]\times[c,d]} \\ \le \; & ||g_n^*(x,y) - \bar{g}_n(x,y)||_{[a,b]\times[c,d]} + ||\bar{g}_n(x,y) - E\bar{g}_n(x,y)||_{[a,b]\times[c,d]} + ||E\bar{g}_n(x,y)||_{[a,b]\times[c,d]}. \end{aligned}$$

We can argue in the same way as Cheng and Lin (1981) did that there exists a sample subspace Ω_0 which satisfies $P(\Omega_0) = 1$ and for any $\omega \in \Omega_0$,

$$\begin{aligned} & \lim_{n\to\infty} \frac{n^\nu}{\beta_n \log n} ||g_n^*(x,y) - \bar{g}_n(x,y)||_{[a,b]\times[c,d]} \\ \le \; & \lim_{n\to\infty} \frac{C_\omega n^\nu}{nh_n p_n \beta_n \log n} ||K^*(x,y)||_{[-1,1]\times[-1,1]} \\ = \; & 0, \end{aligned}$$

where C_ω is a finite constant which may depend on ω. The last equation in the above expression is based on the condition (2) on h_n and p_n given in the theorem.

On the other hand,

$$\begin{aligned}
|E\bar{g}_n(x,y)| &\leq \frac{1}{nh_np_n}||K^*(x,y)||_{[-1,1]\times[-1,1]}\sum_{i=1}^{n}|\int_{|\varepsilon_i|<i^{1/p}}\varepsilon_i dP(\varepsilon_i)| \\
&= \frac{1}{nh_np_n}||K^*(x,y)||_{[-1,1]\times[-1,1]}\sum_{i=1}^{n}|\int_{|\varepsilon_i|>i^{1/p}}\varepsilon_i dP(\varepsilon_i)| \\
&\leq \frac{1}{nh_np_n}||K^*(x,y)||_{[-1,1]\times[-1,1]}\sum_{i=1}^{n} i^{-(1-p)/p}E|\varepsilon_1|^p \\
&= \frac{C'n^{1/p}}{nh_np_n}
\end{aligned}$$

where C' is a constant. By condition (3),

$$\lim_{n\to\infty}\frac{n^\nu}{\beta_n \log n}||E\bar{g}_n(x,y)||_{[a,b]\times[c,d]} \leq \lim_{n\to\infty}\frac{C'n^{\nu+1/p}}{nh_np_n\beta_n\log n} = 0$$

Consequently, we have

$$\lim_{n\to\infty}\frac{n^\nu}{\beta_n \log n}||f_n(x,y) - Ef_n(x,y)||_{[a,b]\times[c,d]} = 0, \quad a.s. \qquad \ldots(A.3)$$

Combining (A.1) and (A.3), the first conclusion of the theorem is proved.

In the following, we investigate the expression

$$\frac{n^\nu}{\beta_n \log n}||f_n(\theta,x,y) - f(x,y)||_{[-\pi/2,\pi/2]\times[a,b]\times[c,d]},$$

where a rotation parameter θ is involved. We can check that all of the above steps except that for S_{2n} will be true after we replace $K^*(x,y)$ by $K(\theta,x,y)$ and after we maximize the corresponding terms with respect to $\theta \in [-\pi/2,\pi/2]$.

Let

$$\begin{aligned}
S_{2n}^* &= \frac{n^\nu}{\beta_n \log n}||\bar{g}_n(\theta,v(x),w(y)) - E\bar{g}_n(\theta,v(x),w(y))||_{[-\pi/2,\pi/2]\times[a,b]\times[c,d]} \\
\bar{g}_n(\theta,x,y) &= \frac{1}{nh_np_n}\sum_{i=1}^{n}\bar{\varepsilon}_iK(\theta,\frac{x_i-x}{h_n},\frac{y_i-y}{p_n}) \\
B_n &= \left\{\frac{i\cdot\pi}{[n^\delta]} - \pi/2, i=1,2,\cdots,[n^\delta]\right\}
\end{aligned}$$

where $\delta > 0$ is a constant. Then for any $\theta \in [-\pi/2,\pi/2]$, there exists $U(\theta) \in B_n$ such that $|\theta - U(\theta)| \leq \pi/[n^\delta]$. It is not difficult to check that

$$||\bar{g}_n(\theta,v(x),w(y)) - E\bar{g}_n(\theta,v(x),w(y))||_{[-\pi/2,\pi/2]\times[a,b]\times[c,d]} \leq T_{1n} + T_{2n} + T_{3n}$$

where

$$\begin{aligned}
T_{1n} &= ||\bar{g}_n(\theta,v(x),w(y)) - \bar{g}_n(U(\theta),v(x),w(y))||_{[-\pi/2,\pi/2]\times[a,b]\times[c,d]} \\
T_{2n} &= ||\bar{g}_n(U(\theta),v(x),w(y)) - E\bar{g}_n(U(\theta),v(x),w(y))||_{[-\pi/2,\pi/2]\times[a,b]\times[c,d]} \\
T_{3n} &= ||E\bar{g}_n(U(\theta),v(x),w(y)) - E\bar{g}_n(\theta,v(x),w(y))||_{[-\pi/2,\pi/2]\times[a,b]\times[c,d]}.
\end{aligned}$$

According to Lemma 2, we have

$$\begin{aligned}
&|\bar{g}_n(\theta, v(x), w(y)) - \bar{g}_n(U(\theta), v(x), w(y))| \\
= \quad & \frac{1}{nh_np_n} |\sum_{i=1}^{n} \bar{\varepsilon}_i [K(\theta, \frac{x_i - v(x)}{h_n}, \frac{y_i - w(y)}{p_n}) - K(U(\theta), \frac{x_i - v(x)}{h_n}, \frac{y_i - w(y)}{p_n})]| \\
\leq \quad & \frac{C}{nh_np_n} \cdot n^{1+1/p} \cdot n^{-\delta} \cdot \frac{\sqrt{h_n^2 + p_n^2}}{h_np_n}, \quad a.s.
\end{aligned}$$

where C is a constant. So we can choose δ large enough such that

$$\lim_{n\to\infty} \frac{n^\nu}{\beta_n \log n} T_{1n} = 0, \quad a.s.$$

By the same argument, we can prove that

$$\lim_{n\to\infty} \frac{n^\nu}{\beta_n \log n} T_{3n} = 0.$$

Finally,

$$\begin{aligned}
& P(\frac{n^\nu}{\beta_n \log n} T_{2n} > \varepsilon) \\
\leq \quad & n^{\delta+\eta+\zeta} \cdot n^{-\varepsilon\beta_n^{1/2}} \exp\left\{ \frac{n^{2\nu}\sigma^2}{n^2 h_n^2 p_n^2 \beta_n} \cdot \max_{(U(\theta), v(x), w(y))} \right. \\
& \left. [\sum_{i=1}^{n} K^2(U(\theta), \frac{x_i - v(x)}{h_n}, \frac{y_i - w(y)}{p_n})] \right\} \\
= \quad & O(n^{\delta+\eta+\zeta} \cdot n^{-\varepsilon\beta_n^{1/2}}).
\end{aligned}$$

So

$$\lim_{n\to\infty} \frac{n^\nu}{\beta_n \log n} T_{2n} = 0, \quad a.s.$$

Combining the above results, the proof is completed.

ACKNOWLEDGEMENTS. The author would like to thank an anonymous referee and the Editor for their careful reading and many helpful comments and suggestions.

References

BESAG, J. (1986). On the statistical analysis of dirty pictures (with discussion), *Jour. Roy. Statist. Soc.* B **48**, 259-302.

BHANDARKAR, S.M., ZHANG, Y. and POTTER, W.D. (1994). An edge detection technique using genetic algorithm-based optimization, *Pattern Recognition*, **27**, 1159-1180.

BHATTACHARYA, P.K. (1994). Some aspects of change-point analysis, *IMS Lecture Notes*, vol.23, *Change-Point Problems* (E. Carlstein, H.G. Müller and D. Siegmund eds.), 28-56.

CANNY, J. (1986). A computational approach to edge detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**, 679-698.

CHEN, X.R. (1988). Hypothesis testing and interval estimation about the model which has only one change-point, *Scientia Sinica*, **8A**, 1-13.

CHEN, M.H., LEE, D. and PAVLIDIS, T. (1991). Residual analysis for feature detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**, 30-40.

CHENG, K.F. and LIN, P.E. (1981). Nonparametric estimation of a regression function, *Z. Wahrscheinlichkeitstheorie view. Gebiete* **57**, 223-233.

CRESSIE, N. (1991). *Statistics for Spatial Data*, New York: John Wiley.

DUAN, N.H. and LI, K.C. (1991). Slicing regression: a link-free regression model, *Ann. Statist.*, **19**, 505-530.

FRIEDMAN, J. and STUETZLE, W. (1981). Projection pursuit regression, *Jour. Amer. Statist. Assoc.*, **76**, 817-823.

GASSER, T. and MÜLLER, H.G. (1979). Kernel estimation of regression functions, In *Smoothing Techniques for Curve Estimation* (Gasser and Rosenblatt eds.), Lecture Notes in Mathematics, **757**, Springer-Verlag: New York.

HALL, P. and TITTERINGTON, M. (1992). Edge-preserving and peak-preserving smoothing, *Technometrics* **34**, 429-440.

HANSEN, F.R. and ELLIOT, H. (1982). Image segmentation using simple Markov field models, *Computer Graphics and Image Processing*, **20**, 101-132.

HÄRDLE, W. (1990). *Applied Nonparametric Regression,* Cambridge University Press, Cambridge.

– – –– (1991). *Smoothing Techniques: with implementation in S*, Springer-Verlag: New York.

HARALICK, R.M. (1984). Digital step edges from zero crossing of second directional derivatives, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 58-68.

HINKLEY, D.V. (1969), Inference about the intersection in two phase regression, *Biometrika*, **56**, 495-504.

–––– (1971). Inference in two phase regression, *Jour. Amer. Statist. Assoc.*, **66**, 736-743.

LI, K.C. (1991). Slicing inverse regression for dimension reduction, *Jour. Amer. Statist. Assoc.*, **86**, 316-327.

MARR, D. and HILDRETH, E. (1980). Theory of edge detection, *Proceedings of the Royal Society of London*, **207**, 187-217.

MCDONALD, J.A. and OWEN, A.B. (1986). Smoothing with split linear fits, *Technometrics*, **28**, 195-208.

MÜLLER, H.G. (1988). Nonparametric Regression Analysis of Longitudinal Data, *Lecture Notes in Statistics*, Springer-Verlag: New York.

– – –– (1992). Change-points in nonparametric regression analysis, *Ann. Statis.*, **20**, 737-761.

PERONA, P. and MALIK, J. (1990). Scale space and edge detection using anisotropic diffusion, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**, 629-639.

QIU, P. (1991). Estimation of a kind of jump regression functions, *Systems Science and Mathematical Sciences*, **4**, 1-13.

– – –– (1994). Estimation of the number of jumps of the jump regression functions, *Communications in Statistics-Theory and Methods*, **23**, 2141-2155.

QIU, P., ASANO, CHI. and LI, X. (1991). Estimation of jump regression functions, *Bulletin of Informatics and Cybernetics*, **24**, 197-212.

QIU, P. and BHANDARKAR, S.M. (1996). An edge detection technique using local smoothing and statistical hypothesis testing, *Pattern Recognition Letters*, **17**, 849-872.

QIU, P. and YANDELL, B. (1997). Jump detection in regression surfaces, *Journal of Computational and Graphical Statistics,* September Issue.

QUANDT, R.E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes, *Jour. Amer. Statist. Assoc.*, **53**, 873-880.

RICE, J. (1984), Boundary modification for kernel regression, *Communications in Statistics - Theory and Methods*, **13**, 893-900.

ROSENFELD, A. and KAK, A.C. (1982). *Digital Picture Processing,* Vols. 1 and 2, Academic Press, New York.

SAINT-MARC, P., CHEN, J. and MEDIONI, G. (1991). Adaptive smoothing: a general tool for early vision, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**, 514-529.

SARKAR, S.S. and BOYER, K.L. (1990). On optimal infinite impulse response edge detection filters, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**, 1154-1171.

SHABAN, S.A. (1980). Change point problem and two-phase regression: an annotated bibliography, *International Statistical Review*, **48,** 83-93.

SHIAU, J.H. (1987). A note on MSE coverage intervals in a partial spline model, *Communications in Statistics-Theory and Methods*, **16**, 1851-1866.

TAN, H.L., GELFAND, S.B. and DELP, E.J. (1989). A comparative cost function approach to edge detection, *IEEE Transaction on Systems, Man, and Cybernetics*, **19**, 1337-1349.

–––– (1991). A cost minimization approach to edge detection using simulated annealing, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**, 3-18.

WAHBA, G. (1986). Partial spline modelling of the tropopause and other discontinuities, *Function Estimate, Contemporary Mathematics 59* (J.S. Marron eds.), 125-135.

WU, J.S. and CHU, C.K. (1993). Kernel type estimators of jump points and values of a regression function, *Ann. Statis.*, **21**, 1545-1566.

YIN, Y.Q. (1988). Detecting of the number, locations and magnitudes of jumps, *Communications in Statistics-Stochastic Models*, **4**, 445-455.

BIOSTATISTICS PROGRAM
THE OHIO STATE UNIVERSITY
M200 STARLING LOVING HALL
320 WEST 10TH AVENUE
COLUMBUS, OH 43210
U.S.A.

Captions of Figures

Figure 1.1: The solid rectangle denotes the non-zero-value domains of definition of $K_1^*(x, y)$ and $K_2^*(x, y)$. The dashed rectangle is the non-zero-value domains of definition of $K_1(\theta, x, y)$ and $K_2(\theta, x, y)$. Point (w, v) is obtained by rotating point (x, y) an angle θ. $K_i(\theta, w, v) = K_i^*(x, y)$ for $i = 1$ and 2.

Figure 2.1: The small dashed rectangle denotes the non-zero-value domain of definition of $K_1(\arctan(\phi'(x_0)), \frac{x-x_0}{h_n}, \frac{y-\phi(x_0)}{p_n})$.

Figure 3.1: The jump location curve is a circuit and it can be expressed by two functions $\phi_1(x) \leq \phi_2(x)$ for $x \in [A, B]$.

Figure 4.1: (a) The true jump location curve; (b) the jump magnitude function; (c) the true regression surface; (d) a set of noisy data with $\varepsilon_1 \sim N(0, 1)$.

Figure 4.2: Some simulation results at $x = 0.2, 0.4, 0.6$ and 0.8. The diamonds represent true values which are connected by solid lines. The triangles denote estimator values when $n = 400$ which are connected by dotted lines. Small squares are the estimator values when $n = 10000$ which are connected by dashed lines. Plot (a) is about $\phi(x)$ and its estimator $\hat{\phi}(x)$; plot (b) is about $\arctan(\phi'(x))$ and its estimator $\theta^*(x, \hat{\phi}(x))$; plot (c) is about $C(x)$ and its estimator $|M_n(\theta^*(x, \hat{\phi}(x)), x, \hat{\phi}(x))|$.