University of Wisconsin — Madison

Department of Statistics

JUMP DETECTION IN REGRESSION SURFACES

Peihua Qiu

Brian Yandell

Department of Statistics

Technical Report # 948

May 1994

Department of Statistics
University of Wisconsin–Madison
1210 West Dayton Street
Madison WI 53706-1685

Phone: 608/262-2598
Fax: 608/262-0032
Internet: yandell@stat.wisc.edu

JUMP DETECTION IN REGRESSION SURFACES

Peihua Qiu and Brian Yandell

Department of Statistics
University of Wisconsin - Madison
1210 West Dayton Street
Madison, WI 53706

ABSTRACT

We consider the problem of locating the jumps in regression surfaces. A least squares plane is fitted in a neighborhood of a design point in question. The gradient vector of this plane contains both the continuous and jump information of the regression surface at this design point. We try to remove the continuous information by using two neighboring least squares planes along the direction of that gradient vector. In such a way, the jump information is extracted, suggesting a jump detection algorithm. This method requires $O(Nk)$ computations, where N is the sample size and k is the window width of the neighborhood. This property makes it possible to handle large data sets. The conditions imposed on the jump location curves, the jump surfaces and the noise are mild enough. We demonstrate this method in detail with a simulation example.

AMS 1991 subject classifications:Primary 62G05; secondary 62G20.

Key Words: Jump surfaces, Jump location curves, Least squares plane, Slopes, Jump detection criterion, Threshold value, Modification procedure, Image processing.

1 Introduction

In computer image analysis, a very important problem involves detecting the edges of objects, or equivalently, detecting the discontinuities of the underlying "intensity function" (the brightness of each point in the image is expressed by this function). In meteorology and oceanography, the equi-temperature surfaces of the high sky and the deep ocean are usually discontinuous. From a statistical point, all of these problems could be regarded as the applications of the estimation of two dimensional (2-D) jump regression surfaces (JRS). The purpose of this paper is to develop a method to detect the jump locations of the 2-D JRS.

The research on jump regression models is currently under rapid development. In the one dimensional (1-D) case, McDonald and Owen (1986) suggested a "split linear smoother" which provided a discontinuity preserving curve estimator. Hall and Titterington (1992) proposed an alternative but simpler method to detect the jumps by establishing some relations among three local linear smoothers. Chen (1988), Müller (1992), Qiu (1991,1994), Qiu et al. (1991), Wu and Chu (1993), Yin (1988), among many others, discussed various kernel-type methods. These methods were all based on the difference of two one-sided kernel smoothers. The jump regression models were related to the kernel density estimation techniques by Chu (1994). Shiau (1985, 1987) regarded the jump regression models as partial linear regression models and fitted them with partial splines. Recently, Eubank and Speckman (1993) and Speckman (1993) treated the 1-D jump regression model as a semi-parametric regression model and proposed estimates of the jump locations and magnitudes. Qiu and Yandell (1994) developed a jump detection algorithm with local polynomial regression.

Comparatively little research has been published on 2-D JRS, due to difficulties to generalize from 1-D to 2-D cases. In 1-D cases, the jump locations are at most a series of points. But in 2-D cases, the jump locations are curves and in most situations these curves can not be expressed by explicit mathematical functions. Some Russian scientists did pioneering theoretical research in this area. See, for example, Brodsky and Darkhovsky (1993), Korostelev and Tsybakov (1993) and the references cited there. Korostelev and Tsybakov (1993) investigated the jump locations detection and the jump surfaces fitting under several kinds of designs and jump boundaries. They suggested approximating the jump location curves by piecewise polynomials and then estimating the coefficients by maximum likelihood estimation. Such method was proved to reach the "minimax

optimal rate of convergence", but its computation complexity is not known. Müller and Song (1994) proposed "maximin" estimators of the jump boundaries of the d-dimensional ($d \geq 1$) jump surfaces under the condition that the number of such jump boundary curves (surfaces) is known and the population of the possible jump boundary curves (surfaces) is also known. Qiu (1992) suggested a so-called *Rotational Difference Kernel Estimator (RDKE)* of the jump location curves of the JRS. Both of the above two methods were based on two one-sided kernel smoothers along a direction and the estimators were obtained by maximizing the jump detection criteria with respect to this direction. This makes the computation very expensive. When the jump location curves are completely known, Shiau(1985) (see Shiau, Wahba and Johnson (1986)) fit the 2-D JRS by the partial spline method.

In this paper we make another attempt to detect the jump locations of the JRS. Our method is based on the local simple least squares (LS) fitting. At a point in question, a LS plane is fitted in a neighborhood of this point. The LS coefficients of this plane give an approximation of the gradient direction of the JRS at this point. They carry both the continuous and the jump information about the JRS. We then try to delete the continuous information from the LS coefficients by considering two small neighborhoods along the approximated gradient direction, on either side of the original neighborhood. In such a way, the jump information is extracted. Based on that a jump detection criterion is derived. This method avoids using the maximization with respect to the direction at each design point. This saves a lot of computation and makes it possible to handle very large data sets. The computation of the jump detection criterion can be updated easily from one point to the next. The whole algorithm requires $O(Nk)$ calculations, with N the sample size and k the window width of the neighborhoods. The conditions needed on the jump location curves are mild enough. We do not assume any distributions on the noise. These will be stated in the next section.

The rest of the paper is organized as follows. In the next section we describe the model and the jump detection method. A numerical example is discussed in section 3. In section 4, we give some concluding remarks. The proof of a theorem is given in the Appendix A.

2 Jump Detection Algorithm

Observations $\{z_{ij}\}$ come from the following model

$$z_{ij} = f(x_i, y_j) + \epsilon_{ij},\ i, j = 1, 2, \cdots, n, \tag{2.1}$$

where $\{(x_i, y_j) = (i/n, j/n), i, j = 1, 2, \cdots, n\}$ are equally spaced design points in $[0,1] \times [0,1]$, $\{\epsilon_{ij}\}$ are iid random numbers with mean 0 and variance σ^2. The sample size is $N = n^2$. The regression function $f(x, y)$ is continuous over $[0,1] \times [0,1]$ except on some curves, which are called the jump location curves (JLCs) hereafter. In the simplest case that $f(x, y)$ has a unique JLC which divides $[0,1] \times [0,1]$ into 2 connected regions Ω_1 and Ω_2 with $\Omega_1 \bigcap \Omega_2 = \emptyset$ and $\Omega_1 \bigcup \Omega_2 = [0,1] \times [0,1]$, $f(x, y)$ can be expressed as

$$f(x, y) = g(x, y) + C(x, y) I_{(x,y) \in \Omega_1},\ (x, y) \in [0,1] \times [0,1], \tag{2.2}$$

where $g(x, y)$ is continuous over $[0,1] \times [0,1]$, $C(x, y)$ is continuous over Ω_1, $\partial\Omega_1 \bigcap \partial\Omega_2$ is the JLC with $\partial\Omega_i$ denoting the boundary of the region $\Omega_i, i = 1, 2$.

The regression function $f(x, y)$ considered in this paper is not restricted to (2.2). But it has the same fashion that it is continuous over some regions and discontinuous on their boundaries (namely, JLCs). The following assumptions (ASs) are imposed on the JLCs.

(AS1) At any point (x, y) on the JLCs, there exist two orthogonal lines which cross at (x, y) such that two vertical quadrants formed by these two lines belong to two different regions in a small neighborhood (c.f. Figure 2.1).

(AS2) In a band with width $3\sqrt{2}k/n$ and centered at any JLC, there is no other JLC, where k is a window width introduced below.

Remark 2.1 (AS1) is a restriction on the curvature of the JLC. Theoretically speaking, it is enough for our method to work if there exists a uniform angle $\phi \in (0, \pi/4]$ (c.f. Figure 2.1) at all points on the JLCs such that the stated property in (AS1) holds. We choose $\phi = \pi/4$ for simplicity.

Remark 2.2 (AS2) requires that JLCs are not very close to each other. Alternatively, it requires that there is enough data (n large, k/n small) to distinguish nearby JLCs. The number $3\sqrt{2}k/n$ will be explained later.

Remark 2.3 (AS1)-(AS2) exclude the case that the JLCs may cross each other. Theoretically, our method may fail only in the neighborhoods of such cross points. If we have enough data (n large)

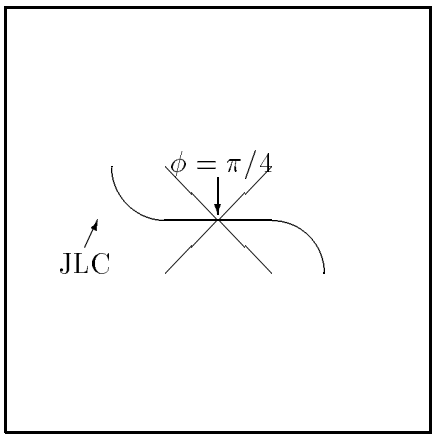

Figure 2.1: At any point on the JLCs, there exist two orthogonal lines crossing at this point and two vertical quadrants formed by these two lines belong to two different regions in a small neighborhood.

and k is relatively small, this would not affect much of the jump detection.

At any design point $(x_i, y_j), \ell + 1 \leq i, j \leq n - \ell$, we consider its neighborhood $N(x_i, y_j)$ with window width $k = 2\ell + 1 \ll n$, where ℓ is a non-negative integer.

$$N(x_i, y_j) := \{(x_{i+s}, y_{j+t}), s, t = -\ell, -\ell + 1, \cdots, 0, \cdots, \ell - 1, \ell\}.$$

A least squares plane is fitted in this neighborhood.

$$\hat{z}_{ij}(x, y) = \hat{\beta}_0^{(i,j)} + \hat{\beta}_1^{(i,j)}(x - x_i) + \hat{\beta}_2^{(i,j)}(y - y_j),\ (x, y) \in N(x_i, y_j).$$

After some calculations, we have

$$\begin{aligned} \hat{\beta}_0^{(i,j)} &= \frac{1}{k^2} z_{\cdot\cdot} \\ \hat{\beta}_1^{(i,j)} &= \frac{1}{k S_x^2} \sum_{s=-\ell}^{\ell} (x_{i+s} - x_i) z_{i+s,\cdot} \\ \hat{\beta}_2^{(i,j)} &= \frac{1}{k S_y^2} \sum_{t=-\ell}^{\ell} (y_{j+t} - y_j) z_{\cdot,j+t} \end{aligned} \tag{2.3}$$

where $z_{\cdot\cdot} = \sum_{s,t=-\ell}^{\ell} z_{i+s,j+t}$, $z_{i+s,\cdot} = \sum_{t=-\ell}^{\ell} z_{i+s,j+t}$, $z_{\cdot,j+t} = \sum_{s=-\ell}^{\ell} z_{i+s,j+t}, S_x^2 = \sum_{s=-\ell}^{\ell} (x_{i+s} - x_i)^2, S_y^2 = \sum_{t=-\ell}^{\ell} (y_{j+t} - y_j)^2$. It is not hard to check that $\hat{\beta}_0^{(i,j)}, \hat{\beta}_1^{(i,j)}$ and $\hat{\beta}_2^{(i,j)}$ are uncorrelated of each other. Furthermore, they have the following property.

Theorem 2.1 In model (2.1), suppose that $f(x,y)$ has the first order partial derivatives at any points $(x_1^*, y_1^*), (x_2^*, y_2^*) \in [0,1]\times[0,1]$ which are not on the JLCs and $f_x'(x_1^*, y_1^*) - f_x'(x_2^*, y_2^*) \to 0$ if the distance between these two points tends to 0. $f_y'(x,y)$ is supposed to have the same property. The JLCs satisfy assumptions (AS1)-(AS2). The window width k satisfies the conditions that $\lim_{n\to\infty}\frac{k}{n} = 0$ and $\lim_{n\to\infty}\frac{k^2}{n} = \infty$. If there is no jump in $N(x_i, y_j)$, then $\hat\beta_1^{(i,j)} \sim f_x'(x_i, y_j)$ and $\hat\beta_2^{(i,j)} \sim f_y'(x_i, y_j)$, with "$\sim$" meaning the difference tends to zero with probability 1. If (x_i, y_j) is on a JLC, then $\hat\beta_1^{(i,j)} \sim f_x'(\tilde x_i, \tilde y_j) + h_1^{(i,j)}C(i,j)$ and $\hat\beta_2^{(i,j)} \sim f_y'(\tilde x_i, \tilde y_j) + h_2^{(i,j)}C(i,j)$, where $(\tilde x_i, \tilde y_j)$ is any point around (x_i, y_j) which neighborhood does not include any jumps and the distance between these two points tends to zero, $C(i,j)$ is the jump magnitude, $h_1^{(i,j)}$ and $h_2^{(i,j)}$ are two constants satisfying $\sqrt{(h_1^{(i,j)})^2 + (h_2^{(i,j)})^2} = O(n/k)$.

(proof is given in the Appendix A.)

From Theorem 2.1, the slopes $\hat\beta_1^{(i,j)}$ and $\hat\beta_2^{(i,j)}$ carry both the continuous and the jump information of the JRS. We try to extract the jump information in the following way. Along the direction of $\vec v_{ij} := (\hat\beta_1^{(i,j)}, \hat\beta_2^{(i,j)})$, two neighboring design points (x_{N1}, y_{N1}) and (x_{N2}, y_{N2}) are determined by the following formulas.

If $\pi/4 \le \theta < 3\pi/4$ or $5\pi/4 \le \theta < 7\pi/4$, then

$$\begin{aligned} x_{N1} &= x_i + \frac{k}{n\cdot\tan\theta}, \quad y_{N1} = y_j + \tfrac{k}{n} \\ x_{N2} &= x_i - \frac{k}{n\cdot\tan\theta}, \quad y_{N2} = y_j - \tfrac{k}{n}; \end{aligned} \tag{2.4}$$

If $-\pi/4 \le \theta < \pi/4$ or $3\pi/4 \le \theta < 5\pi/4$, then

$$\begin{aligned} x_{N1} &= x_i + \frac{k}{n}, \quad y_{N1} = y_j + \tfrac{k}{n}\cdot\tan\theta \\ x_{N2} &= x_i - \frac{k}{n}, \quad y_{N2} = y_j - \tfrac{k}{n}\cdot\tan\theta \end{aligned} \tag{2.5}$$

where $\theta \in [-\pi/4, 7\pi/4]$ is the angle formed by $\vec v_{ij}$ and the positive direction of x-axis. If the two points determined by (2.4)-(2.5) are not exactly the grid points, we just choose two grid points which are closest to them instead.

(x_{N1}, y_{N1}) and (x_{N2}, y_{N2}) have the following properties: (1) they are two design points on the line through (x_i, y_j) and with slope $\hat\beta_2^{(i,j)}/\hat\beta_1^{(i,j)}$; (2) they are closest to (x_i, y_j) among the points on that line which neighborhoods have no overlap with $N(x_i, y_j)$. Notice that $(\hat\beta_1^{(i,j)}, \hat\beta_2^{(i,j)})$

is the gradient vector of the fitted LS plane. The underlying JRS increases (or decreases) most rapidly along a near-by direction. If (x_i, y_j) is on a JLC, then it can be checked that when n is large enough (1) assumption (AS1) guarantees that the JLC which (x_i, y_j) is on could not be in $N(x_{N1}, y_{N1})$ and $N(x_{N2}, y_{N2})$; (2) (AS2) guarantees that no other JLCs are possible in these two neighborhoods. (The number $3\sqrt{2}k/n$ in (AS2) and Remark 2.2 is three times the diagonal width of the neighborhood.) So (x_{N1}, y_{N1}) and (x_{N2}, y_{N2}) are on two different sides of the JLC. We then define the following jump detection criterion δ_{ij}.

$$\delta_{ij} := min\{\|(\hat{\beta}_1^{(i,j)} - \hat{\beta}_1^{(N1)}, \hat{\beta}_2^{(i,j)} - \hat{\beta}_2^{(N1)})\|, \|(\hat{\beta}_1^{(i,j)} - \hat{\beta}_1^{(N2)}, \hat{\beta}_2^{(i,j)} - \hat{\beta}_2^{(N2)})\|\} \tag{2.6}$$

where $\vec{v}_{N1} := (\hat{\beta}_1^{(N1)}, \hat{\beta}_2^{(N1)})$ and $\vec{v}_{N2} := (\hat{\beta}_1^{(N2)}, \hat{\beta}_2^{(N2)})$ are gradient vectors of the fitted LS planes at (x_{N1}, y_{N1}) and (x_{N2}, y_{N2}) respectively and $\|(x, y)\| := \sqrt{x^2 + y^2}$, which is the length of the vector (x, y).

Figure 2.2 gives an intuitive explanation of (2.6). δ_{ij} is the minimum length of vectors $\vec{v}_{ij} - \vec{v}_{N1}$ and $\vec{v}_{ij} - \vec{v}_{N2}$. If there is no jump in these three neighborhoods, then $\vec{v}_{ij}, \vec{v}_{N1}$ and $\vec{v}_{N2}$ should be close to each other. Hence δ_{ij} is close to zero. If (x_i, y_j) is on a JLC, however, δ_{ij} is close to $\sqrt{(h_1^{(i,j)})^2 + (h_2^{(i,j)})^2}C(i,j)$, by Theorem 2.1. So δ_{ij} excludes the continuous information and keeps the jump information at the same time. Theorem 2.1 says that $\sqrt{(h_1^{(i,j)})^2 + (h_2^{(i,j)})^2}C(i,j) = O(n/k)C(i,j)$ which tends to infinity when n increases. Hence δ_{ij} could be used to detect the jumps.

An equivalent and maybe more intuitive explanation of δ_{ij} is as follows. Let $\vec{u}_{ij} = (\hat{\beta}_1^{(i,j)}, \hat{\beta}_2^{(i,j)}, -1)$, $\vec{u}_{N1} = (\hat{\beta}_1^{(N1)}, \hat{\beta}_2^{(N1)}, -1)$ and $\vec{u}_{N2} = (\hat{\beta}_1^{(N2)}, \hat{\beta}_2^{(N2)}, -1)$ be the normal directions of the LS planes at (x_i, y_j), (x_{N1}, y_{N1}) and (x_{N2}, y_{N2}) respectively. δ_{ij} is the minimum length of vectors $\vec{u}_{ij} - \vec{u}_{N1}$ and $\vec{u}_{ij} - \vec{u}_{N2}$. If (x_i, y_j) is on a JLC, then $\vec{u}_{ij}$ has an abrupt change from both $\vec{u}_{N1}$ and $\vec{u}_{N2}$. Hence δ_{ij} is relatively large.

A large value of δ_{ij} indicates a possible jump at (x_i, y_j). For any constant $b > 0$,

$$\begin{aligned} P(\delta_{ij} > b) &\leq P(\|(\hat{\beta}_1^{(i,j)} - \hat{\beta}_1^{(N1)}, \hat{\beta}_2^{(i,j)} - \hat{\beta}_2^{(N1)})\| > b) \\ &= P((\hat{\beta}_1^{(i,j)} - \hat{\beta}_1^{(N1)})^2 + (\hat{\beta}_2^{(i,j)} - \hat{\beta}_2^{(N1)})^2 > b^2) \\ &= E\{P((\hat{\beta}_1^{(i,j)} - \hat{\beta}_1^{(N1)})^2 + (\hat{\beta}_2^{(i,j)} - \hat{\beta}_2^{(N1)})^2 > b^2 | \hat{\beta}_1^{(i,j)}, \hat{\beta}_2^{(i,j)})\}. \end{aligned}$$

For fixed $\hat{\beta}_1^{(i,j)}$ and $\hat{\beta}_2^{(i,j)}$, $((\hat{\beta}_1^{(i,j)} - \hat{\beta}_1^{(N1)})^2 + (\hat{\beta}_2^{(i,j)} - \hat{\beta}_2^{(N1)})^2)/\sigma_{N1}^2$ is approximately χ_2^2 distributed under the assumption that there is no jump in $N(x_i, y_j) \bigcup N(x_{N1}, y_{N1})$. Here $\sigma_{N1}^2 = var(\hat{\beta}_1^{(N1)}) =$

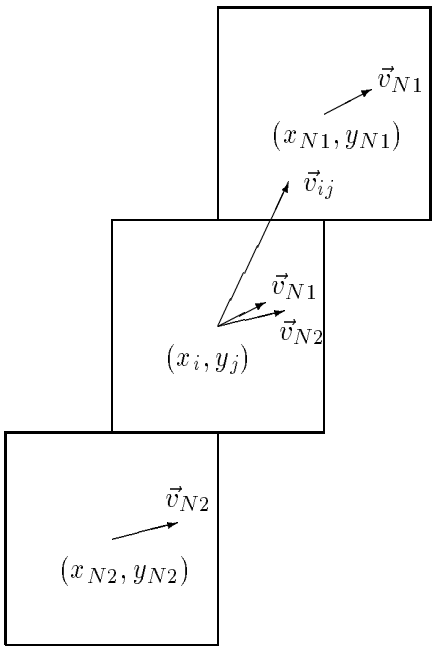

Figure 2.2: At design point (x_i, y_j), the jump detection criterion δ_{ij} is defined as the minimum length of the vectors $\vec{v}_{ij} - \vec{v}_{N1}$ and $\vec{v}_{ij} - \vec{v}_{N2}$, where $\vec{v}_{ij}, \vec{v}_{N1}$ and $\vec{v}_{N2}$ are the gradient vectors of the LS planes at (x_i, y_j) and two neighboring design points (x_{N1}, y_{N1}) and (x_{N2}, y_{N2}) which are defined in (2.4)-(2.5).

$\frac{\sigma^2}{kS_x^2}$. Therefore a natural threshold value of δ_{ij} is

$$b = \sqrt{\chi^2_{2,\alpha} \cdot \frac{\hat{\sigma}^2}{kS_x^2}} = \sqrt{\frac{\chi^2_{2,\alpha}}{kS_x^2}} \cdot \hat{\sigma} \tag{2.7}$$

where $\chi^2_{2,\alpha}$ is a $1-\alpha$ quantile of the χ^2_2 distribution and $\hat{\sigma}$ is a consistent estimator of σ.

The design points $\{(x_i, y_j) : \delta_{ij} > b, i, j = (3k+1)/2, \cdots, n-(3k-1)/2\}$ could be flagged as jump candidates. But two kinds of deceptive jump candidates may also be flagged. One kind is those close to the real jumps. If (x_i, y_j) is flagged, then its neighboring design points will be flagged with high probability. This kind of deceptive candidates could make the detected jump boundaries thick. Another kind of deceptive candidates are from the randomness of the jump detection criterion. (x_i, y_j) could be flagged with probability α even if it is not the jump point. This kind of deceptive candidates scatter in the whole design space. Therefore we need some modification procedures (MPs) to delete these two kinds of candidates. Here are two possible strategies.

(MP1) (to make the detected jump boundaries thin) For each $(3k+1)/2 \le j \le n-(3k-1)/2$, consider the design points $\{(x_i, y_j) : (3k+1)/2 \le i \le n-(3k-1)/2\}$ on line $y = y_j$. Among

them, the flagged candidates which satisfy $-\pi/4 \le \theta < \pi/4$ or $3\pi/4 \le \theta < 5\pi/4$ are denoted as $\{(x_{i_s}, y_j) : 1 \le s \le n_1\}$ where θ is defined as in (2.4)-(2.5). If there are $r_1 < r_2$ such that the increments of the sequence $i_{r_1} < i_{r_1+1} < \cdots < i_{r_2}$ are all less than the window width k, but $i_{r_1} - i_{r_1-1} > k$ and $i_{r_2+1} - i_{r_2} > k$, then we say that $\{(x_{i_s}, y_j) : r_1 \le s \le r_2\}$ forms a *tie* and we select the middle point $((x_{i_{r_1}} + x_{i_{r_2}})/2, y_j)$ as a new candidate to replace the tie set. That is, reduce the candidate set to one representative, the middle point, from each tie set. Along y direction, for each $(3k+1)/2 \le i \le n-(3k-1)/2$, consider the design points $\{(x_i, y_j) : (3k+1)/2 \le j \le n-(3k-1)/2\}$ on line $x = x_i$. We do the same modification as that along x direction except that only those candidates satisfying $\pi/4 \le \theta < 3\pi/4$ or $5\pi/4 \le \theta < 7\pi/4$ are considered this time.

(MP2) (to delete scattered candidates) For any candidate (x_i, y_j), if the number of other jump candidates in $N(x_i, y_j)$ is less than $(k-1)/2$, then we delete (x_i, y_j) from the candidates set.

Remark 2.4 In (MP1), along x direction, the condition $-\pi/4 \le \theta < \pi/4$ or $3\pi/4 \le \theta < 5\pi/4$ implies that the possible JLC forms a less than or equal to $\pi/4$ acute angle with y axis at the candidate point. This avoids cancelling the real jump boundaries that parallel to the x axis.

Remark 2.5 We want to point out that the modification procedures presented above are only two possible ones. Better MPs could be worked out after some more careful consideration. The detected JLCs may also be fragmentary here or there. (This can be seen from the numerical examples in the next section.) Some MPs are necessary to connect the detected JLCs. But this could be a very complicated problem. For more discussions on the MPs, see chapter 6 of Gonzalez and Woods (1992).

We summarize the jump detection method in the following algorithm.

The Jump Detection Algorithm

1. *At any (x_i, y_j) with $\ell+1 \le i, j \le n-\ell$, fit a LS plane in $N(x_i, y_j)$ and calculate its coefficients by formula (2.3).*

2. *Use (2.4)-(2.5) to determine two neighboring design points of $(x_i, y_j), (3k+1)/2 \le i, j \le n-(3k-1)/2$.*

3. *Use formula (2.6) to calculate δ_{ij}.*

4. *Use formula (2.7) to determine the threshold value b.*

5. *Flag the design point (x_i, y_j) as a jump candidate if it satisfies $\delta_{ij} > b$.*

6. *Use modification procedures (MP1) and (MP2) to determine the final candidates.*

Remark 2.6 We use square-shaped neighborhoods of the design points in the method only for convenience of the expressions and calculations. Other kinds of neighborhoods such as ellipses could also be used.

Remark 2.7 The LS planes fitted in the above algorithm can be updated easily from one point to the next since only a few points change. Therefore the algorithm needs $O(Nk)$ calculations. It will simplify the computer program further if we notice the symmetric structures between $\hat{\beta}_1^{(i,j)}$ and $\hat{\beta}_2^{(i,j)}$ (c.f. (2.3)).

Theorem 2.2 If we make the threshold value a little bit larger in the sense that the new threshold value b^* satisfies $b^* = o(n/k)$ and $\lim_{n\to\infty} b^*/b = \infty$, then the detected jumps are a.s. consistent in the Hausdorff distance and the convergence rate could be $O(n^{-1+\epsilon})$, for $\forall\, 0 < \epsilon < 1$.

The proof is based on the following facts. Firstly, from Theorem 2.1, we know that the jump information in the jump detection criterion δ_{ij} is of order $O(n/k)$. Secondly, the threshold value b in (2.7) is of order $O(n/k^2)$. That is also the order of the standard deviation of δ_{ij}. So the jump information dominates the randomness in the jump detection criterion as long as k tends to infinity with n. An outline of the proof is: (1) no jumps could be detected outside the bands centered at the real jump boundaries and with width $\sqrt{2}k/n$ when n is large enough; (2) if the design point (x_i, y_j) is on a JLC, then it could be detected since $\delta_{ij} = O(n/k) > b^*$ with large enough n. The convergence rate of the detected jumps is $O(k/n)$. If $k = O(n^\epsilon)$, for any $0 < \epsilon < 1$, then this rate is $O(n^{-1+\epsilon})$.

3 A Simulation Example

In this section we do some simulations with an artificial example. The regression function $f(x, y)$ has the expression

$$f(x, y) = -0.5 - y + 3(x - 0.5)^2 + I_{\{y > -(x-0.5)^2 + 0.5\}}.$$

There is one JLC $y = -(x - 0.5)^2 + 0.5$ with constant jump magnitude 1. The regression function and the JLC are plotted in Figure 3.1.

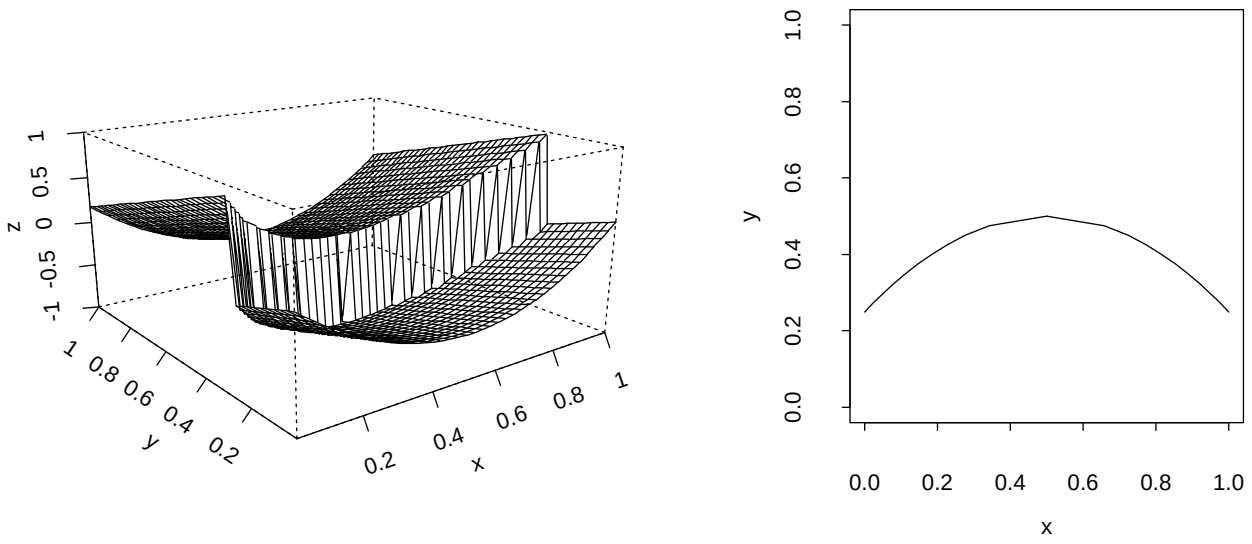

Figure 3.1: (a) The jump regression surface used in the example; (b) the jump location curve.

10000 observations $\{z_{ij}, i, j = 1, 2, \cdots, 100\}$ are generated from $z_{ij} = f(i/n, j/n) + \epsilon_{ij}$ with $n = 100$ and i.i.d. random numbers from $N(0, 0.5^2)$. One realization of the observations are shown in Figure 3.2. We can see that they are noisy enough.

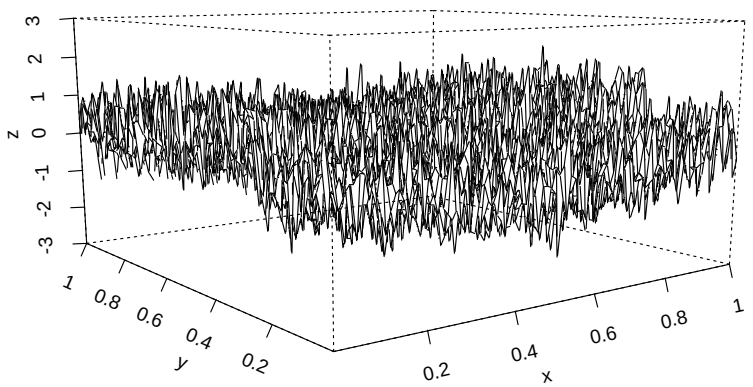

Figure 3.2: One realization of 10000 observations $\{z_{ij}, i, j = 1, 2, \cdots, 100\}$ from $f(i/100, j/100) + \epsilon_{ij}$ with ϵ_{ij} from $N(0, 0.5^2)$.

We then use formulas (2.3)-(2.6) to calculate the jump detection criterion $\{\delta_{ij}\}$, initially with

$k = 7$. $\{\delta_{ij}\}$ is plotted in Figure 3.3(a) (3-D plot) and Figure 3.3(b) (a corresponding image plot). In the image plot, the brightness at each point represents the response value. The darker, the bigger. We use the default legend of the "`image()`" function in `S-plus` to generate the image plots used in this section.

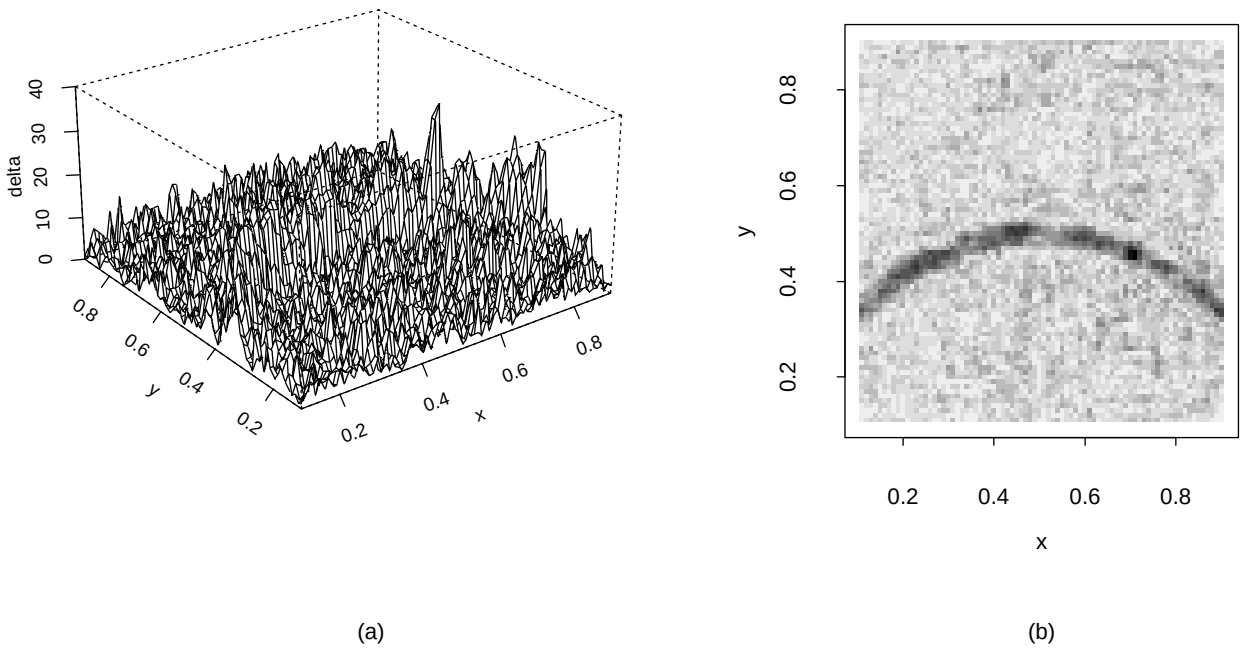

Figure 3.3: (a) The jump detection criterion $\{\delta_{ij}\}$ in 3-D plot; (b) the corresponding image plot of (a). The darker the color in the image, the bigger the criterion value.

Then a threshold is calculated by formula (2.7) with $\alpha = 0.001$ which is the smallest number in most χ^2 tables. The flagged jump candidates are plotted in Figure 3.4(b) by black points. As we mentioned in section 2, the detected jump boundary is quite thick and there are some scattered candidates also. We then use the modification procedure (MP1) and then (MP2) to modify the candidates set. The results are plotted in Figure 3.4(c) and (d). As a comparison, we plot the real JLC in Figure 3.4(a). We notice that there are some breaks here or there in the detected jump boundary in Figure 3.4(d). The detected boundary is not thin enough at some places. These imply that there is some room for our MPs to be improved as we mentioned in Remark 2.5.

The above experiment is then repeated 1000 times. The number of times of each design point to be in the final candidates set is plotted in Figure 3.5(a) and (b) with a 3-D plot and a corresponding image plot. We can see that the results are quite impressive.

Theoretically, we can use the Hausdorff distance to measure the performance of our algorithm. In reality, this distance could be very hard to compute. In the following, we use the average orthogonal distance of the points in the final candidates set to the real JLCs as a performance measurement. This measurement is averaged again for 1000 replications. The results for several n, k and σ^2 values are presented in Figure 3.6. From the plots, we could see that the averaged performance measurement (APM) decreases when n increases for each σ^2 value. This may reflect the consistency of the algorithm. For fixed σ^2 value, the best k (k with smallest APM) does not appear to change with n. That verifies the conclusion in Theorem 2.2 that k should be quite stable when n increases, to achieve the biggest accuracy of the detected jumps. The best k increases with σ^2 value. That implies that for noisier data more observations are needed in each window to reduce the randomness in the jump detection criterion. We also notice that APM is much smaller for smaller σ^2 value. These are intuitively reasonable.

4 Some Concluding Remarks

We presented a jump boundary detection algorithm with local LS plane fitting which is intuitively appealing and simple to use. It can handle relatively large data sets. Simulations show that it works well in practice.

We leave some parameters such as the window width used in the algorithm to be adjustable to the users. Much future research is needed to provide some guidelines on the selection of these parameters. As we mentioned, the modification procedures presented in the paper are only the possible ones. More careful modification procedures are needed to make the detected jump candidates match the real jump boundaries better. Another very important issue is the relationship between jump locations detection and jump surfaces fitting. If we put more structures on the jump locations, then the jump surfaces fitting could be easier. But some real applications are also excluded. It may be important to work out some methods to fit the jump surfaces under mild conditions on the jump locations.

We discussed the jumps in the regression functions in this paper. In some situations jumps in derivatives are also interesting. (The so-called "roof-edges" in image processing correspond to the jumps in the first order derivatives of the regression functions.) We think that the coefficients of

the fitted local polynomials of order $k+1$ contain useful information about the jumps in the k-th derivatives of the underlying regression functions. This kind of relationships need be investigated further. Generalization from 2-D to general d-dimensional cases seems to be straight forward theoretically. But it may not be easy to make the algorithm applicable in high dimensional cases. How to apply some dimension reduction techniques to the jump regression models is another future research topic.

Acknowledgements We thank Professors Charles Dyer, Michael Kosorok, Jun Shao and Grace Wahba for their very helpful comments and suggestions.

APPENDIX

A Proof Of Theorem 2.1

If there is no jump in $N(x_i, y_j)$, then from (2.3),

$$\begin{aligned}\hat{\beta}_1^{(i,j)} &= \frac{1}{kS_x^2}\sum_{s=-\ell}^{\ell}(x_{i+s}-x_i)z_{i+s,\cdot} \\ &= \frac{1}{kS_x^2}\sum_{s=-\ell}^{\ell}(x_{i+s}-x_i)\sum_{t=-\ell}^{\ell}[f(x_{i+s},y_{j+t})+\epsilon_{i+s,j+t}] \qquad (A.1)\end{aligned}$$

According to Lai, Robbins and Wei (1979), under the condition that $\lim_{n\to\infty}\frac{1}{kS_x^2}=0$ (or equivalently $\lim_{n\to\infty}\frac{k^2}{n}=\infty$),

$$\lim_{n\to\infty}\frac{1}{kS_x^2}\sum_{s=-\ell}^{\ell}(x_{i+s}-x_i)\sum_{t=-\ell}^{\ell}\epsilon_{i+s,j+t}=0,\ a.s. \qquad (A.2)$$

By Taylor's expansion,

$$\begin{aligned}f(x_{i+s},y_{j+t}) &= f(x_i,y_j)+f'_x(x_i+\eta(x_{i+s}-x_i),y_j+\eta(y_{j+t}-y_j))(x_{i+s}-x_i)+ \\ &\quad f'_y(x_i+\eta(x_{i+s}-x_i),y_j+\eta(y_{j+t}-y_j))(y_{j+t}-y_j)\end{aligned}$$

where $0<\eta<1$ is a constant. So

$$\begin{aligned}&\frac{1}{kS_x^2}\sum_{s=-\ell}^{\ell}(x_{i+s}-x_i)\sum_{t=-\ell}^{\ell}f(x_{i+s},y_{j+t}) \\ =\ & f'_x(x_i,y_j)+\frac{1}{k}\sum_{t=-\ell}^{\ell}[f'_x(x_i+\eta(x_{i+s}-x_i),y_j+\eta(y_{j+t}-y_j))-f'_x(x_i,y_j)]+ \\ & \frac{1}{kS_x^2}\sum_{s=-\ell}^{\ell}(x_{i+s}-x_i)\sum_{t=-\ell}^{\ell}[f'_y(x_i+\eta(x_{i+s}-x_i),y_j+\eta(y_{j+t}-y_j))-f'_y(x_i,y_j)](y_{j+t}-y_j) \\ \sim\ & f'_x(x_i,y_j) \qquad (A.3)\end{aligned}$$

Combining (A.1)-(A.3), we have

$$\lim_{n\to\infty}\hat{\beta}_1^{(i,j)}=f'_x(x_i,y_j),\ a.s.$$

Similarly,

$$\lim_{n\to\infty}\hat{\beta}_2^{(i,j)}=f'_y(x_i,y_j),\ a.s.$$

Now, if (x_i, y_j) is on a JLC and $N(x_i, y_j)$ is divided into two parts I_1 and I_2 by this JLC (without loss of generality, let us assume that there is a positive jump from I_1 to I_2 at (x_i, y_j)), then from (A.1)-(A.2)

$$\begin{aligned}
\hat{\beta}_1^{(i,j)} &\sim \frac{1}{kS_x^2} \sum_{s=-\ell}^{\ell} (x_{i+s} - x_i) \sum_{t=-\ell}^{\ell} f(x_{i+s}, y_{j+t}), \quad a.s. \\
&= \frac{1}{kS_x^2} \sum_{s=-\ell}^{\ell} (x_{i+s} - x_i) \sum_{t=-\ell}^{\ell} f^*(x_{i+s}, y_{j+t}) + h_1^{(i,j)} \\
&\sim f'_x(\tilde{x}_i, \tilde{y}_j) + h_1^{(i,j)}
\end{aligned}$$

where $(\tilde{x}_i, \tilde{y}_j)$ is some design point around (x_i, y_j) which neighborhood does not include any jumps and the distance between (x_i, y_j) and $(\tilde{x}_i, \tilde{y}_j)$ tends to zero,

$$f^*(x_{i+s}, y_{j+t}) = \begin{cases} f(x_{i+s}, y_{j+t}) & \text{if } (x_{i+s}, y_{j+t}) \in I_1 \\ f(x_{i+s}, y_{j+t}) - C(i,j) & \text{if } (x_{i+s}, y_{j+t}) \in I_2, \end{cases}$$

and

$$h_1^{(i,j)} = \frac{C(i,j)}{kS_x^2} \sum_{(x_{i+s}, y_{j+t}) \in I_2} (x_{i+s} - x_i).$$

Similarly,

$$\hat{\beta}_2^{(i,j)} \sim f'_y(\tilde{x}_i, \tilde{y}_j) + h_2^{(i,j)}, \ a.s.$$

with

$$h_2^{(i,j)} = \frac{C(i,j)}{kS_y^2} \sum_{(x_{i+s}, y_{j+t}) \in I_2} (y_{j+t} - y_j).$$

Next, we notice that (AS1) guarantees the existance of the left and the right derivatives of the JLCs at any point on them. To make this more precise, let us suppose that a JLC has a parametric form $(x(t), y(t)), t \in [T_1, T_2]$. Then it has right and left derivatives at $t_0 \in (T_1, T_2)$, if $\lim_{t \to t_0 + 0} x(t)$, $\lim_{t \to t_0 - 0} x(t)$, $\lim_{t \to t_0 + 0} y(t)$ and $\lim_{t \to t_0 - 0} y(t)$ all exist. Based on this fact, we prove that

$$\sqrt{(h_1^{(i,j)})^2 + (h_2^{(i,j)})^2} = O(n/k) \tag{A.4}$$

by considering the following two cases.

Case 1 The left and right tangent half lines of the JLC at (x_i, y_j) are both under the line $x = x_i$ in $N(x_i, y_j)$ (c.f. Figure A.1). In this case, the assumption (AS1) says that we can find two orthogonal lines crossing each other at (x_i, y_j) and two vertical quadrants formed by them belong to different

parts I_1 and I_2. For convenience, let us assume that the lower quadrant Q belongs to I_2. Clearly, when n is large enough (the difference between the tangent half lines and the real JLC could be neglected in the small neighborhood),

$$\begin{aligned} |h_2^{(i,j)}| &> |\frac{C(i,j)}{kS_y^2} \sum_{(x_{i+s},y_{j+t})\in Q} (y_{j+t} - y_j)| \\ &> |\frac{C(i,j)}{kS_y^2} \sum_{s=0}^{\ell} \sum_{t=-\ell}^{0} (y_{j+t} - y_j)| \\ &= O(n/k) \end{aligned} \tag{A.5}$$

On the other hand,

$$max\{|h_1^{(i,j)}|, |h_2^{(i,j)}|\} \leq 2 \cdot |\frac{C(i,j)}{kS_y^2} \sum_{s=0}^{\ell} \sum_{t=-\ell}^{0} (y_{j+t} - y_j)|$$

Hence (A.4) is true.

Case 2 The left and right tangent half lines of the JLC at (x_i, y_j) are in two vertical quadrants in $N(x_i, y_j)$ (c.f. Figure A.2). Without loss of generality, we assume that one is in the upper-right quadrant and the other is in the lower-left quadrant. The former forms a region J_1 with line $x = x_i$ and the latter forms a region J_2 with line $y = y_j$. If

$$\sum_{(x_{i+s},y_{j+t})\in J_2} |y_{j+t} - y_j| \geq \sum_{(x_{i+s},y_{j+t})\in J_1} |y_{j+t} - y_j|,$$

then (A.5) and consequently (A.4) is true. If

$$\sum_{(x_{i+s},y_{j+t})\in J_2} |y_{j+t} - y_j| < \sum_{(x_{i+s},y_{j+t})\in J_1} |y_{j+t} - y_j|,$$

then by the facts that

$$\sum_{(x_{i+s},y_{j+t})\in J_1} |y_{j+t} - y_j| < \sum_{(x_{i+s},y_{j+t})\in J_1} |x_{i+s} - x_i|$$

$$\sum_{(x_{i+s},y_{j+t})\in J_2} |y_{j+t} - y_j| > \sum_{(x_{i+s},y_{j+t})\in J_2} |x_{i+s} - x_i|,$$

we have

$$\begin{aligned} |h_1^{(i,j)}| &> |\frac{C(i,j)}{kS_x^2} \sum_{s=0}^{\ell} \sum_{t=-\ell}^{0} (x_{i+s} - x_i)| \\ &= O(n/k) \end{aligned}$$

Hence (A.4) is also true.

By the geometric symmetry, the other cases can be discussed in the same way. We finish the proof.

REFERENCES

Brodsky, B.E., and Dharkhovsky, B.S. (1993), *Nonparametric Methods in Change-Point Problems*, Kluwer Academic Publishers, The Netherlands.

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

Chu, C.K. (1994), "Estimation of change-points in a nonparametric regression function through kernel density estimation," *Communications in Statistics-Theory and Methods 23*, 3037-3062.

Eubank, R.L., and Speckman, P.L. (1993),"Nonparametric estimation of functions with jump discontinuities," Department of Statistics, University of Missouri, Columbia, MO 65211.

Gonzalez, R.C., and Woods, R.E. (1992), *Digital Image Processing*, Addison-Wesley Publishing Company, Inc.

Hall, P., and Titterington, M. (1992),"Edge-preserving and peak-preserving smoothing," *Technometrics 34,* 429-440.

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

Lai, T.L., Robbins, H., and Wei, C.Z.(1979),"Strong consistency of least squares estimates in multiple regression II," *Journal of the Multivariate Analysis 9,* 343-362.

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

Müller, H.G.(1992),"Change-points in nonparametric regression analysis," *The Annals of Statistics 20,*737-761.

Müller, H.G., and Song, K.S. (1994), "Maximin estimation of multidimensional boundaries," *Journal of the Multivariate Analysis 50*, 265-281.

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

Qiu, P. (1992), "Nonparametric estimation of the jump surfaces," Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, USA.

Qiu, P.(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 Yandell, B. (1994), "A local polynomial jump detection algorithm in nonparametric regression," Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, USA.

Shiau, J.H.(1985),"Smoothing spline estimation of functions with discontinuities," Technical Report NO. 768, Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, USA.

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

Shiau, J.H., Wahba, G., and Johnson, D.R. (1986), "Partial spline models for the inclusion of tropopause and frontal boundary information in otherwise smooth two- and three-dimensional objective analysis," *Journal of Atmospheric and Oceanic Technology 3*, 714-725.

Speckman, P.L.(1993),"Detection of change-points in nonparametric regression," Department of Statistics, University of Missouri, Columbia, MO 65211.

Wu, J.S., and Chu, C.K.(1993),"Kernel type estimators of jump points and values of a regression function," *The Annals of Statistics 21*, 1545-1566.

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

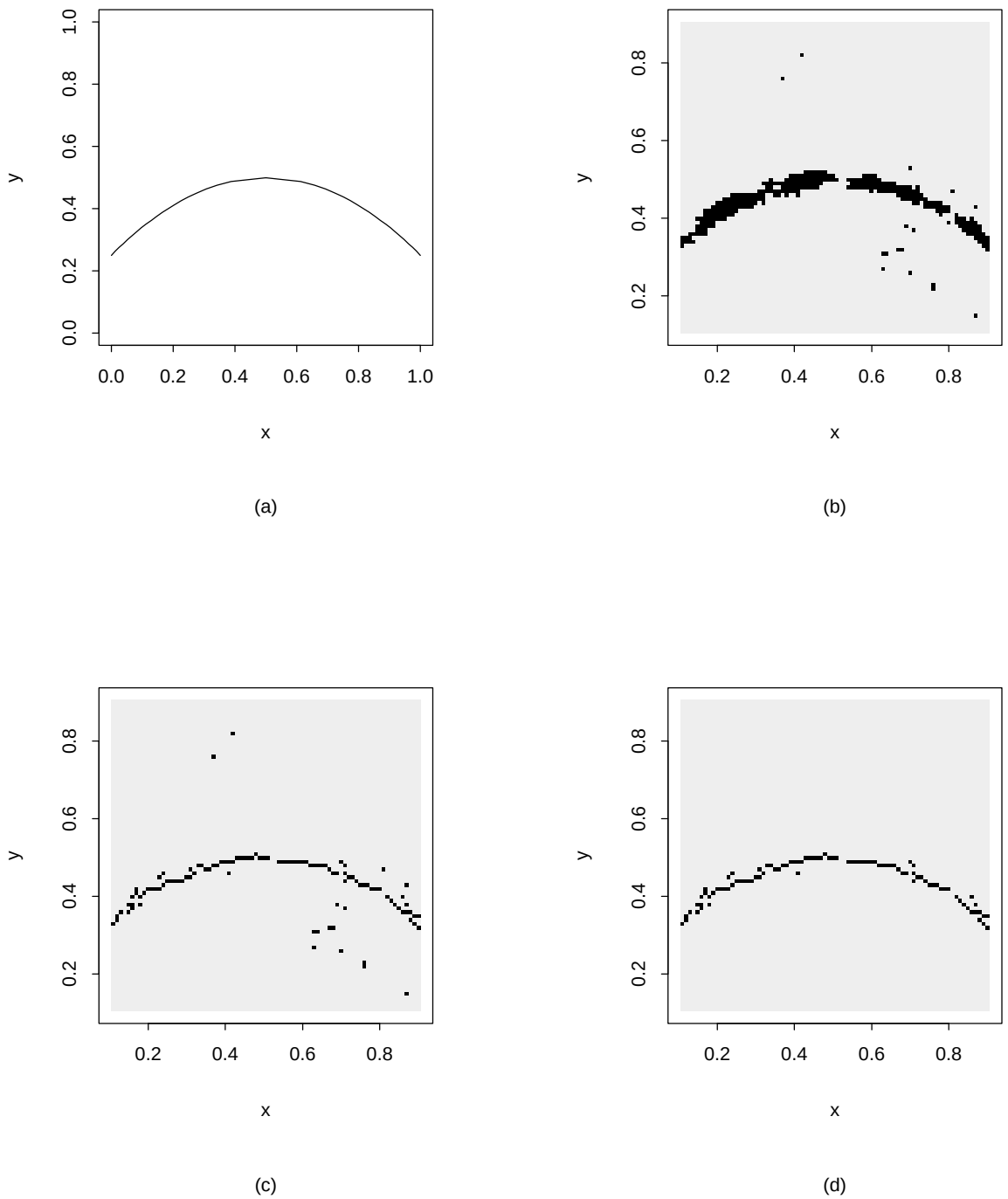

Figure 3.4: (a) The real jump location curve; (b)detected jump candidates from criterion (2.7), which are denoted by black points; (c) the modified jump candidates from those in (b) by the modification procedure (MP1); (d) the modified jump candidates from those in (c) by the modification procedure (MP2).

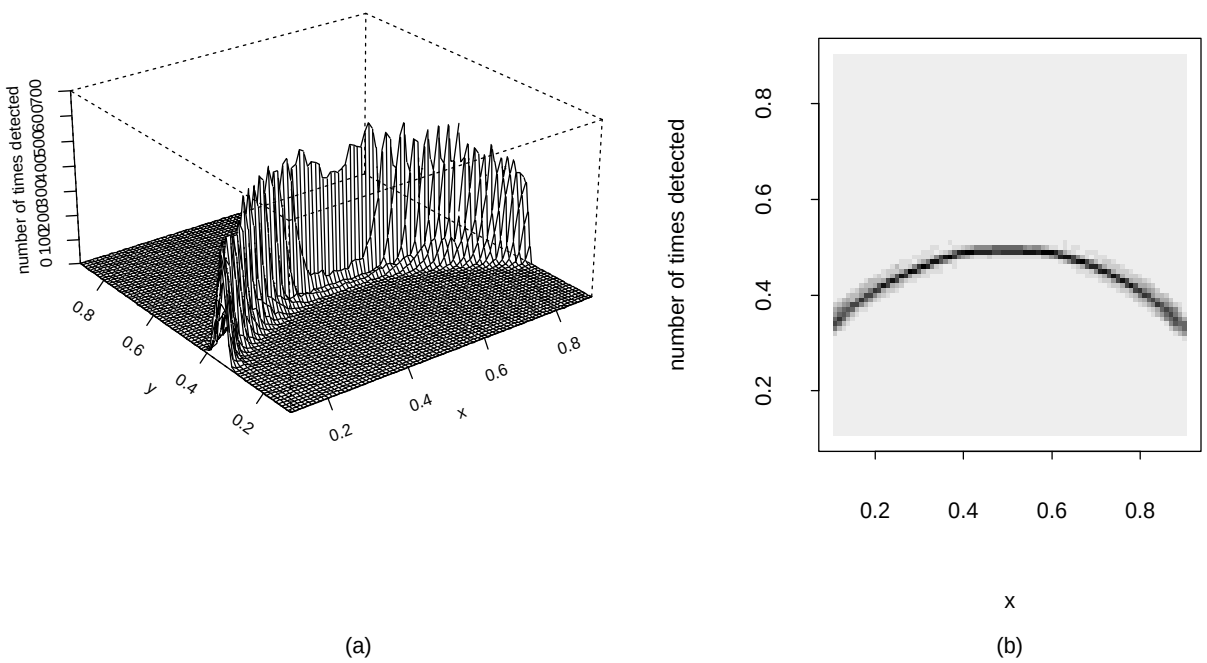

Figure 3.5: (a) The number of times each design point is detected in 1000 replications; (b) the corresponding image plot of (a).

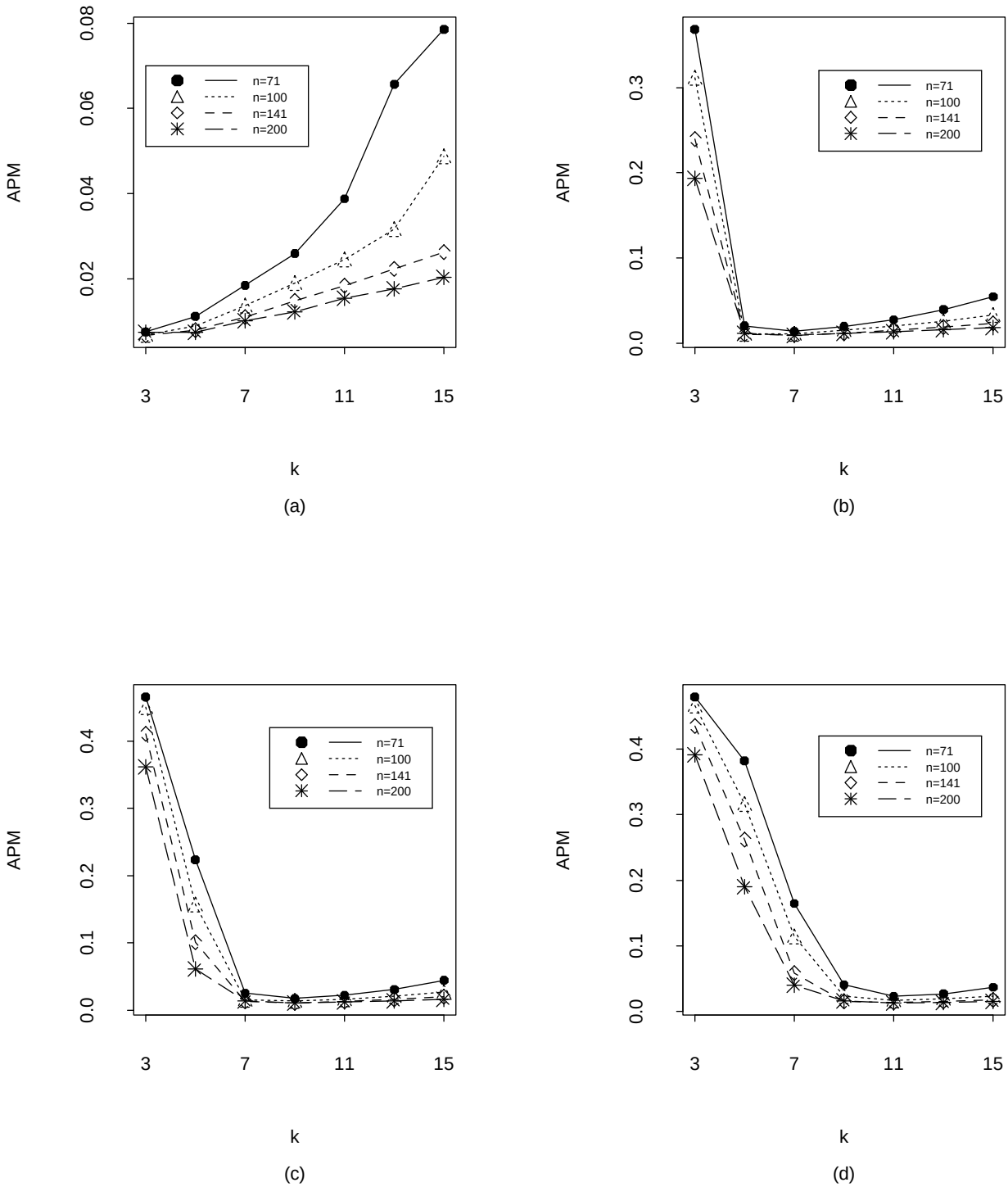

Figure 3.6: Averaged performance measurements (APMs) from 1000 replications with several n, k, and σ^2 values. (a) $\sigma^2 = 0.25$; (b) $\sigma^2 = 0.5$; (c) $\sigma^2 = 0.75$; (d) $\sigma^2 = 1.0$.

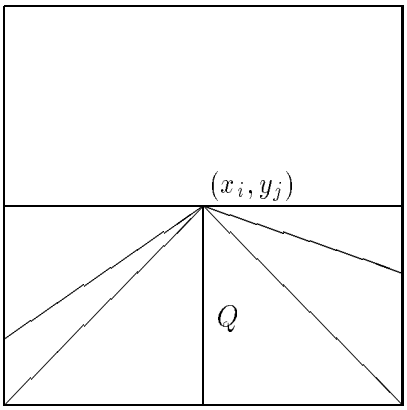

Figure A.1

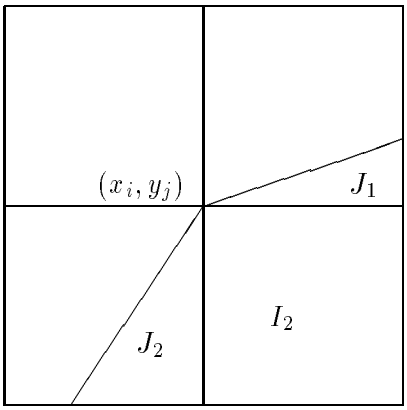

Figure A.2