A COMPARISON OF SMOOTHING PARAMETER CHOICES IN IMAGE RESTORATION

J. W. Kay
Department of Statistics
University of Glasgow
Scotland

ABSTRACT

We discuss the choice of smoothing parameter for the problem of restoring two-dimensional pixellated images which have been degraded by blur and noise. Some asymptotic comparisons are presented and discussed. The methods are applied to a number of test images.

Key words: Image Restoration; Deconvolution; Noise-Removal; Regularization; Optimal Smoothing; Singular-Value Decomposition.

1. Introduction

We consider the problem of restoring two-dimensional pixellated images which have been degraded by a combination of blur and noise processes. We employ the method of regularization in the solution of this problem; see Titterington (1985), Poggio et al. (1985) and Davies and Andersson (1986) for further details. In mathematical terms, this means that we choose as our restoration that f which solves the optimization problem

$$\min_f(\Delta(f,g) + \lambda\Phi(f)\} \tag{1.1}$$

for a given λ, where $\Delta(f,g)$ is a measure of the discrepancy between the "true scene" f and the degraded scene $g, \Phi(f)$ is a roughness penalty and λ is the *smoothing parameter.* The purpose of the ensuing discussion is to compare several ways of choosing λ.

We assume an additive, linear structure for the observed image g i.e.

$$g = \mathbf{H}f + \varepsilon \tag{1.2}$$

where $\mathbf{H}$ is a space-invariant point-spread matrix, ε represents additive, signal-independent, noise and $\operatorname{cov}(\varepsilon) = \sigma^2 I$. (all matrices are of order $n \times n$, where $n = p^2$ and the image is of size $p \times p$)

For the purpose of this discussion we take

$$\Delta(f,g) = \frac{1}{n}||g - \mathbf{H}f||^2 \tag{1.3}$$

and

$$\Phi(f) = f^T C f \tag{1.4}$$

where C is a positive semi-definite smoothing matrix. It follows from an argument due to Bates and Wahba (1983; Section 4) that the optimization problem (1.1), with the specifications (1.2) and (1.3), may be written as

$$\min_\beta\{\frac{1}{n}|||y - \mathbf{M}\beta||^2 + \lambda||\beta||^2\} \tag{1.5}$$

where the vectors y and β are linear transformations of g and f, respectively, and $\mathbf{M}$ is of order $r \times r$, where $r = \operatorname{rank} C(= O(n))$.

The solution of the optimization problem (1.4) is then

$$\widehat{\beta}(\lambda) = (\mathbf{M}^{\mathbf{T}}\mathbf{M} + n\lambda\,\mathbf{I})^{-1}\mathbf{M}^{\mathbf{T}}y \tag{1.6}$$

We now develop some notation which will be useful in the sequel.

$$\begin{aligned}
\widehat{y}(\lambda) &= \mathbf{M}\widehat{\beta}(\lambda) \\
\mathbf{RSS}(\lambda) &= ||y - \widehat{y}(\lambda)||^2 \\
\mathbf{K}(\lambda) &= \mathbf{M}(\mathbf{M}^{\mathbf{T}}\mathbf{M} + n\lambda\mathbf{I})^{-1}\mathbf{M}^{\mathbf{T}} \\
\mathbf{EDF}(\lambda) &= \operatorname{tr}\{\mathbf{I} - \mathbf{K}(\lambda)\} = n - \operatorname{tr}\mathbf{K}(\lambda).
\end{aligned}$$

In Section 2 we define various choices of smoothing parameter while in Section 3 we present the problem in a canonical form which results in the derivation of simple algebraic forms for the various criteria involved. In Section 4 we discuss some asymptotic results; see Kay (1988b) for proofs and further details. Finally in Section 5 we present some numerical experiments using a number of test images; for further details, see Kay (1988a).

2. Smoothing parameter choices

Perhaps the best known method for choosing the smoothing parameter in inverse problems is the *Chi-squared Choice* defined by

$$\mathbf{RSS}(\lambda) = n\sigma^2 \tag{2.1}$$

This is also termed the Discrepancy Principle; see Groetsch (1984).

However this method has been generally recognized to impose too much smoothing; see Hall and Titterington (1986, 1987) who proposed the *Equivalent Degrees of Freedom Choice* (proposed independently by Turchin (1967, 1968)). Here λ is defined as the solution of the equation

$$\mathbf{RSS}(\lambda) = \sigma^2 \mathbf{EDF}(\lambda) \tag{2.2}$$

While both of these methods require the availability of an estimate for σ^2, the *Generalized Cross-Validation Choice*, due to Golub et al. (1979), is entirely data-based; it is defined as the minimizer of

$$\mathbf{GCV}(\lambda) = \mathbf{RSS}(\lambda)/\{\mathbf{EDF}(\lambda)\}^2 \tag{2.3}$$

As a means of comparison we also consider two "optimal" choices namely the Mean Squared Errors for Prediction and Estimation. The *Mean Squared Prediction Error Choice*, $\lambda_{\mathbf{TP}}$, is defined as the solution of

$$\min_{\lambda} \mathbf{E}\{||\mathbf{M}\widehat{\beta}(\lambda) - \mathbf{M}\beta||^2\} \tag{2.4}$$

whereas the *Mean Squared Estimation Error Choice*, $\lambda_{\mathbf{TE}}$, is defined as the solution of

$$\min_{\lambda} \mathbf{E}\{||\widehat{\beta}(\lambda) - \beta||^2\} \tag{2.5}$$

The choices defined in (2.1)–(2.5) will be used in some numerical experiments described in Section 5. In order to develop an asymptotic comparison of the choices we now consider deterministic versions of (2.1), (2.2) and (2.3) and define $\lambda_{\mathbf{CHI}}$ as the solution of

$$\mathbf{E}(\mathbf{RSS}(\lambda)) = n\sigma^2 \tag{2.6}$$

$\lambda_{\mathbf{EDF}}$ as the solution of

$$\mathbf{E}(\mathbf{RSS}(\lambda)) = \sigma^2 \mathbf{EDF}(\lambda) \tag{2.7}$$

and $\lambda_{\mathbf{GCV}}$ as the minimizer of

$$\mathbf{E}(\mathbf{RSS})(\lambda)/\{\mathbf{EDF}(\lambda)\}^2 \tag{2.8}$$

3. Canonical characteristic

We employ a Singular-Value Decomposition

$$\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}$$

where $\mathbf{U}, \mathbf{V}$ are orthogonal; $D = \operatorname{diag}\{m_1, m_2, \ldots, m_r\}$ and $\{m_i^2\}$ are the eigenvalues of $\mathbf{M}^{\mathrm{T}}\mathbf{M}$.

We denote by v_i, y_j and h_k, the ith column of $\mathbf{V}$, the jth element of $\mathbf{U}^{\mathrm{T}}y$ and the kth element of $\mathbf{V}^{\mathrm{T}}\beta$, respectively $(i, j, k = 1, \ldots, r)$.

It is then straightforward to deduce the following simple algebraic forms.

$$\widehat{\beta}(\lambda) = \sum_{i=1}^{r} \frac{m_i g_i}{(m_i^2 + n\lambda)} v_i \tag{3.1}$$

$$\mathrm{RSS}(\lambda) = \sum_{i=1}^{r} \left\{ \frac{n\lambda}{m_i^2 + n\lambda} \right\}^2 y_i^2 \tag{3.2}$$

$$\mathrm{EDF}(\lambda) = \sum_{i=1}^{r} \frac{n\lambda}{m_i^2 + n\lambda} \tag{3.3}$$

$$\mathrm{E}(\mathrm{RSS}(\lambda) = n^2\lambda^2 \sum_{i=1}^{r} \frac{m_i^2 h_i^2}{(m_i^2 + n\lambda)^2} + \sigma^2 n^2 \lambda^2 \sum_{i=1}^{r} (m_i^2 + n\lambda)^{-2} \tag{3.4}$$

$$\mathrm{E}\{||\mathbf{M}\widehat{\beta}(\lambda) - \mathbf{M}\beta||^2\} = n^2\lambda^2 \sum_{i=1}^{r} \frac{m_i^2 h_i^2}{(m_i^2 + n\lambda)^2} + \sigma^2 \sum_{i=1}^{r} \frac{m_i^4}{(m_i^2 + n\lambda)^2} \tag{3.5}$$

$$\mathrm{E}\{||\widehat{\beta}(\lambda) - \beta||^2\} = n^2\lambda^2 \sum_{i=1}^{r} \frac{h_i^2}{(m_i^2 + n\lambda)^2} + \sigma^2 \sum_{i=1}^{r} \frac{m_i^2}{(m_i + n\lambda)^2} \tag{3.6}$$

Using these simple forms, we then consider the situation where n is large and the matrix $\mathbf{M}$ is ill-conditioned – in particular we assume that the eigenvalues of $(1/n)\mathbf{M}^{\mathrm{T}}\mathbf{M}$ tend to zero, as $n \to \infty$, under an algebraic decay. In image processing, n will tend to be large ($\sim 2^{16}$) and there will often be a sufficient degree of blurring to render $\mathbf{M}$ an ill-conditioned matrix.

4. Asymptotic Comparisons

Theorem. *For each method of choice of smoothing parameter, we assume that*

$$m_i^2 = nc^{-1}i^{-\nu}(c > 0)(r > 1)$$

As $n \to \infty$, we assume that $\lambda \to 0$ in such a way that $n\lambda^{1/\nu} \to \infty$. For parts (a) – (b) we assume also that

$$b_1 = c\sum_{i=1}^{\infty} i^{\nu} h_i^2 < \infty.$$

whereas for part (e), we require to assume

$$b_2 = c^2 \sum_{i=1}^{\infty} i^{2\nu} h_i^2 < \infty.$$

Then it follows that:

$$(a) \qquad \lambda_{\mathbf{CHI}} = \left[\frac{\sigma^2(2k_1 - k_2)}{nb_1}\right]^{\nu/(2\nu+1)} \quad (1 + o(1))$$

$$(b) \qquad \lambda_{\mathbf{EDF}} = \left[\frac{\sigma^2(k_1 - k_2)}{nb_1}\right]^{\nu/(2\nu+1)} \quad (1 + o(1))$$

$$(c) \qquad \lambda_{\mathbf{GCV}} = \left[\frac{\sigma^2 k_2}{2n\nu b_1}\right]^{\nu/(2\nu+1)} \quad (1 + o(1))$$

$$(d) \qquad \lambda_{\mathbf{TP}} = \left[\frac{\sigma^2 k_3}{nb_1}\right]^{\nu/(2\nu+1)} \quad (1 + o(1))$$

$$(e) \qquad \lambda_{\mathbf{TE}} = \left[\frac{\sigma^2 k_4}{nb_2}\right]^{\nu/(3\nu+1)} \quad (1 + o(1))$$

where $o(1) \to 0$ as $n \to \infty$, and $k_1 - k_4$ are constants.

It follows, immediately, that

Corollary. *As $n \to \infty$,*

$$(f) \qquad \lambda_{\mathbf{CHI}}/\lambda_{\mathbf{TP}} = \left[\frac{2\nu(\nu+1)}{(\nu-1)}\right]^{\nu/(2\nu+1)} \quad (1 + o(1))$$

$$(g) \qquad \lambda_{\mathbf{EDF}}/\lambda_{\mathbf{TP}} = \left[\frac{2\nu}{(\nu-1)}\right]^{\nu/(2\nu+1)} \quad (1 + o(1))$$

$$(h) \qquad \lambda_{\mathbf{GCV}}/\lambda_{\mathbf{TP}} = 1 + o(1).$$

$$(i) \qquad \lambda_{\mathbf{CHI}}/\lambda_{\mathbf{EDF}} = (\nu+1)^{\nu/(2\nu+1)} \quad (1 + o(1))$$

If we use $\lambda_{\mathbf{TP}}$ as a yardstick, then it is clear that, asymptotically speaking,

(1) $\lambda_{\mathbf{CHI}}$ tends to over-smooth.

(2) $\lambda_{\mathbf{EDF}}$ also tends to impose too much smoothing, but not as much as $\lambda_{\mathbf{CHI}}$.

(3) $\lambda_{\mathbf{GCV}}$ is an asymptotically optimal choice.

(4) $\lambda_{\mathbf{TP}}$ and $\lambda_{\mathbf{TE}}$ are not equivalent, even asymptotically.

Of course, it is not entirely clear that $\lambda_{\mathbf{TP}}$ should be used in this way as an "optimal measure" because the problem of image restoration is really a problem of estimation rather than prediction, as pointed out by Rice (1986). The method of regularization attempts to combine the tasks of noise removal and deconvolution by using a single smoothing parameter. On the other hand one could propose an alternative procedure –

that of *Double Smoothing.* Reverting to the notation of (1.1) – (1.4), if we let $b = \mathbf{H}f$, (1.2) becomes

$$g = b + \varepsilon$$

Now apply the method of regularization to find $\widehat{b}$, given by

$$\widehat{b} = (\mathbf{I} + \lambda_{\mathbf{N}} C)^{-1} g$$

where $\lambda_{\mathbf{N}}$ is the smoothing parameter for *noise-removal.* Now apply the method of regularization again to find $\widehat{f}$, given by

$$\widehat{f} = (\mathbf{H}^{\mathbf{T}}\mathbf{H} + \lambda_{\mathbf{D}}\mathbf{I})^{-1}\mathbf{H}^{\mathbf{T}}\widehat{b}$$

where $\lambda_{\mathbf{D}}$ is the smoothing parameter for the *deconvolution* phase of the restoration process. This proposal is currently under investigation.

We have now considered theoretically several choices of smoothing parameters, so let us now apply these methods to some test images.

5. Some restorations

Four test images were constructed and then degraded both by blurring and by the addition of uncorrelated, homogeneous, Gaussian noise. These test images are displayed in Figures 10.1–10.4.

The images were blurred by convolving each neighborhood with a $b \times b$ blurring mask, except at the boundary of the image where a smaller mask was used. Two masks were used:

(B1): the outer product $\underset{\sim}{u}\underset{\sim}{u}^{\mathbf{T}}$, $\underset{\sim}{u} = [\cdot 3 \cdot 4 \cdot 3]^{\mathbf{T}}$

(B2): the outer product $uu^{\mathbf{T}}$, $\underset{\sim}{u} = [\cdot 04 \cdot 12 \cdot 18 \cdot 32 \cdot 18 \cdot 12 \cdot 04]^{\mathbf{T}}$

Two levels of noise were employed viz. $\sigma = 2.0, 7.0$

The smoothing matrix used was that obtained by passing the 3×3 Laplacian mask

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

over the image and convolving this mask with successive overlapping 3×3 windows of the true scene.

Consequently, the matrices H and C of (1.3) and (1.4) are of Block-Toeplitz form, and each block is also a Toeplitz matrix.

These matrices were approximated by Block Circulant matrices so that Fourier computation might be used. This effectively reduces the computational burden of image restoration to three $2D$ FFTs and an inverse $2D$ FFT; for further details, see Kay (1988a).

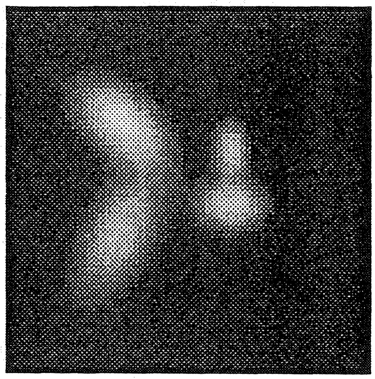

Fig. 10.1

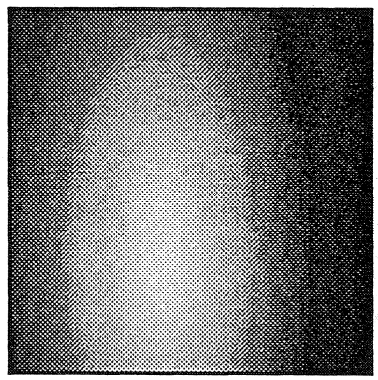

Fig.10.2

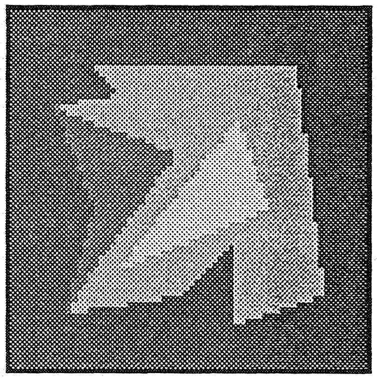

Fig. 10.3

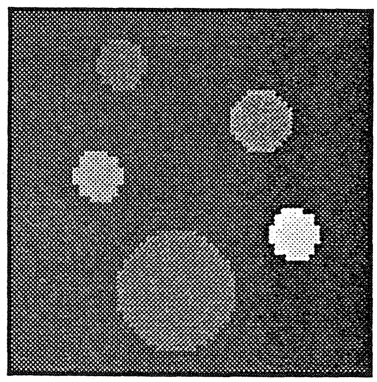

Fig. 10.4

Figures 10.1–10.4. GALAXY, SURFACE, PLANES and CIRCLES: true images. GALAXY is the sum of four bivariate Gaussian probability density functions. SURFACE is a single bivariate Gaussian probability density function. PLANES and CURCLES are piecewise constant images.

The degraded images were restored using each of the five methods described in (2.1–2.5) and also for a range of values of the smoothing parameter λ, including the "No Smoothing" case ($\lambda = 0$).

The noise estimate used for σ was the standard deviation of the data obtained from the random number generator; clearly this tends to favour the CHI and EDF methods. Kay (1988a) proposed a neighborhood noise estimator" for σ and work is currently underway to attempt to eliminate its bias.

The various restorations obtained are shown in Figures. 2.1–9.4. All of the smoothed restorations of GALAXY are similar in quality except for Figure 3.4, which appears to be "over-smoothed" while the TP and TE restorations are reasonable and, perhaps surprisingly, Figures 5.3, 5.4 have a better likeness to the original.

The CHI, EDF and GCV restorations of PLANES are under-smoothed relative to TP and TE, while Figures. 7.3, 7.4 show "over–smoothing". Similar comments apply also to the restorations of CIRCLES.

For these "single–realisation" experiments, EDF tends to impose too little smooth-

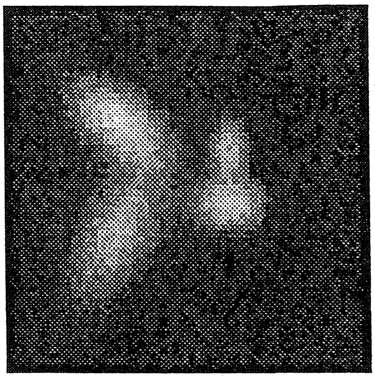

Fig. 1.1

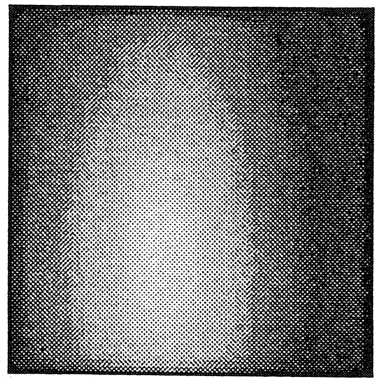

Fig. 1.2

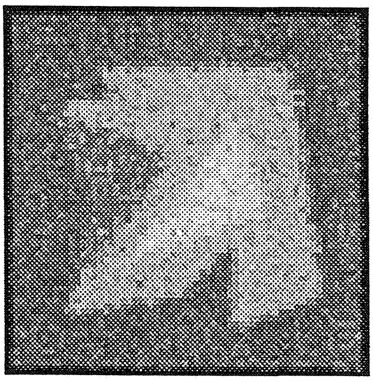

Fig. 1.3

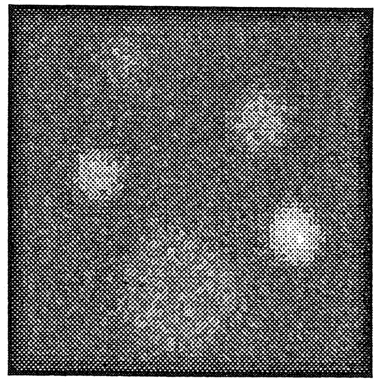

Fig 1.4

Figures 1.1–1.4 GALAXY with blur B1 and $\sigma = 7.0$; SURFACE with blur B2 and $\sigma = 2.0$; PLANES with blur B1 and $\sigma = 7.0$; CIRCLES with blur B1 and $\sigma = 2.0$.

ing; apart from this the restorations obtained using CHI, GCV, TP and TE are of a similar quality – this is also indicated by crude measures of pixel-wise error rates.

It is difficult to escape the conclusion, based on these limited experiments, that while the actual value of the smoothing parameter obtained by the various methods is different – even by a different order of magnitude, in some cases – there appears to be rather little difference in quality between the different methods (apart from EDF). It appears that there is a window of values of λ within which the restorations are of comparable quality. Hence what really matters is obtaining a value of λ in this window. Generally speaking the CHI and GCV methods yields a value of λ within the critical window, which can be rather wide.

Another point to emerge from the experiments was that, contrary to the experiments of Rice (1986), the restorations obtained using TP and TE were generally of similar quality; also the actual values of λ obtained were almost identical.

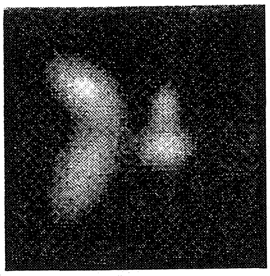

Fig. 2.1

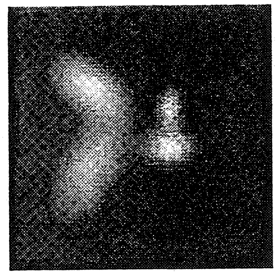

Fig. 3.1

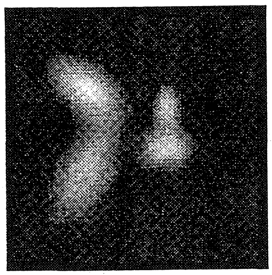

Fig. 2.2

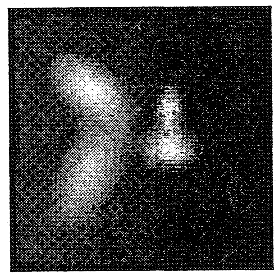

Fig. 3.2

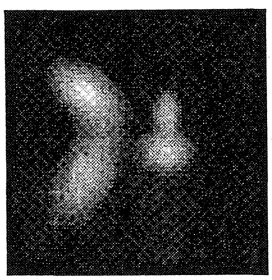

Fig. 2.3

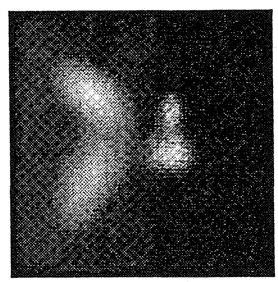

Fig. 3.3

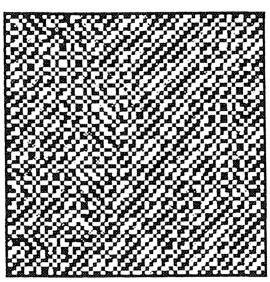

Fig. 2.4

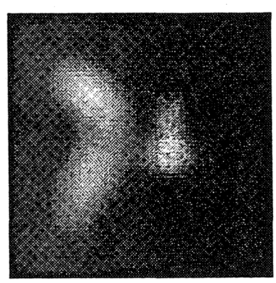

Fig. 3.4

Figures 2.1–2.4. GALAXY restored using CHI, EDF, GCV and No Smoothing.

Figures 3.1–3.4. GALAXY restored using TP, TE, ($\lambda = 2.0$) and ($\lambda = 10.0$).

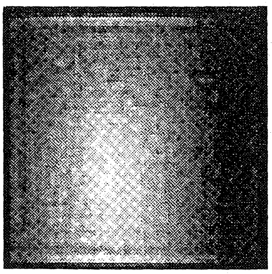

Fig. 4.1

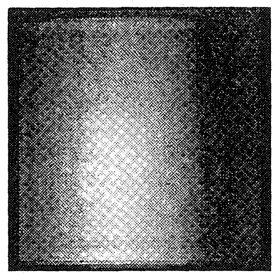

Fig. 5.1

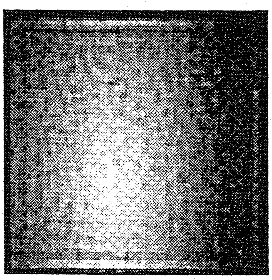

Fig. 4.2

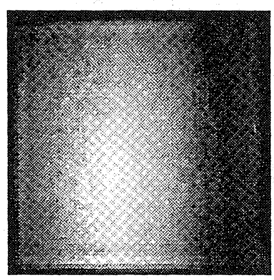

Fig. 5.2

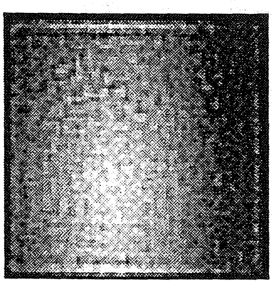

Fig. 4.3

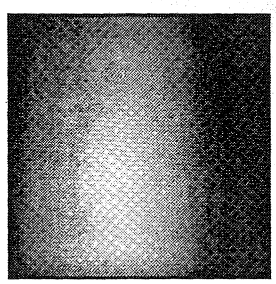

Fig. 5.3

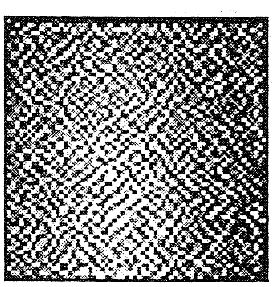

Fig. 4.4

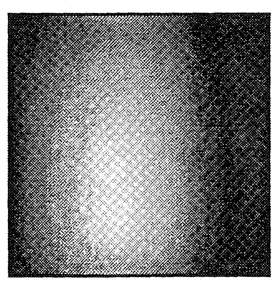

Fig. 5.4

Figures 4.1–4.4. SURFACE restored using CHI, EDFF, GCV and No Smoothing.

Figures 5.1–5.4. SURFACE restored using TP, TE, ($\lambda = 10.0$) and ($\lambda = 100.0$).

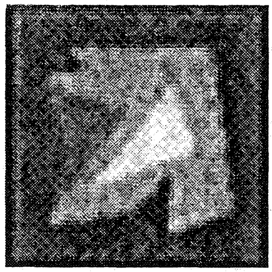

Fig. 6.1

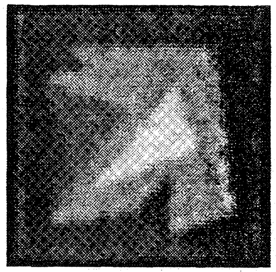

Fig. 7.1

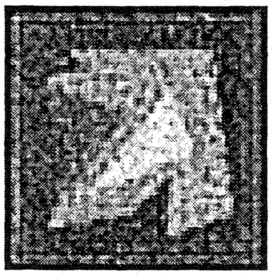

Fig. 6.2

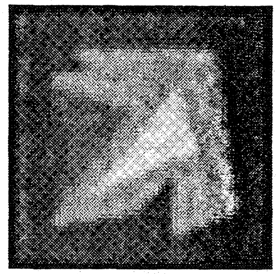

Fig. 7.2

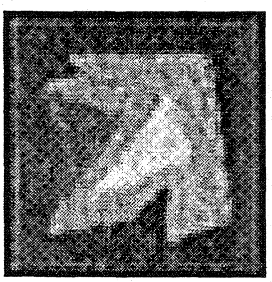

Fig. 6.3

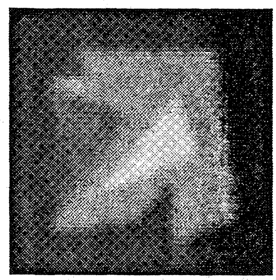

Fig. 7.3

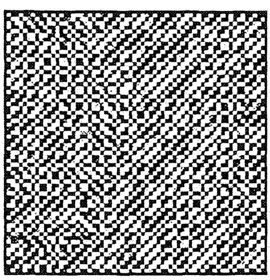

Fig. 6.4

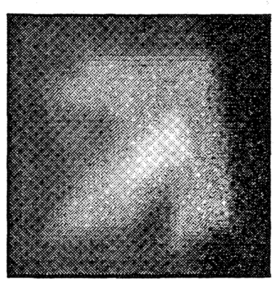

Fig. 7.4

Figures 6.1–6.4 PLANES restored using CHI, EDF, GCV and No Smoothing.
Figures 7.1–7.4 PLANES restored using TP, TE, ($\lambda = 1.0$) and ($\lambda = 10.0$).

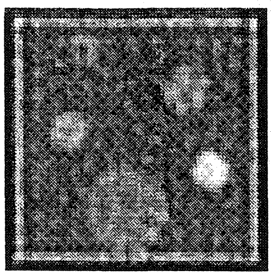

Fig. 8.1

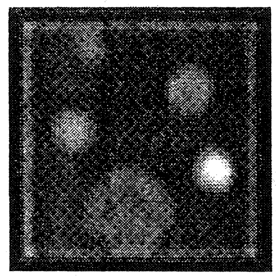

Fig. 9.1

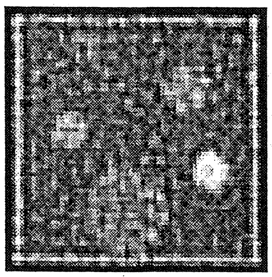

Fig. 8.2

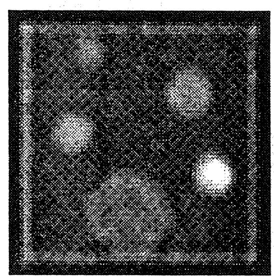

Fig. 9.2

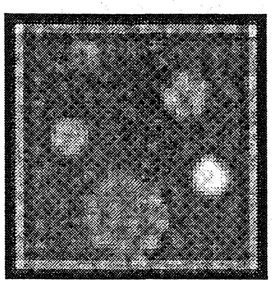

Fig. 8.3

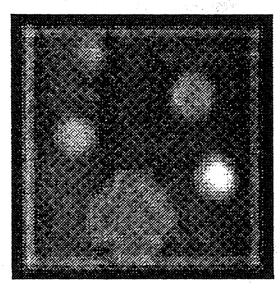

Fig. 9.3

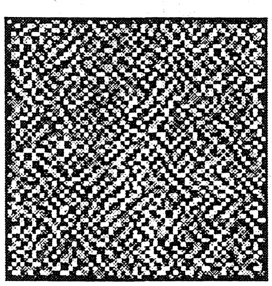

Fig. 8.4

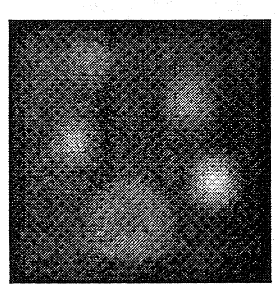

Fig. 9.4

Figures 8.1–8.4 CIRCLES restored using CHI, EDF, GCV and No Smoothing.
Figures 9.1–9.4 CIRCLES restored using TP, TE, ($\lambda = 0.1$) and ($\lambda = 10.0$).

Acknowledgement

I wish to thank D. M. Titterington for several stimulating conversations.

References

Bates, D. and Wahba, G. (1983) A Truncated Singular Value Distribution and other methods for generalized cross-validation Technical Report No. 715 Dept. of Statistics, University of Wisconsin.

Davies, A. R. and Anderssen, R (1986) Optimization in the regularization of ill-posed problems. Journal of the Australian Mathematical Society (Series B), 28, 114–133.

Golub, G. H., Heath, M. and Wahba, G. (1979) Generalized cross-validation as a method of choosing a good ridge parameter. Technometrics 21, 215–223.

Groetsch, C. W. (1984) The theory of Tikhonov regularization for Fredholm equations of the first kind. Research notes in Mathematics No. 105 Pitman.

Hall, P. & Titterington, D. M. (1986) On some smoothing techniques used in image processing. J.R. Statist. Soc. Series B, 48, 330–343.

Hall, P. & Titterington, D.M. (1987) Common structure of techniques for choosing smoothing parameters in regression problems. J. R. Statist. Soc. Series B, 49, 184–198.

Kay, J. W. (1988a) On the choice of regularization parameter in image restoration. Lecture Notes in Computer Science Vol. 301, Springer-Verlag.

Kay, J. W. (1988b) Asymptotic comparison factors for smoothing parameter choices in linear regression. Submitted for publication.

Poggio, T., Torre, V. & Koch, C. (1985) Computational vision and regularization theory. Nature Vol. 317, 314–319.

Rice, J. (1986) Choice of smoothing parameter in deconvolution problems. Contemporary Mathematics Vol. 59, 137–151.

Titterington, D. M. (1985) General structure of regularization procedures in image processing. Astron. Astrophys. 144, 381–387.