DISCONTINUITY DETECTION IN REGRESSION SURFACES

Peihua Qiu and Brian Yandell, University of Wisconsin - Madison
Peihua Qiu, Department of Statistics, University of Wisconsin, Madison, WI 53706

KEY WORDS: Jump surfaces, Least squares plane, Jump detection criterion, Image processing.

ABSTRACT

We consider the problem of locating jumps in regression surfaces. A jump detection algorithm is suggested based on local least squares estimation. 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.

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 view point, all of these problems could be regarded as applications of 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 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. Müller (1992), Qiu (1994), Wu and Chu (1993), among many others, discussed various kernel-type methods. These methods were all based on the difference between two one-sided kernel smoothers. Eubank 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. Wang (1995) proposed detecting jumps with wavelet transformations.

In 2-D case, Russian scientists did much theoretical research in this area. Korostelev and Tsybakov (1993) investigated jump location detection and fitting jump surfaces under several kinds of design and jump boundaries. They suggested approximating jump location curves by piecewise polynomials and then estimating the coefficients by maximum likelihood estimation. O'Sullivan and Qian (1994) suggested detecting object boundaries by defining a contrast statistic. 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. Qiu (1992) suggested a so-called *Rotational Difference Kernel Estimator* 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 quite expensive. Jump detection in regression surfaces is directly related to edge detection in computer image processing. Gonzalez and Woods (1992), Qiu and Bhandarkar (1996), Rosenfeld and Kak (1982) and Torre and Poggio (1986) presented an excellent overview of computer edge detection techniques.

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. 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. 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. Comparing with the existing derivative-based edge detectors in image processing literature, we explicitly characterize the jump information in the edge detection criterion and eliminate the effect of the continuous variation of the intensity function on the edge detection. We also establish the statistical consistency of the edge detection procedure and provide the rate of convergence. The conditions imposed on the edge curves are mathematically explicitly expressed. These efforts, we think, might be helpful to the further development of edge detection techniques.

The rest of the paper is organized as follows. In the next section we describe the model and the jump detection method. Numerical examples are discussed in Section 3. In Section 4, we give some concluding remarks.

2 Jump Detection Algorithm

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

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

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 i.i.d. 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

$$f(x, y) = g(x, y) + C(x, y) I_{(x,y)\in\Omega_1}, \qquad (2.2)$$

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 is similar in that it is continuous over connected regions and discontinuous on their boundaries (namely, JLCs).

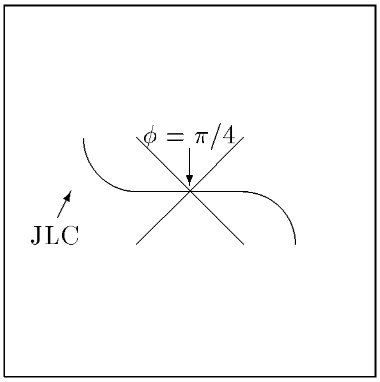

Figure 2.1: At any point on the JLCs, there exist two orthogonal lines crossing at this point and two vertical quadrants formed by these two lines belong to two different regions in a small neighborhood.

The following assumption (AS) is imposed on the JLCs.

<u>**(AS)**</u> 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).

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.

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

A least squares plane is fitted in this neighborhood

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

After some calculations, we have

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

where $z_{..} = \sum_{s,t=-\ell}^{\ell} z_{i+s,j+t}$, $z_{i+s,.} = \sum_{t=-\ell}^{\ell} z_{i+s,j+t}$, $z_{.,j+t} = \sum_{s=-\ell}^{\ell} z_{i+s,j+t}$, $S_x^2 = \sum_{s=-\ell}^{\ell} (x_{i+s} - x_i)^2$, $S_y^2 = \sum_{t=-\ell}^{\ell} (y_{j+t} - y_j)^2$. 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. Furthermore, they have the following property (proof is omitted):

Theorem 2.1 In model (2.1), suppose that $f(x,y)$ has continuous first order partial derivatives over $(0,1)\times(0,1)$ except on the JLCs at which it has the first order right and left partial derivatives. The JLCs satisfy the assumption (AS). The window width k satisfies the conditions that $\lim_{n\to\infty} k=\infty$ and $\lim_{n\to\infty} k/n = 0$. If there is no jump in $N(x_i,y_j)$, then

$$\hat{\beta}_1^{(i,j)} = f'_x(x_i,y_j) + O(\frac{n\sqrt{loglogk}}{k^2}),\quad a.s.$$

$$\hat{\beta}_2^{(i,j)} = f'_y(x_i,y_j) + O(\frac{n\sqrt{loglogk}}{k^2}),\quad a.s.$$

If (x_i,y_j) is on a JLC, then

$$\begin{aligned}\hat{\beta}_1^{(i,j)} &= f'_x(\tilde{x}_i,\tilde{y}_j) + h_1^{(i,j)}C(i,j) + \gamma_1 C_x(i,j) + \\ &\quad O(\frac{n\sqrt{loglogk}}{k^2}),\quad a.s.\end{aligned}$$

$$\begin{aligned}\hat{\beta}_2^{(i,j)} &= f'_y(\tilde{x}_i,\tilde{y}_j) + h_2^{(i,j)}C(i,j) + \gamma_2 C_y(i,j) + \\ &\quad O(\frac{n\sqrt{loglogk}}{k^2}),\quad a.s.\end{aligned}$$

where $(\tilde{x}_i,\tilde{y}_j)$ is some point around (x_i,y_j) which satisfies (i) it is on the same side of the JLC as (x_i,y_j) and (ii) the distance between $(\tilde{x}_i,\tilde{y}_j)$ and (x_i,y_j) tends to zero; $C(i,j), C_x(i,j)$ and $C_y(i,j)$ are the jump magnitudes of $f(x,y)$ and its first order x and y partial derivatives; $h_1^{(i,j)}$ and $h_2^{(i,j)}$ are two constants satisfying $\sqrt{(h_1^{(i,j)})^2+(h_2^{(i,j)})^2} = O(n/k)$; γ_1 and γ_2 are two constants between -1 and 1.

In Theorem 2.1, the term $O(\frac{n\sqrt{loglogk}}{k^2})$ is due to noise. We could see that 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 for a particular lattice point (x_i,y_j). The angle formed by $\vec{v}_{ij} := (\hat{\beta}_1^{(i,j)},\hat{\beta}_2^{(i,j)})$ and the positive direction of x-axis is denoted as $\theta\in[-\pi/4, 7\pi/4]$. 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

$$x_{N1} = x_i + \frac{k}{n\cdot\tan\theta},\quad y_{N1} = y_j + \frac{k}{n}$$

$$x_{N2} = x_i - \frac{k}{n\cdot\tan\theta},\quad y_{N2} = y_j - \frac{k}{n}; \tag{2.4}$$

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

$$x_{N1} = x_i + \frac{k}{n},\quad y_{N1} = y_j + \frac{k}{n}\cdot\tan\theta$$

$$x_{N2} = x_i - \frac{k}{n},\quad y_{N2} = y_j - \frac{k}{n}\cdot\tan\theta \tag{2.5}$$

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 $\vec{v}_{ij}$ is the gradient vector of the fitted LS plane. The underlying JRS increases most rapidly along a near-by direction. If (x_i,y_j) is on a JLC, then assumption (AS) guarantees that the JLC could not be in $N(x_{N1},y_{N1})$ and $N(x_{N2},y_{N2})$ when n is large enough. In other words, (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}.

$$\delta_{ij} := min\{\|\vec{v}_{ij}-\vec{v}_{N1}\|, \|\vec{v}_{ij}-\vec{v}_{N2}\|\} \tag{2.6}$$

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 $\|\cdot\|$ is the Euclidean norm.

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 small. If (x_i,y_j) is on a JLC, by Theorem 2.1, $\delta_{ij}\approx\sqrt{(h_1^{(i,j)})^2+(h_2^{(i,j)})^2}C(i,j)$ $= O(n/k)C(i,j)$ which tends to infinity when n increases. Hence δ_{ij} could be used to detect the jumps.

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

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

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 $\sigma_{N1}^2 = var(\hat{\beta}_1^{(N1)}) = \frac{\sigma^2}{kS_x^2}$. Therefore a natural threshold value of δ_{ij} is

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

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

Suppose that (x_i, y_j) is on a JLC with jump magnitude $C(i,j)$. Then the values of most kernel-type jump detection criteria (e.g. Müller and Song, 1994) are about $C(i,j)$ at this point while our criterion is of order $O(n/k)$ which tends to infinity with the sample size. Hence δ_{ij} is more sensitive to the jumps. This property has two benefits. One is that δ_{ij} visually reveals the jumps better. The other is that our jump detector is more robust to the selection of the threshold. The threshold could be chosen a little bit larger than usual without missing the jumps when the sample size is larger since δ_{ij} is quite large in this case.

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. Two modification procedures (MPs) are also suggested to make the detected jump boundaries thin and to delete some scattered deceptive candidates.

We summarize the jump detection method in the following algorithm.

The Jump Detection Algorithm

1. *At any (x_i, y_j) with $\ell+1 \leq i, j \leq n-\ell$, fit a LS plane in $N(x_i, y_j)$ by formula (2.3).*
2. *Use (2.4)-(2.5) to determine two neighboring design points of $(x_i, y_j), (3k+1)/2 \leq i, j \leq n-(3k-1)/2$.*
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 to determine the final candidates.*

Theorem 2.2 If α_n in (2.7) is chosen such that (i) $\lim_{n\to\infty} \alpha_n = 0$; (ii) $\lim_{n\to\infty} log(\alpha_n)/log(log(k)) = -\infty$; and (iii) $\lim_{n\to\infty} log(\alpha_n)/k^2 = 0$, then the detected jumps are a.s. consistent in the Hausdorff distance and the convergence rate is $O(n^{-1}log(n))$.

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 $O(n\sqrt{-log(\alpha_n)}/k^2)$. (We use the fact that $\chi^2_{2,\alpha_n} = -2log(\alpha_n)$ here.) Thirdly, the order of the standard deviation of δ_{ij} is $O(n\sqrt{log(log(k))}/k^2)$. So the jump information dominates the randomness in the jump detection criterion as long as k tends to infinity with n.

3 Numerical Analysis

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.

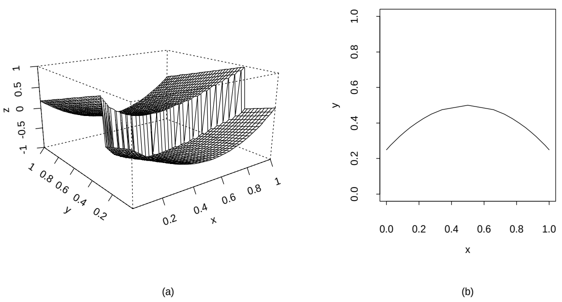

Figure 3.1: (a) The jump regression surface used in the example; (b) the jump location curve.

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 $N(0, 0.5^2)$. We then use formulas (2.3)-(2.6) to calculate the jump detection criterion $\{\delta_{ij}\}$, initially with $k = 7$. The gradient vector $\vec{v}_{ij}$ of the fitted LS plane at each design point is shown by Figure 3.2.

Then a threshold is calculated by formula (2.7) with $\alpha_n = 0.001$ which is the smallest number in most χ^2 tables. The flagged jump candidates are plotted in Figure 3.3(b) by black points. We notice that the detected jump boundary is quite thick and there are some scattered candidates also. We then use two modification procedures to modify the set of candidates. The results are plotted in Figure 3.3(c) and (d). As a comparison, we plot the real JLC in Figure 3.3(a). We notice that there are some breaks here or there in the detected jump boundary in Figure 3.3(d). The detected boundary is not thin

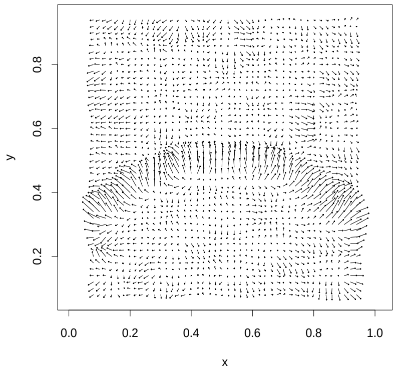

Figure 3.2: The gradient vector $\vec{v}_{ij}$ of the fitted LS plane at each design point.

enough at some places. These imply that there is some room for our MPs to be improved.

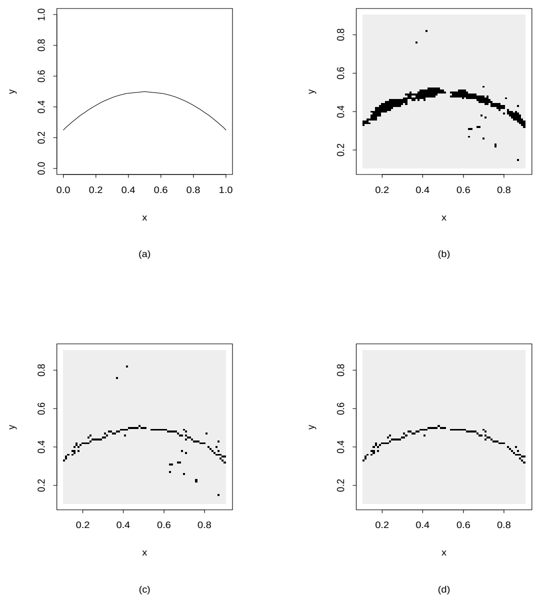

Figure 3.3: (a) The real jump location curve; (b)detected jump candidates by criterion (2.7); (c) the modified jump candidates from those in (b) by the first MP; (d) the modified jump candidates from those in (c) by the second MP.

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 3.4. 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 set of candidates to the real JLCs as a performance measurement. This measurement is averaged again for 1000 replications. The re-

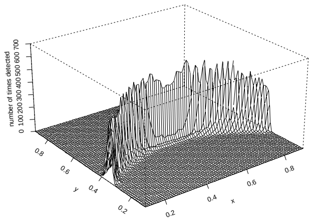

Figure 3.4: The number of times each design point is detected in 1000 replications.

sults for several n, k and σ^2 values are presented in Figure 3.5. 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 much 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.

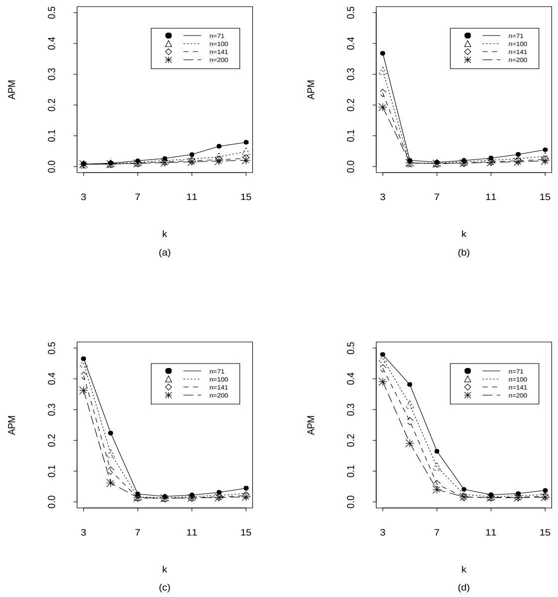

Figure 3.5: Averaged performance measurements (APMs) from 1000 replications with several n, k, and σ^2 values. (a) $\sigma^2 = 0.25$; (b) $\sigma^2 = 0.5$; (c) $\sigma^2 = 0.75$; (d) $\sigma^2 = 1.0$.

4 Some Concluding Remarks

We have 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 two of 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 location detection and jump surface fitting. If we put more structure on the jump locations, then fitting the jump surfaces would 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 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 relationship 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.

REFERENCES

Eubank, R.L., and Speckman, P.L. (1993),"Nonparametric estimation of functions with jump discontinuities," IMS Lecture Notes, vol.23, *Change-Point Problems* (E. Carlstein, H.G. Müller and D. Siegmund eds.).

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.

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.

O'Sullivan, F., and Qian, M. (1994), "A regularized contrast statistic for object boundary estimation-implementation and statistical evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence 16*, 561-570.

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

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

Torre, V., and Poggio, T.A. (1986), "On edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence 8*, 147-163.

Wang, Y. (1995), "Jump and sharp cusp detection by wavelets," *Biometrika 82*, 385-397.

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.