ELSEVIER Computational Statistics & Data Analysis 26 (1998) 393–410

COMPUTATIONAL STATISTICS & DATA ANALYSIS

Maximum likelihood restoration and choice of smoothing parameter in deconvolution of image data subject to Poisson noise

H. Malcolm Hudson*,[1], Thomas C.M. Lee[2]

Department of Statistics, SEFS, Macquarie University, NSW 2109, Australia

Received 1 April 1996; accepted 1 July 1997

Abstract

Image degradation by blurring is a well-known phenomenon often described by the mathematical operation of convolution. Fourier methods are well developed for recovery, or restoration, of the true image from an observed image affected by convolution blur and additive constant variance Gaussian noise. One focus of this paper is to describe another statistical restoration method which is available when the image data exhibits Poisson variability. This is a common situation when counts of recorded activity form the image, as in medical imaging contexts.

We apply Maximum Likelihood (ML) and Maximum Penalized Likelihood (MPL) procedures to deconvolve image data which has been degraded by blurring and Poisson variability in recorded activity.

A second focus is formulation and comparison of automated selection procedures for regularization (smoothing) parameters in this context.

Keywords: Cross-validation; Deconvolution; Fast computation; Regularization; Statistical algorithms

1. Introduction

Mathematical and statistical models of image degradation are essential in modern image processing. Many statistical models for images assume image data y of the form $y = \theta + \epsilon$, where *direct* observations are taken on the unknown image θ with

* Corresponding author. E-mail: malcolm.hudson@mq.edu.au.

[1] Supported in part by Australian Research Council grants.

[2] Also supported by CSIRO Mathematical and Information Sciences.

0167-9473/98/$19.00
PII S0167-9473(97)00041-8

an additive noise component ϵ. Models where data is only *indirectly* related to the image θ are often more realistic. One such form of indirect data results from image blurring; another form is tomography. Image data can be degraded by blurring through movement of the camera, atmospheric effects, etc., and such blurring can be readily incorporated in the model for the data as convolution of the true image with a blurring kernel.

Important advances have occurred in Bayesian formulation of image degradation. Markov Random Field models and Markov Chain Monte Carlo methods have been introduced by Geman and Geman (1984) and Grenander (1981) and there has been substantial further development of these ideas (e.g. Besag and Green, 1993).

Statistical models for blurring and related forms of indirect data have been formulated by many authors in the statistical literature. Image processing applications include Besag (1986), Hall (1987) and Titterington (1985). Applications in tomography were introduced by Vardi et al. (1985), and Bayesian formulations by Geman and McClure (1987) and Hebert and Leahy (1989). For consideration of general issues in density deconvolution see Rice (1986) and Stefanski and Carroll (1990).

Reversing effects of blurring involves deconvolution techniques. An integral part of the formulation of such deconvolution techniques has been the selection of a regularization or smoothing parameter.

Deconvolution procedures apply Fourier filtering (see Jain, 1989, Ch. 8) to deblur and restore images, but other approaches have been proposed in certain cases. For example, in tomography, Bayesian MAP methods for Markov Random Field priors have been developed by Green (1990b) and smoothing in concern with penalized likelihood estimation proposed by Silverman et al. (1990) for integral equations of the first-kind (including applications in stereology and positron emission tomography). Bayesian methods lead to an optimization problem for a penalized log likelihood, similar to the approach to be adopted here. These methods have been found to be superior to traditional Fourier deconvolution (e.g. Chornboy et al., 1990) in tomography.

Our paper will study the effectiveness of penalized likelihood methods in deconvolution problems involving blurring of a *localized* nature and Poisson observation data, thus extending the evaluation of MPL procedures beyond tomography. Poisson noise models are often more realistic than Gaussian models for emitted activity in a region.

MPL has some natural advantages over filtering. MPL based on Poisson model assumptions automatically implements a weighted deconvolution, since accuracy of the observations varies according to the level of counts, unlike Fourier methods for additive noise models. MPL methods preserve natural positivity constraints on a density or an image, a property not shared by Fourier deconvolution methods. Maximum likelihood (ML) and maximum penalized likelihood (MPL) restorations may therefore be preferred for some purposes.

In Section 2 we introduce methods for ML and MPL restoration with an image formed by counts subject to Poisson variability. These methods are formulated for *global* descriptions of the effect of blurring on image observations, but are readily applicable within segmented regions of an image.

MPL restoration involves the choice of a smoothing parameter. In Section 3 we discuss some automatic methods for choosing such a smoothing parameter. Coakley (1991), Coakley and Llacer (1991), Green and Silverman (1994) and Johnson (1994) have studied methods of smoothing parameter selection for deconvolution.

Methods commonly adopted for this task include cross-validation procedures and Mallows C_L. The computation required for these procedures with image data can be impractical. In Section 4 we provide techniques for fast computation.

In Section 5, we apply ML and MPL restoration to test image data, and compare the deblurred results with restorations provided by Fourier filtering. Empirical results demonstrate that MPL restoration is superior to ML restoration and Fourier filtering.

2. Mathematical model, ML and MPL restorations

2.1. Mathematical model

Assume the true image (not directly observable) specifies the intensity of emitted activity in a region of interest. The digitized true image $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$ is a p-vector recording intensities of emitter activity within p pixels comprising this region of interest. To interpret elements of $\boldsymbol{\theta}$ as corresponding to pixel emitter activities, introduce the positivity constraint $\boldsymbol{\theta} \geq 0$.

The true image $\boldsymbol{\theta}$ is further degraded by *blurring* and *Poisson noise* before recording. The resulting data will generally be digitized on the same grid, but need not be. In order to allow some flexibility for interpolative and smoothing methods, we permit digitization with respect to a new grid of N pixels defined on the original image domain.

2.1.1. Blurring

The blurred image $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)^T$ is defined by

$$\boldsymbol{\mu} = \boldsymbol{A}\boldsymbol{\theta}, \tag{1}$$

where $\boldsymbol{A} = \{a_{ij}\}$ is a known $N \times p$ stochastic matrix, which reallocates the initial pixel activity to recordings on other pixels, according to probabilities specified by the elements of the matrix. If the blurring is local in nature, only affecting nearby pixels, $\boldsymbol{A}$ will be sparse, each row containing only relatively few non-zero elements. It is convenient to define the sets I_i and J_j: define I_i as the set of indices of those θ_j's contributing to μ_i and define J_j as the set of indices of those μ_i's to which θ_j contributes. If we insist the blurring introduced by $\boldsymbol{A}$ should preserve the total count number, then $\sum_{i \in J_j} a_{ij} = \sum_{j \in I_i} a_{ij} = 1$.

2.1.2. Poisson noise

We depart from the usual linear filtering formulation that specifies the recorded image as an additive combination of (blurred) signal and white noise. While the additive Gaussian noise model can provide a reasonable approximation to some blurred image contexts, particularly in high count, spatially stationary images, Poisson noise

models are often more suitable. They offer correct physics for random emission phenomena (as in emission tomography), and appropriate modelling of resulting non-stationary errors, with noise variance proportional to the signal (local emitted activity).

The blurred image is instead assumed to be subject to Poisson variability. Observed data are independent Poisson counts $\boldsymbol{y} = (y_1, \ldots, y_N)^{\mathrm{T}}$ with

$$\boldsymbol{y} \sim P(\boldsymbol{\mu}), \tag{2}$$

where $\boldsymbol{\mu}$ is specified in Eq. (1). With this setting, our aim is to restore the true image $\boldsymbol{\theta}$ given an observed image $\boldsymbol{y}$.

In addition, the above model can be easily generalized to handle multiplicative noises by modifying (1) with the use of a log-link:

$$\log(\boldsymbol{\mu}) = \boldsymbol{A}\boldsymbol{\theta}.$$

However, this generalization will not be pursued here.

2.2. ML restoration

The model described by Eqs. (1) and (2) is a generalized linear model with identity link, in the terminology of McCullagh and Nelder (1989). This suggests parameter estimation be conducted using the method of ML. The log likelihood function is

$$\sum_{i=1}^{N} (y_i \log \mu_i - \mu_i) \quad \text{with } \mu_i = \sum_{j \in I_i} a_{ij}\theta_j, \tag{3}$$

and setting the partial derivatives of log likelihood with respect to $\boldsymbol{\theta}$ to zero provides corresponding normal equations:

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{\theta} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{y}, \tag{4}$$

where $\boldsymbol{V} = \mathrm{diag}(\mu_1, \ldots, \mu_N)$. Hence one can obtain the ML estimate $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ either by (i) maximizing the log likelihood function (3) or (ii) directly solving the normal equations (4). We begin by discussing the second approach.

2.2.1. Solving normal equations

If $\boldsymbol{V}$ were known (4) would be normal equations providing a weighted least squares solution $\hat{\boldsymbol{\theta}}$. The weights, diagonal elements of $\boldsymbol{V}^{-1}$, are usually replaced by current estimates in iterative reweighted least squares (see Green, 1984). An initial set of weights is specified by $\hat{\boldsymbol{V}} = \mathrm{diag}(y_1, \ldots, y_N)$. Iterative least squares is a natural approach to obtain a ML solution but may produce negative estimates of some image activity parameters, so the required positivity constraint is not satisfied by direct solution of the normal equations. For this reason, and because the approach is computationally very demanding, we will not consider it any further.

2.2.2. Maximizing log likelihood in the positive orthant

Since the model specified by Eqs. (1) and (2) is formally equivalent to the model for emission tomography of Vardi et al. (1985), the EM algorithm may also be applied to solution of normal equations, according to the method of Shepp and Vardi (1982). The EM algorithm provides an iterative solution that will preserve positivity constraints, and satisfy a similar set of normal equations to (4).

Denote $\boldsymbol{\theta}^{(k)} = (\theta_1^{(k)}, \ldots, \theta_p^{(k)})$ as the kth iterative estimate of $\boldsymbol{\theta}$ and let $\boldsymbol{\mu}^{(k)} = \boldsymbol{A}\boldsymbol{\theta}^{(k)}$. The EM algorithm provides the iteration:

$$\theta_j^{(k+1)} = \begin{cases} \theta_j^{(k)} \sum_{i \in J_j} \dfrac{a_{ij} y_i}{\mu_i^{(k)}} \Big/ \left(\sum_{i \in J_j} a_{ij} \right) & \text{if } \theta_j^{(k)} > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The formal equivalence of our problem to tomography applications remarked on above deserves comment. Both blurring and tomography applications involve large non-negative, sparse model matrices $\boldsymbol{A}$. A common reconstruction technique is desirable, though the different but special structure of the model matrix in each case can often be exploited for effective solution (e.g. Girard, 1987).

It is known that the restoration provided by such iterative algorithms is improved by stopping iterations early. Because of the high dimension of the parameter relative to data dimension and the ill-conditioned nature of the design matrix $\boldsymbol{A}$, the ML solution is often unstable and requires a regularization method, of which stopping iteration early is just one approach.

2.3. Regularization by MPL

The importance of regularization in ill-conditioned inverse problems is well understood, for example see Poggio et al. (1985). We can illustrate the inadequacy of ML restoration in deconvolution by a simple example.

If $\boldsymbol{A}$ is nonsingular, and all components of $\boldsymbol{\mu}$ are strictly positive, the normal equations defining ML simplify to the linear system $\boldsymbol{A}\boldsymbol{\theta} = \boldsymbol{y}$, specifying perfect agreement between fitted data (blur of $\hat{\boldsymbol{\theta}}$) and observed counts. This linear system produces overfitting of data y and an unstable solution, the estimates being very sensitive to the noise in data in the common case where some eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ are near 0.

In order to provide more stable solutions, the MPL procedure adopts a roughness penalty discouraging variability in $\boldsymbol{\theta}$. With a quadratic penalty, $J(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{\theta}$, the penalized log likelihood function is

$$\sum_{i=1}^{N} (y_i \log \mu_i - \mu_i) - \lambda J(\boldsymbol{\theta}), \tag{6}$$

which provides normal equations

$$(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{A} + \lambda \boldsymbol{R})\boldsymbol{\theta} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{y},$$

similar to the estimating equations in ridge regression. Here $\boldsymbol{R}$ is an $p \times p$ matrix of roughness coefficients, assumed positive definite, and $\lambda > 0$ a smoothing parameter

specifying the relative emphasis placed on likelihood and roughness of the estimates of image activity. From a Bayesian point of view, adding the penalty term specified by $\boldsymbol{R}$ can be interpreted as imposing a Gaussian prior on $\boldsymbol{\theta}$.

The iterative weighted least squares procedure remains available for MPL parameter estimations (as before, this normal-equation-solving approach will not be discussed any further, as it does not preserve the positivity property), but the EM algorithm is computationally impractical in general. A modified EM algorithm, the One Step Late (OSL) algorithm of Green (1990a), has been applied to MPL estimation in tomography. For the present problem, the OSL algorithm gives the iteration:

$$\theta_j^{(k+1)} = \begin{cases} \theta_j^{(k)} \left(\sum_{i \in J_j} \frac{a_{ij} y_i}{\mu_i^{(k)}} \right) \Big/ \left(\sum_{i \in J_j} a_{ij} + \lambda (\boldsymbol{R\theta})_j \right) & \text{if } \theta_j^{(k)} > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where $(\boldsymbol{R\theta})_j$ is the jth component of the vector $\boldsymbol{R\theta}$, the gradient of $J(\boldsymbol{\theta})$. Alternative iterative schemes are also available, for example see Hudson et al. (1994).

3. Automated choice of smoothing parameter

In this section we discuss two data-driven methods for choosing the smoothing parameter λ in (6). However, we shall begin with a brief review of some popular automatic smoothing parameter selection methods for additive noise image models.

3.1. A brief review of methods for additive noise image models

Blurring, with additive white noise, is a particular case of the familiar linear model $\boldsymbol{y} = \boldsymbol{A\theta} + \boldsymbol{\epsilon}$. For such image models, popular approaches for automatic smoothing parameter selection include Bayesian, unbiased risk estimation and cross-validation; see the review papers by Hall and Titterington (1987), Thompson et al. (1991) and Chan and Kay (1991). Related papers are Rice (1986) (a pioneering paper for smoothing parameter selection in deconvolution problems), Galatsanos and Katsaggelos (1992) (describing different approaches in the context of image restoration, and proposing a marginal likelihood approach), and Reeves and Mersereau (1992) (proposing methods for image restoration incorporating regularization when the blurring kernel is of a parameterized form). Solo (1988) and Hall and Koch (1992) provide some asymptotic results.

3.2. Choice of optimality criterion

Recent work with additive noise assumptions has investigated the distinction between estimation domain and predictive domain criteria in regard to optimal selection of the smoothing parameter. In the current context, the question is whether λ should

be chosen to minimize the discrepancy between a restoration and the true image (i.e. estimation), or the discrepancy between the blurred restoration and the blurred true image (i.e. prediction). Since blurring is generally a low pass filter on an image, the optimal smoothing parameter based on prediction errors is less than the optimal parameter based on estimation domain errors, implying less smooth restoration.

The example studied by Rice (1986), in a study of deconvolution, suggested that the choice of estimation domain could provide very different results to prediction domain.

Thompson et al. (1991), in a study of regularized solutions, employed the fast Fourier transform to efficiently compute generalized cross-validation, in restorations of simulated blurred images resulting from known source images. They concluded that the predictive mean squared error criterion, using the prediction domain, was an appropriate measure of quality of restoration in such cases.

Hall and Koch (1992) studied deblurring with additive noise. For a specific convolution example, asymptotic results for a "leave-one-out" predictive sum of squares (PRESS) were available for low blur data. Study was directed to asymptotic rates of convergence of minimizers of specified cross-validation functions appropriate for the estimation domain.

Galatsanos and Katsaggelos (1992) suggest prediction domain criteria weight more heavily data points corresponding to large eigenvalues of the blurring operator, and may therefore be preferred, though this runs contrary to findings of Solo (1988).

Recent opinion seems to be forming that while the optimal parameter may be markedly different in prediction and estimation domain, the quality of restoration is not markedly different when employing criteria appropriate to one or the other domain. For example, Thompson et al. (1991), and Chan and Kay (1991), conclude in their examples that minimizers of estimation and prediction errors were of a similar quality, suggesting that the choice was not a practical issue for restoration of large degraded images.

Other metrics that describe information content of an image, and fidelity to the true image, have been suggested, and no single criterion will satisfy all considerations. Some alternatives to global mean square prediction error are described in Hanson (1990) and Wilson et al. (1996).

3.3. *Cross-validation and unbiased risk estimation for Poisson deconvolution*

Now we introduce two automated methods for choosing the smoothing parameter λ in (6): one is based on cross-validation and while the other is based on unbiased risk estimation. The general idea is to choose λ to minimize the *prediction risk*:

$$r_p = E\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}^{\lambda}\|^2,$$

where $\hat{\boldsymbol{\mu}}^{\lambda} = \boldsymbol{A}\hat{\boldsymbol{\theta}}^{\lambda}$ is an estimate of $\boldsymbol{\mu}$ for a given λ. (The norm operator is defined as: $\|\boldsymbol{x}\|^2 = \sum x_i^2$ for $\boldsymbol{x} = (x_1, \ldots, x_n)^T$). One can also use the more appropriate *estimation risk*

$$r_e = E\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^{\lambda}\|^2$$

as a measure for quality of fit. However, as mentioned in Girard (1995), r_e is more difficult to work with. Even though our aim is to estimate θ (i.e. λ should be chosen to minimize r_e), for the reason of simplicity, we focus on r_p. Notice that if the blurring defined by A is local in nature, there should be no big difference between r_p or r_e.

Since r_p is unknown, a standard approach is to construct an unbiased estimator of r_p and choose λ to minimize this risk estimator. Cross-validation aims for this: if $\hat{\mu}^{\lambda}_{i,-i}$ is the "leave-one-out" estimate of μ_i (i.e. remove the observation of the ith pixel when estimating μ_i), then the cross-validation score

$$cv(\lambda) = \sum_i (y_i - \hat{\mu}^{\lambda}_{i,-i})^2$$

should be asymptotically unbiased for r_p (except for a constant term). Our first method for automated choice of λ is to choose λ as the minimizer of $cv(\lambda)$.

One disadvantage of this leave-out-one approach is that, the leaving out of one observation introduces a missing pixel in the image. In principle, the matrix A must be modified; this requires a row elimination in A in Eq. (1). The EM/OSL algorithm may then be applied to the resulting linear system. An equivalent procedure, the two-stage EM algorithm, is provided by Gidas and Hudson (1991). This approach imputes an estimate for the missing data value and applies the standard EM iteration. The value imputed is recomputed in each iteration, being the fitted value for the missing pixel count based on the current image estimates.

Besides the above cross-validation method, other estimators for r_p also exist. Hudson (1974, 1978) derived an *unbiased risk* estimator for estimators of Poisson means. For linear estimators, for which $\hat{\boldsymbol{\mu}}^{\lambda} = \boldsymbol{B}_{\lambda}\boldsymbol{y}$, where $\boldsymbol{B}_{\lambda} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{A} + \lambda\boldsymbol{R})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}$ is the hat matrix, Pawitan and O'Sullivan (1993) have derived the risk estimator

$$\hat{r}_p(\lambda) = \|\boldsymbol{y} - \hat{\boldsymbol{\mu}}^{\lambda}\|^2 + 2\mathrm{tr}\{\boldsymbol{B}_{\lambda}\boldsymbol{D}(\boldsymbol{y})\} - \sum y_i,$$

where $\boldsymbol{D}(\boldsymbol{y})$ is diagonal with elements $\boldsymbol{y}$, and $\hat{r}_p(\lambda)$ is unbiased for r_p. Our second method is to choose λ as the minimizer of $\hat{r}_p(\lambda)$.

Computations for the quantities $cv(\lambda)$ and $\hat{r}_p(\lambda)$ are impractical in MPL as they require repeated non-linear iterative calculations. EM restorations of as many blurred images as there are pixels are required for exact computation, for even a single choice of λ. Consequently, fast techniques for approximating these quantities are required; see next section.

4. Techniques for fast computational approximations

4.1. Techniques for fast $cv(\lambda)$ calculations: Sampling and local approximation

In this section we consider two techniques which can be used to improve the computation speed for $cv(\lambda)$.

The first technique termed *sampling* originated from the work of Kozek (1994). The idea is to approximate/estimate

$$cv(\lambda) = \sum_{i=1}^{N} (y_i - \hat{\mu}_{i,-i}^{\lambda})^2$$

by

$$\hat{cv}(\lambda) = \sum_{i \in S} (y_i - \hat{\mu}_{i,-i}^{\lambda})^2,$$

where S is a subset of the integer set $\{1,\ldots,N\}$. Elements of S can be chosen in a random or a regular fashion from $\{1,\ldots,N\}$. The computational time gained from applying this technique depends on the number of the elements in S. For example, if the number of elements of S is $N/10$, then the computation time of $\hat{cv}(\lambda)$ is only one-tenth that of $cv(\lambda)$, and yet $\hat{cv}(\lambda)$ is very likely to provide a good estimate of $cv(\lambda)$ (especially if N is large).

The second technique termed *local approximation* can be used if (i) the data do not exhibit long range dependence and (ii) the blurring specified by $\boldsymbol{A}$ is local in nature. If these two conditions are satisfied, then ignorance of the data values in regions far from the pixel of interest is not important. Hence one can approximate each $\hat{\mu}_{i,-i}^{\lambda}$ by $\tilde{\mu}_{i,-i}^{\lambda}$, where $\tilde{\mu}_{i,-i}^{\lambda}$ is the leave-one-out estimate of μ_i based *only* on neighbouring pixel observations of pixel i. It is clear that $\tilde{\mu}_{i,-i}^{\lambda}$ can be computed much quicker than $\hat{\mu}_{i,-i}^{\lambda}$ if the size of the pre-chosen neighbourhood is small when compared to the image size. From an implementation point of view, it is convenient to define the neighbourhood as a local rectangular subset of the image. In other words, this permits solution in a local moving window (with size much smaller than the size of the image) centred on the pixel of interest.

One can combine sampling and local approximation techniques. That is, approximate $cv(\lambda)$ by

$$\tilde{cv}(\lambda) = \sum_{i \in S} (y_i - \tilde{\mu}_{i,-i}^{\lambda})^2. \tag{8}$$

4.2. Technique for fast $\hat{r}_p(\lambda)$ calculation: trace approximation

Computation of the risk estimator $\hat{r}_p(\lambda)$ requires computation of the trace

$$\mathrm{tr}\{\boldsymbol{B}_{\lambda}\boldsymbol{D}(\boldsymbol{y})\} = \mathrm{tr}\{\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{A} + \lambda\boldsymbol{R})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{D}(\boldsymbol{y})\}.$$

This is computationally demanding, but an approach (reported with success) suggested by Pawitan and O'Sullivan (1993) in another context can reduce computations. First choose $\boldsymbol{V} = \boldsymbol{D}(\boldsymbol{y})$. Then making the crude substitution, $\bar{y}\boldsymbol{I}$ for $\boldsymbol{V}$ ($\bar{y}$ is the mean of $y_1,\ldots,y_N$ and $\boldsymbol{I}$ is the identity matrix), provides the simplification

$$\mathrm{tr}\{\boldsymbol{B}_{\lambda}\boldsymbol{D}(y)\} = \bar{y}\,\mathrm{tr}\{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \lambda\bar{y}\boldsymbol{R})^{-1}\},$$

When $\boldsymbol{A}$ and $\boldsymbol{R}$ are simultaneously diagonalized by discrete Fourier transforms, evaluation is straightforward (e.g. see Brillinger, 1981). That is, for the reason of fast computation, one can approximate $\hat{r}_p(\lambda)$ by

$$\tilde{r}_p(\lambda) = \|\boldsymbol{y} - \hat{\boldsymbol{\mu}}^{\lambda}\|^2 + 2\bar{y}\,\mathrm{tr}\{\boldsymbol{A}^T\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A} + \lambda\bar{y}\boldsymbol{R})^{-1}\} - \sum y_i. \tag{9}$$

Recently O'Sullivan and Pawitan (1996) provide another approach for tomography application.

5. Example – lung data deconvolution

A case study is provided to illustrate the methods described for deconvolution, and compare MPL restoration solutions afforded by different automatic smoothing parameter selection methods.

5.1. The data

From an initial image, obtained from camera recordings in emission tomography, synthetic data was simulated.

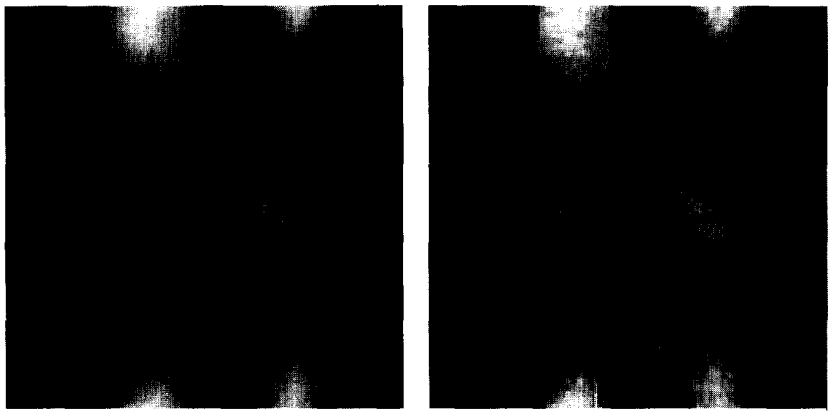
The initial image, corresponding to $\boldsymbol{\theta}$ in the notation of Section 2, is the *sinogram* displayed in Fig. 1(a). Each row of this image represents (on a range between 0 corresponding to black, and 255 corresponding to white) the numbers of recorded photons collected in 64 bins along a line scan on a camera. Each row corresponds to one fixed camera position; in emission tomography the 64×64 sinogram digitization results from rotation of the camera after each of 64 such line scans, the camera rotation around the patient producing the observed sinusoidal effects.

A crude investigation of the effect of imperfect resolution in the camera was undertaken by the creation of a synthetic blurred image. Blurring replaced each pixel activity in the sinogram image by a local average of activity in the first-order neighbourhood of immediate neighbours of that pixel. Neighbours of pixels on the outer borders were defined by wrap around. This is likely to be a reasonable specification because of the periodic nature of the camera rotation (through 360°) and the zero counts at camera edges. Blurring utilised weights 0.5 (central pixel), 0.125 (4 neighbour pixels). Following blurring, the data image was created with independent Poisson distributed activity in pixels with means specified by the blurred image. Total counts of the resulting blurred, noisy image were 202,752. Signal-to-noise ratio (snr) varied substantially throughout the synthetic sinogram, with relatively low snr in regions of low emitter activity. The noisy image is displayed in Fig. 1(b).

5.2. Methods for deconvolution

Several deconvolution methods were considered. They are described below.

Fourier filtering was implemented by discrete Fourier transform diagonalization of the design matrix corresponding to the blurring; the design matrix is banded in



(a) (b)

Fig. 1. Lung sinogram images: (a) true projections of lung, (b) blurred noisy projections.

structure (see Brillinger, 1981). Smoothing was incorporated as a Weiner filter (e.g. Shumway, 1988, Ch. 4.4) with smoothing parameter appropriate for snr of 10:1.

ML restoration of emitter activity were constructed through solution of the EM iteration (5).

MPL restorations were similarly obtained by the OSL iteration (7). The penalty functional was specified as $\lambda\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{\theta}$ with $\boldsymbol{R}$ defined below. We evaluated two forms for $\boldsymbol{R}$: (i) $\boldsymbol{R}=\boldsymbol{I}$, the identity matrix, corresponding to amplitude penalization; (ii) $\boldsymbol{R}$ with additional offdiagonal elements -0.25 for four first-order neighbours. This second choice penalizes discrepancies between estimates in neighbour pixels.

Automated estimation of λ in MPL was considered. The optimal data specific choice λ was established from MPL restorations, calculating average squared errors in the estimation domain (AESE, see below), for λ in an interval $[0, 0.025]$.

The cross-validation and unbiased risk estimation criteria $cv(\lambda)$ and $\hat{r}_{\mathrm{p}}(\lambda)$ of Section 3.3 were studied. Since calculations of these criteria are computational demanding, we adopted the fast computational techniques described in Section 4. That is, we approximated $cv(\lambda)$ and $\hat{r}_{\mathrm{p}}(\lambda)$ by $\tilde{cv}(\lambda)$ and $\tilde{r}_{\mathrm{p}}(\lambda)$, respectively.

When computing $\tilde{cv}(\lambda)$, the set S, see (8), comprised 180 randomly selected pixels within the region of activity (i.e. "non-dark" region) of the image and the local windows for computing those $\tilde{\mu}_{i,-i}$'s are of size 15×15 pixels.

For all iterative restorations mentioned above, our criterion for convergence was the mean of the squared pixelwise differences between two successive restorations less than 0.01.

5.3. Results

Fig. 2(a) displays the restored image (from the blurred noisy image in Fig. 1(b)) using the Wiener filter with snr set to be 10. Visual inspection tells that the restoration is unsatisfactory. This may be attributed to: (i) the erroneous assumption of identi-

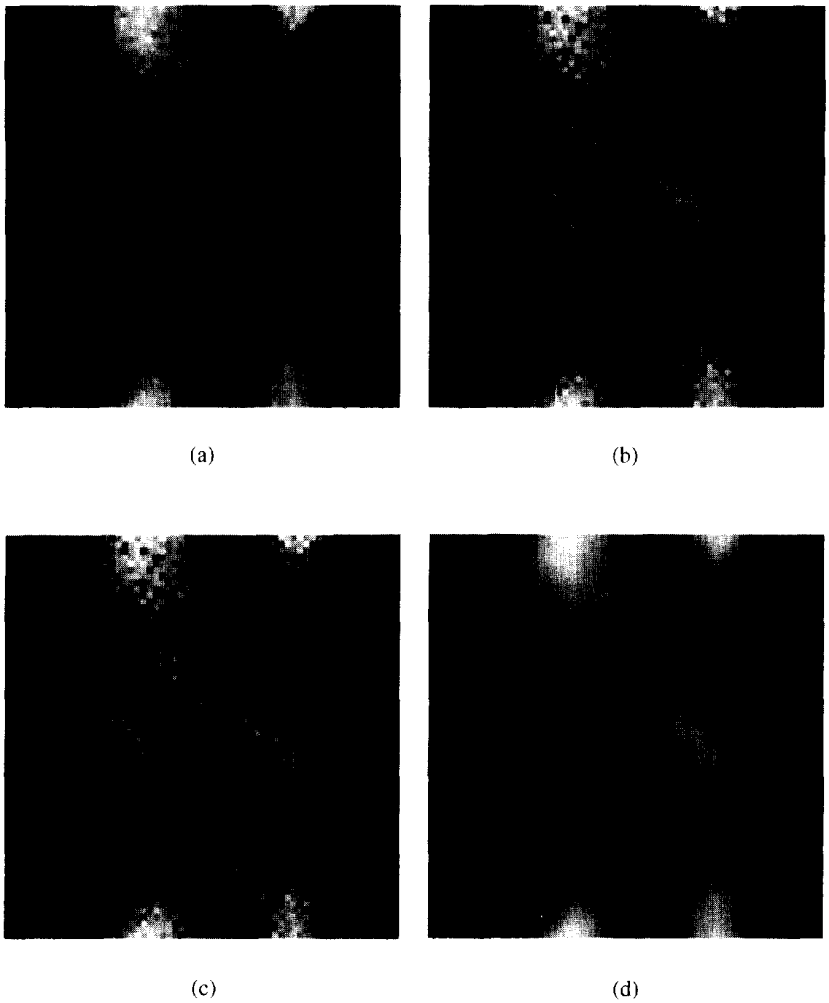


Fig. 2. 4 different restorations: (a) Fourier, (b) ML, (c) MPL with $\boldsymbol{R} = \boldsymbol{I}$, (d) MPL with local roughness penalty.

cally distributed errors in the additive noise model; (ii) not imposing the essential positivity constraint.

The ML restoration is displayed in Fig. 2(b). Again, visual inspection suggests that the restoration is unsatisfactory. The EM algorithm required 49 iterations for convergence.

Figs. 2(c) and (d) are the MPL restoration using two choices for the penalty matrix $\boldsymbol{R}$. The smoothing parameter λ was chosen automatically by minimizing $c\tilde{v}(\lambda)$. The selections were $\lambda = 0.002$ and $\lambda = 0.008$ for the two roughness matrix choices. These methods required 49 and 12 iterations, respectively. The approximation $\tilde{r}_p(\lambda)$ attained its minimum at $\lambda = 0$. That is, it produced the ML restoration.

Table 1
Various summary statistics of the restorations

Restoration	AESE	APSE	RSS
Fourier filtering	128	67.2	55.7
ML	294	44.6	2.9
MPL ($\boldsymbol{R} = \boldsymbol{I}$)	293	44.4	2.9
MPL (local roughness)	23	11.6	29.5

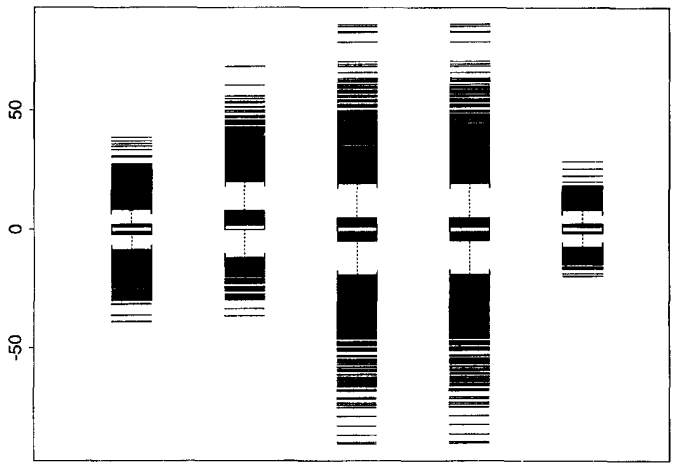


Fig. 3. Boxplots of the differences between the true (unknown) image and (from left to right) (i) noisy blurred image, (ii) Fourier restoration, (iii) ML restoration, (iv) MPL with $\boldsymbol{R} = \boldsymbol{I}$ restoration, (v) MPL with local roughness penalty restoration.

The true (but unobservable) average estimation squared error (AESE) $\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}\|^2$, the true (but unobservable) average prediction squared error (APSE) $\|\hat{\boldsymbol{\mu}}^{\lambda} - \boldsymbol{\mu}\|^2$ and the (observable) residual sum of squares (RSS) $\|\boldsymbol{y} - \hat{\boldsymbol{\mu}}^{\lambda}\|^2$ of the restorations are given in Table 1. Both AESE and APSE indicate that the MPL with local roughness penalty approach provided the best restoration. See also Fig. 3.

Hence MPL restoration with local roughness penalty and choice of λ as the minimizer of $c\tilde{v}(\lambda)$ was the most successful real-time automated restoration approach amongst approaches considered. The accuracy of estimating λ by this approach depends on the choice of those sample pixels and the size of the local window. From our experience, when the number of sample pixels (i.e. number of elements in the set S) and the size of the local window were sufficiently large, the minimizer of $c\tilde{v}(\lambda)$ always provided a reliable estimate for the optimal λ. For the example presented in this paper, we found that 180 sample pixels and an 15×15 local window were sufficiently large to ensure accurate choice of λ (see Section 5.4 for some empirical results). These figures represent a sampling fraction $180/4096 < 5\%$ and reduction of image dimensions to $225/4096 < 6\%$ for restoration and cross-validation

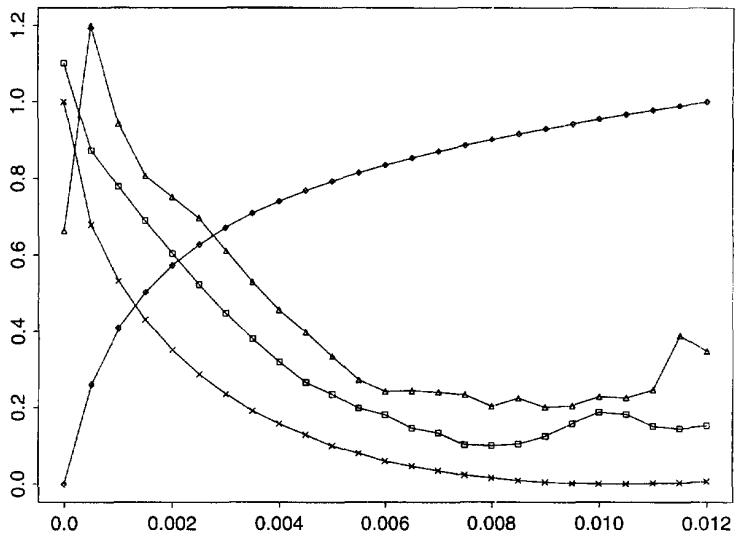


Fig. 4. Plots of different sum of squares statistics. (i) ×: true but unknown AESE, (ii) □: exact cross-validation score $cv(\lambda)$, (iii) △: approximate cross-validation score $\tilde{cv}(\lambda)$, and (iv) ◇: $\tilde{r}_p(\lambda)$ with trace approximation. Scaling of the y-axis is arbitrary, and specific to each function.

Table 2
Averages ($n = 100$) of relative errors $(\tilde{\lambda} - \lambda_0)/\lambda_0$

Number of sample pixels s	Width of local window w				
	7	11	15	19	23
100	0.221	0.296	0.250	0.292	0.259
140	0.212	0.266	0.319	0.299	0.263
180	0.217	0.264	0.293	0.244	0.295
220	0.222	0.268	0.299	0.283	0.276
260	0.219	0.278	0.270	0.255	0.224

residual computations. Together, they provide an essential major reduction of computations.

Fig. 4 displays the plot of a number of estimation and prediction domain statistics for the MPL restoration with local roughness penalty. These include: (i) the true but unknown AESE, (ii) $cv(\lambda)$, (iii) $\tilde{cv}(\lambda)$, and (iv) $\tilde{r}_p(\lambda)$. For display purposes the values of functions have been given the same scale. The selected minimizer $\lambda = 0.008$ of $\tilde{cv}(\lambda)$ can be seen to be close to the optimal smoothing parameter near $\lambda = 0.010$. However $\tilde{r}_p(\lambda)$ does not parallel AESE(λ). Other computations indicated that the trace approximation is inaccurate in our context.

We have also applied those methods discussed in this section to other data sets and similar results are obtained.

Table 3
Standard deviations ($n = 100$) of relative errors $(\tilde{\lambda} - \lambda_0)/\lambda_0$

Nmber of sample pixels s	Width of local window w				
	7	11	15	19	23
100	0.327	0.330	0.347	0.364	0.357
140	0.309	0.310	0.335	0.379	0.309
180	0.279	0.347	0.266	0.317	0.338
220	0.292	0.302	0.298	0.309	0.297
260	0.311	0.325	0.336	0.256	0.251

Table 4
Average computation time for $\tilde{\lambda}$

Number of sample pixels s	Width of local window w				
	7	11	15	19	23
100	6	17	37	63	110
140	8	24	49	86	152
180	11	31	63	114	188
220	13	38	77	136	227
260	16	44	86	162	275

5.4. *Effect of local window size and sampling fraction*

An important factor to the success of the MPL restoration with local roughness penalty is the quality of the selected smoothing parameter, $\tilde{\lambda}$, defined to be the minimizer of $c\tilde{v}(\lambda)$. It is obvious that the quality of $\tilde{\lambda}$ is dependent on the size of the local window, w, and the number of sample pixels, s. We studied the effect of different choices of (w,s) on $\tilde{\lambda}$.

From the true lung sinogram $\boldsymbol{\theta}$ displayed in Fig. 1(a), 100 blurred noisy images similar to the one displayed in Fig. 1(b) were simulated. For each of these noisy images, we computed the optimal smoothing parameter, λ_0, which minimizes the (unobservable) AESE $\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}\|^2$. For each noisy image, we also computed 25 $\tilde{\lambda}$'s corresponding to 25 different combinations of (w,s). In other words, for each of the 25 (w,s) combinations, we have computed 100 λ_0's as well as 100 $\tilde{\lambda}$'s. These computed values provide some understanding to the statistical behaviour of $\tilde{\lambda}$ as an estimate of λ_0 under different (w,s) combinations. The 25 selected (w,s)'s are provided in Table 2.

Bias of $\tilde{\lambda}$: For each (w,s), we computed the average of 100 values of $(\tilde{\lambda} - \lambda_0)/\lambda_0$, an estimate of the bias of $\tilde{\lambda}$ relative to the optimal value for that data. Results are given in Table 2: $\tilde{\lambda}$ shows consistent positive bias, the reconstruction is slightly oversmoothed. Unexpectedly, the bias of $\tilde{\lambda}$ depends little on (w,s), though it seems to be smaller when the size of the local window is smallest $(w = 7)$!.

Variability of $\tilde{\lambda}$: For each (w,s), we also computed the sample standard deviation of $(\tilde{\lambda}-\lambda_0)/\lambda_0$, a measure of the variability of $\tilde{\lambda}$; see Table 3. As expected, the larger are w and s, the smaller is the variability of $\tilde{\lambda}$.

Computational time of $\tilde{\lambda}$: For comparative purposes, we list in Table 4 some typical times (in seconds) required for computing $\tilde{\lambda}$ on a Sparc-10 workstation. These figures represent a huge reduction of computational effort, as the time for computing a full scale cross-validation λ (without fast approximation) is about three days on the same machine.

6. Discussion

While many extensive studies have now been conducted examining choice of smoothing parameter in stationary additive noise models, there has been little investigation of parameter selection for Poisson counts, except in the tomography context.

For our data the optimal choice of λ was similar for either estimation or prediction domain. Similar smoothing parameter choices were expected because blurring was very localized. In such circumstances, the choice of a predictive domain risk minimizer seems acceptable.

The 3×3 blurring in our example is very localized, but heavy in that even with no noise, the original image is unrecoverable. This singularity of solution has not prevented MPL methods providing adequate restorations of the source image in the presence of additional Poisson variability. Furthermore, we have been successful in obtaining a data determined choice of smoothing parameter in MPL (based on a cross-validation score) very near the optimal value. We were able to speed up the calculation of this cross-validation score using techniques discussed in Section 4, and obtained acceptable smoothing parameter estimates.

We recently discussed work by Girard (1995) which illustrates Monte-Carlo cross-validation and C_L procedures for linear estimators in a general context. In this approach synthetic white noise was used for necessary trace components which determine PRESS curves. The restriction to constant variance noise is not essential to Girard's method, so that the method appears to be applicable for a linearized form of MPL restoration. The promise of Girard's method for automated smoothing parameter estimation is left for future evaluation.

Acknowledgements

We thank Peter Hall and Andrzej Kozek for useful discussions, and the referees for helpful comments.

References

Besag, J., 1986. On the statistical analysis of dirty pictures. J. Roy. Statist. Soc. Ser. B 48, 259–302.

Besag, J., Green, P., 1993. Spatial statistics and Bayesian computation. J. Roy. Statist. Soc. Ser. B 55 25–38.

Brillinger, D., 1981. Time Series: Data Analysis and Theory. Holden-Day, San Francisco.
Chan, K.P.-S., Kay, J.W., 1991. Smoothing parameter selection in image restoration. In: Roussas, G. (Ed.), Nonparametric Functional Estimation and Related Topics, NATO Advanced Study Institute, Kluwer, Dortrecht, pp. 201–211.
Chornboy, E., Chen, C., Miller, M., Miller, T., Snyder, D., 1990. An evaluation of maximum likelihood reconstruction for SPECT. IEEE Trans. Med. Imaging MI-9, 99–110.
Coakley, K.J., 1991. A cross-validation procedure for stopping the EM algorithm and deconvolution of neutron depth profiling spectra. IEEE Trans. Nucl. Sci. 38, 9–15.
Coakley, K., Llacer, J., 1991. The use of Cross-Validation procedure as a stopping rule in Emission Computed Tomography image reconstruction. In: SPIE Proc. Image Phys: Med. Imaging V, vol. 1443, pp. 226–233.
Galatsanos, N.P., Katsaggelos, A.K., 1992. Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. IEEE Trans. Image Proc. 1, 322–336.
Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Machine Intell. 6, 721–735.
Geman, S., McClure, D., 1987. Statistical methods for tomographic image reconstruction. In: Proceedings of the 46th Session ISI, Bulletin ISI, vol. 52.
Gidas, B., Hudson, H.M., 1991. A two stage EM algorithm with applications in emission tomography, Technical Report 39, Brown University, Complex System Report.
Girard, D.A., 1987. Optimal regularized reconstruction in computerized tomography. SIAM J. Sci. Statist. Comput. 8(6), 934–950.
Girard, D.A., 1995. The fast Monte-Carlo cross-validation and C_L procedures: Comments, new results and applications to image recovery problems. Comput. Statist. 10, 205–231.
Green, P., 1984. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives (with discussion). J. Roy. Statist. Soc. Ser. B 46, 149–192.
Green, P., 1990a. Bayesian reconstruction from emission tomography data using a modified EM algorithm. IEEE Trans. Med. Imaging MI-9, 84–93.
Green, P., 1990b. On use of the EM algorithm for penalized likelihood estimation. J. Roy. Statist. Soc. Ser. B 52, 443–452.
Green, P., Silverman, B., 1994. Nonparametric Regression and Generalized Linear Models, Chapman & Hall, London.
Grenander, U., 1981. Abstract Inference, Wiley, New York.
Hall, P., 1987. On the processing of a motion-blurred image. SIAM J. Appl. Math. 47, 441–453.
Hall, P., Koch, I., 1992. On the feasibility of cross-validation in image analysis. SIAM J. Appl. Math. 52, 259–313.
Hall, P., Titterington, D., 1987. Common structure of techniques for choosing smoothing parameters in regression problems. J. Roy. Statist. Soc. Ser. B 49, 184–198.
Hanson, K.M., 1990. Method of evaluating image-recovery algorithms based on task performance. J. Opt. Soc. Amer. 7, 1124–1302.
Hebert, T., Leahy, R., 1989. A generalized EM algorithm for 3-D Bayesian reconstruction from poisson data using Gibbs priors. IEEE Trans. Med. Imaging MI-8, 194–202.
Hudson, H.M., 1974. Empirical Bayes estimation, Ph.D. thesis, Stanford University, Technical Report No. 58, Dept. Statistics.
Hudson, H.M., 1978. A natural identity for exponential families with applications in multiparameter estimation. Ann. Statist. 6, 473–484.
Hudson, H.M., Ma, J., Green, P., 1994. Fisher's method of scoring in statistical image reconstruction: comparison of Jacobi and Gauss-Seidel iterative schemes. Statist. Methods Med. Res. 3, 41–61.
Jain, A., 1989. Fundamentals of Digital Image Processing, Prentice-Hall, Englewood Cliffs, NJ.
Johnson, V., 1994. A note on stopping Rules for EM-ML algorithm reconstructions of ECT images. IEEE Trans. Med. Imaging 13, 569–571.
Kozek, A.S., 1994. On a partial cross validation in nonparametric regression. In: Proceedings of the Interface, 26, pp. 187–191.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, 2nd ed., Chapman & Hall, London.
O'Sullivan, F., Pawitan, Y., 1996. Bandwidth selection for indirect density estimation based on corrupted histogram data. J. Amer. Statist. Assoc. 91 (434), 610–626.
Pawitan, Y., O'Sullivan, F., 1993. Data dependent bandwidth selection for emission computed tomography reconstruction. IEEE Trans. Med. Imaging 12, 167–172.
Poggio, T., Torre, V., Koch, C., 1985. Computational vision and regularization theory. Nature 317, 314–319.
Reeves, S.J., Mersereau, R.M., 1992. Blur identification by the method of Generalized Cross Validation. IEEE Trans. Image Proc. 1, 301–311.
Rice, J., 1986. Choice of smoothing parameter in deconvolution problems. Contemp. Math. 59, 137–151.
Shepp, L., Vardi, Y., 1982. Maximum likelihood reconstruction for emission tomography. IEEE Trans. Med. Imaging MI-2, 113–122.
Shumway, R.H., 1988. Applied Statistical Time Series Analysis, Prentice Hall, Englewood Cliffs, NJ.
Silverman, B.W., Jones, M.C., Wilson, J.D., Nychka, D.W., 1990. A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography (with discussion). J. Roy. Statist. Soc. Ser. B 52, 271–324.
Solo, V., 1988. Linear image restoration: An analytical study, Technical Report 88-VS-May-1988, Dept. of ECE, The Johns Hopkins Univ., Baltimore, MD.
Stefanski, L., Carroll, R.J., 1990. Deconvoluting kernel density estimators. Statistics 21, 169–184.
Thompson, A.M., Brown, J.C., Kay, J.W., Titterington, D.M., 1991. A study of methods of choosing the smoothing parameter in image restoration by regularization. IEEE Trans. Pattern Anal. Machine Intell. 13, 326–339.
Titterington, D.M., 1985. Common structure of smoothing techniques in statistics. Internat. Statist. Rev. 53, 141–170.
Vardi, Y., Shepp, L., Kaufman, A., 1985. A statistical model for positron emission tomography (with discussion). J. Amer. Statist. Assoc. 80, 8–20.
Wilson, D., Baddeley, A., Owens, R., 1997. A new metric for grey-scale image comparison. Internat. J. Comput. Vision, 24, 5–17.