The Local Piecewisely Linear Kernel Smoothing Procedure For Fitting Jump Regression Surfaces

Peihua Qiu
School of Statistics
University of Minnesota
313 Ford Hall
224 Church St. S.E.
Minneapolis, MN 55455

Abstract

It is known that a surface fitted by conventional local smoothing procedures is not statistically consistent at the jump locations of the true regression surface. In this paper, a procedure is suggested for modifying conventional local smoothing procedures such that the modified procedures can fit the surface with jumps preserved automatically. Taking the local linear kernel smoothing procedure as an example, in a neighborhood of a given point, we fit a bivariate piecewisely linear function with possible jumps along the boundaries of four quadrants. The fitted function provides four estimators of the surface at the given point, which are constructed from observations in the four quadrants, respectively. When the difference among the four estimators is smaller than a threshold value, the given point is most likely a continuous point and the surface at that point is then estimated by the average of the four estimators. When the difference is larger than the threshold value, the given point is likely a jump point and at least one of the four estimators estimates the surface well under some regularity conditions. By comparing the weighted residual sums of squares of the four estimators, the best one is selected to define the surface estimator at the given point. Like most conventional estimators, the current surface estimator has an explicit mathematical formula. Therefore it is easy to compute and convenient to use. It can be applied directly to image reconstruction problems and other jump surface estimation problems including mine surface estimation in geology and equi-temperature surface estimation in meteorology and oceanography.

Key Words: Denoising; Edge detection; Image reconstruction; Jump location curves; Jump-preserving surface estimation; Local linear kernel estimation; Nonparametric regression.

1 Introduction

We discuss regression surface estimation when the surface has jumps in the design space. This problem has broad applications. For example, an image can be regarded as a surface of the image intensity at each pixel. Such a surface has jumps (called edges in the image processing literature) at the outlines of objects. Edge-preserving image reconstruction is an important research problem for several purposes including denoising, changing the image resolution and data compression (Gonzalez and Woods 1992). Similar jump surface estimation problems can be found in other application fields such as geology, meteorology and oceanography. In this paper, a procedure is suggested for estimating jump surfaces with the jumps preserved.

Suppose that the related regression model is

$$z_{ij} = f(x_i, y_j) + \epsilon_{ij},\ i, j = 1, 2, \cdots, n, \tag{1.1}$$

where $\{(x_i, y_j) = (i/n, j/n), i, j = 1, 2, \cdots, n\}$ are equally spaced design points in the design space $[0, 1] \times [0, 1]$, $\{\epsilon_{ij}\}$ are i.i.d. random errors with mean 0 and variance σ^2, f is an unknown nonparametric regression function which is continuous in the entire design space except on some curves which are called the jump location curves (JLCs) hereafter, and $N = n^2$ is the sample size.

It can be checked that the surface fitted by most conventional local smoothing procedures is not statistically consistent at the JLCs. For example, the local linear kernel (LK) estimator $\widehat{f}_{LK}(x, y)$ of $f(x, y)$, for some $(x, y) \in [0, 1] \times [0, 1]$, is defined by the solution for α in the following minimization problem (see e.g., Fan and Gijbels 1996; Ruppert and Wand 1994):

$$\min_{\alpha,\beta_1,\beta_2} \sum_{i=1}^{n} \sum_{j=1}^{n} \{z_{ij} - \alpha - \beta_1(x_i - x) - \beta_2(y_j - y)\}^2 K(\frac{x_i - x}{h_n}, \frac{y_j - y}{p_n}), \tag{1.2}$$

where K is a bivariate kernel function with support $[-1/2, 1/2] \times [-1/2, 1/2]$, h_n and p_n are two bandwidth parameters. It can be checked that $\widehat{f}_{LK}(x, y)$ does not converge to $f(x, y)$ if the point (x, y) is on a JLC.

A main reason why conventional surface estimators are inconsistent around the JLCs is that a "continuous" function, which is a linear function in the case of (1.2), is used for estimating a jump function. A natural way to overcome this limitation is to fit a piecewisely linear function in (1.2), which may have jumps along the boundaries of the four quadrants $Q_{11}(x, y), Q_{12}(x, y), Q_{21}(x, y)$ and $Q_{22}(x, y)$ of the neighborhood $[x - h_n/2, x + h_n/2] \times [y - p_n/2, y + p_n/2]$ of a given point (x, y)

as shown in Figure 1.1. The fitted function then provides four estimators of the surface at (x, y), which correspond to the four pieces of fitted planes in $Q_{11}(x, y), Q_{12}(x, y), Q_{21}(x, y)$ and $Q_{22}(x, y)$, respectively. In the case when (x, y) is on a single JLC and the JLC has right and left tangent lines at (x, y) (a JLC with parametric expressions $x = x(t)$ and $y = y(t)$ has a right (left) tangent line at a given point if the right (left) derivatives of $x(t)$ and $y(t)$ both exist at that point), at least one of the four estimators should estimate $f(x, y)$ well, as demonstrated by Figure 1.2. In Figure 1.2(a), the right and left tangent lines of the JLC at (x, y) are located in a single quadrant. In such a case, the three estimators constructed from the other three quadrants should all estimate $f(x, y)$ well. In Figure 1.2(b), the right and left tangent lines are located in two different quadrants which are next to each other. In this case, the two estimators constructed from the other two quadrants should estimate $f(x, y)$ well. In the last situation shown by Figure 1.2(c), the two tangent lines are in two opposite quadrants. In this case, only one estimator constructed from one of the other two quadrants estimates $f(x, y)$ well. So in all three cases at least one of the four estimators estimates $f(x, y)$ well. The best one of the four, which can be determined by comparing the weighted residual sums of squares of the four fitted planes in the four quadrants, can be used as the fitted value of $f(x, y)$. On the other hand, if there are no jumps in the neighborhood of (x, y), then all four estimators estimate $f(x, y)$ well and their average should provide a reasonable estimator of $f(x, y)$. In applications, however, it is often unknown whether or not (x, y) is a jump point. Therefore a data-based mechanism is needed for making such a judgment. To this end, a threshold parameter is introduced in our procedure and the point (x, y) is regarded as a jump point if the range of the four estimators exceeds the threshold value.

In the literature, there are several conventional local smoothing procedures available which include running averages (Tukey 1977), the locally weighted scatter plot smoothing procedure (Cleveland 1979), kernel smoothing procedures (Härdle 1990), local polynomial kernel smoothing procedures (Fan and Gijbels 1996) and several others. In this paper, we focus mainly on modifying the local linear kernel smoothing procedure to accommodate jumps. The other conventional local smoothing procedures can be modified in a similar way.

There are some existing procedures in the statistical literature for fitting jump regression surfaces. Donoho and Johnstone (1994) pointed out that jump curves/surfaces could be estimated well by discrete wavelet transformation (DWT) and thresholding procedures. Qiu (1998) proposed a three-stage procedure for fitting jump surfaces. In the first stage, jump candidate points were

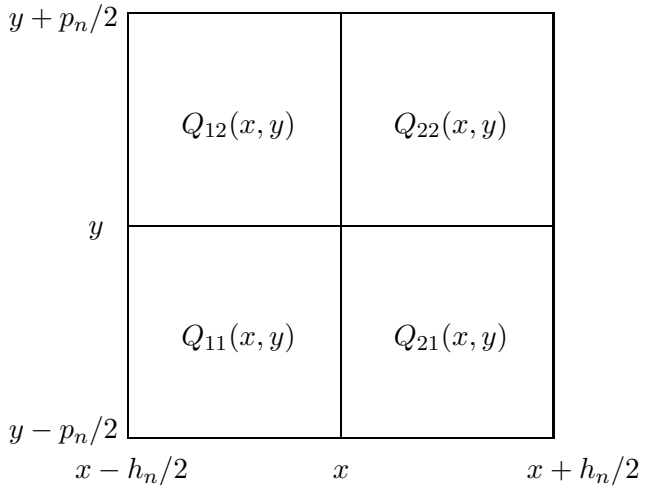

Figure 1.1: The neighborhood of a given point (x, y) consists of four quadrants $Q_{11}(x, y), Q_{12}(x, y), Q_{21}(x, y)$ and $Q_{22}(x, y)$.

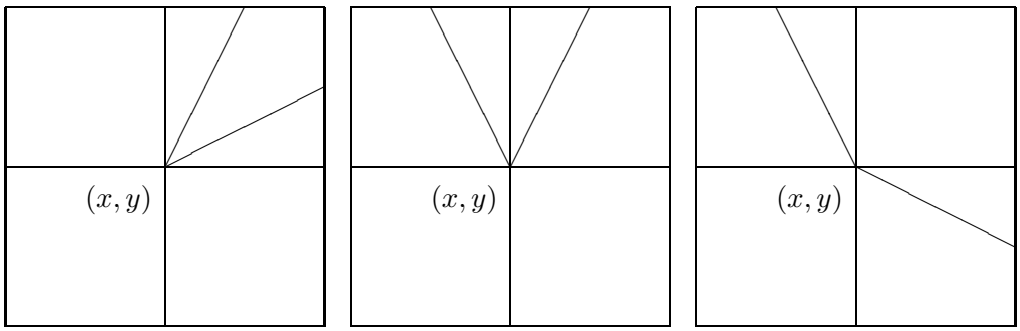

Figure 1.2: (a) The right and left tangent lines of the JLC at (x, y) are located in a single quadrant; (b) the two tangent lines are in two different quadrants which are next to each other; (c) the two tangent lines are in two opposite quadrants.

detected by a jump detector. A local principal component line was then fitted through these points in a neighborhood of a given point. Finally, observations on the same side of the line as the given point were combined using a weighted average procedure to fit the surface at that point. Chu *at al.* (1998) studied two types of jump-preserving smoothers: the sigma filter and the M smoother. Both of them were based on the idea of robust estimation. Polzehl and Spokoiny (2000) suggested an adaptive weights smoothing algorithm for estimating jump surfaces when they contain large homogeneous regions. When the number of JLCs is assumed to be known and the JLCs satisfy some smoothing conditions, Korostelev and Tsybakov (1993), Müller and Song (1994), O'Sullivan and Qian (1994), Qiu (1997), Wang (1998) and several others suggested various two-stage procedures: the JLCs were estimated first and then the regression surface was fitted in design sub-spaces

separated by the estimated JLCs. Several papers including Hall and Rau (2000), Hall, Peng and Rau (2001), Qiu (2002), Qiu and Bhandarkar (1996) and Qiu and Yandell (1997) focus mainly on estimation of the JLCs.

Edge-preserving image reconstruction is essentially the same problem as jump-preserving regression surface fitting. When the image has a limited number of colors, several authors including Switzer (1983), Switzer *et al.* (1982) and Owen (1984) suggested using discriminant analysis to classify each pixel of the image as the most probable color type. Image reconstruction based on Markov random field (MRF) models is an active research area in recent years. The true image is assumed to be a MRF, or equivalently, to have a Gibbs distribution. Geman and Geman (1984) suggested reconstructing the image by maximizing *a posteriori* (MAP) with a restoration algorithm which is based on stochastic relaxation and annealing. Besag (1986) suggested maximizing the marginal posterior distribution at each pixel by using the Iterated Conditional Modes (ICM) algorithm. See Besag *et al.* (1995), Fessler *et al.* (2000), Li (1995) and Marroquin *et al.* (2001) for discussions of recent developments in this area, especially for image reconstruction with Markov chain Monte Carlo (McMC).

Compared to existing procedures, the current procedure has several advantages. Many jump-preserving surface reconstruction procedures impose various restrictive assumptions on the model. For example, O'Sullivan and Qian (1994) assumed that the JLCs were "smooth, simple and closed" curves. Müller and Song (1994) assumed that the population of the JLCs was known. Qiu (1997) assumed that there was only one JLC and the JLC satisfied a Lipschitz condition. All these methods required that the number of JLCs was known beforehand. As more structure is assumed for the model, the reconstruction of the jump surface becomes easier. But some real applications are excluded at the same time. The current procedure can work well under milder conditions on the model. Therefore it can be applied to some real-life surface estimation problems that cannot be handled directly by these other procedures.

From a computational viewpoint, many existing procedures are relatively complicated. For example, the procedure by Qiu (1998) consists of three steps. Before the surface estimator is defined at a given point, the jump positions need to be detected and a principal component line needs to be fitted in a neighborhood of the given point. The M smoother discussed by Chu *et al.* (1998), the additive weights smoothing procedure by Polzehl and Spokoiny (2000) and the MRF

procedures are all based on iterative algorithms. In comparison, the fitted surface of the current procedure has an explicit mathematical formula, making it simple to compute and convenient to use.

Because of the iterative nature of some existing procedures (e.g., Geman and Geman 1984; Polzehl and Spokoiny 2000), we do not know much about their theoretical properties. Although the performance of these procedures can be evaluated by numerical experiments based on visual impression, we believe that theoretical justifications can help us understand their strengths and limitations so that they can be further improved. With the current procedure, some mild conditions on the model will be given explicitly so that users know beforehand where in the design space this method could work well and where it may fail to reconstruct the surface properly.

The rest of the article is organized as follows. In next section, our jump-preserving surface fitting procedure is introduced in some detail. Selection of the related procedure parameters is discussed in Section 3. In Section 4, some numerical examples are presented regarding the numerical performance of the proposed procedure. Several remarks conclude the article in Section 5. Some statistical properties of the fitted surface of the proposed procedure are given in Appendix A.

2 The Jump-Preserving Surface Fitting Procedure

In a neighborhood $N_n(x,y) = [x - h_n/2, x + h_n/2] \times [y - p_n/2, y + p_n/2]$ of a given point $(x,y) \in [0,1] \times [0,1]$, a local piecewisely linear function is fitted by the following minimization problem:

$$\begin{aligned}\min_{a_{s_1s_2},b_{s_1s_2},c_{s_1s_2};\, s_1,s_2=1,2} \quad & \sum_{i=1}^{n}\sum_{j=1}^{n} \{z_{ij} - [\, a_{11} + b_{11}(x_i - x) + c_{11}(y_j - y) + \\ & (a_{21} - a_{11})I(x_i - x) + (b_{21} - b_{11})(x_i - x)I(x_i - x) + \\ & (c_{21} - c_{11})(y_j - y)I(x_i - x) + \\ & (a_{12} - a_{11})I(y_j - y) + (b_{12} - b_{11})(x_i - x)I(y_j - y) + \\ & (c_{12} - c_{11})(y_j - y)I(y_j - y) + \\ & (a_{22} - a_{21} - a_{12} + a_{11})I(x_i - x)I(y_j - y) + \\ & (b_{22} - b_{21} - b_{12} + b_{11})(x_i - x)I(x_i - x)I(y_j - y) + \\ & (c_{22} - c_{21} - c_{12} + c_{11})(y_j - y)I(x_i - x)I(y_j - y)\,]\}^2 \\ & K(\frac{x_i - x}{h_n}, \frac{y_j - y}{p_n}),\end{aligned} \tag{2.1}$$

where $I(\cdot)$ is an indicator function defined by $I(x) = 1$ if $x \geq 0$ and 0 otherwise. It can be checked that the procedure (2.1) is equivalent to fitting a linear function $a_{s_1s_2} + b_{s_1s_2}(u-x) + c_{s_1s_2}(v-y)$, which is regarded as a function of (u, v) in $Q_{s_1s_2}(x, y)$ (cf. Figure 1.1), by the weighted least squares procedure (1.2) for $s_1, s_2 = 1, 2$. By some routine algebraic manipulations, the solution of (2.1) is

$$\begin{aligned}
\widehat{a}_{s_1s_2}(x,y) &= \frac{\sum_{i=1}^n \sum_{j=1}^n \{A^{(1)}_{s_1s_2} + A^{(2)}_{s_1s_2}(x_i - x) + A^{(3)}_{s_1s_2}(y_j - y)\} z_{ij} K(\frac{x_i - x}{h_n}, \frac{y_j - y}{p_n})}{\Delta_{s_1s_2}}, \\
\widehat{b}_{s_1s_2}(x,y) &= \frac{\sum_{i=1}^n \sum_{j=1}^n \{A^{(2)}_{s_1s_2} + A^{(4)}_{s_1s_2}(x_i - x) + A^{(5)}_{s_1s_2}(y_j - y)\} z_{ij} K(\frac{x_i - x}{h_n}, \frac{y_j - y}{p_n})}{\Delta_{s_1s_2}}, \\
\widehat{c}_{s_1s_2}(x,y) &= \frac{\sum_{i=1}^n \sum_{j=1}^n \{A^{(3)}_{s_1s_2} + A^{(5)}_{s_1s_2}(x_i - x) + A^{(6)}_{s_1s_2}(y_j - y)\} z_{ij} K(\frac{x_i - x}{h_n}, \frac{y_j - y}{p_n})}{\Delta_{s_1s_2}}, \\
&\quad \text{for } s_1, s_2 = 1, 2, \qquad (2.2)
\end{aligned}$$

where

$$\Delta_{s_1s_2} = \begin{vmatrix} B^{(00)}_{s_1s_2} & B^{(10)}_{s_1s_2} & B^{(01)}_{s_1s_2} \\ B^{(10)}_{s_1s_2} & B^{(20)}_{s_1s_2} & B^{(11)}_{s_1s_2} \\ B^{(01)}_{s_1s_2} & B^{(11)}_{s_1s_2} & B^{(02)}_{s_1s_2} \end{vmatrix},$$

$$A^{(1)}_{s_1s_2} = B^{(20)}_{s_1s_2} B^{(02)}_{s_1s_2} - B^{(11)}_{s_1s_2} B^{(11)}_{s_1s_2},\ A^{(2)}_{s_1s_2} = B^{(01)}_{s_1s_2} B^{(11)}_{s_1s_2} - B^{(10)}_{s_1s_2} B^{(02)}_{s_1s_2},$$

$$A^{(3)}_{s_1s_2} = B^{(10)}_{s_1s_2} B^{(11)}_{s_1s_2} - B^{(01)}_{s_1s_2} B^{(20)}_{s_1s_2},\ A^{(4)}_{s_1s_2} = B^{(00)}_{s_1s_2} B^{(02)}_{s_1s_2} - B^{(01)}_{s_1s_2} B^{(01)}_{s_1s_2},$$

$$A^{(5)}_{s_1s_2} = B^{(01)}_{s_1s_2} B^{(10)}_{s_1s_2} - B^{(00)}_{s_1s_2} B^{(11)}_{s_1s_2},\ A^{(6)}_{s_1s_2} = B^{(00)}_{s_1s_2} B^{(20)}_{s_1s_2} - B^{(10)}_{s_1s_2} B^{(10)}_{s_1s_2},$$

$$B^{(r_1r_2)}_{s_1s_2} = \sum_{i=1}^n \sum_{j=1}^n (x_i - x)^{r_1} (y_j - y)^{r_2} K_{s_1s_2}(\frac{x_i - x}{h_n}, \frac{y_j - y}{p_n}),$$

and $K_{s_1s_2}(x/h_n, y/p_n) = K(x/h_n, y/p_n)$ if $(x, y) \in Q_{s_1s_2}(0, 0)$ and 0 otherwise, for $s_1, s_2 = 1, 2$ and $r_1, r_2 = 0, 1, 2$.

For simplicity of presentation, a point (x, y) is called *a nonsingular point* of the JLCs if it is on a single JLC and the JLC has left and right tangent lines at (x, y). All other points on the JLCs are called *singular points*. Obviously a point (x, y) is a singular point if it is a cross point of several JLCs or it is on a single JLC but the JLC does not have left or right tangent lines at (x, y).

The quantities $\widehat{a}_{11}(x, y), \widehat{a}_{12}(x, y), \widehat{a}_{21}(x, y)$ and $\widehat{a}_{22}(x, y)$ defined in (2.2) provide four estimators of $f(x, y)$. If there are no jumps in $N_n(x, y)$, then all of them should estimate $f(x, y)$ well and it is reasonable to use their average to estimate $f(x, y)$. If the point (x, y) is a nonsingular point of the JLCs, then at least one of the four estimators estimates $f(x, y)$ well as explained in Section 1 (cf. Figure 1.2). A natural way to estimate $f(x, y)$ in the latter case is to choose one of

$\widehat{a}_{11}(x,y), \widehat{a}_{12}(x,y), \widehat{a}_{21}(x,y)$ and $\widehat{a}_{22}(x,y)$ based on the weighted residual sums of squares (RSS) of the four pieces of fitted planes in the four quadrants of $N_n(x,y)$. In order to tell whether or not there are jumps in $N_n(x,y)$, a threshold parameter u_n is introduced and it is concluded that there are no jumps in $N_n(x,y)$ if

$$range\,(\widehat{a}_{11}(x,y), \widehat{a}_{12}(x,y), \widehat{a}_{21}(x,y), \widehat{a}_{22}(x,y)) \leq u_n. \tag{2.3}$$

Our piecewisely linear kernel (PLK) estimator of $f(x,y)$ is then defined by

$$\widehat{f}_{PLK}(x,y) = \begin{cases} \frac{1}{4}(\widehat{a}_{11}(x,y) + \widehat{a}_{12}(x,y) + \widehat{a}_{21}(x,y) + \widehat{a}_{22}(x,y)), & \text{if (2.3) is true,} \\ \tilde{a}(x,y), & \text{otherwise,} \end{cases} \tag{2.4}$$

where $\tilde{a}(x,y)$ is one of $\widehat{a}_{11}(x,y), \widehat{a}_{12}(x,y), \widehat{a}_{21}(x,y)$ and $\widehat{a}_{22}(x,y)$ which satisfies

$$RSS(\tilde{a}(x,y)) = \min\{RSS(\widehat{a}_{11}(x,y)), RSS(\widehat{a}_{12}(x,y)), RSS(\widehat{a}_{21}(x,y)), RSS(\widehat{a}_{22}(x,y))\}$$

and $RSS(\widehat{a}_{s_1s_2}(x,y))$ is the weighted residual sum of squares defined by

$$RSS(\widehat{a}_{s_1s_2}(x,y)) = \sum_{i=1}^{n}\sum_{j=1}^{n} \Big\{z_{ij} - [\widehat{a}_{s_1s_2}(x,y) + \widehat{b}_{s_1s_2}(x,y)(x_i - x) + \widehat{c}_{s_1s_2}(x,y)(y_j - y)]\Big\}^2 K_{s_1s_2}(\frac{x_i - x}{h_n}, \frac{y_j - y}{p_n}) \tag{2.5}$$

for $s_1, s_2 = 1, 2$. If there are two or more of $\widehat{a}_{11}(x,y), \widehat{a}_{12}(x,y), \widehat{a}_{21}(x,y)$ and $\widehat{a}_{22}(x,y)$ with the same and smallest RSS value, then $\tilde{a}(x,y)$ is defined by their simple average.

The entire jump-preserving surface fitting procedure can be summarized as follows:

- At a given point (x,y), compute the four estimators $\widehat{a}_{11}(x,y), \widehat{a}_{12}(x,y), \widehat{a}_{21}(x,y)$ and $\widehat{a}_{22}(x,y)$ by (2.2).
- Compute the range of the four estimators and compare the range to the threshold value u_n.
- If the range is smaller than or equal to u_n, then use the simple average of the four estimators as the surface estimator at (x,y).
- If the range is larger than u_n, then compute $RSS(\widehat{a}_{11}(x,y))$, $RSS(\widehat{a}_{12}(x,y))$, $RSS(\widehat{a}_{21}(x,y))$ and $RSS(\widehat{a}_{22}(x,y))$ by (2.5) and determine $\tilde{a}(x,y)$ which is one of $\widehat{a}_{11}(x,y), \widehat{a}_{12}(x,y), \widehat{a}_{21}(x,y)$ and $\widehat{a}_{22}(x,y)$ with the smallest RSS value. The surface estimator at (x,y) is then defined by $\tilde{a}(x,y)$.

By using the above procedure, the regression surface is estimated with jumps preserved automatically. Unlike two-stage or three-stage procedures such as those suggested by Müller and Song (1994), O'Sullivan and Qian (1994) and Qiu (1997, 1998), explicit jump detection is avoided in the current procedure. There are several benefits by doing so. One is that the current procedure is very easy to use because its surface estimator has a mathematical formula and can be calculated in a single stage like most conventional surface estimators. The second major benefit is that we do not need any restrictive assumptions on the JLCs. In contrast, some existing procedures (e.g. O'Sullivan and Qian 1994; Qiu 1997) use smooth curves to estimate JLCs. These procedures usually assume that the number of JLCs is known (e.g. there is only one JLC) and the JLCs are smooth.

In (2.4), there are several alternative ways to define $\widehat{f}_{PLK}(x,y)$ when (2.3) is true. For example, $\widehat{f}_{PLK}(x,y)$ could be defined by the weighted average

$$\frac{\sum_{s_1=1}^{2}\sum_{s_2=1}^{2}\widehat{a}_{s_1s_2}(x,y)/RSS(\widehat{a}_{s_1s_2}(x,y))}{\sum_{s_1=1}^{2}\sum_{s_2=1}^{2}1/RSS(\widehat{a}_{s_1s_2}(x,y))}.$$

It can also be defined by the conventional local linear kernel estimator constructed from the entire neighborhood $N_n(x,y)$. However, we need to compute $RSS(\widehat{a}_{11}(x,y))$, $RSS(\widehat{a}_{12}(x,y))$, $RSS(\widehat{a}_{21}(x,y))$ and $RSS(\widehat{a}_{22}(x,y))$ if the first alternative approach is used, which is not necessary by using the definition (2.4). Since most design points are continuity points of the regression surface at which (2.3) is likely true, this extra computation is quite expensive. To use the second alternative approach, we need to compute the conventional local linear kernel estimator of $f(x,y)$ in the entire neighborhood $N_n(x,y)$ after $\widehat{a}_{11}(x,y)$, $\widehat{a}_{12}(x,y)$, $\widehat{a}_{21}(x,y)$ and $\widehat{a}_{22}(x,y)$ are computed. Therefore these two alternative approaches require much extra computation compared to the approach of (2.4). If the extra computation involved is not an issue for a specific application problem, then the alternative approaches might be more reasonable to use.

From Figures 1.2(a) and 1.2(b), it seems that a potential improvement of (2.4) is to use more than one of $\widehat{a}_{11}(x,y)$, $\widehat{a}_{12}(x,y)$, $\widehat{a}_{21}(x,y)$ and $\widehat{a}_{22}(x,y)$ for estimating $f(x,y)$ when (2.3) is not true. For example, in the case of Figure 1.2(a), $\widehat{f}_{PLK}(x,y)$ could be defined as the average of $\widehat{a}_{11}(x,y)$, $\widehat{a}_{12}(x,y)$ and $\widehat{a}_{21}(x,y)$. However, it is a challenging task to distinguish the three cases demonstrated in Figure 1.2 based on data, and this approach is left for future research.

In model (1.1), it is assumed that the design points are equally spaced in the design space, which is true in many applications including the image reconstruction problem. From the construction

of $\widehat{f}_{PLK}(x,y)$, it can be seen that this procedure will also work when the design points are not equally spaced. In such cases, if the design points have some homogeneity, which is often assumed in multivariate nonparametric regression (cf. Müller 1988, Chapter 6, for related discussion), then constant bandwidths may be still appropriate to use. If the design points have no homogeneity at all, then variable bandwidths might be more appropriate. In regions where the design points are sparse, the bandwidths need to be relatively large to include enough observations for estimating f. In regions where the design points are dense, the bandwidths need to be relatively small.

Like most local smoothing estimators, the estimator $\widehat{f}_{PLK}(x,y)$ is defined only for $(x,y) \in [h_n/2, 1-h_n/2] \times [p_n/2, 1-p_n/2]$. It is not well defined in the boundary regions of the design space. This is the notorious "boundary problem" in the literature. There are several existing proposals to overcome this problem. For example, most DWT software packages use periodic or symmetric "padding" methods to define neighborhoods in the boundary regions (Nason and Silverman 1994). In this paper, the symmetric "padding" method is used in all numerical examples. It should be noted that different padding procedures may introduce different artificial jumps at the border of the design space. But this would not affect the estimated surface of (2.4) much due to the fact that procedure (2.4) is jump preserving.

3 Parameter Selection

For a given point (x,y) in the design space, if $u_n = \infty$ in (2.4), then $\widehat{f}_{PLK}(x,y)$ is close to the conventional local linear kernel estimator $\widehat{f}_{LK}(x,y)$, which removes noise well in the continuity regions of f. But it blurs the jumps at the same time. On the other hand, if $u_n = 0$, then $\widehat{f}_{PLK}(x,y) = \tilde{a}(x,y)$, which preserves jumps well. But its ability to remove noise is limited because it is constructed in a single quadrant of $N_n(x,y)$ only. The main purpose for using the threshold parameter u_n is to make the estimator $\widehat{f}_{PLK}(x,y)$ able to preserve the jumps around the JLCs and remove noise efficiently as well. It is therefore important to choose its value properly.

From (2.2), the four estimators $\widehat{a}_{11}(x,y), \widehat{a}_{12}(x,y), \widehat{a}_{21}(x,y)$ and $\widehat{a}_{22}(x,y)$ are obviously independent of each other. They are all linear estimators. Therefore they are asymptotically normally distributed when the sample size tends to infinity. It can be checked that when there are no jumps

in $N_n(x,y)$, their asymptotic means all equal $f(x,y)$ and their asymptotic variances (AV) are

$$AV(\widehat{a}_{s_1s_2}(x,y)) = \frac{\sigma^2}{n^2h_np_n\widetilde{\Delta}^2_{s_1s_2}}\int_{-1/2}^{1/2}\int_{-1/2}^{1/2}(\widetilde{\gamma}^{(1)}_{s_1s_2}+\widetilde{\gamma}^{(2)}_{s_1s_2}u+\widetilde{\gamma}^{(3)}_{s_1s_2}v)^2K^2_{s_1s_2}(u,v)\,dudv, \quad \text{for } s_1,s_2=1,2,$$

where

$$\widetilde{\Delta}_{s_1s_2} = \begin{vmatrix} \widetilde{\beta}^{(00)}_{s_1s_2} & \widetilde{\beta}^{(10)}_{s_1s_2} & \widetilde{\beta}^{(01)}_{s_1s_2} \\ \widetilde{\beta}^{(10)}_{s_1s_2} & \widetilde{\beta}^{(20)}_{s_1s_2} & \widetilde{\beta}^{(11)}_{s_1s_2} \\ \widetilde{\beta}^{(01)}_{s_1s_2} & \widetilde{\beta}^{(11)}_{s_1s_2} & \widetilde{\beta}^{(02)}_{s_1s_2} \end{vmatrix},$$

$$\widetilde{\gamma}^{(1)}_{s_1s_2} = \widetilde{\beta}^{(20)}_{s_1s_2}\widetilde{\beta}^{(02)}_{s_1s_2} - \widetilde{\beta}^{(11)}_{s_1s_2}\widetilde{\beta}^{(11)}_{s_1s_2},\ \widetilde{\gamma}^{(2)}_{s_1s_2} = \widetilde{\beta}^{(01)}_{s_1s_2}\widetilde{\beta}^{(11)}_{s_1s_2} - \widetilde{\beta}^{(10)}_{s_1s_2}\widetilde{\beta}^{(02)}_{s_1s_2},$$

$$\widetilde{\gamma}^{(3)}_{s_1s_2} = \widetilde{\beta}^{(10)}_{s_1s_2}\widetilde{\beta}^{(11)}_{s_1s_2} - \widetilde{\beta}^{(01)}_{s_1s_2}\widetilde{\beta}^{(20)}_{s_1s_2},\ \widetilde{\gamma}^{(4)}_{s_1s_2} = \widetilde{\beta}^{(00)}_{s_1s_2}\widetilde{\beta}^{(02)}_{s_1s_2} - \widetilde{\beta}^{(01)}_{s_1s_2}\widetilde{\beta}^{(01)}_{s_1s_2},$$

$$\widetilde{\gamma}^{(5)}_{s_1s_2} = \widetilde{\beta}^{(01)}_{s_1s_2}\widetilde{\beta}^{(10)}_{s_1s_2} - \widetilde{\beta}^{(00)}_{s_1s_2}\widetilde{\beta}^{(11)}_{s_1s_2},\ \widetilde{\gamma}^{(6)}_{s_1s_2} = \widetilde{\beta}^{(00)}_{s_1s_2}\widetilde{\beta}^{(20)}_{s_1s_2} - \widetilde{\beta}^{(10)}_{s_1s_2}\widetilde{\beta}^{(10)}_{s_1s_2},$$

and

$$\widetilde{\beta}^{(r_1r_2)}_{s_1s_2} = \int_{-1/2}^{1/2}\int_{-1/2}^{1/2}u^{r_1}v^{r_2}K_{s_1s_2}(u,v)\,dudv, \quad \text{for } s_1,s_2=1,2;\ r_1,r_2=0,1,2.$$

Please notice that although $\widetilde{\gamma}^{(4)}_{s_1s_2}, \widetilde{\gamma}^{(5)}_{s_1s_2}$ and $\widetilde{\gamma}^{(6)}_{s_1s_2}$ are not used in the expression of $AV(\widehat{a}_{s_1s_2}(x,y))$, they will be used in Appendix A. So their definitions are also given here.

For simplicity, let us assume that the kernel function $K(\cdot,\cdot)$ is symmetric in the sense that $K(x,y) = K(y,x) = K(-x,y)$ for any $(x,y) \in [-1/2,1/2]\times[-1/2,1/2]$. For example, the commonly used bivariate Epanechnikov kernel function $K(x,y) = \frac{144}{121}(1-x^2)(1-y^2)$ satisfies this condition. It can be checked that $AV(\widehat{a}_{11}(x,y)) = AV(\widehat{a}_{12}(x,y)) = AV(\widehat{a}_{21}(x,y)) = AV(\widehat{a}_{22}(x,y))$ $= \sigma^2\phi^2/(n^2h_np_n)$, where

$$\phi^2 = \frac{1}{\widetilde{\Delta}^2_{11}}\int_{-1/2}^{1/2}\int_{-1/2}^{1/2}(\widetilde{\gamma}^{(1)}_{11}+\widetilde{\gamma}^{(2)}_{11}u+\widetilde{\gamma}^{(3)}_{11}v)^2K^2_{11}(u,v)\,dudv.$$

If we define $\xi_{s_1s_2}(x,y) = (\widehat{a}_{s_1s_2}(x,y) - f(x,y))/(\sigma\phi/nh_n^{1/2}p_n^{1/2})$ for $s_1,s_2 = 1,2$, then $\xi_{11}(x,y)$, $\xi_{12}(x,y)$, $\xi_{21}(x,y)$ and $\xi_{22}(x,y)$ are asymptotically i.i.d. with a common asymptotic distribution $N(0,1)$. Obviously,

$$\begin{aligned} & range(\widehat{a}_{11}(x,y),\widehat{a}_{12}(x,y),\widehat{a}_{21}(x,y),\widehat{a}_{22}(x,y)) \\ = \ & range(\xi_{11}(x,y),\xi_{12}(x,y),\xi_{21}(x,y),\xi_{22}(x,y))\cdot(\sigma\phi/nh_n^{1/2}p_n^{1/2}). \end{aligned}$$

So a natural choice for u_n is

$$u_n = \widehat{\sigma}\phi R_{\alpha_n}/(nh_n^{1/2}p_n^{1/2}), \tag{3.1}$$

Table 3.1: Several quantiles of the limiting distribution of $range(\xi_{11}(x,y),\ \xi_{12}(x,y),\ \xi_{21}(x,y),\ \xi_{22}(x,y))$.

$1-\alpha_n$	R_{α_n}	$1-\alpha_n$	R_{α_n}
0.5000	1.9746	0.9250	3.3959
0.6000	2.2070	0.9500	3.6253
0.7000	2.4596	0.9750	3.9647
0.8000	2.7785	0.9900	4.3911
0.8500	2.9747	0.9990	5.2724
0.9000	3.2314	0.9999	5.8557

where R_{α_n} is the $1-\alpha_n$ quantile of the limiting distribution of $range(\xi_{11}(x,y),\ \xi_{12}(x,y),\ \xi_{21}(x,y),\ \xi_{22}(x,y))$, $0 \le \alpha_n \le 1$ is a significance level and $\widehat{\sigma}$ is a consistent estimator of σ.

A formula for the limiting distribution of $range(\xi_{11}(x,y), \xi_{12}(x,y), \xi_{21}(x,y), \xi_{22}(x,y))$ can be found in Balakrishnan and Cohen (1991). Based on that formula, several values of R_{α_n} are calculated and presented in Table 3.1.

In the literature, there are several existing data-driven bandwidth selection procedures such as plug-in procedures, the cross-validation procedure, the Mallow's C_p criterion and Akaike's information criterion (cf. Loader 1999). Since explicit expressions for the mean and variance of $\widehat{f}_{PLK}(x,y)$ are not available yet, plug-in procedures are not considered in this paper. In the numerical examples presented in Section 4, the bandwidth parameters h_n and p_n are selected by the cross-validation procedure. That is, the optimal values of h_n and p_n are chosen by minimizing the following cross-validation (CV) criterion:

$$CV(h_n, p_n) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(z_{ij} - \widehat{f}_{-i,-j}(x_i, y_j) \right)^2, \tag{3.2}$$

where $\widehat{f}_{-i,-j}(x,y)$ is the "leave-one-out" estimator of $f(x,y)$. Namely, the observation (x_i, y_j, z_{ij}) is left out in constructing $\widehat{f}_{-i,-j}(x,y)$, for $i,j = 1,2,\cdots,n$. We would like to mention that bootstrap techniques may also be used for choosing procedure parameters such as h_n and p_n, although more computation is involved in these resampling techniques.

As discussed at the beginning of this section, the threshold parameter u_n controls the trade-off between noise removal and jump preservation. By using (3.1) and Table 3.1, we know the asymptotic probability that a continuity point is treated as a jump point by procedure (2.4). For example if we use $R_{\alpha_n} = 4.3911$ in (3.1), then asymptotically about one percent of the continuity points are treated as jump points. The consequence of this type of mistake is that the surface

estimator loses some efficiency around those continuity points that are treated as jumps points, which is not serious because the surface estimator is still statistically consistent at those points. The threshold parameter u_n, or equivalently the quantile parameter R_{α_n}, can also be chosen by the cross-validation criterion. Let us re-write the left hand side of (3.2) as $CV(R_{\alpha_n}, h_n, p_n)$. Then the optimal values of R_{α_n}, h_n and p_n can be determined by minimizing $CV(R_{\alpha_n}, h_n, p_n)$.

4 Numerical Examples

We present some numerical examples in three parts. Section 4.1 includes several examples concerning the numerical performance of the proposed procedure (2.4). The procedure (2.4) is compared to some existing jump-preserving surface fitting procedures in Section 4.2. Then the related procedures are all applied to the elevation data of South America in Section 4.3.

4.1 Numerical performance of the procedure (2.4)

We first consider the regression function $f(x,y) = -4(x-0.5)^2 - 4(y-0.5)^2$ when $0 \le x < 0.5$ and $0 \le y < 0.75 - 0.5x$; and $f(x,y) = -4(x-0.5)^2 - 4(y-0.5)^2 + 1$ otherwise. This function has a single JLC: $\{(x,y) : y = 0.75 - 0.5x, 0 \le x \le 0.5\} \bigcup \{(x,y) : x = 0.5, 0 \le y \le 0.5\}$, which has an angle at the point $(0.5, 0.5)$. A set of observations are generated from the model (1.1) with $\epsilon_{11} \sim N(0, \sigma^2)$, $\sigma = 0.25$ and $n = 100$. The true regression function and the observations are presented in Figures 4.1(a) and 4.1(b), respectively.

The jump-preserving surface fitting procedure (2.4) is then applied to the data set shown in Figure 4.1(b). In the procedure, the kernel function is chosen to be the bivariate Epanechnikov function $K(x,y) = \frac{144}{121}(1-x^2)(1-y^2)$ if $(x,y) \in [-1/2, 1/2] \times [-1/2, 1/2]$ and 0 otherwise. The significance level α_n is fixed at 0.01. The bandwidths h_n and p_n are assumed to be equal. By the cross-validation procedure (3.2), they are chosen to be 0.16. The surface fitted by procedure (2.4) is presented in Figure 4.1(c). As a comparison, the surface fitted by the conventional local linear kernel smoothing procedure (1.2) with the same kernel function is presented in Figure 4.1(d). In this conventional procedure, the bandwidth is chosen to be 0.05 by the cross-validation procedure. From the plots, it can be seen that the jumps are preserved reasonably well by the current procedure and they are blurred by the conventional procedure as expected. It can also be noticed that procedure

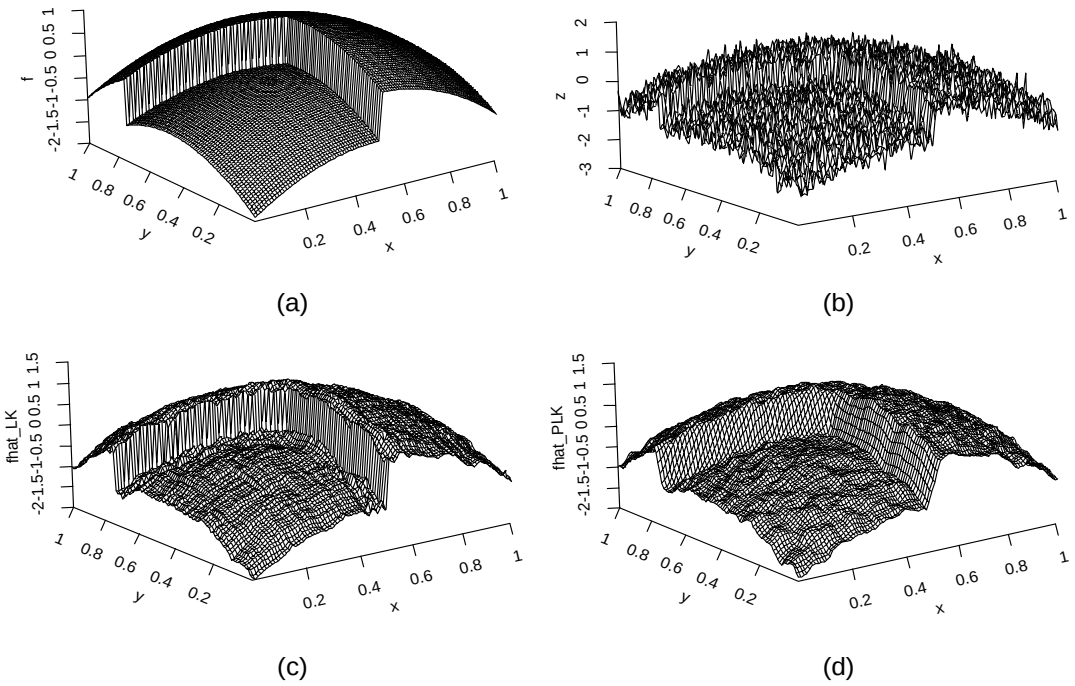

Figure 4.1: (a) The true regression surface; (b) a set of observations with $\sigma = 0.25$ and $n = 100$; (c) the reconstructed surface by the procedure (2.4) with $h_n = p_n = 0.16$; (d) the reconstructed surface by the conventional local linear kernel procedure with $h_n = p_n = 0.05$.

(2.4) performs pretty well in the boundary regions.

The above simulation is then repeated 100 times. The 2.5 and 97.5 percentiles of the 100 surface fits by procedure (2.4) in the cross section of $y = 0.25$ are presented in Figure 4.2(a) by the lower and upper dashed curves, respectively. In the plot, the solid curve represents the cross section of the true surface. The dotted curve is the average of the 100 fits. The corresponding results using the conventional local linear kernel procedure are presented in Figure 4.2(b). It can be seen that the conclusions drawn from Figures 4.1(c) and 4.1(d) are also true here. We can further notice that the conventional estimator is slightly biased in the boundary regions, which is caused by the symmetric "padding" procedure.

In Section 3, we mentioned that the window widths and the quantile parameter R_{α_n} used in (3.1) could be chosen by the cross-validation procedure simultaneously. The next example investigates the accuracy of the selected optimal values of these parameters. In order to see the effect of the JLCs on the selected parameters, we use a more flexible regression function $f(x, y) = -4(x-0.5)^2 - 4(y-0.5)^2$ when $0 \le x < 0.5$ and $0 \le y < c-(2c-1)x$; and $f(x, y) = -4(x-0.5)^2 -$

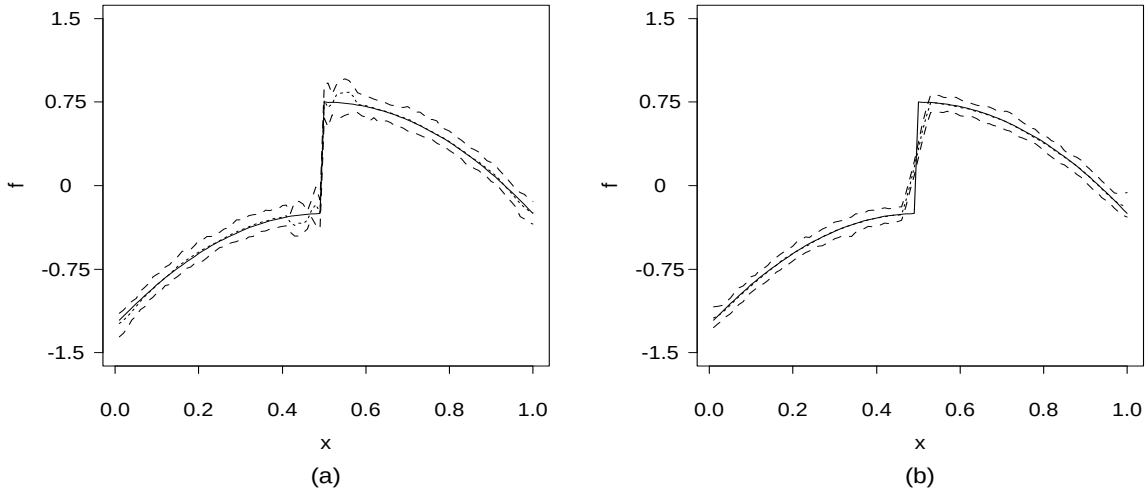

Figure 4.2: (a) The lower and upper dashed curves represent the 2.5 and 97.5 percentiles of the 100 replications of the fitted surface by the procedure (2.4) in the cross section of $y = 0.25$; (b) the corresponding results by the conventional local linear kernel procedure. In the two plots, the solid curves represent the cross section of the true surface and the dotted curves are the averaged fits.

Table 4.1: In each entry, the first line gives the searched values of h_n, R_{α_n} by the CV procedure and the corresponding CV score; the second line gives the corresponding results by the MSE criterion.

c	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.3$	$\sigma = 0.4$	$\sigma = 0.5$
0	.12,4.1,.0161	.16,4.0,.0478	.18,4.1,.1004	.18,4.3,.1738	.20,4.7,.2667
	.12,3.7,.0016	.13,3.4,.0043	.17,3.5,.0082	.17,3.9,.0124	.20,4.6,.0160
.5	.14,4.1,.0108	.20,4.2,.0418	.23,3.9,.0932	.27,4.3,.1648	.30,4.7,.2569
	.14,3.6,.0008	.20,3.9,.0019	.25,3.9,.0034	.26,4.2,.0054	.30,4.7,.0077
1.0	.17,4.6,.0153	.20,4.3,.0467	.22,4.2,.0995	.23,4.2,.1730	.25,4.3,.2668
	.16,4.0,.0008	.18,3.7,.0031	.22,4.0,.0070	.23,3.9,.0114	.23,4.1,.0158

$4(y - 0.5)^2 + 1$ otherwise. It has a single JLC: $\{(x, y) : y = c - (2c - 1)x, 0 \le x \le 0.5\} \bigcup \{(x, y) : x = 0.5, 0 \le y \le 0.5\}$. The value of c controls the angle of the JLC at $(0.5, 0.5)$. When $c = 0.75$, this function is exactly the same as the one presented in Figure 4.1(a). For simplicity, we assume that $h_n = p_n$ as before. Let $n = 100$, c take values 0, 0.5 and 1.0, and σ take values .1, .2, .3, .4 and .5. When $c = .5$, the two JLC segments are parallel to the x and y axes. When $c = 1$ and 0, the JLC has an obtuse and acute angle, respectively, at (.5,.5). For each combination of c and σ, h_n and R_{α_n} are determined by the cross-validation and MSE criteria, respectively, based on 100 replications. The results are presented in Table 4.1.

From Table 4.1, it can be seen that (1) the parameters selected by the cross-validation procedure match those by the MSE criterion reasonably well; (2) when the data are noisier (i.e., σ is larger), the window widths need to be chosen larger; (3) when the curvature of the JLC is larger (i.e., c is smaller in this example), the window widths need to be chosen smaller to accommodate the jumps.

However, when the JLCs are parallel to the x and y axes (the case when $c = .5$), the window widths can still be chosen relatively large because of the square-shaped windows used in (2.4) (cf. Figure 1.1). From this example and our experience, it is often appropriate to choose $(1-\alpha_n) \in [0.9, 0.999]$ (or equivalently, $R_{\alpha_n} \in [3.2314, 5.2724]$).

The surface estimator (2.4) is based on local linear kernel estimation, which accommodates the effect of the surface slope on the estimator. But the curvature of the true surface might still have some effect on the estimator. To investigate this issue, let us consider the following regression function: $f(x,y) = d(x-0.5)^2 + d(y-0.5)^2$ when $0 \le x < 0.5$ and $0 \le y < 0.75 - 0.5x$; and $f(x,y) = d(x-0.5)^2 + d(y-0.5)^2 + 1$ otherwise, where d can take values -1,-2,-3,-4 and -5. When $d = -4$, this function is the same as the one in Figure 4.1(a). Obviously $f''_{xx}(x,y) = f''_{yy}(x,y) = 2d$ and $f''_{xy}(x,y) = 2d(x+y-1)$ when (x,y) is a continuity point. Therefore the curvature of f is larger when $|d|$ is larger. Suppose that $n = 100$ and $\sigma = 0.25$. For each d value, the window width h_n $(= p_n)$ and the quantile parameter R_{α_n} are determined by the cross-validation procedure. The corresponding MSE values are presented in Figure 4.3. It can be seen that the curvature of f does have some effect on the estimated surface. But the effect is small: the MSE value only increases about 10 percent when d changes from -1 to -5.

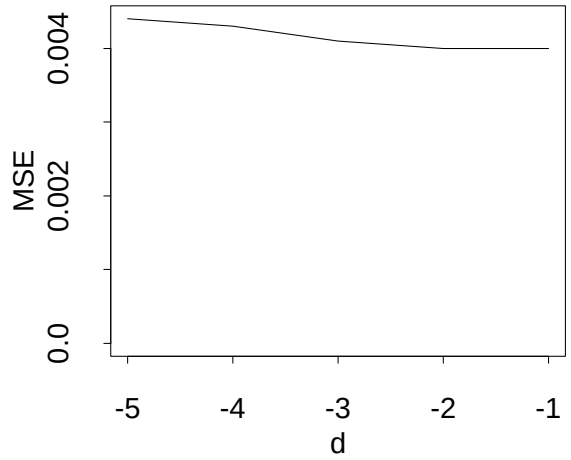

Figure 4.3: The MSE values of the estimated surfaces when d changes from -1 to -5.

4.2 Some numerical comparisons

In this part, we compare the current procedure with three existing procedures in some numerical examples. The existing procedures considered are the MRF image reconstruction procedure

suggested by Godtliebsen and Sebastiani (1994), the local median smoothing procedure and the discrete wavelet transformation procedure included in the R package wavethresh (Nason and Silverman 1994). The MRF procedure has three positive parameters α, β and λ. Godtliebsen and Sebastiani (1994) showed that it performed better than some well-known image reconstruction procedures in some cases. The surface estimator of the local median smoothing procedure is defined by the sample median of the observations in a neighborhood of a given point. Because it is robust to outliers and has some ability to preserve the jumps, it is widely used in industry (cf. Gonzalez and Woods 1992, Chapter 4). The discrete wavelet transformation procedure requires that several parameters be determined before it can be used. In the numerical examples below, we use the default family of wavelets (which is Daubechies' "extremal phase" wavelet) and the "symmetric" boundary handling condition. The parameter "filter.number", which determines the regularity of the wavelet, can vary from 1 to 10. The thresholding "policy" is either "hard" or "soft". The "levels" to be thresholded could be $r : s$ where 2^{s+1} is the sample size and r is an integer number ranging from 1 to s.

The true regression function considered is the one used in the example of Table 4.1. The value of c is either 0.75 or 0.25. The angle of the JLC at the point $(0.5, 0.5)$ is obtuse when $c = 0.75$ and acute when $c = 0.25$. Since the discrete wavelet transformation procedure can only handle the case where n is a power of 2, n is chosen to be 128 (i.e., 2^7). The value of σ can vary among 0.1, 0.25, 0.4 and 0.75. The noise level of the data is low when $\sigma = 0.1$, it is moderate when σ is 0.25 or 0.4, and it is considered high when $\sigma = 0.75$. In each case, the optimal parameter values of the four procedures are determined by searches based on the criterion of MSE and 100 replications. For the PLK and the local median smoothing procedures, it is assumed that $h_n = p_n$ and the optimal value of h_n is determined by searching among all values of k/n where $1 < k < n$ is an odd integer number. This search is almost exhaustive because the MSE value when $(k-2)/n < h_n \leq k/n$ is about the same as its value when $h_n = k/n$. For the PLK procedure, the optimal values of R_{α_n} and h_n are searched simultaneously. The optimal value of R_{α_n} is searched in steps of 0.1. For the MRF procedure, its parameters α, β and λ are also searched in steps of 0.1.

The optimal MSE values of the four procedures when c is 0.75 and 0.25 are presented in Figures 4.4(a) and 4.4(b), respectively. From Figure 4.4(a), it can be seen that the current procedure outperforms the local median smoothing procedure and the discrete wavelet transformation procedue uniformly. It performs better than the MRF procedure when the noise level is moderate to high.

The MRF procedure performs the best when the noise level is low. Similar conclusions can be drawn from Figure 4.4(b).

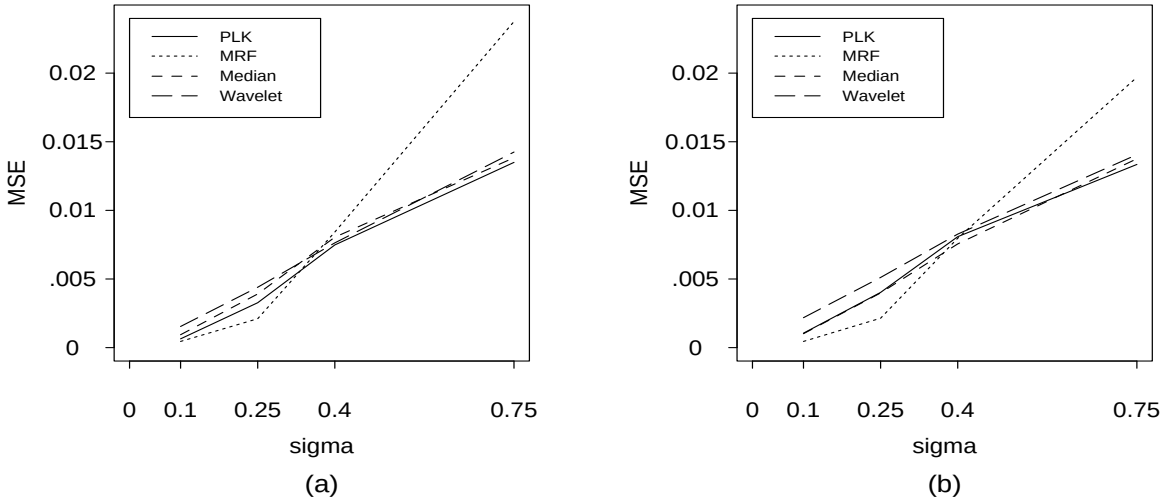

Figure 4.4: (a) The MSE values of the four procedures when $c = 0.75$; (b) the corresponding results when $c = 0.25$.

4.3 Application to the elevation data of South America

Figure 4.5(a) shows elevation data for South America, contaminated by i.i.d. random noise with distribution $N(0, 1000^2)$. The resolution of the data is 1 degree $\times$ 1 degree in latitude and longitude and the sample size is $64^2 = 4096$. Because of the low resolution of the data, jumps exist between land and sea, especially along the west coast of South America which is the famous Andes mountain area.

We first applied the MRF procedure discussed in the previous part to the contaminated elevation data. Since the cross-validation procedure is not appropriate for choosing the parameters α, β and λ, we tried many combinations of these parameters and the reconstructed surface with the best visual impression is presented in Figure 4.5(b). From the plot, it can be seen that some noise is removed by this procedure. But the jumps along the west coast of South America are slightly blurred. Then the local median smoothing procedure is applied to the same data. By the cross-validation procedure, its window width is chosen to be 0.0625 and the reconstructed surface is shown in Figure 4.5(c). It can be seen from the plot that the noise is removed well but the jumps are mostly blurred. The discrete wavelet transformation procedure is then used. Since the cross-validation procedure is not appropriate in this case either, its parameters (i.e., "filter.number",

"levels" and "policy") are chosen for good visual impression. The fitted surface is presented in Figures 4.5(d). We can see that the estimated surface is quite noisy and the jumps along the west coast of South America are not preserved well either. Next we apply procedure (2.4) to the dataset with the same kernel function and significance level as those in the example of Figure 4.1. The window widths are determined to be $h_n = p_n = 0.1875$ (it is assumed that $h_n = p_n$ as before). Its estimated surface is displayed in Figures 4.5(e). We can see that the jumps along the west coast of South America are preserved reasonably well by this procedure and the noise is removed well too.

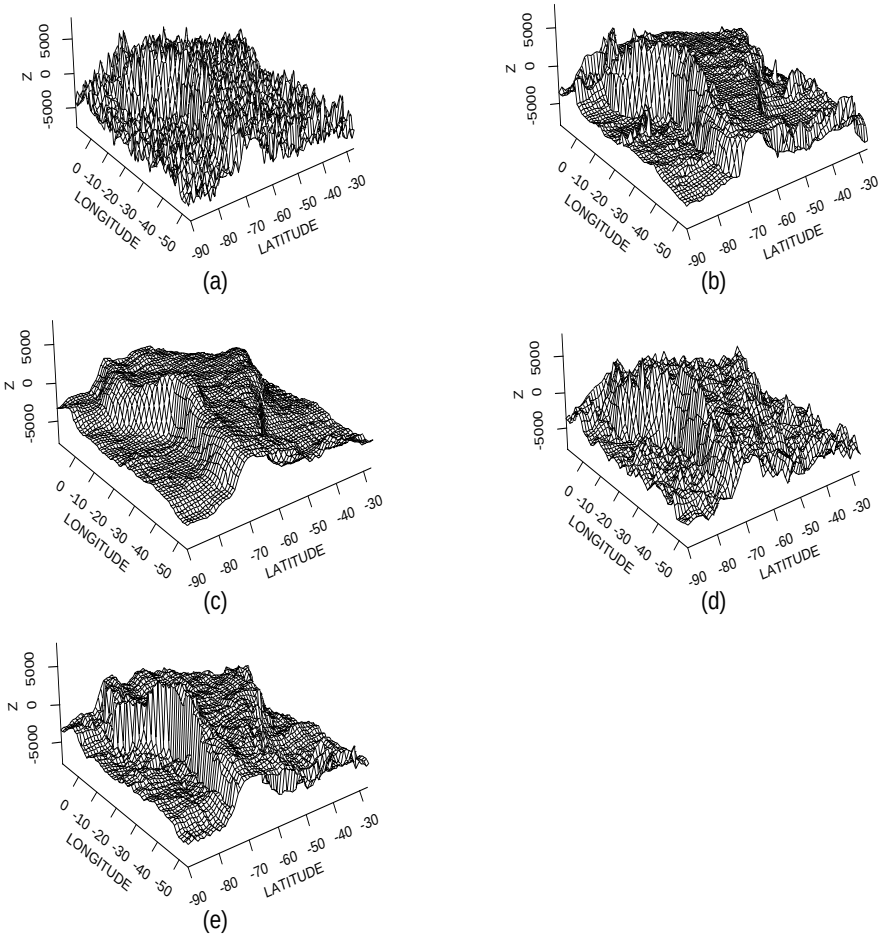

Figure 4.5: (a) The noisy elevation data of South America; (b)-(e) the reconstructed surfaces by the MRF procedure, the local median smoothing procedure, the discrete wavelet transformation procedure and the procedure (2.4), respectively. In the plots, negative latitudes denote latitudes west and negative longitudes denote logitudes south.

5 Concluding Remarks

We have presented a procedure for modifying conventional local smoothing procedures such that the modified procedures can estimate regression surfaces with possible jumps preserved. Although the local linear kernel smoothing procedure is our focus in the discussion, other conventional local smoothing procedures can be handled in a similar way. One advantage of the current procedure is that its surface estimator is expressed explicitly by a mathematical formula like most conventional local smoothing estimators. Therefore it is convenient to use and easy to compute. Another advantage of the current procedure is that it requires mild conditions on the model. It has been shown that all types of jumps except those around the singular points of the JLCs can be preserved well by this procedure. Further improvements of the current procedure are possible. For example, when (2.3) is true, other types of averages (other than the simple average used in (2.4)) of $\widehat{a}_{11}(x,y), \widehat{a}_{12}(x,y), \widehat{a}_{21}(x,y)$ and $\widehat{a}_{22}(x,y)$ could be used for estimating $f(x,y)$ if the extra computation involved is not a big issue. When (2.3) is not true, it is possible to use more than one of $\widehat{a}_{11}(x,y), \widehat{a}_{12}(x,y), \widehat{a}_{21}(x,y)$ and $\widehat{a}_{22}(x,y)$ for defining $\widehat{f}_{PLK}(x,y)$. To this end, we need a data-driven machenism to distinguish the three cases demonstrated by Figure 2.1, which might be a challenging task.

Acknowledgements We would like to thank the editor, the associate editor and two referees for some constructive comments and suggestions which greatly improved the article.

Appendix

A Some Statistical Properties Of The Procedure (2.4)

In this appendix, we discuss some properties of the fitted surface of the jump-preserving surface fitting procedure (2.4). First, we have the following result.

Theorem A.1 Suppose that the regression function $f(\cdot,\cdot)$ has continuous second-order derivatives in the design space $[0,1]\times[0,1]$; the kernel function $K(\cdot,\cdot)$ is Lipschitz (1) continuous; $E(\epsilon_{11}^2)<\infty$; $h_n=O(n^{-1/3})$; and $p_n=O(n^{-1/3})$. Then

$$\frac{n^{2/3}}{\log n}\|\widehat{a}_{11}-f\|_{[h_n/2,1]\times[p_n/2,1]}=O(1),\ \ a.s.;\ \ \frac{n^{2/3}}{\log n}\|\widehat{a}_{12}-f\|_{[h_n/2,1]\times[0,1-p_n/2]}=O(1),\ \ a.s.;$$

$$\frac{n^{2/3}}{\log n}\|\widehat{a}_{21}-f\|_{[0,1-h_n/2]\times[p_n/2,1]}=O(1),\ \ a.s.;\ \ \frac{n^{2/3}}{\log n}\|\widehat{a}_{22}-f\|_{[0,1-h_n/2]\times[0,1-p_n/2]}=O(1),\ \ a.s.$$

where $\|f\|_\Omega=\max_{(x,y)\in\Omega}|f(x,y)|$.

The above theorem establishes the uniformly strong consistency of the four estimators $\widehat{a}_{11}(x,y)$, $\widehat{a}_{12}(x,y)$, $\widehat{a}_{21}(x,y)$ and $\widehat{a}_{22}(x,y)$ of $f(x,y)$ when the true regression function f is continuous in the entire design space. It can be proved by some arguments similar to those in the proof of Theorem 2.1 of Qiu (1997). Based on this theorem, the four estimators should be all close to $f(x,y)$ when n is large enough and the point (x,y) is located in a continuous region of f. Consequently $\widehat{f}_{PLK}(x,y)$ is close to $f(x,y)$ in such a case. A formal statement of this result is included in Theorem A.3 below. When the point (x,y) is around the JLCs, we have the following result about $RSS(\widehat{a}_{11}(x,y))$.

Theorem A.2 In model (1.1), suppose that the regression function f has continuous second-order derivatives in each of its continuous regions in the design space $[0,1]\times[0,1]$; $E(\epsilon_{11}^4)<\infty$; the kernel function $K(\cdot,\cdot)$ is Lipschitz (1) continuous; $h_n=O(n^{-1/3})$; and $p_n=O(n^{-1/3})$. For a given point $(x,y)\in[h_n/2,1]\times[p_n/2,1]$, if there are no jumps in $Q_{11}(x,y)=[x-h_n/2,x)\times[y-p_n/2,y)$, then

$$RSS(\widehat{a}_{11}(x,y))=\widetilde{\beta}_{11}^{(00)}\sigma^2n^2h_np_n+o(n^2h_np_n),\ \ a.s.; \tag{A.1}$$

if there is a single JLC in $Q_{11}(x,y)$, then

$$RSS(\widehat{a}_{11}(x,y))=(\widetilde{\beta}_{11}^{(00)}\sigma^2+d_\tau^2C_\tau^2)n^2h_np_n+o(n^2h_np_n),\ \ a.s., \tag{A.2}$$

where $d_\tau > 0$ denotes the jump magnitude of f at the point (x_τ, y_τ) which is a point on the JLC in $Q_{11}(x,y)$ that is closest to (x,y),

$$
\begin{aligned}
C_\tau^2 = \; & \tfrac{1}{\widetilde{\Delta}_{11}^2} \textstyle\int\int_{Q_{11}^{(1)}(0,0)} \{ \quad \int\int_{Q_{11}^{(2)}(0,0)} (\widetilde{\gamma}_{11}^{(1)} + \widetilde{\gamma}_{11}^{(2)} u + \widetilde{\gamma}_{11}^{(3)} v) K_{11}(u,v)\, dudv + \\
& \qquad x \int\int_{Q_{11}^{(2)}(0,0)} (\widetilde{\gamma}_{11}^{(2)} + \widetilde{\gamma}_{11}^{(4)} u + \widetilde{\gamma}_{11}^{(5)} v) K_{11}(u,v)\, dudv + \\
& \qquad y \int\int_{Q_{11}^{(2)}(0,0)} (\widetilde{\gamma}_{11}^{(3)} + \widetilde{\gamma}_{11}^{(5)} u + \widetilde{\gamma}_{11}^{(6)} v) K_{11}(u,v)\, dudv \}^2 K_{11}(x,y)\, dxdy + \\
& \tfrac{1}{\widetilde{\Delta}_{11}^2} \textstyle\int\int_{Q_{11}^{(2)}(0,0)} \{ \quad \int\int_{Q_{11}^{(1)}(0,0)} (\widetilde{\gamma}_{11}^{(1)} + \widetilde{\gamma}_{11}^{(2)} u + \widetilde{\gamma}_{11}^{(3)} v) K_{11}(u,v)\, dudv + \\
& \qquad x \int\int_{Q_{11}^{(1)}(0,0)} (\widetilde{\gamma}_{11}^{(2)} + \widetilde{\gamma}_{11}^{(4)} u + \widetilde{\gamma}_{11}^{(5)} v) K_{11}(u,v)\, dudv + \\
& \qquad y \int\int_{Q_{11}^{(1)}(0,0)} (\widetilde{\gamma}_{11}^{(3)} + \widetilde{\gamma}_{11}^{(5)} u + \widetilde{\gamma}_{11}^{(6)} v) K_{11}(u,v)\, dudv \}^2 K_{11}(x,y)\, dxdy,
\end{aligned}
$$

$Q_{11}^{(1)}(x,y)$ and $Q_{11}^{(2)}(x,y)$ denote two different parts of $Q_{11}(x,y)$ separated by the JLC with a positive jump at (x_τ, y_τ) from $Q_{11}^{(1)}(x,y)$ to $Q_{11}^{(2)}(x,y)$, $Q_{11}^{(1)}(0,0) = \{(u,v) : (uh_n + x, vp_n + y) \in Q_{11}^{(1)}(x,y)\}$ and $Q_{11}^{(2)}(0,0) = \{(u,v) : (uh_n + x, vp_n + y) \in Q_{11}^{(2)}(x,y)\}$.

Outline of the Proof We first prove the equation (A.1) in the case that there are no jumps in $Q_{11}(x,y)$. By the definition of $RSS(\widehat{a}_{11}(x,y))$, it can be written as

$$
\begin{aligned}
& RSS(\widehat{a}_{11}(x,y)) \\
= \; & \sum_{i=1}^n \sum_{j=1}^n \epsilon_{ij}^2 K_{11}(\frac{x_i - x}{h_n}, \frac{y_j - y}{p_n}) + \\
& 2 \sum_{i=1}^n \sum_{j=1}^n \epsilon_{ij} [f(x_i, y_j) - \widehat{a}_{11}(x,y) - \widehat{b}_{11}(x,y)(x_i - x) - \widehat{c}_{11}(x,y)(y_j - y)] K_{11}(\frac{x_i - x}{h_n}, \frac{y_j - y}{p_n}) + \\
& \sum_{i=1}^n \sum_{j=1}^n [f(x_i, y_j) - \widehat{a}_{11}(x,y) - \widehat{b}_{11}(x,y)(x_i - x) - \widehat{c}_{11}(x,y)(y_j - y)]^2 K_{11}(\frac{x_i - x}{h_n}, \frac{y_j - y}{p_n}) \\
=: \; & I_1 + I_2 + I_3
\end{aligned}
$$

It can be checked that

$$
I_1 = \sigma^2 \widetilde{\beta}_{11}^{(00)} n^2 h_n p_n + o(n^2 h_n p_n), \; a.s.. \tag{A.3}
$$

By the Taylor's expansion,

$$
\begin{aligned}
I_2 \; = \; & 2(f(x,y) - \widehat{a}_{11}(x,y)) \sum_{i=1}^n \sum_{j=1}^n \epsilon_{ij} K_{11}(\frac{x_i - x}{h_n}, \frac{y_j - y}{p_n}) + \\
& 2(f_x'(x,y) - \widehat{b}_{11}(x,y)) \sum_{i=1}^n \sum_{j=1}^n \epsilon_{ij} (x_i - x) K_{11}(\frac{x_i - x}{h_n}, \frac{y_j - y}{p_n}) +
\end{aligned}
$$

$$
\begin{aligned}
& 2(f'_y(x,y) - \widehat{c}_{11}(x,y)) \sum_{i=1}^{n}\sum_{j=1}^{n} \epsilon_{ij}(y_j - y) K_{11}(\frac{x_i - x}{h_n}, \frac{y_j - y}{p_n}) + o(n^2 h_n p_n) \\
= \;& o(n^2 h_n p_n), \; a.s. \qquad (A.4)
\end{aligned}
$$

Similarly it can be checked that

$$
I_3 = o(n^2 h_n p_n), \; a.s. \qquad (A.5)
$$

After combining (A.3)-(A.5), the equation (A.1) is proved.

Next we assume that there is a JLC in $Q_{11}(x,y)$ which does not have any singular points in $Q_{11}(x,y)$. The JLC devides $Q_{11}(x,y)$ into two parts: $Q_{11}^{(1)}(x,y)$ and $Q_{11}^{(2)}(x,y)$. The limits of f at (x_τ, y_τ) from $Q_{11}^{(1)}(x,y)$ and $Q_{11}^{(2)}(x,y)$ are denoted by $f_-(x_\tau,y_\tau)$ and $f_+(x_\tau,y_\tau)$, respectively. Then it can be checked that (A.3) and (A.4) are still true and

$$
\begin{aligned}
E(\widehat{a}_{11}(x,y)) &= \frac{f_-(x_\tau,y_\tau)}{\widetilde{\Delta}_{11}} \int\int_{Q_{11}^{(1)}(0,0)} (\widetilde{\gamma}_{11}^{(1)} + \widetilde{\gamma}_{11}^{(2)} x + \widetilde{\gamma}_{11}^{(3)} y) K_{11}(x,y) \, dxdy + \\
& \quad \frac{f_+(x_\tau,y_\tau)}{\widetilde{\Delta}_{11}} \int\int_{Q_{11}^{(2)}(0,0)} (\widetilde{\gamma}_{11}^{(1)} + \widetilde{\gamma}_{11}^{(2)} x + \widetilde{\gamma}_{11}^{(3)} y) K_{11}(x,y) \, dxdy + o(1); \qquad (A.6) \\
E(\widehat{b}_{11}(x,y)) &= \frac{f_-(x_\tau,y_\tau)}{\widetilde{\Delta}_{11} h_n} \int\int_{Q_{11}^{(1)}(0,0)} (\widetilde{\gamma}_{11}^{(2)} + \widetilde{\gamma}_{11}^{(4)} x + \widetilde{\gamma}_{11}^{(5)} y) K_{11}(x,y) \, dxdy + \\
& \quad \frac{f_+(x_\tau,y_\tau)}{\widetilde{\Delta}_{11} h_n} \int\int_{Q_{11}^{(2)}(0,0)} (\widetilde{\gamma}_{11}^{(2)} + \widetilde{\gamma}_{11}^{(4)} x + \widetilde{\gamma}_{11}^{(5)} y) K_{11}(x,y) \, dxdy + o(1/h_n); \qquad (A.7) \\
E(\widehat{c}_{11}(x,y)) &= \frac{f_-(x_\tau,y_\tau)}{\widetilde{\Delta}_{11} p_n} \int\int_{Q_{11}^{(1)}(0,0)} (\widetilde{\gamma}_{11}^{(3)} + \widetilde{\gamma}_{11}^{(5)} x + \widetilde{\gamma}_{11}^{(6)} y) K_{11}(x,y) \, dxdy + \\
& \quad \frac{f_+(x_\tau,y_\tau)}{\widetilde{\Delta}_{11} p_n} \int\int_{Q_{11}^{(2)}(0,0)} (\widetilde{\gamma}_{11}^{(3)} + \widetilde{\gamma}_{11}^{(5)} x + \widetilde{\gamma}_{11}^{(6)} y) K_{11}(x,y) \, dxdy + o(1/p_n). \qquad (A.8)
\end{aligned}
$$

By combining the equations (A.6)-(A.8), we have

$$
I_3 = d_\tau^2 C_\tau^2 n^2 h_n p_n \qquad (A.9)
$$

Then the equation (A.2) is obtained after combining (A.3), (A.4) and (A.9).

Similar results can be derived for $\widehat{a}_{12}(x,y), \widehat{a}_{21}(x,y)$ and $\widehat{a}_{22}(x,y)$. Based on Theorems A.1 and A.2, the strong consistency of $\widehat{f}_{PLK}$ is established in the following theorem.

Theorem A.3 If R_{α_n} defined in (3.1) satisfies the conditions that (i) $\lim_{n\to\infty} R_{\alpha_n} n^{-2/3} = 0$ and (ii) $\lim_{n\to\infty} R_{\alpha_n} (\log n)^{-1} = \infty$, then under the assumptions in Theorem A.2, we have

(i) $\|\widehat{f}_{PLK} - f\|_{\mathcal{D}_\varepsilon} = O(n^{-2/3} \log n), \; a.s.$;

(ii) for any $(x,y) \in \mathcal{D}_\varepsilon^*$, $|\widehat{f}_{PLK}(x,y) - f(x,y)| = O(n^{-2/3}\log n),\ a.s.$,

where $\varepsilon > 0$ is an arbitrarily small number, $\mathcal{D}_\varepsilon = \{(x,y) : \sqrt{(x-x^*)^2+(y-y^*)^2} \geq \varepsilon$, where (x^*,y^*) is any point on the JLCs or on the border of the design space$\}$, $\mathcal{D}_\varepsilon^* = L_\varepsilon \backslash S_\varepsilon$, $L_\varepsilon = \{(x,y) : \sqrt{(x-x^*)^2+(y-y^*)^2} < \varepsilon,$ where (x^*,y^*) is any point on the JLCs$\}$ and $S_\varepsilon = \{(x,y) : \sqrt{(x-x^*)^2+(y-y^*)^2} < \varepsilon$, where (x^*,y^*) is any singular point of the JLCs or a point on the border of the design space$\}$.

Proof For any point $(x,y) \in \mathcal{D}_\varepsilon$, if $\lim_{n\to\infty} R_{\alpha_n}(\log n)^{-1} = \infty$, then when n is large enough we have

$$range(\widehat{a}_{11}(x,y), \widehat{a}_{12}(x,y), \widehat{a}_{21}(x,y), \widehat{a}_{22}(x,y)) < u_n. \tag{A.10}$$

By (2.4) and Theorem A.1, we have

$$\frac{n^{2/3}}{\log n}|\widehat{f}_{PLK}(x,y) - f(x,y)| = O(1),\ a.s. \tag{A.11}$$

Since the expressions (A.10) and (A.11) are uniformly true with regard to $(x,y) \in \mathcal{D}_\varepsilon$,

$$\frac{n^{2/3}}{\log n}\|\widehat{f}_{PLK} - f\|_{\mathcal{D}_\varepsilon} = O(1),\ a.s.$$

When $(x,y) \in \mathcal{D}_\varepsilon^*$, we only conside the case that the point (x,y) is on a JLC and the two tangent lines of the JLC at (x,y) are located in $Q_{11}(x,y)$ and $Q_{12}(x,y)$, respectively. Other cases can be discussed similarly. Then

$$E(\widehat{a}_{11}(x,y)) = f_-(x,y) + \frac{d(x,y)}{\widetilde{\Delta}_{11}} \int\int_{Q_{11}^{(2)}(0,0)} (\widetilde{\gamma}_{11}^{(1)} + \widetilde{\gamma}_{11}^{(2)}x + \widetilde{\gamma}_{11}^{(3)}y) K_{11}(x,y)\, dxdy + o(h_n^2),$$

where $d(x,y)$ is the jump size of f at (x,y). Similarly, $E(\widehat{a}_{22}(x,y)) = f(x,y) + O(h_n^2)$. Therefore

$$range(\widehat{a}_{11}(x,y), \widehat{a}_{12}(x,y), \widehat{a}_{21}(x,y), \widehat{a}_{22}(x,y)) \geq |\widehat{a}_{11}(x,y) - \widehat{a}_{22}(x,y)| > u_n$$

under the condition that $\lim_{n\to\infty} R_{\alpha_n} n^{-2/3} = 0$. Therefore $\widehat{f}_{PLK}(x,y) = \tilde{a}(x,y)$ by (2.4). By Theorem A.2, $\tilde{a}(x,y)$ must be one of $\widehat{a}_{21}(x,y)$ and $\widehat{a}_{22}(x,y)$ when n is large enough because their RSS values are smaller than the RSS values of $\widehat{a}_{11}(x,y)$ and $\widehat{a}_{12}(x,y)$ in such a case. By Theorem A.1, we have $|\widehat{f}_{PLK}(x,y) - f(x,y)| = O(n^{-2/3}\log n),\ a.s.$ This finishes the proof.

Theorem A.3 says that the fitted surface of the procedure (2.4) is uniformly strongly consistent in the continuous regions of f and the convergence rate is the same as that of the conventional local linear kernel estimator. It is also pointwisely strongly consistent around the JLCs.

References

Balakrishnan, N., and Cohen, A.C. (1991), *Order Statistics and Inference: Estimation Methods*, Academic Press: Boston.

Besag, J. (1986), "On the statistical analysis of dirty pictures (with discussion)", *Journal of the Royal Statistical Society - B,* **48**, 259-302.

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995), "Bayesian computation and stochastic systems (with discussion)", *Statistical Science 10*, 3-66.

Chu, C.K., Glad, I.K., Godtliebsen, F., and Marron, J.S. (1998), "Edge-preserving smoothers for image processing (with discussion)", *Journal of the American Statistical Association*, **93**, 526-556.

Cleveland, W.S. (1979), "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association*, **74**, 828-836.

Donoho, D.L., and Johnstone, I.M. (1994), "Ideal spatial adaptation by wavelet shrinkage," *Biometrika 81*, 425–455.

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications,* Chapman & Hall: London.

Fessler, J.A., Erdogan, H., and Wu, W.B. (2000), "Exact distribution of edge-preserving MAP estimators for linear signal models with Gaussian measurement noise," *IEEE Transactions on image processing 9*, 1049-1055.

Geman, S., and Geman, D. (1984), "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **6**, 721-741.

Glasbey, C., and Horgan, G. (1995), *Image Analysis for the Biological Sciences*, John Wiley & Sons: New York.

Godtliebsen, F., and Sebastiani, G. (1994), "Statistical methods for noisy images with discontinuities," *Journal of Applied Statistics*, **21**, 459-476.

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

Hall, P., Peng, L., and Rau, C. (2001), "Local likelihood tracking of fault lines and boundaries," *Journal of the Royal Statistical Society - B(63)*, 569-582.

Hall, P., and Rau, C. (2000), "Tracking a smooth fault line in a response surface", *The Annals of Statistics*, **28**, 713-733.

Härdle, W. (1990), *Applied Nonparametric regression*, Oxford University Press.

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

Li, S.Z. (1995), *Markov random field modeling in computer vision,* New York: Springer-Verlag.

Loader, C.R. (1999), "Bandwidth selection: Classical or plug-in?", *The Annals of Statistics*, **27**, 415-438.

Marroquin, J.L., Velasco, F.A., Rivera, M., and Nakamura, M. (2001), "Gauss-Markov measure field models for low-level vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence 23*, 337-347.

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

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

Nason, G., and Silverman, B. (1994), "The discrete wavelet transform in S," *Journal of Computational and Graphical Statistics 3,* 163-191.

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.

Owen, A. (1984), "A neighborhood-based classifier for Landsat data," *Canadian Journal of Statistics*, **12**, 191-200.

Polzehl, J., and Spokoiny, V.G. (2000), "Adaptive weights smoothing with applications to image restoration," *Journal of the Royal Statistical Society - B*, **62**, 335-354.

Qiu, P. (1997), "Nonparametric estimation of jump surface," *Sankhyā (A)*, **59**, 268-294.

Qiu, P. (1998), "Discontinuous regression surfaces fitting," *The Annals of Statistics*, **26**, 2218-2245.

Qiu, P. (2002), "A nonparametric procedure to detect jumps in regression surfaces," *Journal of Computational and Graphical Statistics*, **11**, 799-822.

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*, **6**, 332-354.

Ruppert, D., and Wand, M.P. (1994), "Multivariate locally weighted least squares regression," *The Annals of Statistics*, **22**, 1346-1370.

Switzer, P. (1983), "Some spatial statistics for the interpretation of satellite data", *Bulletin of the International Statistical Institute*, **50**, Book 2, 962-972.

Switzer, P., Kowalik, W.S., and Lyon, R.J.P. (1982), "A prior method for smoothing discriminant analysis classification maps", *Journal of the International Association for Mathematical Geology,* **14**, 433-444.

Tukey, J.W. (1977), *Exploratory Data Analysis,* Reading, MA: Addison-Wesley.

Wang, Y. (1998), "Change curve estimation via wavelets," *Journal of the American Statistical Association*, **93**, 163-172.