Statistica Sinica **17**(2007), 1483-1509

BLIND DECONVOLUTION AND DEBLURRING IN IMAGE ANALYSIS

Peter Hall and Peihua Qiu

The University of Melbourne and University of Minnesota

Abstract: Blind deconvolution problems arise in image analysis when both the extent of image blur, and the true image, are unknown. If a model is available for at least one of these quantities then, in theory, the problem is solvable. It is generally not solvable if neither the image nor the point-spread function, which controls the extent of blur, is known parametrically. In this paper we develop methods for solution when a model is known for the point-spread function, but the image is assessed only nonparametrically. We assume that the image includes sharp edges — mathematically speaking, lines of discontinuity of image brightness. However, the locations, shapes and other properties of these lines are not needed for our algorithm. Our technique involves mathematically "focussing" the restored image until the edges are as sharp as possible, with sharpness being measured using a difference-based approach. We pay special attention to the Gaussian point-spread function. Although this context is notorious in statistical deconvolution problems on account of the difficulty of finding a usable solution, it is arguably the central, and the most important, setting for restoring blurred images. Numerical simulation, application to an image, and theoretical analysis demonstrate the effectiveness of our approach.

Key words and phrases: Adaptive estimation, blur, deblurring, Gaussian blur, ill-posed problem, image restoration, inverse problem, noise, nonparametric regression, point-spread function, regularisation, test pattern.

1. Introduction

Photographic images, whether recorded by digital or analogue means, have imperfections which prevent them from conveying the "true" scene. These degradations have a variety of causes but two types, blurring and noise, are especially common. The removal of blur, in the presence of noise, is generally an ill-posed deconvolution problem, the solution of which requires inversion of the blur operator followed by a smoothing step.

Although challenging, this type of problem is quite well understood if the extent of blur can be described in precise mathematical terms. However, there is a rapidly increasing interest in problems where the mathematical operation of blurring is known only approximately, for example in terms of a function

which depends on unknown parameters that have to be computed from image data. This is a blind deconvolution problem and is, of course, significantly more challenging than its more conventional, non-blind counterpart. See, for example, work of Kundur and Hatzinakos (1998), Carasso (2001), Galatsanos, Mesarović, Molina, Katsaggelos and Mateos (2002) and Figueiredo and Nowak (2003).

In this paper we suggest a new technique for double-blind deconvolution when the point-spread function, describing the manner in which blur degrades the true image, is available only up to unknown parameters. Our approach is of interest because it is closely linked to the physically meaningful notion of adjusting the parameters until the image is sharpest. In contrast to some earlier methods, ours does not require detailed information about the "true" image. Instead, we recognise that the sort of image that would be used to determine a point-spread function would usually be one that has relatively sharp boundaries, such as those in a photographic test pattern. We introduce a difference-based approximation, D say, to the derivative of such an image, and observe that in the absence of error, D is mathematically virtually identical to the integral, along all boundaries, of the squared height of jump discontinuities in image intensity. We suggest gradually altering the inverse of a candidate point-spread function, by steadily increasing its spread, until a place is reached where D increases sharply from a low to a high value; and then taking the corresponding version of the function to be the correct one.

An algorithm of this type can be implemented using a threshold argument, as follows. Gradually increase the spread of the point-spread function until D exceeds the threshold, and then stop. In operation, this technique is rather like focusing a lens and stopping when the image is at its sharpest point, except that the extent to which the image is blurred is determined not by eye but by the mathematical criterion D. Moreover, in the present problem, if the degree of spread of the candidate point-spread function is increased beyond the point where it is exactly correct, the deblurred image becomes extremely erratic, and in fact the value of D would be theoretically infinite if we were to go beyond the critical point-spread parameter, were it not for the influence of noise and discretisation error.

The ideal choice of threshold is the value of the integral of the square of the jump in intensity along edges, mentioned two paragraphs above. In some instances, the value of this integral may be known within a range. We might take the threshold to be toward the upper end of this range, although in theory, statistically consistent estimators can be obtained for any value in the range; see the theorem in Section 3. More generally, experimentation with different images that have approximately the same sort of discontinuities and noise levels

as the one under investigation, and where the point-spread function is similar but known, can assist in threshold choice.

Thus our technique avoids the need for information about the true image, by using a mathematical description of the extent to which the image is too "soft" through the effect of blur. However, some parametric forms of point spread functions differ very little, from one parameter value to another, in terms of the effects they have on the extent to which an image is blurred. A technique based on assessing the sharpness of a test pattern cannot be expected to perform well in such circumstances. This issue is the subject of particular discussion in Section 2, which gives general theoretical background for our technique. For simplicity, we do not consider the effects of noise there; this matter is taken up in mathematical detail in Section 3. A summary of numerical properties of our method is given in Section 4.

We focus on the problem of estimating the point-spread function, rather than the obviously related one of of image restoration, since there are a great many approaches to solving the latter problem, and our objectives would have to narrow if we were to treat just one of them. Moreover, there is usually intrinsic interest in the point-spread function, not the least because knowing the nature of that function is an important step in determining the blurring mechanism, so as to improve performance of the imaging device.

There is a substantial, recent statistical literature on deconvolution problems, including four discussion papers (Cornford, Csató, Evans and Opper (2004), Johnstone, Kerkyacharian, Picard and Raimondo (2004), Haario, Laine, Lehtinen, Saksman and Tamminen (2004) and Wolfe, Godsill and Ng (2004)) in a single issue of the *Journal of the Royal Statistical Society*. Recent statistical, or closely related, work on blind deconvolution includes that of Gassiat and Gautherat (1999), Poskitt, Dougancay and Chung (1999), Higdon and Yamamoto (2001), Li and Shedden (2001), Zhang, Amari and Cichocki (2001), Doucet, Godsill and Robert (2002), Ellis (2002), May, Stathaki and Katsaggelos (2003), Rosec, Boucher, Nsiri and Chonavel (2003), Sotthivirat and Fessler (2003), Carasso (2004), Erdogmus, Hild, Principe, Lazaro and Santamaria (2004), and Likas and Galatsanos (2004).

2. Models and Methodology

2.1. Models for point-spread function

Denote by $X(x, y)$ the brightness of an image at a point (x, y) in the plane $\mathbb{R}^2$. We consider the function X to be "blurred" by a point-spread function, ϕ^θ, where θ denotes the value of an unknown (possibly vector valued) parameter

on which the point-spread function depends. The result is the observed, blurred image

$$Y^\theta(x,y) = \iint \phi^\theta(u,v) X(x-u, y-v)\, du\, dv\,, \quad (x,y) \in \mathbb{R}^2. \tag{2.1}$$

(Here and below, unqualified integrals are taken over the whole real line or the entire plane.) Importantly, in none of the work in this paper do we assume that X is known. This is especially relevant when the image represented by X is degraded by stochastic noise as well as by systematic blur; see Section 3 for discussion of the case of noise.

We may view Y^θ as the result $\phi^\theta X$, say, of operating on X by the linear transform with kernel ϕ^θ. (For economy of notation we use the same notation for both a point-spread function and the associated linear operator.) For the most part we assume θ is univariate. We suppose too that ϕ^θ preserves average brightness, in the sense that for each θ,

$$\iint \phi^\theta(x,y)\, dx\, dy = 1. \tag{2.2}$$

Examples include the Gaussian point-spread function

$$\phi^\theta_{\rm Gau}(x,y) = \frac{1}{2\pi\theta} \exp\Big\{ - \frac{\frac{1}{2}(x^2+y^2)}{\theta} \Big\}\,, \quad (x,y) \in \mathbb{R}^2. \tag{2.3}$$

Although, as statisticians know from experience in errors-in-variables problems, this type of blur is especially challenging for deconvolution, it is especially common in applications of image analysis. One of the reasons is the central, canonical role played by the Gaussian density function, which comes about in part through arguments based on the Central Limit Theorem.

The circular-exponential point-spread function,

$$\phi^\theta_{\rm exp}(x,y) = \frac{1}{2\pi\,\theta^2} \exp\Big\{ - \theta^{-1}(x^2+y^2)^{\frac{1}{2}} \Big\}\,, \quad (x,y) \in \mathbb{R}^2, \tag{2.4}$$

where $\theta > 0$, is occasionally discussed in image analysis. One generalisation of $\phi^\theta_{\rm Gau}$ is to the circular, "stable" point-spread function,

$$\phi^\theta_{\rm sta}(x,y) = \theta_2^{-2} \psi^{\theta_1}_{\rm sta}\Big\{ \frac{(x,y)}{\theta_2} \Big\}\,, \quad (x,y) \in \mathbb{R}^2, \tag{2.5}$$

where $\theta = (\theta_1, \theta_2)$, $0 < \theta_1 \le 2$, $\theta_2 > 0$, and $\psi^{\theta_1}_{\rm sta}$ denotes the density of a bivariate, symmetric, stable law with characteristic function

$$\begin{aligned} &\iint \exp\{-(isx+ity)\}\, \psi^{\theta_1}_{\rm sta}(x,y)\, dx\, dy \\ &\quad = \exp\Big\{ - (s^2+t^2)^{\frac{\theta_1}{2}} \Big\},\ (s,t) \in \mathbb{R}^2. \end{aligned}$$

The Gaussian case is obtained by taking, in (2.5), $\theta_1 = 2$ and θ_2 proportional to the standard deviation.

2.2. Models for images

The type of image that is most likely to be useful for determining a point-spread function is one that includes sharp boundaries. That is, if we represent an image as a function, X, from the plane to the real line, then the function should include jump discontinuities. Such functions can be thought of as modelling the intensity of light reflected from a typical test-pattern used in photography. Rows of lines in the test-pattern are represented mathematically as long, narrow, closely-spaced rows of rectangular blocks, on each of which the function takes a particular value, say 1, and off which it takes a different value, say 0.

An example to which we pay special attention, because of its mathematical simplicity, is the square-block test-pattern, where the image is represented as the function

$$X(x,y) = \begin{cases} 1 & \text{if } |x| \le 1 \text{ and } |y| \le 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.6}$$

Note, however, that we do not need to know the true image; that is, the function X is not assumed known. Our approach supposes only that X is a function with fault-type discontinuities; we do not require information about their location, height or extent.

2.3. Effect of deblurring on discontinuities

If a function X of the type discussed in Section 2.2 is smoothed by the operator ϕ^{θ_0}, producing the blurred function Y^{θ_0} defined at (2.1), then vertical fault-type discontinuities in X will generally be converted to sloping surfaces in $Y^{\theta_0} = \phi^{\theta_0} X$. Conversely, if we attempt to recover the image represented by X by applying the inverse of ϕ^θ to Y^{θ_0}, then in many cases we should be able to determine the correct value of θ (that is, $\theta = \theta_0$) by considering which values of θ produce a function

$$Z^\theta = \bigl(\phi^\theta\bigr)^{-1} Y^{\theta_0} = \bigl(\phi^\theta\bigr)^{-1} \phi^{\theta_0} X \tag{2.7}$$

that has had fault-type discontinuities restored to it.

Since we do not assume that the true image represented by X is known, then we do not know where to look for the fault lines, or how to assess how high the jump discontinuities might be. Thus, rather than make a subjective assessment of image smoothness we require a mathematical criterion, based on differentiation or differencing. Such an approach is discussed in the next section.

To appreciate in more detail the effect that the choice of θ has on properties of Z^θ, defined at (2.7), assume that θ is a scalar parameter and that increasing θ increases the extent of "spread" of ϕ^θ. Examples include the Gaussian and circular-exponential point-spread functions, at (2.3) and (2.4), respectively, and also the circular-stable point-spread function at (2.5) if θ_1 is fixed and $\theta = \theta_2$ is varied, or if θ_2 is fixed and $\theta = \theta_1$ is altered. In such cases, if the univariate parameter θ is too small then applying the inverse operator, $(\phi^\theta)^{-1}$, to $Y^{\theta_0} = \phi^{\theta_0} X$, as at (2.7), will not fully compensate for the blurring action that took X to Y^{θ_0}. However, increasing θ all the way to $\theta = \theta_0$ will cause the image to "snap into focus," and allow us to recover X perfectly. On the other hand, increasing θ beyond θ_0 will cause Z^θ to fluctuate in a highly erratic manner, especially if some noise is present. By assessing these situations mathematically, we have an opportunity to determine the correct value of θ.

Let $\phi^\theta_{\rm ft}$ denote the Fourier transform of ϕ^θ:

$$\phi^\theta_{\rm ft}(s,t) = \frac{1}{(2\pi)^2} \iint \phi^\theta(x,y) \exp\{-(isx + ity)\}\, dx\, dy. \tag{2.8}$$

Mathematically, the point-spread functions considered in the previous paragraph, with θ univariate and with the exception of the circular-exponential function, all have the following properties:

$$\begin{aligned} &\text{if } 0 < \theta_1 < \theta_2 \text{ then for each } C > 0, \left|\phi^{\theta_2}_{\rm ft}(s,t)/\phi^{\theta_1}_{\rm ft}(s,t)\right| \text{ is bounded above} \\ &\text{by a constant multiple of } (1 + |s| + |t|)^{-C}, \text{ uniformly in } (s,t) \in \mathbb{R}^2. \end{aligned} \tag{2.9}$$

Since C in (2.9) can be chosen arbitrarily large, then the implications of (2.9) may be expressed in words as follows: If $0 < \theta_1 < \theta_2$ then the effect that ϕ^{θ_2} has on dampening down high-frequency portions of the function X is exponentially greater, as a function of those frequencies, than the effect had by ϕ^{θ_1}. Here we interpret "exponentially greater" as meaning "greater than any polynomial function," although in the examples in the previous paragraph it actually is a conventional exponential function.

2.4. Measuring the sharpness of a blurred image

If the extent of deblurring is too small, then the deblurred image will be relatively "soft." In particular, if ϕ^θ satisfies (2.9) and θ is strictly less than the correct value θ_0 then, even when X contains fault-type discontinuities, Z^θ (representing our attempt at deblurring the blurred version, Y^{θ_0}, of X) will be differentiable. On the other hand, provided X contains fault-type discontinuities, Z^θ will fail to be differentiable if we deblur by exactly the right amount, or if we deblur too much. These considerations suggest that we might measure

the sharpness of the deblurred image by assessing the extent to which Z^θ is differentiable.

Of course, in practice we have to do all our calculations on a discrete grid, representing the pixel grid on which images are digitised. Therefore, we are interested in difference operators as much as we are in differentiation. The methodology discussed below is strongly influenced by this consideration.

Let $h > 0$ denote a small positive quantity, let $Z = Z^\theta$ represent our attempt at restoring the function Y^{θ_0}, obtained by smoothing X, and define

$$A(h) = (2h)^{-1} \iint \max_\omega \Big\{ Z(x + h\cos\omega, y + h\sin\omega) - Z(x - h\cos\omega, y - h\sin\omega) \Big\}^2 dx\, dy, \tag{2.10}$$

$$A_1(h) = (2h)^{-1} \iint \Big[\{Z(x+h, y) - Z(x-h, y)\}^2 + \{Z(x, y+h) - Z(x, y-h)\}^2 \Big] dx\, dy. \tag{2.11}$$

Clearly,

$$A_1(h) \le 2A(h), \tag{2.12}$$

and so if discontinuities in Z were evident from the criterion $A_1(h)$, they would also be apparent from $A(h)$.

We show in Appendix A.1 that, if Z is smooth except for fault-type discontinuities along a collection, $\mathcal{C}$, of curves in the plane, and if $f(s)$ denotes the jump height at $s \in \mathcal{C}$, then as $h \to 0$,

$$A(h) \to \int_{\mathcal{C}} f(s)^2 ds, \tag{2.13}$$

the latter denoting the line integral of f^2 along $\mathcal{C}$. Of course, $\mathcal{C}$ does not need to be a single connected curve, and need not be smooth; it may contain corners.

The limit for $A_1(h)$ is less explicit than that for $A(h)$, but always lies between 1 and $\sqrt{2}$ times the limit for $A(h)$: as $h \to 0$,

$$A_1(h) \to \alpha \int_{\mathcal{C}} f(s)^2 ds, \tag{2.14}$$

where $1 \le \alpha \le \sqrt{2}$. In general, α will depend on $\mathcal{C}$ and f in a complex manner. To obtain (2.13) and (2.14) in Appendix A.1 we assume that, at points away from $\mathcal{C}$, the function X is smooth, in the sense of having a bounded derivative.

Results (2.13) and (2.14) demonstrate the ability of the criteria A and A_1 to discern places in Z where brightness changes rapidly. In particular, A and A_1 are not appreciably affected by smooth parts of Z, and $A(h)$ is asymptotically equal to the integrated squared height of jump discontinuities.

2.5. Properties of A and A_1 when applied to Z^θ

Let us assume that a graph of the surface $z = X(x, y)$ shows a nondegenerate collection, $\mathcal{C}$, of fault-type discontinuities, but that X is smooth elsewhere and, in addition, is compactly supported so that it is square-integrable on $\mathbb{R}^2$. (The square-block test-pattern, for which X is given at (2.6), is a case in point.) If we have chosen $\theta = \theta_0$ exactly right in the definition of $Z = Z^\theta$ at (2.7), then $Z = X$, and so by (2.13) and (2.14), $A(h)$ and $A_1(h)$ both converge to a strictly positive constant as $h \to 0$.

On the other hand, if the point-spread function ϕ^θ satisfies (2.9) and if $\theta < \theta_0$, then it can be proved, using a simple argument based on Fourier transforms, that the function $Z = Z^\theta$ has infinitely many square-integrable derivatives on $\mathbb{R}^2$. In consequence, the values of $A(h)$ and $A_1(h)$ for this Z both converge to zero as $h \to 0$. Therefore, the fact that we have used a value of θ that is too small can be recognised from the property that $A(h)$ and $A_1(h)$ are also unduly small.

Moving to the other side of θ_0, if ϕ^θ satisfies (2.9) and if $\theta > \theta_0$, then $A(h)$ and $A_1(h)$ are both infinite, for each $h > 0$. (In a practical setting, where noise and digitisation error are involved, and A and A_1 are computed by a regularisation procedure, "infinite" should be replaced by "very large.") This follows from Parseval's identity and the fact that if X has a fault-type discontinuity running through it, then its Fourier transform decreases to zero only polynomially quickly as frequency increases whereas, by (2.9), the ratio $|\phi_{\rm ft}^{\theta_0}(s,t)/\phi_{\rm ft}^{\theta}(s,t)|$ diverges faster than any polynomial as $s, t \to \infty$. In Appendix A.2 we give a rigorous proof for the square-block test-pattern function defined at (2.6); other functions with fault-type discontinuities can be treated in the same fashion.

These considerations suggest that we can estimate θ by constructing either $A(h)$ or $A_1(h)$ for a small value of θ, and steadily increasing θ until the criterion changes sharply, from a small to a very large value. The appropriate threshold can be determined through experimentation. Section 3 discusses implementation of this rule in the presence of noise, and Section 4 reports the results of a numerical study of properties of the rule.

In this discussion we have, for the sake of simplicity, confined attention to the case where θ is univariate and satisfies (2.9). However, the results described in the last two paragraphs can be generalised to a range of other settings. The case of the circular-stable point-spread function (2.5), where θ is bivariate, is a case in point. There, with $\theta = (\theta_1, \theta_2)$ denoting a general parameter choice and $\theta_0 = (\theta_{01}, \theta_{02})$ representing the true value, the following results are readily proved. If either $\theta_1 < \theta_{01}$, or both $\theta_1 = \theta_{01}$ and $\theta_2 < \theta_{02}$, then $Z = Z^\theta$ is infinitely differentiable and, therefore, $A(h)$ and $A_1(h)$ both converge to zero as $h \to 0$. If either $\theta_1 > \theta_{01}$, or both $\theta_1 = \theta_{01}$ and $\theta_2 > \theta_{02}$, then $A(h) = A_1(h) = \infty$

for $h > 0$. Therefore the true value of θ can be identified by examining either A or A_1 for a range of choices of θ.

2.6. Limitations to using differences as a means of determining a point-spread function

The principle on which our methodology is founded is that, if θ lies on one side of θ_0, then the partially restored function Z^θ is relatively smooth and, in particular, does not contain the fault-type discontinuities in the "true" function $X = Z_0^\theta$; and if θ lies on the other side, then Z^θ is so rough that it is not well-defined in the sense that the integral of its squared difference is infinite. If these properties do not hold then a method based on assessing the roughness of Z^θ, whether by using the criteria A and A_1 or by employing another approach, is unlikely to be effective.

Consider, for example, the case of a point-spread function which has the property that a fault-type discontinuity can persist in the restored function Z^θ no matter on what side of θ_0 the value of θ lies. In this context the presence of a discontinuity in Z^θ cannot be used to characterise the correct choice of θ. We show in Appendix A.2 that the circular-exponential function $\phi^\theta_{\rm exp}$, defined at (2.4), is of this type. This function does not satisfy (2.9), because altering the value of θ does not have a very large effect on the impact that smoothing has on high-frequency parts of the image. For example, if X denotes the square-block test-pattern function defined at (2.6), if $\theta_0 > 0$ is fixed, if $Z = Z^\theta$ is defined by (2.7), and if $A(h)$ and $A_1(h)$ are given by (2.10) and (2.11), then no matter what the value of θ, $A(h)$ and $A_1(h)$ converge to strictly positive constants as $h \to 0$. The function Z^θ, given at (2.7), still contains a fault-type discontinuity, no matter what the values of θ and θ_0.

This problem is intrinsic to methodologies that use differentiation, or differencing, to assess the efficacy of restoration. It is not an artifact of the particular techniques we have chosen. If using the incorrect amount of deblurring does not result in either marked undersmoothing (e.g., rendering a discontinuous function smooth) or oversmoothing (making a function extremely rough, in the sense that it is no longer square-integrable), then it is very difficult to use differentiation or differencing as a means of determining the extent of deblurring.

3. Effects of Noise

3.1. Main results

In practice the blurred image, represented in Section 2 by the function $Y^{\theta_0} = \phi^{\theta_0} X$, is not observed exactly. In particular, there is a degree of noise at the level of the pixel grid. In this section we investigate the impact of that source of error.

If noise is added to $\phi^{\theta_0}X$ then we need to re-define the function Y^{θ_0} that represents the observed image. Rather than treat pixels explicitly, we use a bivariate form of the white-noise model:

$$Y^{\theta_0}(x,y) = \big(\phi^{\theta_0}X\big)(x,y) + \delta\, dW(x,y), \tag{3.1}$$

where W denotes a standard Wiener process in the plane. White-noise models, as problems with Weiner-process noise are often called, are used in statistics as a means of ensuring theoretical justification for methodology, while retaining the insight and clarity that only a relatively incisive argument can provide. The equivalence of white-noise problems to discrete-data problems is well established in a range of settings. See, for example, Brown and Low (1996), Nussbaum (1996) and Brwon, Cai, Low and Zhang (2002).

In more conventional applications of the white-noise model in statistics, involving samples of size n, the "scale" δ is taken proportional to $n^{-1/2}$. In the context of data on a pixel grid, "sample size" can be represented by the number of pixels per unit area of the plane, N say, and δ is proportional to $N^{-1/2}$. Therefore, our asymptotic theory will be based on δ decreasing to zero.

In order to counteract the effects of noise we compute Fourier transforms in an increasing, although compact, region that for simplicity is assumed to be the disc $\mathcal{R}_r$ of radius r centred at the origin. (An alternative approach would be to use a compactly supported stochastic process in place of W, at (3.1), but that would arguably not be a realistic assumption.) To distinguish truncated Fourier transforms from regular ones, we use the subscript "tft," rather than "ft," to indicate the former. In particular, the truncated Fourier transform of Y^{θ_0}, at (3.1), is

$$\begin{aligned} Y^{\theta_0}_{\rm tft}(s,t) = \frac{1}{(2\pi)^2} \iint_{\mathcal{R}_r} \big(\phi^{\theta_0}X\big)(x,y) \exp\{-(isx+ity)\} dx\, dy \\ + \frac{\delta}{(2\pi)^2} \iint_{\mathcal{R}_r} \exp\{-(isx+ity)\} dW(x,y). \end{aligned} \tag{3.2}$$

The dependence of $Y^{\theta_0}_{\rm tft}$ on θ_0, here and below, serves only to indicate that $Y^{\theta_0}_{\rm tft}$ is based on the actual signal Y^{θ_0}, computed for the true value of the parameter. It does not mean that the true parameter value has to be known in order to compute $Y^{\theta_0}_{\rm tft}$.

We "deblur" Y^{θ_0} by dividing $Y^{\theta_0}_{\rm tft}$ by $\phi^{\theta}_{\rm ft}$, treating the ratio as a Fourier transform, and inverting it using a hard-thresholding approach to regularisation, obtaining

$$V^{\theta}(x,y) = \iint \frac{Y^{\theta_0}_{\rm tft}(s,t)}{\phi^{\theta}_{\rm ft}(s,t)} \exp\{i(sx+ty)\} I\big(s^2+t^2 \le \lambda^2\big) ds\, dt, \tag{3.3}$$

where $\lambda > 0$ denotes the threshold. If there were no noise, in particular if $\delta = 0$ in (3.1), and if we were to take $r = \lambda = \infty$ in (3.2) and (3.3), then V^θ would be identical to Z^θ defined at (2.7). When $\delta \neq 0$, finite values of r and λ are necessary in order to reduce the effects of noise. In such cases, the function V^θ at (3.3) is generally not real-valued. However this is not a problem, since we are not interested in V^θ for its own sake, but instead in using it to construct versions of the difference operators $A(h)$ and $A_1(h)$.

Indeed, writing $|z|$ for the absolute value of a complex number z, we construct the following generalised forms of $A(h)$ and $A_1(h)$ at (2.10) and (2.11), respectively:

$$B^\theta(h) = (2h)^{-1} \iint \max_\omega \Big| V^\theta(x + h\cos\omega, y + h\sin\omega) - V^\theta(x - h\cos\omega, y - h\sin\omega) \Big|^2 \, dx\, dy,$$

$$B_1^\theta(h) = (2h)^{-1} \iint \Big\{ \big| V^\theta(x+h, y) - V^\theta(x-h, y) \big|^2 + \big| V^\theta(x, y+h) - V^\theta(x, y-h) \big|^2 \Big\} \, dx\, dy.$$

For simplicity we work only with the latter, which by Parseval's identity is

$$B_1^\theta(h) = 2h^{-1} \frac{1}{(2\pi)^2} \iint \left| \frac{Y_{\mathrm{tft}}^{\theta_0}(s,t)}{\phi_{\mathrm{ft}}^\theta(s,t)} \right|^2 (\sin^2 sh + \sin^2 th) I(s^2 + t^2 \le \lambda^2) ds dt. \quad (3.4)$$

If there is no noise and we take $r = \lambda = \infty$, then B_1^θ is identical to A_1 at (2.11). Note particularly that, as indicated by (3.4), smoothing is undertaken in the frequency rather than the spatial domain. As a result, smoothing has very little deleterious impact on jump discontinuities or similar non-regular features of the true image.

The discussion in Section 2 suggests that simple thresholding rules, based on the value of $B_1^\theta(h)$, may be used to determine θ_0. Two such rules are given below.

> Let C be strictly positive but otherwise arbitrary, and let $\hat\theta$ denote either the least value of θ for which $B_1^\theta(h) \ge C$, or the least value of θ such that $B_1^{\theta_1}(h) \ge C$ for all $\theta_1 \ge \theta$. (3.5)

We claim that, for a range of point-spread functions which includes those satisfying (2.9), both algorithms in (3.5) produce statistically consistent estimators of θ_0. The convergence rate of estimators of θ_0 can generally be improved by using refined versions of those algorithms, but that will not concern us here; for

simplicity we focus only on consistency. Moreover, since the assumptions imposed on h, r and λ will depend on choice of the parametric class of point-spread function, we reduce the length of our treatment by considering only the Gaussian case, at (2.3). Our method of proof is also valid for circular-stable point-spread functions, at (2.5).

We assume that the image function X satisfies the following:

X is bounded and compactly supported, and its Fourier transform $X_{\rm ft}$ decreases to zero at only a polynomial rate as s, t increase, in the sense that there exist constants $C_1, C_2 > 0$ and $0 < c_1 < c_2 < \infty$ such that the interval (c_1, c_2) does not contain an integer multiple of π^2, and (3.6)

$$\iint |X_{\rm ft}(s,t)|^2 \, I\big(c_1 u \le s^2 + t^2 \le c_2 u\big) ds\, dt \ge C_1 u^{-C_2}$$

for all sufficiently large u.

Condition (3.6) asks that the image X not be too smooth. To appreciate why, note that the displayed integral in (3.6) decreases to zero very fast, as a function of increasing u, only if X is smooth. Only in such cases does the inequality in (3.6) fail. In particular, (3.6) holds for images that include jump discontinuities.

However, (3.6) also holds for many functions X that are quite smooth. For example, consider the case

$$X(x,y) = \begin{cases} C_3(1-x^2)^k(1-y^2)^k & \text{if } |x| \le 1 \text{ and } |y| \le 1 \\ 0 & \text{otherwise,} \end{cases} \tag{3.7}$$

with integer $k \ge 0$ and a C_3 nonzero constant. Then (3.6) is satisfied. This choice of X at (3.7) has k bounded derivatives in the plane, and in particular does not, unless $k = 0$, have the jump discontinuities that a typical test-pattern function would enjoy. Nevertheless, our technique gives consistent estimation in the case of such an X. The square-block test-pattern function (2.6) is an example of the case $k = 0$.

We also need to prescribe the manner in which h, in the definition of $B_1^\theta(h)$, and the parameters r and λ of the regularisation vary together. We consider each of h, r and λ to be a function of δ, and assume that r and λ diverge to infinity and that

$$\frac{\lambda}{r}\,, \frac{\lambda}{|\log \delta|^{\frac12}} \text{ and } \frac{(\log r)}{\lambda} \text{ all converge to 0, and } h\lambda \to \infty, \text{ as } \delta \to 0. \tag{3.8}$$

Therefore, r can be exponentially large as a function of λ, and λ can be exponentially large as a function of δ, before our method fails in a serious way. This

robustness against values of these parameters will be explored numerically in Section 4.

Theorem. *Assume* (3.6) *and* (3.8)*, and that the point-spread function* ϕ^θ *is given by* (2.3)*. Let* $\hat{\theta}$ *be defined by either of the algorithms at* (3.5)*. Then, with probability 1,* $\hat{\theta} \to \theta_0$ *as* $\delta \to 0$*.*

A proof of the theorem is given in Appendix A.3.

4. Numerical Properties

4.1. Square-block test pattern

Here we summarise the results of a simulation study where the true image $X(x, y)$ was defined in the unit square $[0, 1] \times [0, 1]$, $X(x, y) = 2 - 2(x - 0.5)^2 - 2(y - 0.5)^2$ when (x, y) was in the central square $[0.25, 0.75] \times [0.25, 0.75]$, and $X(x, y) = 1 - 2(x - 0.5)^2 - 2(y - 0.5)^2$ when (x, y) was outside the central square. Observations were generated from the model

$$Y(\frac{i}{n}, \frac{j}{n}) = (\phi^{\theta_0} X)(\frac{i}{n}, \frac{j}{n}) + \epsilon_{ij}, \quad i, j = 1, \ldots, n,$$

where the true point-spread function ϕ^θ was assumed to be ϕ^θ_{Gau}, at (2.3), the true value of θ was $\theta_0 = 0.03^2$, the errors ϵ_{ij} were independent and identically distributed as normal $\text{N}(0, \sigma^2)$, and n^2 was the sample size. When $n = 128$ and $\sigma = 0.02$, the original image, its blurred version, and the blurred-and-noisy version are shown in Figure 1(a)−1(c), respectively.

Using the inversion formula (3.3), when the parameters (θ, λ) were chosen to be $(0.02^2, 60)$, $(0.03^2, 60)$ or $(0.04^2, 15)$, the restored images obtained from the image in Figure 1(b) are shown in Figure 1(d)−1(f), respectively. Note that, in this example, the domain of definition of the true image X is finite. Therefore, $\mathcal{R}_r$ in (3.2) can be simply chosen to be the whole domain of definition of X (i.e., $[0, 1] \times [0, 1]$). From Figure 1, it can be seen that when $\theta < \theta_0$, the image is under-deblurred; it is perfectly deblurred when $\theta = \theta_0$; and it is over-deblurred when $\theta > \theta_0$. In the first two cases, theoretically speaking, λ could have been chosen equal to infinity. We selected $\lambda = 60$ so as to avoid underflow in computation. For the same reason, a smaller λ was chosen in the third case. The corresponding restored images from the blurred-and-noisy image in Figure 1(c) are shown in Figure 1(g)−1(i), where λ was chosen to be 16, 16 and 15, respectively. Smaller λ values were chosen in this case due to noise. It can be seen that image restoration is more difficult in the presence of noise. The periodic artifacts seen in some panels of Figure 1 are the result of oscillating trigonometric terms included in the Fourier transformation. These are mainly caused by the thresholding scheme used in the image estimator (see (3.3)) and by noise.

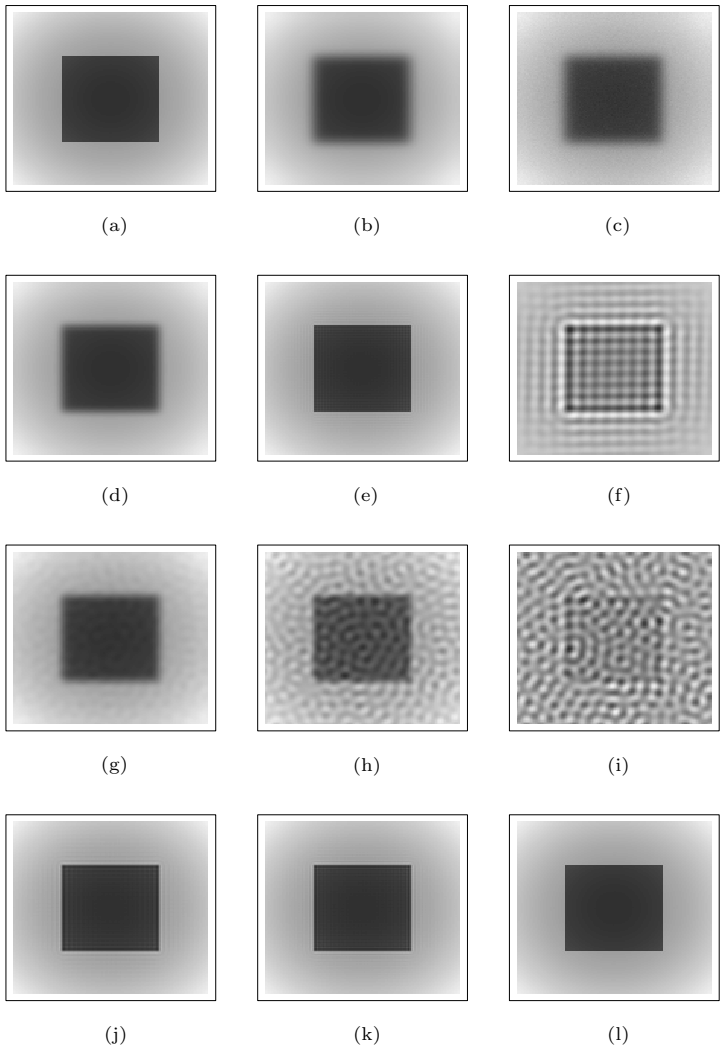
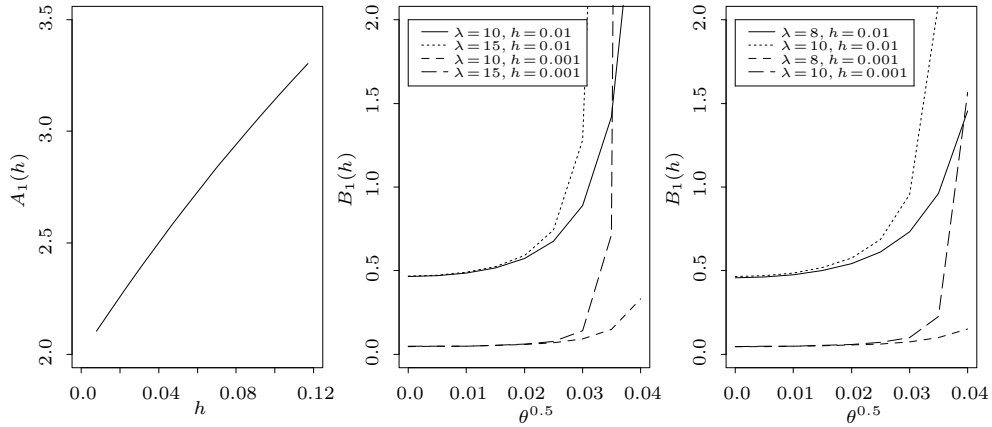


Figure 1. Graphs of X, ϕX, Y, and deblurred images. Panels (a)−(c) show the true image X, its blurred version ϕX, and the blurred-and-noisy version, respectively. Panels (d)−(f) show the deblurred images, based on (3.3) and constructed from (b), when θ in the deblurring operator is smaller than θ_0, equal to θ_0, and larger than θ_0, respectively. Panels (g)−(i) show the corresponding deblurred images from (c). Panels (j)−(l) show the deblurred images, constructed from (b), when $\theta = 0.03^2$ and λ equals 40, 50, and 70, respectively.

This example also suggests a practical approach to choosing λ and θ in (3.3). If the computer program performs well for a particular value of λ when

θ is smaller than a specific value, say θ^*, but has an underflow problem when θ is larger than θ^*, then the true value of θ should be close to θ^*, and λ should be chosen accordingly for best visual impression. Near θ^*, the restored image is quite robust to λ. To see this, the restored images obtained from the image in Figure 1(b) are shown in Figure 1(j)−1(l), respectively, when $\theta = 0.03^2$ and λ equals 40, 50, and 70. It can be seen that results with different values of λ are almost identical.



(a) (b) (c)

Figure 2. Graphs of $A_1(h)$ and $B_1^\theta(h)$. Panel (a) shows $A_1(h)$. Panel (b) graphs $B_1^\theta(h)$ when $\theta^{1/2}$ varies within $[0, 0.04]$, when $\sigma = 0$, and when (λ, h) takes one of the values $(10, 0.01)$, $(15, 0.01)$, $(10, 0.001)$, and $(15, 0.001)$. Panel (c) depicts the corresponding results when $\sigma = 0.02$ and (λ, h) takes one of the values $(8, 0.01)$, $(10, 0.01)$, $(8, 0.001)$, and $(10, 0.001)$.

The image restoration procedure (3.3), and the procedure (3.5) for estimating θ_0, are based on the properties (2.13) and (2.14) of $A(h)$ and $A_1(h)$. In the case of the square-block test pattern considered above, $A(h) = A_1(h)$, and $\int_{\mathcal{C}} f(s)^2\, ds = 2$ if $Z = X$ in (2.13) and (2.14). The plot of $A_1(h)$ is shown in Figure 2(a). It can be seen that $A_1(h)$ tends to $\int_{\mathcal{C}} f(s)^2\, ds$ when h converges to zero. When noise is involved, the criterion $A_1(h)$ should be replaced by $B_1^\theta(h)$, defined at (3.4). Figure 2(b) plots $B_1^\theta(h)$ when $\theta^{1/2}$ varies within the range $[0, 0.04]$, when $\sigma = 0$, and when (λ, h) takes one of the values $(10, 0.01)$, $(15, 0.01)$, $(10, 0.001)$, and $(15, 0.001)$. The corresponding results when $\sigma = 0.02$ and (λ, h) takes one of the values $(8, 0.01)$, $(10, 0.01)$, $(8, 0.001)$, and $(10, 0.001)$ are presented in Figure 2(c). It can be seen that: (a) $B_1^\theta(h)$ increases with θ and λ; (b) if we decrease the value of h, the value of θ at which $B_1^\theta(h)$ would start to change dramatically with λ would be closer to θ_0; and (c) when σ is larger, the value of λ should be chosen relatively small, which is consistent with results presented in Table 1 below. These facts should be helpful for determining $\hat{\theta}$ in applications.

Table 1. Optimal values of λ and h in (3.5). Tabulated are the smallest values of λ and h such that the mean squared error of $\hat{\theta}$ is zero, based on 100 simulations.

	$n = 64$			$n = 128$		
C	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$
1	(12,0.01)	(12,0.01)	(12,0.01)	(12,0.01)	(12,0.01)	(12,0.01)
10	(18,0.01)	(17,0.01)	(15,0.01)	(19,0.01)	(18,0.01)	(16,0.01)
100	(19,0.01)	(18,0.01)	(17,0.01)	(20,0.01)	(19,0.01)	(18,0.01)

To estimate θ_0 using (3.5), we considered cases where n was either 64 or 128, σ changed its value among 0.01, 0.02 and 0.05, and the threshold C in (3.5) was 1, 10 or 100. In each case, the optimal values of λ and h were sought in $[1, 25]$ and among 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4}, respectively, for minimising the mean squared error of $\hat{\theta}$ based on 100 replications. We found that (a) the two versions of $\hat{\theta}$ in (3.5) gave exactly the same results in all cases, and (b) in each case we could always find values of λ and h such that mean squared error vanished (i.e., $\hat{\theta} = \theta_0$) in all 100 replications. The smallest values of λ and h for which mean squared error vanished are listed in Table 1. It can be seen that the optimal value of h is stable when n, σ and C change. The value of λ should be chosen larger when n is larger, when σ is smaller, or when C is larger.

Image analysis enjoys an advantage that smoothing methods applied to related problems do not: even if an observer has never seen the true image, he or she often has an accurate impression, gained from extensive experience, of the scene to which the degraded image is an approximation. It is appropriate, and very common in practice, to exploit this information and choose tuning parameters so as to produce a restoration that gives the best visual impression. This approach has the advantage that it is not limited by metric-based, technical accounts of performance. While commonly used in theoretical work, those approaches are well-known not to respect perceived visual fidelity. See, for example, Marron and Tsybakov (1995).

4.2. Lena image example

Next we applied our image restoration procedure, described by (3.3)−(3.5), to the Lena image. The original Lena image shown in Figure 3(a) has 256×256 pixels, with grey levels in the range $[0, 255]$. Its version degraded by the Gaussian point-spread function (2.3), with $\theta_0 = 0.015^2$, and by independent and identically distributed noise from the $N(0, 1)$ distribution, is illustrated in Figure 3(b).

The criterion $B_1^\theta(h)$ is shown in Figure 3(c) when $\lambda = 20$ or 25, and $h = 0.01$ or 0.001. It can be seen from the plot that $B_1^\theta(h)$ starts to increase dramatically around $\theta = 0.015^2$ when λ increases from 20 to 25, and when $h = 0.001$. In

(3.5), if we chose $C = 1,000$ then we obtained $\hat{\theta} = 0.015^2$ when $\lambda \geq 20$ and $h = 0.001$. Combining these two facts, it can be concluded that θ_0 is close to 0.015^2. In applications, we can also obtain a reasonable estimator of θ_0 by trying several values of θ. The restored Lena images, using the method at (3.3) when $\lambda = 1,000$ and $\theta = 0.01^2$, 0.014^2, 0.015^2, 0.016^2 or 0.02^2, are shown in Figure 3(d)−(h), respectively. From these plots it can be seen that the restored images are quite good when θ is close to θ_0; compare Figure 3(e)−(g).

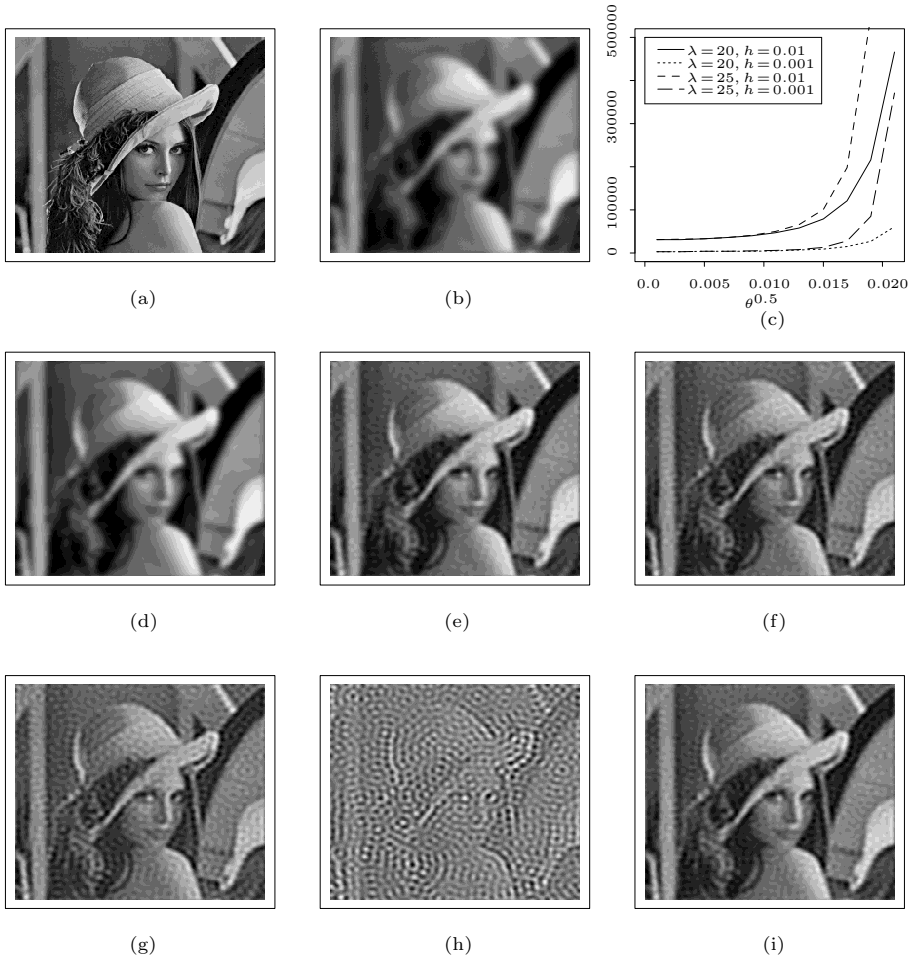


Figure 3. *Lena image example.* Panel (a) shows the original Lena image. Panel (b) depicts the blurred-and-noisy Lena image, obtained via the Gaussian point-spread function (2.3) with $\theta_0 = 0.015^2$, further degraded by independent and identically distributed noise from the $N(0, 1)$ distribution. Panel (c) shows the criterion $B_1^{\theta}(h)$ when $\lambda = 20$ or 25, and $h = 0.01$ or 0.001. Panels (d)−(h) show the degraded images using (3.3)−(3.5) with $\lambda = 1,000$ and θ equal to 0.01^2, 0.014^2, 0.015^2, 0.016^2 or 0.02^2, respectively. Panel (i) shows the image restored using the Wiener filter.

The restored image defined at (3.3) was obtained using an inverse filter with hard thresholding. In the literature there are several other procedures for recovering X from Y when the point-spread function is specified. For example, the restored image using a Wiener filter is defined by

$$\widehat{X}(x,y) = \frac{1}{(2\pi)^2} \Re\biggl[\iint \frac{\bar{\phi}^\theta_{\rm ft}(s,t)}{|\phi^\theta_{\rm ft}(s,t)|^2 + \alpha(s^2+t^2)^{\frac{\beta}{2}}} Y_{\rm ft}(s,t) \exp\{i(sx+ty)\} ds\, dt \biggr],$$

where $\Re$ denotes "real part," $\bar{\phi}^\theta_{\rm ft}$ is the complex conjugate of $\phi^\theta_{\rm ft}$, and $\alpha, \beta > 0$ are parameters. The Wiener filter is derived by minimising the mean integrated squared error of the restored image, under the assumption that the error distribution is Gaussian; see Gonzalez and Woods (1992, Chap. 5). The restored Lena image is shown in Figure 3(i) in the case $\alpha = 5 \times 10^{-6}$, $\beta = 1$ and $\theta = 0.015^2$. It can be seen that the result is similar to that shown in Figure 3(f), obtained there using the inverse filter with hard thresholding.

For real images, such as the Lena photograph, errors due to the impact of noise are more clearly visible in regions where a viewer feels the grey shade should be constant, than they are in other areas. However, in quantitative terms the performance of our method is no less good in regions of perceived constant grey shade than it is in the vicinity of discontinuities.

5. Generalisations and New Directions

Our results can be extended in a number of directions. For instance, the theorem in Section 3 can be generalised so that it applies to a much wider variety of point-spread functions than the spherically symmetric Gaussian one treated there. In particular, it is straightforward to show that the theorem holds for asymmetric Gaussian point-spread functions. There, the scale on each principal axis, and the angles of orientation of each these axes, can be estimated consistently from data. Some methods for manufacturing optical lenses can result in asymmetries of this type.

Cases where the point-spread function varies spatially in a reasonably straightforward manner can also be treated. For example, the instance where θ in (2.3) alters with location can be treated, provided the variation is particularly smooth.

The theorem in Section 3 can also be generalised to a substantially larger class of point-spread functions, where the characteristic function $\phi_{\rm ft}(t)$ decays exponentially fast as the norm, $\|t\|$, of frequency diverges to infinity. Examples include spherically symmetric cases where the profile of the point-spread function is the density of a univariate stable law.

It is also possible to explore, in more detail, cases where our method does not perform so well, for example those discussed in Section 2.6. Take, for example,

the case where the point-spread function is "rough but not too rough," in the sense that it is not infinitely differentiable but (unlike the circular-exponential example discussed at (2.4)) has a number of bounded derivatives. The extent to which our method works can be quantified theoretically by taking first the limit as δ converges to zero, and then permitting the number of derivatives of the point-spread function to increase. On the other hand, if the number of derivatives is not large then fitting the wrong model (e.g., a Gaussian model, even if it is incorrect) may sometimes give acceptable results.

More broadly, limitations of the models and methodology we have developed can be explored. Note that the point-spread function model exemplified by (2.1) and (3.1) is appropriate only in the setting of a linear filter. Non-linear filters cannot be removed so readily; in general, our approach would find it difficult to distinguish non-linear filters from noise. Conversely, our approach would confuse noise added to the image, before blur, with the true image, and would fail to suppress noise in such cases. Methods quite different from those that we have discussed are needed to deal with problems such as this.

Appendix A.1. Integrals of Squared Differences of Images

Let $\mathcal{C}$ denote a curve in the plane, having a continuously turning tangent, except possibly for a finite number of jump discontinuities in the tangent (giving corners of $\mathcal{C}$); assume $\mathcal{C}$ is of finite length and, for the present, is neither disconnected nor self-intersecting; let f denote a continuous function from $\mathcal{C}$ to the real line; and, for points in $\mathbb{R}^2$ close to but not on $\mathcal{C}$, identify two sides, $\mathcal{S}_1$ and $\mathcal{S}_2$ say, of $\mathcal{C}$. Let X be a compactly supported function from $\mathbb{R}^2$ to $\mathbb{R}$, with the property that (a) both partial derivatives of X are bounded uniformly in $(x, y) \in \mathbb{R}^2 \setminus \mathcal{C}$, and (b) if $s \in \mathcal{C}$ then $X(s_1) - X(s_2) \to f(s)$ as $s_j \to s$ for $j = 1, 2$, with s_j on side $\mathcal{S}_j$ of $\mathcal{C}$. Define $A(h)$ as at (2.10), although for the function X instead of Z. Then (2.13) holds.

More generally, we do not need to confine attention to instances where the image X includes a single, non-self intersecting curve along which a fault-type discontinuity occurs. General cases may be constructed by taking $X = X_0$ to be the sum of any finite number of functions $X_1, \ldots, X_m$ say, each of which has the properties of X given in the previous paragraph, with respective fault-line curves $\mathcal{C}_1, \ldots, \mathcal{C}_m$ and jump-size functions $f_1, \ldots, f_m$. In such cases the fault line along which X_0 has jump discontinuities will be $\mathcal{C}_0 = \cup_{1 \le j \le m} \mathcal{C}_j$; it may be self-intersecting and is not necessarily connected. If each triple $(X_j, \mathcal{C}_j, f_j)$ satisfies the conditions given for $(X, \mathcal{C}, f)$ in the previous paragraph, then $X_0 = \sum_{1 \le j \le m} X_j$ has a fault-type discontinuity along $\mathcal{C}_0$, with jump size $f_0(s)$, say, at

$s \in \mathcal{C}_0$; and results (2.13) continue to hold, in the forms,

$$A(h) \equiv (2h)^{-1} \iint \max_{\omega} \{X_0(x + h\cos\omega, y + h\sin\omega) - X_0(x - h\cos\omega, y - h\sin\omega)\}^2 \, dx\, dy \to \int_{\mathcal{C}_0} f_0(s)^2 ds,$$

$$A_1(h) \equiv (2h)^{-1} \iint \Big[\{X_0(x + h, y) - X_0(x - h, y)\}^2 + \{X_0(x, y + h) - X_0(x, y - h)\}^2\Big] dx\, dy \to \alpha \int_{\mathcal{C}_0} f_0(s)^2 ds,$$

where $1 \le \alpha \le \sqrt{2}$.

Our proofs of these properties will treat the simpler setting discussed two paragraphs above. Let $\mathcal{C}(h)$ denote the set of all points (x, y) that are distant less than h from at least one point of $\mathcal{C}$. Note that, by assumption (a) two paragraphs above, the integrand of (2.10) equals $O(h^2)$ uniformly in $(x, y) \notin \mathcal{C}(h)$. This result, and the compact support of X, imply that, if the integral on the right-hand side of (2.10) were taken over $(x, y) \notin \mathcal{C}(h)$, then the right-hand side of (2.10) would converge to zero as $h \to 0$. Therefore, to prove (2.13) it suffices to establish that result when $A(h)$ there is replaced by $A_2\{h, \mathcal{C}(h)\}$:

$$A_2\{h, \mathcal{C}(h)\} = (2h)^{-1} \iint_{\mathcal{C}(h)} \max_{\omega} \Big\{X(x + h\cos\omega, y + h\sin\omega) - X(x - h\cos\omega, y - h\sin\omega)\Big\}^2 dx\, dy.$$

Since $\mathcal{C}$ has a continuously turning tangent in a piecewise sense then, to prove the just-mentioned result, it suffices to treat the case where $\mathcal{C}$ consists of a finite sequence of nondegenerate line segments; all other cases can be addressed by approximation. In the line-segment context, suppose $\mathcal{C}$ is the union of disjoint, nondegenerate line segments $\mathcal{D}_1, \ldots, \mathcal{D}_k$. Then, since $\mathcal{C}(h) \setminus \cup_j \mathcal{D}_j(h)$ is empty, and since the measure of the set of points (x, y) that are in two or more $\mathcal{D}_j(h)$'s equals $O(h^2)$, we have:

$$A_2\{h, \mathcal{C}(h)\} = \sum_{j=1}^{k} A_2\{h, \mathcal{D}_j(h)\} + o(1).$$

Therefore, it suffices to prove (2.13) with $A(h)$ and $\mathcal{C}$ there replaced by $A_2\{h, \mathcal{D}_j(h)\}$ and $\mathcal{D}_j$, respectively. Call this result (R_j).

It is straightforward to prove that if $\mathcal{E}_j(h)$ denotes the set of points in the plane that are not on $\mathcal{D}_j$ but are distant strictly less than h from $\mathcal{D}_j$ and no closer than h to either end of $\mathcal{D}_j$; if $\mathcal{F}_j(h)$ is the set of points that are within h of

either end of $\mathcal{D}_j$; and if $s(x,y)$ denotes the point on $\mathcal{D}_j$ that is nearest to (x,y); then, as $h \to 0$,

$$\begin{aligned} Y(x,y) &\equiv \max_{\omega} \Big\{ X(x + h\cos\omega, y + h\sin\omega) - X(x - h\cos\omega, y - h\sin\omega) \Big\}^2 \\ &= f\{s(x,y)\}^2 + o(1) \end{aligned} \tag{A.1}$$

uniformly in $(x,y) \in \mathcal{E}_j(h)$, and

$$Y(x,y) = O(1) \tag{A.2}$$

uniformly in $(x,y) \in \mathcal{F}_j(h)$. Since the area of $\mathcal{F}_j(h)$ equals $O(h^2)$ then, by adding the integrals of $Y(x,y)$ over $\mathcal{E}_j(h)$ and $\mathcal{F}_j(h)$, we conclude that (R_j) holds. This concludes our derivation of (2.13).

The proof of (2.14) is similar. Arguing as above we see that, if we define $\mathcal{C}(h)$ as before, and re-define

$$\begin{aligned} A_2\{h, \mathcal{C}(h)\} = (2h)^{-1} \iint_{\mathcal{C}(h)} \Big[&\{X(x+h,y) - X(x-h,y)\}^2 \\ &+ \{X(x,y+h) - X(x,y-h)\}^2 \Big] dx\, dy, \end{aligned}$$

then it suffices to prove that (2.14) holds if $A(h)$ there is replaced by $A_2\{h, \mathcal{C}(h)\}$. To establish this result it is enough to consider the case where $\mathcal{C}$ is piecewise-linear, with components $\mathcal{D}_1, \ldots, \mathcal{D}_k$, and prove that for each j,

$$A_2\{h, \mathcal{D}_j(h)\} \to \alpha_j \int_{\mathcal{D}_j} f_j(s)^2 ds, \tag{A.3}$$

where $0 \le \alpha_j \le \sqrt{2}$ and $f_j(s)$ denotes the jump height at $s \in \mathcal{D}_j$.

Let $\mathcal{E}_j(h)$ and $\mathcal{F}_j(h)$ be as in the proof of (2.13), and introduce $\mathcal{G}_j^V(h)$ and $\mathcal{G}_j^H(h)$ as the sets of points in $\mathcal{E}_j(h)$ that are distant strictly less than h from $\mathcal{D}_j$ in the vertical and horizontal directions, respectively, and $\mathcal{H}_j^V(h)$ and $\mathcal{H}_j^H(h)$ as the sets of points in $\mathcal{E}_j(h)$ that are strictly further than h from $\mathcal{D}_j$ in the vertical and horizontal directions, respectively. Recall that $s(x,y)$ denotes the point on $\mathcal{D}_j$ that is nearest to (x,y). Then, in place of (A.1),

$$\begin{aligned} &\{X(x+h,y) - X(x-h,y)\}^2 = f\{s(x,y)\}^2 + o(1), \text{ uniformly in } (x,y) \in \mathcal{G}_j^H(h), \\ &\{X(x,y+h) - X(x,y-h)\}^2 = f\{s(x,y)\}^2 + o(1), \text{ uniformly in } (x,y) \in \mathcal{G}_j^V(h), \end{aligned}$$

and in place of (A.2),

$$\{X(x+h,y) - X(x-h,y)\}^2 + \{X(x,y+h) - X(x,y-h)\}^2 = O(1)$$

uniformly in $(x, y) \in \mathcal{F}_j(h)$. Additionally,

$$\{X(x+h,y) - X(x-h,y)\}^2 = O\big(h^2\big), \text{ uniformly in } (x,y) \in \mathcal{H}_j^H(h),$$
$$\{X(x,y+h) - X(x,y-h)\}^2 = O\big(h^2\big), \text{ uniformly in } (x,y) \in \mathcal{H}_j^V(h).$$

Combining these results, and defining $\mathcal{D}_j^H$ [respectively, $\mathcal{D}_j^V$] to be the set of x (respectively, y) such that $(x, y) \in \mathcal{D}_j$ for some y (for some x), we deduce that

$$\begin{aligned}
&(2h)^{-1} \iint_{\mathcal{D}_j(h)} \{X(x+h,y) - X(x-h,y)\}^2 dx\, dy \\
&\quad = (2h)^{-1} \iint_{\mathcal{G}_j^H(h)} \{X(x+h,y) - X(x-h,y)\}^2 dx\, dy + o(1) \\
&\quad = \int_{\mathcal{D}_j^H} f\{s(x,y)\}^2 dx + o(1), \qquad \text{(A.4)} \\
&(2h)^{-1} \iint_{\mathcal{D}_j(h)} \{X(x,y+h) - X(x,y-h)\}^2 dx\, dy \\
&\quad = \int_{\mathcal{D}_j^V} f\{s(x,y)\}^2 dy + o(1). \qquad \text{(A.5)}
\end{aligned}$$

Assume $\mathcal{D}_j(h)$ is inclined at angle $\omega \in [0, \pi/2]$ to the positive direction of the horizontal axis. (Other orientations may be treated similarly.) Combining (A.4) and (A.5) we deduce that

$$\begin{aligned}
A_2\{h, \mathcal{D}_j(h)\} &\to \int_{\mathcal{D}_j^H} f\{s(x,y)\}^2 dx + \int_{\mathcal{D}_j^V} f\{s(x,y)\}^2 dy \\
&= (\cos\omega + \sin\omega) \int_{\mathcal{D}_j} f\{s(x,y)\}^2 ds. \qquad \text{(A.6)}
\end{aligned}$$

Since $1 \le \cos\omega + \sin\omega \le \sqrt{2}$ then (A.3) follows from (A.6).

Appendix A.2. Proofs for Square-Block Test-Pattern

Using Parseval's identity it may be proved that if the function Z has Fourier transform Z_{ft}, and if $A_1(h)$ is given by (2.11), then

$$A_1(h) = 2h^{-1}(2\pi)^2 \iint |Z_{\text{ft}}(s,t)|^2 \big(\sin^2 sh + \sin^2 th\big) ds\, dt. \qquad \text{(A.7)}$$

If we take $Z = X$, where X denotes the function defining the square-block test-pattern, then the fact that the sides of the pattern are parallel to the coordinate axes, and of equal length and height, implies that $A_1 = A$, and so $A(h)$ is also given by (A.7).

The Fourier transform of the square-block test-pattern function X, given at (2.6), is

$$X_{\rm ft}(s,t) = \frac{\sin s \, \sin t}{\pi^2 st},$$

and the Fourier transform of Z^θ, given by (2.7), is

$$Z^\theta_{\rm ft} = \frac{\phi^{\theta_0}_{\rm ft} X_{\rm ft}}{\phi^\theta_{\rm ft}}. \tag{A.8}$$

Hence by (A.7), the version of $A_1(h)$ for Z^θ is

$$A_1(h) = \frac{16}{h\pi^2} \iint \left(\frac{\sin s \, \sin t}{st} \right)^2 \left\{ \frac{\phi^{\theta_0}_{\rm ft}(s,t)}{\phi^\theta(s,t)} \right\}^2 (\sin sh)^2 ds \, dt. \tag{A.9}$$

We stated in Section 2.5 that if the function X contains fault-type discontinuities, if ϕ^θ satisfies (2.9), and if $\theta > \theta_0$, then $A(h) = A_1(h) = \infty$ for each $h > 0$. We prove this result here for the square-block test-pattern, showing that if $\theta > \theta_0$ then $A_1(h) = \infty$. The infiniteness of $A(h)$ then follows via (2.12). Now, (2.9) implies that if $\theta > \theta_0$ then for each $C > 0$,

$$A_1(h) > \text{const.} \iint \left(\frac{\sin s \, \sin t}{st} \right)^2 (1 + |s| + |t|)^C (\sin sh)^2 ds \, dt,$$

which is clearly infinite if $C \geq 2$.

Next we prove that if $\phi^\theta = \phi^\theta_{\rm exp}$ denotes the circular-exponential point-spread function at (2.4), if θ_0 is positive but otherwise arbitrary, and if we apply the difference operators $A(h)$ and $A_1(h)$ to $Z = Z^\theta$ given at (2.7), then (a) for each $\theta > 0$, Z^θ is square-integrable, and (b) $A_1(h) \to C_1(\theta/\theta_0)^6$ as $h \to 0$, where $C_1 > 0$ does not depend on θ or θ_0. We know from (2.12) that $A_1(h)$ is a lower bound to $2\,A(h)$, and in fact a longer argument will show that $A(h) \to C_2\,(\theta/\theta_0)^6$, where $C_2 > 0$.

It can be proved that for each $\theta > 0$,

$$\begin{aligned} &\phi^\theta_{\rm ft}(s,t) \text{ is real-valued, strictly positive for all } (s,t), \text{ and satisfies} \\ &\phi^\theta_{\rm ft}(s,t) \sim C_3 \theta^{-3} (s^2 + t^2)^{-\frac{3}{2}} \text{ as } |s| + |t| \to \infty, \text{ where } C_3 > 0 \text{ de-} \\ &\text{notes a constant.} \end{aligned} \tag{A.10}$$

This property and (A.8) imply that $Z^\theta_{\rm ft}$, and hence (by Parseval's identity) Z^θ itself, are square-integrable, thus giving (a) above.

Define $J = \int (t^{-1} \sin t)^2 dt$, where, here and below, integrals are taken over the whole real line. Combining (A.9) and (A.10) we deduce that for a constant

$C_4 > 0$ not depending on θ or θ_0,

$$A_1(h) \sim \frac{C_4\theta^6}{h{\theta_0}^6} \iint \left(\frac{\sin s\ \sin t}{st}\right)^2 (\sin sh)^2 ds\, dt$$
$$= \frac{C_4 J\theta^6}{{\theta_0}^6} \int \left(\frac{\sin s}{s}\right)^2 \sin^2(\frac{s}{h}) ds \to \frac{C_4 J^2\theta^6}{2{\theta_0}^6}.$$

That is, $A_1(h) \to C(\theta/\theta_0)^6$ as $h \to 0$, where $C > 0$, and so (b) above is true.

Appendix A.3. Proof of Theorem

Let $\phi^\theta_{\rm ft}$ denote the Fourier transform defined at (2.8), and let $X_{\rm ft}$ be the conventional Fourier transform of X:

$$X_{\rm ft}(s,t) = \frac{1}{(2\pi)^2} \iint X(x,y) \exp\{-(isx + ity)\} dx\, dy.$$

The Fourier transform of $\phi^\theta_0 X$ equals the product, $\phi^{\theta_0}_{\rm ft} X_{\rm ft}$ and differs from

$$\big(\phi^{\theta_0} X\big)_{\rm tft}(s,t) = \frac{1}{(2\pi)^2} \iint_{\mathcal{R}_r} \big(\phi^{\theta_0} X\big)(x,y) \exp\{-(isx + ity)\} dx\, dy$$

only in terms of order $\exp(-c_1 r^2)$, i.e., for a constant $c_1 > 0$:

$$\sup_{(s,t)\in\mathbb{R}^2} \big|(\phi^{\theta_0} X)_{\rm tft} - \phi^{\theta_0}_{\rm ft}(s,t) X_{\rm ft}(s,t)\big| = O\big\{\exp\big(-c_1 r^2\big)\big\}. \tag{A.11}$$

Taking the truncated Fourier transform of both sides of (3.1), we deduce that

$$Y^{\theta_0}_{\rm tft} = \big(\phi^{\theta_0} X\big)_{\rm tft} + \delta w_{\rm tft}, \tag{A.12}$$

where

$$w_{\rm tft}(s,t) = \frac{1}{(2\pi)^2} \iint_{\mathcal{R}_r} \exp\{-(isx + ity)\} dW(x,y). \tag{A.13}$$

Put $g_1(s,t) = \phi^{\theta_0}_{\rm ft}(s,t) X_{\rm ft}(s,t)/\phi^\theta_{\rm ft}(s,t)$ and

$$g_2(s,t) = \delta \frac{w_{\rm tft}(s,t)}{\phi^\theta_{\rm ft}(s,t)} + O\Big[\exp\big\{\frac{1}{2}\theta(s^2+t^2) - c_1 r^2\big\}\Big], \tag{A.14}$$

where the big-oh term is interpreted as purely deterministic. Then, combining (A.11) and (A.12), we may write:

$$\frac{Y^{\theta_0}_{\rm tft}(s,t)}{\phi^\theta_{\rm ft}(s,t)} = g_1(s,t) + g_2(s,t), \tag{A.15}$$

uniformly in $(s,t) \in \mathbb{R}^2$.

For $j = 1, 2$ define

$$D_j = D_j^\theta(h) = 2h^{-1}(2\pi)^2 \int\int |g_j(s,t)|^2 \big(\sin^2 sh + \sin^2 th\big) I\big(s^2+t^2 \le \lambda^2\big)\, ds\, dt. \quad \text{(A.16)}$$

In view of (3.4) and (A.15),

$$\big|B_1^\theta - D_1^\theta\big| \le D_2^\theta + 2\big(D_1^\theta D_2^\theta\big)^{\frac{1}{2}}. \quad \text{(A.17)}$$

By treating separately the real and imaginary parts of w_{tft}, it may be proved from (A.13) that for a constant $C > 0$, and with probability 1,

$$\sup_{s^2+t^2 \le \lambda^2} |w_{\mathrm{tft}}(s,t)| = O\big\{(r|\log\delta|)^C\big\}.$$

Using this bound, and the fact that

$$\sup_u \left[\big\{\min\big(1,u^2\big)\big\}^{-1} \sin^2 u\right] < \infty, \quad \text{(A.18)}$$

we may deduce from (A.14) and (A.16) that for each $c_2 > 0$ and some $C > 0$, with probability 1,

$$D_2^\theta(h) = O\Big[\delta r^C \exp\big\{(\theta+c_2)\lambda^2\big\} + \exp\big\{(\theta+c_2)\lambda^2 - 2c_1 r^2\big\}\Big] \to 0, \quad \text{(A.19)}$$

uniformly in $0 < \theta < \theta_1$ for any $\theta_1 > 0$, where the limit result follows from (3.8).

Furthermore,

$$D_1^\theta(h) = 2h^{-1}(2\pi)^2 \iint |X_{\mathrm{ft}}(s,t)|^2 \exp\big\{(\theta-\theta_0)(s^2+t^2)\big\} \times \big(\sin^2 sh + \sin^2 th\big) I\big(s^2+t^2 \le \lambda^2\big)\, ds\, dt.$$

This quantity is clearly a monotone increasing function of θ. Since X is bounded and compactly supported then X_{ft} is bounded, and so by (A.18),

$$h^{-1}D_1^\theta(h) \text{ is bounded uniformly in } h, \lambda > 0 \text{ and in } 0 \le \theta \le \theta_1, \text{ for any } \theta_1 \in [0, \theta_0). \quad \text{(A.20)}$$

Suppose, on the other hand, that $\theta_2 > \theta_0$ and $0 < c_1 < c_2 < \infty$. Then, noting that $h\lambda \to \infty$, we see that if δ is sufficiently small we have for all $\theta \ge \theta_2$,

$$D_1^\theta(h) \ge 2h^{-1}(2\pi)^2 \iint |X_{\mathrm{ft}}(s,t)|^2 \exp\big\{(\theta_2-\theta_0)\big(s^2+t^2\big)\big\} \times \big(\sin^2 sh + \sin^2 th\big) I\big\{c_1 \le \big(s^2+t^2\big)h^2 \le c_2\big\}\, ds\, dt.$$

Choose c_1, c_2 so that the interval $[c_1, c_2]$ does not contain an integer multiple of π^2. Then there exists $C_1 > 0$ such that, for all sufficiently small δ, $\sin^2(sh) + \sin^2(th) \geq C_1$ for all (s,t) that satisfy $c_1 \leq (s^2+t^2)h^2 \leq c_2$. Therefore, with $C_2 = 8\pi^2 C_1$ we have:

$$D_1^\theta(h) \geq C_2 h^{-1} \exp\{(\theta_2 - \theta_0)h^{-2}c_1\} \times \iint |X_{\rm ft}(s,t)|^2 I\{c_1 \leq (s^2+t^2)h^2 \leq c_2\} ds\, dt.$$

This bound and (3.6) imply that for constants $C_3 > 0$ and C_4,

$$D_1^\theta(h) \geq C_3 h^{C_4} \exp\{(\theta_2 - \theta_0)h^{-2}c_1\} \text{ uniformly in } \theta \geq \theta_2. \tag{A.21}$$

The theorem follows from (A.17) and (A.19)−(A.21).

References

Brown, L. D., Cai, T. T., Low, M. G. and Zhang, C. H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *Ann. Statist.* **30**, 688-707.

Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24**, 2384-2398.

Carasso, A. S. (2001). Direct blind deconvolution. *SIAM J. Appl. Math.* **61**, 1980-2007.

Carasso, A. S. (2004). Singular integrals, image smoothness, and the recovery of texture in image deblurring. *SIAM J. Appl. Math.* **64**, 1749-1774.

Cornford, D., Csató, L., Evans, D. J. and Opper, M. (2004). Bayesian analysis of the scatterometer wind retrieval inverse problem: some new approaches. *J. Roy. Statist. Soc. Ser. B* **66**, 609-626.

Doucet, A., Godsill, S. J. and Robert, C. P. (2002). Marginal maximum a posteriori estimation using Markov chain Monte Carlo. *Statist. Comput.* **12**, 77-84.

Ellis, S. P. (2002). Blind deconvolution when noise is symmetric: existence and examples of solutions. *Ann. Inst. Statist. Math.* **54**, 758-767.

Erdogmus, D., Hild, K. E., Principe, J. C., Lazaro, M. and Santamaria, I. (2004). Adaptive blind deconvolution of linear channels using Renyi's entropy with Parzen window estimation. *IEEE Trans. Signal Process.* **52**, 1489-1498.

Figueiredo, M. A. T. and Nowak, R. D. (2003). An EM algorithm for wavelet-based image restoration. *IEEE Trans. Image Process.* **12**, 906-916.

Galatsanos, N. P., Mesarović, V. Z., Molina, R., Katsaggelos, A. K. and Mateos, J. (2002). Hyperparameter estimation in image restoration problems with partially known blurs. *Opt. Eng.* **41**, 1845-1854.

Gassiat, E. and Gautherat, E. (1999). Speed of convergence for the blind deconvolution of a linear system with discrete random input. *Ann. Statist.* **27**, 1684-1705.

Gonzalez, R. C. and Woods, R. E. (1992). *Digital Image Processing.* Addison-Wesley, New York.

Haario, H., Laine, M., Lehtinen, M., Saksman, E. and Tamminen, J. (2004). Markov chain Monte Carlo methods for high dimensional inversion in remote sensing. *J. Roy. Statist. Soc. Ser. B* **66**, 591-607.

Higdon, D. and Yamamoto, S. (2001). Estimation of the head sensitivity function in scanning magnetoresistance microscopy. *J. Amer. Statist. Assoc.* **96**, 785-793.

Johnstone, I. M., Kerkyacharian, G., Picard, D. and Raimondo, M. (2004). Wavelet deconvolution in a periodic setting. *J. Roy. Statist. Soc. Ser. B* **66**, 547-573.

Kundur, D. and Hatzinakos, D. (1998). A novel blind deconvolution scheme for image restoration using recursive filtering. *IEEE Trans. Signal Process.* **46**, 375-389.

Li, K. C. and Shedden, K. (2001). Monte Carlo deconvolution of digital signals guided by the inverse filter. *J. Amer. Statist. Assoc.* **96**, 1014-1021.

Likas A. C. and Galatsanos, N. P. (2004). A variational approach for Bayesian blind image deconvolution. *IEEE Trans. Signal Process.* **52**, 2222-2233.

Marron, J. S. and Tsybakov, A. B. (1995). Visual error criteria for qualitative smoothing. *J. Amer. Statist. Assoc.* **90**, 499-507.

May, K. L., Stathaki, T. and Katsaggelos, A. K. (2003). Spatially adaptive intensity bounds for image restoration. *EURASIP J. Appl. Sig. Process.* **2003**, 1167-1180.

Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24**, 2399-2430.

Poskitt, D. S., Dougancay, K. and Chung, S. H. (1999). Double-blind deconvolution: the analysis of post-synaptic currents in nerve cells. *J. Roy. Statist. Soc. Ser. B* **61**, 191-212.

Rosec, O., Boucher, J. M., Nsiri, B. and Chonavel, T. (2003). Blind marine seismic deconvolution using statistical MCMC methods. *IEEE J. Oceanic Eng.* **28**, 502-512.

Sotthivirat, S. and Fessler, J. A. (2003). Relaxed ordered-subset algorithm for penalized-likelihood image restoration. *J. Opt. Soc. Amer. Ser. A* **20**, 439-449.

Wolfe, P. J., Godsill, S. J. and Ng, W.-J. (2004). Bayesian variable selection and regularization for time-frequency surface estimation. *J. Roy. Statist. Soc. Ser. B* **66**, 575-589.

Zhang, L. Q., Amari, S. and Cichocki, A. (2001). Semiparametric model and superefficiency in blind deconvolution. *Sig. Process.* **81**, 2535-2553.

Department of Mathematics and Statistics, The University of Melbourne, Melbourne, VIC 3010, Australia.

E-mail: p.hall@stat.umn.edu.au

School of Statistics, University of Minnesota, 389 Ford Hall, 224 Church Street SE, Minneapolis, MN 55455, USA.

E-mail: qiu@stat.umn.edu

(Received May 2005; accepted March 2006)