The Annals of Probability
1990, Vol. 18, No. 2, 887–900

OPTIMAL CONVERGENCE RATES IN SIGNAL RECOVERY

BY PETER HALL[1]

Brown University

In the context of image analysis, the method of Fourier-domain processing is shown to yield restored signals of optimal quality. This confirms conjectures of statistical optimality that have been made in the past. Quality is measured in terms of convergence rates, and the influence of image smoothness on convergence rates is quantified. This influence is particularly interesting in the case of motion blur, where there is a critical degree of image smoothness (approximately four derivatives of the image) beyond which no improvement in restored image quality may be obtained by passing to smoother images. That is in marked contrast to the case of out-of-focus blur, where restored image quality is always greater for smoother images.

1. Introduction and results.

1.1. *Introduction and summary.* Let t denote a signal, to be distorted by a Toeplitz transformation and further corrupted by stochastic noise. A practical illustration of this phenomenon is image processing, where t represents the true scene, the transformation describes blurring effects such as incorrect focussing and subject motion, and the stochastic component models electronic noise at a pixel level. It is desired to recover t from the blurred, noisy recorded signal, in the sense of estimating t by a data-based signal $\tilde{t}$. A relatively old yet still commonly used method of signal recovery is Fourier-domain processing, in which one operates numerically on the Fourier transform of the recorded image to correct for the effect of blur and noise and then constructs $\tilde{t}$ by Fourier inversion. It is often alleged that this technique yields "a restoration that is optimal in a statistical sense" [Cannon and Hunt (1981)], although very little theoretical evidence is available to support this claim. In the present paper we show that, under appropriate models for the true signal and signal degradation, Fourier-domain processing yields optimal convergence rates of $E(\tilde{t} - t)^2$ to zero.

We also give a concise description of the effects on optimal image restoration of image smoothness and smoothness of the blurring transformation. The effect of image smoothness is particularly interesting, for the following reasons. In the case of out-of-focus blur, increasing smoothness of the true image always results in improved restoration. However, this is not the case for motion blur. There we prove that there is a critical degree of image smoothness, corresponding approximately to four derivatives of the true image,

Received April 1988; revised January 1989.
[1]On leave from The Australian National University.
AMS 1980 *subject classifications.* Primary 60G35; secondary 62G05.
Key words and phrases. Blur, convergence rates, image processing, image restoration, point-spread function, signal processing.

beyond which no improvements in restored image quality are obtainable by passing to smoother true images.

The method of Fourier-domain image processing may be described very simply as follows. Since the effect of the blur is to convolve the true image with a known function, called the point-spread function, then the Fourier transformation of the blurred image (neglecting the noise) is just the product of the Fourier transforms τ and χ of the true image t and the point-spread function h, respectively. If we divide this product by χ and invert, we should obtain precisely t. However, in the presence of noise, serious problems arise in the vicinity of frequencies where χ is very small or zero. There we are, in effect, dividing by zero, and so the influence of noise at such frequencies will be greatly increased. The solution is to dampen down the contribution of those frequencies when inverting the Fourier transform. In this paper, we assign zero weight to such frequencies in the inversion, although it will be clear that one could obtain identical convergence rates with a smoother compromise.

Cannon and Hunt [(1981), pages 143 and 144] give an equivalent but slightly less technical account of Fourier-domain image processing. See also Rosenfeld and Kac [(1982), Chapter 6].

Mathematical models for images, image degradation and partial Fourier inversion will be described in Sections 1.2, 1.3 and 1.4, respectively. Our results will be stated in Section 1.5 and related to existing literature in Section 1.6. Proofs will be given in Section 2.

1.2. *Image models.* To model the effect of discrete image digitisation, we take t to be a real-valued function on the d-dimensional integer lattice $\mathbf{Z}^d$ and define the discrete Fourier transforms of t by

$$\tau(\theta) \equiv \sum_j t(j) e^{ij^T\theta}, \qquad \theta \in (-\pi, \pi)^d.$$

We might think of $t(j)$ as the grey-scale intensity of the true scene at pixel j.

Our theory models the performance of image restoration procedure on grids which become increasingly fine. In effect, we treat $t(j)$ as though it were the value $f(j/n)$ of a function f, where n increases as the grid becomes finer. There is no requirement for n to be an integer, but connections with statistical theory are stronger if we think of n^d as "sample size"; see the statistical work cited in Section 1.6.

If t were continuous and $t(x) \equiv f(x/n)$, then the (continuous) Fourier transform of t would be $\tau(\theta) = n^d \phi(n\theta)$ where ϕ is the Fourier transform of f. Therefore we may encapsulate our model for images on an increasingly fine grid by indexing their discrete Fourier transform τ according to "smoothness classes" $\mathscr{E}(a, A)$, defined by

$$\mathscr{E}(a, A) \equiv \left\{t\colon |\tau(\theta)| \le A n^d (1 + \|n\theta\|)^{-a}, \theta \in (-\pi, \pi)^d\right\}, \qquad a > 1 \text{ and } A > 0. \tag{1.1}$$

Here, $\|\theta\| \equiv (\theta_1^2 + \cdots + \theta_d^2)^{1/2}$.

1.3. *Image degradation models.* Let H denote the Toeplitz transformation describing blur, and let h (a nonnegative function) be the corresponding point-spread function. The effect of the blur is to spread the light which would have been recorded at pixel j over other pixels, in such a way that a proportion $h(j+k)$ falls at pixel $j+k$. Thus, the blurred image is $b = Ht$, where

$$b(j) = \sum_k h(k)t(j+k), \qquad j \in \mathbf{Z}^d. \tag{1.2}$$

To ensure that total image intensity is preserved, we ask that $\sum h(j) = 1$, so that h is a probability density function on $\mathbf{Z}^d$.

We consider two types of point-spread function, one describing out-of-focus blur and the other, motion blur. In the out-of-focus case, take

$$h_{(1)}(j) \equiv \{(1-\rho)/(1+\rho)\}^d \rho^{|j|} \qquad j \in \mathbf{Z}^d,$$

where $0 < \rho < 1$ and $|(j_1, \dots, j_d)| = \sum_l |j_l|$. This point-spread function describes defocussing when the out-of-focus effect decreases exponentially quickly away from the origin. Let $h_{(\nu)}$ denote the ν-fold convolution of $h_{(1)}$. The index ν describes the degree of smoothness of the distortion—the higher the value of ν, the smoother is H. To incorporate our asymptotic model for the increasingly fine pixel grid we should take $\rho = 1 + cn^{-1} + o(n^{-1})$ as $n \to \infty$, where $c > 0$.

Motion blur is a one-dimensional phenomenon, to a first approximation, and so there we take $d = 1$. Let $m \geq 1$ be an integer, put

$$h_0(j) \equiv \begin{cases} (2m-1)^{-1} & \text{if } |j| \leq m, \\ 0 & \text{otherwise,} \end{cases}$$

and define b by (1.2) with $h = h_0$. The point-spread function h_0 models image degradation caused by the subject moving a distance of approximately $2m/n$ during exposure. To incorporate our asymptotic model, we should take $m \sim cn$ as $n \to \infty$, where $c > 0$. For both motion blur and out-of-focus blur, our definitions of h ensure that average image intensity is preserved, since $\sum_j h(j) = 1$.

Let the noise N be independent and identically distributed on pixels, with variance η^2. The observed, degraded image is

$$X \equiv b + N = Ht + N.$$

1.4. *Partial Fourier inversion.* Discrete Fourier transforms of the point-spread functions $h_{(\nu)}$ and h_0 are

$$\begin{aligned} \chi_{(\nu)}(\theta) &\equiv \sum_j e^{it^T\theta} h_{(\nu)}(\theta) \\ &= \sum_{l=1}^d \left\{1 + 2\rho(1-\rho)^{-2}(1-\cos\theta_l)\right\}^{-\nu}; \qquad \theta \in (-\pi, \pi)^d, \\ \chi_0(\theta) &\equiv \sum_j e^{ij\theta} h_0(\theta) \\ &= \sin\{(2m-1)\theta/2\}/\{(2m-1)\sin(\theta/2)\}, \qquad \theta \in (-\pi, \pi), \end{aligned}$$

respectively. Suppose the degraded image X is recorded within a large but finite region $\mathscr{R}$. Put

$$\zeta(\theta) \equiv \sum_{j \in \mathscr{R}} e^{it^T\theta} X(j),$$

let Θ be a subset of $(-\pi, \pi)^d$, and define

$$\hat{t}_{\mathscr{R}}(j) \equiv (2\pi)^{-d} \operatorname{Re} \int_{\Theta} e^{-ij^T\theta} \zeta(\theta) \{\chi(\theta)\}^{-1} \, d\theta, \tag{1.3}$$

where Re denotes real part. This is a partial Fourier inversion estimate of t, "partial" because Θ usually does not include all of $(-\pi, \pi)^d$. Selection of Θ is analogous to choice of smoothing parameter in statistical curve estimation, and so Θ might be called a smoothing set.

We do not treat the problem of boundary effects, and so work with

$$E\{\hat{t}(j) - t(j)\}^2 \equiv \lim_{\mathscr{R} \uparrow \mathbf{Z}^d} E\{\hat{t}_{\mathscr{R}}(j) - t(j)\}^2.$$

1.5. *Results.* It is convenient to think of noise variance η^2 as being indexed by n. Now, it is relatively easy to show that mean-square consistency is possible if and only if η and n vary together in such a manner that $n^{-d}\eta^2 \to 0$ as $n \to \infty$. Therefore we assume throughout that $\xi \equiv n^{-d}\eta^2 \to 0$. It turns out that optimal convergence rates depend on n and η^2 only through ξ.

In the results below, part (i) always refers to out-of-focus blur and part (ii) to motion blur. The set $\mathscr{C}(a, A)$ of images t is defined in (1.1), the point-spread functions $h_{(\nu)}$ and h_0 are defined as in Section 1.3, and $\alpha(\xi)$, $\lambda(\xi)$ and $s(\xi)$ are defined by

$$\alpha(\xi) \equiv \begin{cases} \xi^{(a+2)/2(2a+1)}, \\ (\xi/|\log \xi|)^{1/3}, \\ \xi^{1/3}, \end{cases} \qquad \lambda(\xi) \equiv \begin{cases} \xi^{-1/(2a+1)} & \text{if } 1 < a < 4, \\ \xi^{-1/9} & \text{if } a = 4, \\ \infty & \text{if } a > 4, \end{cases}$$

$$s(\xi) \equiv \begin{cases} \xi^{(2a-1)/(2a+1)} & \text{if } 1 < a < 4, \\ \xi^{2/3}|\log \xi|^{4/3} & \text{if } a = 4, \\ \xi^{2/3} & \text{if } a > 4, \end{cases}$$

respectively.

First we present upper bounds to convergence rates for partial Fourier inversion estimates.

THEOREM 1. (i) *Suppose $d \geq 1$, $a > d$, $v \geq 1$ and n, η vary together such that $\xi \to 0$ as $n \to \infty$. Then in the case of out-of-focus blur, if we choose the smoothing set Θ to be*

$$\Theta \equiv \left\{\theta \in (-\pi, \pi)^d : |\theta_l| \leq n^{-1}\xi^{-1/(2a+4dv-d)} \text{ for each } l,\ 1 \leq l \leq d\right\},$$

we have

$$\sup_{t\in\mathscr{C}(a,A)}\ \sup_{j\in\mathbf{Z}^d} E\{\hat{t}(j)-t(j)\}^2 = O\big(\xi^{2(a-d)/(2a+4dv-d)}\big).$$

(ii) *Suppose* $d=1$, $a>1$ *and* n,η *vary together such that* $\xi\to 0$. *Then in the case of motion blur, if we take the smoothing set* Θ *to be*

$$\Theta \equiv \Big\{\theta\in(-\pi,\pi)\colon n^{-1}\alpha(\xi)\le|\theta|\le n^{-1}\lambda(\xi) \text{ and } \big|\theta-2j\pi(2m-1)^{-1}\big| > n^{-1}\alpha(\xi)\,j^{(a+2)/2} \text{ for all } j\ne 0\Big\},$$

we have

$$\sup_{t\in\mathscr{C}(a,A)}\ \sup_{j\in\mathbf{Z}^d} E\{\hat{t}(j)-t(j)\}^2 = O\{s(\xi)\}.$$

Next we present lower bounds to convergence rates of arbitrary estimators. These results show that the upper bounds described in Theorem 1 are best possible with respect to the image classes $\mathscr{C}(a,A)$. Let $\mathscr{T}$ denote the class of all possible d-dimensional image restoration methods. Assume in Theorem 2 that N is Gaussian.

THEOREM 2. (i) *Under the conditions of part* (i) *of Theorem* 1,

$$\liminf_{\xi\to 0} \xi^{-2(a-d)/(2a+4dv-d)} \inf_{\tilde{t}\in T}\ \sup_{t\in\mathscr{C}(a,A)} E\{\tilde{t}(0)-t(0)\}^2 > 0.$$

(ii) *Under the conditions of part* (ii) *of Theorem* 1,

$$\liminf_{\xi\to 0} s(\xi)^{-1} \inf_{\tilde{t}\in T}\ \sup_{t\in\mathscr{C}(a,A)} E\{\tilde{t}(0)-t(0)\}^2 > 0.$$

REMARKS. (a) The "optimal" smoothing sets Θ in Theorem 1 are designed to prevent the function $\{\chi(\theta)\}^{-1}$, appearing in the definition (1.3) of the estimator $\hat{t}_{\mathscr{R}}$, from becoming too large. In the case of out-of-focus blur, the function $\{\chi(\theta)\}^{-1}$ only takes large values when θ is some distance from the origin and therefore, Θ is defined so as to avoid giving emphasis to the tails. However, in the case of motion blur, $\{\chi(\theta)\}^{-1}$ equals infinity whenever θ is an integer multiple of $2\pi/(2m-1)$, and so Θ is designed to avoid those points as well as θ-values in the tails.

(b) In the case of out-of-focus blur, observe that

$$\xi^{2(a-d)/(2a+4dv-d)} \to \xi$$

as $a\to\infty$. This means that, for very smooth images [i.e., images from $\mathscr{C}(a,A)$ for large a], the optimal convergence rate is approximately ξ. The rate steadily improves as a increases.

(c) The property evinced in Remark (b) does not hold for motion blur. There, the optimal convergence rate for all images with $a>4$ is $\xi^{2/3}$. This may be explained as follows. In the case of out-of-focus blur, the main difficulties in restoring images are due to high-frequency components, because it is only for θ-values some distance from zero that the function $\{\chi(\theta)\}^{-1}$ is large. When images are smooth, relatively little information is present in high frequencies and so restoration is relatively simple. However, in the case of motion blur there are many *low* frequencies where restoration is difficult—these frequencies are those θ-values for which $\chi(\theta) = 0$. Of course, they persist even if the image is quite smooth, and so there is a point (corresponding to $a = 4$, i.e., to approximately four derivatives of the true image) beyond which no improvement in resolution is obtainable by passing to smoother images.

(d) In the case of out-of-focus blur, the convergence rate becomes worse as ν increases or, equivalently, as the point-spread function becomes less smooth. This may be explained much as in (c) above—for smooth point-spread functions, $\{\chi(\theta)\}^{-1}$ is larger and so image restoration is correspondingly more difficult.

1.6. *Alternative models and related literature.* As we pointed out at the very beginning, our aim in this paper has been to verify claims that have been made concerning the statistical optimality of Fourier-domain image processing. Our choice of models for images and image degradation has reflected this goal. One can draw somewhat different conclusions under different models. For example, one could adopt a Bayesian viewpoint, place a prior on the class of images and compute the Bayes risk and Bayes estimate. If the prior is Gaussian, then an exact optimal estimator, the Wiener filter, results. This solution is classical.

Our discrete model and our assumption of additive, uncorrelated noise are good approximations to reality in the case of electronic recording devices of the still-video type. However, if the image is recorded by analogue means and if digital processing takes place on a pixel grid much finer than the noise structure, then it is necessary to use continuum models for both image and noise, and it is essential to allow a correlation structure for noise. Continuous models with correlated noise will be treated by Hall and Koch (1989).

Even within the confines of the basic models treated in this paper, alternative approaches are possible. For example, instead of imposing a condition directly on the size of the Fourier transform τ, such as $|\tau(\theta)| \le An^d(1 + \|n\theta\|)^{-a}$, one could place a bound on the integral of $|\tau|^2$, such as

$$\int \left| n^{-d}\tau(\theta/n) \right|^2 (1 + \|\theta\|)^{2b}\, d\theta \le A.$$

The analogue of this approach in statistical curve estimation would be to ask that the function g being estimated satisfy $\int |g^{(k)}|^2 \le A$ instead of $|g^{(k)}| \le A$, where $g^{(k)}$ denotes the kth derivative of g. The square integral type of constraint is usually favoured in orthogonal series curve estimation. In other contexts Kuks and Olman (1972), Li (1982) and Speckman (1985) give formulae for the exact linear minimax estimator of a linear functional when the function class is determined by a square integral constraint. See also Sacks and Ylvisaker (1978, 1981) and Donoho and Liu (1987).

Our emphasis has been on rates of convergence, not on the precise manner of convergence. That is, if $E(\tilde{t} - t)^2 \sim Cn^{-c}$ then only the value of c, not that of C, is of interest to us. Again this is a feature of our decision to focus on the problem of Fourier-domain image processing, which is optimal in the sense that it yields the best c but not optimal in the sense of giving the best C.

In the explicit context of image processing, some authors have studied the problems of distinguishing between a single object and two closely-spaced objects, of image information and of diffraction-limited resolution [e.g., Barnes (1966); Buck and Gastincic (1967); Frieden (1970); Cunningham, Laramore and Barrett (1976); Gonsalves (1976)]. Hall (1987, 1988) has described performance of restoration methods from the viewpoint of consistency. In the present paper we focus instead on the *rate* of consistency, using image models entirely different from those considered earlier by Hall. Our present approach is most closely related to work on statistical function estimation, where the emphasis is on rates of convergence of function estimates. See for example, Farrell (1972), Wahba (1975) or Stone (1980, 1982, 1983).

2. Proofs. Throughout, the symbols $C, C_1, C_2, \dots$ denote positive generic constants not depending on j, n or θ.

PROOF OF THEOREM 1. The following results will be used to prove both parts of the theorem: Mean-squared error between $\hat{t}(j)$ and $t(j)$ is given by

$$E\{\hat{t}(j) - t(j)\}^2 = B(j)^2 + V, \tag{2.1}$$

where

$$B(j) \equiv (2\pi)^{-d} \operatorname{Re} \int_\Theta e^{-ij^T\theta}\tau(\theta)\, d\theta - t(j) \tag{2.2}$$

denotes bias,

$$V \equiv \eta^2 (2\pi)^{-d} \int_\Theta \chi(\theta)^{-2}\, d\theta \tag{2.3}$$

equals variance and $\tilde{\Theta}$ is the complement of Θ in $(-\pi, \pi)^d$.

(i) In the case of out-of-focus blur, take

$$\Theta \equiv \left\{\theta \in (-\pi, \pi)^d \colon \|\theta_l\| \le \varepsilon \text{ for } 1 \le l \le d\right\}.$$

Then for $t \in \mathscr{C}(a, A)$ we see from (2.2) that

$$\begin{aligned}
(2\pi)^d |B(j)| &\le \sum_{l=1}^{d} \int_{|\theta_l| > \varepsilon;\, |\theta_m| \le \pi,\ \text{all } m} |\tau(\theta)|\, d\theta \\
&\le dAn^d \int_{|\theta_1| > \varepsilon;\, |\theta_m| \le \pi,\ \text{all } m} (1 + \|n\theta\|)^{-a}\, d\theta \\
&\le 2^d dA \int_{\varepsilon n}^{\infty} d\theta_1 \int_{(0,\infty)^{d-1}} (1 + \theta_1 + \cdots + \theta_d)^{-a}\, d\theta_2 \cdots d\theta_d \\
&\le C(a, A, d)(\varepsilon n)^{d-a}.
\end{aligned} \tag{2.4}$$

Furthermore, since $\chi(\theta) = \Pi_l\{1 + 2\rho(1-\rho)^{-2}(1-\cos\theta_l)\}^{-v}$, we see from (2.3) that

$$\begin{aligned} V &= \pi^{-d}\eta^2\left[\int_0^{\varepsilon}\{1 + 2\rho(1-\rho)^{-2}(1-\cos x)\}^{2v}\,dx\right]^d \\ &\le C_1(c,d)\eta^2\left\{\int_0^{\varepsilon}(1+n^2x^2)^{2v}\,dx\right\}^d \\ &\le C_2(c,d)\eta^2\{n^{-1}(\varepsilon n)^{4v+1}\}^d \end{aligned} \tag{2.5}$$

if $\varepsilon n \ge$ const.

Combining (2.1), (2.4) and (2.5), we deduce that

$$\begin{aligned} E\{\hat{t}(j) - t(j)\}^2 &= B(j)^2 + V \\ &\le C(a,A,c,d)\left[(\varepsilon n)^{2(d-a)} + \eta^2\{n^{-1}(\varepsilon n)^{4v+1}\}^d\right]. \end{aligned}$$

Taking $\varepsilon n = (n^d\eta^{-2})^{1/(2a+4dv-d)}$, we obtain

$$E\{\hat{t}(j) - t(j)\}^2 \le 2C(a,A,c,d)(n^{-d}\eta^2)^{2(a-d)/(2a+4dv-d)}.$$

(ii) In the case of motion blur, take the smoothing set Θ to be the one given in Theorem 1. Then for $t \in \mathscr{C}(a,A)$ we see from (2.2) that

$$\begin{aligned} 2\pi|B(j)| &\le \int_{\tilde{\Theta}}|\tau(\theta)|\,d\theta \le A\int_{n\tilde{\Theta}}(1+|\theta|)^{-1}\,d\theta \\ &\le C_1\sum_{|j|\le\lambda}\int_{|\theta-\{2nj\pi/(2m-1)\}|\le\varepsilon_j}(1+|\theta|)^{-a}\,d\theta \\ &\quad + C_1\int_{|\theta|>n\varepsilon}(1+|\theta|)^{-a}\,d\theta \\ &\le C_2\left\{\sum_{1\le j\le\lambda}j^{-a}\varepsilon_j + (n\varepsilon)^{-a+1}\right\}. \end{aligned} \tag{2.6}$$

Since $\chi(\theta) = \sin\{(2m-1)\theta/2\}/\{(2m-1)\sin(\theta/2)\}$, then by (2.3),

$$\begin{aligned} V &\le \pi^{-1}(2m-1)^{-1}\eta^2\int_{(2m-1)\Theta/2}\theta^2(\sin\theta)^{-2}\,d\theta \\ &\le C_1n^{-1}\eta^2\sum_{|j|\le\lambda}j^2\int_{|\theta-j\pi|>(2m-1)\varepsilon_j/2n\pi}(\sin\theta)^{-2}\,d\theta \\ &\le C_2n^{-1}\eta^2\sum_{1\le j\le\lambda}j^2\varepsilon_j^{-1}. \end{aligned} \tag{2.7}$$

Combining (2.1), (2.6), and (2.7), we conclude that

$$\begin{aligned} T_1 &\equiv E\{\hat{t}(j) - t(j)\}^2 \\ &\le T_2 \\ &\equiv C_3(a,A)\left\{\left(\sum_{1\le j\le\lambda}j^{-a}\varepsilon_j + \lambda^{-a+1}\right)^2 + \xi\sum_{1\le j\le\lambda}j^2\varepsilon_j^{-1}\right\}, \end{aligned} \tag{2.8}$$

where $\xi \equiv n^{-1}\eta^2$.

Define

$$S(\lambda) \equiv \begin{cases} \lambda^{(4-a)/2} & \text{if } 1 < a < 4, \\ \log \lambda & \text{if } a = 4, \\ 1 & \text{if } a > 4, \end{cases} \tag{2.9}$$

and $\varepsilon_j \equiv \alpha j^{(a+2)/2}$, where $\alpha = \alpha(\xi)$ will be chosen shortly. Then

$$\sum_{1 \le j \le \lambda} j^{-a} \varepsilon_j = \alpha \sum_{1 \le j \le \lambda} j^{(2-a)/2} \le C_4 \alpha S(\lambda),$$

$$\sum_{1 \le j \le \lambda} j^2 \varepsilon_j^{-1} = \alpha^{-1} \sum_{1 \le j \le \lambda} j^{(2-a)/2} \le C_4 \alpha^{-1} S(\lambda).$$

Hence,

$$T_2 \le C_5(a, A)\{\xi \alpha^{-1} S(\lambda) + \alpha^2 S(\lambda)^2 + \lambda^{-2(a-1)}\}. \tag{2.10}$$

If $1 < a < 4$, take $\alpha = \xi^{(a+2)/2(2a+1)}$ and $\lambda = \xi^{-1/(2a+1)}$ in (2.8) and (2.10), to deduce that

$$T_1 \le T_2 \le C_6(a, A) \xi^{2(a-1)/(2a+1)}.$$

If $a = 4$, take $\alpha = (\xi / |\log \xi|)^{1/3}$ and $\lambda = \xi^{-1/9}$, obtaining

$$T_1 \le T_2 \le C_6(a, A) \xi^{2/3} |\log \xi|^{4/3}.$$

If $a > 4$, take $\alpha = \xi^{1/3}$ and $\lambda = \infty$, obtaining

$$T_1 \le T_2 \le C_6(a, A) \xi^{2/3}.$$

□

Proof of Theorem 2. The following argument is used to prove both parts of the theorem. Let t_1 and t_2 be two images from $\mathscr{C}(a, A)$, let $b_r \equiv H t_r$ denote the blur of t_r and let $X = b + N$ where b is either b_1 or b_2. Likelihood is proportional to

$$\exp\left[-\tfrac{1}{2} \sum_j \{X(j) - b_r(j)\}^2\right], \tag{2.11}$$

assuming t_r is the true image. The likelihood-ratio, or Bayes, rule for choosing between t_1 and t_2 is to decide in favor of t_2 if and only if

$$\sum_j \{X(j) - b_2(j)\}^2 \le \sum_j \{X(j) - b_1(j)\}^2.$$

The probability of incorrectly deciding in favor of t_2, given that the true image is t_1, equals

$$\begin{aligned} p &\equiv P\left[\sum_j \{b_1(j) + N(j) - b_2(j)\}^2 \le \sum_j \{b_1(j) + N(j) - b_1(j)\}^2\right] \\ &= P\left[2Z \ge \left\{\eta^{-2} \sum_j (b_1(j) - b_2(j))^2\right\}^{1/2}\right], \end{aligned} \tag{2.12}$$

where Z is Normal $N(0, 1)$. If $\tilde{t}$ is an estimator of t, define $D = 1$ if

$|\tilde{t}(0) - t_1(0)| \le |\tilde{t}(0) - t_2(0)|$ and put $D = 2$ otherwise. Then D is a decision rule for choosing between t_1 and t_2. Let D_0 be the Bayes rule defined above. By the Neyman–Pearson lemma,

$$P_{t_1}(D = 2) + P_{t_2}(D = 1) \ge P_{t_1}(D_0 = 2) + P_{t_2}(D_0 = 1) \ge P_{t_1}(D_0 = 2) = p.$$

Therefore,

$$\begin{aligned} 2 \max_{t=t_1, t_2} E_t\{\tilde{t}(0) - t(0)\}^2 &\ge E_{t_1}\{\tilde{t}(0) - t_1(0)\}^2 + E_{t_2}\{\tilde{t}(0) - t_2(0)\}^2 \\ &\ge \tfrac{1}{4}\{t_1(0) - t_2(0)\}^2\{P_{t_1}(D = 2) + P_{t_2}(D = 1)\} \\ &\ge \tfrac{1}{4}p\{t_1(0) - t_2(0)\}^2. \end{aligned} \tag{2.13}$$

Suppose we prove that t_1, t_2 (depending on n and η) may be chosen from $\mathscr{C}(a, A)$ such that $p \ge C_1$ and

$$\{t_1(0) - t_2(0)\}^2 \ge C_2 s, \tag{2.14}$$

where C_1 and C_2 are fixed constants. Then it will follow from (2.13) that

$$\sup_{t \in \mathscr{C}} E_t\{\tilde{t}(0) - t(0)\}^2 \ge \tfrac{1}{8}C_1 C_2 s,$$

as claimed by Theorem 2. In view of (2.12), $p \ge C_1$ with $C_1 \equiv P(Z > \frac{1}{2}C_3^{1/2})$ will follow if we show that

$$\eta^{-2} \sum_j \{b_1(j) - b_2(j)\}^2 \le C_3. \tag{2.15}$$

Therefore we shall establish (2.14) and (2.15).

Summations in formulae such as (2.11) should, strictly speaking, be taken over a bounded set so as to ensure finiteness. However, the left-hand side of (2.15) is finite even for an infinite summation. This minor problem evaporates if we work with the data $\{X(j),\ j \in \mathscr{S}\}$, where $\mathscr{S}$ is a large but bounded set, until we get to (2.13) and at that point let $\mathscr{S} \to \mathbf{Z}^d$.

We shall in fact take $t_1 \equiv 0$, the identically zero image, and t_2 to be an image whose Fourier transform is $n^d \delta(n\theta)$, for an appropriate real-valued, symmetric function δ.

(i) Let $\lambda > 0$ be a positive function of n and η, to be specified shortly. Put

$$\delta(\theta) \equiv \begin{cases} A(1 + \|\theta\|)^{-a} & \text{if } |\theta_l| > \lambda \text{ for each } l, \\ 0 & \text{otherwise,} \end{cases}$$

let $t_1 \equiv 0$ and let t_2 be the image whose Fourier transform is $n^d \delta(n\theta)$. Then

$$|t_1(0) - t_2(0)| = \pi^{-d} A \int_\lambda^\infty \cdots \int_\lambda^\infty (1 + \|\theta\|)^{-a}\, d\theta = C\lambda^{d-a}$$

and

$$\sum_j \{b_1(j) - b_2(j)\}^2 = \int |\tau\chi|^2$$

$$\leq C_2 n^d \int_{|\theta_l| > \lambda,\ \text{all } l} (1 + \|\theta\|)^{-2a} \prod_{l=1}^{d} (1 + |\theta_l|)^{-4v}\, d\theta$$

$$\leq C_3 n^d \lambda^{-(2a + 4dv - d)}$$

if we take $\Delta = \lambda^{-2(a-d)}$ and

$$\lambda = \left(n^d \eta^{-2}\right)^{1/(2a+4dv-d)}.$$

(ii) Let $t_1 \equiv 0$ and t_2 be the image whose real-valued, symmetric, nonnegative Fourier transform is $n\delta(n\theta)$. Since

$$|t_1(0) - t_2(0)| = \pi^{-1} \int_0^{n\pi} \delta(\theta)\, d\theta$$

and

$$\sum_j \{b_1(j) - b_2(j)\}^2 = \pi^{-1} n \int_0^{n\pi} \delta(\theta)^2 \chi(\theta/n)^2\, d\theta,$$

then we must prove that a sequence of δ's may be chosen so that

$$0 \leq \delta(\theta) \leq A(1 + \theta)^{-a} \quad \text{for } \theta > 0,$$

$$s^{-1/2} \int_0^{n\pi} \delta(\theta)\, d\theta \geq C_1, \tag{2.16}$$

$$\xi^{-1} \int_0^{n\pi} \delta(\theta)^2 \chi(\theta/n)^2\, d\theta \leq C_2, \tag{2.17}$$

where $\xi \equiv n^{-1}\eta^2$ and $s = s(\xi)$ is as defined in Section 1.

Remembering that $\chi(\theta) = \sin\{(2m-1)\theta/2\}/\{(2m-1)\sin(\theta/2)\}$, we have

$$\int_0^{n\pi} \delta(\theta)^2 \chi(\theta/n)^2\, d\theta \leq C_3 \int_0^{\infty} \theta^{-2} \sin^2\{(2m-1)\theta/2n\} \delta(\theta)^2\, d\theta. \tag{2.18}$$

For ease of notation we shall replace $(2m-1)\theta/2n$ by θ on the right-hand side and seek a δ vanishing on $(0, 1)$ and satisfying $0 \leq \delta(\theta) \leq A(1 + \theta)^{-a}$, (2.16) and

$$\xi^{-1} \int_1^{\infty} \theta^{-2} (\sin^2 \theta) \delta(\theta)^2\, d\theta \leq C_4,$$

the latter instead of (2.17).

Put

$$I_1(\delta) \equiv \int_1^{n\pi} \delta(\theta)\, d\theta, \qquad I_2(\delta) \equiv \int_1^{\infty} \theta^{-2} (\sin^2 \theta) \delta(\theta)^2\, d\theta.$$

We wish to prove that the maximum of $I_1(\delta)$ over δ's which satisfy $0 \le \theta^a\delta(\theta) \le 1$ and $I_2(\delta) = \xi$, is of order at least $s^{1/2}$ as $\xi \to 0$. Now,

$$I_1(\delta) \ge \sum_{j=1}^{n-1} \{(j+\tfrac{1}{2})\pi\}^{-a} \int_{j\pi-(\pi/2)}^{j\pi+(\pi/2)} \theta^a\delta(\theta)\, d\theta,$$

$$I_2(\delta) \le \sum_{j=1}^{n-1} \{(j-\tfrac{1}{2})\pi\}^{-2(a+1)} \int_{j\pi-(\pi/2)}^{j\pi+(\pi/2)} (\theta - j\pi)^2 \{\theta^a\delta(\theta)\}^2\, d\theta.$$

Let $\psi(\varepsilon)$ denote the maximum of

$$\int_0^{\pi/2} \Delta(\theta)\, d\theta$$

over all functions Δ which satisfy $0 \le \Delta(\theta) \le 1$ and

$$\int_0^{\pi/2} \theta^2 \Delta(\theta)^2\, d\theta = \varepsilon.$$

It suffices to prove that the maximum of

$$\sum_{j=1}^{n-1} j^{-a}\psi(\varepsilon_j)$$

over all nonnegative ε_j's satisfying

$$\sum_{j=1}^{\infty} j^{-2(a+1)}\varepsilon_j = \xi, \tag{2.19}$$

is of order at least $s^{1/2}$ as $\xi \to 0$.

The function $\psi(\varepsilon)$ is of order at least $\varepsilon^{1/3}$ as $\varepsilon \to 0$, as may be seen by taking $\Delta(\theta) \equiv 1$ for $0 < \theta \le \varepsilon^{1/3}$ and $\Delta(\theta) \equiv 0$ for $\varepsilon^{1/3} < \theta < 1$. Therefore, it suffices to show that the maximum of

$$\sum_{j=1}^{n-1} j^{-a}\varepsilon_j^{1/3}$$

over all nonnegative ε_j's satisfying (2.19), is of order at least $s^{1/2}$.

Take $\varepsilon_j = (j/l)^{3(a+2)/2}$ for $1 \le j \le l$ and $\varepsilon_j = 1$ for $j > l$, where $l \le n/2$. Then, defining $S(\lambda)$ as at (2.9),

$$\begin{aligned} T_1 \equiv \sum_{j=1}^{\infty} j^{-2(a+1)}\varepsilon_j &= l^{-3(a+2)/2} \sum_{j=1}^{l} j^{(2-a)/2} + \sum_{j=l+1}^{\infty} j^{-2(a+1)} \\ &\le C_1\{l^{-3(a+2)/2}S(l) + l^{-(2a+1)}\}, \end{aligned}$$

$$\begin{aligned} T_2 \equiv \sum_{j=1}^{n-1} j^{-a}\varepsilon_j^{1/3} &= l^{-(a+2)/2} \sum_{j=1}^{l} j^{(2-a)/2} + \sum_{j=l+1}^{n-1} j^{-a} \\ &\ge C_2\{l^{-(a+2)/2}S(l) + l^{-(a-1)}\}. \end{aligned}$$

If $1 < a < 4$ then $T_1 \le 2C_1 l^{-(2a+1)}$ and $T_2 \ge C_2 l^{-(a-1)}$. Hence, taking l to be of size $\xi^{-1/(2a+1)}$, we deduce that the maximum of T_2 subject to $T_1 = \xi$ is at least of order $\xi^{(a-1)/(2a+1)} = s^{1/2}$. However, this argument fails if $\xi^{-1/(2a+1)} > n$, for then it violates our assumption that $l \le \frac{1}{2}n$. Should $\xi^{-1/(2a+1)} > n$ or, equivalently, $n^{-1} > \xi^{1/(2a+1)}$, take l to equal the integer part of $\frac{1}{2}n$, from which it follows that

$$T_2 \ge C_2 n^{-(a-1)} \ge C_3 \xi^{(a-1)/(2a+1)} = C_3 s^{1/2}.$$

If $a = 4$ then $T_1 \le 2C_1 l^{-9} \log l$ and $T_2 \ge C_2 l^{-3} \log l$. Hence, taking l to be of size $(\xi^{-1}|\log \xi|)^{1/9}$, we deduce that the maximum of T_2 subject to $T_1 = \xi$ is at least of order $\xi^{1/3}|\log \xi|^{2/3} = s^{1/2}$. This argument fails if $(\xi^{-1}|\log \xi|)^{1/9} > n$, and in that case we take l to be the integer part of $\frac{1}{2}n$, implying that

$$T_2 \ge C_3 n^{-3} \log n = C_3\{n^{-3}(\log n)^{1/3}\}(\log n)^{2/3} \ge C_4 \xi^{1/3}|\log \xi|^{2/3} = C_4 s^{1/2}.$$

If $a > 4$ then $T_1 \le 2C_1 l^{-3(a+2)/2}$ and $T_2 \ge C_2 l^{-(a+2)/2}$. Hence, taking l to be of size $\xi^{-2/3(a+2)}$, we deduce that the maximum of T_2 subject to $T_1 = \xi$ is at least of order $\xi^{1/3}$. This argument fails if $\xi^{-2/3(a+2)} > n$, in which case we take l to be the integer part of $\frac{1}{2}n$, implying that

$$T_2 \ge C_3 n^{-(a+2)/2} \ge C_4 \xi^{1/3}. \qquad \square$$

Acknowledgments. The paper has benefited from constructive comments of a referee and an Associate Editor.

REFERENCES

BARNES, C. W. (1966). Object restoration in a diffraction-limited imaging system. *J. Opt. Soc. Amer.* **56** 575–578.

BUCK, E. J. and GASTINCIC, J. J. (1967). Resolution limitations of a finite aperture. *IEEE Trans. Antennas and Propagation* **AP-15** 376–381.

CANNON, T. M. and HUNT, B. R. (1981). Image processing by computer. *Sci. American* **245** (4) 214–225.

CUNNINGHAM, D. R., LARAMORE, R. D. and BARRETT, E. (1976). Detection in image-dependent noise. *IEEE Trans. Inform. Theory* **IT-22** 603–610.

DONOHO, D. L. and LIU, R. C. (1987). Geometrizing rates of convergence. (Unpublished).

FARRELL, R. H. (1972). On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Statist.* **43** 170–180.

FRIEDEN, B. R. (1970). Information and the restorability of images. *J. Opt. Soc. Amer.* **60** 575–577.

GONSALVES, R. A. (1976). Cramér–Rao bounds on mensuration errors. *Appl. Optics* **15** 1270–1275.

HALL, P. (1987). On the amount of detail that can be recovered from a degraded signal. *Adv. in Appl. Probab.* **19** 371–395.

HALL, P. (1988). On the processing of a motion-blurred image. *SIAM J. Appl. Math.* **47** 441–453.

HALL, P. and KOCH, I. (1989). On continuous image models and image analysis in the presence of correlated noise. *Adv. in Appl. Prob.* To appear.

KUKS, J. and OLMAN, V. (1972). Linear minimax estimation of regression coefficients. *Eesti NSV Tead. Akad. Toimetised Füüs.-Mat.* **21** 66–72. (In Russian.)

LI, K.-C. (1982). Minimaxity of the method of regularization on stochastic processes. *Ann. Statist.* **10** 937–942.

ROSENFELD, A. and KAC, A. I. (1982). *Digital Picture Processing*. Academic, New York.

SACKS, J. and YLVISAKER, D. (1978). Linear estimation for approximately linear models. *Ann. Statist.* **6** 1122–1137.

SACKS, J. and YLVISAKER, D. (1981). Asymptotically optimum estimation for density estimation at a point. *Ann. Statist.* **9** 334–346.

SPECKMAN, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.* **13** 970–983.

STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

STONE, C. J. (1983). Optimal uniform rate of convergence for nonparametric estimators of a density function or its derivatives. In *Recent Advances in Statistics* (*papers in honor of H. Chernoff*) (H. Rizvi, J. Rustagi and D. Siegmund, eds.) 393–406. Academic, New York.

WAHBA, G. (1975). Optimal convergence properties of variable knot, kernel and orthogonal series methods for density estimation. *Ann. Statist.* **3** 15–29.

DEPARTMENT OF STATISTICS
AUSTRALIAN NATIONAL UNIVERSITY
G.P.O. BOX 4
CANBERRA A.C.T. 2601
AUSTRALIA