

Boise State University

ScholarWorks

---

Mathematics Faculty Publications and  
Presentations

Department of Mathematics

---

11-2018

## Efficient Blind Image Deblurring Using Nonparametric Regression and Local Pixel Clustering

Yicheng Kang  
*Bentley University*

Partha Sarathi Mukherjee  
*Boise State University*

Peihua Qiu  
*University of Florida*

# Efficient Blind Image Deblurring Using Nonparametric Regression And Local Pixel Clustering

Yicheng Kang

Department of Mathematical Sciences, Bentley University

Partha Sarathi Mukherjee

Department of Mathematics, Boise State University

and

Peihua Qiu

Department of Biostatistics, University of Florida

November 27, 2017

## Abstract

Blind image deblurring is a challenging ill-posed problem. It would have an infinite number of solutions even in cases when an observed image contains no noise. In reality, however, observed images almost always contain noise. The presence of noise would make the image deblurring problem even more challenging because the noise can cause numerical instability in many existing image deblurring procedures. In this paper, a novel blind image deblurring approach is proposed, which can remove both pointwise noise and spatial blur efficiently without imposing restrictive assumptions on either the point spread function (psf) or the true image. It even allows the psf to be location dependent. In the proposed approach, a local pixel clustering procedure is used to handle the challenging task of restoring complicated edge structures that are tapered by blur, and a nonparametric regression procedure is used for removing noise at the same time. Numerical examples show that our proposed method can effectively handle a wide variety of blur and it works well in applications.

*Keywords:* Blind image deblurring; Clustering; Deconvolution; Denoising; Edges; Image reconstruction; Smoothing; Surface estimation; Nonparametric regression.

# 1 Introduction

Observed images are not always faithful representations of the scenes that we see. As a matter of fact, some sort of degradation often arises when recording a digital image. For instance, in astronomical imaging, the incoming light in the telescope is often bent by atmospheric turbulence. In aerial reconnaissance, the optical system in camera lens could be out of focus. In our daily life, image distortion often arises in cases when there is a relative motion between a camera and an object. Environmental effects such as scattered and reflected light also degrade images. Other sources of degradations include device noise (e.g., charge-coupled device sensor and circuitry) and quantization noise. See Bates and McDonnell (1989) and Gonzalez and Woods (2008) for a detailed discussion about formation and description of various degradations. Classically, image degradation is modeled as the result of two phenomena (Aubert and Kornprobst, 2006). The first one is related to the image acquisition (e.g., *blur* created by motion). The second one is random and corresponds to the *noise* coming from signal transmission.

In the literature, a commonly used model for describing the relationship between the true image  $f$  and its degraded version  $Z$  is as follows.

$$Z(x, y) = G\{f\}(x, y) + \varepsilon(x, y), \quad \text{for } (x, y) \in \Omega, \quad (1)$$

where  $G\{f\}(x, y) = \int \int_{\mathcal{R}^2} g(u, v; x, y) f(x - u, y - v) dudv$  denotes the convolution between a 2-D point spread function (psf)  $g$  and a true image intensity function  $f$ ,  $\varepsilon(x, y)$  is the pointwise noise at  $(x, y)$ , and  $\Omega$  is the design space of the image. In model (1), it is assumed that the true image  $f$  is degraded spatially by  $g$  and pointwise by  $\varepsilon$ , the spatial blur is linear, and the pointwise noise is additive. In most references, it is further assumed that the psf  $g$ , which describes the blurring mechanism, is location (or spatially) invariant. That is,  $g(u, v; x, y)$  does not depend on  $(x, y)$ .

Blind image deblurring (BID) is for estimating  $f$  from  $Z$  when the psf  $g$  is not completely specified. This problem is ill-posed in nature because only  $Z$  is observed in (1), all  $g$ ,  $f$  and  $\varepsilon$  are unobservable, and it is impossible to distinguish  $(g, f)$  from  $(ag, a^{-1}f)$  based on the observed image  $Z$  alone, for any constant  $a \neq 0$ . This ill-posed nature would get even worse in cases when  $g$  changes over location. In the literature, some image deblurring

procedures have been developed under the assumption that the psf  $g$  is completely known. Such procedures are often referred to be *non-blind*. The main difficulty in non-blind image deblurring lies behind the removal of blur in presence of noise (cf., Qiu (2005), Chapter 7). To overcome this difficulty, a number of image deblurring techniques have been proposed using the regularization framework (e.g., Chan and Wong (1998), Figueiredo and Nowak (2003), Oliveira et al. (2009), Rudin et al. (1992), You and Kaveh (1996)). In practice, however, it is hard to specify the psf  $g$  completely. In cases when the assumed psf is different from the true psf, it has been shown that the deblurred image could be seriously distorted (cf., Qiu (2005), Chapter 7). To avoid such limitations, a number of BID methods have been developed in the literature. Some of them assume that  $g$  follows a parametric model with one or more unknown parameters, and these parameters are estimated together with the true image  $f$  by certain algorithms (e.g., Carasso (2001), Carasso (2003), Hall and Qiu (2007b), Joshi and Chaudhuri (2005), Katsaggelos and Lay (1990)). Some others assume that the true image  $f$  has one or more regions with certain known edge structures or the image's edge structures can be estimated reasonably well (e.g., Hall and Qiu (2007a), Kang and Qiu (2014), Kundur and Hatzinakos (1998), Qiu (2008), Qiu and Kang (2015), Yang et al. (1994)). Some BID methods adopt the Bayesian framework to make the originally ill-posed BID problem well-posed by imposing some prior information on the psf or on the true image (e.g., Fergus et al. (2006), Miskin and MacKay (2000), Skilling (1989)). Some other BID methods estimate both  $g$  and  $f$  in an alternating fashion, using the iterative Richardson-Lucy scheme (e.g., Biggs and Andrews (1997), Jansson (1997)).

This paper proposes an alternative approach to the BID problem based on the observation that *spatial blur alters the image structure most dramatically around step edges and least dramatically at places where the true image intensity surface is straight*. Based on this observation, our proposed approach focuses on deblurring around step edges. More specifically, it works as follows. In a neighborhood of a given pixel, if we conclude based on a data-driven criterion that there could be step edges in the neighborhood, then all pixels are clustered into two groups. In such cases, the image intensity at the given pixel is estimated by a weighted average of all image intensities in the group that the given pixel belongs to. If we conclude that there are no step edges in the neighborhood, then

the image intensity at the given pixel is estimated by a weighted average of all image intensities in the entire neighborhood. One major feature of this approach is that it does not require any restrictive assumptions on either  $g$  or  $f$ . It even allows  $g$  to vary over location. Numerical comparisons with some representatives of the state-of-the-art image deblurring methods show that the proposed method is capable of handling a wide variety of blur and it works well in various applications. The proposed method can be accomplished by the functions `surfaceCluster()` and `surfaceCluster_bandwidth()` in the R-package *DRIP* (<https://cran.r-project.org/web/packages/DRIP/>). The test images used in Section 3 are also included in the package.

The rest part of the paper is organized as follows. Our proposed methodology is described in detail in Section 2. Some numerical examples are presented in Section 3. Several remarks conclude the paper in Section 4.

## 2 Methodology

We describe our proposed BID method in two parts. In Subsection 2.1, our new method is described in detail. In Subsection 2.2, selection of procedure parameters is discussed.

### 2.1 Proposed BID Method

Assume that an observed image follows the discretized version of model (1)

$$Z_{ij} = G\{f\}(i, j) + \varepsilon_{ij}, \quad \text{for } i, j = 1, 2, \dots, n,$$

where  $(i, j)$  denote the  $(i, j)$ -th equally spaced pixel (i.e., the pixel located at  $(i/n, j/n)$ ) in the design space  $\Omega = [0, 1] \times [0, 1]$ ,  $\{Z_{ij}, i, j = 1, 2, \dots, n\}$  are observed image intensities, and  $\{\varepsilon_{ij}, i, j = 1, 2, \dots, n\}$  are independent and identically distributed (i.i.d.) random errors with mean 0 and unknown variance  $\sigma^2$ . It is further assumed that  $f$  is continuous in  $\Omega$  except on some edge curves (see Qiu (1998) for a mathematical definition of edge curves).

For the  $(x, y)$ -th pixel, let us consider its circular neighborhood

$$O(x, y; k, n) = \left\{ (i, j) : \sqrt{(x - i)^2 + (y - j)^2} \leq k, \quad 1 \leq i, j \leq n \right\},$$

where the positive integer  $k \leq n$  is a bandwidth parameter and  $(x, y)$  denotes the two-dimensional index of the design point (we will also refer to the pixel or its location as  $(x, y)$  and its meaning should be no ambiguity from the context). In this neighborhood, a local plane is fitted by the following local linear kernel (LLK) smoothing procedure (cf., Fan and Gijbels (1996)):

$$\min_{a,b,c} \left\{ \sum_{i=1}^n \sum_{j=1}^n \left[ Z_{ij} - a - \frac{b}{n}(i-x) - \frac{c}{n}(j-y) \right]^2 K \left( \frac{i-x}{k}, \frac{j-y}{k} \right) \right\}, \quad (2)$$

where  $K$  is a circularly symmetric bivariate density kernel function with its support on the unit disk. The above LLK smoothing procedure approximates the image intensity surface locally by a plane and uses the kernel function  $K$  to control the weights in the weighted least squares procedure (2). Usually,  $K$  is chosen such that pixels closer to  $(x, y)$  receive more weights, which is intuitively reasonable because pixels closer to  $(x, y)$  should provide more information about the image intensity at  $(x, y)$ . Let  $(\hat{a}_n(x, y), \hat{b}_n(x, y), \hat{c}_n(x, y))$  denote the solution to the minimization problem (2). The mathematical expressions are shown in (12) - (14) in the appendix. Then,  $\hat{a}_n(x, y)$  in (12) is the LLK estimator of  $f(x, y)$ , and  $\hat{b}_n(x, y)$  and  $\hat{c}_n(x, y)$  are the LLK estimators of the  $x$  and  $y$  derivatives of  $f$  at  $(x, y)$ , respectively, in cases when such derivatives exist.

The LLK estimator removes noise but also blurs edges at the same time. Center weighted median (CWM) filtering is a useful method in image processing and it can preserve edges to some extent (Ko and Lee, 1991; Sun et al., 1994). Next, a CWM filter with center weight  $W_{0,0}$  is applied to  $O(x, y; k, n)$  and the filter output at  $(x, y)$  is denoted by  $\tilde{a}_n(x, y)$ . The residual at  $(x, y)$  is defined by

$$e_n(x, y) = \hat{a}_n(x, y) - \tilde{a}_n(x, y). \quad (3)$$

If  $(x, y)$  is in the continuity region of  $f$ , then the image structure within  $O(x, y; k, n)$  should be approximated well by the local plane described by  $(\hat{a}_n(x, y), \hat{b}_n(x, y), \hat{c}_n(x, y))$ . Thus,  $e_n(x, y)$  should be relatively small. On the other hand, if  $O(x, y; k, n)$  contains edge curves, then the fitted local plane cannot well describe the image structure within  $O(x, y; k, n)$ . Consequently, the value of  $e(x, y)$  should be relatively large. Therefore,  $e_n(x, y)$  can be used to judge whether the neighborhood  $O(x, y; k, n)$  contains any edge curves. More

specifically, if

$$|e_n(x, y)| > u_n, \quad (4)$$

then we can conclude that there are edge curves in  $O(x, y; k, n)$ , where  $u_n$  is a threshold value. In such a case, we can cluster the pixels in  $O(x, y; k, n)$  into two groups based on their CWM outputs. Intuitively, pixels on the same side of an edge curve have similar CWM outputs. So, they can be put in the same group. Pixels on different sides of the edge curve have quite different CWM outputs, and they should be put in different groups. Of course, it is not easy to specify the exact position of the edge curve within  $O(x, y; k, n)$  and define the two groups of pixels accordingly. But, an informative pixel clustering procedure can generate groups such that pixels within a group are similar in their CWM outputs and pixels in different groups have quite different CWM outputs. Such a pixel clustering procedure can reflect the local edge structure well without imposing restrictive conditions on the smoothness or shape of the edge curve. In this paper, we suggest a simple but effective pixel clustering procedure which uses a cut-off constant  $c$  to define the two clusters in  $O(x, y; k, n)$ . More specifically, the two clusters are defined to be

$$\begin{aligned} O_1(x, y; k, n, c) &= \{(i, j) \in O(x, y; k, n) : \tilde{a}_n(i, j) \leq c\}, \\ O_2(x, y; k, n, c) &= \{(i, j) \in O(x, y; k, n) : \tilde{a}_n(i, j) > c\}, \end{aligned}$$

where  $c \in R(x, y; k, n)$ , and  $R(x, y; k, n)$  is the range of the image intensity values in  $O(x, y; k, n)$  defined to be

$$R(x, y; k, n) = \min_{O(x, y; k, n)} \tilde{a}_n(i, j), \max_{O(x, y; k, n)} \tilde{a}_n(i, j) .$$

So, it is obvious that both  $O_1(x, y; k, n, c)$  and  $O_2(x, y; k, n, c)$  are non-empty sets for any constant  $c \in R(x, y; k, n)$ ,  $O(x, y; k, n, c) = O_1(x, y; k, n, c) \cup O_2(x, y; k, n, c)$ , and  $O_1(x, y; k, n, c) \cap O_2(x, y; k, n, c) = \emptyset$ . Let  $c_0$  be the maximizer to the following maximization problem:

$$\max_{c \in R(x, y; k, n)} \frac{|O_1(x, y; k, n, c)|(\bar{\eta}_1 - \bar{\eta})^2 + |O_2(x, y; k, n, c)|(\bar{\eta}_2 - \bar{\eta})^2}{\sum_{O_1(x, y; k, n, c)} (\tilde{a}_n(i, j) - \bar{\eta}_1)^2 + \sum_{O_2(x, y; k, n, c)} (\tilde{a}_n(i, j) - \bar{\eta}_2)^2}, \quad (5)$$

where  $|A|$  denotes the number of elements in the pointset  $A$ ,  $\bar{\eta}_s$  denotes the sample mean of the CWM outputs within  $O_s(x, y; k, n, c)$ , for  $s = 1, 2$ , and  $\bar{\eta}$  denotes the sample mean

of the CWM outputs within  $O(x, y; k, n)$ . In (5), the numerator measures the dissimilarity between the two groups, and the denominator measures the dissimilarity within each of the two groups. Thus, it is reasonable to cluster the pixels in  $O(x, y; k, n, c)$  by maximizing their ratio. It can be checked that (5) is actually the one dimensional version of the well-known clustering criterion proposed by Friedman and Rubin (1967). Note that there are only finitely many cut-off constants that can result in different partitions. Namely, it is sufficient to evaluate (5) on the finite set of  $\{\tilde{a}_n(i, j) : (i, j) \in O(x, y; k, n)\}$ . Therefore, the maximization problem can be solved by exhaustive search.

Without loss of generality, assume that  $(x, y) \in O_1(x, y; k, n, c_0)$ . Then, a weighted average of observations in  $O_1(x, y; k, n, c_0)$  should provide a good estimate for  $f(x, y)$  when there is no blurring involved, as discussed in the image denoising literature (cf., Qiu 1998). In cases when the observed image contains blur, if the intensity value of a pixel is closer to the cut-off constant  $c_0$ , then it should receive less weight in the weighted average since it is more likely that that pixel has blur involved. To address this issue related to the image blur, besides a bivariate kernel function used in the conventional kernel smoothing procedure to assign more weights to pixels closer to  $(x, y)$ , a univariate kernel function is used to assign less weights to pixels whose intensity values are closer to  $c_0$ . Then, our proposed BID estimator  $\hat{f}_n(x, y)$  is defined to be the solution to  $a_0$  in the following local constant kernel (LCK) smoothing procedure:

$$\min_{a_0 \in \mathcal{R}} \sum_{O_1(x, y; k, n, c_0)} (Z_{ij} - a_0)^2 K \left( \frac{i-x}{k}, \frac{j-y}{k} \right) L \left( \frac{|\tilde{a}_n(i, j) - c_0|}{|\tilde{a}_{min}^{(1)} - c_0|} \right), \quad (6)$$

where  $L$  is a univariate increasing density kernel function with the support  $[0, 1]$ , and  $\tilde{a}_{min}^{(1)}$  denotes the minimum CWM output in  $O_1(x, y; k, n, c_0)$ . It is easy to check that  $\hat{f}_n(x, y)$  has the following expression:

$$\hat{f}_n(x, y) = \frac{\sum_{O_1(x, y; k, n, c_0)} Z_{ij} K \left( \frac{i-x}{k}, \frac{j-y}{k} \right) L \left( \frac{|\tilde{a}_n(i, j) - c_0|}{|\tilde{a}_{min}^{(1)} - c_0|} \right)}{\sum_{O_1(x, y; k, n, c_0)} K \left( \frac{i-x}{k}, \frac{j-y}{k} \right) L \left( \frac{|\tilde{a}_n(i, j) - c_0|}{|\tilde{a}_{min}^{(1)} - c_0|} \right)}. \quad (7)$$

In cases when  $(x, y) \in O_2(x, y; k, n, c_0)$ ,  $\hat{f}_n(x, y)$  can be defined in the same way except that  $O_1(x, y; k, n, c_0)$  and  $\tilde{a}_{min}^{(1)}$  in (7) should be replaced by  $O_2(x, y; k, n, c_0)$  and  $\tilde{a}_{max}^{(2)}$ , respectively, where  $\tilde{a}_{max}^{(2)}$  denotes the maximum CWM output in  $O_2(x, y; k, n, c_0)$ .



To demonstrate the efficacy of the image deblurring procedure (7), a cross section of an image around a step edge, a blurred version, a blurred-and-noisy version, and the deblurred version by (7) when  $K$  and  $L$  are chosen to be the ones used in Section 3 are shown in plots (a)-(d) of Figure 1, respectively. From plot (d), it can be seen that (7) can restore the blurred edge structure to some extent while removing the noise at the same time.

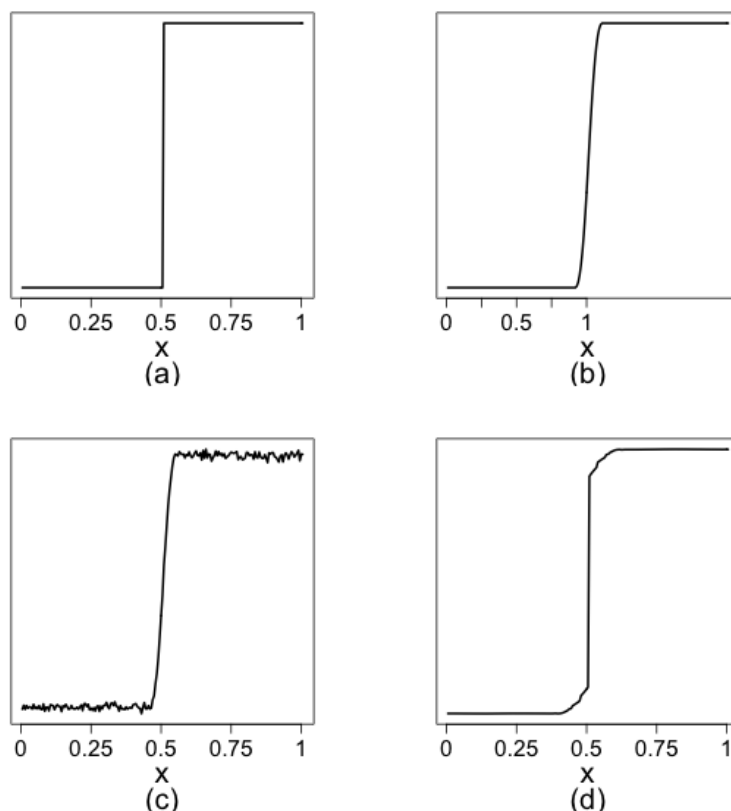


Figure 1: (a): A cross section of an image intensity surface around a step edge; (b): A blurred version of (a); (c): A blurred-and-noisy version of (a); (d): The deblurred version from (c) by the BID procedure (7).

In cases when (4) is not satisfied, it is likely that the pixel  $(x, y)$  is in a continuity region of  $f$ . In such cases, the spatial blur would not alter the image intensity surface much, as discussed in Section 1. So, we suggest estimating  $f(x, y)$  by the conventional LLK estimator  $\hat{a}_n(x, y)$  defined in (12). There are two benefits of doing this. First, it has been well demonstrated in the literature that the LLK estimator has less bias compared to the LCK estimator in continuity regions of  $f$  (cf., Fan and Gijbels (1996)). Second,

since  $\widehat{a}_n(x, y)$  has already been computed before we compute  $\widehat{f}_n(x, y)$  in (10), it saves much computation.

## 2.2 Parameter Selection

In the proposed BID procedure (5)–(7), there are two parameters to choose, including the threshold value  $u_n$  in (4) and the bandwidth parameter  $k$  in (2). To choose a reasonable value for  $u_n$ , we need to derive the asymptotic distribution of  $e_n(x, y)$  defined in (3). Based on (3), we have the following proposition.

**Proposition 1.** *Let  $\Psi(\cdot)$  and  $\psi(\cdot)$  denote the cumulative distribution function (cdf) and the probability density function of  $\varepsilon_{11}$ , respectively. Under the regularity conditions (B1) – (B5) in Appendix B and assume that  $(x, y)$  is a continuity point, then*

$$k \left[ \begin{pmatrix} \widehat{a}_n(x, y) \\ \widetilde{a}_n(x, y) \end{pmatrix} - \begin{pmatrix} G\{f\}(x, y) \\ G\{f\}(x, y) \end{pmatrix} \right] \xrightarrow{d} N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right),$$

where

$$\begin{aligned} \Sigma_{11} &= \sigma^2 \int \int_{x^2+y^2 \leq 1} K(x, y)^2 dx dy, \\ \Sigma_{12} &= \Sigma_{21} = \frac{E|\varepsilon_{11}|}{2\pi\psi(0)}, \\ \Sigma_{22} &= \frac{1}{4\pi\psi(0)^2}, \end{aligned}$$

and  $\xrightarrow{d}$  denotes convergence in distribution as  $n \rightarrow \infty$ .

The proof of Proposition 1 is provided in Appendix B. If  $(x, y)$  is a continuity point, it follows from Proposition 1 that

$$k(\widehat{a}_n(x, y) - \widetilde{a}_n(x, y)) \xrightarrow{d} N(0, \Sigma_{11} + \Sigma_{22} - 2\Sigma_{12}).$$

So a reasonable choice for the threshold in (4) would be

$$u_n = Z_{1-\alpha/2}(\Sigma_{11} + \Sigma_{22} - 2\Sigma_{12})/k, \tag{8}$$

where  $Z_{1-\alpha/2}$  denotes the  $1 - \alpha/2$  percentile of the standard normal distribution and  $\alpha$  is the significance level. There are still three unknown quantities in (8),  $\psi(0)$ ,  $\sigma^2$ , and

$E|\varepsilon_{11}|$ , to be estimated. We suggest applying a surface estimator (e.g., Qiu 2009) to obtain residuals  $\{\widehat{\varepsilon}_{11}, \dots, \widehat{\varepsilon}_{nn}\}$ . Then  $\psi(0)$ ,  $\sigma^2$ , and  $E|\varepsilon_{11}|$  can be estimated by

$$\widehat{\psi(0)} = \frac{1}{n^2 h_n} \sum_{i,j=1}^n K_1 \left( \frac{\widehat{\varepsilon}_{ij} - 0}{h_n} \right), \quad \widehat{\sigma^2} = \frac{1}{n^2} \sum_{i,j=1}^n \widehat{\varepsilon}_{ij}^2, \quad \widehat{E|\varepsilon_{11}|} = \frac{1}{n^2} \sum_{i,j=1}^n |\widehat{\varepsilon}_{ij}|, \quad (9)$$

where  $K_1(\cdot)$  is the one-dimensional Gaussian kernel and  $h_n = 1.06\widehat{\sigma}n^{-2/5}$  (Wand and Jones 1994, Chapter 2).

Next, we discuss the selection of the bandwidth parameter  $k$ . In numerical simulations, the true image is often known. In such cases,  $k$  can be chosen by minimizing

$$\text{MSE}(f, \widehat{f}; k) = \frac{1}{n^2} \sum_{i,j=1}^n \left[ f(i, j) - \widehat{f}_n(i, j) \right]^2, \quad (10)$$

where  $\widehat{f}_n$  is the deblurred image. In practice,  $f$  is usually unknown. In such cases, the cross validation (CV) approach is natural to consider (cf., Qiu 2005, Chapter 2). In the image deblurring problem, however, the mean response is  $G\{f\}$ , instead of  $f$ . In such cases, the CV approach is inappropriate to use because the chosen parameter is for approximating  $G\{f\}$ . To overcome this limitation of the conventional CV approach, we propose the following modified cross validation (MCV) approach.

$$\text{MCV}(k) = \frac{w}{|\Omega \setminus J|} \sum_{\Omega \setminus J} \left( Z_{ij} - \widehat{f}_{-Z_{ij}}(i, j) \right)^2 + \frac{1-w}{|J|} \sum_J \left( \widetilde{Z}_{ij} - \widehat{f}_{-\widetilde{Z}_{ij}}(i, j) \right)^2, \quad (11)$$

where  $w \in [0, 1]$  is a constant,  $J = \{(i, j) : |e_n(i, j)| > u_n\}$ ,  $\widehat{f}_{-Z_{ij}}(x, y)$  denotes the proposed the BID estimate at  $(x, y)$  with the observation  $Z_{ij}$  held out, and  $\widetilde{Z}_{ij}$  is the image intensity whose CWM output equal to  $\tilde{a}_{min}^{(1)}$  in (7) or  $\tilde{a}_{max}^{(2)}$  in (7)'s alternative form in the cases when  $(i, j) \in O_2(i, j; k, n, c_0)$ . The rationale behind (11) is again based on our key observation that spatial blur alters image most dramatically around step edges and least dramatically at places where the true image intensity surface is straight. More specifically, the intensities of the pixels in the continuity region are not altered much by blur. Thus the first term in (11) uses the conventional leave-one-out CV approach. As for the pixels around step edges, their observed image intensities are no longer representative of  $f$ . We approximate them with a nearby pixel's image intensity that is not affected dramatically by blur (i.e.,  $\widetilde{Z}_{ij}$  in the second term of (11)). MCV is a weighted average of the two and  $w$  represents the relative importance of the first term. It needs to be specified by the user.

By taking into account all these considerations, our proposed BID procedure is summarized below.

### Proposed Blind Image Deblurring Procedure

1. For a given pixel  $(x, y)$ , solve the minimization problem (2) and compute its solution by (12)-(14).
2. Apply CWM filter and obtain  $\tilde{a}_n(x, y)$
3. Compute the residual  $e_n(x, y)$  in (3).
4. If (4) holds, then execute the local clustering procedure by solving the maximization problem (5), and estimate  $f(x, y)$  by (7). Otherwise, estimate  $f(x, y)$  by (12).

## 3 Numerical Study

In this section, we discuss several numerical examples concerning the performance of the proposed BID procedure and the MCV bandwidth selection procedure. Throughout this section, the center weight  $W_{0,0}$  in the CWM filter is chosen to be 3, the significance level  $\alpha$  is chosen to be 0.001, the relative weight in (11) is 0.5, the two dimensional kernel function  $K$  used in (2) and (7) is chosen to be  $(2/\pi)(1-x^2-y^2)I(x^2+y^2 \leq 1)$ , and the one dimensional kernel function  $L$  used in (7) is chosen to be  $(1/1.194958) \exp(x^2/2)I(0 \leq x \leq 1)$ . We choose these two kernel functions because the former is the Epanechnikov kernel function, which is a standard choice in the statistical literature, and the latter is a truncated Gaussian kernel function, which is commonly used in the computer science literature.

### 3.1 Numerical Experiment with Lena Image

We denote the proposed BID procedure as NEW and compare it with three other popular methods. The first existing method considered here is the one accomplished by the MATLAB blind deconvolution routine *deconvblind*, which is based on the method discussed by Biggs and Andrews (1997) and Jansson (1997) under the framework of Richardson-Lucy (RL) algorithm. The second existing method is the total variation (TV) image deblurring

method proposed by Oliveira et al. (2009). The third existing method is the blind image deconvolution procedure developed under the Bayesian framework by Fergus et al. (2006). These three existing methods are denoted as RL, TV and Bayes, respectively. It should be pointed out that both RL and Bayes are blind image deblurring schemes, but TV is designed for non-blind image deblurring. Two versions of TV, denoted as  $TV_1$  and  $TV_2$ , distinguished by how the psf  $g$  is specified, are considered. The specific description of  $TV_1$  and  $TV_2$  will be given later. The bandwidth  $k$  used in (2) is chosen by minimizing (10).

The Lena test image has  $512 \times 512$  pixels. The following two psf's are considered:

$$g_1(u, v; x, y) = \begin{cases} \frac{1}{C_1(x,y)} \exp\{-\frac{u^2+v^2}{2}\} I(u^2 + v^2 \leq 0.01^2) & \text{if } y > 0.5, \\ \delta_0(u)\delta_0(v) & \text{otherwise;} \end{cases}$$

$$g_2(u, v; x, y) = \begin{cases} \frac{1}{C_2(x,y)} I(|u| \leq 0.01)\delta_0(v) & \text{if } |x - 0.5| \leq 0.3 \text{ and } |y - 0.5| \leq 0.3, \\ \frac{1}{C_2(x,y)} \delta_0(u) I(|v| \leq 0.1) & \text{otherwise,} \end{cases}$$

where  $C_j(x, y)$  is the standardization constant such that  $\int_{\mathcal{R}^2} g_j(u, v; x, y) dudv = 1$ , for any  $(x, y) \in \Omega$  and  $j = 1, 2$ , and  $\delta_0(\cdot)$  is the delta function with the point mass at 0. The random noise is generated from the normal distribution  $N(0, \sigma^2)$ , and two different noise levels  $\sigma = 5$  and  $10$  are considered. From the above expression, we can see that  $g_1$  is a truncated Gaussian blur for the upper half of the image and there is no blur for the lower half;  $g_2$  is a horizontal motion blur for the central part of the image and is a vertical motion blur for the rest part of the image. In the case when psf is  $g_1$ ,  $TV_1$  and  $TV_2$  denotes the TV method when the psf is specified as the Gaussian blur of  $g_1$  and the delta function (i.e., no blur) of  $g_1$ , respectively. In the case when psf is  $g_2$ ,  $TV_1$  and  $TV_2$  denotes the TV method when the psf is specified as the horizontal motion blur of  $g_2$  and the vertical motion blur of  $g_2$ , respectively.

Figure 2(a)-(c) present the original Lena image, its blurred version with  $g_2$ , and its blurred-and-noisy version with  $g_2$  and  $\sigma = 10$ , respectively. Figure 2(d)-(h) present the deblurred images by NEW, RL,  $TV_1$ ,  $TV_2$  and Bayes, respectively. It should be pointed out that the support of the psf needs to be specified when using RL and the true support of  $g_2$  is used in this example to show its best performance, and a subregion defined by the coordinates  $[86/512, 214/512] \times [293/512, 421/512]$  is prespecified for Bayes, as suggested

in Fergus et al. (2006) that their algorithm would perform better and run faster if a smaller patch, rich in edge structure, is manually selected. From Figure 2, it can be seen that (i) NEW removes noise and the blur well, (ii) there are many artifacts in the deblurred image of RL and the noise has not been reduced much, and (iii) TV generates many artifacts at places where the psf is misspecified.



Figure 2: (a)–(c): Original Lena image, its blurred version and its blurred-and-noisy version, respectively. The RMSE of (c) and (d) is 15.12 and 12.51, respectively. (d)–(h): Deblurred images by NEW, RL,  $TV_1$ ,  $TV_2$  and Bayes, respectively.

Next, we compare the five methods quantitatively. Table 1 presents the values of the root mean squared error (RMSE) defined to be  $RMSE = \sqrt{\frac{1}{n} \sum_{i,j=1}^n [f(i,j) - \hat{f}_n(i,j)]^2}$  of the five methods for each case considered based on 100 replicated simulations. The number in each parenthesis represents the standard error of the corresponding RMSE. From Table 1, it can be seen that NEW outperforms all the other four methods.

Table 1: Estimated values of RMSE of the five image deblurring methods in the Lena image example based on 100 replicated simulations. The numbers in the parentheses are the standard errors of RMSE.

Methods	$g_1$		$g_2$	
	$\sigma = 5$	$\sigma = 10$	$\sigma = 5$	$\sigma = 10$
New	11.06 (0.02)	11.47 (0.02)	12.28 (0.02)	12.53 (0.02)
RL	19.80 (0.03)	30.73 (0.05)	26.61 (0.04)	29.67 (0.08)
TV <sub>1</sub>	15.83 (0.10)	16.33 (0.27)	24.21 (0.16)	26.31 (0.39)
TV <sub>2</sub>	12.25 (0.09)	12.61 (0.02)	13.00 (0.04)	13.89 (0.15)
Bayes	26.43 (1.54)	38.11 (1.08)	27.93 (1.11)	40.91 (1.95)

### 3.2 Numerical Experiment with Peppers Image

Next, we discuss the second numerical example, in which the test image of peppers with  $256 \times 256$  pixels is used. The psf  $g$  considered has the expression:

$$g(u, v; x, y) = \frac{3}{\pi r^2(x, y)} \left( 1 - \sqrt{\frac{u^2}{r^2(x, y)} + \frac{v^2}{r^2(x, y)}} \right) I(u^2 + v^2 \leq r^2(x, y)),$$

where  $r(x, y) > 0$  may change over location and it is the radius of the circular support of  $g$ . In this paper,  $r(x, y)$  is called the *blur extent* function. Three blur extent functions,  $r_1(x, y) = 0.03(1 - (x - 0.5)^2 - (y - 0.5)^2)$ ,  $r_2(x, y) = 0.03x$ ,  $r_3(x, y) = 0.02$ , and two noise levels,  $\sigma = 5$ ,  $\sigma = 10$ , are considered. Clearly,  $r_1(x, y)$  and  $r_2(x, y)$ , are location variant., and  $r_3(x, y)$  is location invariant. In the case with  $r_3(x, y)$ , the blur described by  $g(u, v; x, y)$  is homogeneous across the entire image, which is the case discussed by most references. As in the previous example, the noise is generated from the distribution  $N(0, \sigma^2)$ . Regarding the four image deblurring methods, we would like to make the following remarks. (i) RL requires the blur extent function to be constant (i.e., location invariant) and completely specified. So, in this example, we searched the value of  $r$  to achieve the minimum RMSE such that RL performs the best. (ii) TV requires the psf  $g$  to be completely specified and the blur extent function needs to be constant as well. In this example the value of  $r$  is searched to achieve the minimum RMSE values for TV, for which the parametric

form of  $g$  is correctly specified. (iii) The prespecified subregion for Bayes is chosen to be  $[78/256, 206/256] \times [42/256, 170/256]$ .

The results in the same setup as Figure 2 are shown in Figure 3, where the blur extent function  $r_2(x, y)$  and  $\sigma = 10$  are considered. From the figure, it can be seen that (i) the blur gets more severe when moving from the left side of the image to the right side (cf., plot(b)), (ii) NEW deblurs the image and removes the noise well, (iii) RL performs poorly, (iv) the middle part of the deblurred image by TV looks good but the places near the boundary contain many artifacts because TV cannot handle location variant blur, (v) Bayes performs poorly in this example. It is worth noting that the RMSE of the deblurred image is larger than that of the observed image. The reason is as follows. The blur extent changes rapidly as the pixel location moving from the left to the right. At the places close to the left boundary of the image, where there is little blur involved, our deblurring procedure is still carried out nonetheless. And it results in large RMSE in those areas. It can be seen from  $\text{RMSE}(f, \hat{f}_n)_{|x < 0.25} = \frac{\sum_{i < n/4} \sum_{j=1}^n (f(i, j) - \hat{f}_n(i, j))^2 / (n^2/4)}{n^2/4} = 20.75$  and  $\text{RMSE}(f, Z)_{|x < 0.25} = 10.37$ . On the other hand, at the places where the blur is severe (i.e., close to the right boundary of the image), our deblurring procedure does improve on the observed image with  $\text{RMSE}(f, \hat{f}_n)_{|x > 0.75} = 21.64$  and  $\text{RMSE}(f, Z)_{|x > 0.75} = 22.02$ . This reveals a limitation of the proposed deblurring method that it is nonadaptive in the sense that it does not adjust for different blur extents at different pixel locations.

In cases when  $r_3(x, y)$  (i.e., blur is location invariant) and  $\sigma = 10$  are considered, the results are shown in Figure 4. In this case, TV method makes full use of the completely specified blurring mechanism and thus can be considered as the gold standard. From Figure 4, we can see that (i) both RL and Bayes perform poorly, (ii) TV performs well as expected and (iii) NEW still gives a comparable performance to TV despite it uses much less prior information.

The quantitative performance measures of these methods in the same setup as that of Table 1 are presented in Table 2. It can be seen from Table 2 that (i) NEW works stably as the blur extent function and noise level change, (ii) TV, which requires the parametric form of the psf is correctly specified, works slightly better than NEW in a few cases, (iii) in the cases when the blur extent function is location varying  $r_1(x, y)$ , TV is still performing



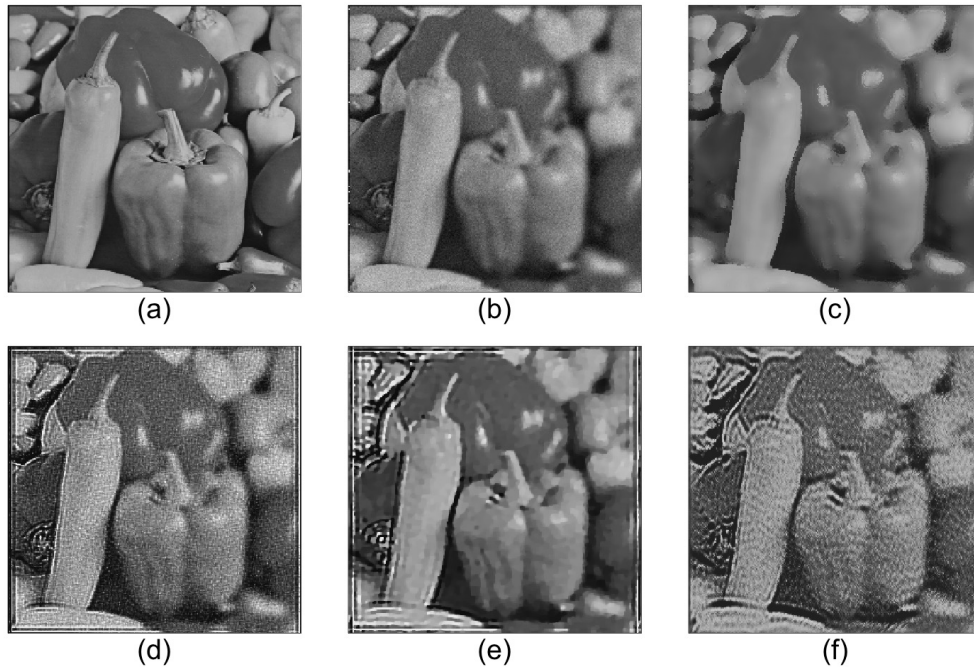


Figure 3: (a)–(b): Original peppers image and its blurred-and-noisy version, respectively. The RMSE of (b) and (c) is 16.99 and 19.43, respectively. (c)–(f): Deblurred images by NEW, RL, TV, and Bayes, respectively.

because  $r_1(x, y)$  changes slowly across the image, whereas its performance deteriorates significantly as the blur function changes a little more rapidly (i.e., when the blur function is  $r_2(x, y)$ ), and (iii) RL and Bayes both perform poorly.

### 3.3 Numerical Experiment with Brain Image

Next, we consider an example with a brain test image. Figure 5(a) shows an observed brain image with  $217 \times 217$  pixels which seems to have some blur involved. Its noisy version is shown in Figure 5(b), where the noise is generated from  $N(0, 7^2)$ . Figure 5(c)–(f) present the deblurred images by NEW, RL, TV and Bayes, respectively. The bandwidth in NEW is chosen to be  $4/217$ . The support of the psf for RL is chosen to give its best visual impression. For TV, the psf is specified as a horizontal motion blur and the blur extent is chosen to give the best visual impression. We also tried several other forms of psf for TV but they did not provide significant improvements. The prespecified subregion required by

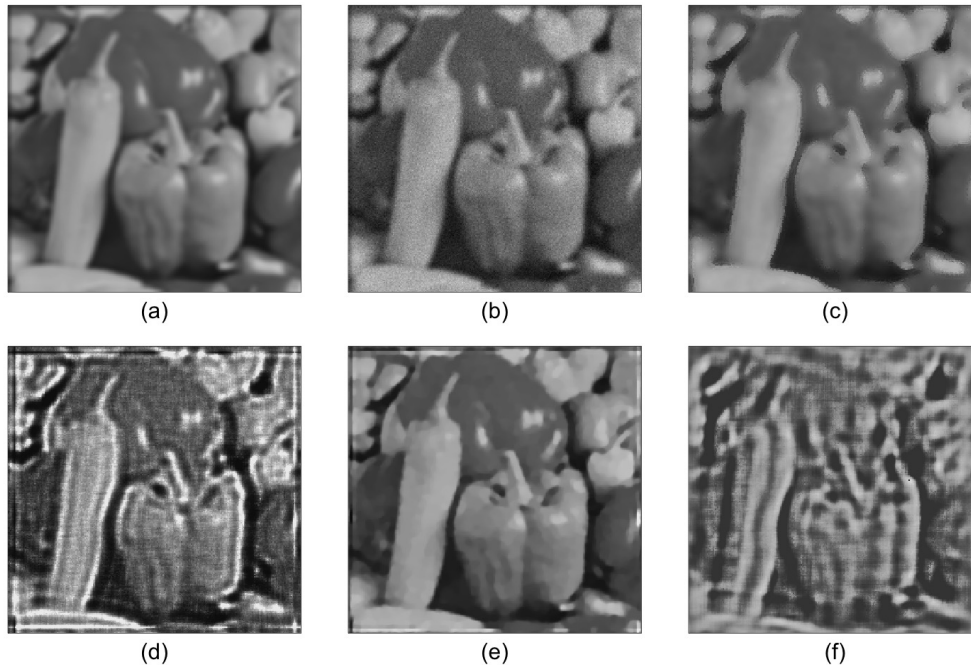


Figure 4: (a)–(b) Blurred peppers image and its blurred-and-noisy version in the case when the blur extent function is  $r_3(x, y)$  and  $\sigma = 10$ . The RMSE of (b) and (c) is 19.69 and 19.21, respectively. (c)–(f): Deblurred images by NEW, RL, TV and Bayes, respectively.

Bayes is chosen to be  $[84/217, 138/217] \times [22/217, 76/217]$ . It can be seen from Figure 5 that (i) NEW sharpens the image and removes the noise efficiently, (ii) both RL and Bayes generate many artifacts in their deblurred images around edges, and (iii) the deblurred image by TV does not seem to be improved much compared to the observed image.

### 3.4 Bandwidth Selection and Comparison with Wavelet Based Image Deblurring

Wavelet based image deblurring methods are well received in the literature. In this subsection, we compare the numerical performance of our BID method with the wavelet based method proposed by Beck and Teboulle (2009) (denoted as WAV) using a simulated example. In the simulation, the proposed bandwidth selection procedure is evaluated as well.

Table 2: Estimated values of RMSE of the four image deblurring methods in the Peppers image example based on 100 replicated simulations. The numbers in the parentheses are the standard errors of RMSE.

Methods	$r_1(x, y)$		$r_2(x, y)$		$r_3(x, y)$	
	$\sigma = 5$	$\sigma = 10$	$\sigma = 5$	$\sigma = 10$	$\sigma = 5$	$\sigma = 10$
New	19.69 (0.03)	20.66 (0.04)	19.12 (0.03)	19.43 (0.05)	18.00 (0.03)	19.18 (0.05)
RL	27.03 (0.06)	34.73 (0.12)	42.79 (1.46)	47.16 (0.69)	37.91 (0.10)	40.15 (0.18)
TV	19.49 (0.03)	19.52 (0.08)	45.52 (0.78)	46.27 (0.87)	17.39 (0.03)	17.44 (0.06)
Bayes	29.19 (6.75)	45.05 (3.89)	34.80 (9.00)	43.48 (7.07)	28.28 (10.00)	42.89 (7.68)

The true image intensity has the following expression (its image is shown in Figure 6(a)):

$$f(x, y) = \begin{cases} 3 & , \text{ if } (x - 0.25)^2 + (y - 0.75)^2 \leq 0.15^2 \text{ and } y \geq x. \\ 2 & , \text{ if } (x - 0.25)^2 + (y - 0.75)^2 > 0.15^2 \text{ and } y \geq x. \\ 1 & , \text{ if } (x - 0.75)^2 + (y - 0.25)^2 \leq 0.15^2 \text{ and } y < x. \\ 0 & , \text{ if } (x - 0.75)^2 + (y - 0.25)^2 > 0.15^2 \text{ and } y < x. \end{cases}$$

Throughout this subsection, we consider Gaussian blur with a location invariant blur extent  $r(x, y) = 0.02$ . The blurred image is shown in Figure 6(b). The comparison results are reported in Table 3, which includes the cases when  $n = 256, 512$  and  $\sigma = 0.015, 0.05, \text{ and } 0.1$ . Let  $k_0$  and  $\hat{k}_0$  denote the optimal bandwidth parameter that minimizes the MSE and the bandwidth selected by the proposed MCV procedure, respectively. Thus,  $|k_0 - \hat{k}_0|/n$  measures the performance of our bandwidth selection procedure. The values of MSE of NEW and WAV are shown for each combination of sample size  $n$  and noise level  $\sigma$ . The numbers in the parentheses are the standard error for the corresponding MSE. From Table 3, it can be seen that (i) NEW works stably and outperforms WAV, (ii) WAV works reasonably well when the noise level is low but its performance deteriorates rapidly as the noise level increases, and (iii) the MCV bandwidth selection procedure selects the bandwidth parameter close to  $k_0$ .

The first row in Figure 7 shows the observed images when the noise level is 0.015, 0.05, and 0.1. The second and third row shows the corresponding deblurred images by NEW

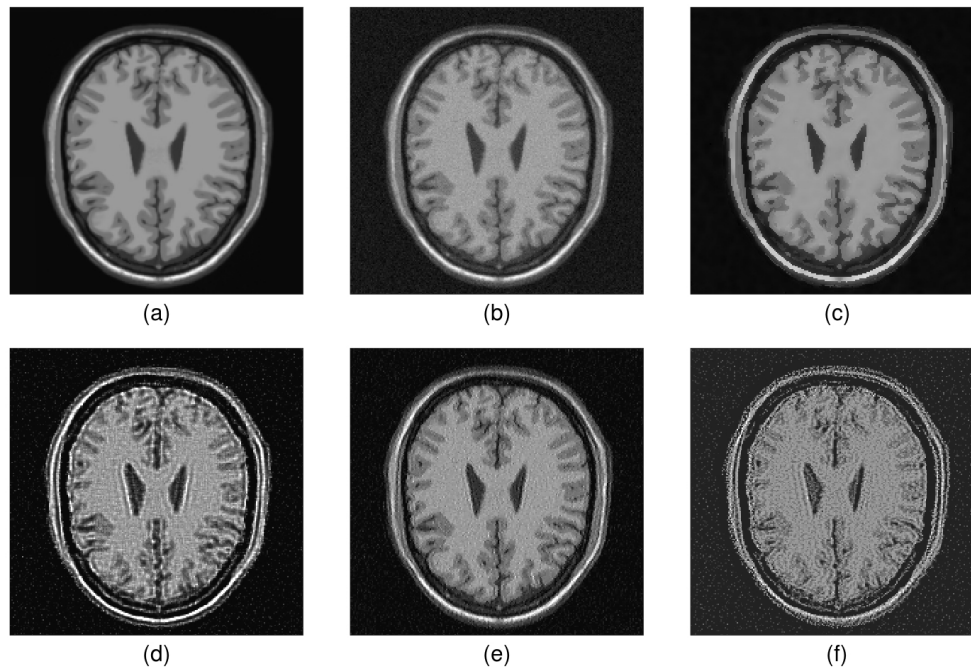


Figure 5: (a): A brain image with some blurring involved. (b): A noisy version of (a). (c)–(f): Deblurred images by NEW, RL, TV and Bayes, respectively.

and WAV, respectively. It can be seen that WAV does a decent job when the noise level is low but start to introduce artifacts as the observed image gets noisier. In comparison, NEW works well across different noise levels. This is consistent with the results in Table 3.

## 4 Concluding Remarks

We have proposed a blind image deblurring method which simultaneously removes spatial blur and pointwise noise from an observed image without imposing restrictive assumptions on the blurring mechanism. It even allows the psf to vary over location. This method is based on our observation that spatial blur alters the image structure significantly around step edges, but it does not change the image structure much in continuity regions of the image intensity surface. The challenging task of restoring complicated edge structures tapered by blur is accomplished by a local clustering procedure and by a weighted local smoothing. A data-driven bandwidth selection procedure is proposed along the BID method as well.

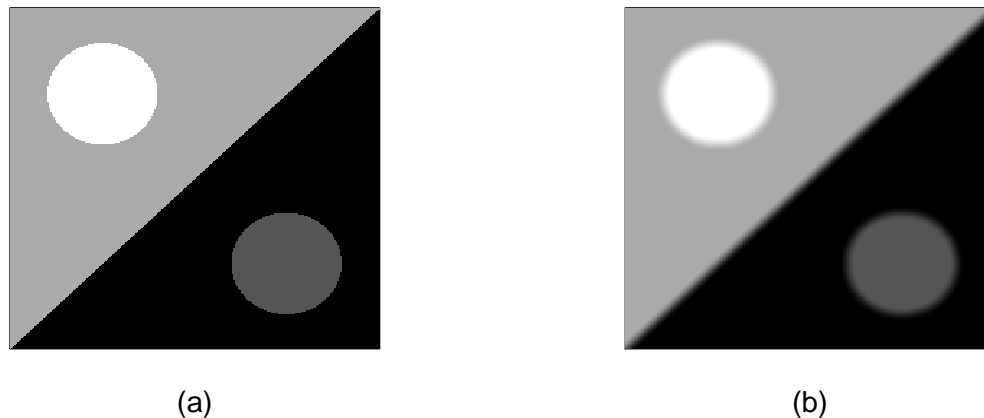


Figure 6: (a): The original image of the simulated example; (b): The blurred image by Gaussian blur with blur extent  $r(x, y) = 0.02$ .

Table 3: Comparison with WAV and numerical study of the proposed bandwidth selection procedure based on 100 replicated simulations.  $k_0$  and  $\hat{k}_0$  denote the bandwidth that minimizes the MSE and the bandwidth selected by MCV, respectively. The values of MSE are shown for each combination of  $n$  and  $\sigma$  and the numbers in the parentheses are the standard error for its corresponding MSE. All numbers except those under column  $\frac{k_0 - \hat{k}_0}{n}$  are in the unit of  $10^{-3}$ .

$n$	$\sigma = 0.015$			$\sigma = 0.05$			$\sigma = 0.1$		
	$\frac{ k_0 - \hat{k}_0 }{n}$	NEW	WAV	$\frac{ k_0 - \hat{k}_0 }{n}$	NEW	WAV	$\frac{ k_0 - \hat{k}_0 }{n}$	NEW	WAV
256	$\frac{3.00}{256}$	5.49 (0.08)	9.30 (0.16)	$\frac{1.99}{256}$	7.22 (0.25)	58.20 (1.20)	$\frac{1.68}{256}$	8.77 (0.48)	165.9 (4.80)
512	$\frac{3.00}{512}$	5.68 (0.06)	10.10 (0.08)	$\frac{5.90}{512}$	6.67 (0.14)	46.40 (0.68)	$\frac{2.59}{512}$	7.62 (0.22)	151.5 (2.60)

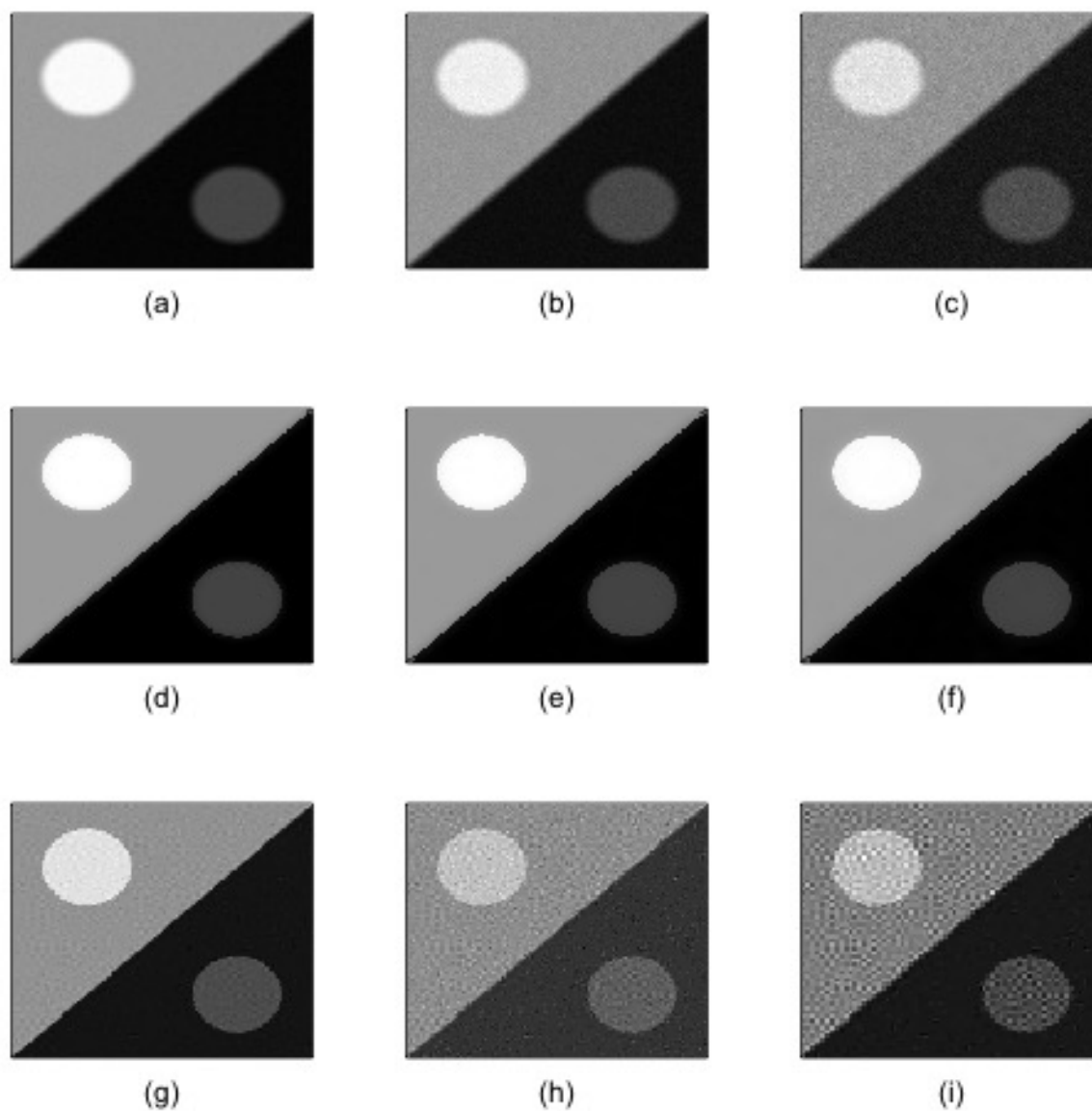


Figure 7: (a) – (c): Blurred-noisy images with noise level  $\sigma = 0.015$ ,  $0.05$  and  $0.1$ , respectively. (d) – (f): Deblurred image by NEW when the observed image is (a), (b) and (c), respectively. (g) – (i): Deblurred image by WAV when the observed image is (a), (b) and (c), respectively.

Numerical comparison with some state-of-the-art image deblurring methods shows that our proposed procedure can do a better job in removing a wide variety of different blur and in removing noise at different levels as well.

There is much room for improvement of our proposed method. First, this paper focuses on removing blur around step edges because those places dominate human visual perception. In other words, our proposed method only removes noise and does not attempt to deblur in the continuity regions. However, features in the continuity regions (e.g., roof/valley edges, peaks, etc.) ought to be restored even though they are less visually dominant. A natural improvement is to properly deblur the observed image to recover these features too. Second, we used a single bandwidth for local smoothing in the current method. The idea of multilevel smoothing that uses variable bandwidths can be incorporated into the proposed method. Third, the proposed bandwidth selection procedure works well in our simulation studies. Some theoretical justification for the asymptotic properties of the selected bandwidth would be another improvement of the current method. Finally, as seen in the numerical example with the Peppers image, our method carries out the deblurring procedure even in places where there is little blur involved and that could result in relatively large RMSE. Having the deblurring method adaptive to the blur extent would be an interesting theme for future research.

## References

- Aubert, G. and Kornprobst, P. (2006). *Mathematical problems in image processing: partial differential equations and the calculus of variations*, volume 147. Springer Science & Business Media.
- Bates, R. H. and McDonnell, M. J. (1989). *Image restoration and reconstruction*. Clarendon Press Oxford.
- Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm with application to wavelet-based image deblurring. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 693–696. IEEE.

- Biggs, D. S. and Andrews, M. (1997). Acceleration of iterative image restoration algorithms. *Applied optics*, 36(8):1766–1775.
- Carasso, A. S. (2001). Direct blind deconvolution. *SIAM Journal on Applied Mathematics*, 61(6):1980–2007.
- Carasso, A. S. (2003). The apex method in image sharpening and the use of low exponent lévy stable laws. *SIAM Journal on Applied Mathematics*, 63(2):593–618.
- Chan, T. F. and Wong, C.-K. (1998). Total variation blind deconvolution. *IEEE Transactions on Image Processing*, 7(3):370–375.
- Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability*, volume 66. CRC Press.
- Fergus, R., Singh, B., Hertzmann, A., Roweis, S. T., and Freeman, W. T. (2006). Removing camera shake from a single photograph. *ACM Transactions on Graphics (TOG)*, 25(3):787–794.
- Figueiredo, M. A. and Nowak, R. D. (2003). An em algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 12(8):906–916.
- Friedman, H. P. and Rubin, J. (1967). On some invariant criteria for grouping data. *Journal of the American Statistical Association*, 62(320):1159–1178.
- Gonzalez, R. C. and Woods, R. E. (2008). *Digital Image Processing*. Pearson Prentice Hall, 3 edition.
- Hall, P. and Qiu, P. (2007a). Blind deconvolution and deblurring in image analysis. *Statistica Sinica*, 17(4):1483.
- Hall, P. and Qiu, P. (2007b). Nonparametric estimation of a point-spread function in multivariate problems. *The Annals of Statistics*, 35(4):1512–1534.
- Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30.



- Jansson, P. (1997). *Deconvolution of spectra and images*. Academic Press Inc., New York.
- Joshi, M. V. and Chaudhuri, S. (2005). Joint blind restoration and surface recovery in photometric stereo. *Journal of Optical Society of America, Series A*, 22(6):1066–1076.
- Kang, Y. and Qiu, P. (2014). Jump detection in blurred regression surfaces. *Technometrics*, 56(4):539–550.
- Katsaggelos, A. K. and Lay, K.-T. (1990). Image identification and restoration based on the expectation-maximization algorithm. *Optical Engineering*, 29(5):436–445.
- Ko, S.-J. and Lee, Y. H. (1991). Center weighted median filters and their applications to image enhancement. *IEEE transactions on circuits and systems*, 38(9):984–993.
- Kundur, D. and Hatzinakos, D. (1998). A novel blind deconvolution scheme for image restoration using recursive filtering. *IEEE Transactions on Signal Processing*, 46(2):375–390.
- Miskin, J. and MacKay, D. J. (2000). Ensemble learning for blind image separation and deconvolution. In *Advances in independent component analysis*, pages 123–141. Springer.
- Oliveira, J. P., Bioucas-Dias, J. M., and Figueiredo, M. A. (2009). Adaptive total variation image deblurring: a majorization–minimization approach. *Signal Processing*, 89(9):1683–1693.
- Qiu, P. (1998). Discontinuous regression surfaces fitting. *The Annals of Statistics*, 26(6):2218–2245.
- Qiu, P. (2005). *Image processing and jump regression analysis*, volume 599. John Wiley & Sons.
- Qiu, P. (2008). A nonparametric procedure for blind image deblurring. *Computational Statistics & Data Analysis*, 52(10):4828–4841.
- Qiu, P. (2009). Jump-preserving surface reconstruction from noisy data. *Annals of the Institute of Statistical Mathematics*, 61(3):715–751.

Qiu, P. and Kang, Y. (2015). Blind image deblurring using jump regression analysis. *Statistica Sinica*, 25(3):879–899.

Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268.

Skilling, J. (1989). Classic maximum entropy. In *Maximum Entropy and Bayesian Methods*, pages 45–52. Springer.

Sun, T., Gabbouj, M., and Neuvo, Y. (1994). Center weighted median filters: some properties and their applications in image processing. *Signal processing*, 35(3):213–229.

Wand, M. P. and Jones, M. C. (1994). *Kernel smoothing*. CRC Press.

Yang, Y., Galatsanos, N. P., and Stark, H. (1994). Projection-based blind deconvolution. *Journal of Optical Society of America, Series A*, 11(9):2401–2409.

You, Y.-L. and Kaveh, M. (1996). A regularization approach to joint blur identification and image restoration. *IEEE Transactions on Image Processing*, 5(3):416–428.

## Appendix

### A Local Linear Kernel Estimates

By standard algebraic manipulations, the solution to (2) is as follows.

$$\hat{a}_n(x, y) = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}^{(1)}(x, y) Z_{ij}}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}^{(1)}(x, y)}, \quad (12)$$

$$\hat{b}_n(x, y) = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}^{(2)}(x, y) Z_{ij}}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}^{(2)}(x, y)}, \quad (13)$$

$$\hat{c}_n(x, y) = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}^{(3)}(x, y) Z_{ij}}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}^{(3)}(x, y)}, \quad (14)$$

where

$$\begin{aligned}
 w_{ij}^{(1)}(x, y) &= A_{11}(x, y) + A_{12}(x, y) \frac{i}{n} - \frac{x}{n} + A_{13}(x, y) \frac{j}{n} - \frac{y}{n} K \frac{i-x}{k}, \frac{j-y}{k}, \\
 w_{ij}^{(2)}(x, y) &= A_{21}(x, y) + A_{22}(x, y) \frac{i}{n} - \frac{x}{n} + A_{23}(x, y) \frac{j}{n} - \frac{y}{n} K \frac{i-x}{k}, \frac{j-y}{k}, \\
 w_{ij}^{(3)}(x, y) &= A_{31}(x, y) + A_{32}(x, y) \frac{i}{n} - \frac{x}{n} + A_{33}(x, y) \frac{j}{n} - \frac{y}{n} K \frac{i-x}{k}, \frac{j-y}{k}, \\
 A_{11}(x, y) &= r_{20}(x, y)r_{02}(x, y) - r_{11}(x, y)r_{11}(x, y), \\
 A_{12}(x, y) &= r_{01}(x, y)r_{11}(x, y) - r_{10}(x, y)r_{02}(x, y), \\
 A_{13}(x, y) &= r_{10}(x, y)r_{11}(x, y) - r_{01}(x, y)r_{20}(x, y), \\
 A_{21}(x, y) &= r_{01}(x, y)r_{11}(x, y) - r_{10}(x, y)r_{02}(x, y), \\
 A_{22}(x, y) &= r_{00}(x, y)r_{02}(x, y) - r_{01}(x, y)r_{01}(x, y), \\
 A_{23}(x, y) &= r_{01}(x, y)r_{10}(x, y) - r_{00}(x, y)r_{11}(x, y), \\
 A_{31}(x, y) &= r_{10}(x, y)r_{11}(x, y) - r_{20}(x, y)r_{01}(x, y), \\
 A_{32}(x, y) &= r_{01}(x, y)r_{10}(x, y) - r_{00}(x, y)r_{11}(x, y), \\
 A_{33}(x, y) &= r_{00}(x, y)r_{20}(x, y) - r_{10}(x, y)r_{10}(x, y), \\
 r_{s_1, s_2}(x, y) &= \sum_{i=1}^n \sum_{j=1}^n \frac{i}{n} - \frac{x}{n}^{s_1} \frac{j}{n} - \frac{y}{n}^{s_2} K \frac{i-x}{k}, \frac{j-y}{k}, \text{ for } s_1, s_2 = 0, 1, 2.
 \end{aligned}$$

## B Technical Details

In this section, we provide the proof of Proposition 1. First, let us introduce the following notations. Let  $\gamma(\cdot|x, y)$  denote the pdf of  $Z$  with respect to the  $(x, y)$ -th pixel, with corresponding cdf  $\Gamma(\cdot|x, y)$ . Let  $\xi_p(x_0, y_0)$  be the  $p$ -quantile of  $Z$  with respect to the  $(x_0, y_0)$ -th pixel. Since  $p = 1/2$  and  $(x_0, y_0)$  will remain fixed throughout our discussion, we shall write  $\xi_p(x_0, y_0) = \xi$ . Let

$$\begin{aligned}
 \gamma(\cdot | i + x_0, j + y_0) &= \gamma_{n, ij}(\cdot), \quad \Gamma(\cdot | i + x_0, j + y_0) = \Gamma_{n, ij}(\cdot) \\
 \bar{\gamma}_{nk}(\cdot) &= \sum_{i^2+j^2 \leq k^2} w_{nk, ij} \gamma_{n, ij}(\cdot), \quad \bar{\Gamma}_{nk}(\cdot) = \sum_{i^2+j^2 \leq k^2} w_{nk, ij} \Gamma_{n, ij}(\cdot),
 \end{aligned} \tag{15}$$

where  $w_{nk, ij}$  are positive numbers representing the weights in the weighted median filter and  $\sum_{i^2+j^2 \leq k^2} w_{nk, ij} = 1$ . The kernel empirical cdf of  $Z$  (with respect to  $(x_0, y_0)$ ) is defined

as

$$\widehat{\Gamma}_{nk}(z) = \sum_{i^2+j^2 \leq k^2} w_{nk,ij} \mathbf{1}(Z_{n,ij}^* \leq z), \quad (16)$$

where  $Z_{n,ij}^* = Z_{nx_0+i,ny_0+j}$  and  $\mathbf{1}(A)$  denotes the indicator of the event  $A$ . The weighted median filter output (i.e., the kernel estimator of  $\xi$ ) can be expressed as the  $p$ -quantile of  $\widehat{\Gamma}_{nk}$ , i.e.,

$$\widehat{\xi}_{nk} = \inf\{z : \widehat{\Gamma}_{nk}(z) \geq p\}. \quad (17)$$

Then  $\xi_{nk}$ , which is the target of  $\widehat{\xi}_{nk}$ , is given by

$$\bar{\Gamma}_{nk}(\xi_{nk}) = p = \Gamma(\xi), \quad (18)$$

where  $\Gamma(\xi) = \Gamma(\xi|x_0, y_0)$ . Also let  $\gamma(\xi) = \gamma(\xi|x_0, y_0)$ .

Next, the following regularity conditions are assumed.

- (B1)  $\gamma(\xi) > 0$  and  $\Gamma(\xi) = p$ .
- (B2) The partial derivatives  $\gamma_z(z|x, y)$ ,  $\gamma_{xx}(z|x, y)$ , and  $\gamma_{yy}(z|x, y)$  of  $\gamma(z|x)$  and  $\Gamma_{xx}(z|x, y)$  and  $\Gamma_{yy}(z|x, y)$  of  $\Gamma(z|x, y)$  exist in a neighborhood of  $(x_0, y_0, \xi)$ ,  $N(x_0, y_0, \xi)$ . And there exists  $M < \infty$  such that any  $(x, y, z) \in N(x_0, y_0, \xi)$ , we have
- $$\begin{aligned} |\gamma_z(z|x, y)| &\leq M, & |\gamma_x(z|x_0, y_0)| &\leq M, & |\gamma_y(z|x_0, y_0)| &\leq M, \\ |\gamma_{xx}(z|x_0, y_0)| &\leq M, & |\gamma_{yy}(z|x_0, y_0)| &\leq M, \\ |\gamma_{xx}(z|x, y) - \gamma_{xx}(z|x_0, y_0)| &\leq M(|x - x_0| + |y - y_0|), \\ |\gamma_{yy}(z|x, y) - \gamma_{yy}(z|x_0, y_0)| &\leq M(|x - x_0| + |y - y_0|). \end{aligned}$$
- (B3)  $f$  is piecewise continuous and has continuous second order derivatives in each closed set of the design space.  $\{\varepsilon_{ij}, i, j = 1, \dots, n\}$  are i.i.d. random variables with mean 0, median 0 and variance  $\sigma^2$ .
- (B4)  $K$  is a Lipschitz-1 continuous and radially symmetric bivariate density function on the unit disk.
- (B5) The bandwidth parameter  $k$  satisfies that  $c_1 \leq k/n^\alpha \leq c_2$ , where  $c_1$  and  $c_2$  are some positive constants and  $\alpha \in (0, 1/3)$ .

**Lemma 1.** *There exists a positive constant  $C$  such that*

$$|\xi_{nk} - \xi| \leq C \frac{k^2}{n^2}.$$

*Proof.* First, for  $\delta_n > 0$ , we have the following equivalent conditions:

$$|\xi_{nk} - \xi| \leq \delta_n \iff \bar{\Gamma}_{nk}(\xi - \delta_n) \leq \bar{\Gamma}(\xi_{nk}) = \Gamma(\xi) \quad \& \quad \bar{\Gamma}_{nk}(\xi + \delta_n) \geq \bar{\Gamma}(\xi_{nk}) = \Gamma(\xi)$$

Next, we do Taylor expansion on  $\bar{\Gamma}_{nk}(\xi + \delta_n)$ .

$$\begin{aligned} & \bar{\Gamma}_{nk}(\xi + \delta_n) \\ = & \Gamma(\xi + \delta_n) + \sum_{i^2+j^2 \leq k^2} w_{nk,ij} \frac{i}{n} \Gamma_x(\xi + \delta_n) + \frac{j}{n} \Gamma_y(\xi + \delta_n) + \frac{1}{2} \frac{i^2}{n^2} \Gamma_{xx}(\xi + \delta_n) \\ & + \frac{1}{2} \frac{j^2}{n^2} \Gamma_{yy}(\xi + \delta_n) + \frac{ij}{n^2} \Gamma_{xy}(\xi + \delta_n) + O \frac{k^3}{n^3} \\ = & \Gamma(\xi + \delta_n) + O \frac{k^2}{n^2} \\ \geq & \Gamma(\xi) + \frac{1}{2} \gamma(\xi) \delta_n + O \frac{k^2}{n^2} \end{aligned}$$

Therefore, there exists a positive constant  $C$  such that

$$\delta_n = C \frac{k^2}{n^2}, \quad \frac{1}{2} \gamma(\xi) \delta_n + O \frac{k^2}{n^2} \geq 0.$$

Similarly, we can show that

$$\bar{\Gamma}_{nk}(\xi - \delta_n) \leq \Gamma(\xi).$$

And the proof is completed. □

**Lemma 2.**

$$P \left| \hat{\xi}_{nk} - \xi_{nk} \right| > C k^{-1} \sqrt{\log n \log \log n}, i.o. = 0,$$

where  $C$  is some positive constant.

*Proof.* Let  $a_n = k^{-1} \sqrt{\log n \log \log n}$ . If  $\hat{\xi}_{nk} - \xi_{nk} < -a_n$ , then

$$\hat{\Gamma}_{nk}(\xi_{nk} - a_n) \geq p.$$

Then,

$$\begin{aligned}
& \sum_{i^2+j^2 \leq k^2} w_{nk,ij} [\mathbf{1}(Z_{ij} \leq \xi_{nk} - a_n) - \Gamma_{n,ij}(\xi_{nk} - a_n)] \\
& \geq p - \bar{\Gamma}_{nk}(\xi_{nk} - a_n) \\
& = \bar{\Gamma}_{nk}(\xi_{nk}) - \bar{\Gamma}_{nk}(\xi_{nk} - a_n) \\
& = \sum_{i^2+j^2 \leq k^2} w_{nk,ij} [\Gamma(\xi_{nk}) - \Gamma(\xi_{nk} - a_n)] + O \frac{k^2}{n^2} \\
& = \gamma(\xi)a_n + O a_n^2 + O \frac{k^2}{n^2} \\
& \geq \frac{1}{2}\gamma(\xi)a_n \quad \text{when } n \text{ is large enough.}
\end{aligned}$$

Lemma 1 was used in the second last line. By Theorem 2 in Hoeffding (1963), we have

$$\begin{aligned}
& P \left( \frac{1}{k^2} \sum_{i^2+j^2 \leq k^2} k^2 w_{nk,ij} [\mathbf{1}(Z_{ij} \leq \xi_{nk} - a_n) - \Gamma_{n,ij}(\xi_{nk} - a_n)] \geq \frac{1}{2}\gamma(\xi)a_n \right) \\
& \leq e^{-Ck^2 a_n^2} \\
& = n^{-C \log \log n},
\end{aligned}$$

where  $C$  is some positive constant. And this completes the proof.  $\square$

### Lemma 3.

$$P \quad \hat{\Gamma}_{nk}(\hat{\xi}_{nk}) - \hat{\Gamma}(\xi_{nk}) - \bar{\Gamma}_{nk}(\hat{\xi}_{nk}) - \bar{\Gamma}_{nk}(\xi_{nk}) > Ck^{-(1+\delta)}(\log n)^{\frac{3}{4}}(\log \log n)^{\frac{1}{4}}, i.o. = 0,$$

where  $\delta \in (0, 1/2]$  and  $C$  is some positive constant.

*Proof.* Let

$$\begin{aligned}
H_{nk}(z) &= \hat{\Gamma}_{nk}(\hat{\xi}_{nk}) - \hat{\Gamma}(\xi_{nk}) - \bar{\Gamma}_{nk}(\hat{\xi}_{nk}) - \bar{\Gamma}_{nk}(\xi_{nk}) \\
&= \sum_{i^2+j^2 \leq k^2} w_{nk,ij} [(\mathbf{1}(Z_{ij} \leq z) - \mathbf{1}(Z_{ij} \leq \xi_{nk})) - (\Gamma_{n,ij}(z) - \Gamma_{n,ij}(\xi_{nk}))] \\
&= \sum_{i^2+j^2 \leq k^2} w_{nk,ij} (U_{nk,ij} - \mu_{nk,ij}),
\end{aligned}$$

where  $U_{nk,ij} = \mathbb{1}(Z_{ij} \leq z) - \mathbb{1}(Z_{ij} \leq \xi_{nk})$  and  $\mu_{nk,ij} = \Gamma_{n,ij}(z) - \Gamma_{n,ij}(\xi_{nk})$ . Next, we have

$$\begin{aligned} \mu_{nk,ij} &= \gamma_{n,ij}(\theta_{nk,ij})(z - \xi_{nk}) \\ &= \gamma'_{nk,ij}(\theta_{nk,ij}^*)(\theta_{nk,ij} - \xi) + \gamma_{n,ij}(\xi) - \gamma(\xi) + \gamma(\xi) (z - \xi_{nk}) \\ &= \gamma(\xi)(z - \xi_{nk}) + (z - \xi_{nk}) \cdot O(|z - \xi_{nk}| + |\xi_{nk} - \xi|) + \frac{k}{n} \end{aligned}$$

Choosing  $z = \widehat{\xi}_{nk}$ , by Lemma 2,

$$|\mu_{nk,ij}| = O(k^{-1} \sqrt{\log n \log \log n}).$$

For  $M_{nk} > 0$ , by Bernstein Inequality, we have

$$\begin{aligned} P(|H_{nk}(\widehat{\xi}_{nk})| > M_{nk}) &= P\left(\sum_{i^2+j^2 \leq k^2} w_{nk,ij}(U_{nk,ij} - \mu_{nk,ij}) > M_{nk}\right) \\ &\leq 2 \exp\left[-\frac{\frac{1}{2}M_{nk}^2}{\sum_{i^2+j^2 \leq k^2} w_{nk,ij}^2 |\mu_{nk,ij}| (1 - |\mu_{nk,ij}|) + C \frac{1}{k^2} M_{nk}}\right] \\ &\leq 2 \exp\left[-\frac{\frac{1}{2}M_{nk}^2}{k^{-1} \sqrt{\log n \log \log n} \cdot k^{-2} C_1 + C_2 k^{-2} M_{nk}}\right], \end{aligned}$$

where  $C$ ,  $C_1$  and  $C_2$  are positive constants. Let  $M_{nk} = k^{-(1+\delta)}(\log n)^{3/4}(\log \log n)^{1/4}M$ , where  $\delta \in (0, 1/2]$  and  $M$  is some positive constant to be determined later. Then,

$$\frac{M_{nk}}{k^{-1} \sqrt{\log n \log \log n}} \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Hence,

$$\begin{aligned} P(|H_{nk}(\widehat{\xi}_{nk})| > M_{nk}) &\leq 2 \exp\left[-\frac{1}{2} \cdot \frac{\frac{1}{2}M_{nk}^2}{k^{-3} \sqrt{\log n \log \log n} C_1}\right] \\ &= 2 \exp\{-C^* M^2 k^{1-2\delta} \log n\} \\ &\leq 2 \exp\{-2 \log n\} = 2n^{-2}, \end{aligned}$$

where  $M$  is chosen such that  $M \cdot C^* > 2$  in the case when  $\delta = 1/2$ . Then Lemma 3 follows immediately.  $\square$

## ***Proof of Proposition 1.***

$$\begin{aligned}
& p - \widehat{\Gamma}_{nk}(\xi_{nk}) \\
&= H_{nk} \widehat{\xi}_{nk} + \bar{\Gamma}_{nk} \widehat{\xi}_{nk} - \bar{\Gamma}_{nk}(\xi_{nk}) \\
&= H_{nk} \widehat{\xi}_{nk} + \widehat{\xi}_{nk} - \xi_{nk} \sum_{i^2+j^2 \leq k^2} \gamma_{nk,ij}(z_{nk,ij}) w_{nk,ij} \\
&= H_{nk} \widehat{\xi}_{nk} + \widehat{\xi}_{nk} - \xi_{nk} \sum_{i^2+j^2 \leq k^2} (\gamma_{nk,ij}(z_{nk,ij}) - \gamma_{nk,ij}(\xi) + \gamma_{nk,ij}(\xi) - \gamma(\xi) + \gamma(\xi)) w_{nk,ij} \\
&= k^{-(1+\delta)} (\log n)^{\frac{3}{4}} (\log \log n)^{\frac{1}{4}} + \gamma(\xi) \widehat{\xi}_{nk} - \xi_{nk} \\
&\quad + \widehat{\xi}_{nk} - \xi_{nk} \sum_{i^2+j^2 \leq k^2} O \left( k^{-1} \frac{1}{\log n \log \log n} + \frac{k^2}{n^2} \right) + \gamma_{nk,ij}(\xi) - \gamma(\xi) w_{nk,ij} \\
&= k^{-(1+\delta)} (\log n)^{\frac{3}{4}} (\log \log n)^{\frac{1}{4}} + \gamma(\xi) \widehat{\xi}_{nk} - \xi_{nk} \\
&\quad + O \left( k^{-1} \frac{1}{\log n \log \log n} \right) k^{-1} \frac{1}{\log n \log \log n} + \frac{k^2}{n^2} + O \left( k^{-1} \frac{1}{\log n \log \log n} \right) \frac{k^2}{n^2}.
\end{aligned}$$

Then,

$$\widehat{\xi}_{nk} - \xi_{nk} = O \left( k^{-(1+\delta)} (\log n)^{\frac{3}{4}} (\log \log n)^{\frac{1}{4}} \right) + O \left( \frac{k}{n^2} \frac{1}{\log n \log \log n} \right) + p - \widehat{\Gamma}_{nk}(\xi_{nk}) \frac{1}{\gamma(\xi)}.$$

By Lemma 1, we have

$$\begin{aligned}
\widehat{\xi}_{nk} - \xi &= O \left( \frac{k^2}{n^2} \right) + O \left( k^{-(1+\delta)} (\log n)^{\frac{3}{4}} (\log \log n)^{\frac{1}{4}} \right) + O \left( \frac{k}{n^2} \frac{1}{\log n \log \log n} \right) \\
&\quad + \frac{1}{\gamma(\xi)} \sum_{i^2+j^2 \leq k^2} w_{nk,ij} (p - \mathbf{1}(Z_{ij} \leq \xi_{nk})). \tag{19}
\end{aligned}$$

Next, let  $Z_{ij}^* = \Gamma^{-1} \circ \Gamma_{n,ij}(Z_{ij})$ . Then,

$$\begin{aligned}
& \sum_{i^2+j^2 \leq k^2} w_{nk,ij} (p - \mathbf{1}(Z_{ij} \leq \xi_{nk})) - \sum_{i^2+j^2 \leq k^2} w_{nk,ij} (\Gamma(\xi) - \mathbf{1}(Z_{ij}^* \leq \xi)) \\
&= \sum_{i^2+j^2 \leq k^2} w_{nk,ij} (\Gamma_{n,ij}(\xi_{nk}) - \mathbf{1}(Z_{ij} \leq \xi_{nk})) - \sum_{i^2+j^2 \leq k^2} w_{nk,ij} (\Gamma(\xi) - \mathbf{1}(Z_{ij}^* \leq \xi)) \\
&= \sum_{i^2+j^2 \leq k^2} w_{nk,ij} [(\mathbf{1}(Z_{ij}^* \leq \xi) - \mathbf{1}(Z_{ij} \leq \xi_{nk})) - (\Gamma(\xi) - \Gamma_{n,ij}(\xi_{nk}))].
\end{aligned}$$

Since  $|\Gamma(\xi) - \Gamma_{n,ij}(\xi_{nk})| \leq |\Gamma(\xi) - \Gamma(\xi_{nk})| + |\Gamma(\xi_{nk}) - \Gamma_{n,ij}(\xi_{nk})| = O(k^2/n^2 + k/n)$ , we have



(by Bernstein inequality), for  $A_{nk} > 0$ ,

$$\begin{aligned}
 & P \left( \sum_{i^2+j^2 \leq k^2} w_{nk,ij} \mathbf{1}(Z_{ij}^* \leq \xi) - \mathbf{1}(Z_{ij} \leq \xi_{nk}) - (\Gamma(\xi) - \Gamma_{n,ij}(\xi_{nk})) > A_{nk} \right) \\
 & \leq 2 \exp - \frac{\frac{1}{2} A_{nk}^2}{\sum_{i^2+j^2 \leq k^2} w_{nk,ij}^2 C_1 \frac{k}{n} + C_2 k^{-2} A_{nk}} \\
 & = 2 \exp - \frac{\frac{1}{2} A_{nk}^2}{C_1 k^{-2} \frac{k}{n} + C_2 k^{-2} A_{nk}} . \tag{20}
 \end{aligned}$$

Choose  $A_{nk} = (k/n)^{1-\eta} (\log n)^{1/2} M$ , where  $\eta = (1 - 3\alpha)/(2 - 2\alpha)$  and  $M$  is some positive constant to be determined later. Then,

$$\begin{aligned}
 (20) & \leq 2 \exp - \frac{1}{4} \frac{k}{n} \frac{2^{-2\eta} (\log n) M^2}{C_1 k^{-1} n^{-1}} \\
 & = 2 \exp - C^* M^2 \log n ,
 \end{aligned}$$

where  $C^*$  is a positive constant and  $M$  is chosen such that  $C^* M^2 \geq 2$ . Then,

$$\begin{aligned}
 \widehat{\xi}_{nk} - \xi & = O \frac{k^2}{n^2} + O k^{-(1+\delta)} (\log n)^{\frac{3}{4}} (\log \log n)^{\frac{1}{4}} + O \frac{k}{n}^{1-\eta} (\log n)^{\frac{1}{2}} \\
 & \quad + \frac{1}{\gamma(\xi)} \sum_{i^2+j^2 \leq k^2} w_{nk,ij} (\Gamma(\xi) - \mathbf{1}(Z_{ij}^* \leq \xi)) . \\
 k \widehat{\xi}_{nk} - \xi & = O \frac{k^3}{n^2} + O k^{-\delta} (\log n)^{\frac{3}{4}} (\log \log n)^{\frac{1}{4}} + O \frac{k^{2-\eta}}{n^{1-\eta}} (\log n)^{\frac{1}{2}} \\
 & \quad + \frac{k}{\gamma(\xi)} \sum_{i^2+j^2 \leq k^2} w_{nk,ij} (\Gamma(\xi) - \mathbf{1}(Z_{ij}^* \leq \xi)) .
 \end{aligned}$$

Note that  $k^{2-\eta}/n^{1-\eta} = n^{-(1-\alpha)^2}$ . So,

$$k \widehat{\xi}_{nk} - \xi = \frac{k}{\gamma(\xi)} \sum_{i^2+j^2 \leq k^2} w_{nk,ij} [\mathbf{1}(\varepsilon_{ij} \leq 0) - (1-p)] + o_p(1). \tag{21}$$

Thus, the joint asymptotic normality in Proposition 1 follows from (21) and (12).  $\square$