Jump-Preserving Surface Reconstruction From Noisy Data

(Short title: Jump-Preserving Surface Reconstruction)

Peihua Qiu
School of Statistics
University of Minnesota
Minneapolis, MN 55455

Abstract

A new local smoothing procedure is suggested for jump-preserving surface reconstruction from noisy data. In a neighborhood of a given point in the design space, a plane is fitted by local linear kernel smoothing, giving the conventional local linear kernel estimator of the surface at the point. The neighborhood is then divided into two parts by a line passing through the given point and perpendicular to the gradient direction of the fitted plane. In the two parts, two half planes are fitted, respectively, by local linear kernel smoothing, providing two one-sided estimators of the surface at the given point. Our surface reconstruction procedure then proceeds in the following two steps. First, the fitted surface is defined by one of the three estimators, i.e., the conventional estimator and the two one-sided estimators, depending on the weighted residual means of squares of the fitted planes. The fitted surface of this step preserves the jumps well, but it is a bit noisy, compared to the conventional local linear kernel estimator. Second, the estimated surface values at the original design points obtained in the first step are used as new data, and the above procedure is applied to this data in the same way except that one of the three estimators is selected based on their estimated variances. Theoretical justification and numerical examples show that the fitted surface of the second step preserves jumps well and also removes noise efficiently. Besides two window widths, this procedure does not introduce other parameters. Its surface estimator has an explicit formula. All these features make it convenient to use and simple to compute.

Key Words: Denoising; Edge detection; Image reconstruction; Gradient; Local linear kernel estimation; Local smoothing; Neighborhood; One-sided estimators; Surface estimation.

1 Introduction

Local smoothing techniques are popular in applications for reconstructing regression curves and surfaces from noisy data. Conventional local smoothing procedures, including running averages (Tukey (1977)), locally weighted scatter plot smoothing (Cleveland (1979)), kernel smoothing (Härdle

(1990)), local polynomial kernel smoothing (Fan and Gijbels (1996)), and several others, are appropriate for estimating continuous regression functions. When the underlying regression function has jumps, the estimated functions by conventional procedures are not statistically consistent at the jump positions. However, the problem of estimating jump regression functions is important because true regression functions are often discontinuous in applications. For example, the image intensity function of an image is discontinuous at the outlines of objects, and equi-temperature surfaces in high sky or deep ocean are often discontinuous too. This article introduces a new jump-preserving surface reconstruction procedure, which has good statistical properties and is convenient to use.

A main reason why most conventional surface estimators are not statistically consistent around the jumps, I think, is that: a *continuous* function has been used by these procedures for estimating a *jump* function. To preserve jumps while estimating a surface, the estimation procedure should adapt itself to the jump structure of the surface, which is a major challenge in this research problem, because the jump structure is unobservable and it is often too complicated to be expressed by any mathematical formulas. In the image processing literature, most Bayesian estimation methods use a so-called *line process* to denote the unobservable jump structure of the image intensity surface (Geman and Geman (1984)). Each element in this process is binary and denotes the existence of jumps between two neighboring pixels, with 1 denoting presence of jumps and 0 absence. The true image intensity function and the line process are then assumed to form a joint Markov random field (MRF). Under the assumption that observed image intensities follow a known distribution (e.g., Normal distribution), the true image is then estimated by maximizing *a posteriori* (MAP). The MAP process is accomplished computationally by using the stochastic relaxation and simulated annealing. This kind of procedures, however, require expensive computation, although some simplifications and modifications exist (e.g., Besag (1986), Besag *et al.* (1995), Fessler *et al.* (2000), Godtliebsen and Sebastiani (1994), Li (1995), Marroquin *et al.* (2001), Sebastiani and Godtliebsen (1997), Titterington (1985), and Yi and Chelberg (1995)). Due to their iterative nature, it is also difficult to study their statistical properties, including the consistency of the reconstructed images.

Another existing jump surface estimation method is based on bilateral filtering (Tomasi and Manduchi (1998)). Its major idea is that: observations which are more different from the observation at a given point should receive less weights in the local averaging for estimating the surface at the given point, because it is more likely that they are located on a different side of the related jump location curve (JLC), relative to the given point. This strategy is accomplished by using two kernel functions: one for pixel locations, and the other one for observations of the response. Because observations on the different side of the JLC, relative to the given point, still receive *some* weights, although such weights are generally small, some blurring may still happen around the

jumps, using this method. Furthermore, blurring can also be generated by large surface slopes, because in such cases observations on the different side of the JLC, relative to the given point, could be even closer to the observation at the given point, compared to some observations on the same side. A closely related procedure to this bilateral filtering procedure is the adaptive weights smoothing (AWS) algorithm, suggested by Polzehl and Spokoiny (2000). In the AWS algorithm, bilateral filtering is performed iteratively. Numerical studies show that it performs well when the true surface has large homogeneous regions. Other related methods in the literature include the nonlinear diffusion filter by Perona and Malik (1990) (also see Keeling and Stollberger (2002) and Weickert *et al.* (1998) for modifications and generalizations), the sigma filter by Chu *et al.* (1998), the three-stage procedure by Qiu (1998), and several others. See Qiu (2007) for a more detailed discussion.

In special cases when the number of JLCs is known and the JLCs satisfy some smoothness conditions, Müller and Song (1994), O'Sullivan and Qian (1994), Qiu (1997), and several others suggest various two-stage procedures, by which the JLCs are first estimated by some curves, and then the regression surface is fitted as usual in design sub-spaces separated by the estimated JLCs. In applications, however, it is often difficult to obtain information about the number of JLCs and their smoothness. Therefore, these procedures may not be appropriate for many applications. Recently, Qiu (2004) suggests a jump surface estimation procedure by fitting four local planes in four quadrants of a neighborhood of a given point. The best one of the four surface estimators defined by the four fitted local planes is chosen by a criterion to be the final surface estimator at the given point. It has been shown that this procedure preserves jumps well; but it is relatively noisy around the JLCs, because only one quadrant of the neighborhood is actually used in defining its surface estimator. Several papers, including Carlstein and Krishnamoorthy (1992), Hall and Molchanov (2003), Hall, Peng and Rau (2001), Hall and Rau (2000), Qiu (2002), Qiu and Yandell (1997), and Sun and Qiu (2007), focus mainly on estimation of the JLCs. For an overview of the methods mentioned above, see Qiu (2005).

The major goal of this paper is to suggest a jump surface estimator, which would have an explicit mathematical formula like most conventional surface estimators and which can preserve jumps well and remove noise efficiently. Generally speaking, there are two major benefits for a surface estimator to have an explicit mathematical formula. One is that such an estimator is often easy to use and simple to compute; the other major benefit is that it is usually easier for us to study its statistical properties. To achieve these goals, we suggest a three-stage procedure briefly described below. In the first stage, a local plane is fitted in a neighborhood of a given point by the local linear kernel smoothing, from which the conventional local linear kernel estimator of the

surface at the given point can be obtained. In the second stage, the neighborhood is divided into two parts by a line passing through the given point and perpendicular to the gradient direction of the fitted local plane. Two one-sided local linear kernel estimators of the surface are then constructed, respectively, in the two parts. The fitted surface is defined by one of the three estimators (i.e., the conventional estimator and the two one-sided estimators), depending on the weighted residual means of squares of the corresponding fitted planes: the smaller, the better. To further remove noise, in the third stage, the estimated surface values at the original design points obtained in the second stage are used as a new dataset, and the surface estimation procedure of the second stage is applied to this dataset in the same way except that one of the three estimators is selected as the final surface estimator based on their estimated variances in this step. It will be shown that the final surface estimator preserves jumps well and removes noise efficiently.

The idea to define an estimator of a jump regression function by one of the three estimators (i.e., the conventional estimator and the two one-sided estimators) has been discussed by Gijbels *et al.* (2007) and Qiu (2003) in 1-D cases. However, 2-D problems are much more challenging than their 1-D counterparts due mainly to the following two facts. First, in 1-D cases, jump locations are at most a series of points in the design interval; in 2-D cases, jump locations are often curves without any mathematical expressions. Second, in 1-D cases, two different sides of a given point are well defined; in 2-D cases, two different sides of a given point can be defined along a specific direction, but the direction parameter would add much complexity to 2-D problems. Recently, Gijbels *et al.* (2006) generalizes these 1-D methods to 2-D cases. In that paper, the surface estimator at a given point is defined in a single step, by one of the three estimators constructed directly from the observations. To choose among the three estimators, a threshold parameter is introduced and used for comparing their weighted residual means of squares. That procedure is then modified for preserving corners of edges. The random design case is also discussed in that paper along with many numerical examples using various test images. But, theoretical properties of that procedure are not derived. As a comparison, the proposed procedure in the current paper has three steps, motivated by our theoretical study about the properties of the three local estimators, especially about the specific relationship between their performance and the signal-to-noise ratio in the data (see Sections 2 and 3 for detailed discussion). Besides the two bandwidths, no other parameters, such as threshold values, are used in the current procedure. Its estimated surface is proved to be strong consistent.

The rest of the article is organized as follows. In next section, the proposed jump-preserving surface estimation procedure is introduced in some details. Some of its statistical properties are discussed in Section 3. Several numerical examples are presented in Section 4. Some technical

details, including two propositions and proofs of several theorems, are provided in Section 5. Some remarks conclude the article in Section 6.

2 The Jump-Preserving Surface Reconstruction Procedure

In this section, we introduce our jump surface estimation procedure in detail. To make the introduction easier to understand, we keep it intuitive here, and provide some mathematically more rigorous justifications in Section 3.

Suppose that the regression model is

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

where f is a bivariate nonparametric regression function, $\{(x_i, y_j) = (i/n, j/n), i, j = 1, 2, \ldots, n\}$ are equally spaced design points in $[0, 1] \times [0, 1]$, $\{\varepsilon_{ij}\}$ are i.i.d. random errors with mean 0 and variance σ^2, $\{z_{ij}\}$ are observations, and $N = n^2$ is the sample size. The regression function f is assumed to be continuous in the design space except on some curves, which are called the jump location curves (JLCs). A formal definition of the JLCs can be found in Qiu (1998).

At a given point (x, y) in the design space, we consider a neighborhood $M_n(x, y) = \{(u, v) : (u, v) \in [0, 1] \times [0, 1], \sqrt{(u - x)^2 + (v - y)^2} \le h_n\}$, where $h_n \le 1$ is a window width parameter. In $M_n(x, y)$, a local plane is fitted by the local linear kernel smoothing:

$$\min_{a,b,c \in R} \sum_{i=1}^{n} \sum_{j=1}^{n} \{z_{ij} - [a + b(x_i - x) + c(y_j - y)]\}^2 K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n}\right), \tag{2}$$

where $K(\cdot, \cdot)$ is an isotropic, bivariate density kernel function with support $\{(x, y) : x^2 + y^2 \le 1\}$. The solution of (2) is denoted by $\widehat{a}(x, y), \widehat{b}(x, y)$, and $\widehat{c}(x, y)$; $\widehat{a}(x, y)$ is the conventional local linear kernel estimator of $f(x, y)$. Its weighted residual mean square (WRMS) is

$$e(x, y) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \left\{z_{ij} - [\widehat{a}(x, y) + \widehat{b}(x, y)(x_i - x) + \widehat{c}(x, y)(y_j - y)]\right\}^2 K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n}\right)}{\sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n}\right)}. \tag{3}$$

The gradient direction of the fitted plane is $G(x, y) = (\widehat{b}(x, y), \widehat{c}(x, y))$, and it is a good estimator of the gradient direction of f.

A major consideration for using circular neighborhoods here is their rotation-invariance property. For some applications in geology, meteorology, and oceanography, a pre-specified coordinate system is not essential to the surface reconstruction problem. In such cases, it is desirable to have

the estimated surface coordinate-free, and circular neighborhoods can make this possible. For some other applications, such as image reconstruction, the coordinate system is well defined. In such cases, square-shaped or other types of neighborhoods can also be used.

If the point (x, y) is on a JLC, then according to Corollary 1 in Section 3, $G(x, y)$ also well indicates the direction orthogonal to the JLC tangent at (x, y). In the literature, there are several different ways to estimate the gradient direction of f. For example, in the image processing literature (cf. Gonzalez and Woods (1992)), people often use two discrete difference operators in the x and y directions for estimating the gradient direction. Here, the gradient direction of f is estimated by $G(x, y)$ based on two considerations. One is computational. Since $G(x, y)$ is a by-product of the minimization problem (2), no extra computation is required for obtaining $G(x, y)$ after the problem (2) is solved. The other consideration is theoretical. It can be shown that $G(x, y)$ converges to the gradient direction of f with optimal convergence rate under some regularity conditions (cf., Fan and Gijbels (1996)).

In order to accommodate the jump structure of the surface in surface fitting, the neighborhood $M_n(x, y)$ is then divided into two parts: $M_n^{(1)}(x, y)$ and $M_n^{(2)}(x, y)$, by a line which passes through (x, y) and is perpendicular to $G(x, y)$, as shown in Figure 1.

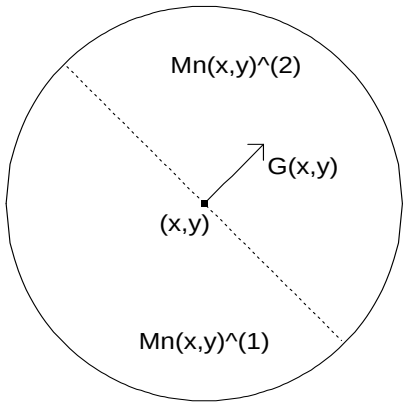

Figure 1: The neighborhood $M_n(x, y)$ is divided into two parts $M_n^{(1)}(x, y)$ and $M_n^{(2)}(x, y)$, along a direction perpendicular to the gradient direction $G(x, y)$.

Then, in $M_n^{(1)}(x, y)$ and $M_n^{(2)}(x, y)$, two one-sided local planes are fitted, respectively, by the local linear kernel smoothing:

$$\min_{a,b,c\in R} \sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \{z_{ij} - [a + b(x_i - x) + c(y_j - y)]\}^2 K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n}\right), \text{ for } \ell = 1, 2. \quad (4)$$

The solution of (4) is denoted by $(\widehat{a}^{(\ell)}(x, y), \widehat{b}^{(\ell)}(x, y), \widehat{c}^{(\ell)}(x, y))$, for $\ell = 1$ and 2. The WRMS values

of the two fitted one-sided planes are

$$e^{(\ell)}(x,y)=\frac{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)}\left\{z_{ij}-[\widehat{a}^{(\ell)}(x,y)+\widehat{b}^{(\ell)}(x,y)(x_i-x)+\widehat{c}^{(\ell)}(x,y)(y_j-y)]\right\}^2 K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right)}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right)} \tag{5}$$

for $\ell=1$ and 2.

Intuitively, a comparison of the three estimators $\widehat{a}(x,y)$, $\widehat{a}^{(1)}(x,y)$ and $\widehat{a}^{(2)}(x,y)$ can be done to determine the best overall estimator of $f(x,y)$, based on whether or not $M_n(x,y)$ contains any jump points. When there are no jumps in $M_n(x,y)$, all three estimators are statistically consistent for estimating $f(x,y)$ under some regularity conditions. Then, the conventional estimator $\widehat{a}(x,y)$ is preferred, because it averages more observations. When there is a single JLC in $M_n(x,y)$ and the JLC has a unique tangent line on each of its points in the neighborhood, $\widehat{a}(x,y)$ does not estimate $f(x,y)$ well, because observations on both sides of the JLC are averaged and jumps are blurred. In such cases, one of the two parts of $M_n(x,y)$ should be mostly on one side of the JLC, and the one-sided estimator constructed in that part should be able to estimate the surface well. So, in both cases, at least one of the three estimators $\widehat{a}(x,y)$, $\widehat{a}^{(1)}(x,y)$ and $\widehat{a}^{(2)}(x,y)$ estimates $f(x,y)$ well when the sample size is reasonably large.

In applications, however, it is often unknown whether or not a given point (x,y) is close to a JLC. So a data-driven mechanism is needed to choose one of the three estimators for estimating $f(x,y)$, which leads to three possible solutions, described below.

To estimate $f(x,y)$ properly, one possibility is to choose one of the three estimators based on the corresponding WRMS values, defined by (3) and (5). According to Proposition 1 given in Section 5, $e(x,y)\geq\min(e^{(1)}(x,y),e^{(2)}(x,y))$, for any $(x,y)\in[0,1]\times[0,1]$. Therefore, by this proposal, $\widehat{a}(x,y)$ can never be selected, and the surface estimator is actually defined by

$$\widehat{f}_1(x,y;\underline{z})=\begin{cases}\widehat{a}^{(1)}(x,y), & \text{if } e^{(1)}(x,y)<e^{(2)}(x,y)\\ \widehat{a}^{(2)}(x,y), & \text{if } e^{(1)}(x,y)>e^{(2)}(x,y)\\ (\widehat{a}^{(1)}(x,y)+\widehat{a}^{(2)}(x,y))/2 & \text{if } e^{(1)}(x,y)=e^{(2)}(x,y),\end{cases} \tag{6}$$

where $\underline{z}$ denotes the vector of all observations $\{z_{ij},i,j=1,2,\cdots,n\}$. As explained above, $\widehat{f}_1(x,y;\underline{z})$ should preserve the jump well at each point of the JLCs at which the JLCs have a unique tangent line and the jump magnitude is non-zero. The "unique tangent line" requirement is mainly due to the fact that the neighborhood $M_n(x,y)$ is divided into two half circles $M_n^{(1)}(x,y)$ and $M_n^{(2)}(x,y)$ by a *line* in (4). This requirement can be made more flexible, if, for example, two opposite sectors in $M_n(x,y)$ along a direction perpendicular to $G(x,y)$ are used in (4), for constructing the two one-sided estimators of $f(x,y)$.

Like most local smoothing estimators, the estimator $\widehat{f}_1(x,y;\underline{z})$ is defined for $(x,y) \in H_n = [h_n, 1-h_n] \times [h_n, 1-h_n]$ only. 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 partially overcome this problem. For example, most discrete wavelet transformation (DWT) software packages use periodic or symmetric extension methods to define neighborhoods in the boundary regions (Nason and Silverman (1994)). In this paper, the symmetric extension method is used in all numerical examples.

While the surface estimator $\widehat{f}_1(x,y;\underline{z})$ should preserve jumps well, it is expected to be relatively noisy, compared to the conventional estimator $\widehat{a}(x,y)$, because it is defined by only half observations in $M_n(x,y)$. The major reason why $\widehat{f}_1(x,y;\underline{z})$ is noisy is that $\widehat{a}(x,y)$ can not be selected in its definition, even in continuity regions of f. One natural idea to overcome this limitation is to choose one of $\widehat{a}(x,y)$, $\widehat{a}^{(1)}(x,y)$ and $\widehat{a}^{(2)}(x,y)$ for estimating $f(x,y)$ based on their variances, because the variance of $\widehat{a}(x,y)$ would be smaller than the variances of $\widehat{a}^{(1)}(x,y)$ and $\widehat{a}^{(2)}(x,y)$ in the continuity regions of f due to the fact that the former averages more observations. As a matter of fact, when the kernel function K is a constant (i.e., the three estimators are local constant kernel estimators), it can be checked that $Var(\widehat{a}^{(1)}(x,y)) \approx Var(\widehat{a}^{(2)}(x,y)) \approx 2Var(\widehat{a}(x,y)) \approx C^*(nh_n)^{-2}\sigma^2$, when there are no jumps in $M_n(x,y)$, where C^* is a constant and "$\approx$" denotes asymptotic equality. In applications, σ^2 is often unknown. But it can be estimated by the WRMS values $e(x,y), e^{(1)}(x,y)$ and $e^{(2)}(x,y)$ for the three estimators. Based on these considerations, the surface estimator can be defined by

$$\widehat{f}_2(x,y;\underline{z}) = \begin{cases} \widehat{a}(x,y), & \text{if } e(x,y)/2 \le \min(e^{(1)}(x,y), e^{(2)}(x,y)) \\ \widehat{a}^{(1)}(x,y), & \text{if } e^{(1)}(x,y) < e(x,y)/2 \text{ and } e^{(1)}(x,y) < e^{(2)}(x,y) \\ \widehat{a}^{(2)}(x,y), & \text{if } e^{(2)}(x,y) < e(x,y)/2 \text{ and } e^{(2)}(x,y) < e^{(1)}(x,y) \\ (\widehat{a}^{(1)}(x,y) + \widehat{a}^{(2)}(x,y))/2 & \text{if } e^{(1)}(x,y) = e^{(2)}(x,y) < e(x,y)/2. \end{cases} \tag{7}$$

From Proposition 1 in Section 5, $e(x,y)$ can never be the smallest one among $e(x,y), e^{(1)}(x,y)$, and $e^{(2)}(x,y)$). So, by comparing (6) and (7), we can see that the estimator $\widehat{f}_1(x,y;\underline{z})$ can also be obtained by equation (7), after the quantity $e(x,y)/2$ is replaced by $e(x,y)$. This connection between (6) and (7) is helpful for computer programming.

In continuity regions of f, $e(x,y)/2$ is less than both $e^{(1)}(x,y)$ and $e^{(2)}(x,y)$ when n is large, because all of $e(x,y), e^{(1)}(x,y)$ and $e^{(2)}(x,y)$ are consistent estimators of σ^2. In this case, $\widehat{a}(x,y)$ is selected in (7), and, consequently, $\widehat{f}_2(x,y;\underline{z})$ can remove noise well. When there are jumps in $M_n(x,y)$, $e(x,y)$ would be relatively large, because of the jumps. If the jump size is large compared to σ, then it can happen that $e(x,y)/2 > \min(e^{(1)}(x,y), e^{(2)}(x,y))$. In such a case, $\widehat{f}_2(x,y;\underline{z})$ equals

$\widehat{f}_1(x,y;\underline{z})$, and consequently the jumps are preserved well. If the jump size is small compared to σ, however, $e(x,y)/2$ could be smaller than both $e^{(1)}(x,y)$ and $e^{(2)}(x,y)$. In such a case, the jumps are blurred by (7). These facts will be formally justified in Section 3 when we discuss statistical properties of $\widehat{f}_2(x,y;\underline{z})$.

From the above description, we notice that procedure (7) provides a good surface estimator when the ratio of the jump size to σ (which is called the signal-to-noise ratio (SNR) hearafter) is large, and procedure (6) is preferred when this ratio is small. Based on this observation, we suggest the following procedure: procedure (6) is first applied to the original data to decrease the noise level, and then procedure (7) is applied to the estimated surface of the previous step to further remove noise. That is, the surface estimator is defined by

$$\widehat{f}(x,y;\underline{z}) = \widehat{f}_2(x,y;\underline{\widehat{f}}_1) \tag{8}$$

where $\underline{\widehat{f}}_1$ denotes the vector of $\{\widehat{f}_1(i/n, j/n;\underline{z}), i,j = 1,2,\ldots,n\}$. It should be noticed that the window widths used in the two steps of (8) could be different. They are denoted by h_{n1} and h_{n2} hereafter. Since both steps of (8) involve local smoothing only, the computation of (8) is quite straightforward. Its computational complexity is $O(N^2h_{n1}^2) + O(N^2h_{n2}^2)$.

In applications, the window widths h_{n1} and h_{n2} can be determined by minimizing the following cross-validation criterion

$$CV(h_{n1},h_{n2}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(z_{ij} - \widehat{f}_{-i,-j}(x_i,y_j;\underline{z})\right)^2, \tag{9}$$

where $\widehat{f}_{-i,-j}(x,y;\underline{z})$ is the "leave-one-out" estimator of $f(x,y)$. Namely, the observation (x_i,y_j,z_{ij}) is left out in constructing $\widehat{f}_{-i,-j}(x,y;\underline{z})$.

3 Some Statistical Properties of the Fitted Surfaces

In this section, we discuss some statistical properties of the estimated surfaces of procedures (6)-(8). For simplicity of presentation, a point on the JLCs is called a *nonsingular* point below, if the JLCs have a unique tangent line at this point and the jump magnitude is non-zero. Other points of the JLCs are called *singular* points. Obviously, a singular point of the JLCs is: (i) a cross point of several JLCs, or (ii) a point on a single JLC at which the JLC does not have a unique tangent line, or (iii) a point on a single JLC at which there exists a neighborhood such that the jump magnitude is zero at the given point but non-zero at any other points of the JLC in the neighborhood.

The first theorem below is about the conventional local linear kernel estimators $\widehat{a}(x,y), \widehat{b}(x,y)$ and $\widehat{c}(x,y)$. Proofs of original theorems are given in Section 5.

Theorem 1 (This theorem can be proven similarly to some closely related theorems in Qiu (1997) and Qiu and Yandell (1997)) Suppose that the regression function f has continuous second order derivatives in each closed set of the design space in which it is continuous; $E(\varepsilon_{11}^3) < \infty$; the kernel function $K(x,y)$ is a Lipschitz-1 continuous, isotropic, bivariate density function; and h_n satisfies the condition that $\log^2(n)/(nh_n^3) = O(1)$. Then

$$\|\widehat{a} - f\|_{D_{h_n}} = O(h_n^2) + O\left(\frac{\log(n)}{nh_n}\right) \ a.s., \tag{10}$$

$$\|\widehat{b} - f'_x\|_{D_{h_n}} = O(h_n) + O\left(\frac{\log(n)}{nh_n^2}\right) \ a.s., \tag{11}$$

$$\|\widehat{c} - f'_y\|_{D_{h_n}} = O(h_n) + O\left(\frac{\log(n)}{nh_n^2}\right) \ a.s., \tag{12}$$

where $D_{h_n} = H_n \backslash J_{h_n}$, J_{h_n} is a band of the JLCs with radius h_n, and $\|g\|_{D_{h_n}}$ denotes $\sup_{(x,y)\in D_{h_n}} |g(x,y)|$. For a given point $(x_\tau, y_\tau) \in J \backslash S$, where J denotes the set of points on the JLCs and S denotes the set of all singular points of the JLCs, if the projection of a point $(x,y) \in J_{h_n}$ to J is (x_τ, y_τ) and the Euclidean distance between the two points is ch_n, where $0 < c < 1$ is a constant (note: the point (x,y) depends on n although it is not explicit in notation), then

$$\widehat{a}(x,y) = f_-(x_\tau, y_\tau) + d_\tau \int\int_{Q^{(2)}} K(u,v)\ dudv + o(1)\ a.s., \tag{13}$$

$$\widehat{b}(x,y) = \frac{d_\tau}{\widetilde{\beta}_{02} h_n} \int\int_{Q^{(2)}} uK(u,v)\ dudv + o(1/h_n)\ a.s., \tag{14}$$

$$\widehat{c}(x,y) = \frac{d_\tau}{\widetilde{\beta}_{20} h_n} \int\int_{Q^{(2)}} vK(u,v)\ dudv + o(1/h_n)\ a.s., \tag{15}$$

where $d_\tau > 0$ is the jump magnitude of f at (x_τ, y_τ), which is assumed to be finite, $Q_n^{(1)}(x,y)$ and $Q_n^{(2)}(x,y)$ are two different parts of $M_n(x,y)$ separated by the JLC with a positive jump at (x_τ, y_τ) from $Q_n^{(1)}(x,y)$ to $Q_n^{(2)}(x,y)$, $Q^{(1)}$ and $Q^{(2)}$ are the two corresponding parts of the support of $K(\cdot,\cdot)$, $f_-(x_\tau, y_\tau)$ is the limit of f at (x_τ, y_τ) from $Q_n^{(1)}(x,y)$, and $\widetilde{\beta}_{s_1 s_2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u^{s_1} v^{s_2} K(u,v)\ dudv$, for $s_1, s_2 = 0, 1, 2$.

In Theorem 1, if $h_n \sim n^{-1/3}\log^{2/3}(n)$, then $\|\widehat{a} - f\|_{D_{h_n}} = o(1)$, $\|\widehat{b} - f'_x\|_{D_{h_n}} = o(1)$, and $\|\widehat{c} - f'_y\|_{D_{h_n}} = o(1)$. Therefore, $\widehat{a}(x,y), \widehat{b}(x,y)$ and $\widehat{c}(x,y)$ converge uniformly to $f(x,y)$, $f'_x(x,y)$ and $f'_y(x,y)$, respectively, in D_{h_n} where f is continuous. Theorem 3.1 also shows that $\widehat{a}(x,y), \widehat{b}(x,y)$ and $\widehat{c}(x,y)$ are affected by the jumps around the JLCs. In a special case when the point (x,y) is a nonsingular point of the JLCs, it can be seen from (13) that $\widehat{a}(x,y)$ does not converge to $f(x,y)$, as mentioned in Section 1.

Corollary 1 If $(x, y) \in H_n$ is a non-singular point on a JLC and the JLC has a unique tangent line at (x, y) with the direction of θ, where $\theta \in [0, \pi]$, then it can be checked that: under all other conditions in Theorem 1, we have

$$\widehat{b}(x,y) = -\frac{d_\tau C}{\widetilde{\beta}_{02} h_n}\sin(\theta) + o(1/h_n), \;\; \widehat{c}(x,y) = \frac{d_\tau C}{\widetilde{\beta}_{20} h_n}\cos(\theta) + o(1/h_n), \; a.s.,$$

where C is a constant, $d_\tau, \widetilde{\beta}_{02}$ and $\widetilde{\beta}_{20}$ are defined in Theorem 1.

Corollary 1 is a direct conclusion of equations (14) and (15) in Theorem 1. It implies that the gradient direction $G(x, y) = (\widehat{b}(x, y), \widehat{c}(x, y))$ is approximately in the direction of $(-\sin(\theta), \cos(\theta))$, which is perpendicular to the direction of the tangent line of the JLC at (x, y). Therefore, $G(x, y)$ indeed indicates the orientation of the JLC at (x, y) well, which is one of the foundations of our jump surface estimation procedures introduced in Section 2.

Theorem 2 Suppose that the conditions stated in Theorem 1 are all satisfied, and $E(\varepsilon_{11}^4) < \infty$. Then

$$\begin{aligned} \|e - \sigma^2\|_{D_{h_n}} &= o(1) \; a.s., \\ \|e^{(\ell)} - \sigma^2\|_{D_{h_n}} &= o(1) \; a.s., \text{ for } \ell = 1, 2. \end{aligned} \tag{16}$$

For a given point $(x_\tau, y_\tau) \in J \backslash S$, if the projection of a point $(x, y) \in J_{h_n}$ to J is (x_τ, y_τ) and the Euclidean distance between the two points is ch_n, where $0 < c < 1$ is a constant, then

$$\begin{aligned} e(x,y) &= \sigma^2 + d_\tau^2 C_\tau^2 + o(1) \; a.s., \\ e^{(\ell)}(x,y) &= \sigma^2 + d_\tau^2 (C_\tau^{(\ell)})^2 + o(1) \; a.s., \text{ for } \ell = 1, 2, \end{aligned} \tag{17}$$

where

$$\begin{aligned} C_\tau^2 = \; & \frac{1}{\widetilde{\beta}_{02}^2} \int\int_{Q^{(1)}} \left\{ \int\int_{Q^{(2)}} (\widetilde{\beta}_{02} + us + vt) K(u,v) \, dudv \right\}^2 K(s,t) \, dsdt + \\ & \frac{1}{\widetilde{\beta}_{20}^2} \int\int_{Q^{(2)}} \left\{ \int\int_{Q^{(1)}} (\widetilde{\beta}_{20} + us + vt) K(u,v) \, dudv \right\}^2 K(s,t) \, dsdt, \end{aligned}$$

$d_\tau, Q^{(1)}$ and $Q^{(2)}$ are defined in Theorem 1, and $(C_\tau^{(1)})^2$ and $(C_\tau^{(2)})^2$ are defined in Section 5, similarly to C_τ^2.

Theorem 2 says that the three WRMS values are all consistent estimators of σ^2 in continuity regions of f. When the point (x, y) is close to a JLC, the WRMS values are affected by the jumps. Based on Theorems 1 and 2, the strong consistency of the estimated surfaces of procedures (6) and (7) is established below.

Theorem 3 Under the conditions in Theorem 2, we have

$$\|\widehat{f}_1 - f\|_{D_{h_n}} = O(h_n^2) + O\left(\frac{\log(n)}{nh_n}\right) \ a.s. \tag{18}$$

Theorem 4 Under the conditions in Theorem 2,

$$\|\widehat{f}_2 - f\|_{D_{h_n}} = O(h_n^2) + O\left(\frac{\log(n)}{nh_n}\right) \ a.s. \tag{19}$$

For a given point $(x_\tau, y_\tau) \in J \backslash S$, if the projection of a point $(x, y) \in J_{h_n}$ to J is (x_τ, y_τ) and the Euclidean distance between the two points is ch_n, where $0 < c < 1$ is a constant, then
(i) if $d_\tau/\sigma > 1/C_\tau$, then

$$\widehat{f}_2(x, y) = f(x, y) + o(1), \ a.s. \tag{20}$$

(ii) if $d_\tau/\sigma \leq 1/C_\tau$, then

$$\widehat{f}_2(x, y) = f(x, y) + d_\tau \int\int_{Q^{(2)}} K(u, v) \, dudv + o(1) \ a.s., \tag{21}$$

where d_τ, C_τ and $Q^{(2)}$ are defined in Theorems 1 and 2.

Theorem 3 says that $\widehat{f}_1$ is uniformly consistent in regions where f is continuous. With regard to $\widehat{f}_2$, Theorem 4 says that it is uniformly consistent in regions where f is continuous. In regions around the JLCs, $\widehat{f}_2$ is consistent only when the SNR is larger than a certain value.

Theorem 5 Under the conditions in Theorem 2, if $h_{n2} \sim h_{n1}$, then

$$\|\widehat{f} - f\|_{D_{h_{n1}+h_{n2}}} = O(h_{n1}^2) + O\left(\frac{\log(n)}{nh_{n1}}\right) \ a.s. \tag{22}$$

Theorem 5 establishes the strong consistency of the estimated surface of procedure (8) in continuity regions of f. By comparing (22) with (10), it can be seen that $\widehat{f}$ has the same convergence rate in the continuity regions of f, as the rate of the conventional estimator $\widehat{a}$, which is $O(n^{-2/3}\log^2(n))$ when $h_{n1} \sim n^{-1/3}\log(n)$. According to Stone (1982), this rate is optimal, up to a logarithmic factor.

For most existing jump surface estimation procedures in the literature, we do not know much of their theory yet, mainly due to their iterative nature which makes it hard to study their statistical properties under reasonably flexible assumptions. For instance, Geman and Geman (1984) proves that the image estimator obtained by their simulated annealing algorithm converges in probability to the MAP estimator, under some quite restrictive assumptions (cf. Section 1 for some related discussion). But it is still unknown at this moment whether the MAP estimator is a consistent estimator or not of the original image. It is not clear either, where in the design space the MAP estimator would work well and where it may estimate the original image poorly.

4 Numerical Study

In this section, we present some numerical examples for evaluating the performance of procedures (6)-(8). The examples are organized in four parts. Those regarding procedures (6) and (7) are discussed in Section 4.1. The numerical performance of (8) is investigated in Section 4.2. In Section 4.3, procedures (6)-(8) are compared to the local median smoother, a DWT procedure and a MRF procedure by some examples. The related procedures are then applied to a test image in Section 4.4.

4.1 Numerical performance of procedures (6) and (7)

Assume that the true regression surface is $f(x,y) = -2(x-0.5)^2 - 2(y-0.5)^2 + 1$ if $(x-0.5)^2 + (y-0.5)^2 < 0.25^2$; and $f(x,y) = -2(x-0.5)^2 - 2(y-0.5)^2$ otherwise. It has one JLC which is a circle with constant jump size 1. We first apply procedures (6) and (7) to this example. The observations are generated by (1) with $n = 100$ and $\sigma = 0.2$, 0.5 or 0.8. We let the window width h_n in both procedures change from 0.02 to 0.2 with step 0.01. The MSE values of the fitted surfaces in various combinations of σ and h_n are shown in Figure 2. If there is no further specification, all MSE values presented in this section are averages of 100 replications. From the plots, it can be seen that for each σ, the MSE values of each procedure first decrease and then increase when h_n increases from 0.02 to 0.2. The optimal value of h_n gets larger when σ is larger, which implies that h_n needs to be chosen larger for noisier data. By comparing the two plots, we can see that h_n needs to be chosen a little larger in (6) than its value in (7). This can be explained by the fact that procedure (6) uses only half observations in $M_n(x,y)$ for constructing its surface estimator while procedure (7) uses more observations.

We then concentrate on the case when $\sigma = 0.5$. The 2.5 and 97.5 percentiles of 100 replications of the fitted surface by (6) in the cross section of $y = 0.5$ are presented in Figure 3(a) by the lower and upper dashed curves, respectively. In (6), h_n is chosen to be 0.11 which is optimal according to Figure 2(a). In Figure 3(a), the solid curve denotes the true surface in the cross section of $y = 0.5$ and the dotted curve denotes the averaged surface fit. The corresponding results of (7) are presented in Figure 3(b). As demonstrated by Theorem 4, the performance of (7) depends on the value of SNR. To further see this, we present the corresponding results of (7) when $\sigma = 0.2$ in Figure 3(c). The window width h_n is 0.06 in Figure 3(b) and 0.05 in Figure 3(c). Both of them are optimal according to Figure 2(b). Figure 3 shows that procedure (6) preserves the jumps well but its estimated surface is relatively noisy (since its 95% confidence interval is relatively wide)

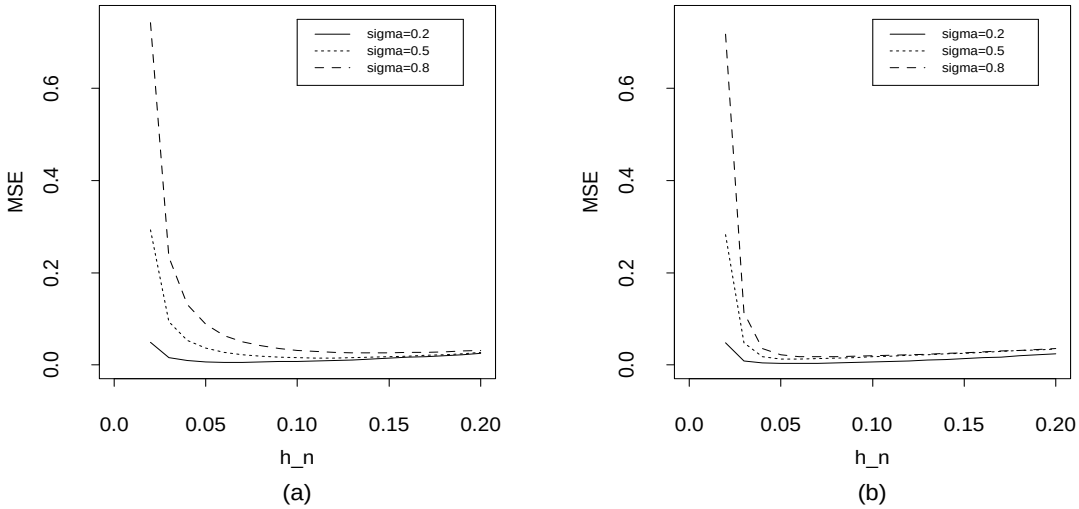

Figure 2: The MSE values of the fitted surfaces of procedures (6) (plot (a)) and (7) (plot (b)).

compared to the estimated surface of (7). However, procedure (7) blurs the jumps when $\sigma = 0.5$ (the SNR value is relatively small in this case) and it preserves the jumps well when $\sigma = 0.2$ (the SNR value is relatively large in this case).

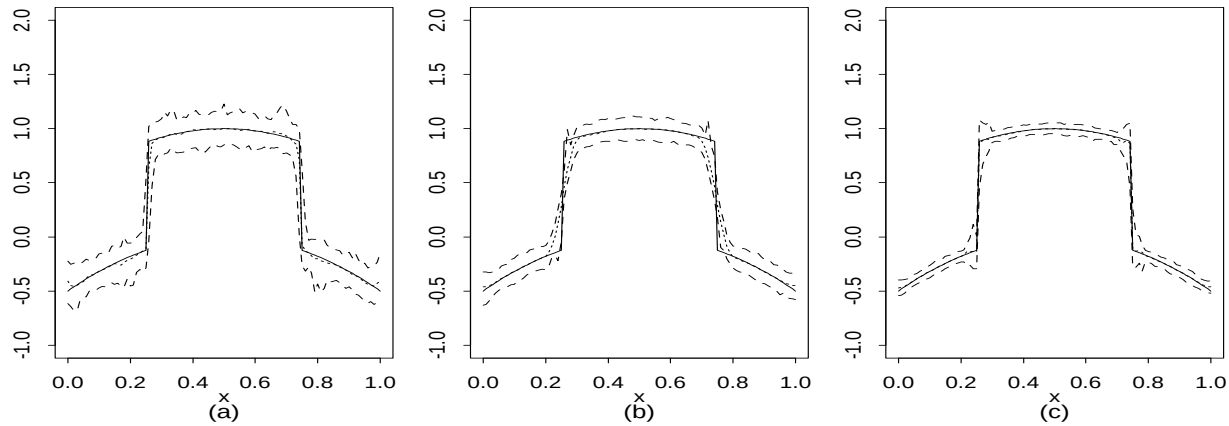

Figure 3: In each plot, the dashed curves denote the 2.5 and 97.5 percentiles of the 100 replications of the fitted surface, the solid curve denotes the true surface and the dotted curve denotes the averaged surface fit in the cross section of $y = 0.5$. (a) Procedure (6); (b) procedure (7) when $\sigma = 0.5$; (c) procedure (7) when $\sigma = 0.2$.

The MSE value of the estimated surface of (7) is 0.0127 (cf. Figure 2) in the case of Figure 3(b), which is smaller than the MSE value (=0.0151) of the estimated surface of (6) in the same case. So the MSE value does not reflect the blurring phenomenon seen in Figure 3(b). To better see this, part of the two dotted curves (corresponding to $\sigma = 0.5$) in Figures 2(a) and 2(b) has been put together in Figure 4(a). One explanation of this result is that MSE is computed from the entire design space in which the blurring effect of (7) around the JLC is attenuated by the relatively

small variability of its estimated surface in regions where f is continuous. To further check this issue, local MSE values of the estimated surfaces of (6) and (7) are computed in a local band of the JLC with radius h_n. They are presented in Figure 4(b). We can see that the local MSE value of (7) is much larger than that of (6) in the range of $0.06 \leq h_n \leq 0.2$. This example shows that the conventional MSE is a good measurement of global performance of the surface reconstruction procedures while the local MSE is a better measurement of jump preservation of these procedures.

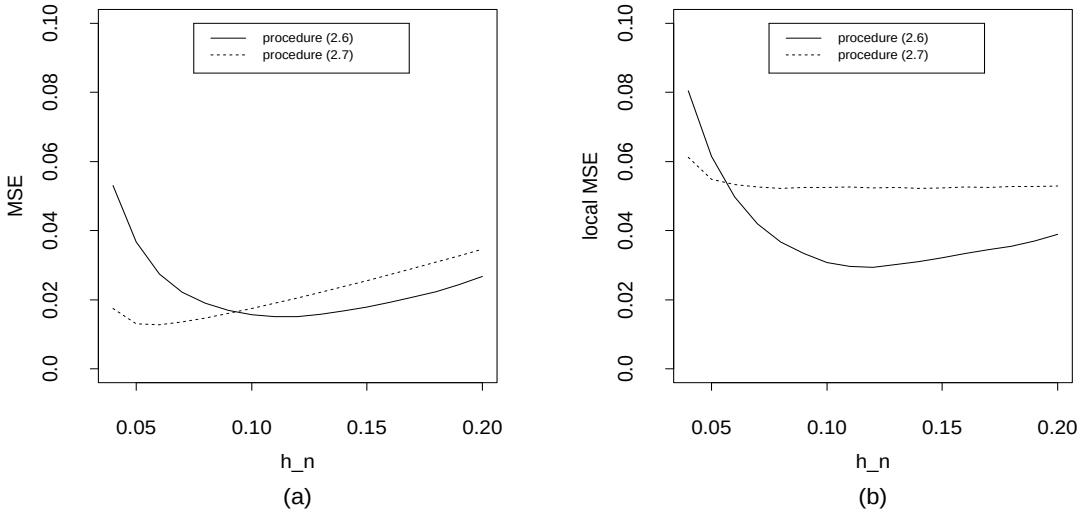

Figure 4: (a) The MSE values of the estimated surfaces of (6) and (7) when $\sigma = 0.5$; (b) the local MSE values of the two procedures calculated in a local band of the JLC with radius h_n.

4.2 Numerical performance of procedure (8)

Procedure (8) is a combination of (6) and (7), each of which has a window width. In the previous part, the impact of the two window widths on the performance of the two individual procedures has been studied. In the next example, we study their joint impact on the performance of (8). Suppose that the true surface is the one used in the previous example, $n = 100$, and $\sigma = 0.5$. The two window widths h_{n1} and h_{n2} in (8) can both vary from 0.03 to 0.16 with step 0.01. The averaged MSE value of the estimated surface of (8) based on 10 replications is presented in Figure 5(a). This plot shows that neither h_{n1} nor h_{n2} should be chosen too large or too small. The optimal values of h_{n1} and h_{n2} are 0.05 and 0.08, respectively, with MSE=0.0112. Compared to the results shown in Figure 2, i.e., the optimal values of h_{n1} and h_{n2} are respectively 0.11 and 0.06 with MSE=0.0150 and MSE=0.0127 for procedures (6) and (7) when $\sigma = 0.5$, we can see that: (1) h_{n1} should be chosen larger than h_{n2} when the two procedures (6) and (7) are used separately, and (2) h_{n1} should be chosen smaller than h_{n2} when the two procedures are used jointly in (8). With the optimal

window widths, the 2.5 and 97.5 percentiles of 100 replications of the estimated surface of (8) in the cross section of $y = 0.5$ are presented in Figure 5(b) by the lower and upper dashed curves, respectively. Compared to Figures 3(a) and 3(b), we can see that the estimated surface of (8) is smoother than that of (6), especially in regions where f is continuous, and it preserves the jumps better than that of (7) as well.

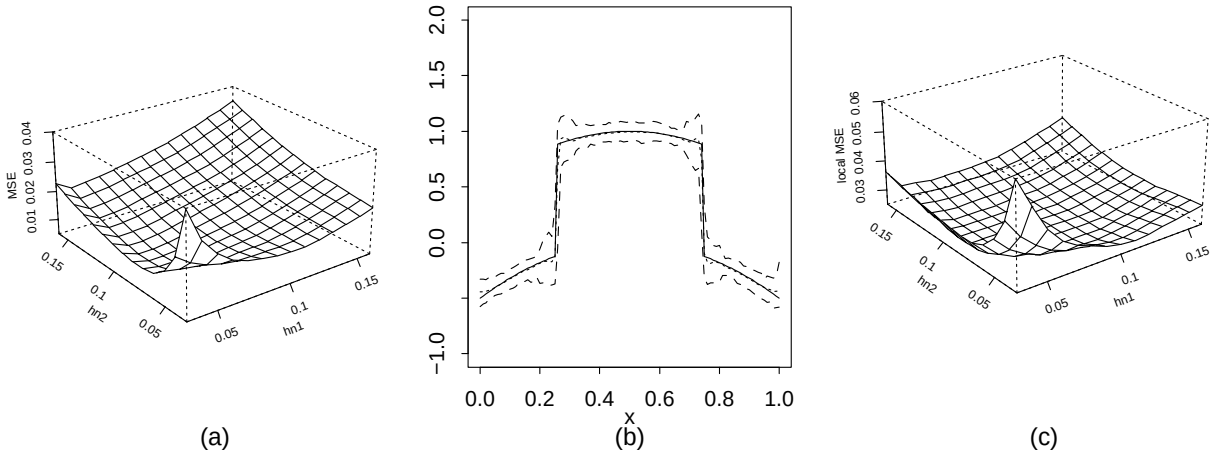

Figure 5: (a) The MSE values of the estimated surfaces of (8) are presented. (b) With the optimal window widths, the 2.5 and 97.5 percentiles of 100 replications of the estimated surface of (8) in the cross section of $y = 0.5$ are denoted by the lower and upper dashed curves, the true surface and the averaged surface estimator are denoted by the solid and dotted curves, respectively. (c) The local MSE values of the estimated surfaces of (8) are presented.

The averaged local MSE value of the estimated surface of (8) based on 10 replications is presented in Figure 5(c). The local MSE is computed in a band of the true JLC with width $h_{n1} + h_{n2}$. From the plot, it can be seen that the local MSE has a similar pattern to that of the global MSE shown in Figure 5(a), except that the local MSE seems more robust to selection of h_{n1} and h_{n2} in the sense that its value is minimal or close to minimal for more combinations of h_{n1} and h_{n2}. The minimum value 0.0221 of the local MSE is reached when h_{n1} and h_{n2} are 0.05 and 0.09, respectively. It can be seen that the optimal values of h_{n1} and h_{n2} by the local MSE criterion are similar to those by the MSE criterion. Also, the relationship between these local optimal bandwidth values and the local optimal bandwidth values $h_{n1} = 0.12$ with local MSE=0.0293 and $h_{n2} = 0.08$ with local MSE=0.0522 (cf., Figure 4(b)) when procedures (6) and (7) are used separately is similar to that under the MSE criterion discussed in the previous paragraph.

We then consider three σ values: 0.2, 0.5 and 0.8; and three n values: 100, 200 and 500. For each combination of σ and n, the optimal window widths of (8) are searched with step 0.01 based on 10 replications and on the MSE criterion. The results are presented in Table 1. It can be seen that the optimal value of h_{n1} is smaller than the optimal value of h_{n2} in all cases. When n is larger, both h_{n1} and h_{n2} need to be smaller and both nh_{n1} and nh_{n2} need to be larger, which is often

Table 1: In each entry, the first two numbers are the optimal h_{n1} and h_{n2} and the last number is the corresponding MSE value.

	σ		
n	0.2	0.5	0.8
100	0.03, 0.06, 0.0051	0.05, 0.08, 0.0112	0.06, 0.10, 0.0175
200	0.02, 0.04, 0.0022	0.03, 0.06, 0.0048	0.04, 0.07, 0.0078
500	0.01, 0.02, 0.0008	0.02, 0.03, 0.0019	0.02, 0.04, 0.0030

true for local smoothing procedures (Härdle 1990). When σ is larger, both h_{n1} and h_{n2} need to be chosen larger, which is intuitively reasonable.

4.3 Some numerical comparisons

In this part, the three procedures (6)-(8) are compared to each other and to several existing procedures in some numerical examples. Three existing procedures are considered here. The first one is the local median smoothing procedure, by which the surface estimator is defined by the sample median of the observations in a neighborhood of a given point. Because this procedure is simple and has some ability to preserve jumps while removing noise, it is widely used in the image reconstruction literature (cf. Gonzalez and Woods (1992), Chapter 4). The second one is the DWT procedure implemented by the R package wavethresh (see Nason and Silverman (1994) for detailed introduction). It has several parameters to determine before it can be used for image reconstruction. In this paper, the default family of wavelets (which is Daubechies' "extremal phase" wavelet) and the "symmetric" boundary handling condition are used. The parameter "filter.number" can vary from 1 to 10. The thresholding "policy" is either "hard" or "soft". The "levels" to be thresholded are $r : s$, where 2^{s+1} is the sample size and r is an integer ranging from 1 to s. The third existing procedure is the MRF procedure suggested by Godtliebsen and Sebastiani (1994). It combines the idea of discontinuity labeling process (Geman and Geman (1984)) and the iterated conditional modes algorithm (Besag (1986)). This procedure assumes that a binary line component exists between any two vertically or horizontally neighboring pixels, with 1 denoting an existing edge between the two pixels and 0 denoting no edge. In a 3×3 neighborhood of a given pixel, there are 12 line components and 2^{12} possible configurations of these components. To use this procedure, probabilities of the 2^{12} possible line configurations need to be specified. In this and the next subsections, these probabilities are estimated from the true regression function values at the design points, which is in favor of this procedure. Besides the line configurations, it has three positive procedure parameters α, β and λ. The estimated surfaces by the three existing procedures are denoted by $\widehat{f}_m, \widehat{f}_w$ and $\widehat{f}_{mrf}$, respectively.

To compare their performance, we consider two regression functions $f^{(1)}$ and $f^{(2)}$, where $f^{(1)}$ is the one used in the previous two subsections, $f^{(2)}(x,y) = -2(x-0.5)^2 - 2(y-0.5)^2 + g(x,y)$ where $g(x,y) = 0$,1 and 2 when (x,y) belongs to $[0,0.5] \times [0,0.5]$, $([0,0.5] \times (0.5,1]) \bigcup ((0.5,1] \times [0,0.5])$ and $(0.5,1] \times (0.5,1]$, respectively. So, $f^{(1)}$ has one JLC without any singular points, and $f^{(2)}$ has four JLC segments with one singular point at $(0.5, 0.5)$. Observations are generated by model (1) with n=100 and $\sigma = 0.2$, 0.5 or 0.8. Table 2 presents the optimal MSE values of the six procedures in various cases along with the corresponding values of the procedure parameters. In order to see the performance of the procedures around the JLCs, their local MSE values are also presented. For each of $\widehat{f}_1, \widehat{f}_2$ and $\widehat{f}_m$, the local MSE value is computed in a local band of the JLCs with radius equal to the optimal window width. For $\widehat{f}$, its local MSE value is computed in a local band of the JLCs with radius equal to the sum of the two optimal window widths of (8). For $\widehat{f}_w$ and $\widehat{f}_{mrf}$, no window widths are involved. To make the results comparable, their local MSE values in the same local band of the JLCs as that for $\widehat{f}$ are also presented in the table.

Table 2: In each entry, the numbers in the first two lines are the minimum MSE value and the corresponding local MSE value, respectively. The third line gives the optimal values of the procedure parameters (i.e., the optimal window widths for $\widehat{f}, \widehat{f}_1, \widehat{f}_2$ and $\widehat{f}_m$; the optimal "filter.number", "policy" and "levels" for $\widehat{f}_w$; and the optimal α, β and λ for $\widehat{f}_{mrf}$).

	$f^{(1)}$			$f^{(2)}$		
σ	0.2	0.5	0.8	0.2	0.5	0.8
$\widehat{f}$	0.0051	0.0112	0.0175	0.0022	0.0085	0.0168
	0.0157	0.0228	0.0295	0.0045	0.0142	0.0229
	0.03,0.06	0.05,0.08	0.06,0.10	0.04,0.06	0.06,0.10	0.08,0.13
$\widehat{f}_1$	0.0055	0.0151	0.0260	0.0029	0.0114	0.0223
	0.0153	0.0297	0.0431	0.0053	0.0186	0.0313
	0.07	0.11	0.14	0.09	0.14	0.18
$\widehat{f}_2$	0.0031	0.0127	0.0177	0.0021	0.0147	0.0205
	0.0151	0.0533	0.0553	0.0080	0.0507	0.0574
	0.05	0.06	0.08	0.06	0.05	0.07
$\widehat{f}_m$	0.0036	0.0135	0.0207	0.0039	0.0169	0.0245
	0.0189	0.0584	0.0670	0.0163	0.0605	0.0686
	0.04	0.05	0.07	0.04	0.05	0.07
$\widehat{f}_w$	0.0077	0.0146	0.0204	0.0057	0.0151	0.0212
	0.0209	0.0295	0.0302	0.0159	0.0294	0.0291
	3,hard,5:6	5,hard,4:6	5,soft,4:6	4,hard,4:6	4,hard,3:6	5,soft,4:6
$\widehat{f}_{mrf}$	0.0013	0.0132	0.0220	0.0009	0.0160	0.0225
	0.0016	0.0219	0.0317	0.0008	0.0227	0.0286
	3.8,41.5,1.9	7.5,5.9,0.2	17,2.5,6.0	35,33,0.4	3.9,2.8,6.7	7.7,2.1,5.1

First, let us compare the performance of (6)-(8). It can be seen from the table that the global and local MSE values of (8) are smaller than the corresponding values of (6) and (7) when σ is

0.5 or 0.8, in both cases of $f^{(1)}$ and $f^{(2)}$. So it might be safe to say that when the noise level is moderate to high, procedure (8) outperforms procedures (6) and (7). When $\sigma = 0.2$, procedure (7) performs better than both (6) and (8) in the case of $f^{(1)}$. In the case of $f^{(2)}$, the three procedures perform similarly well. Therefore procedure (7) is competitive when the noise level is low, which is consistent with the results found in Figure 3.

Next, let us compare the performance of (8) with the performance of $\widehat{f}_m, \widehat{f}_w$ and $\widehat{f}_{mrf}$. Table 2 shows that procedure (8) performs uniformly better than $\widehat{f}_w$. It performs better than $\widehat{f}_m$ in all cases except the case when $f = f^{(1)}$ and $\sigma = 0.2$, in which $\widehat{f}_m$ performs better than $\widehat{f}$ in MSE, but $\widehat{f}$ is slightly better in terms of the local MSE. Compared to $\widehat{f}_{mrf}$, $\widehat{f}$ performs better when σ is either 0.5 or 0.8. When $\sigma = 0.2$, $\widehat{f}_{mrf}$ performs the best among all procedures.

4.4 Application to a test image

In this part, we apply all the related procedures discussed in the previous part to a test image posted on the Waterloo Research Group's web page http://links.uwaterloo.ca/bragzone.base.html. The original image has four grey levels: 20, 75, 150 and 235; and several circular edges with three different jump magnitudes: 75, 160 and 215. Its resolution is 256×256. We then add i.i.d. noise with distribution $N(0, 75^2)$ to the image, and the noisy image is shown in Figure 6(a). The darker the color, the larger the image grey level.

Since the intensity function of this test image is piecewisely constant and it has large homogeneous regions, this image is ideal for the adaptive weights smoothing (AWS) procedure by Polzehl and Spokoiny (2000), which adapts itself to edge structures iteratively in surface reconstruction. We then use the AWS procedure as a golden standard in this example, and compare the others with it. The AWS procedure has a number of parameters and functions to specify. We use their default values specified in the R package aws(). The local median smoothing procedure, the DWT procedure, the MRF procedure, the AWS procedure, and procedures (6)-(8) are then applied to the noisy test image, and their reconstructed images are presented in Figures 6(b)-6(h), respectively. For the local median smoothing procedure and the procedures (6)-(8), their window widths are chosen to be 0.02, 0.05, 0.02 and (0.02,0.04), respectively, by the cross-validation procedure (cf. (9)). For the DWT procedure, all possible combinations of its parameters are tried and the combination with the best visual impression is selected, which turn out to be: filter.number=5, policy="soft" and levels=6:7. For the MRF procedure, as requested by one referee, the version by Sebastiani and Godtliebsen (1997) is used here, which searches for the estimate of the true image by MCMC and which chooses its parameter values automatically.

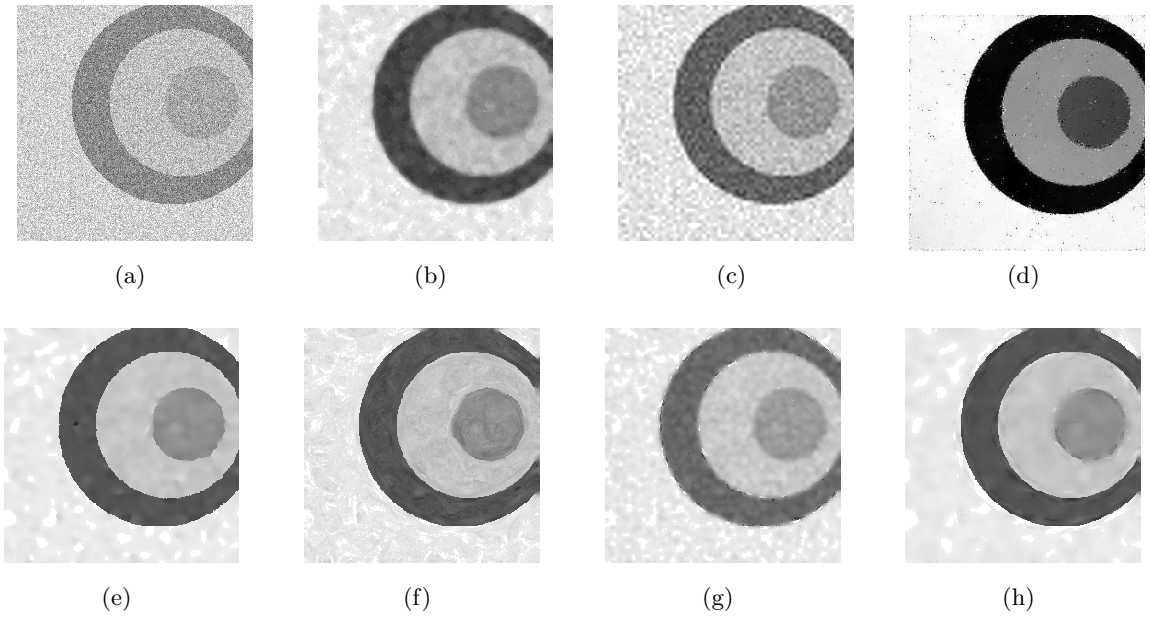

Figure 6: Plot (a): The noisy test image. Plots (b)-(h): the reconstructed images by the local median smoothing procedure, the DWT procedure, the MRF procedure, the AWS procedure, and procedures (6)-(8), respectively.

From the plots, it can be seen that the local median smoothing procedure and procedure (7) blur the edges much, especially in the region around the smallest circle. The reconstructed image of the DWT procedure looks quite noisy and some edges are also blurred. The MRF procedure can preserve the edges well, although its results include certain salt and pepper noise. The reconstructed images by procedures (6) and (8) look compatible to that by the AWS procedure. If we compare them carefully, then we can see that the reconstructed image of (8) is smoother than that of (6). We can also see that both procedures slightly blur some parts of the edges close to the image border. That is because they both use the "symmetric" extension procedure to handle the boundary regions and these parts are close to some singular points of the edges artificially created by the "symmetric" extension procedure. From Figure 6, it seems that procedure AWS does a good job in removing noise and preserving edges, compared to all other procedures. That might due to the facts that procedure AWS adapts its degree of smoothing to local features of the observed image by constantly changing the neighborhood size at a given point, while all other procedures considered here use constant procedure parameters for the entire image, and that procedure AWS smooths the observed data iteratively, while procedure (8) is non-iterative. This example shows that a potential improvement to the proposed procedure (8) is to use variable bandwidth and to apply procedure (6) iteratively, which is left to our future research.

5 Technical Details

This section mainly gives proofs of the theorems presented in Section 3. First, we give two propositions.

Proposition 1 For $e(x,y), e^{(1)}(x,y)$ and $e^{(2)}(x,y)$ defined by (3) and (5), there exists the following relationship:

$$e(x,y) \geq \min\left(e^{(1)}(x,y), e^{(2)}(x,y)\right),$$

for any $(x,y) \in [0,1] \times [0,1]$.

Proof. By the definition of $e(x,y)$, we have

$$\begin{aligned}
e(x,y) &= \frac{\sum_{(x_i,y_j)\in M_n^{(1)}(x,y)} \left\{z_{ij} - [\widehat{a}(x,y) + \widehat{b}(x,y)(x_i - x) + \widehat{c}(x,y)(y_j - y)]\right\}^2 K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)}{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} \\
&+ \frac{\sum_{(x_i,y_j)\in M_n^{(2)}(x,y)} \left\{z_{ij} - [\widehat{a}(x,y) + \widehat{b}(x,y)(x_i - x) + \widehat{c}(x,y)(y_j - y)]\right\}^2 K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)}{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} \\
&\geq \frac{\sum_{(x_i,y_j)\in M_n^{(1)}(x,y)} \left\{z_{ij} - [\widehat{a}^{(1)}(x,y) + \widehat{b}^{(1)}(x,y)(x_i - x) + \widehat{c}^{(1)}(x,y)(y_j - y)]\right\}^2 K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)}{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} \\
&+ \frac{\sum_{(x_i,y_j)\in M_n^{(2)}(x,y)} \left\{z_{ij} - [\widehat{a}^{(2)}(x,y) + \widehat{b}^{(2)}(x,y)(x_i - x) + \widehat{c}^{(2)}(x,y)(y_j - y)]\right\}^2 K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)}{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} \\
&= \frac{e^{(1)}(x,y)\sum_{(x_i,y_j)\in M_n^{(1)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right) + e^{(2)}(x,y)\sum_{(x_i,y_j)\in M_n^{(2)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)}{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} \\
&\geq \min(e^{(1)}(x,y), e^{(2)}(x,y)).
\end{aligned}$$

The first inequality in the above expressions is due to the definition (cf. equation (4)) of the one-sided estimators $(\widehat{a}^{(\ell)}(x,y), \widehat{b}^{(\ell)}(x,y), \widehat{c}^{(\ell)}(x,y))$, for $\ell = 1$ and 2, and the second inequality is due to the fact that a weighted average of two numbers must be larger than or equal to the smaller one of the two numbers.

Proposition 2 Under the conditions in Theorem 1, we have

$$\left\| \frac{1}{n^2 h_n^2} \sum_{i=1}^n \sum_{j=1}^n K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n}\right) - 1 \right\|_{D_{h_n}} = O\left(\frac{1}{nh_n}\right), \tag{23}$$

and

$$\left\| \frac{1}{n^2 h_n^2} \sum_{i=1}^n \sum_{j=1}^n \varepsilon_{ij} K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n}\right) \right\|_{D_{h_n}} = o\left(\frac{\beta_n \log(n)}{nh_n}\right), \ a.s. \tag{24}$$

for any positive sequence β_n that diverges to infinity as $n \to \infty$.

Proof. First, we prove equation (23). For any $(x, y) \in D_{h_n}$, we can write

$$
\begin{aligned}
& \left| \frac{1}{n^2 h_n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n} \right) - 1 \right| \\
= & \left| \frac{1}{n^2 h_n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n} \right) - \frac{1}{h_n^2} \int \int_{B_{h_n}(x,y)} K\left(\frac{u - x}{h_n}, \frac{v - y}{h_n} \right) \, dudv \right| \\
= & \left| \frac{1}{n^2 h_n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n} \right) - \frac{1}{h_n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \int \int_{B_{h_n}(x,y) \bigcap \Delta_{ij}} K\left(\frac{u - x}{h_n}, \frac{v - y}{h_n} \right) \, dudv \right| \\
= & \left| \frac{1}{h_n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[\frac{1}{n^2} K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n} \right) - \int \int_{B_{h_n}(x,y) \bigcap \Delta_{ij}} K\left(\frac{u - x}{h_n}, \frac{v - y}{h_n} \right) \, dudv \right] \right| \\
= & \left| \frac{1}{h_n^2} O\left(\frac{h_n}{n} \right) + \frac{1}{h_n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n} \right) \int \int_{B_{h_n}(x,y) \bigcap \Delta_{ij}} dudv - \right. \right. \\
& \left. \left. \int \int_{B_{h_n}(x,y) \bigcap \Delta_{ij}} K\left(\frac{u - x}{h_n}, \frac{v - y}{h_n} \right) \, dudv \right] \right| \\
= & \left| \frac{1}{h_n^2} O\left(\frac{h_n}{n} \right) + \frac{1}{h_n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \int \int_{B_{h_n}(x,y) \bigcap \Delta_{ij}} \left[K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n} \right) - K\left(\frac{u - x}{h_n}, \frac{v - y}{h_n} \right) \right] \, dudv \right| \\
\leq & \left| \frac{1}{h_n^2} O\left(\frac{h_n}{n} \right) \right| + \frac{1}{h_n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \int \int_{B_{h_n}(x,y) \bigcap \Delta_{ij}} \left| K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n} \right) - K\left(\frac{u - x}{h_n}, \frac{v - y}{h_n} \right) \right| \, dudv \\
\leq & \left| \frac{1}{h_n^2} O\left(\frac{h_n}{n} \right) \right| + \frac{1}{h_n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\sqrt{2} C_K}{n h_n} \int \int_{B_{h_n}(x,y) \bigcap \Delta_{ij}} dudv \\
= & \left| \frac{1}{h_n^2} O\left(\frac{h_n}{n} \right) \right| + \frac{1}{h_n^2} \frac{\sqrt{2} C_K}{n h_n} \int \int_{B_{h_n}(x,y)} dudv \\
= & \left| \frac{1}{h_n^2} O\left(\frac{h_n}{n} \right) \right| + \frac{1}{h_n^2} \frac{\sqrt{2} C_K}{n h_n} \pi h_n^2 = O\left(\frac{1}{n h_n} \right),
\end{aligned}
$$

where $B_{h_n}(x, y)$ is the circle centered at (x, y) with radius h_n, $\Delta_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$, $x_0 = y_0 = 0$, and $C_K > 0$ is the Lipschitz constant that satisfies $|K(x, y) - K(x', y')| \leq C_K \sqrt{(x - x')^2 + (y - y')^2}$. So, equation (23) is true.

To prove (24), let us first define

$$
\begin{aligned}
\bar{\varepsilon}_{ij} &= \varepsilon_{ij} I_{|\varepsilon_{ij}| \leq (t_{ij} \log^2(t_{ij}))^{1/3}}, \quad i, j = 1, 2, \cdots, n \\
\bar{g}_n(x, y) &= \frac{1}{n^2 h_n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \bar{\varepsilon}_{ij} K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n} \right) \\
g_n^*(x, y) &= \frac{1}{n^2 h_n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \varepsilon_{ij} K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n} \right),
\end{aligned}
$$

where $t_{ij} = (i-1)^2+j$ if $j \leq i$ and $t_{ij} = j^2-(i-1)$ otherwise. By the definition, $t_{ij} = 1, 2, 3, 4, 5, \cdots$, respectively, when $(i,j) = (1,1), (2,1), (2,2), (1,2), (3,1), \cdots$. When (i,j) changes by a bijection from (1,1) to (n,n), t_{ij} changes from 1 to n^2.

Let $A_n = \{(i/[n^\eta], j/[n^\eta]) : i = 1, 2, \cdots, [n^\eta], j = 1, 2, \cdots, [n^\eta]\}$, where η is a positive constant and $[x]$ is the integer part of x. Then, there are $[n^\eta]^2$ points in A_n, and for any $(x,y) \in [0,1] \times [0,1]$, there exists $(v(x), w(y)) \in A_n$ such that $|x - v(x)| \leq 1/[n^\eta]$ and $|y - w(y)| \leq 1/[n^\eta]$. Let $\beta_n > 0$ be an arbitrary sequence of constants diverging to infinity as $n \to \infty$. Then, we have

$$\frac{nh_n}{\beta_n \log n} \|\bar{g}_n(x,y) - E\bar{g}_n(x,y)\|_{D_{h_n}} \leq S_{1n} + S_{2n} + S_{3n},$$

where

$$\begin{aligned}
S_{1n} &= \frac{nh_n}{\beta_n \log n} \|\bar{g}_n(x,y) - \bar{g}_n(v(x), w(y))\|_{D_{h_n}} \\
S_{2n} &= \frac{nh_n}{\beta_n \log n} \|\bar{g}_n(v(x), w(y)) - E\bar{g}_n(v(x), w(y))\|_{D_{h_n}} \\
S_{3n} &= \frac{nh_n}{\beta_n \log n} \|E\bar{g}_n(v(x), w(y)) - E\bar{g}_n(x,y)\|_{D_{h_n}}.
\end{aligned}$$

We can choose η large enough such that

$$\begin{aligned}
S_{1n} &= \frac{nh_n}{\beta_n \log n} \left\| \frac{1}{n^2 h_n^2} \sum_{i=1}^n \sum_{j=1}^n \bar{\varepsilon}_{ij} \left[K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n} \right) - K\left(\frac{x_i - v(x)}{h_n}, \frac{y_j - w(y)}{h_n} \right) \right] \right\|_{D_{h_n}} \\
&\leq \frac{nh_n}{\beta_n \log n} \cdot \frac{n^2}{n^2 h_n^2} \cdot \left(n^2 \log^2(n^2) \right)^{1/3} \cdot \frac{C_K \sqrt{2}}{[n^\eta] h_n} = o(1).
\end{aligned}$$

By similar arguments, we have $S_{3n} = o(1)$. Now, for any $(x,y) \in D_{h_n}$,

$$\begin{aligned}
\bar{g}_n(x,y) - E\bar{g}_n(x,y) &= \sum_{i=1}^n \sum_{j=1}^n \frac{1}{n^2 h_n^2} (\bar{\varepsilon}_{ij} - E\bar{\varepsilon}_{ij}) K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n} \right) \\
&=: \sum_{i=1}^n \sum_{j=1}^n \tilde{g}_{n,ij}(x,y)
\end{aligned}$$

For any small constant $\epsilon > 0$, when n is large enough, we have

$$\begin{aligned}
& P\left(\frac{nh_n}{\beta_n \log n} [\bar{g}_n(x,y) - E\bar{g}_n(x,y)] > \epsilon \right) \\
\leq\ & \exp\left(-\epsilon \beta_n^{1/2} \log n \right) E\left\{ \Pi_{i=1}^n \Pi_{j=1}^n \exp\left[(nh_n/\beta_n^{1/2}) \tilde{g}_{n,ij}(x,y) \right] \right\} \\
\leq\ & n^{-\epsilon \beta_n^{1/2}} \Pi_{i=1}^n \Pi_{j=1}^n \left[1 + (n^2 h_n^2/\beta_n) \mathrm{Var}(\tilde{g}_{n,ij}(x,y)) \right] \\
\leq\ & n^{-\epsilon \beta_n^{1/2}} \Pi_{i=1}^n \Pi_{j=1}^n \exp\left((n^2 h_n^2/\beta_n) \mathrm{Var}(\tilde{g}_{n,ij}(x,y)) \right) \\
=\ & n^{-\epsilon \beta_n^{1/2}} \exp\left((n^2 h_n^2/\beta_n) \sum_{i=1}^n \sum_{j=1}^n \mathrm{Var}(\tilde{g}_{n,ij}(x,y)) \right) \qquad (25)
\end{aligned}$$

In (25), the second line is obtained by using the Chebyshev's inequality of the exponential form. The third line is obtained by the facts that

$$(nh_n/\beta_n^{1/2})\tilde{g}_{n,ij}(x,y) \leq (nh_n/\beta_n^{1/2})\frac{2(t_{nn}\log^2(t_{nn}))^{1/3}\|K\|}{n^2h_n^2} = O\left(\frac{1}{\beta_n^{1/2}}\left(\frac{\log^2(n)}{nh_n^3}\right)^{1/3}\right) = o(1)$$

and that $\exp(x) \leq 1+x+x^2$ when $|x| \leq 1/2$, where $\|K\|$ denotes the maximum value of the kernel function K in its support. The fourth line in (5.3) is based on the fact that $\exp(x) \geq 1+x$ when $x \geq 0$. Now,

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Var}(\tilde{g}_{n,ij}(x,y)) \leq \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\sigma^2}{(n^2h_n^2)^2}K^2\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right) = O\left(\frac{1}{n^2h_n^2}\right).$$

Combining this result and (25), we have

$$P\left(\frac{nh_n}{\beta_n\log n}\left[\bar{g}_n(x,y)-E\bar{g}_n(x,y)\right] > \epsilon\right) \leq O\left(n^{-\epsilon\beta_n^{1/2}}\right),$$

which is uniformly true for $(x,y) \in D_{h_n}$. By similar arguments, we have

$$\begin{aligned} & P\left(\frac{nh_n}{\beta_n\log n}\left[(-\bar{g}_n(x,y))-E(-\bar{g}_n(x,y))\right] > \epsilon\right) \\ = \;& P\left(\frac{nh_n}{\beta_n\log n}\left[\bar{g}_n(x,y)-E(\bar{g}_n(x,y))\right] < -\epsilon\right) \\ \leq \;& O\left(n^{-\epsilon\beta_n^{1/2}}\right). \end{aligned}$$

So,

$$P\left(\frac{nh_n}{\beta_n\log n}\left|\bar{g}_n(x,y)-E\bar{g}_n(x,y)\right| > \epsilon\right) \leq O\left(n^{-\epsilon\beta_n^{1/2}}\right),$$

and

$$P\left(S_{2n} > \epsilon\right) \leq [n^{\eta}]^2 \cdot O\left(n^{-\epsilon\beta_n^{1/2}}\right) = O\left(n^{-\epsilon\beta_n^{1/2}+2\eta}\right).$$

Consequently, $\sum_{n=1}^{\infty}P\left(S_{2n} > \epsilon\right) < \infty$. By the Borel-Cantelli Lemma, we have

$$S_{2n} = o(1),\ a.s.$$

Therefore

$$\frac{nh_n}{\beta_n\log n}\|\bar{g}_n(x,y)-E\bar{g}_n(x,y)\|_{D_{h_n}} = o(1),\quad a.s. \tag{26}$$

Next, we write

$$\|g_n^*(x,y)\|_{D_{h_n}} \leq \|g_n^*(x,y)-\bar{g}_n(x,y)\|_{D_{h_n}} + \|\bar{g}_n(x,y)-E\bar{g}_n(x,y)\|_{D_{h_n}} + \|E\bar{g}_n(x,y)\|_{D_{h_n}}. \tag{27}$$

Since $P(|\varepsilon_{ij}| > (t_{ij}\log^2(t_{ij}))^{1/3}) \leq (t_{ij}\log^2(t_{ij}))^{-1}E(|\varepsilon_{11}|^3)$ and

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(t_{ij}\log^2(t_{ij}))^{-1} = \sum_{\ell=1}^{\infty}(\ell\log^2(\ell))^{-1} < \infty,$$

we have

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P(|\varepsilon_{ij}| > (t_{ij}\log^2(t_{ij}))^{1/3}) < \infty.$$

By the Borel-Cantelli lemma, $P(|\varepsilon_{ij}| > (t_{ij}\log^2(t_{ij}))^{1/3}, i.o.) = 0$. Therefore, there exists a full set Ω_0 such that for each $\omega \in \Omega_0$ there exists a finite positive integer N_ω and when $\max(i,j) \geq N_\omega$ we have

$$\varepsilon_{ij}(\omega) = \bar{\varepsilon}_{ij}(\omega).$$

Hence, for any $(x,y) \in D_{h_n}$, when $n \geq N_\omega$,

$$\begin{aligned}
& |g_n^*(x,y) - \bar{g}_n(x,y)| \\
= & \left| \frac{1}{n^2h_n^2} \sum_{i=1}^{n}\sum_{j=1}^{n} (\varepsilon_{ij} - \bar{\varepsilon}_{ij}) K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n}\right) \right| \\
= & \left| \frac{1}{n^2h_n^2} \sum_{i=1}^{N_\omega}\sum_{j=1}^{N_\omega} (\varepsilon_{ij} - \bar{\varepsilon}_{ij}) K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n}\right) \right| \\
\leq & \sum_{i=1}^{N_\omega}\sum_{j=1}^{N_\omega} |\varepsilon_{ij} - \bar{\varepsilon}_{ij}| \frac{\|K\|}{n^2h_n^2}.
\end{aligned}$$

So,

$$\frac{nh_n}{\beta_n \log n} \|g_n^*(x,y) - \bar{g}_n(x,y)\|_{D_{h_n}} \leq \frac{nh_n}{\beta_n \log n} \cdot \frac{C(N_\omega, K)}{n^2h_n^2} = o(1), \tag{28}$$

where $C(N_\omega, K) = \|K\| \sum_{i=1}^{N_\omega}\sum_{j=1}^{N_\omega} |\varepsilon_{ij} - \bar{\varepsilon}_{ij}|$ is a factor depending on both N_ω and K.

Finally, for any $(x,y) \in D_{h_n}$,

$$\begin{aligned}
|E\bar{g}_n(x,y)| & \leq \frac{\|K\|}{n^2h_n^2} \sum \sum_{(x_i,y_j)\in B_{h_n}(x,y)} |E\bar{\varepsilon}_{ij}| \\
& = \frac{\|K\|}{n^2h_n^2} \sum \sum_{(x_i,y_j)\in B_{h_n}(x,y)} \left| E\varepsilon_{ij} I_{|\varepsilon_{ij}| \leq (t_{ij}\log^2(t_{ij}))^{1/3}} \right| \\
& = \frac{\|K\|}{n^2h_n^2} \sum \sum_{(x_i,y_j)\in B_{h_n}(x,y)} \left| E\varepsilon_{ij} I_{|\varepsilon_{ij}| > (t_{ij}\log^2(t_{ij}))^{1/3}} \right| \\
& \leq \frac{\|K\|}{n^2h_n^2} \sum \sum_{(x_i,y_j)\in B_{h_n}(x,y)} E\left(|\varepsilon_{ij}| I_{|\varepsilon_{ij}| > (t_{ij}\log^2(t_{ij}))^{1/3}} \right) \\
& \leq \frac{\|K\|}{n^2h_n^2} \sum \sum_{(x_i,y_j)\in B_{h_n}(x,y)} E|\varepsilon_{ij}|^3 (t_{ij}\log^2(t_{ij}))^{-2/3} \\
& \leq \frac{\|K\|}{n^2h_n^2} \sum \sum_{(x_i,y_j)\in B_{h_n}(h_n,h_n)} E|\varepsilon_{ij}|^3 (t_{ij}\log^2(t_{ij}))^{-2/3} \\
& \leq \frac{E(\varepsilon_{11}^3)\|K\|}{n^2h_n^2} \sum_{i=1}^{2nh_n+1} \sum_{j=1}^{2nh_n+1} (t_{ij}\log^2(t_{ij}))^{-2/3}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{E(\varepsilon_{11}^3)\|K\|}{n^2h_n^2} \sum_{\ell=1}^{(2nh_n+1)^2} (\ell \log^2(\ell))^{-2/3} \\
&= O\left(\frac{(nh_n)^{2/3}}{n^2h_n^2}\right),
\end{aligned}
$$

which is uniformly true for $(x,y) \in D_{h_n}$. In the above expression, $B_{h_n}(x,y)$ denotes a circle centered at (x,y) with radius h_n, $\log(1)$ is defined to be 1 in lines 5–8 for simplicity of presentation, and line 6 is obtained based on the facts that $x \geq h_n$ and $y \geq h_n$ when $(x,y) \in D_{h_n}$ and $\sum\sum_{(x_i,y_j)\in B_{h_n}(x,y)}(t_{ij}\log^2(t_{ij}))^{-2/3} \leq \sum\sum_{(x_i,y_j)\in B_{h_n}(h_n,h_n)}(t_{ij}\log^2(t_{ij}))^{-2/3}$ in such cases. Therefore,

$$
\frac{nh_n}{\beta_n \log n}\|E\bar{g}_n(x,y)\|_{D_{h_n}} = O\left(\frac{1}{\beta_n \log n}\cdot\frac{1}{(nh_n)^{1/3}}\right) = o(1). \tag{29}
$$

By combining (26)–(29), we have $\|g_n^*(x,y)\|_{D_{h_n}} = o\left(\frac{\beta_n \log n}{nh_n}\right),\ a.s.$.

Remark 1: For any $(x,y) \in D_{h_n}$, we have

$$
\begin{aligned}
&\frac{\sum_{i=1}^n\sum_{j=1}^n \varepsilon_{ij}K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right)}{\sum_{i=1}^n\sum_{j=1}^n K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right)} \\
= &\frac{\frac{1}{n^2h_n^2}\sum_{i=1}^n\sum_{j=1}^n \varepsilon_{ij}K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right)}{1+\left(\frac{1}{n^2h_n^2}\sum_{i=1}^n\sum_{j=1}^n K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right)-1\right)} \\
= &\frac{1}{n^2h_n^2}\sum_{i=1}^n\sum_{j=1}^n \varepsilon_{ij}K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right)\left[1-\left(\frac{1}{n^2h_n^2}\sum_{i=1}^n\sum_{j=1}^n K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right)-1\right)+o\left(\frac{1}{nh_n}\right)\right].
\end{aligned}
$$

So, by equations (23) and (24), we have

$$
\left\|\frac{\sum_{i=1}^n\sum_{j=1}^n \varepsilon_{ij}K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right)}{\sum_{i=1}^n\sum_{j=1}^n K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right)}\right\|_{D_{h_n}} = O\left(\frac{\log(n)}{nh_n}\right),\ a.s. \tag{30}
$$

By similar arguments to those in Proposition 2, it can be shown that, under conditions stated in Theorem 2, we have

$$
\left\|\frac{\sum_{i=1}^n\sum_{j=1}^n \left(\varepsilon_{ij}^2-\sigma^2\right)K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right)}{\sum_{i=1}^n\sum_{j=1}^n K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right)}\right\|_{D_{h_n}} = o(1),\ a.s. \tag{31}
$$

Proof of Theorem 2. First, we prove the second and third equations of (16). The first equation of (16) can be proved in a similar way. For $\ell = 1, 2$ and any $(x,y) \in D_{h_n}$, based on (5), we have

$$
e^{(\ell)}(x,y)
$$

$$
\begin{aligned}
&= \frac{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \left\{\varepsilon_{ij} + f(x_i,y_j) - \widehat{a}^{(\ell)}(x,y) - \widehat{b}^{(\ell)}(x,y)(x_i-x) - \widehat{c}^{(\ell)}(x,y)(y_j-y)\right\}^2 K(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n})}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} \\
&= \frac{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \varepsilon_{ij}^2 K(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n})}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} + \\
&\quad \frac{2\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \varepsilon_{ij}\left[f(x_i,y_j) - \widehat{a}^{(\ell)}(x,y) - \widehat{b}^{(\ell)}(x,y)(x_i-x) - \widehat{c}^{(\ell)}(x,y)(y_j-y)\right] K(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n})}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} + \\
&\quad \frac{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \left[f(x_i,y_j) - \widehat{a}^{(\ell)}(x,y) - \widehat{b}^{(\ell)}(x,y)(x_i-x) - \widehat{c}^{(\ell)}(x,y)(y_j-y)\right]^2 K(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n})}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} \\
&=: I_1^{(\ell)}(x,y) + I_2^{(\ell)}(x,y) + I_3^{(\ell)}(x,y) \qquad (32)
\end{aligned}
$$

By similar results to (31), we have

$$\|I_1^{(\ell)} - \sigma^2\|_{D_{h_n}} = o(1);\ a.s. \qquad (33)$$

By Taylor's expansion of $f(x_i,y_j)$ at the point (x,y), we have

$$
\begin{aligned}
&I_3^{(\ell)}(x,y) \\
&= \left(f(x,y) - \widehat{a}^{(\ell)}(x,y)\right)^2 + \\
&\quad \left(f_x'(x,y) - \widehat{b}^{(\ell)}(x,y)\right)^2 h_n^2 \frac{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \left(\frac{x_i-x}{h_n}\right)^2 K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} + \\
&\quad \left(f_y'(x,y) - \widehat{c}^{(\ell)}(x,y)\right)^2 h_n^2 \frac{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \left(\frac{y_j-y}{h_n}\right)^2 K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} + \\
&\quad 2\left(f(x,y) - \widehat{a}^{(\ell)}(x,y)\right)\left(f_x'(x,y) - \widehat{b}^{(\ell)}(x,y)\right) h_n \frac{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \frac{x_i-x}{h_n} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} + \\
&\quad 2\left(f(x,y) - \widehat{a}^{(\ell)}(x,y)\right)\left(f_y'(x,y) - \widehat{c}^{(\ell)}(x,y)\right) h_n \frac{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \frac{y_j-y}{h_n} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} + \\
&\quad 2\left(f_x'(x,y) - \widehat{b}^{(\ell)}(x,y)\right)\left(f_y'(x,y) - \widehat{c}^{(\ell)}(x,y)\right) h_n^2 \frac{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \frac{(x_i-x)(y_j-y)}{h_n^2} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} + \\
&\quad o(h_n^2)
\end{aligned}
$$

By equations (10)–(12) and the fact that the absolute value of each ratio of two summations appeared on the right hand side of the above equation is less than or equal to 1, we have

$$\|I_3^{(\ell)}\|_{D_{h_n}} = o(1),\ a.s. \qquad (34)$$

For $I_2^{(\ell)}$, by Taylor's expansion and equations (10)-(12), we have

$$\begin{aligned}
\|I_2^{(\ell)}\|_{D_{h_n}} \quad \leq \quad & 2\|f-\widehat{a}\|_{D_{h_n}} \left\| \frac{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \varepsilon_{ij} K(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n})}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} \right\|_{D_{h_n}} + \\
& 2h_n\|f'_x-\widehat{b}\|_{D_{h_n}} \left\| \frac{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \varepsilon_{ij} \frac{x_i-x}{h_n} K(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n})}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} \right\|_{D_{h_n}} + \\
& 2h_n\|f'_y-\widehat{c}\|_{D_{h_n}} \left\| \frac{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \varepsilon_{ij} \frac{y_j-y}{h_n} K(\frac{x_i-x}{h_n}, \frac{y_j-y}{p_n})}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} \right\|_{D_{h_n}} + \\
& o(h_n) \left\| \frac{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \varepsilon_{ij} K(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n})}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} \right\|_{D_{h_n}} \qquad (35)
\end{aligned}$$

Using similar arguments to those in the proof of Proposition 2, we can prove that

$$\left\| \frac{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \varepsilon_{ij} \frac{x_i-x}{h_n} K(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n})}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} \right\|_{D_{h_n}} = o(1), \ a.s.$$

and

$$\left\| \frac{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \varepsilon_{ij} \frac{y_j-y}{h_n} K(\frac{x_i-x}{h_n}, \frac{y_j-y}{p_n})}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} \right\|_{D_{h_n}} = o(1), \ a.s.$$

Therefore, by combining these results with (10)–(12), (30) and (35), we have

$$\|I_2^{(\ell)}\|_{D_{h_n}} = o(1), \ a.s. \qquad (36)$$

By (32)–(34) and (36), the second and third equations of (16) are proved.

Next, we assume that (x_τ, y_τ) is a given nonsingular point of a JLC, the projection of a point $(x,y) \in J_{h_n}$ to the JLC is (x_τ, y_τ) and the Euclidean distance between the two points is ch_n with $0 < c < 1$ a constant. In such cases, as in (5.10), we still write

$$e^{(\ell)}(x,y) = I_1^{(\ell)}(x,y) + I_2^{(\ell)}(x,y) + I_3^{(\ell)}(x,y).$$

Then, by (33), we have

$$I_1^{(\ell)}(x,y) = \sigma^2 + o(1), \ a.s. \qquad (37)$$

For $I_2^{(\ell)}(x,y)$, it can be written as

$$\begin{aligned}
I_2^{(\ell)}(x,y) \quad = \quad & \frac{2}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} \left(\sum_{(x_i,y_j)\in Q_n^{(1)}(x,y) \bigcap M_n^{(\ell)}(x,y)} + \sum_{(x_i,y_j)\in Q_n^{(2)}(x,y) \bigcap M_n^{(\ell)}(x,y)} \right) \\
& \varepsilon_{ij} \left[f(x_i,y_j) - \widehat{a}^{(\ell)}(x,y) - \widehat{b}^{(\ell)}(x,y)(x_i-x) - \widehat{c}^{(\ell)}(x,y)(y_j-y) \right] K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right) \\
=: \quad & I_{21}^{(\ell)}(x,y) + I_{22}^{(\ell)}(x,y). \qquad (38)
\end{aligned}$$

By equations (13)-(15),

$$
\begin{aligned}
I_{21}^{(\ell)}(x,y) &= \frac{2\sum_{(x_i,y_j)\in Q_n^{(1)}(x,y)\bigcap M_n^{(\ell)}(x,y)} \varepsilon_{ij}\left[f(x_i,y_j)-f_-(x_\tau,y_\tau)\right] K(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n})}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right)} - \\
&\quad \frac{(C_1+o(1))\sum_{(x_i,y_j)\in Q_n^{(1)}(x,y)\bigcap M_n^{(\ell)}(x,y)} \varepsilon_{ij} K(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n})}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right)} - \\
&\quad \frac{(C_2+o(1))\sum_{(x_i,y_j)\in Q_n^{(1)}(x,y)\bigcap M_n^{(\ell)}(x,y)} \varepsilon_{ij} \frac{x_i-x}{h_n} K(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n})}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right)} - \\
&\quad \frac{(C_3+o(1))\sum_{(x_i,y_j)\in Q_n^{(1)}(x,y)\bigcap M_n^{(\ell)}(x,y)} \varepsilon_{ij} \frac{y_j-y}{h_n} K(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n})}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right)} \\
&= o(1),\ a.s., \qquad (39)
\end{aligned}
$$

where C_1, C_2 and C_3 are constants. In (39), we have used the results that the ratio of two summations appeared in each of the four terms on the right hand side of the first equation is of order $o(1)$ almost surely, which is a consequence of (30). Similarly,

$$I_{22}^{(\ell)}(x,y) = o(1),\ a.s. \qquad (40)$$

Now, let $\widetilde{M}^{(\ell)}$, for $\ell = 1, 2$, be two halves of the support of the kernel function K separated by a line passing the center of the circular support in the direction perpendicular to the asymptotic direction of $(\widehat{b}(x,y), \widehat{c}(x,y))$, which is $\vec{g} = (\frac{1}{\widetilde{\beta}_{02}} \int\int_{Q^{(2)}} uK(u,v)\ dudv, \frac{1}{\widetilde{\beta}_{20}} \int\int_{Q^{(2)}} vK(u,v)\ dudv)$ (cf., equations (14) and (15)). Define $Q^{(1\ell)} = Q^{(1)} \bigcap \widetilde{M}^{(\ell)}$, $Q^{(2\ell)} = Q^{(2)} \bigcap \widetilde{M}^{(\ell)}$, $A_{0\ell} = \int\int_{Q^{(2\ell)}} K(u,v)\ dudv$, $A_{1\ell} = \int\int_{Q^{(2\ell)}} uK(u,v)\ dudv$, and $A_{2\ell} = \int\int_{Q^{(2\ell)}} vK(u,v)\ dudv$, where $Q^{(1)}$ and $Q^{(2)}$ are defined in Theorem 1. By the isotropic property of K and by similar arguments to those in the proof of equation (23), we have

$$\left\| \frac{1}{n^2h_n^2} \sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right) - \frac{1}{2} \right\|_{D_{h_n}} = o(1). \qquad (41)$$

Also, for a function $\phi(x,y)$ satisfying the condition that $\sup_{x^2+y^2\le 1} |\phi(x,y)| \le b_\phi < \infty$, we have

$$
\begin{aligned}
&\left| \frac{1}{n^2h_n^2} \sum_{(x_i,y_j)\in Q_n^{(1)}(x,y)\bigcap M_n^{(\ell)}(x,y)} \phi\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right) K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right) - \right. \\
&\left. \frac{1}{n^2h_n^2} \sum_{(x_i,y_j)\in Q_n^{(1)}(x,y)\bigcap \widetilde{M}_n^{(\ell)}(x,y)} \phi\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right) K\left(\frac{x_i-x}{h_n},\frac{y_j-y}{h_n}\right) \right|
\end{aligned}
$$

$$
\begin{aligned}
&\leq b_\phi \|K\| \frac{1}{n^2 h_n^2} \sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)\Delta \widetilde{M}_n^{(\ell)}(x,y)} 1 \\
&= O(\theta_n) = o(1)\ a.s., \qquad (42)
\end{aligned}
$$

where $M_n^{(\ell)}(x,y)\Delta\widetilde{M}_n^{(\ell)}(x,y) = (M_n^{(\ell)}(x,y)\backslash\widetilde{M}_n^{(\ell)}(x,y))\bigcup(\widetilde{M}_n^{(\ell)}(x,y)\backslash M_n^{(\ell)}(x,y))$, $\widetilde{M}_n^{(\ell)}(x,y)$, for $\ell = 1,2$, are two halves of the neighborhood $M_n(x,y)$ separated by a line passing the center of $M_n(x,y)$ in the direction perpendicular to $\vec{g}$, and θ_n is the acute angle between the two directions $(\widehat{b}(x,y),\widehat{c}(x,y))$ and $\vec{g}$. The last equation in (42) is a direct conclusion of equations (14) and (15).

Using (41), (42), and results about $\widehat{a}^{(\ell)}(x,y), \widehat{b}^{(\ell)}(x,y)$ and $\widehat{c}^{(\ell)}(x,y)$ similar to those in (13)–(15), we have

$$
\begin{aligned}
& I_3^{(\ell)}(x,y) \\
= & \frac{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \left[f(x_i,y_j) - \widehat{a}^{(\ell)}(x,y) - \widehat{b}^{(\ell)}(x,y)(x_i-x) - \widehat{c}^{(\ell)}(x,y)(y_j-y)\right]^2 K(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n})}{\sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} K\left(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}\right)} \\
= & \frac{2}{n^2h_n^2} \sum_{(x_i,y_j)\in M_n^{(\ell)}(x,y)} \left[f(x_i,y_j) - f_-(x_\tau,y_\tau) - d_\tau A_{0\ell} - \frac{d_\tau A_{1\ell}}{\widetilde{\beta}_{02}}\frac{x_i-x}{h_n} - \right. \\
& \left. \frac{d_\tau A_{2\ell}}{\widetilde{\beta}_{20}}\frac{y_j-y}{h_n}\right]^2 K(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}) + o(1) \\
= & \frac{2}{n^2h_n^2} \left(\sum_{(x_i,y_j)\in Q_n^{(1)}(x,y)\bigcap M_n^{(\ell)}(x,y)} + \sum_{(x_i,y_j)\in Q_n^{(2)}(x,y)\bigcap M_n^{(\ell)}(x,y)} \right) \\
& \left[f(x_i,y_j) - f_-(x_\tau,y_\tau) - d_\tau A_{0\ell} - \frac{d_\tau A_{1\ell}}{\widetilde{\beta}_{02}}\frac{x_i-x}{h_n} - \frac{d_\tau A_{2\ell}}{\widetilde{\beta}_{20}}\frac{y_j-y}{h_n}\right]^2 K(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}) + o(1) \\
= & \frac{2}{n^2h_n^2} \left(\sum_{(x_i,y_j)\in Q_n^{(1)}(x,y)\bigcap \widetilde{M}_n^{(\ell)}(x,y)} + \sum_{(x_i,y_j)\in Q_n^{(2)}(x,y)\bigcap \widetilde{M}_n^{(\ell)}(x,y)} \right) \\
& \left[f(x_i,y_j) - f_-(x_\tau,y_\tau) - d_\tau A_{0\ell} - \frac{d_\tau A_{1\ell}}{\widetilde{\beta}_{02}}\frac{x_i-x}{h_n} - \frac{d_\tau A_{2\ell}}{\widetilde{\beta}_{20}}\frac{y_j-y}{h_n}\right]^2 K(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}) + o(1) \\
= & \frac{2}{n^2h_n^2} \sum_{(x_i,y_j)\in Q_n^{(1)}(x,y)\bigcap \widetilde{M}_n^{(\ell)}(x,y)} \left[-d_\tau A_{0\ell} - \frac{d_\tau A_{1\ell}}{\widetilde{\beta}_{02}}\frac{x_i-x}{h_n} - \frac{d_\tau A_{2\ell}}{\widetilde{\beta}_{20}}\frac{y_j-y}{h_n}\right]^2 K(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}) + \\
& \frac{2}{n^2h_n^2} \sum_{(x_i,y_j)\in Q_n^{(2)}(x,y)\bigcap \widetilde{M}_n^{(\ell)}(x,y)} \left[d_\tau - d_\tau A_{0\ell} - \frac{d_\tau A_{1\ell}}{\widetilde{\beta}_{02}}\frac{x_i-x}{h_n} - \frac{d_\tau A_{2\ell}}{\widetilde{\beta}_{20}}\frac{y_j-y}{h_n}\right]^2 K(\frac{x_i-x}{h_n}, \frac{y_j-y}{h_n}) \\
& + o(1) \\
= & 2d_\tau^2 \int\int_{Q^{(1\ell)}} \left[A_{0\ell} + \frac{A_{1\ell}s}{\widetilde{\beta}_{02}} + \frac{A_{2\ell}t}{\widetilde{\beta}_{20}}\right]^2 K(s,t)\ dsdt +
\end{aligned}
$$

$$
\begin{aligned}
& 2d_\tau^2 \int\int_{Q^{(2\ell)}} \left[1 - A_{0\ell} - \frac{A_{1\ell}s}{\widetilde{\beta}_{02}} - \frac{A_{2\ell}t}{\widetilde{\beta}_{20}}\right]^2 K(s,t)\, dsdt + o(1) \\
= \;& d_\tau^2 \left(C_\tau^{(\ell)}\right)^2 + o(1),\ a.s. \qquad (43)
\end{aligned}
$$

where

$$
\begin{aligned}
C_\tau^{(\ell)} \;=\; & \left\{ 2\int\int_{Q^{(1\ell)}} \left[A_{0\ell} + \frac{A_{1\ell}s}{\widetilde{\beta}_{02}} + \frac{A_{2\ell}t}{\widetilde{\beta}_{20}}\right]^2 K(s,t)\, dsdt + \right. \\
& \left. 2\int\int_{Q^{(2\ell)}} \left[1 - A_{0\ell} - \frac{A_{1\ell}s}{\widetilde{\beta}_{02}} - \frac{A_{2\ell}t}{\widetilde{\beta}_{20}}\right]^2 K(s,t)\, dsdt \right\}^{1/2}.
\end{aligned}
$$

In the fifth equation of (43), we have used the facts that (i) $f(x_i,y_j) - f_-(x_\tau,y_\tau) = O\left(\frac{1}{nh_n}\right)$ when $(x_i,y_j) \in Q_n^{(1)}(x,y)$, and (ii) $f(x_i,y_j) - f_-(x_\tau,y_\tau) = d_\tau + O\left(\frac{1}{nh_n}\right)$ when $(x_i,y_j) \in Q_n^{(2)}(x,y)$.

By expressions (37)–(40) and (43), the second and third equations of (17) are proved. The first equation of (17) can be proved in a similar way, after using the properties of the kernel function that (i) $\int\int_{Q^{(1)}} K(u,v)\ dudv = 1 - \int\int_{Q^{(2)}} K(u,v)\ dudv$, (ii) $\int\int_{Q^{(1)}} uK(u,v)\ dudv = -\int\int_{Q^{(2)}} uK(u,v)\ dudv$, and (iii) $\int\int_{Q^{(1)}} vK(u,v)\ dudv = -\int\int_{Q^{(2)}} vK(u,v)\ dudv$.

Proof of Theorem 3. The estimator $\widehat{f}_1(x,y)$ is one of $\widehat{a}^{(1)}(x,y)$ and $\widehat{a}^{(2)}(x,y)$ when $(x,y) \in D_{h_n}$. By results similar to those in Theorem 1, both of them are uniformly consistent for estimating f in D_{h_n}. Therefore, equation (18) is true.

Proof of Theorem 4. The estimator $\widehat{f}_2(x,y)$ is one of $\widehat{a}(x,y)$, $\widehat{a}^{(1)}(x,y)$, $\widehat{a}^{(2)}(x,y)$, and $(\widehat{a}^{(1)}(x,y) + \widehat{a}^{(2)}(x,y))/2$, all of which are uniformly consistent for estimating f in D_{h_n} (cf., (10)). Therefore, equation (19) is true.

For a given nonsingular point (x_τ,y_τ) on a JLC, suppose that the projection of a point (x,y) to the JLC is (x_τ,y_τ), the Euclidean distance between the two points is ch_n with $0 < c < 1$ a constant, without loss of generality the tangent line of the JLC at (x_τ,y_τ) is assumed to be parallel to the y-axis, and (x,y) is on the left side of (x_τ,y_τ). Then, by (17), we have $e^{(1)}(x,y) = \sigma^2 + o(1),\ a.s.$, $e^{(2)}(x,y) = \sigma^2 + d_\tau^2(C_\tau^{(2)})^2 + o(1),\ a.s.$, and $e(x,y) = \sigma^2 + d_\tau^2 C_\tau^2 + o(1),\ a.s.$, where $(C_\tau^{(2)})^2$ and C_τ^2 are positive constants. Therefore, when $d_\tau/\sigma > 1/C_\tau$ and n is large enough, $e^{(1)}(x,y) < e(x,y)/2$ and $e^{(1)}(x,y) < e^{(2)}(x,y)$. Consequently, $\widehat{f}_2(x,y)$ equals $\widehat{a}_1(x,y)$. So, equation (20) is true by results similar to (10). When $d_\tau/\sigma \le 1/C_\tau$, $e(x,y)/2 \le \min(e^{(1)}(x,y), e^{(2)}(x,y))$. By (7), $\widehat{f}_2(x,y) = \widehat{a}(x,y)$ in such a case. So equation (21) is true according to (13).

Proof of Theorem 5. For a given point $(x,y) \in D_{h_{n1}+h_{n2}}$, we can write

$$
\widehat{a}(x,y;\underline{\widehat{f}}_1) - f(x,y) = \left(\widehat{a}(x,y;\underline{\widehat{f}}_1) - \widehat{a}(x,y;\underline{f})\right) + \left(\widehat{a}(x,y;\underline{f}) - f(x,y)\right)
$$

where $\underline{f}$ denotes the vector of $\{f(i/n, j/n), i, j = 1, 2, \ldots, n\}$. So

$$\|\widehat{a}(x,y;\widehat{\underline{f}}_1) - f(x,y)\|_{D_{h_{n1}+h_{n2}}} \leq \|\widehat{a}(x,y;\widehat{\underline{f}}_1) - \widehat{a}(x,y;\underline{f})\|_{D_{h_{n1}+h_{n2}}} + \|\widehat{a}(x,y;\underline{f}) - f(x,y)\|_{D_{h_{n1}+h_{n2}}}. \tag{44}$$

Now, the local linear kernel estimator $\widehat{a}(x,y;\underline{z})$, which uses kernel function K and bandwidth h_n, has the following expression (cf., e.g., equation (2.2) in Qiu (2004)):

$$\widehat{a}(x,y;\underline{z}) = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(x,y) z_{ij}}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(x,y)},$$

where

$$\begin{aligned}
w_{ij}(x,y) &= \left[A^{(1)}(x,y) + A^{(2)}(x,y)(x_i - x) + A^{(3)}(x,y)(y_j - y)\right] K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n}\right) \\
A^{(1)}(x,y) &= B^{(20)}(x,y)B^{(02)}(x,y) - B^{(11)}(x,y)B^{(11)}(x,y) \\
A^{(2)}(x,y) &= B^{(01)}(x,y)B^{(11)}(x,y) - B^{(10)}(x,y)B^{(02)}(x,y) \\
A^{(3)}(x,y) &= B^{(10)}(x,y)B^{(11)}(x,y) - B^{(01)}(x,y)B^{(20)}(x,y) \\
B^{(r_1 r_2)}(x,y) &= \sum_{i=1}^n \sum_{j=1}^n (x_i - x)^{r_1} (y_j - y)^{r_2} K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n}\right), \text{ for } r_1, r_2 = 0, 1, 2.
\end{aligned}$$

Similar to (23), we have, for $r_1, r_2 = 0, 1, 2$,

$$\left\| \frac{1}{n^2 h_n^2} \sum_{i=1}^n \sum_{j=1}^n \left(\frac{x_i - x}{h_n}\right)^{r_1} \left(\frac{y_j - y}{h_n}\right)^{r_2} K\left(\frac{x_i - x}{h_n}, \frac{y_j - y}{h_n}\right) - \widetilde{\beta}_{r_1 r_2} \right\|_{D_{h_{n1}+h_{n2}}} = O\left(\frac{1}{n h_n}\right), \tag{45}$$

where $\widetilde{\beta}_{r_1 r_2} = \int_{-1}^1 \int_{-1}^1 u^{r_1} v^{r_2} K(u,v)\, du dv$. Since the kernel function K is assumed to be an isotropic bivariate density function, we have $\widetilde{\beta}_{01} = \widetilde{\beta}_{10} = 0$. By these results, we have

$$\begin{aligned}
\left\| \frac{1}{n^4 h_n^8} A^{(1)}(x,y) - C_K^* \right\|_{D_{h_{n1}+h_{n2}}} &= o(1), \\
\left\| \frac{1}{n^4 h_n^7} A^{(2)}(x,y) \right\|_{D_{h_{n1}+h_{n2}}} &= o(1), \\
\left\| \frac{1}{n^4 h_n^7} A^{(3)}(x,y) \right\|_{D_{h_{n1}+h_{n2}}} &= o(1),
\end{aligned} \tag{46}$$

where $C_K^* = \widetilde{\beta}_{20}\widetilde{\beta}_{02} - \widetilde{\beta}_{11}^2 > 0$. By these results and equation (18), we have

$$\begin{aligned}
& \left\| \widehat{a}(x,y;\widehat{\underline{f}}_1) - \widehat{a}(x,y;\underline{f}) \right\|_{D_{h_{n1}+h_{n2}}} \\
= & \left\| \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(x,y) \left(\widehat{f}_1(x_i,y_j) - f(x_i,y_j)\right)}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(x,y)} \right\|_{D_{h_{n1}+h_{n2}}} \\
= & \left\| \frac{\sum_{i=1}^n \sum_{j=1}^n \left[n^4 h_{n2}^8 C_K^* + o(n^4 h_{n2}^8)\right] \left(\widehat{f}_1(x_i,y_j) - f(x_i,y_j)\right) K(\frac{x_i - x}{h_{n2}}, \frac{y_j - y}{h_{n2}})}{\sum_{i=1}^n \sum_{j=1}^n \left[n^4 h_{n2}^8 C_K^* + o(n^4 h_{n2}^8)\right] K(\frac{x_i - x}{h_{n2}}, \frac{y_j - y}{h_{n2}})} \right\|_{D_{h_{n1}+h_{n2}}} \\
\leq & \left\| \widehat{f}_1 - f \right\|_{D_{h_{n1}+h_{n2}}} (1 + o(1)) = O(h_{n1}^2) + O\left(\frac{\log(n)}{n h_{n1}}\right) \ a.s.,
\end{aligned} \tag{47}$$

where h_{n1} and h_{n2} are two bandwidths used in the two steps of (8). Since $f(x,y)$ is assumed to have continuous second order derivatives at any $(x,y) \in D_{h_{n1}+h_{n2}}$, by Taylor's expansion of $f(x_i,y_j)$ at (x,y) and by (45) and (46), we have

$$
\begin{aligned}
& \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x,y)\left[f(x_i,y_j)-f(x,y)\right] \\
= \;& n^4 h_{n2}^8 C_K^* \sum_{i=1}^{n}\sum_{j=1}^{n}(1+o(1))\left[f(x_i,y_j)-f(x,y)\right] K\left(\frac{x_i-x}{h_{n2}},\frac{y_j-y}{h_{n2}}\right) \\
= \;& n^4 h_{n2}^8 C_K^* \sum_{i=1}^{n}\sum_{j=1}^{n}(1+o(1))\left[f_x'(x,y)(x_i-x)+f_y'(x,y)(y_j-y)\right] K\left(\frac{x_i-x}{h_{n2}},\frac{y_j-y}{h_{n2}}\right)+ \\
& \frac{1}{2} n^4 h_{n2}^8 C_K^* \sum_{i=1}^{n}\sum_{j=1}^{n}(1+o(1))\left[f_{xx}''(x,y)(x_i-x)^2+2f_{xy}''(x,y)(x_i-x)(y_j-y)+\right. \\
& \left. f_{yy}''(x,y)(y_j-y)^2\right] K\left(\frac{x_i-x}{h_{n2}},\frac{y_j-y}{h_{n2}}\right)+n^4 h_{n2}^8 C_K^*(1+o(1))o(n^2h_{n2}^4) \\
= \;& n^4 h_{n2}^8 C_K^*(1+o(1)) n^2 h_{n2}^3 \left[f_x'(x,y)\frac{1}{n^2h_{n2}^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{x_i-x}{h_{n2}}\right) K\left(\frac{x_i-x}{h_{n2}},\frac{y_j-y}{h_{n2}}\right)+\right. \\
& \left. f_y'(x,y)\frac{1}{n^2h_{n2}^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{y_j-y}{h_{n2}}\right) K\left(\frac{x_i-x}{h_{n2}},\frac{y_j-y}{h_{n2}}\right)\right]+ \\
& \frac{1}{2} n^4 h_{n2}^8 C_K^*(1+o(1)) n^2 h_{n2}^4 \left[f_{xx}''(x,y)\frac{1}{n^2h_{n2}^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{x_i-x}{h_{n2}}\right)^2 K\left(\frac{x_i-x}{h_{n2}},\frac{y_j-y}{h_{n2}}\right)+\right. \\
& 2f_{xy}''(x,y)\frac{1}{n^2h_{n2}^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{x_i-x}{h_{n2}}\right)\left(\frac{y_j-y}{h_{n2}}\right) K\left(\frac{x_i-x}{h_{n2}},\frac{y_j-y}{h_{n2}}\right)+ \\
& \left. f_{yy}''(x,y)\frac{1}{n^2h_{n2}^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{y_j-y}{h_{n2}}\right)^2 K\left(\frac{x_i-x}{h_{n2}},\frac{y_j-y}{h_{n2}}\right)\right]+n^4 h_{n2}^8 C_K^*(1+o(1))o(n^2h_{n2}^4) \\
= \;& n^4 h_{n2}^8 C_K^*(1+o(1)) n^2 h_{n2}^3 \left[f_x'(x,y)(\widetilde{\beta}_{10}+O(\frac{1}{nh_{n2}}))+f_y'(x,y)(\widetilde{\beta}_{01}+O(\frac{1}{nh_{n2}}))\right]+ \\
& \frac{1}{2} n^4 h_{n2}^8 C_K^*(1+o(1)) n^2 h_{n2}^4 \left[f_{xx}''(x,y)\widetilde{\beta}_{20}+2f_{xy}''(x,y)\widetilde{\beta}_{11}+f_{yy}''(x,y)\widetilde{\beta}_{02}+o(1)\right]+ \\
& n^4 h_{n2}^8 C_K^*(1+o(1))o(n^2h_{n2}^4) \\
= \;& n^4 h_{n2}^8 C_K^*(1+o(1)) n^2 h_{n2}^3\, o(h_{n2})+\frac{1}{2} n^4 h_{n2}^8 C_K^*(1+o(1)) n^2 h_{n2}^4 \left[f_{xx}''(x,y)\widetilde{\beta}_{20}+\right. \\
& \left. 2f_{xy}''(x,y)\widetilde{\beta}_{11}+f_{yy}''(x,y)\widetilde{\beta}_{02}+o(1)\right]+n^4 h_{n2}^8 C_K^*(1+o(1))o(n^2h_{n2}^4). \qquad (48)
\end{aligned}
$$

In the last equation of (48), we have used the results $\widetilde{\beta}_{01} = \widetilde{\beta}_{10} = 0$ and the condition that $1/nh_{n2}^2 = o(1)$. Obviously, (48) is uniformly true for $(x,y) \in D_{h_{n1}+h_{n2}}$. Similarly, we have

$$
\begin{aligned}
& \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x,y) \\
= \;& n^4 h_{n2}^8 C_K^*(1+o(1)) n^2 h_{n2}^2 \frac{1}{n^2h_{n2}^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{x_i-x}{h_{n2}},\frac{y_j-y}{h_{n2}}\right)
\end{aligned}
$$

$$= \quad n^4 h_{n2}^8 C_K^* (1+o(1)) n^2 h_{n2}^2 (1+o(1)), \tag{49}$$

which is uniformly true for $(x,y) \in D_{h_{n1}+h_{n2}}$. By (48) and (49), we have

$$\begin{aligned} & \|\widehat{a}(x,y;\underline{f}) - f(x,y)\|_{D_{h_{n1}+h_{n2}}} \\ = & \left\| \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(x,y)\,[f(x_i,y_j) - f(x,y)]}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(x,y)} \right\|_{D_{h_{n1}+h_{n2}}} \\ = & \quad O(h_{n2}^2). \end{aligned} \tag{50}$$

By (44), (47), and (50) and the condition that $h_{n2} \sim h_{n1}$, we have

$$\left\|\widehat{a}(x,y;\widehat{\underline{f}}_1) - f(x,y)\right\|_{D_{h_{n1}+h_{n2}}} = O(h_{n1}^2) + O\left(\frac{\log(n)}{nh_{n1}}\right),\ a.s.$$

Similar results can be derived for $\widehat{a}^{(1)}(x,y;\widehat{\underline{f}}_1)$ and $\widehat{a}^{(2)}(x,y;\widehat{\underline{f}}_1)$. Since $\widehat{f}(x,y;\widehat{\underline{f}}_1)$ is one of $\widehat{a}(x,y;\widehat{\underline{f}}_1)$, $\widehat{a}^{(1)}(x,y;\widehat{\underline{f}}_1)$, $\widehat{a}^{(2)}(x,y;\widehat{\underline{f}}_1)$, and $(\widehat{a}^{(1)}(x,y;\widehat{\underline{f}}_1) + \widehat{a}^{(2)}(x,y;\widehat{\underline{f}}_1))/2$, equation (22) is proved.

6 Concluding Remarks

We have introduced three jump-preserving surface reconstruction procedures (6)-(8). Procedure (6) preserves the jumps well but its estimated surface is quite noisy compared to the estimated surfaces of the other two procedures. Procedure (7) preserves the jumps well and also smooths away the noise efficiently when the signal-to-noise ratio is high. When this ratio is low, its ability to preserve jumps is limited. Procedure (8) is a combination of procedures (6) and (7). Numerical examples show that it preserves the nonsingular points of the JLCs well and also smooths away the noise efficiently. Theoretically, it has been proved that its estimated surface is uniformly, strongly consistent in continuity regions of f.

Acknowledgments: The author thanks two referees for many constructive comments and suggestions which greatly improved the quality of the paper. This research is partially supported by a grant from the National Security Agency and a grant from the National Science Foundation of USA.

References

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.

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

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.

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.

Gijbels, I., Lambert, A., and Qiu, P. (2006), "Edge-preserving image denoising and estimation of discontinuous surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **28**, 1075–1087.

Gijbels, I., Lambert, A., and Qiu, P. (2007), "Jump-preserving regression and smoothing using local linear fitting: a compromise," *Annals of the Institute of Statistical Mathematics*, **59**, 235–272.

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., and Molchanov (2003), "Sequential methods for design-adaptive estimation of discontinuities in regression curves and surfaces," *The Annals of Statistics*, **31**, 921–941.

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.

Keeling, S.L., and Stollberger, R. (2002), "Nonlinear anisotropic diffusion filtering for multiscale edge enhancement," *Inverse Problems*, **18**, 175–190.

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

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., 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.

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

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. (2003), "A jump-preserving curve fitting procedure based on local piecewise-linear kernel estimation," *Journal of Nonparametric Statistics*, **15**, 437–453.

Qiu, P. (2004), "The local piecewisely linear kernel smoothing procedure for fitting jump regression surfaces," *Technometrics*, **46**, 87–98.

Qiu, P. (2005), *Image Processing and Jump Regression Analysis,* New York: John Wiley & Sons.

Qiu, P. (2007), "Jump surface estimation, edge detection, and image restoration," *Journal of the American Statistical Association*, **102**, 745–756.

Qiu, P., and Yandell, B. (1997), "Jump detection in regression surfaces," *Journal of Computational and Graphical Statistics*, **6**, 332–354.

Sebastiani, G., and Godtliebsen, F. (1997), "On the use of Gibbs priors for Bayesian image restoration," *Signal Processing*, **56**, 111–118.

Stone, C. (1982), "Optimal global rates of convergence for nonparametric regression," *The Annals of Statistics*, **10**, 1040–1053.

Sun, J., and Qiu, P. (2007), "Jump detection in regression surfaces using both first-order and second-order derivatives," *Journal of Computational and Graphical Statistics*, **16**, 289–311.

Titterington, D.M. (1985), "Common structure of smoothing techniques in statistics", *International Statistical Review*, **53**, 141–170.

Tomasi, C., and Manduchi, R. (1998), "Bilateral Filtering for Gray and Color Images," *Proceedings of the 1998 IEEE International Conference on Computer Vision*, 839–846, Bombay, India.

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

Weikert, J., ter Haar Romeny, B.M., and Viergever, M. (1998), "Efficient and reliable schemes for nonlinear diffusion filtering," *IEEE Transactions on Image Processing*, **7**, 398–410.

Yi, J.H., and Chelberg, D.M. (1995), "Discontinuity-preserving and viewpoint invariant reconstruction of visible surfaces using a first order regularization", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**, 624–629.