②

AD-A205 643

AD

Statistical smoothing methods and image analysis

Final Technical Report

by

C. Jennison & B. W. Silverman

December 1988

United States Army

EUROPEAN RESEARCH OFFICE OF THE U.S. ARMY

London England

CONTRACT NUMBER DAJA45-86-C-0054

University of Bath

Approved for Public Release; distribution unlimited

89 3 03 035

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188
Exp Date Jun 30, 1986

1a REPORT SECURITY CLASSIFICATION: Unclassified

1b. RESTRICTIVE MARKINGS:

2a SECURITY CLASSIFICATION AUTHORITY:

2b DECLASSIFICATION / DOWNGRADING SCHEDULE:

3 DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution Unlimited

4 PERFORMING ORGANIZATION REPORT NUMBER(S):

5. MONITORING ORGANIZATION REPORT NUMBER(S): R&D 5421-MA-01

6a. NAME OF PERFORMING ORGANIZATION: University of Bath

6b. OFFICE SYMBOL (If applicable):

7a NAME OF MONITORING ORGANIZATION: European Research Office USARDSG-UK

6c. ADDRESS (City, State, and ZIP Code):
School of Mathematical Sciences
Bath, BA2 7AY
UK

7b. ADDRESS (City, State, and ZIP Code):
Box 65
FPO NY 09510-1500

8a. NAME OF FUNDING / SPONSORING ORGANIZATION: USARDSG-UK ARO-E

8b OFFICE SYMBOL (If applicable):

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: DAJA45-86-C-0054

8c. ADDRESS (City, State, and ZIP Code):
Box 65
FPO NY 09510-1500

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
61103A	1L161103BH5	7 03	

11. TITLE (Include Security Classification): (U) Statistical Smoothing Methods and Image Analysis

12. PERSONAL AUTHOR(S): C. Jennison and B.W. Silverman

13a. TYPE OF REPORT: Final

13b. TIME COVERED: FROM 1/10/86 TO 31/9/88

14 DATE OF REPORT (Year, Month, Day): 1988/12/31

15 PAGE COUNT: 265

16. SUPPLEMENTARY NOTATION:

17 COSATI CODES

FIELD	GROUP	SUB-GROUP
12	01	
09	04	

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number):

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

The research reported here covers a number of different areas. The methodology of density estimation has been considered with particular reference to the smoothed bootstrap. A new method of fitting parsimonious additive models has been devised. The topic of statistical integral equations has been investigated in detail and algorithms for two main cases of particular interest have been developed and investigated. Applications to image analysis have been considered. Contributions to the theory of estimation from indirect information have been made. There has been careful consideration of the appropriate way to penalise an edge process model in an image reconstruction. The methodology of nonparametric discriminant analysis, with particular reference to the CART approach, has been the subject of considerable attention. The ICM method of image reconstruction has been studied. A new method of image refinement has been developed.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☒ DTIC USERS

21 ABSTRACT SECURITY CLASSIFICATION: Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL: Dr. Gerald R. Andersen

22b TELEPHONE (Include Area Code): 01-409 4423

22c OFFICE SYMBOL: AMXSN-UK-RP

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

SUMMARY

The ~~research reported here~~ covers a number of different areas. The methodology of density estimation has been considered with particular reference to the smoothed bootstrap. A new method of fitting parsimonious additive models has been devised. The topic of statistical integral equations has been investigated in detail and algorithms for two main cases of particular interest have been developed and investigated. Applications to image analysis have been considered. Contributions to the theory of estimation from indirect information have been made. There has been careful consideration of the appropriate way to penalise an edge process model in an image reconstruction. The methodology of nonparametric discriminant analysis, with particular reference to the CART approach, has been the subject of considerable attention. The ICM method of image reconstruction has been studied. A new method of image refinement has been developed.

KEYWORDS: density estimation; smoothed bootstrap; nonparametric regression; computer algebra; parsimonious additive models; piecewise linear fitting; statistical integral equations; stereology; tomography; EM algorithms; missing data; indirectly observed images; image analysis; information theory; edge process; nonparametric discriminant analysis; classification and regression trees; Markov random fields; iterated conditional modes; image refinement.

CONTENTS

APPENDICES

[1] Silverman, B. W. & Young, G. A. The bootstrap: to smooth or not to smooth?

[3] Jennison, C. & Jubb, M. Statistical image restoration and refinement.

[4] Silverman, B.W., Jennison, C. & Brown, T.C. The specification of edge penalties for regular and irregular pixel images.

[5] Friedman, J.H. & Silverman, B.W. Flexible parsimonious smoothing and additive modeling.

[6] Wilson, J.D. A smoothed EM algorithm for the solution of Wicksell's corpuscle problem.

[7] Glendinning, R.H. An evaluation of the ICM algorithm for image reconstruction.

[9] Johnstone, I.M. & Silverman, B.W. Speed of estimation in positron emission tomography.

[11] Silverman, B.W., Jones, M.C, Wilson, J.D. & Nychka, D.W. A smoothed EM approach to a class of problems in image analysis and integral equations.

[12] Jones, M.C. & Silverman, B.W. An orthogonal series density estimation approach to reconstructing positron emission tomography images.

[14] Jubb, M. & Jennison, C. Aggregation and refinement in binary image restoration.

1. Introduction

Most of the work conducted under the aegis of the project has now been written up in the form of papers submitted for publication. These are listed in Section 10 below. In this report, a brief description of the work done will be given under a number of headings, and fuller details are given in the papers, most of which are attached as appendices. Against each head are given the numbers of the relevant papers in the publication list. The same numbering system is used for the appendices.

2. Density Estimation [1,13]

One of the aims of the project was the extension of the existing density estimation methodology in various directions. Among these was the use of density estimation in techniques such as the smoothed bootstrap. A criterion has been developed for deciding whether smoothing is worth performing in any particular bootstrap situation. For full details, see [1]. One novel feature was the use of Computer Algebra to solve this statistical problem, and Professor Silverman gave an extremely well received presentation on this aspect to a Royal Statistical Society workshop on Computer Algebra in Statistics.

In 1951, Fix and Hodges wrote a technical report which contained prophetic work on nonparametric discriminant analysis and density estimation. The report introduced several important concepts for the first time, and was never published. It is not just of historical interest, but contains much material of contemporary relevance. A commentary [13] has been written placing the paper in context and interpreting its ideas in the light of more modern developments. The commentary has been submitted for publication together with the paper itself.

3. Parsimonious additive models [5,10]

A very simple and powerful new method for fitting nonlinear regression models was devised and investigated by Professor Silverman in collaboration with J.H. Friedman of Stanford. The basic idea is to fit a sequence of segmented linear regressions on single variables to the data and then to use a suitable stopping rule to decide when to stop elaborating the model. Finally a backward elimination step is used to resimplify up to an appropriate point. The paper [5] on this material was selected by the editors of *Technometrics* to be the special discussion paper at the 1988 ASA meetings and will shortly appear with discussion and rejoinder in that journal.

4. Solution of Statistical Integral Equations [6,8,11,15]

A considerable amount of work has been carried out on the general topic of the solution of statistical integral equations. The main aim has been the development of a general approach that can be applied to any problem where the model for the observed data is obtained by applying a (known) compact linear operator A to the function f of real interest. There are two cases of main interest, where the data arise as observations of Af at known points subject to error ("regression dependence") and where the data are observations from a non-homogeneous Poisson process with intensity Af ("density dependence".) We have developed methodology for both of these cases, as discussed separately below.

Regression dependence

Suppose the data are of the form $Y_i = Af(t_i) + \varepsilon_i$ where $Af(t) = \int A(t,u)f(u)du$ and the ε_i are uncorrelated errors with mean zero. A natural estimate of f is given by constrained penalised least squares, where one finds $\hat{f}$ to minimise $S(f) = \|Af-Y\|^2 + \alpha\int f''^2$, subject to any relevant linear constraints on f, such as positivity. In all the applications of interest, positivity is a constraint on f and in some cases f is constrained in addition to integrate to 1. Particular practical problems of interest arose from consultation with materials scientists. Another practical problem considered in detail was the determination of the ventilation/perfusion distribution over the human lung given data on inert gas elimination (Evans & Wagner, *J. Appl. Physiol.* **42**, 889-898, 1977).

The approach adopted was to apply quadratic programming to a discretised version of $S(f)$, as follows: the function f was approximated by a vector of values $\mathbf{f}$ on a grid; the vector $Af(t_i)$ of the values of Af at each of the values t_i can then be expressed by a simple quadrature rule as $K\mathbf{f}$, where K is a suitable matrix. The roughness penalty $\int f''^2$ is approximated by a quadratic form $\mathbf{f}^T D\mathbf{f}$. One then minimises $(\mathbf{Y}-K\mathbf{f})^T(\mathbf{Y}-K\mathbf{f}) + \alpha\mathbf{f}^T D\mathbf{f}$ subject to $\mathbf{f} \geq 0$ and any other relevant constraints. The quadratic programming method used was that of Wolfe (*Econometrica* **27**, 382-398, 1959) which involves similar manipulations to the simplex algorithm for linear programming. The use of Wolfe's algorithm has the advantage that the final simplex tableau gives additional information that is of use as explained below. The broad conclusions of the work were as follows; some of these are treated in detail in [15]. It is intended that one or two papers will be written based on this work.

(a) The quadratic programming algorithm terminated in a reasonable number of steps in all the applications and simulations tried. For any particular data set, the number of pivots — and hence the time taken to find the solution — is approximately constant as the smoothing parameter α varies.

(b) The positivity constraint alone is not sufficient to provide a properly regularised solution; some smoothing (i.e. $\alpha > 0$) is required in addition. This casts a little doubt on some of the existing methodology in this field, in which one chooses a control parameter to get a non-negative solution and then assumes implicitly that this solution will be sufficiently smoothed.

(c) By considering the special case (linear nonparametric regression) where the minimisation of $\|\mathbf{Y}-Af(\mathbf{t})\| + \alpha\int f''^2$ can be carried out explicitly, it appears that the discretisation has a negligible effect, except where α is chosen inappropriately small.

(d) The final QP tableau makes it possible to draw approximate Bayesian posterior confidence intervals for the curve f with only trivial additional computational effort. This procedure has been implemented and investigated. One particular point of interest is the frequentist behaviour of the Bayesian confidence intervals, which has been considered (for example by Wahba) in the nonparametric regression case. In our more general setting, it is clear that the choice of smoothing parameter is crucial to this frequentist behaviour. Wahba has suggested that a smoothing parameter chosen by generalised cross-validation will give Bayesian posterior intervals which are also (pointwise) frequentist confidence intervals. In the case of ill-posed problems, our work casts doubt on the generalisability of this claim, because in practice varying α between quite wide limits produces curves which are completely different in appearance but which fit the data almost equally well; the difference between the solutions lies in a space spanned by singular functions of A with extremely small singular values. Yet while the goodness-of-fit to the data remains almost unchanged, the width and frequentist coverage probabilities of the Bayesian intervals changes dramatically. This general behaviour also shows that it is more appropriate to make use of prior information rather than attempting to choose the smoothing parameter automatically. One extremely useful aspect of the posterior intervals was in demonstrating to a materials science client the effect of the ill-posed nature of his particular problem. It was immediately clear that the detailed question he was asking was not resolvable on the basis of the experiment conducted.

Density dependence

Suppose now that the function A is a non-negative function satisfying $\int A(t,u)dt = 1$ for all u and that the available data consist of independent observations Y_i drawn from the probability density Af. We might think of these observations as being "indirect" observations from the density f of real interest, since in many of the practical problems of this type, there is an unobservable sample X_i drawn from f itself, and each Y_i is drawn from the density $A(y,X_i)$. Examples of this situation in practice include the classical stereology problem of determining the particle-size distributions from data collected on plane sections through a composite medium, and the problem in image processing of reconstructing a section through the human body by means of positron emission tomography. Vardi, Shepp and Kaufman (*J. Amer. Statist. Assoc.* **80**, 8-37, 1985) describe this latter problem in detail and give an approach based on the EM algorithm that aims towards a maximum likelihood estimate of f in any problem of this kind. In the positron emission tomography problem they consider in detail, the EM algorithm does not actually converge in a reasonable number of steps, and so they propose stopping after a finite number of steps thereby obtaining a smoother estimate of f than would be given by maximum likelihood, but one which depends on the starting point of the iterations and which is not a limit point of any iterative procedure.

We have developed a general approach in which a smoothing step is introduced between each EM iteration. The smoothing part of each iteration involves very little computational effort. A wide variety of linear and non-linear smoothers have been tried and the conclusion is that best results are obtained by a simple local averaging procedure; furthermore the effect of quite a small amount of smoothing is quite dramatic. The effect of the discrete grey level nature of the images was also considered and it was found that the best results are obtained by working in continuous values for the level of the images and only discretising at the display stage. On all our empirical evidence, the smoothed EM procedure converges in a reasonable number of iterations, and furthermore the limit point of the procedure does not depend on the starting configuration. We have demonstrated heuristically that the smoothed EM approach corresponds to a classical EM algorithm applied to a *penalised* maximum likelihood problem, where the likelihood is penalised by a term depending quadratically on the square root of the function of interest. The paper [6] dealing with the specific application to the classical stereology problem is already in press; a more general discussion, and several particular points concerning the positron emission tomography problem, is given in the paper [11].

It is also possible to apply the smoothed EM approach to problems of the regression dependence kind. Some comparisons have been made in [15] between this approach and the quadratic programming method for one-dimensional problems. The general conclusion is that the results obtained are very similar, and the quadratic programming method has the advantage of providing approximate posterior confidence intervals at negligible cost. For problems in which f is a function on a higher dimensional space, the quadratic programming approach could not be applied, because the computational cost of each iteration depends quadratically on the number of pixels or bins, and the number of pivots required is bounded below by the number of bins in which the solution is non-zero.

5. Availability of information in indirect observation problems [9,12]

In any indirect observation problem, it is of interest to ask how much information is actually available in a sample of given size, as compared to an experiment in which "direct" observations are available from the density function being estimated. We have concentrated on the positron emission tomography problem, but the general methodology is applicable to any indirect estimation problem where the singular value decomposition of the integral operator can be expressed explicitly, and also, as explained in Section 6 of the paper, to a wider class of related problems.

Given a large sample $\{Y_i\}$ of indirect observations, we consider the size of the equivalent sample $\{X_i\}$ of observations, whose original exact positions would allow equally accurate estimation of the image of interest. Both for indirect and for direct observations, we establish exact minimax rates of convergence of estimation, for all possible estimators, over suitable smoothness classes of functions. For indirect data and (in practice unobservable) direct data in a two-dimensional version of the PET problem, the rates for mean integrated square error are $n^{-p/(p+2)}$ and $(n/\log n)^{-p/(p+1)}$ respectively, for densities in a class corresponding to bounded square-integrable pth derivatives. We obtain numerical values for equivalent sample sizes for minimax linear estimators using a slightly modified error criterion.

One of the technical tools used in the paper is an orthogonal series approach based on the singular value decomposition of the integral operator. Although this originally arose for theoretical reasons, it is shown in [9] that it yields estimates that are in a sense rate-optimal. Although this estimator can only be constructed in the special case where the SVD is explicitly available, its calculation in this case can be carried out quickly, and so it was of interest to explore its practical behaviour. In [12] an investigation of this kind was conducted. The method has the advantages of speed and of independence of any pixellation; it has the disadvantages of ignoring the positivity constraint and of making rapid changes in value more difficult to achieve than the EMS algorithm. Nevertheless it is clear that the method is certainly useful as a "quick and dirty" approach in those cases where the SVD is tractable.

6. Edge process models [2,4]

One of the ingredients of recent methodology in statistical image restoration is the idea of introducing a system of "edges" between pixels in the image. See, for example, Geman and Geman (*IEEE Trans. PAMI*-**6**, 721-741, 1984). If an edge is present between two contiguous pixels then they are not considered as neighbours in the restoration procedure. The use of such a process is likely to be of value in restoring images which consist of a number of regions within each of which the value varies smoothly. In penalized maximum likelihood estimation of the image, the number and configuration of the edges is controlled by a penalty term; in model-based restoration using Markov random fields there is an analogous penalty term in the energy function of the Gibbs distribution for the edge process. We have investigated how some geometrical insights can be used to provide penalties for the various edge configurations in a way that is roughly independent of the pixel discretisation. The penalties we obtained are consistent over pixels of different sizes, shapes and orientations, even if these occur in the same pattern; pixel grids consisting of pixels of different sizes are a key element in the work on positron emission tomography discussed in [11]. The cases of square, rectangular, hexagonal and irregular pixels are considered.

7. Nonparametric Discriminant Analysis

A great deal of work has been carried out on the Classification and Regression Tree (CART) approach to nonparametric discriminant analysis. This has not yet been written up in the form of papers, but will first appear in detail in P.C. Taylor's (1989) PhD thesis.

We have designed and implemented an entirely novel method of displaying the classification tree making use of sophisticated colour graphics. This method produces "block tree diagrams" which have great practical value in explaining what the procedure is doing, and methodological value in pointing out ways in which the current algorithm is working well and badly. Another area of attention has been alternative splitting criteria with particular reference to the problems raised when there is a large number of classes to be considered. In addition to the Gini criterion and the twoing procedure suggested by Breiman et al, we have investigated five new suggested splitting criteria some (but not all!) of which appear to have great promise. The next main contribution has been in looking at adaptive "anti end-cut factors" which work to prevent the introduction of large numbers of splits that remove very small parts of the data. Such factors need to depend adaptively on such things as the number of cases at the current node and the number of species represented, and these ideas have been incorporated into the procedures.

Further refinements have been made to the display program for presenting the successive splits carried out by CART on a colour display. In particular ideas for dealing with categorical variables have been included in the package. The algorithm itself has been enhanced to include a surrogate splits option, which allows the program to cope with missing values in the predictor variables. Surrogate splits can also be used to rank the importance of each predictor variable in terms of their discriminatory usefulness. Ongoing activity is in two main areas. The first is aimed at reducing the amount of pruning required to create a classification tree. The second is an attempt to detect hierarchies in the class structure. For example, when discriminating between different types of vehicle, we may hope that tracked and wheeled vehicles could be distinguished near the root of the tree.

8. Image Refinement [3,14]

A consequence of the use of a statistical model for a true scene is the possibility of producing restored images on a finer pixel grid than that on which the signal is originally collected. This fact, pointed out by Jennison in the discussion of Besag's paper (*J. Royal Statist. Soc.*, **48**, 288-289), has formed the basis of a very promising avenue of research. There are immediate potential applications in LANDSAT imaging and other forms of aerial photography where, because of the large pixel size, the proportion of "mixed" pixels can be disturbingly high; a proper subdivision of such pixels into regions of more than one type should improve classification rates considerably. More generally, methods which do not impose the unrealistic assumption that a scene is uniform within each pixel offer a possibility of more accurate restoration in all image problems.

Work in this area has been carried out with the assistance of M. Jubb, whose (1989) PhD Thesis will contain a full account of progress thus far. It became apparent at an early stage of our work that there would be insuperable computational problems associated with pixel subdivision beyond a 2×2 refinement. However, the limiting case in which arbitrary boundaries are allowed within each pixel was found to be quite tractable. Our initial work was to implement a method for computing an approximation to the solution of this limiting case problem in which straight line edges were allowed within each pixel. This procedure was found to be very effective in the presence of low levels of additive noise; details and examples appear in [3].

Further work has tackled the same problem in the presence of greater noise levels. An important technique for producing starting values which can then be updated iteratively by the edge fitting algorithm is signal aggregation: by adding together signals from groups of pixels a signal on a coarser pixel grid but with greater signal to noise ration is obtained. A cascade algorithm, in which a series of restorations are obtained at successively lower levels of signal aggregation has been developed. This has been found to produce good restorations for very noisy data which existing methods fail to handle at all well. Details of this algorithm are given in [14].

Our research in this area is still continuing. In particular, we are considering the additional problems associated with grey-level data and true images which contain objects separated by sharp boundaries but also with smooth changes in colour within an object.

9. Markov random field algorithms for image restoration [7]

The iterated conditional modes (ICM) approach of Besag and the annealing approach of Geman and Geman have been investigated. A suite of programs and algorithms implementing these approaches to image analysis was written in order to give a basis for experimentation and improvement. A large simulation study was then carried out on some aspects of these approaches. One particular aspect of interest has been the investigation of the appropriate choice of interaction parameter(s) in the Markov random field model as used in the prior for the images. A theoretical argument demonstrates that an appealing procedure is to weight diagonal neighbours of each pixel by $2^{-\frac{1}{2}}$ the amount used to weight direct neighbours. Such a scheme should produce reconstructions that are are largely unaffected by the way in which the pixel grid is placed on the true underlying image. The broad conclusions of the simulation study were that worthwhile gains can be achieved using an 'optimal' value of of Besag's parameter β rather than the portmanteau value 1.5, and that in the absence of specific prior knowledge about the corrupted scene a second order neighbourhood model with down-weighted diagonals should be used, for example the one suggested by the theoretical arguments referred to above. For full details see [7].

10. Publications

The following publications were produced under the aegis of the project.

[1] Silverman, B. W. and Young, G. A. (1987). The bootstrap: to smooth or not to smooth? *Biometrika*, 74, 469-479.

[2] Brown, T. C. and Silverman, B. W. Edge process models for regular and irregular pixels. Technical Report No. 267, Department of Statistics, Stanford University, Stanford, California. Subsumed within [4] below.

[3] Jennison, C. and Jubb, M. (1988). Statistical image restoration and refinement. *Information Processing in Medical Imaging*, (C.N. deGraaf & M.A. Viergever, eds.) pp. 255-262.

[4] Silverman, B.W., Jennison, C. & Brown, T.C. The specification of edge penalties for regular and irregular pixel images. (33 pp.) Submitted for publication.

[5] Friedman, J.H. and Silverman, B.W. (1988). Flexible parsimonious smoothing and additive modeling. *Technometrics*, to appear. Presented as a special discussion paper at the 1988 American Statistical Association Annual Meetings.

[6] Wilson, J.D. (1988). A smoothed EM algorithm for the solution of Wicksell's corpuscle problem. (27 journal pages) *J. Statist. Comput. Simul.*, to appear.

[7] Glendinning, R.H. An evaluation of the ICM algorithm for image reconstruction. (42 pp) Submitted for publication.

[8] Jones, M.C., Silverman, B.W. & Wilson, J.D. A smoothed EM approach to a class of problems in image analysis and integral equations. (31 pp, 8 plates) Subsumed within [11] below.

[9] Johnstone, I.M. and Silverman, B.W. Speed of estimation in positron emission tomography. (44 pp) Submitted for publication.

[10] Friedman, J.H. and Silverman, B.W. Authors' response to the discussion of "Flexible parsimonious smoothing and additive modeling." (9 pp) To appear in *Technometrics*.

[11] Silverman, B.W., Jones, M.C, Wilson, J.D. & Nychka, D.W. A smoothed EM approach to a class of problems in image analysis and integral equations. (43 pp) Submitted for publication.

[12] Jones, M.C. & Silverman, B.W. An orthogonal series density estimation approach to reconstructing positron emission tomography images. (22 pp) Submitted for publication.

[13] Silverman, B.W. & Jones, M.C. E. Fix and J.L. Hodges (1951): an important unpublished contribution to nonparametric discriminant analysis and density estimation. (8 pp) Submitted for publication.

[14] Jubb, M. and Jennison, C. Aggregation and refinement in binary image restoration. (11pp) To appear in *Proceedings of AMS-IMS-SIAM Summer Research Conference on Spatial Statistics and Imaging, Brunswick, 1988.*

[15] Wilson, J.D. A statistical perspective on the solution of integral equations of the first kind. Ph.D. thesis, University of Bath, 1988. (225 pages).

Appendix 1

The bootstrap: to smooth or not to smooth?

by

B. W. Silverman and G. A. Young

Biometrika (1987), **74**, 3, pp. 469–79
Printed in Great Britain

The bootstrap: To smooth or not to smooth?

BY B. W. SILVERMAN
School of Mathematical Sciences, University of Bath, Bath BA2 7AY, U.K.
AND G. A. YOUNG
Statistical Laboratory, University of Cambridge, Cambridge CB2 1SB, U.K.

SUMMARY

The bootstrap and smoothed bootstrap are considered as alternative methods of estimating properties of unknown distributions such as the sampling error of parameter estimates. Criteria are developed for determining whether it is advantageous to use the smoothed bootstrap rather than the standard bootstrap. Key steps in the argument leading to these criteria include the study of the estimation of linear functionals of distributions and the approximation of general functionals by linear functionals. Consideration of an example, the estimation of the standard error in the variance-stabilized sample correlation coefficient, elucidates previously-published simulation results and also illustrates the use of computer algebraic manipulation as a useful technique in asymptotic statistics. Finally, the various approximations used are vindicated by a simulation study.

Some key words: Bootstrap; Computer algebra; Density estimation; Kernel; Resampling; Smoothed bootstrap.

1. INTRODUCTION

1·1. *The standard bootstrap*

The bootstrap is an appealing nonparametric approach to the assessment of errors and related quantities in statistical estimation. The method is described and explored in detail by Efron (1979, 1982). A typical context in which the bootstrap is used is in assessing the sampling mean squared error $\alpha(F)$ of an estimate $\hat{\theta}(X_1, \ldots, X_n)$ of a parameter $\theta(F)$ based on a sample $X_1, \ldots, X_n$ drawn from an unknown distribution F. If F were known, α might be most easily estimated by repeatedly simulating samples from F. The standard bootstrap technique is to estimate $\alpha(F)$ by the sampling method, but with the samples being drawn not from F itself but from the empirical distribution function F_n of the observed data $X_1, \ldots, X_n$. A sample from F_n is generated by successively selecting uniformly with replacement from $\{X_1, \ldots, X_n\}$ to construct a bootstrap sample $\{X_1^*, \ldots, X_n^*\}$. For each bootstrap sample, the estimate $\hat{\theta}(X_1^*, \ldots, X_n^*)$ of the quantity $\theta(F_n)$ is calculated. Since arbitrarily large numbers of bootstrap samples can be constructed, $\alpha(F_n)$ can easily be estimated to any reasonable required accuracy from the simulations. The quantity $\alpha(F_n)$ is then used as an estimate of $\alpha(F)$.

The bootstrap method thus consists of two main elements, which are often confused. There is first the idea of estimating a functional $\alpha(F)$ by its empirical version $\alpha(F_n)$, and secondly the observation that $\alpha(F_n)$ can in very many contexts be constructed by repeated resampling from the observed data. The resampling idea is an extremely important one, but it has, perhaps, been overstressed at the expense of the underlying estimation step. Once the two steps are conceptually separated it becomes easier to gain a fuller understanding of how the bootstrap actually works. In particular it becomes clear that there is nothing special about estimating functionals $\alpha(F)$ that are themselves sampling properties of parameter estimates; the bootstrap idea can be applied to any functional $\alpha(F)$ of interest.

1·2. *The smoothed bootstrap*

Because the empirical distribution F_n is a discrete distribution, samples constructed from F_n in the bootstrap simulations will have some rather peculiar properties. All the values taken by the members of the bootstrap samples will be drawn from the original sample values, and nearly every sample will contain repeated values. The smoothed bootstrap (Efron, 1979) is a modification to the bootstrap procedure to avoid samples with these properties. The essential idea of the smoothed bootstrap is to perform the repeated sampling not from F_n itself, but from a smoothed version $\hat{F}$ of F_n. Two possible versions of the smoothed bootstrap will be described in more detail below; whatever method of smoothing is used, the net effect of using the smoothed bootstrap is to estimate the functional $\alpha(F)$ by $\alpha(\hat{F})$.

The main aim of this paper is to investigate some properties of the smoothed bootstrap, in order to give some insight into circumstances when the smoothed bootstrap will give better results than the standard bootstrap. As an important by-product, the value of computer algebraic manipulation as a tool in asymptotic statistics will be demonstrated.

Efron (1982) considered the application of the bootstrap, and various other techniques, to the estimation of the standard error of the variance-stabilized transformed correlation coefficient. He illustrated by direct simulation that in a particular case a suitable smoothed bootstrap gave better estimates of standard error than the standard bootstrap. We shall discuss Efron's example later in the present paper and demonstrate how his results can be elucidated and extended by using a suitable approximation argument.

Before going on to discuss the estimation of general functionals $\alpha(F)$, we shall first consider the estimation of functionals α that are linear in F. For such functionals we shall obtain simple sufficient conditions under which using the smoothed bootstrap can decrease the mean squared error in the estimation of $\alpha(F)$.

We close this section by giving details of the two kinds of smoothed bootstrap considered in later discussion. Suppose $X_1, \ldots, X_n$ is a set of r-dimensional observations drawn from some r-variate density f and that V is the variance matrix of f, or a consistent estimator of this variance matrix, such as the sample variance matrix of the data. Choose a kernel function K such that K is a symmetric probability density function of an r-variate distribution with unit variance matrix, for example the standard unit r-variate normal density.

Define the kernel estimate $\hat{f}_h(x)$ of $f(x)$ by

$$\hat{f}_h(x) = |V|^{-\frac{1}{2}} n^{-1} h^{-r} \sum_{i=1}^{n} K\{h^{-1} V^{-\frac{1}{2}}(x - X_i)\}, \tag{1·1}$$

and the shrunk kernel estimate $\hat{f}_{h,s}(x)$ by

$$\hat{f}_{h,s}(x) = (1 + h^2)^{\frac{1}{2}r} \hat{f}_h\{(1 + h^2)^{\frac{1}{2}} x\}. \tag{1·2}$$

Density estimates in general are discussed, for example, by Silverman (1986). The smoothing parameter h determines the amount by which the data are smoothed to provide estimates. Estimates of the form (1·2) have the property that the density $\hat{f}_{h,s}$ has the same variance structure as the original data, if V is taken to be the sample variance matrix.

Given any functional $\alpha(F)$ of an r-variate distribution F, the unshrunk smoothed bootstrap estimate of $\alpha(F)$ is defined to be $\alpha(\hat{F}_h)$ and the shrunk smoothed bootstrap estimate is $\alpha(\hat{F}_{h,s})$, where $\hat{F}_h$ and $\hat{F}_{h,s}$ are the distribution functions corresponding to $\hat{f}_h$ and $\hat{f}_{h,s}$ respectively. It is easy to simulate either from $\hat{f}_h$ or from $\hat{f}_{h,s}$ by sampling with replacement from the original data and perturbing each sampled point appropriately; for details see Efron (1979) or Silverman (1986, § 6.4). Hence values of $\alpha(\hat{F}_h)$ and $\alpha(\hat{F}_{h,s})$ can be obtained in practice by simulation if necessary.

2. LINEAR FUNCTIONALS

In this section we consider the estimation of a linear functional $A(F)$. Because A is linear, standard calculus demonstrates the existence of a function $a(t)$ such that

$$A(F) = \int a(t)\, dF(t).$$

The standard bootstrap estimate $\hat{A}_0(F)$ will satisfy

$$\hat{A}_0(F) = A(F_n) = \int a(t)\, dF_n(t) = n^{-1} \sum_{i=1}^{n} a(X_i).$$

The unshrunk smoothed bootstrap estimate $\hat{A}_h(F)$ will satisfy

$$\hat{A}_h(F) = \int a(t) \hat{f}_h(t)\, dt,$$

and the shrunk smoothed bootstrap estimate $\hat{A}_{h,s}(F)$ will satisfy

$$\hat{A}_{h,s}(F) = \int a(t) \hat{f}_{h,s}(t)\, dt,$$

with $\hat{f}_h$ and $\hat{f}_{h,s}$ as defined in (1·1) and (1·2) above.

In the discussion that follows we assume that the function a has continuous derivatives of all orders required. *All unspecified integrals are taken over the whole of r-dimensional space.* Assume that V is fixed and define the differential operator D_V by

$$D_V a = \sum_{i=1}^{r} \sum_{j=1}^{r} V_{ij} \partial^2 a / \partial x_i\, \partial x_j .$$

Our first theorem gives a criterion for smoothing, without shrinkage, to be of potential value in the bootstrap estimation process.

THEOREM 1. *Suppose $a(X)$ and $D_V a(X)$ are negatively correlated. Then the mean squared error of $\hat{A}_h(F)$ can be reduced below that of $\hat{A}_0(F)$ by choosing a suitable $h > 0$.*

Proof. Assume without loss of generality that $A(F) = 0$, by replacing $a(t)$ by $a(t) - \int a(x) f(x)\, dx$ if necessary. By this assumption,

$$\text{MSE}\,\{\hat{A}_h(F)\} = E\{\hat{A}_h(F)^2\} = \text{var}\,\{\hat{A}_h(F)\} + [E\{\hat{A}_h(F)\}]^2. \qquad (2\cdot1)$$

Now, by some easy manipulations, $\hat{A}_h(F) = n^{-1} \sum w(X_i)$, say, where the sum is over $i = 1, \ldots, n$, and where

$$w(x) = \int a(t) h^{-r} |V|^{-\frac{1}{2}} K\{h^{-1} V^{-\frac{1}{2}} (t - x)\}\, dt = \int K(\xi) a(x + hV^{\frac{1}{2}} \xi)\, d\xi \qquad (2\cdot2)$$

on making the substitution $t = x + hV^{\frac{1}{2}}\xi$. A Taylor expansion gives

$$a(x+hV^{\frac{1}{2}}\xi) = a(x) + h(V^{\frac{1}{2}}\xi)^{\mathrm{T}}\{a(x)\}^2 + \tfrac{1}{2}h^2(V^{\frac{1}{2}}\xi)^{\mathrm{T}}H_a(x)(V^{\frac{1}{2}}\xi) + O(h^4),$$

where $H_a(x)_{ij} = \partial^2 a(x)/\partial x_i\,\partial x_j$. By our assumptions on the kernel K it follows that

$$w(x) = a(x) + \tfrac{1}{2}h^2 D_V a(x) + O(h^4), \tag{2·3}$$

$$E\{\hat{A}_h(F)\} = E\{w(X)\} = \tfrac{1}{2}h^2 \int f(x) D_V a(x)\,dx + O(h^4), \tag{2·4}$$

since $\int a(x)f(x)\,dx = 0$. Also, since $X_1, \ldots, X_n$ are independent,

$$n \operatorname{var}\{\hat{A}_h(F)\} = \operatorname{var}\{w(X)\} = \int a(x)^2 f(x)\,dx + h^2 \int a(x) D_V a(x) f(x)\,dx + O(h^4) \tag{2·5}$$

using (2·3). Combining (2·4) and (2·5) gives the mean squared error

$$\text{MSE}\{\hat{A}_h(F)\} = n^{-1}\int a(x)^2 f(x)\,dx + n^{-1}h^2 \int a(x) D_V a(x) f(x)\,dx + O(h^4). \tag{2·6}$$

For fixed n, the equation (2·6) demonstrates that, under the assumption that $a(X)$ and $D_V a(X)$ are negatively correlated, the mean squared error of $\hat{A}_h(F)$ will, at least for small h, be smaller than that of $\hat{A}_0(F)$, completing the proof of the theorem. □

The next theorem gives the corresponding criterion for smoothing with shrinkage to lead to more accurate bootstrap estimation. Define $a^*(X)$ by

$$a^*(X) = D_V a(X) - X.\nabla a(X).$$

THEOREM 2. *Suppose $a(X)$ and $a^*(X)$ are negatively correlated. Then the mean squared error of $\hat{A}_{h,s}(F)$ can be reduced below that of $\hat{A}_{0,s}(F) = \hat{A}_0(F)$ by choosing a suitable $h > 0$.*

Proof. As before assume without loss of generality that $A(F) = 0$. We have by similar manipulations to those used above, $\hat{A}_{h,s}(F) = n^{-1}\Sigma\, w^*(X_i)$, where

$$w^*(x) = (1+h^2)^{\frac{1}{2}r} \int a(t) h^{-r}|V|^{-\frac{1}{2}} K[h^{-1}V^{-\frac{1}{2}}\{x - (1+h^2)^{\frac{1}{2}}t\}]\,dt$$

$$= \int a\{(1+h^2)^{-\frac{1}{2}}(x + hV^{\frac{1}{2}}\xi)\} K(\xi)\,d\xi,$$

on making the substitution $t = (x + hV^{\frac{1}{2}}\xi)/(1+h^2)^{\frac{1}{2}}$. Now, for h small, $(1+h^2)^{-\frac{1}{2}} \simeq 1 - \frac{1}{2}h^2$, so

$$w^*(x) \simeq \int a(x + hV^{\frac{1}{2}}\xi - \tfrac{1}{2}h^2 x) K(\xi)\,d\xi.$$

A Taylor expansion of a about x, and our assumptions on the kernel K give

$$w^*(x) = a(x) + \tfrac{1}{2}h^2 a^*(x) + O(h^4). \tag{2·7}$$

Now we have

$$E\{\hat{A}_{h,s}(F)\} = E\{w^*(X)\} = \tfrac{1}{2}h^2 \int f(x) a^*(x)\,dx + O(h^4),$$

and, on using (2·7),

$$n \operatorname{var}\{\hat{A}_{h,s}(F)\} = \int a(x)^2 f(x)\,dx + h^2 \int a(x)a^*(x)f(x)\,dx + O(h^4).$$

The proof of Theorem 2 is completed in the same way as that of Theorem 1. □

As a simple illustration, consider the estimation of the sixth moment $\int x^6 f(x)\,dx$ of a univariate density. It is not immediately clear whether smoothing is worthwhile in this case. In the notation used above, $a(x) = x^6$, $D_V a(x) = 30\,Vx^4$ and $a^*(x) = 30\,Vx^4 - 6x^6$. It follows that, setting $\mu_r = EX^r$,

$$\operatorname{cov}\{a(X), a^*(X)\} = -6\mu_{12} + 30\,V\mu_{10} + 6\mu_6^2 - 30\,V\mu_4\mu_6.$$

If, for example, X has a normal distribution with mean zero and variance V, we have $\mu_{2j} = V^j 2^{-j}(2j)!/j!$ and hence $\operatorname{cov}\{a(X), a^*(X)\} = -34020\,V^6 < 0$.

Thus a shrunk-smoothed estimate $\int x^6 \hat{f}_{h,s}(x)\,dx$ will always, for a suitably chosen value of h, give a more accurate estimate of $E(X^6)$ than will the raw sixth moment if X is drawn from a normal distribution. Similar calculations for other distributions show that the same conclusion holds under a wide variety of distributional assumptions for X.

The results obtained by applying the criteria can sometimes be a little surprising. Suppose X is drawn from a standard normal distribution. Application of the criterion for estimation by unshrunk smoothing demonstrates that, for small h, this will have a deleterious effect in the estimation of either $E(X^4)$ or $E(X^2)$ alone. However, for the linear combination of moments $E(X^4 - cX^2)$, unshrunk smoothing will be worth performing provided $c > 6$. Details of this somewhat counter-intuitive result are left to the reader to reconstruct.

We do not, in this paper, devote much attention to the question of how much smoothing should be applied in cases where smoothing is worth performing; that problem is left for future work. However, the last example of this section demonstrates that the question of how much to smooth can be a rather delicate one. In this example, let ϕ_σ denote the density of the normal distribution with mean zero and variance σ^2. Let

$$A_\varepsilon(F) = \int \phi_\varepsilon(t)\,dF(t),$$

and suppose that the quantity ε converges to zero as the sample size increases. Assume that F has a smooth density f with derivatives of all orders required. Consider the estimation of $A_\varepsilon(F)$ by the unshrunk smoothed estimator $\hat{A}_h(F)$, constructed using the normal density as the kernel. We shall investigate the optimal large-sample behaviour of the smoothing parameter h. Assume throughout that h is small for large n and that $f(0) > 0$.

Setting $\delta^2 = h^2 + \varepsilon^2$ and performing some simple manipulations, we have

$$\hat{A}_h(F) = \int \phi_\varepsilon(t)\hat{f}_h(t)\,dt = n^{-1}\sum_i \phi_\delta(X_i).$$

Hence, substituting $u = t\delta$ and performing a Taylor series expansion,

$$E\{\hat{A}_h(F)\} = \int \phi_\delta(t)f(t)\,dt = \int \phi(u)f(u\delta)\,du = f(0) + \tfrac{1}{2}\delta^2 f''(0) + O(\delta^4).$$

Since, by a similar argument,

$$A_\varepsilon(F) = \int \phi_\varepsilon(t)f(t)\,dt = f(0) + \tfrac{1}{2}\varepsilon^2 f''(0) + O(\varepsilon^4),$$

it follows that

$$E\{\hat{A}_h(F)\} - A_r(F) = \tfrac{1}{2}h^2 f''(0) + O(\delta^4).$$

By standard arguments

$$\operatorname{var}\{\hat{A}_h(F)\} = n^{-1}\operatorname{var}\{\phi_\delta(X)\} = n^{-1}f(0)/(2\delta\pi^{\frac{1}{2}})\{1+O(\delta)\}.$$

Thus the mean squared error of $\hat{A}_h(F)$ will be, asymptotically, given by

$$\text{MSE}\,\{\hat{A}_h(F)\} \simeq n^{-1}f(0)/(2\delta\pi^{\frac{1}{2}}) + \tfrac{1}{4}h^4 f''(0)^2 = n^{-1}f(0)/(2\delta\pi^{\frac{1}{2}}) + \tfrac{1}{4}(\delta^2-\varepsilon^2)^2 f''(0)^2,$$

where the terms neglected are of order $n^{-1}+\delta^6$. This approximate mean squared error is a convex function of δ, and its minimizer will satisfy $\delta^3(\delta^2-\varepsilon^2) = C(f)n^{-1}$, where $C(f) = f(0)/\{2\pi^{\frac{1}{2}}f''(0)^2\}$, or, in terms of h and ε,

$$(1+h^2/\varepsilon^2)^{3/2}h^2/\varepsilon^2 = C(f)n^{-1}\varepsilon^{-5}. \tag{2·8}$$

Denote by $\psi(R)$ the root in $[0,\infty)$ of the equation

$$(1+\psi^2)^{3/2}\psi^2 = R;$$

then by simple calculus $\psi(R) \sim R^{\frac{1}{2}}$ as $R \to 0$, and $\psi(R) \sim R^{1/5}$ as $R \to \infty$. The asymptotically optimal h for the estimation of A_r will satisfy, from (2·8),

$$h_{\text{opt}} = \varepsilon\psi\{C(f)n^{-1}\varepsilon^{-5}\}.$$

If $n^{-1}\varepsilon^{-5} \to \infty$ then $h_{\text{opt}} \sim \varepsilon C(f)^{1/5}n^{-1/5}\varepsilon^{-1} = C(f)^{1/5}n^{-1/5}$.

Standard density estimation theory (Parzen, 1962) shows that this is the asymptotically optimal smoothing parameter for the estimation of the density at zero. Thus, in this case, the best estimate of A_r will be based on the best estimate of the density.

Unfortunately this will by no means always be the case. If $n^{-1}\varepsilon^{-5} \to 0$, we will have

$$h_{\text{opt}} \sim \varepsilon C(f)^{\frac{1}{2}}n^{-\frac{1}{2}}\varepsilon^{-5/2} = C(f)^{\frac{1}{2}}n^{-\frac{1}{2}}\varepsilon^{-3/2}$$

and if $n^{-1}\varepsilon^{-5} \to a$, where $0 < a < \infty$, $h_{\text{opt}} \sim \varepsilon\psi\{aC(f)\}$.

In neither of these cases will it be optimal to construct an optimal estimate of f in order to estimate $A_r(f)$, since the optimal choice of h will be smaller, in order of magnitude in the first case, than that required for the estimation of f itself. Thus the optimal estimate of $A_r(F)$ will be based on an undersmoothed estimate of the underlying density. This example is, of course, rather artificial, but it does illustrate the likely difficulty of obtaining general rules for deciding how much to smooth when estimating functionals of a density. Even in cases where smoothing is advantageous, the amount of smoothing required may be quite different from that needed for the estimation of the density itself.

3. MORE GENERAL FUNCTIONALS

3·1. *Linear approximation*

In this section, the work of § 2 is extended, by considering local linear approximations, to more general functionals of an unknown distribution. When an explicit bootstrap method is being used the functional being estimated is unlikely to be linear, and so a more general theory is necessary. Local linear approximations to functionals of distributions have also been used by Hinkley & Wei (1984) and Withers (1983).

Consider the estimation of a functional $\alpha(F_0)$ of an unknown distribution F_0 underlying a set of sample data. Suppose that α admits a linear von Mises expansion about F_0 given by

$$\alpha(F) \simeq \alpha(F_0) + A(F - F_0), \tag{3·1}$$

where the linear functional A is representable as an integral

$$A(F - F_0) = \int a(t)\, d(F - F_0)(t). \tag{3·2}$$

A detailed discussion of differentiation of functionals and general von Mises approximation is given by Fernholz (1983). The precise accuracy of the expansion (3·1) depends on the detailed properties of α, but the error will in general be of order $\sup|F - F_0|^2$.

The expansion (3·1) gives an obvious approximation to the bootstrap estimate of $\alpha(F_0)$. If $\hat{F}$ is an estimate of F_0, then we will in general have, provided $\sup|\hat{F} - F_0|$ is $O_p(n^{-\frac{1}{2}})$,

$$\alpha(\hat{F}) = \alpha(F_0) + A(\hat{F}) - A(F_0) + O_p(n^{-1}),$$

and so the sampling properties of $\alpha(\hat{F})$ will be approximately the same as those of $A(\hat{F})$. The criteria of § 2 can then be applied to the linear functional A. If using an appropriate smoothed bootstrap will improve the estimation of $A(F_0)$ then, neglecting any errors inherent in the linear approximation (3·1), the smoothed bootstrap will be worth using in the estimation of $\alpha(F_0)$.

3·2. *The transformed sample correlation coefficient*

In this section we consider application of the linear approximation procedure to estimation of the sampling standard deviation of the variance-stabilized sample correlation coefficient. Suppose F_0 is a bivariate distribution with mean zero and correlation coefficient ρ, and let $\zeta = \tanh^{-1}\rho$. Let r be the computed sample correlation coefficient based on a sample of n independent observations from F_0, and let $z = \tanh^{-1} r$ be the sample estimate of ζ. Then the functional of interest is $\alpha_n(F_0) = \{\mathrm{var}(z)\}^{\frac{1}{2}}$. Efron (1982) devoted considerable attention to the estimation of $\alpha_n(F_0)$ by a variety of methods, including the smoothed bootstrap, for the specific case of F_0 bivariate normal, with marginals of unit variance and $\rho = \frac{1}{2}$, and for sample size $n = 14$.

A key step in our investigation of the estimation of $\alpha_n(F_0)$ will be an approximate formula, given by Kendall & Stuart (1977, p. 251). Let

$$\alpha(F_0) = \left[\frac{\rho^2}{(1-\rho^2)^2}\left\{\frac{\mu_{22}}{\mu_{11}^2} + \frac{1}{4}\left(\frac{\mu_{40}}{\mu_{20}^2} + \frac{\mu_{04}}{\mu_{02}^2} + \frac{2\mu_{22}}{\mu_{20}\mu_{02}}\right) - \left(\frac{\mu_{31}}{\mu_{11}\mu_{20}} + \frac{\mu_{13}}{\mu_{11}\mu_{02}}\right)\right\}\right]^{\frac{1}{2}}, \tag{3·3}$$

where μ_{ij} is the (i, j)th moment given by $\mu_{ij} = \int x_1^i x_2^j \, dF_0(x)$. Here and subsequently in this section unsubscripted letters x will denote vectors (x_1, x_2). Kendall & Stuart give

$$\alpha_n(F_0) = n^{-\frac{1}{2}}\alpha(F_0) + O(n^{-3/2}),$$

so that estimation of $\alpha_n(F_0)$ is approximately equivalent to that of $\alpha(F_0)$.

Consider now the calculation of the function $a(t)$ defined in (3·2). For fixed t let δ_t be the distribution function of a point mass at t and, for any $\varepsilon > 0$ let F_ε be the improper distribution $F_0 + \varepsilon\delta_t$. Then simple calculus combining (3·1) and (3·2) gives

$$a(t) = [(d/d\varepsilon)\alpha(F_\varepsilon)]_{\varepsilon=0}. \tag{3·4}$$

Our functional $\alpha(F)$ is defined for improper distributions, as well as for probability distribution functions, and hence there is no need when calculating $a(t)$ to consider the more complicated perturbation $\varepsilon(\delta_t - F_0)$ to F_0 used by Hinkley & Wei (1984). The actual algebraic manipulations required in the calculation of $a(t)$ from (3·4) and (3·3) are extremely laborious. However, it is relatively easy to write a program in a computer algebraic symbolic manipulation language, such as MACSYMA, to perform the necessary differentiations and substitutions. The function $a(t)$ itself is a fourth-order polynomial in t_1 and t_2 whose coefficients depend on the moments of F_0. It is only used as an intermediate step, in the special cases considered below, in the calculation of the criteria derived from Theorems 1 and 2, and the calculation of these criteria was also performed by computer algebra. Further details of the manipulations are available from the authors.

To complete this section we consider the results of the application of the computer algebraic manipulation procedure to the functional (3·3) for two special cases. Further details of the results discussed will be given in § 3·3 below. Let $A_{SB}(F_0)$ be the criterion obtained from Theorem 2 for the shrunk smoothed bootstrap to be advantageous in the estimation of the functional $A(F_0)$. Recall that $A_{SB}(F_0) < 0$ means that some smoothing at least is worthwhile.

Suppose, first, that the distribution of the data can be reduced by an affine transformation to a radially symmetric distribution $F^\dagger$. Without loss of generality it can be assumed that $F^\dagger$ has unit marginal variances. Let R be the radial component of $F^\dagger$, and denote by s_j the jth central moment of R^2. Computer algebra shows that the criterion $A_{SB}(F_0)$ reduces, in this case, to

$$A_{SB}(F_0)\beta_0^{-2} = -\{3s_4 + (4-3s_2)s_3 + s_2^3 + 2s_2^2 + 24s_2 + 16\}/32, \tag{3·5}$$

where β_0 is the positive quantity $\frac{1}{2}\alpha(F_0)^{-1}$. Using the standard inequality $s_3^2 \leqslant s_2 s_4$, we have

$$\begin{aligned}-32A_{SB}(F_0)\beta_0^{-2} &\geqslant 3s_4 - 4s_4^{\frac{1}{2}}s_2^{\frac{1}{2}} - 3s_4^{\frac{1}{2}}s_2^{3/2} + s_2^3 + 2s_2^2 + 24s_2 + 16\\ &= 3(s_4^{\frac{1}{2}} - \tfrac{1}{2}s_2^{3/2} - 2s_2^{\frac{1}{2}}/3)^2 + \tfrac{1}{4}s_2^3 + 68s_2/3 + 16 \geqslant 16.\end{aligned}$$

This gives the general conclusion that $A_{SB}(F_0) \leqslant -\frac{1}{2}\beta_0^2$ for any distribution F_0 which can be affinely transformed to radial symmetry.

Another class of distributions for which $A_{SB}(F_0)$ is guaranteed not to be positive is the class for which a particular affine transformation of F_0 to unit variance-covariance matrix yields a distribution with independent marginals. Let X be a random vector with distribution F_0, and let $\sigma_1^2 = \mathrm{var}(X_1)$, $\sigma_2^2 = \mathrm{var}(X_2)$ and $\rho = \mathrm{corr}(X_1, X_2)$. Define a matrix S by

$$S = \begin{bmatrix}\sigma_1 & 0\\ 0 & \sigma_2\end{bmatrix}\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}^{\frac{1}{2}}; \tag{3·6}$$

here the power $\frac{1}{2}$ denotes the symmetric positive-definite square root. Define a bivariate distribution F^* by $F^*(u) = F_0(Su)$ for all 2-vectors u. A random vector $Y = S^{-1}X$ with distribution F^* and unit variance-covariance matrix can be obtained by first rescaling the marginals of X to have unit variance and then rescaling the principal components of the resulting vector to have unit variances. If this natural affine transform of F_0 has independent marginals, then an argument given in § 3·3 below demonstrates that $A_{SB}(F_0) \leqslant 0$, with equality only if X has a uniform discrete distribution giving probability $\frac{1}{4}$ to each of four points.

In summary, we have derived the following conclusion. Provided all the approximations we have made are reasonable, using a shrunk smoothed bootstrap, with an appropriate smoothing parameter, will give improved estimation of $\alpha_n(F_0)$ over that obtained by the standard bootstrap, if either F_0 is an affine transformation of a radially symmetric distribution or F_0 is an affine transformation, of a particular kind, of a distribution with independent marginals, and F_0 is not a uniform four-point discrete distribution. In practice the underlying distribution F_0 will not be known. An obvious topic for future investigation is the construction of empirical versions of the criteria of Theorems 1 and 2, on the basis of which a decision whether or not to smooth can be made for each data set encountered. Some preliminary simulations along these lines have been encouraging.

3·3. *Some technical details*

Throughout this section, define the matrix S as in (3·6), and suppose that X is a random vector with distribution F_0. Let $Y=S^{-1}X$ as in § 3·2, and let $F^*(y)=F_0(Sy)$ be the distribution of Y. It is easily seen that the existence of an affine transformation reducing F_0 to radial symmetry is equivalent to the radial symmetry of the particular affine transformation F^*.

Define $a_S(u)=a(Su)$ and let $k_{ij}=E(Y_1^i Y_2^j)$, where $Y=S^{-1}X$. In both of the two special cases considered in § 3·2, $k_{13}=k_{31}=0$, and computer algebraic manipulation showed that $a_S(u)$ reduces to the simple form

$$a_S(u)=\{u_1^2u_2^2-k_{22}(u_1^2+u_2^2)\}\beta_0.$$

The criterion given in Theorem 2 also reduces to a simple form when expressed in terms of a_S. We have, by standard calculus,

$$a^*(X)=D_V a(X)-X.\nabla a(X)=\nabla^2 a_S(Y)-Y.\nabla a_S(Y)=a_S{}^*(Y),$$

say, where $a_S^*(u)=\{2(1+k_{22})(u_1^2+u_2^2)-4k_{22}-4u_1^2u_2^2\}\beta_0$.

Since, by definition, $a(X)=a_S(Y)$, it follows that

$$\begin{aligned} A_{SB}(F_0)&=\operatorname{cov}\{a(X),a^*(X)\}=\operatorname{cov}\{a_S(Y),a_S^*(Y)\}\\ &=E\{a_S(Y)+\beta_0k_{22}\}a_S^*(Y) \end{aligned} \tag{3·7}$$

since it is immediate that $E\{a_S(Y)\}=-\beta_0k_{22}$.

Suppose, now, that the distribution of Y is radially symmetric, so that $Y^{\mathrm{T}}=(R\cos\Theta,R\sin\Theta)$ with Θ uniformly distributed on $(0,2\pi)$. The form (3·7) for $A_{SB}(F_0)$ can be expressed in terms of even moments of Y up to order 8, and each of these moments can be expressed in terms of the moments of R^2. For example

$$k_{22}=E(R^4\sin^2\Theta\cos^2\Theta)=E(R^4/8)=(s_2+4)/8,$$

where, as in § 3·2, $s_j=E(R^2-2)^j$ is the jth central moment of R^2; the assumption that $E(Y_1^2)=E(Y_2^2)=1$ implies that R^2 has mean 2. Performing all these substitutions, by computer algebra, yields the form (3·5) for $A_{SB}(F_0)$ and hence the conclusion given in § 3·2 for distributions that can be transformed to radial symmetry.

Now suppose that Y_1 and Y_2 are independent, but that Y is not necessarily radially symmetric. It will then be the case that $k_{22}=E(Y_1^2)E(Y_2^2)=1$ and hence

$$a_S^*(u)=-4\beta_0(u_1^2u_2^2-u_1^2-u_2^2+1)=-4\{a_S(u)+\beta_0\}.$$

It follows that $A_{SB}(F_0)=-4\operatorname{var}\{a_S(y)\}$. Since Y_1 and Y_2 are independent, the only way $\operatorname{var}\{a_S(y)\}$ can be zero is for Y to have the four point distribution giving probability $\frac{1}{4}$ to each of the points $(\pm1,\pm1)$; otherwise $a_S(Y)$ has positive variance, and $A_{SB}(F_0)<0$.

4. SIMULATION STUDY

The discussion in § 3 above involved heavy dependence on two approximations, one of them specific to the example under consideration and the other a key feature of our proposed general methodology. In this section, we investigate both of these approximations by a simulation study which extends the one carried out by Efron (1982, Table 5.2). All our simulations are carried out under the assumptions of Efron's model, that F_0 is the bivariate normal distribution with unit marginal variances and correlation $\frac{1}{2}$. Efron considered only samples of size 14, though we consider here larger sample sizes as well. We follow Efron in using the values 0 and $\frac{1}{2}$ for the smoothing parameter h.

For each sample size n, the accuracy of the bootstrap and smoothed bootstrap estimates of the sampling standard deviation $\alpha_n(F_0)$ of the variance-stabilized correlation coefficient was assessed in three different ways. First, a direct simulation of the bootstrap procedure itself was carried out; two hundred data sets were generated from F_0 and for each one $\alpha_n(F_0)$ was estimated by the usual resampling procedure, using two hundred resampled data sets of size n in each case. The true value of $\alpha_n(F_0)$ is known and so it is possible to estimate the root mean squared error of the direct bootstrap procedures from our simulations. The results thus obtained are labelled 'direct' in Table 1.

Table 1. *Estimates of root mean squared errors of bootstrap estimates of sampling standard deviations of variance-stabilized and untransformed correlation coefficients; sample sizes n and smoothing parameters h.*

		Variance-stabilized			Untransformed		
n	h	Direct	Linear	Delta	Direct	Linear	Delta
14	0	0·075	0·071	0·077	0·070	0·076	0·060
	½	0·045	0·046	0·057	0·057	0·055	0·052
20	0	0·049	0·050	0·053	0·046	0·053	0·044
	½	0·033	0·032	0·037	0·045	0·039	0·041
30	0	0·029	0·033	0·033	0·033	0·036	0·030
	½	0·019	0·021	0·022	0·027	0·026	0·027
40	0	0·024	0·025	0·025	0·024	0·027	0·027
	½	0·015	0·016	0·017	0·021	0·019	0·020
50	0	0·020	0·020	0·021	0·020	0·021	0·019
	½	0·013	0·013	0·014	0·019	0·015	0·018
100	0	0·011	0·010	0·010	0·010	0·011	0·010
	½	0·008	0·006	0·007	0·009	0·008	0·008

Secondly, in order to investigate the accuracy of our linear approximation $\hat{A}_{h,s}(F_0)$, some analytic calculations were carried out, making use of computer algebra. By this means, the behaviour of the approximation can be studied without recourse to any simulation. For the bivariate normal population under consideration, the standard deviation of $\hat{A}_{h,s}(F_0)$ was found to be $n^{-1}(1+h^2)^{-2}$. This quantity is referred to as the 'linear' estimate of the root mean squared error of the bootstrap procedure. Closeness of the 'linear' and 'direct' estimates of root mean squared error would vindicate our proposed procedure of studying the sampling properties of the bootstrap by means of linear approximations.

Our development of the linear approximation involved the intermediate step of approximating $\alpha_n(F_0)$ by $n^{-\frac{1}{2}}\alpha(F_0)$, as given in (3·3). This intermediate approximation raises the possibility of studying the sampling properties of the smoothed bootstrap by considering those of the approximation (3·3), with F_0 replaced by $\hat{F}_{h,s}$. This corresponds to substituting the moments of $\hat{F}_{h,s}$, which are easily calculated from the sample, into (3·3). By analogy with § 6.5 of Efron (1982), we refer to this procedure as the nonparametric delta approximation to the smoothed bootstrap. For each of two hundred simulated samples from F_0 this approximation was calculated. From the values thus obtained, a third estimate of the root mean squared error of the smoothed bootstrap procedure was found. This is labelled 'delta' in Table 1.

The analogous investigation was carried out for the untransformed correlation coefficient r, in the context of the same bivariate normal model. The factor $(1-\rho^2)^{-2}$ is omitted from (3·3) in this case; otherwise the same algebraic manipulations and simulations were performed as for the variance-stabilized coefficient z. The 'linear' estimate of the root mean squared error is now $\frac{3}{4}n^{-1}(1+h^2)^{-2}(2+2h^2+h^4)^{\frac{1}{2}}$. The results are presented in the last three columns of Table 1.

The broad conclusions to be drawn from Table 1 are the same for both correlation coefficients. Even for the small sample size considered by Efron (1982), our linear approximation procedure gives good estimates of the accuracy of the full bootstrap procedure, and the relative improvement due to smoothing is well predicted. Efron's conclusions could have been obtained without recourse to any simulation. On the whole the delta procedure, which itself involves some simulation, gives slightly inferior estimates of the bootstrap's accuracy.

It is known (Davison, Hinkley & Schechtman, 1986) that the variance-stabilized correlation coefficient is highly correlated with its linear approximation, but the untransformed correlation coefficient is in general not. The suspicion expressed by a referee that this may have a deleterious effect on our approximations in the untransformed case does not appear to have been borne out by the simulation study, except that the beneficial effects of smoothing the bootstrap were systematically slightly exaggerated by the linear method in this case.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the referee's helpful suggestions and the financial support of the Science and Engineering Research Council and the U.S. Army European Research Office.

REFERENCES

DAVISON, A. C., HINKLEY, D. V. & SCHECHTMAN, E. (1986). Efficient bootstrap simulation. *Biometrika* **73**, 555-66.

EFRON, B. (1979). Bootstrap methods; another look at the jackknife. *Ann. Statist.* **7**, 1-26.

EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* Philadelphia: SIAM.

FERNHOLZ, L. T. (1983). *Von Mises Calculus for Statistical Functionals.* New York: Springer Verlag.

HINKLEY, D. V. & WEI, B.-C. (1984). Improvements of jackknife confidence limit methods. *Biometrika* **71**, 331-9.

KENDALL, M. G. & STUART, A. (1977). *The Advanced Theory of Statistics*, **1**. 4th ed. London: Griffin.

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065-76.

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* London; Chapman and Hall.

WITHERS, C. S. (1983). Expansions for the distribution and quantiles of a regular functional of the empirical distribution with applications to nonparametric confidence intervals. *Ann. Statist.* **11**, 577-87.

[*Received September* 1986. *Revised October* 1986]

Appendix 3

Statistical image restoration and refinement.

by

C. Jennison and M. Jubb

STATISTICAL IMAGE RESTORATION AND REFINEMENT

Christopher Jennison and Michael Jubb

School of Mathematical Sciences

University of Bath, BATH BA2 7AY, U.K.

SUMMARY

We consider the problem of reconstructing an image from a noisy record. We describe existing methods due to Geman and Geman (1984) and Besag (1986) which use a Markov random field model for the true scene but assume that each pixel consists of a single colour. In order to improve the quality of the restoration at the boundary of regions of different colours we extend these methods to allow pixels to contain two regions of colour separated by a single straight line. An algorithm for performing the reconstruction is presented and illustrated by an example.

INTRODUCTION

We consider a rectangular region partitioned into pixels labelled $1, 2, \ldots, n$. Each pixel is coloured black or white and the colour of pixel i is denoted by x_i which takes the value 0 for white and 1 for black. The x_i are unobserved. We work instead from the observed record y_i which consists of x_i plus added noise. We denote the whole scene by $x = \{x_i;\ i=1,\ldots,n\}$ and the set of records by $y = \{y_i;\ i=1,\ldots,n\}$. The noise distribution will be assumed to be known but if this were not the case, it could be established by studying training data.

Recent developments in statistical restoration methods use a Bayesian approach. The maximum a posteriori (MAP) estimate of the true scene is the value of x which maximises $P(x|y)$, the conditional probability of x given the record y. By Bayes' theorem

$$P(x|y) \propto l(y|x)\, p(x), \qquad (1)$$

where $l(y|x)$ is the conditional likelihood of the observed record, y, given the true colouring, x, and $p(x)$ is the prior probability of x.

We assume the conditional density function $f(y_i|x_i)$ to be known and for the remainder of this paper we shall assume that the records, y_i, are independently distributed as Gaussian with mean x_i and variance σ^2. Thus,

$$l(y|x) = \prod_{i=1}^{n} f(y_i|x_i) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\{ \frac{-1}{2\sigma^2} \sum_{i=1}^{n} (y_i - x_i)^2 \} .$$

To obtain a valid formula for $p(x)$, we assume that the true scene corresponds to a locally dependent Markov random field (MRF) with respect to a specified neighbourhood system, that is, the conditional distribution of pixel i given the colourings of all other pixels' depends *only* on the neighbours of pixel i. We shall use a second order neighbourhood system in which pixels are considered to be neighbours if they are horizontally, vertically or diagonally adjacent to each other. A detailed definition and further examples of Markov random fields may be found in Besag (1974).

The form of $p(x)$ is determined by the nature of the Markov random field. In our case, we have

$$p(x) \propto e^{-\beta Z(x)} ,$$

where $Z(x)$ is the number of discrepant pairs in the scene, x, i.e. the number of pairs of neighbours which are of opposite colour, and β is a fixed positive constant (normally chosen to be between 0.5 and 1.5).

The maximisation of $P(x|y)$ now corresponds to the minimisation of

$$\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - x_i)^2 + \beta Z(x) \qquad (2)$$

over values of $x = \{x_i ; i=1,\ldots,n\}$.

This expression may be regarded as a penalty, the first term penalising any difference between the record and the fitted value, the second term penalising excessive roughness in the reconstruction. Clearly, with 2^n possible values for x this is a computationally large problem and necessitates the use of a sophisticated algorithm.

Geman and Geman (1984) use the method of simulated annealing which attempts to find the MAP estimate of x given the record y. Their method is computationally extravagant and more recent developments by Greig, Porteous and Seheult (1986) show that the MAP estimate of any two colour scene may be found exactly using the Ford-Fulkerson labelling algorithm for maximising flow through a network.

Besag (1986) proposed the computationally simpler method of iterated conditional modes (ICM) which updates each pixel in turn, choosing for it the most likely colour based on its record and the current colouring of its neighbours. In updating pixel i the new x_i is chosen to minimise the sum of terms involving x_i in the penalty (2), i.e.

$$\frac{1}{2\sigma^2}(y_i - x_i)^2 + \beta Z(x_i)$$

where $Z(x_i)$ is the number of neighbours of pixel i in the current restoration which are of the opposite colour to x_i. The method proceeds by scanning the scene, successively updating each pixel until convergence is reached. This will normally occur at a local rather than global maximum of $P(x|y)$, but, given the possibility of undesirable long range dependencies in the MRF model, this is not a serious drawback and might even be an advantage.

SPLIT PIXELS

So far we have considered scenes in which each pixel is coloured wholly one colour. We now allow pixels in the true scene to be coloured partly black and partly white. Each record y_i is distributed as Gaussian with variance σ^2 and mean p_i, the proportion of pixel i which is coloured black. The restoration methods that we have previously discussed can be used for this problem by proceeding as if the pixels were only of one colour but the quality of the restoration at the edges of objects or regions will obviously be poor. Instead, we can allow pixels in the restored image to be coloured partly black and partly white. The simplest form of this is to quarter each pixel and allow it to be filled with the most likely of the 2^4 configurations. This method, proposed by Jennison (1986) uses a modified version of ICM, firstly iterating at full pixel size and subsequently restoring the quarters; in the second stage the same form of MRF model is used for the subpixels as is originally used for full pixels This method appears to work well and has prompted work into the further breakdown of pixels.

For further refinement we can either (i) consider an $m \times m$ breakdown of each pixel or (ii) use continuous lines within the pixel to represent the edge. The implementation of (i) requires the minimisation of

$$\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \frac{1}{m^2}\sum_{j=1}^{m}\sum_{k=1}^{m} x_{ijk})^2 + \tfrac{1}{2}\beta\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{m} Z(x_{ijk}),$$

where the subscript ijk refers to subpixel j,k within pixel i; x_{ijk} takes value 0 or 1 and $Z(x_{ijk})$ is the number of subpixel neighbours of subpixel ijk in the current restoration which are of the opposite colour to x_{ijk} (the factor $\frac{1}{2}$ is needed as each discordant pair is counted twice). Note that subpixels at the edge of a pixel will have some subpixel neighbours contained in an adjacent pixel. We can see that as m increases this minimisation becomes computationally cumbersome. Also, it offers only an approximation to (ii) and it turns out to be easier to pass to the limit and work directly with continuous solutions.

The most basic form of (ii) allows a single straight line edge within each pixel and it is the implementation of this that we shall describe. It is no longer meaningful to talk of discrepant pixel or subpixel pairs and we replace the second term of (2) by a multiple of the total length of edge in the reconstruction x. Thus, the restored image is chosen to minimise

$$\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - p_i(x))^2 + \beta' L(x), \tag{3}$$

over images x made up of pixels x_i, $i=1,\ldots,n$, either of a single colour or divided into two regions of different colours by a single straight line; $p_i(x)$ denotes the proportion of black in pixel i; $L(x)$ is the total edge length in scene x and β' is a fixed constant related to the β used earlier.

An advantage of edge length as a measure is that the penalty is rotationally invariant, i.e. remains constant throughout all rotations of the scene within the region. This could not be obtained using discrepant pairs as a measure although it has been shown by our colleague Robin Sibson that this variability can be minimised using a down weighting of $1/\sqrt{2}$ for the diagonal adjacencies.

THE RESTORATION ALGORITHM

The restoration is done in three stages, the first two of which have already been described :

Stage 1 : ICM to convergence on full size pixel grid.

Stage 2 : ICM to convergence on 2×2 pixel grid.

Stage 3 : Updating process on the line segments representing the edges.

Stage 3 requires that we now regard the reconstruction as a series of line segments separating the two colours. An initial representation is obtained in a straightforward way from the end product of Stage 2. The updating process treats pixels in pairs, selecting the best place for two edges to meet, given the current restoration of neighbouring pixels.

As an example, consider the configuration at pixels i and j shown in Figure 1. The distances a and b are determined by the current colouring of neighbouring pixels and treated as constant for the moment. The distance W is chosen to minimise the contribution from pixels i and j to the total penalty (3), i.e.

$$g(W) = \frac{1}{2\sigma^2}\sum_{k=i,j}(y_k - p_{kW})^2 + \beta'(e_{iW} + e_{jW}), \tag{4}$$

where e_{kW} is the length of edge in pixel k when the join is at W and p_{kW} is the proportion of black in pixel k when the join is at W.

For the case shown in Figure 1, this penalty is

$$g_1(W) = \frac{1}{2\sigma^2}\{(y_i - a - \tfrac{1}{2}(W-a))^2 + (y_j - b - \tfrac{1}{2}(W-b))^2\}$$
$$+\ \beta'\{\sqrt{1+(W-a)^2} + \sqrt{1+(W-b)^2}\}.$$

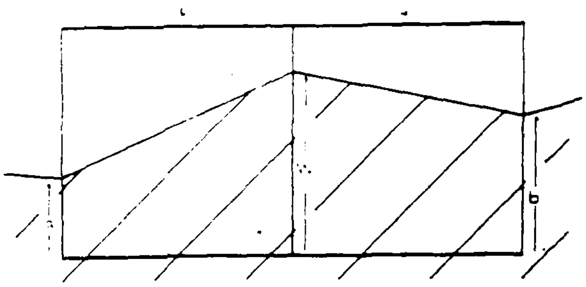


Fig. 1. Updating the position of edges in pixels i and j.

This can not be minimised directly but the form of

$$\frac{dg_1(W)}{dW} = \frac{1}{4\sigma^2}(2W+a-2y_i+b-2y_j) + \beta'\left[\frac{(W-a)}{\sqrt{1+(W-a)^2}}+\frac{(W-b)}{\sqrt{1+(W-b)^2}}\right]$$

suggests an iterative approach. Given an approximate solution W_{s-1} we solve

$$\frac{1}{4\sigma^2}(2W_s+a-2y_i+b-2y_j) + \beta'\left[\frac{(W_s-a)}{\sqrt{1+(W_{s-1}-a)^2}}+\frac{(W_s-b)}{\sqrt{1+(W_{s-1}-b)^2}}\right] = 0$$

to obtain

$$W_s = \frac{4\sigma^2\beta'\left[\dfrac{a}{\sqrt{1+(W_{s-1}-a)^2}}+\dfrac{b}{\sqrt{1+(W_{s-1}-b)^2}}\right] + (2y_i-a+2y_j-b)}{2+4\sigma^2\beta'\left[\dfrac{1}{\sqrt{1+(W_{s-1}-a)^2}}+\dfrac{1}{\sqrt{1+(W_{s-1}-b)^2}}\right]}.$$

Starting from any sensible initial value, W_0, accuracy to 3 decimal places was achieved after at most four iterations. In practice we take W_0 to be the value of W prior to this update.

Different forms of (4) are possible depending on which neighbours of pixels i and j contain both colours. There are only four distinct cases that may arise and these are shown in Figure 2.

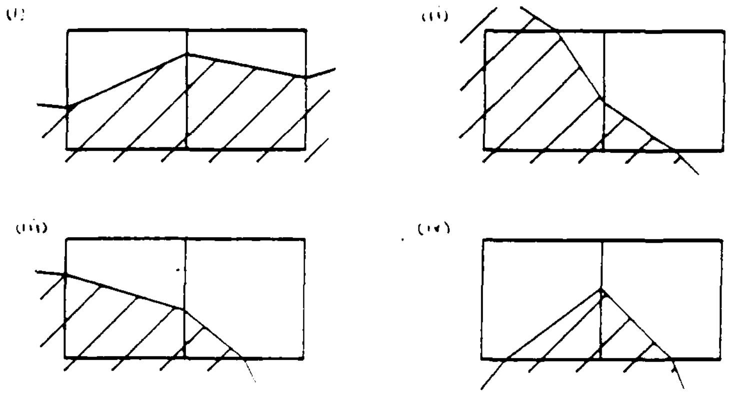


Fig. 2. Possible configurations of edges in two neighbouring pixels.

We have shown the method of solution for case (i) and cases (ii) - (iv) are solved in a similar way. All other cases can be reduced to one of the above by means of exchanging and/or inverting the pixels and their colours.

The most natural order of updating the edge pixels would seem to be to follow an edge around, updating each join in turn, completing circuits of the edge until convergence. An alternative method is to update every k^{th} join around the circuit, therefore completing k laps before each pixel has been updated once. Initial results suggest that this provides additional stability in the updating process; we have found the value $k = 3$ to give particularly good results.

AN EXAMPLE

We illustrate the methods we have described with an artificial example. Figure 3a shows a true image and the superimposed pixel grid. The record from which a restored image was constructed was obtained by generating a Gaussian random variable for each pixel with mean equal to the proportion of the pixel coloured black in the true image and variance 0.01^2. Figure 3b is the reconstruction after stage 1, in which the ICM method with $\beta=1$ has been used, treating each pixel as either completely black or completely white. Note that this is a rather poor approximation to the true image but it is the best that can be done without dividing pixels. Subdividing each pixel into four in stage 2 produces the reconstruction in Figure 3c: the amounts of black in each full pixel are now much closer to the corresponding records and the divisions of split pixels match up well with the true image. Proceeding to stage 3, we found that using $\beta'=50$ gave better results than those obtained using lower values of β'. The final reconstruction is shown in Figure 3d. Despite the coarseness of the original pixel grid and the addition of noise to the record, this reconstruction is barely distinguishable from the true image.

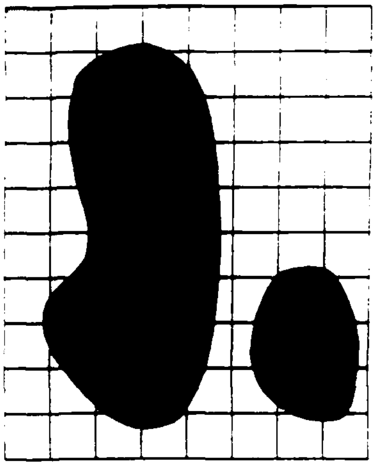


Fig 3a True image

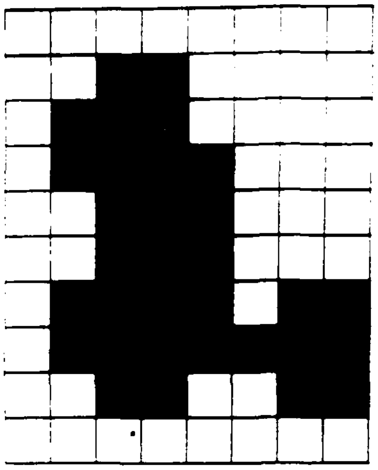


Fig 3b Reconstruction after stage 1

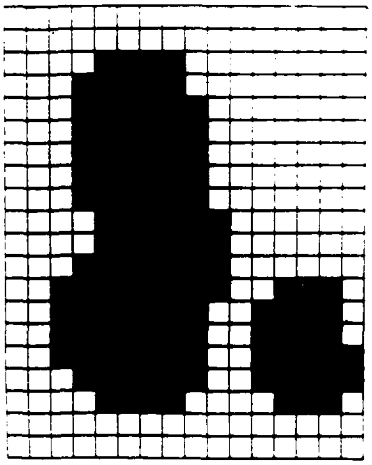


Fig 3c Reconstruction after stage 2

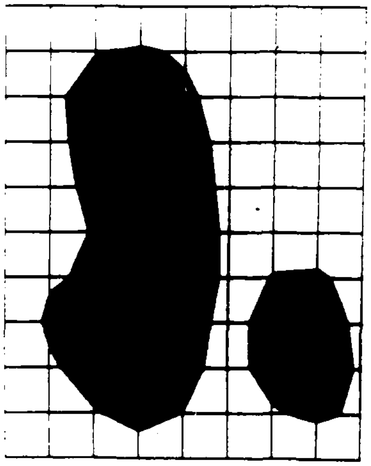


Fig 3d Final reconstruction

FURTHER EXTENSIONS

(a) Consider a pixel which has true colouring as shown in Figure 4. Clearly the straight line approximation to this edge will be poor and could have an adverse effect on the reconstruction of neighbouring pixels and pixels further along the edge. This may be overcome using a more intricate restoration method, e.g. allowing two straight lines meeting at some point within a pixel.

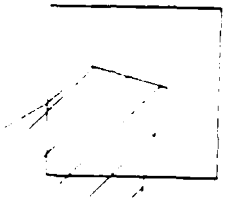


Fig. 4. A pixel containing a boundary that can not be approximated well by a single straight line.

(b) The method presented in this paper can be extended to scenes containing more than two different colours. Where any two regions meet we can adjust the algorithm to provide a continuous line join. More computation is required to find the best colouring for a pixel in which three or more regions meet.

REFERENCES

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). J. Royal Statist. Soc. B, 36, 192-236.

Besag, J. E. (1986). *On the statistical analysis of dirty pictures (with discussion).* J. Royal Statist. Soc. B, 48, 259-302.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Machine Intell., 6, 721-741.

Greig, D. M., Porteous, B.T. and Seheult, A.H. (1986). Contribution to Discussion of Besag (1986). J. Royal Statist. Soc. B, 48, 282-284.

Jennison, C. (1986). Contribution to Discussion of Besag (1986). J. Royal Statist. Soc. B, 48, 288-289.

Appendix 4

The specification of edge penalties for regular and irregular pixel images.

by

B. W. Silverman, C. Jennison and T. C. Brown

The Specification of Edge Penalties for Regular and Irregular Pixel Images

Bernard W. Silverman, Christopher Jennison and Timothy C. Brown

Abstract

One of the ingredients of recent methodology in statistical image restoration is the idea of introducing a system of "edges" between pixels in the image. If an edge is present between two contiguous pixels then they are not considered as neighbours in the restoration procedure. In penalized maximum likelihood estimation of the image, the number and configuration of the edges is controlled by a penalty term; in model-based restoration using random fields there is an analogous penalty term in the energy function of the Gibbs distribution for the edge process. In this paper we show how some geometrical insights can be used to provide penalties for the various edge configurations in a way that is roughly independent of the pixel discretisation. The penalties obtained are consistent over pixels of different sizes, shapes and orientations, even if these occur in the same pattern. The cases of square, rectangular, hexagonal and irregular pixels are considered.

B. W. Silverman and C. Jennison
School of Mathematical Sciences
University of Bath
Bath, BA2 7AY
UNITED KINGDOM

T. C. Brown
Department of Mathematics
University of Western Australia
Nedlands, WA 6009
AUSTRALIA

December 1988

Please direct all correspondence to B. W. Silverman and C. Jennison

Index Terms: discretisation, edge process, Euler-Poincaré characteristic, statistical image restoration, hexagonal pixels, irregular pixels, line length, Markov random field, penalized likelihood, stochastic geometry, tessellations.

I. INTRODUCTION

Geman and Geman [2] discussed a methodology for pixel image restoration which depended on the idea of modelling the true image by a Markov random field. A key feature of their approach was the possible placing of "edge elements" at "line sites" between pixels of the image. Although the idea of introducing an edge process was introduced in the Markov random field context, its applicability is by no means confined to model-based methods of image restoration and it is important that the construction of the process should be given careful consideration.

The edge process idea corresponds to the notion that the image is segmented into regions over each of which its behaviour is relatively homogeneous, or at least is not subject to abrupt changes; from one region to another, however, large differences in behaviour are possible. The changes in behaviour may relate either to overall grey level or colour, or to more subtle properties such as texture. Of course, the basic motivation for this kind of segmentation of the image is that the true scene is itself segmented into regions, and the edge process in the model is an attempt to approximate boundaries that are present in the true scene. For example, in the context of remote sensing of a rural area, the boundaries would correspond to topographic features like rivers and field boundaries. Our aim in this paper is to investigate the consequences of thinking of the edge process as being a discretised version of an underlying "true" pattern of boundaries. In particular we are interested in the calculation of quantitative summaries of the discretised edge process that have genuine meaning in terms of properities of the underlying boundary pattern, for example the total boundary length and the complexity of the pattern of regions defined by the boundaries.

In Geman and Geman's approach, a prior distribution for the true image is constructed by first constructing a prior Gibbs distribution for the process of edge elements and then specifying the prior for the pixels themselves conditional on the edge process. In the specification of the pixel process, contiguous pixels separated by a line site at which an edge element is actually present are not considered as neighbours, and so are allowed to have quite different grey levels without incurring any penalty in the prior likelihood.

An alternative approach in which an edge process is equally important is penalised maximum likelihood; for background reading see, for example, [4]. In the image analysis context, the image is considered as a high-dimensional unknown parameter, and a penalised log likelihood is constructed by subtracting from the log likelihood of the image given the observed data a penalty term based on the "dirtyness" of the image. The idea of penalised likelihood is that there are two conflicting aims in image restoration; one is to obtain a faithful fit to the data, as measured by the likelihood, while the other is to obtain a "clean" image, corresponding to a low value of the penalty. For reasons we shall discuss in Section 2 below, a convenient penalty to use is the energy function of the prior Gibbs distribution for the image as considered as a realisation from a random prior process. In that case the method of maximum a posteriori estimation as proposed in [2] yields exactly the same restored image as the penalised maximum likelihood approach, even though the philosophy behind the two approaches is different.

In this paper we shall focus attention on the specification of a suitable penalty for the edge process. We shall show how various geometrical insights suggest how such a penalty should be constructed. Our discussion will suggest relative costs for possible configurations somewhat different from those proposed by Geman and Geman [2]. In addition our scheme will provide methods for dealing with rectangular, hexagonal and irregular pixel patterns.

For any given penalty function the Gibbs distribution with energy equal to the penalty defines a stochastic model for the edge process. However, we stress that our interest is in developing the penalty for use in image restoration algorithms, rather than in studying the theory of stochastic models for the edge process. Apart from our intended application to image restoration, the problem of estimating the underlying edge length for a discretized image is of interest in its own right; see, for example, Dorst and Smeulders [1].

II. LOCALLY BASED PENALTY FUNCTIONS

The Gibbs distribution approach constructs a prior likelihood for the edge process by first defining a set of *cliques* of line sites. Each clique C consists of a small set of sites; in the Geman and Geman paper the cliques are the collections of four line sites with a common vertex. The Gibbs model then gives as the prior probability of any configuration ω

$$\pi(\omega) = Z^{-1}\exp\{-U(\omega)\}$$

where Z is a constant and the *energy function* U satisfies

$$U(\omega) = \sum_{\text{cliques } C} V_C(\omega).$$

Each V_C is a function which depends only on those elements of ω that correspond to sites in the clique C. Each clique consists of a set of sites all of which are "neighbours" of one another in some suitable sense, and hence the energy function $U(\omega)$ can be constructed by looking at cliques individually; looking ahead to the prospect of large scale parallel processing, this localisation property is likely to be of extreme importance in the future.

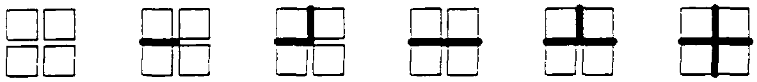
In practical applications the observed record often consists of the true image observed indirectly and subject to the addition of random noise. Maximum a posteriori likelihood estimation of the underlying true image is achieved by maximizing over possible images the likelihood of the observed record given a true image multiplied by the prior probability of that image. Equivalently, one maximizes the sum of the log prior likelihood $-U(\omega)$ and the log likelihood of the record given the true image ω.

The philosophical approach we shall follow is to consider $U(\omega)$ not directly as a prior negative log likelihood, but rather as a *penalty function* for a given configuration ω. The penalty function is subtracted from the log likelihood of the record given the true image to give a *penalised log likelihood*, maximisation of which corresponds to maximum a posteriori estimation in the Bayesian model. Compare the spline smoothing approach to nonparametric regression where the penalty term $\int g''^2$ can be considered either as a direct "roughness penalty" or as a term in a prior log likelihood; for bibliography on spline smoothing, see, for example, Silverman [5]. As already mentioned above, the use of a "locally computable" penalty like $U(\omega)$ has enormous potential advantages in an array processing computer environment, and it is on such penalties that we shall concentrate in this paper.

It is implicitly assumed in the usual restoration methods that each pixel of the true image consists of a single grey level and edges of regions or objects lie along pixel boundaries. It is more realistic to assume the existence of a real image in the plane not necessarily related to any pixel grid; the so called "true" pixel image is then a discretization of the real image. We shall assume in addition that the real image is made up of a number of regions divided by boundaries and that the edge process in the "true" pixel image is constructed to approximate the real boundaries as closely as possible. Our approach is to attempt to specify the form of the function V_C in such a way that the penalty associated with a pixel edge process in the "true" pixel image will, as far as possible, not depend on the way the pixels are constructed or placed on the real image; instead the value of the penalty $U(\omega)$ will give a cost based at least approximately on the real underlying boundary pattern. Particular concerns will be to eliminate, as far as possible, the effect of the position and orientation of a square lattice; to discuss how to modify the penalty if the lattice is refined; and to devise appropriate penalties for irregular pixel patterns. The penalty we shall use will have as one ingredient an estimate of the total boundary length in the underlying "real" image, and so has relevance to the problem discussed in [1].

III. Square Lattices

Let us turn first to the case of the square lattice, considered by Geman and Geman [2]. Suppose that the gauge of the lattice is h, and that each clique consists of the four line-sites meeting at a particular vertex. The possible configurations and the costs ascribed to these configurations by Geman and Geman [2] are shown in Figure 3.1.



	(no lines)	(ending)	(turn)	(continuation)	(branch)	(crossing)
Type	0	1	2	3	4	5
Cost	0	2.7	1.8	0.9	1.8	2.7

Figure 3.1: Possible types of configuration for regular edge process, and the costs ascribed to them by Geman and Geman [2].

Note that the low cost of a continuation relative to the cost of an ending, branch or crossing is intended to favour a small number of long straight edges over complex, meandering edge systems. However, we shall see that this fails to provide an adequate treatment of long straight edges at orientations away from the horizontal and vertical. We shall write v_i for the cost of a configuration of type i, and explore the consequences of various choices of v_i.

A. Boundary Length Considerations

Consider, now, the cost of a very simple pattern, consisting of an infinitely long straight line placed at angle θ to one of the edge directions of the lattice; without loss of generality $0 \leq \theta \leq \frac{\pi}{4}$. The discretization will replace the line by a stepped pattern of the form shown in Figure 3.2.

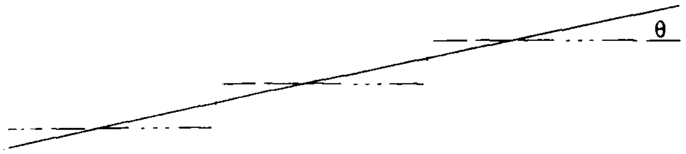


Figure 3.2: A line boundary and its discretization.

Because $\theta \le \frac{\pi}{4}$, vertical segments will always be separated by one or more horizontal segments. Over a long distance L in the x-direction, the number n_x of horizontal segments will be asymptoticaly Lh^{-1}, and the number n_y of vertical segments asymptotically $Lh^{-1}\tan\theta$. The number of configurations of type 2 will be $2n_y$ and the number of type 3 will be $n_x - n_y$. Thus the total cost will be

$$2n_y v_2 + (n_x - n_y)v_3 = Lh^{-1}\{v_3 + (2v_2 - v_3)\tan\theta\}.$$

The total length of underlying boundary is $L\sec\theta$, and so the cost $c(\theta)$ per unit length of underlying boundary is, for large L,

$$c(\theta) = h^{-1}\{v_3\cos\theta + (2v_2 - v_3)\sin\theta\}. \tag{3.1}$$

The ideal situation would be for $c(\theta)$ to be independent of θ, but (3.1) makes it clear that this is impossible. Define $\alpha = v_2/v_3$. A natural index of how far $c(\theta)$ falls short of ideal is given by the ratio

$$I(\alpha) = \max_{0\le\theta\le\pi/4} c(\theta) \Big/ \min_{0\le\theta\le\pi/4} c(\theta).$$

This ratio depends only on α. If $\alpha \ge 1$, $c(\theta)$ is monotonically increasing in $[0, \frac{\pi}{4}]$, and so $I(\alpha) = c(\frac{\pi}{4})/c(0) = \alpha\sqrt{2}$. If $\alpha \le \frac{1}{2}$, $c(\theta)$ is monotonically decreasing in $[0, \frac{\pi}{4}]$, and $I(\alpha) = c(0)/c(\frac{\pi}{4}) = 1/(\alpha\sqrt{2})$. To deal with $\frac{1}{2} < \alpha < 1$, define $\theta_0 = \tan^{-1}(2\alpha - 1)$ and rewrite

$$\begin{aligned} c(\theta) &= h^{-1}v_3\sec\theta_0\{\cos\theta_0\cos\theta + \sin\theta_0\sin\theta\} \\ &= h^{-1}v_3\sec\theta_0\cos(\theta - \theta_0). \end{aligned} \tag{3.2}$$

Since for $\frac{1}{2} < \alpha < 1$ we have $0 < \theta_0 < \frac{\pi}{4}$, it follows that, for α in this range, $c(\theta)$ has a maximum at θ_0 and that $I(\alpha) = \max\{\sec\theta_0, \sec(\frac{\pi}{4} - \theta_0)\}$. Hence $I(\alpha)$ is minimized by setting $\theta_0 = \frac{\pi}{8}$.

The minimum value $\sec\frac{\pi}{8} = (4 - 2\sqrt{2})^{1/2} = 1.082$. Thus it follows that the minimax score $I(\alpha)$ is optimized by setting $2\alpha - 1 = \tan\frac{\pi}{8}$, which implies that $\alpha = \frac{1}{2}(1 + \tan\frac{\pi}{8}) = 1/\sqrt{2}$. If this value of α is used, then lines parallel to the lattice directions or those at 45° to these directions will cost the same amount per unit length, while the most expensive lines will be those at $22\frac{1}{2}°$ to the axis directions, which will cost about 8% more. It is interesting to note that the Geman and Geman value $\alpha = 2$ yields $I(\alpha) = 2\sqrt{2} \approx 2.83$, a much larger value.

It can also be shown, by somewhat tedious algebra, that $\alpha = 1/\sqrt{2}$ also minimizes other criteria of variability of $c(\theta)$, for example the coefficient of variation of $c(\theta)$ with θ uniformly distributed over $[0, \frac{\pi}{4}]$.

The arguments of this section make it possible to settle on a charge for configurations of types 2 and 3. Suppose it is intended to penalize boundaries in the underlying picture by an amount β per unit length. In an ideal world we would like to choose v_2 and v_3 in (3.1) to ensure that $c(\theta) = \beta$ for all θ. As we have seen, this cannot be attained exactly for all θ, but setting $v_2/v_3 = 2^{-1/2}$ will minimize the variability of $c(\theta)$ as θ varies. Having settled the ratio v_2/v_3, it is natural to choose v_3 to ensure that $(2\pi)^{-1}\int_0^{2\pi} c(\theta)d\theta = \beta$. By simple algebra, from (3.2),

$$(2\pi)^{-1}\int_0^{2\pi} c(\theta)d\theta = 4\pi^{-1}\int_0^{\pi/4} h^{-1}v_3 \sec(\frac{\pi}{8}) \cos(\theta - \frac{\pi}{8})d\theta$$

$$= 8\pi^{-1}h^{-1}v_3 \tan(\frac{\pi}{8}) = v_3 h^{-1}k^{-1}$$

where the constant $k = \frac{\pi}{8}/\tan(\frac{\pi}{8}) \approx 0.948$.

It follows that setting $v_3 = k\beta h$ and $v_2 = 2^{-1/2}k\beta h$ will ensure that, while $c(\theta)/\beta$ is only exactly 1 for certain values of θ, it will be the case that $c(\theta)/\beta$ lies between 0.948 and 1.027 for all θ and furthermore that the average value of $c(\theta)$ over (uniformly distributed) θ is precisely β.

The above results are also relevant to the problem of estimating the underlying edged length from a discrtized image as posed in [1]. Suppose a line of fixed length placed at orientation θ uniformly distributed over $[0, 2\pi]$ has N_2 turns and N_3 continuations in its discretized form. Then $2^{-\frac{1}{2}}khN_2 + khN_3$ is an unbiased estimate of the line's original length; it is the minimum variance unbiased estimator among estimators of the form $aN_2 + bN_3$, as a consequence of the fact that $\alpha = 1/\sqrt{2}$ minimizes the coefficient of variation of $c(\theta)$.

Turning now to the question of how much to charge for branches and crossings, we shall explain in the next section how a simple argument concerned with counting the number of regions in the pattern leads to a paradigm for dealing with these configurations.

B. Counting Regions

Suppose that, in the original pattern, the plane is divided up into a number of simply connected regions, and that the edge process is an approximation to the boundaries between regions in this configuration. Assume that the pixel size is sufficiently small relative to the scale of regions in the pattern that each region is represented by a single connected set of pixels in the discretized image.

Apart from the total edge length, a natural measure of the complexity of the pattern of regions is given by the number of regions. If the region boundaries form a connected set, or equivalently if the regions are simply connected, the number of regions can be counted simply by counting the number of "branches" and "crossings" in the edge pattern. To do this, the Euler-Poincaré formula [3, p.241] is used.

Suppose the original process is observed on a window W in the plane and at least one boundary intersects the window edge. Define a *vertex* to be a point where three or more regions meet, or where the boundary between two regions meets the edge of the window. Define a ***boundary section*** to be the piece of boundary or of window edge between two vertices. Let n_v be the number of vertices in the pattern, n_e the number of boundary sections and n_f the number of regions. The Euler-Poincaré formula gives the equation

$$n_v - n_e + n_f = 1$$

and hence

$$n_f = 1 + n_e - n_v .$$

Now both n_e and n_v can be found by counting the number of branches and crossings in the pattern, provided that points where an edge meets the edge of the window count as branches. Let n_b be the number of branches and n_c the number of crossings. It is immediate that

$$n_v = n_b + n_c . \tag{3.3}$$

In order to count the number of boundary sections, notice that three boundary sections meet at each branch and four at each crossing. Thus the number of ends of boundary sections in $3n_b + 4n_c$, and since each boundary section has two ends, we have

$$n_e = \tfrac{3}{2} n_b + 2n_c . \tag{3.4}$$

Substituting (3.3) and (3.4) into (3.5) yields

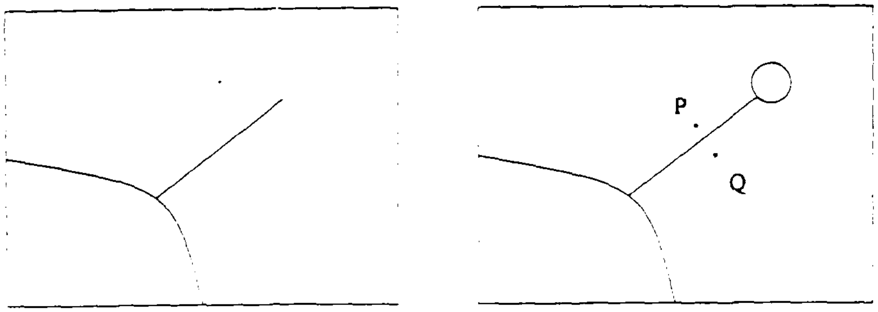
$$n_f = 1 + \tfrac{1}{2} n_b + n_c . \tag{3.5}$$

Formula (3.5) gives a natural price to be charged for branches and crossings. If it is desired to penalize an amount ρ for each region in the pattern, then one should charge $\frac{1}{2}\rho$ for each branch point and ρ for each crossing. If the edge configuration gives rise to regions that are not simply connected the right hand side of (3.5) must be increased by 1 for each connected set of edges which does not intersect the window edge. The charge $(\frac{1}{2}n_b + n_c)\rho$ can be considered in its own right as a penalty for the complexity of the edge pattern which is calculable from local properties. The extra cost of ρ for each isolated connected set of edges cannot be calculated from local properties and, thus, cannot be included in a restoration algorithm that operates entirely by local updating; such an algorithm might, however, be extended to investigate the complete removal of a small connected set of edges in the later stages of reconstruction.

This scheme of charging for branches and crossings does not include a cost for the boundary length involved. We shall return to this point in Section 5 after the necessary tools have been developed.

C. Endings

A pattern made up of disjoint regions cannot, of course, have a configuration of edges containing any endings at all. Therefore the philosophy that we are adopting would naturally lead to an infinite charge for configurations of type 1 in Figure 3.1. This might still not be completely acceptable: although the configuration in Figure 3.3a is prohibited, that in Figure 3.3b is still allowed but note that points P and Q, which lie close together in the same region, are separated by an edge. It must also be remembered that to set any penalty value to infinity may lead to algorithmic difficulties in using the model in practice. Also, a prior model for the edge process under which some configurations have probability zero violates the condition of positive probability for all configurations under which the theory and practice of Markov random fields are developed; see, for example, Section 4 of [2]. In any case, it seems excessively dogmatic to exclude certain configurations completely, since there may be good physical reasons for a boundary to peter out in the middle of a region. Therefore an approach that is likely to be more satisfactory is to ascribe a cost λ to each "loose end" in the boundary pattern, where λ is set to a relatively large value. In fact, there is no advantage in setting λ much greater than $\frac{1}{2}\rho$ since a clever reconstruction algorithm can simply build a small loop of edges onto a loose end at a cost of $\frac{1}{2}\rho$ for the branch plus the cost of the edge length involved.



(a) (b)

Figure 3.3: Possible edge configurations.

D. Summary and Example

We now summarize the appropriate relative costs of different configurations. Let β be the desired cost per unit length of edge, ρ the cost per region of the pattern and λ the cost per loose end. Then the programme we have set out gives as the costs ascribed to possible configurations the costs set out in Table 3.1. Costs for edge length and region counting appear separately; the costs for edge length associated with configurations 3, 4 and 5 will be derived in Section 5. As explained in Section 3*C* the value $\lambda = \frac{1}{2}\rho$ is a reasonable choice but there is no obvious relation between ρ and β. The interpretation of β as a cost per unit length of boundary makes it possible to adjust the scores in a reasonable way if the pixel grid is refined, since the cost of edge length in configurations 1 to 5 is adjusted automatically.

Type of configuration	Cost
0 (none)	0
1 (ending)	$0.412\,h\beta + \frac{1}{2}\rho$
2 (turn)	$0.670\,h\beta$
3 (continuation)	$0.948\,h\beta$
4 (branch)	$1.4\,h\beta + \frac{1}{2}\rho$
5 (crossing)	$1.94\,h\beta + \rho$

Table 3.1. Proposed costs for the configurations of Fig. 3.1.

In order to provide an illustration of the results derived in this section, some costs for the pixellated edge patterns shown in Figure 3.4 were calculated. The pixel size for Figures 3.4c and 3.4d is half that used in Figures 3.4a and 3.4b, and the unit of length is taken such that $h = 1$ in Figures 3.4a and 3.4b. The costs are presented in Table 3.2. Our costs are given in terms of β and ρ, and are also evaluated for the case $\beta = 1$, $\rho = 50$. It can be seen that rotating the pattern affects the Geman and Geman costs quite substantially but has very little effect on the costs calculated using our methods. It can also be seen that our costs maintain consistency across the different pixel sizes. The slightly larger costs obtained for the smaller pixels is presumably due to a "fractal" effect in the discretisation of the coastline.

Figure	Pixel size	Number of cliques			Geman cost	Proposed cost	Proposed cost with $\beta=1$, $\rho=50$
		type 2	type 3	type 4			
3.4a	$h=1$	142	153	8	453	$251.4\beta+4\rho$	451.4
3.4b	$h=1$	237	95	8	585	$260.0\beta+4\rho$	460.0
3.4c	$h=0.5$	333	329	8	1011	$273.1\beta+4\rho$	473.1
3.4d	$h=0.5$	493	221	8	1223	$275.5\beta+4\rho$	475.5

Table 3.2. Costs of edge patterns shown in Fig. 3.4

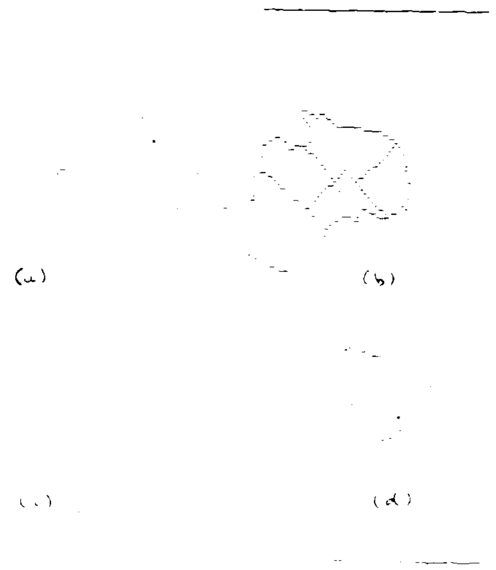


Fig. 3.4. Four discretisations of the same edge pattern.

IV. IRREGULAR AND UNEVEN PIXEL ARRAYS

In this section we turn to the case of arrays of pixels that are no longer based on a regular square lattice. One such case arises if a pixel pattern based on polar coordinates is used, as shown in Figure 4.1. Such circular pixel patterns arise very naturally in the restoration of images observed by positron emission tomography; see Silverman et al. [6] for an application of the circular pixel patterns and Vardi, Shepp and Kaufman [7] for a general discussion of the positron emission tomography problem.

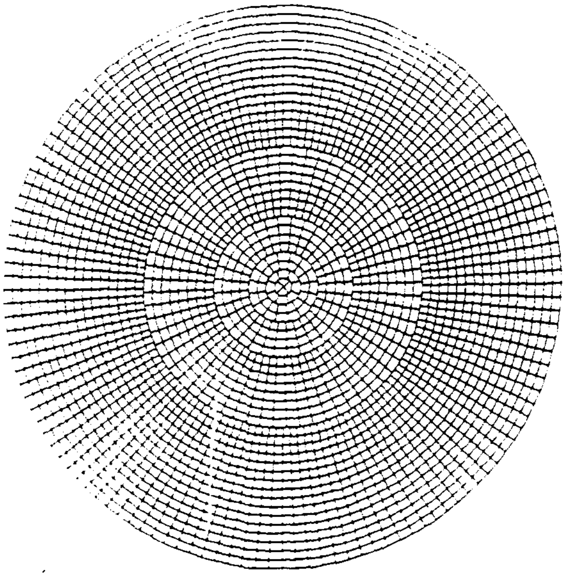


Fig. 4.1. A circular pixel array useful for positron emission tomography images.

In general, the pixels might be more irregular still, and might even themselves be generated by a random process. This is unlikely to be the case where the experimenter has control over the pixel pattern. However, irregular pixels may well occur, for example, in geographical applications, where the observed "image" is made up of measurements averaged or cumulated over small irregularly-shaped regions, and it is not felt desirable to superimpose a regular grid on the existing irregular pixel pattern.

A. *Cliques for Irregular Edge Processes*

We shall assume, for the moment, that the pixel pattern forms a tessellation of the plane or a portion of the plane, and that except at the boundary of the pattern, exactly three pixels meet at each vertex of the tessellation. This assumption is, of course, violated for the circular lattice of Figure 4.1, since some of its vertices are of degree three and some of degree four. It does, however, hold (with probability 1) for many randomly generated pixel models, for example if the pixels are the Voronoi polygons of a homogeneous planar Poisson process.

As in the case of the square lattice, the line-sites in the edge process will be the boundary sections of the pixel array, and we shall suppose that each clique of the edge process consists of the three line-sites meeting at a particular vertex. There are now four possible types of configurations for a particular clique in the edge process, depending on how many edges are present in the clique. We shall say that the configuration is of type k for $k = 0,1,2,3$ if k of the three line-sites in the clique are actually occupied by edges. These configurations are illustrated in Figure 4.2.

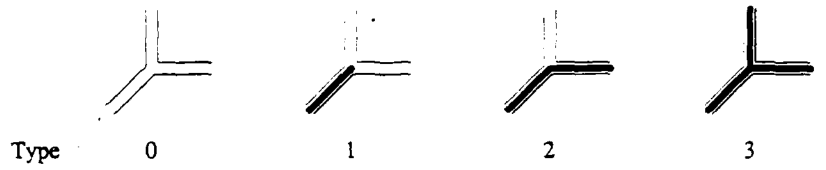


Figure 4.2: Configurations for a vertex of degree 3.

Although, in contrast with the case of square pixels, there are fewer types of configuration to consider, the irregularity of the pixels means that it is no longer necessarily the case that all configurations of a particular type should attract the same penalty.

The first stage in the assignment of costs to various configurations is to use the same region counting arguments as in the square lattice case to assign charges 0, $\frac{1}{2}\rho$ and $\frac{1}{2}\rho$ to configurations of types 0, 1 and 3 respectively. It remains to ascribe costs for the edge length associated with each configuration. In order to do this, construct a dual edge pattern by placing a point in each cell of the original pixel array, and joining points if their corresponding pixels have some boundary in common. The vertices of the dual array can, in principle, be placed anywhere in their corresponding pixels, but in practice they will have a natural position. For example if the pixels are constructed as the Voronoi polygons of a point process then the points of the process will themselves be the vertices of the dual array.

Our assumption that exactly three pixels meet at each vertex of the original tessellation implies that the dual edge pattern will be a triangulation of the plane. In the case of the square pixel array the cost of "continuation" configurations was determined by considering a pattern with a single long straight edge, suitably discretized to fit the pixel pattern. In the more general case, it is no longer quite so clear how this discretization should be performed. One natural way to proceed is to prescribe that an edge segment will be present in the edge process if and only if the corresponding dual edge is intersected by the straight line boundary. We assume, if necessary giving the line an infinitesimal displacement perpendicular to its direction, that no vertices of the dual triangulation lie exactly on the line.

Any edge process in the original tessellation corresponds to an edge process in the dual triangulation in the natural way, a dual edge being present in the process whenever the corresponding original line site is occupied. Each clique of the original edge process will correspond to a triangle in the dual triangulation; the original edge configuration will be a "continuation" if and only if exactly two of the edges are present in the corresponding dual clique. Every triangle intersected by the straight line boundary will give rise to a continuation clique, since exactly two of its edges are necessarily intersected by the line. We shall now describe two possible approaches to charging for continuation configurations. The first of these appears more natural at first sight, but it leads to much more complicated formulas; we shall also show that the second has the additional advantage of being a genuine generalization of our square pixel formulas.

B. Two Possible Length Penalties

Let us concentrate on a single triangle i in the dual triangulation. The notation for this triangle will be as in Figure 4.3. The capital letters A, B, C will be used for both the vertices themselves and for the angles at these vertices.

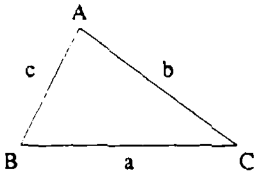


Figure 4.3: A triangle in the dual triangulation.

The lower case letters refer to the sides themselves and to the lengths of these sides. We shall derive possible ways of charging for the "continuation" configuration bc given by the presence of the edges dual to b and c and the absence of the edge dual to a. These costs will be based on the general idea that boundaries should cost an amount β per unit length; for notational simplicity we shall assume henceforth that

$\beta = 1$, and note that the costs obtained should be multiplied by β in the general case.

Let l be a random line of fixed length L randomly situated in the plane, in a sense that will be made precise below. Let L_i be the length of the intersection of l with the triangle i. Then our first possible cost for the continuation configuration bc is

$$V_1(i,b,c) = E(L_i \mid l \text{ intersects triangle } i \text{ through } b \text{ and } c).$$

The motivation for this definition is straightforward. Let $I(i,e_1,e_2)$ be the indicator variable taking the value 1 if l intersects sides e_1 and e_2 of triangle i and 0 otherwise. Let $L(i,e_1,e_2)$ be the length of the intersection of l with triangle i if $I(i,e_1,e_2) = 1$ and 0 otherwise. Now, ignoring end effects, the total line length $L = \sum L(i,e_1,e_2)$ where the summation is taking over all triangles i and pairs of edges (e_1,e_2). The total charge for the line l is then

$$S_1 = \sum_{(i,e_1,e_2)} I(i,e_1,e_2)\, V_1(i,e_1,e_2).$$

Thus, up to the approximation involved in ignoring end effects,

$$\begin{aligned} E(S_1) &= E[\sum I(i,e_1,e_2)\, E\{L(i,e_1,e_2) \mid I(i,e_1,e_2) = 1\}] \\ &= E[\sum E\{L(i,e_1,e_2)\}] \\ &= E\{\sum L(i,e_1,e_2)\} = E(L) = L \qquad (4.1) \end{aligned}$$

and S_1 is an unbiased estimate of the line length, L. It is clear that the above argument will also hold if $V_1(i,e_1,e_2)$ is replaced by $E\{g(i,e_1,e_2) \mid I(i,e_1,e_2) = 1\}$ for any function $g(i,e_1,e_2)$ for which, apart from end effects, $\sum g(i,e_1,e_2) = L$. Our second proposed cost is also of this general form.

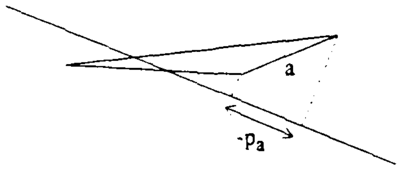


Figure 4.4: A side with negative projected length.

Let p_a be the projected length of the side a on the line l; this length is to be counted as negative if l makes an angle of more than $\frac{\pi}{2}$ with a, in the sense shown in Figure 4.4. The second proposed cost is

$$V_2(i,b,c) = E(\tfrac{1}{2}p_a \mid l \text{ intersects triangle } i \text{ through } b \text{ and } c).$$

To justify this definition, consider the union of all the triangles intersected by the line l. This forms an irregular strip in the plane. The two edges of this strip, one on each side of l, are made up by those edges not intersected by l. The total projection length of these edges on l, neglecting end effects, is equal to twice the length of l, and the sum of all the quantities like $\frac{1}{2}p_a$ is equal to L. Hence $S_2 = \sum I(i,e_1,e_2)\, V_2(i,e_1,e_2)$ is an unbiased estimate of L.

The sense in which l is a random line is as follows. Choose an origin O in the plane and let R be large enough to ensure that the triangle ABC is entirely enclosed within the circle centre O and radius R. Now construct l such that the perpendicular from O to l has orientation uniformly distributed on $(0,2\pi)$ and length uniformly distributed on $(0,R)$. This is the distribution of l conditional on l intersecting triangle i through sides b and c if the pixel grid and associated triangulation is placed down in a random position and at a random orientation relative to the line l. By standard stochastic geometry, the quantities V_1 and V_2 will be independent of the choice of R.

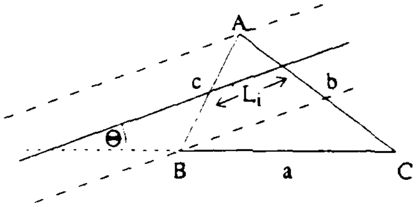


Figure 4.5: A random line intersecting the dual triangle.

We now calculate V_1 and V_2. Let Θ be the angle between l and BC, measured as shown in Figure 4.5. The first step is to find the density $f(\theta)$ of Θ conditional on l intersecting b and c. Note first that, of necessity, $-C < \Theta < B$; consider first the range $0 < \theta < B$.

For such θ, l will intersect c and b if and only if it intersects c. The set of lines at orientation θ that intersect c make up a strip of width $c \sin(B-\theta)$ and so we have

$$f(\theta) \propto c \sin(B-\theta) \quad \text{for } 0 < \theta < B.$$

For $-C < \theta < 0$, a similar argument yields

$$f(\theta) \propto b \sin(C+\theta) \quad \text{for } 0 < -\theta < C.$$

To calculate the constant of proportionality, we note that

$$\int_0^B c \sin(B-\theta)d\theta + \int_{-C}^0 b \sin(C+\theta)d\theta = c\int_0^B \sin\varphi\, d\varphi + b\int_0^C \sin\varphi\, d\varphi$$

$$= b + c - c\cos B - b\cos C = b + c - a,$$

and hence we have

$$f(\theta) = \begin{cases} c \sin(B-\theta)/(b+c-a) & 0 < \theta < B \\ b \sin(C+\theta)/(b+c-a) & -C < \theta < 0 \\ 0 & \text{otherwise} \end{cases}$$

To calculate V_1, consider first $\theta > 0$. Given that $\Theta = \theta$ and that l intersects c and b, the expected value of L_i is half its value when $\Theta = \theta$ and l passes through B. This length is, by the sine formula, equal to $c \sin A/\sin(C+\theta)$. Hence we have

$$\begin{aligned} V_1 &= \int_0^B \{\tfrac{1}{2}c \sin A/\sin(C+\theta)\} f(\theta)d\theta \\ &+ \int_{-C}^0 \{\tfrac{1}{2}b \sin A/\sin(B-\theta)\} f(\theta)d\theta \qquad (4.2) \end{aligned}$$

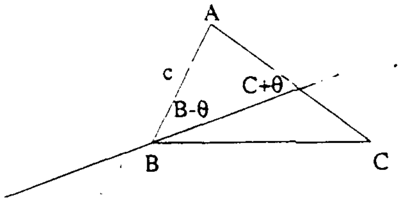


Figure 4.6: The random line passes through B.

To calculate these integrals, first substitute $\varphi = C + \theta$ and use the fact that $A + B + C = \pi$ to give

$$\begin{aligned} \int_0^B \sin(B-\theta)/\sin(C+\theta)d\theta &= \int_C^{B+C} \sin(B+C-\varphi)/\sin\varphi\, d\varphi \\ &= \sin A \int_C^{B+C} \cot\varphi\, d\varphi + B\cos A = \sin A(\log\sin A - \log\sin C) + B\cos A. \end{aligned}$$

Substituting this and the corresponding formula for the second integral into (4.2) gives, after some trigonometry,

$$\begin{aligned} (b+c-a)V_1 &= \tfrac{1}{4}(c^2B + b^2C)\sin 2A \\ &+ \tfrac{1}{2}\sin^2 A\{(b^2 + c^2)\log\sin A - c^2\log\sin C - b^2\log\sin B\}. \qquad (4.3) \end{aligned}$$

This formula is complicated and inelegant, and it turns out that V_2 is much more simply expressed. Given that $\Theta = \varphi$, we have $p_a = a \cos\varphi$, and hence

$$\begin{aligned} V_2 &= \tfrac{1}{2}a \int_{-C}^B \cos\theta\, f(\theta)d\theta \\ &= (b+c-a)^{-1}\left\{\int_0^B \tfrac{1}{2}ac\cos\theta\sin(B-\theta)d\theta + \int_0^C \tfrac{1}{2}ab\cos\theta'\sin(C-\theta')d\theta'\right\}. \end{aligned}$$

(4.4)

The first integral in (4.4) is equal to

$$\int_0^B \tfrac{1}{4}ac\{\sin B + \sin(B-2\theta)\}\,d\theta = \tfrac{1}{4}ac\,B\sin B = \tfrac{1}{2}\Delta B$$

where Δ is the area of the triangle i, and hence

$$V_2 = \tfrac{1}{2}(B+C)\Delta/(b+c-a) \tag{4.5}$$

Thus it is clear that the formula for V_2 is very simple and more appealing than that for V_1.

A second reason for preferring V_2 to V_1 will be elaborated in Section 5 below. It is shown there that, for square pixel arrays, the projection approach produces the minimum variance unbiased estimate of line length.

C. Line Length Associated with Endings and Branches

We now turn to the problem of ascribing a cost for the edge length associated with configurations 1 and 3 of Figure 4.2. From now on we shall restrict our attention to the 'projection' cost V_2. As previously explained, the union of all triangles intersected by the line l forms an irregular strip in the plane and the sum of the projection lengths of the edges of this strip in the direction of l is approximately twice the length of l. This approximation becomes exact if the strip is terminated with edges AF and BF, as shown in Figure 4.7, and the corresponding edges at its other end.

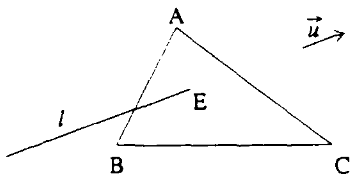


Figure 4.7: An end of line l in the dual triangle ABC.
The vector $\vec{u}$ is a unit vector in the direction of l.

Let p_F be the sum of the projections of AF and BF in the direction of l or, more formally,

$$p_F = (\vec{AF} + \vec{BF}).\vec{u},$$

where $\vec{u}$ is a unit vector in the direction of l (see Figure 4.7) and $\vec{AF}$ and $\vec{BF}$ are vectors. This definition automatically provides a correct treatment of any negative projections. We define

$$V_2(i,c) = E(\tfrac{1}{2}p_F \mid l \text{ intersects } c \text{ and terminates in triangle } i),$$

where the distribution of l and its end point are as described shortly. Repeating the argument leading up to (4.1) we see that the sum of costs V_2 is now an exactly unbiased estimate of the length of line l when the line is placed at random on an infinite pixel grid. (An infinite grid is needed to avoid problems at the window edge.)

In calculating V_2 for this case the distribution of line l is as described previously but with an extra multiplicative factor proportional to the length of the intersection of l and the triangle ABC, only lines intersecting AB are considered and the right hand end of the line is distributed uniformly along the length of the line l inside the triangle. Again, this corresponds to the conditional distribution of the line and its end, given that the line enters triangle ABC through edge c and terminates inside the triangle, when the pixel grid is placed in a random position and orientation. A long and tedious calculation gives the value of $E(\frac{1}{2}p_F)$ for this case

$$V_2 = \tfrac{1}{6}K\{2a^3+a^2b+ab^2+2b^3-3(a^2+b^2)c-(a-b)(b\cos A-a\cos B)c$$
$$-3a^2c\cos A\log\tan\tfrac{1}{2}A-3b^2c\cos B\log\tan\tfrac{1}{2}B-3c(a^2\cos A+b^2\cos B)\log\tan\tfrac{1}{2}C\}$$

where

$$K=\{a^2(B\cot A+\log\frac{a}{c})+b^2(A\cot B+\log\frac{b}{c})+abC\operatorname{cosec}C\}^{-1}.$$

This formula simplifies in special cases, for example, on a regular hexagonal grid in which all dual triangles are equilateral of side a, $V_2=(3\sqrt{3}/8\pi)a\log 3$.

A similar calculation could be performed for the branch in configuration 3 of Figure 4.2. We shall not complete such a calculation but we shall describe the general approach. A typical configuration in the dual space is depicted in Figure 4.8 and the appropriate definition of p_F is

$$p_F=(\vec{AF}+\vec{BF}).\vec{u}_1+(\vec{CF}+\vec{AF}).\vec{u}_2+(\vec{BF}+\vec{CF}).\vec{u}_3.$$

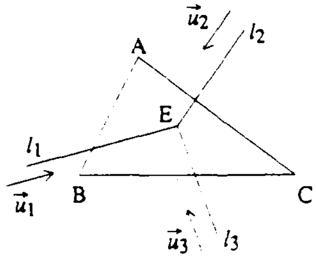


Figure 4.8: Three lines meeting in the dual triangle ABC and their associated unit vectors.

When calculating $V_2 = E(\frac{1}{2}p_F)$ it should be noted that configurations such as that in Figure 4.9 also give rise to the same configuration of pixel edges; this does not cause a serious problem and the total edge length will be estimated correctly as long as these cases are treated as meeting in ABC. Note that for $E(\frac{1}{2}p_F)$ to be properly defined it is necessary to introduce a joint distribution for the angles between the three lines, preferably by appealing to specialised knowledge of the image in question.

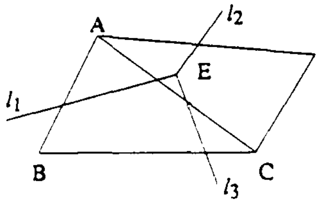


Figure 4.9: Three lines meeting outside the dual triangle ABC but still producing three edges meeting at the vertex associated with ABC.

V. REGULAR ARRAYS REVISITED

In the last section we defined two different ways of obtaining penalties for continuation configurations. One of these was based on the length of the *intersection* of a region in the dual triangulation with a random line, and the other on the length of *projection* of such a region on a random line. It turned out that the projection penalty gave a much more elegant result. In this section, we shall apply the intersection and projection ideas to the regular square lattice considered earlier, and to rectangular and hexagonal lattices.

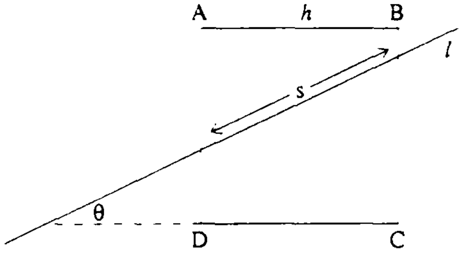


Figure 5.1: A random line intersecting a square in the dual lattice.

A. *Square Lattices*

Our aim is to obtain costs for the "turn" and "continuation" configurations as illustrated in Figure 3.1. The dual of the square lattice is itself a square lattice, and the part of the dual corresponding to a clique is a single square of side h as in Figure 5.1. The configuration of edges in the original clique will be a straight continuation if l crosses AD and BC. We find the distribution of θ conditional on l being a random line under this additional condition.

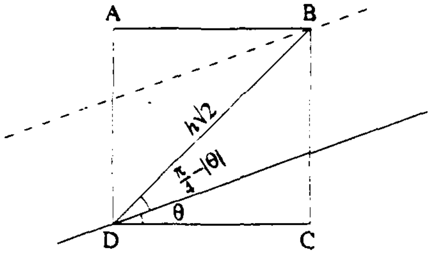


Figure 5.2: Lines of inclination θ crossing AD and BC form a strip of width $\sqrt{2}h\sin(\frac{\pi}{4}-|\theta|)$.

For $-\frac{\pi}{4} < \theta < \frac{\pi}{4}$, the set of lines crossing AD and BC will be a strip of width $h\sqrt{2}\sin(\frac{\pi}{4} - |\theta|)$ by some easy trigonometry. Hence the density $f_1(\theta)$ of θ conditional on l crossing AD and BC will satisfy

$$f_1(\theta) = (2-\sqrt{2})^{-1} \sin(\tfrac{\pi}{4}-|\theta|), \qquad -\tfrac{\pi}{4} < \theta < \tfrac{\pi}{4}$$

using simple calculus to find the constant of proportionality. The intersection length s is equal to $h \sec\theta$, and hence the expected intersection length is

$$\begin{aligned}\int_{-\pi/4}^{\pi/4} h\sec\theta f_1(\theta)d\theta &= \int_0^{\pi/4}(1-\tfrac{1}{2}\sqrt{2})^{-1} \sec\theta \sin(\tfrac{\pi}{4}-\vartheta)d\theta \\ &= (\sqrt{2}-1)^{-1} \int_0^{\pi/4}(1-\tan\theta)d\theta = (\sqrt{2}-1)^{-1}\left[\theta-\log\sec\theta\right]_0^{\pi/4} \\ &= (\sqrt{2}-1)^{-1}(\tfrac{\pi}{4}-\tfrac{1}{2}\log 2).\end{aligned}$$

Thus the "intersection" penalty for a configuration of type 3 in Figure 3.1 would be $(\frac{\pi}{4}-\frac{1}{2}\log 2)\, h/(\sqrt{2}-1) = 1.06\,h$.

To find the "projection" penalty for such a configuration, note that the appropriate generalization of the projection argument given in Section 4 is to take as penalty $\frac{1}{2}$(projection of AB and DC) because both AB and DC will be edges of the irregular strip formed by the union of those dual squares intersected by l. Both AB and DC have projection length $h\cos\theta$ on l, and so the "projection" penalty for a configuration of type 3 will be

$$\int_{-\pi/4}^{\pi/4} h \cos\theta f_1(\theta) d\theta = (1-\tfrac{1}{2}\sqrt{2})^{-1}\ h \int_0^{\pi/4} \cos\theta \sin(\tfrac{\pi}{4}-\theta) d\theta$$
$$= (2-\sqrt{2})^{-1} h \int_0^{\pi/4} \{\sin\tfrac{\pi}{4} - \sin(2\theta-\tfrac{\pi}{4})\}\ d\theta$$
$$= \frac{1}{8}\pi h/(\sqrt{2}-1) = kh = 0.95\ h$$

where $k = \frac{\pi}{8} \tan\frac{\pi}{8}$ as defined in Section 3.

To find the penalties for "turn" configurations, the work of Section 4 can be used almost directly, by noticing that both the "intersection" and "projection" penalties will be the same as those obtained there, for the case of a line crossing the two short sides of an isosceles right-angled triangle. Thus we set $a = h\sqrt{2}$, $b = c = h$, $B = C = \frac{\pi}{4}$ and $A = \frac{\pi}{2}$ in the formulas (4.3) and (4.5).

We obtain as the intersection penalty for the turn configuration $V_1 = \frac{1}{2}h \log 2/(2-\sqrt{2}) = 0.59\ h$ and for the projection penalty $V_2 = 2^{-1/2}kh = 0.67\ h$. It is noteworthy that the projection approach yields penalties for the two configurations that are identical to those obtained in Section 3. Thus, by the argument of Section 4, up to the approximation of ignoring end effects, the projection penalty is the minimum variance unbiased estimate of line length calculated from cliques of four line sites only. The intersection approach yields a higher cost for straight continuation and a lighter cost for turns and, thus, has greater than necessary variability with orientation in the cost of a long straight line boundary.

Following the development of Section 4, we now use the projection cost to assign costs for edge length in configurations 1, 4, and 5 of Figure 3.1. For an ending (configuration 1), as shown in Figure 5.3, the required cost is

$$V_2 = E(\tfrac{1}{2}(\vec{AF} + \vec{DF}).\vec{u} \mid l \text{ crosses } AD \text{ and terminates in } ABCD).$$

The joint distribution of l and F is essentially as for the case of an ending in a triangle treated in Section 4.3 and routine calculation gives

$$V_2 = \{(\sqrt{2} - 1) + \log(\sqrt{2} + 1)\} h/\pi.$$

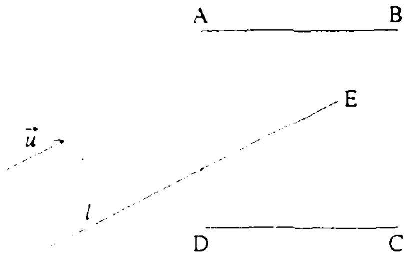


Figure 5.3: A line l ending in the square $ABCD$ in the dual lattice. The unit vector $\vec{u}$ is in the direction of l.

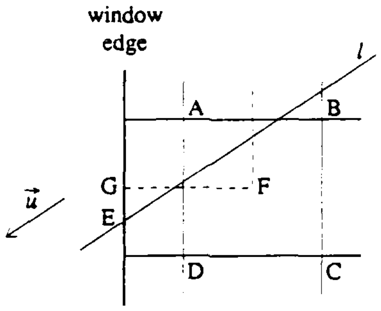


Figure 5.4: Line l crossing dual edge AD and leaving the window.
Dotted lines show edges present in the edge process.
The unit vector $\vec{u}$ is in the direction of l.

A case not yet mentioned is that of a line ending at the edge of the window. Figure 5.4 shows a line meeting the window edge after crossing edge AD of the dual square $ABCD$. To complete the irregular strip containing that part of l within the window we need to add edges AF and DF. Thus the cost of edge GF in the one edge clique associated with G is

$$V_2 = E(\tfrac{1}{2}(\vec{AF} + \vec{FD}).\vec{u} \mid l \text{ crosses } AD \text{ and then leaves the window}).$$

Strictly speaking, this depends on the position of G relative to the corners of the window. The calculation is simplified if we assume an infinite window edge, in which case $V_2 = \pi h/4$.

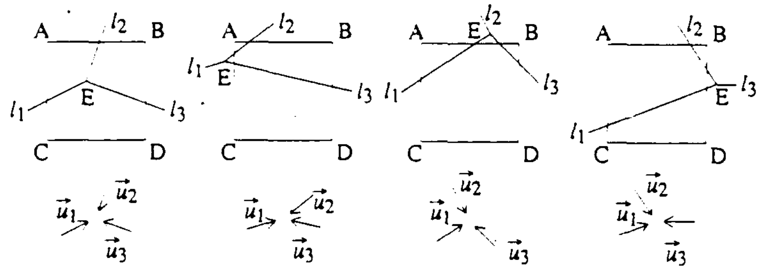


Figure 5.5: Three lines meeting in or near dual square $ABCD$
and associated unit vectors.

A branch (configuration 4) arises when three lines meet. In all four cases shown in Figure 5.5 the branch is associated with dual square $ABCD$ and the sum of the projection lengths required to close off the irregular strips containing l_1, l_2 and l_3 is

$$p_F = (\vec{DF} + \vec{AF}).\vec{u}_1 + (\vec{AF} + \vec{BF}).\vec{u}_2 + (\vec{BF} + \vec{CF}).\vec{u}_3 .$$

In calculating $V_2 = E(\frac{1}{2}p_F)$, the expectation is with respect to a uniform distribution of the point of intersection F in the plane and of the orientation of the set of lines l_1, l_2 and l_3, conditional on sides DA, AB and BC but not CD being intersected. For a given set of angles between the three lines containing no acute angle, calculation of V_2 by numerical integration or Monte Carlo estimation is straightforward. Note, however, that if the lines do meet at acute angles the associated edge process can be more complex than a single branch, as shown in Figure 5.6.

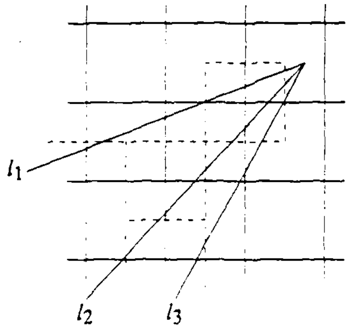


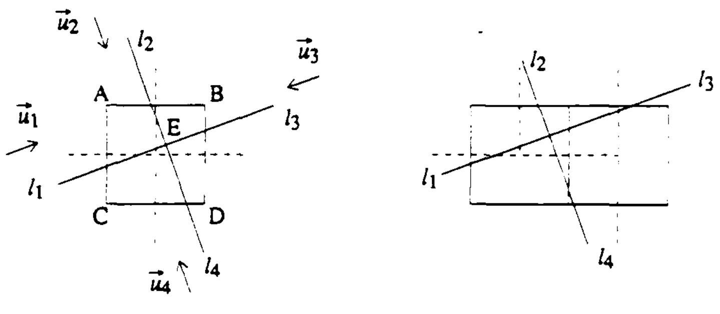
Figure 5.6: Lines l_1, l_2 and l_3 meeting at a point. The dotted lines representing elements of the associated edge process include a crossing and three branches.

For the special cases of three angles of $2\pi/3$ and angles of $\pi/2$, $\pi/2$ and π between the lines l_1, l_2 and l_3, $V_2 = 1.32h$ and $1.45h$ respectively. The value $V_2 = 1.4h$ associated with configuration 4 in Table 3.1 was chosen as a compromise between these two cases.

A crossing (configuration 5) can arise when four lines meet but this will not always be the case. In Figure 5.7b the meeting of four lines produces two adjacent branches rather than a crossing in the edge process: the projection costs calculated for a branch formed by the meeting of three lines are inappropriate in this case but a proper treatment would be possible if the clique size were enlarged. When the meeting of four lines does produce a crossing in the edge process the configuration must be of the type shown in Figure 5.7a, and the projection cost is

$$V_2 = E[\tfrac{1}{2}\{(\vec{DF} + \vec{AF}).\vec{u}_1 + (\vec{AF} + \vec{BF}).\vec{u}_2 + (\vec{BF} + \vec{CF}).\vec{u}_3 + (\vec{CF} + \vec{DF}).\vec{u}_4\}]$$

where F is distributed uniformly over the interior of $ABCD$ and the orientation of the set of four lines is uniform conditional on one line intersecting each edge. For the case of four lines meeting at right angles and producing a crossing in the edge process numerical integration gives $V_2 = 1.94h$ and this is the value used for configuration 5 in Table 3.1.



(a) (b)

Figure 5.7: Four lines meeting at a point. Dotted lines show elements of the associated edge process. Unit vectors $\vec{u}_1,...,\vec{u}_4$ are in the directions of lines $l_1,...,l_4$.

B. Rectangular Pixels

In this section we consider rectangular pixels of length h_1 and breadth h_2. Corresponding to the six possible types of configurations of edges in Figure 3.1 there are now nine possible essentially different configurations, since there are two types each of endings, continuations and branches. For brevity we shall concentrate on the costs for continuations and turns. The cost of a turn is calculated by applying formula (4.5) to a right-angled triangle with short sides h_1 and h_2, yielding the quantity $\frac{\pi}{8}h_1h_2/\{h_1 + h_2 - (h_1^2+h_2^2)^{1/2}\}$.

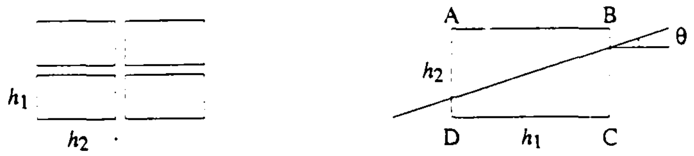


Figure 5.8: A configuration for a clique in the rectangular pixel case.

The cost of a continuation of the kind shown in Figure 5.8 is calculated as in Section 5.1. Let $\theta_0 = \tan^{-1}(h_2/h_1)$. For $|\theta| < \theta_0$ the set of lines of inclination θ intersecting AD and BC forms a strip of width proportional to $\sin(\theta_0-|\theta|)$, and the projection length of AB and CD on a line of inclination θ is $h_1 \cos\theta$. Hence arguments exactly analogous to those of Section 5A give as the cost of a "continuation" as shown in Figure 5.8 the quantity

$$\begin{aligned}
h_1 \int_0^{\theta_0} \cos\theta \sin(\theta_0-\theta) d\theta \Big/ \int_0^{\theta_0} \sin(\theta-\theta_0) d\theta \\
&= \tfrac{1}{2} h_1 \int_0^{\theta_0} \{\sin\theta_0 + \sin(\theta_0-2\theta)\} d\theta \Big/ \int_0^{\theta_0} \sin\theta' \, d\theta' \\
&= \tfrac{1}{2} h_1 \theta_0 \sin\theta_0 / (1-\cos\theta_0). \qquad (5.1)
\end{aligned}$$

The other type of continuation, consisting of two edges of length h_2, will cost an amount obtained by substituting h_2 for h_1 and $\pi - \theta_0$ for θ_0 in (5.1), viz. $\tfrac{1}{2}h_2(\pi-\theta_0) \cos\theta_0/(1-\sin\theta_0)$.

C. Hexagonal Pixels

The presence of a single type of first order neighbour makes the use of hexagonal pixel grids appealing, particularly in applications such as tomography where physical properties of the imaging system do not define a natural pixel grid. The dual space of a grid of regular hexagons with sides of length l contains equilateral triangles of side $l\sqrt{3}$. Applying the formulas of Section 4, the projection penalty and the edge length penalty for a continuation in the dual space are both equal to $\tfrac{1}{2}\pi l$. Since only one form of branching is possible, the region counting penalty and the expected projected edge length penalty for a branch will always appear together and it does not help to evaluate the latter quantity. Thus, the penalties for cliques of type 0, 1, 2 and 3, as depicted in Figure 4.2, are 0, $\tfrac{1}{2}\rho$, $\tfrac{1}{2}\pi l\beta$ and $\tfrac{1}{2}\rho$ respectively.

VI. CONCLUSION

Some simple geometrical considerations have made it possible to define edge process penalties which are approximately invariant to the scale and orientation of the pixel grid, and which can in addition be generalised to irregular pixels. The general idea of evolving penalties based on a conditional expected projection length has the advantage that consistent penalties can be written down for cliques of different kinds that appear in different parts of the same pattern.

ACKNOWLEDGMENTS

The financial support of the U. K. Science and Engineering Research Council, the European Research Office of the U. S. Army and the U. S. National Science Foundation are gratefully acknowledged. The authors warmly thank R. Sibson for his stimulating remarks on the use of the prior log probability as a penalty function, and R. H. Glendinning and G. Stone for their assistance with the examples and illustrations. Parts of this work were carried out during visits by T. C. Brown to the University of Bath and by B. W. Silverman to Stanford University.

REFERENCES

[1] L. Dorst and A. W. M. Smeulders, "Best linear unbiased estimators for properties of digitized straight lines," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 276-282, 1986.

[2] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721-741, 1984.

[3] J. G. Hocking and G. S. Young, *Topology*, Reading MA: Addison-Wesley, 1961.

[4] B. W. Silverman, "Penalized maximum likelihood estimation," *Encyclopedia of Statistical Sciences*, (S. Kotz and N. L. Johnson, editors) New York: Wiley-Interscience, vol. 6, pp. 664-667, 1985.

[5] B. W. Silverman, "Some aspects of the spline smoothing approach to nonparametric regression curve fitting," *J. Roy. Statist. Soc. Ser.. B*, vol. 47, pp. 1-52, 1985.

[6] B. W. Silverman, M. C. Jones, J. D. Wilson, and D. W. Nychka, "A smoothed EM approach to a class of problems in image analysis and integral equations," submitted for publication, 1988.

[7] Y. Vardi, L. A. Shepp, and L. Kaufman, "A statistical model for positron emission tomography," *J. Amer. Statist. Assoc.*, vol. 80, pp. 8-37, 1985.

Appendix 5

Flexible parsimonious smoothing and additive modeling.

by

J. H. Friedman and B. W. Silverman

1. Introduction

In this paper we shall develop an approach to regression fitting based on an extremely simple idea. Consider first of all the univariate case where one has N pairs of measurements (y_i, x_i), $i = 1, \cdots, N$, and it is supposed that, as usual,

$$Y = f(X) + \text{error} \tag{1}$$

where f is a function to be estimated, and the error is assumed to have zero mean; its distribution may well depend on the value of X.

Regression, or curvefitting, is performed for a number of reasons. The value $f(X)$ is the conditional expectation of Y given the value X, and so may be used as an estimate of the response Y for future observations where only the value of the predictor variable X is measured. The function f can also be studied to try to gain insight into the predictive relationship between Y and X. By far the most commonly used approach is, of course, *linear regression*. It is assumed - rightly or wrongly - that f is a linear function $f(X) = aX + b$, and then the parameters a and b are estimated by least squares.

What should be done if the data are not well approximated by a straight line fit? One way forward is to allow f to be a *piecewise linear* function, made up of straight line pieces that join together continuously at points called *knots*. If the knot positions are fixed before looking at the data response values y_i, then, at the expense of introducing more parameters into the problem, we will be able to fit a wider range of data sets reasonably well, while still including simple linear regression as a special case. Furthermore all the necessary parameters can be found and inference performed using standard linear regression methods (see Agarwal and Studden, 1980).

In terms of flexibility, much greater dividends arise if the knot positions are not fixed in advance, but are themselves allowed to depend on the data, including the response values. In this case an enormously wide range of models can be closely approximated using piecewise linear functions f with a small number of knots. There is a computational penalty to be paid, because some sort of search procedure needs to be used to find suitable positions for the knots. In this paper, we describe a stepwise procedure that makes it feasible to fit piecewise linear models with knot positions determined by the data, and we also discuss practical strategies for deciding how many knots to use.

One of the attractive features of our method is that it can very easily be extended to the multivariate case. Suppose that the observations are of the form $(y_i, \mathbf{x}_i)$ where each $\mathbf{x}_i$ is now a p-vector $(x_{1i}, x_{2i}, \cdots, x_{pi})$. It is assumed, as before, that the variable Y depends on $\mathbf{X}$ by a relation of the form

$$Y = f(\mathbf{X}) + \text{error} = f(X_1, X_2, \cdots, X_p) + \text{error}.$$

The way that we make use of our ideas about piecewise linear fitting is to concentrate on the case where f is a sum of functions of the individual components of $\mathbf{X}$,

$$f(\mathbf{X}) = f_1(X_1) + f_2(X_2) + \cdots + f_p(X_p). \tag{2}$$

This approach is known as *additive regression* or *additive modeling*, and replaces the problem of estimating a function f of a p-dimensional variable $\mathbf{X}$ by one of estimating p separate one-dimensional functions f_j. Although not completely general, additive models are often effective; they are easy to interpret, and represent a very important step beyond the simple linear model.

It turns out that our piecewise linear fitting method can be applied easily in the additive modeling context. Each of the individual functions f_j can be modeled as being piecewise linear with knots that depend on the data, including the response values. Our stepwise fitting procedure enables all the functions f_j to be constructed together, at little extra cost than for a univariate problem.

The paper is set out as follows. In Section 2.0 we give a general discussion of smoothing methods. We go on in Sections 2.1 and 2.2 to develop our approach in the univariate case. Computational aspects are discussed in Section 2.3. The important question of model selection – how many knots to use – is considered in Section 2.4. In Section 2.5 we provide a simple extension that produces models with continuous first derivatives (if desired). In Section 3 we explain how the additive modeling approach enables our method to be applied in the multivariate case, and in Section 4 we demonstrate how confidence intervals for the estimated function(s) can be obtained. Finally in Section 5 a number of practical examples display the scope and power of our method as a data-analytic tool.

2.0 Smoothing

We first consider the case of a single predictor variable, $p = 1$. The smoothing problem has been the subject of considerable study, especially in recent years. The lack of flexibility (ability to closely approximate a wide variety of predictive relationships) associated with global fitting

$$f_J(x) = a_0 + \sum_{j=1}^{J} a_j P_j(x) \tag{3}$$

where the P_j are predefined functions (usually involving increasing powers of x) has led to developments in two general directions: piecewise polynomials and local averaging. The basic idea of piecewise polynomials is to replace the single prescribed function $f_J(x)$ (of possibly high order J) defined over the entire range of X values, with several generally low order polynomials, each defined over a different subinterval of the range of X. The points that delineate the subintervals are called knots. The greater flexibility of the piecewise polynomial approach is gained at some expense in terms of local smoothness. The global function is generally taken to be continuous and have continuous derivatives to all orders. Piecewise polynomials on the other hand are permitted to have discontinuities in low order derivatives (and sometimes even the function itself) at the knots. The tradeoff between smoothness and flexibility is controlled by the number of knots at which discontinuities are permitted and the order of the lowest derivative allowed to be discontinuous. The most popular piecewise polynomial fitting procedures are based on splines (de Boor, 1978). An M-spline consists of piecewise polynomials of degree M constrained to be continuous and have continuous derivatives through order $M-1$. Smith (1982) presented an adaptable knot placement strategy for spline fitting based on forward/backwards variable subset selection.

Local averaging smoothers directly use the fact that $f(x)$ is intended to estimate a conditional expectation, $E(Y|x)$. These estimates take the form

$$f(x) = \sum_{i=1}^{N} H(x, x_i) y_i \tag{4}$$

where $H(x, x')$ (called the kernel function) usually has its maximum value at $x' = x$ with its absolute value decreasing as $|x' - x|$ increases. Therefore, $f(x)$ is taken to be a weighted average of the y_i, where the weights are larger for those observations that are close or local to x. A characteristic quantity associated with a local averaging procedure is the local span $s(x)$, defined to be the range centered at x over which a given proportion of the averaging takes place,

$$\int_{x-s(x)/2}^{x+s(x)/2} H(x, x')dx' = \alpha,$$

with α a predefined constant fraction (i.e., $\alpha = 0.68$ or 0.95). If the defining property holds for more than one value of $s(x)$, then the smallest such value is taken. Many local averaging smoothers take the span to be constant over the entire range of x, $s(x) = \lambda$, (Rosenblatt, 1971). Others take it to be inversely proportional to the local density of x values, $s(x) = \lambda/p(x)$ (Cleveland, 1979). Smoothing splines (Reinsch, 1967) are in fact local averaging procedures where the span turns out to be approximately $s(x) \simeq \lambda/[p(x)]^{1/4}$ (see Silverman, 1984, 1985). (The quantity λ represents a parameter of these procedures.) Recently, adaptable span local averaging smoothers have been introduced that estimate optimal local span values based on the values of the responses.

y_i. (Friedman and Stuetzle, 1982, Friedman, 1984). The span function $s(x)$ controls the continuity-flexibility tradeoff for local averaging smoothers. For the nonadaptable smoothers this is in turn regulated by λ, the smoothing parameter of the procedure.

There is, of course, a connection between the piecewise polynomial and local averaging approaches to smoothing. For a given knot placement, piecewise polynomial curve estimates can also be expressed in the form given by (4) (as can global fits). There will be a characteristic local span associated with the corresponding kernel. The more flexible the smoother is to local variation, the smaller will be the span. The basic difference between the two approaches has to do with how the span is specified. With local averaging smoothers the span parameter λ usually enters fundamentally into the definition of the kernel function (or some other aspect of the definition of the smoother) and is either directly set by the user or some automated procedure (i.e. cross-validatory choice) is employed for its selection. For piecewise polynomial smoothers it is indirectly regulated by the choice of the number and placement of the knots, and the degree of continuity required at the knot positions.

The trade-off between continuity and local flexibility is a fundamental one that directly affects the statistical performance of the smoother as a curve estimator. If one assumes that there exists a population from which the data can be regarded as a random sample, then the goal is to estimate the conditional expectation $E(Y|X = x)$ for the population. Even if this is not the case the goal is usually to obtain curve estimates $f(x)$ that have good (future) prediction ability for new observations not part of the training sample used to obtain the estimate.

Increased flexibility provides the smoothing procedure with an increased ability to fit the data at hand more closely. This may or may not be good, depending on the extent to which this training sample is representative of the population of future observations to be predicted. It is often the case that fitting the training data too closely results in degraded estimates with poor future performance. This phenomenon is called "over-fitting" and can be quantified through the bias-variance trade-off. The (future) expected-squared-error can be expressed as

$$E[f^*(x) - f(x)]^2 = [f^*(x) - Ef(x)]^2 + \text{Var} f(x). \tag{5}$$

where $f^*(x) = E(Y|X = x)$ for the population (future observations). The expected values in (5) are over repeated replications of the training sample. The first term on the right hand side of (5) is the squared distance of the average (expected) curve estimate from the truth. It is referred to as the "bias-squared" of the estimate. As the smoother is given more flexibility to fit the data, the bias-squared generally decreases while the variance increases. Thus, for each situation there is a (usually different) optimal flexibility. If a smoothing procedure is to provide good performance over a wide variety of situations, it must be able to effectively adjust its flexibility-continuity trade off for each particular application.

Motivated by the work of Smith (1982), we present an adaptable piecewise polynomial smoothing algorithm. It uses the data to automatically select the number and positions of the knots, and to some extent the degree-of-continuity imposed at the knots as well. Although quite simple the method has both operational and performance characteristics that are quite similar to the recently proposed adaptable span local averaging smoothers (Friedman and Stuetzle, 1981, Friedman, 1984). It appears to have superior performance in low sample size and/or high noise situations.

Our focus is on accurate estimation of the curve itself and not necessarily its derivatives. We therefore restrict our attention to low order polynomials with weak continuity requirements at the knots. This has the effect of minimizing the average effective span (see above) for a given number of knots. This is important if accurate solutions with a small number of knots are required. This will be the case in high noise small sample environments. Our simplest method employs piecewise linear fitting where only the function itself is required to be continuous. We also describe a companion method that fits with piecewise cubic functions where continuous first - but not second - derivatives are imposed. This has the advantage of producing curves that are more cosmetically appealing, if less interpretable. It may sometimes, but not always, produce slightly more accurate estimates in situations where the second derivative of the underlying true curve is nowhere rapidly varying.

Our estimate of future prediction error – to be minimized – is based on the generalized cross-validation measure (Craven and Wahba, 1979). A brief explanation of generalized cross-validation (GCV) is given by Silverman (1985, Section 4.1). To explain GCV it is first necessary to mention cross-validation (CV). Let K be the number of knots in the fitted model. The CV score is given by

$$CV = \frac{1}{N}\sum_{i=1}^{N}[y_i - f_{-i}(x_i)]^2$$

where f_{-i} is the estimate calculated with the current values of the control parameters (in our case the number of knots) from all the data points except the ith. The cross-validation score is then a function of K, and gives a measure of future prediction error that may unfortunately be laborious to calculate.

GCV can be thought of as an appropriate version of CV that has better computational properties. For a suitable increasing function $d(K)$ of the number of knots, the GCV score is defined by

$$GCV = \frac{1}{N}\sum_{i=1}^{N}[y_i - f(x_i)]^2/[1 - \frac{d(K)}{N}]^2. \tag{6}$$

If the knot placement values do not depend upon the sample response values y_i, then it can be shown that an appropriate choice of $d(K)$ is

$$d(K) = \sum_{i=1}^{N} H(x_i, x_i)$$

where H is the kernel function (4). For piecewise linear fitting by least squares with K knots, this turns out to be $d(K) = K + 1$. It can be shown that this choice of $d(K)$ makes GCV and CV identical in certain special cases.

For adaptable span smoothers, such as those we introduce in the present paper, the approximation is no longer good because of the additional flexibility given by the free choice of knot positions. To compensate for this, we use (6) as an approximation with $d(K)$ taken to be a more rapidly increasing function of K; we discuss our choice of $d(K)$ in Section 2.4 below.

2.1 Piecewise linear smoothing

We describe first piecewise linear fitting. For a fixed number of knots K, we aim to place the knots to give the minimum possible value of the average-squared-residual (ASR)

$$ASR = \frac{1}{N}\sum_{i=1}^{N}[y_i - f(x_i)]^2$$

for estimates $f(x)$ chosen to be continuous and piecewise linear with the given knots. Given a set of knot positions there are a number of ways to construct the corresponding piecewise linear fit that minimizes the ASR. These involve choosing a set of basis functions $b_k(x), 1 \leq k \leq K$, parameterized by the knot locations, that have the required continuity properties. The curve estimate is then taken to be

$$f(x) = a_0 + \sum_{k=1}^{K} a_k b_k(x). \tag{7}$$

The values of the coefficients $a_0, \cdots, a_k$ corresponding to the piecewise linear curve that minimizes the ASR, are obtained by a $(K + 1)$-parameter linear least-squares fit of the response Y on the basis function set $b_k(x)$.

There are a variety of basis function sets with the proper continuity properties for piecewise linear fitting. The most convenient for our purposes is the set

$$b_k(x) = (x - t_k)^+ \tag{8}$$

where t_k is the location of the kth knot and the superscript indicates the nonnegative part. The convenience of this basis stems from the fact that each basis function is parameterized by a single knot. Thus, adding, deleting, or changing the position of a knot affects only one basis function.

Optimizing the ASR over all possible (unequal) locations for the K knots is a fairly difficult computational task. We therefore consider the subset of locations defined by the distinct values realized by the data set. This has the effect of providing more potential knot locations, and thus more potential flexibility, in regions of higher data density and correspondingly less potential flexibility in sparser regions. This attempts to control the variance, since regions where the ratio of data points to knots is low can give rise to locally high variance in the curve estimate.

Even the (combinatorial) optimization of the ASR over this restricted set of locations is formidable owing to the large number, N, of potential basis functions from which the optimizing K must be chosen. We therefore adopt a stepwise strategy for knot placement. The first knot ($k = 1$) is placed at the position that yields the best corresponding piecewise linear fit. Thereafter, each additional knot is placed at the location that gives the best piecewise linear fit involving it and the $k - 1$ knots that have already been placed. Knots are added in this manner until some maximum number of knots ($K_{\max}$) have been positioned. This process yields a sequence of $K_{\max}$ models, each one with one more knot that the previous one in the sequence. That model in the sequence with smallest GCV as defined in equation (6) is chosen for further consideration. The number, $K_{\max}$, of models to be considered should be chosen so that the model minimizing the GCV is not too close to the end of the sequence. Owing to the forward stepwise nature of the procedure, it is possible for the GCV sometimes to increase locally as the sequence proceeds and then begin to decrease again. The bound $K_{\max}$ should be large enough so that the GCV associated with the last model is substantially larger than the minimizing one in the sequence.

The model (with K^* knots; $0 \leq K^* < K_{\max}$) that was found to minimize the GCV is next subjected to a backwards stepwise deletion strategy. Each of its knots is in turn deleted and the corresponding $K^* - 1$ knot model is fitted. If any of these fits results in an improved GCV, the one with the smallest is chosen, permanently deleting the corresponding knot. This procedure is then repeated on the new $K^* - 1$ knot model, deleting a knot if a better model is found. This continues until the deletion of any remaining knot results in a curve with higher GCV.

This knot deletion strategy can sometimes result in an improved model because of the nature of forward stepwise procedures. The first few knots must deal with the global nature of the curve without the benefit of the additional knots that come later. They are, therefore, forced to ignore the fine structure. Knots that are added later in order to model the fine structure can in aggregate also account for the global structure, thereby causing the initial few knots to be redundant.

Knot deletion as described above seldom results in a dramatic improvement in GCV. It is worth doing for the small to moderate improvement it sometimes provides, because it adds almost nothing to the computational burden. All necessary calculations can be done using summary statistics (basis covariance matrix and response covariance vector) already calculated for the original K^*-knot model. No further passes over the data are required.

2.2 Minimum span

A natural strategy would be to make every distinct observation abscissa value a candidate location for knot positioning. This would correspond to allowing the minimum local effective span to include only a single observation. In low noise situations such a strategy can give reasonable results. In high noise environments, however, this can lead to unacceptably high local variance. A solution is to impose a minimum effective span by restricting the eligible knot locations. The simplest implementation is to make every (distinct) Mth observation (in order of ascending x-value) eligible for knot placement. This implementation also reduces computation by a factor of N/M in the absence of ties.

A reasonable value for M, as a function of N, can be obtained by a simple coin tossing argument. Suppose $y_i = f^*(x_i) + \varepsilon_i$, $1 \leq i \leq N$, where ε_i is a mean zero random variable with a symmetric distribution. Then ε_i has an equal chance of being positive or negative. A smoother will be resistant to a run of length L of either positive or negative errors so long as its span in the region of the run is large compared to L. If not, the smoother will tend to follow the run and hence incur increased (variance) error. A piecewise linear smoother can completely respond to a run without degrading the fit in any other region (irrespective of the placement of the other knots) if it can place three knots within its length. It can partially respond with two knots in the run for an unfavorable placement of the other knots (i.e. one of them close to the start or end of the run). This would suggest that the minimum knot increment M should satisfy $M > L_{\max}/3$ (or $M > L_{\max}/2.5$ to be conservative) where $L_{\max}$ is the largest positive or negative run to be expected in N binomial trials.

Let $Pr(L)$ be the probability of observing a run of length L or longer in N tosses of a fair coin. For small values of this probability a close upper bound is given by

$$Pr(L) = 2^{1-N} \sum_{j=L}^{N} \sum_{i=1}^{j/L} (-1)^{i+1} \binom{N-j+1}{i} \binom{N-iL}{N-j} \qquad (9)$$

(Bradley, 1968). One can choose a value α for this probability

$$Pr(L) = \alpha \tag{10}$$

(say $\alpha = 0.05$ or 0.01) and solve (9), (10) for the corresponding length $L(\alpha)$. Setting $M = L(\alpha)/2.5$ would (with probability α) give resistance to a run of positive or negative error values. Solving (9), (10) for $L(\alpha)$ would have to be done numerically. It turns out that the simple formula

$$L(\alpha) = -\log_2[-\frac{1}{N} ln(1-\alpha)]$$

approximates the solution quite closely (within a few percent) for $\alpha < 0.1$ and $N \geq 15$. This suggests that a conservative increment for knot placement is given by

$$M(N, \alpha) = -\log_2[-\frac{1}{N} \ell n(1-\alpha)]/2.5 \tag{11}$$

with $0.05 \leq \alpha \leq 0.01$.

2.3 Computational Considerations

For each $k > 0$, at the kth step in the forward stepwise procedure described in Section 2.1 it is necessary to optimize the position of the kth knot (over all eligible locations) given the positions of the $k-1$ previously placed knots. For a given knot placement increment M there are (in the absence of ties) $N/M - k + 1$ eligible places to position the kth knot. (The positions of the $k-1$ previously placed knots are not eligible.) At each such potential new knot location a linear least-squares fit must be performed to obtain the corresponding piecewise linear smooth and its associated ASR. Thus approximatley N/M linear least-squares fits must be computed to place each knot. If this were implemented in a straightforward manner it would give rise to prohibitive computation in all but the richest computing environments. Enormous computational gains can be realized, however, by examining the set of eligible knot locations in a special order that permits the use of rapid updating formulae associated with the basis (8). This strategy involves visiting the potential knot positions in descending abscissa value and taking advantage of the fact that (for $t' \geq t''$)

$$(x - t'')^+ - (x - t')^+ = \begin{cases} 0 & x \leq t'' \\ x - t'' & t'' \leq x \leq t' \\ t' - t'' & x > t' \end{cases} \tag{12}$$

The linear least-squares fit for the kth knot (located at $t_k = t''$) can be accomplished by solving the normal equations

$$Ba = c \tag{13}$$

where B is the $k \times k$ covariance matrix of the k basis functions (8),

$$B_{j\ell} = \sum_{i=1}^{N} b_\ell(x_i)[b_j(x_i) - \bar{b}_j], \tag{14a}$$

and c is the k-dimensional covariance vector of the response with each basis function

$$c_j = \sum_{i=1}^{N} (y_i - \bar{y}) b_j(x_i). \tag{14b}$$

Here $\bar{b}_j$ and $\bar{y}$ represent the averages of the corresponding quantities. The solution vector $a = (a_1, \cdots, a_k)$ represents the coefficients corresponding to the optimizing piecewise linear fit (7) given the knot locations $t_1, \cdots, t_k$. The ASR of the fit is then given by

$$ASR = \operatorname{Var}(Y) - \sum_{j=1}^{p} a_j c_j / N. \tag{14c}$$

Using (13), (14) as prescriptions for computing the corresponding quantities at each potential knot location leads to the prohibitive computation mentioned above. The first thing to notice in attempting to save computation is that only c_k and B_{jk}, $1 \le j \le k$ need to be recomputed since only the kth knot location is changing. (This reduces the computation by a factor of k.) The next thing to note is that if these quantities have already been computed for a knot located at $t_k = t'$ then, from (12), a simple series of updates gives them for a knot located at $t_k = t''$ $(t'' < t')$. Let

$$s_0 = \sum_{x_i \ge t'} (y_i - \bar{y}),$$

$$s_j = \sum_{x_i \ge t'} (b_j(x_i) - \bar{b}_j), \quad 1 \le j \le k - 1,$$

$$u = \sum_{x_i \ge t'} 1, \quad \text{and} \quad v = \sum_{x_i \ge t'} x_i.$$

Then

$$c_k(t'') = c_k(t') + (t' - t'')s_0 + \sum_{t'' \le x_i < t'} (x_i - t'')(y_i - \bar{y}),$$

$$B_{jk}(t'') = B_{jk}(t') + (t' - t'')s_j + \sum_{t'' \le x_i < t'} (x_i - t'')(b_j(x_i) - \bar{b}_j), \quad 1 \le j \le k - 1$$

$$B_{kk}(t'') = B_{kk}(t') - (t'^2 - t''^2)u + 2(t' - t'')v + \sum_{t'' \le x_i < t'} (x_i - t'')^2$$

gives the quantities that enter into the normal equations (13) for $t_k = t''$, given their values at $t_k = t'$. All values are initialized to zero (i.e. $c_k(x_N) = B_{jk}(x_N) = 0$, $1 \le j \le k$).

These updating formulae provide the ingredients for the normal equations (13) at all potential knot locations with total computation of order kN. It remains to solve the normal equations at the (approximately N/M) eligible locations for knot placement. This can be done most rapidly by using the Cholesky decomposition of B followed by back-substitution (see Golub and Van Loan, 1983). Since only the last row and column of B are changing, its Cholesky decomposition can be updated with k^2 computations (Golub and Van Loan, 1983). The back substitution can also be performed in k^2 computation. Therefore the dominant part of the computation for optimizing the ASR with respect to the position of the kth knot is of order k^2N/M. The computation associated with a *single* linear least-squares fit is of order k^2N. Therefore, the updating strategy permits the implicit evaluation of N/M linear least squares fits with less computation than a single such fit. The entire procedure for placing all $K_{\max}$ knots in the forward stepwise procedure requires roughly the same computation as $K_{\max}/3$ linear least squares fits with $K_{\max}$ variables.

The computational strategy outlined above emphasizes speed over numerical stability. First of all, the one sided basis (8) is known to have poor numerical properties compared to other possible representations of piecewise linear functions (de Boor, 1978). Their advantage lies in the fact that each basis function is characterized by a single knot. This leads to the simple and rapidly computable updating formulae derived above. A second compromise is the choice of the normal equations with the Cholesky decomposition of the basis covariance matrix to perform each linear least-squares fit. It is well known that using the QR decomposition of the basis "data" matrix would provide superior numerical properites (see Golub and Van Loan, 1983). Unfortunately, updating the QR decomposition requires computation proportional to kN (compared to k^2 for the Cholesky strategy) which would cause the total computation to be proportional to N^2.

Potential numerical difficulties associated with this particular strategy are mitigated by two factors. First the minimal span requirement (11) limits somewhat the correlation between basis functions (8) associated with adjacent knots. Second, for sample sizes that are not extremely large, the number of knots is generally quite small, keeping the size of the associated least-squares problem small. Numerical problems tend only to arise when this strategy is applied to very large problems (typically $N > 500$) for which the resulting solution is a very complex curve requiring a great many knots. For these cases numerical stability can be achieved by slightly deoptimizing the least-squares fit (13) at each potential location for the kth knot. The basis coefficients $a = (a_1, \cdots, a_k)$ of the piecewise linear fit are taken to be the solution to

$$(B + \epsilon I)a = c,$$

with I being the $k \times k$ identity matrix, and the value of ϵ chosen to be just large enough to maintain numerical stability. Although these coefficient values can be somewhat different from those produced by (13) in highly collinear settings, they produced nearly identical curve estimates (7). The criterion used to select the best knot location is still the ASR. Typically, taking

$$\epsilon = 10^{-5} \text{ trace } B/k$$

maintains stable computation while having very little effect on the resulting curve estimate.

2.4 Model Selection

In order to implement the forwards/backwards stepwise knot placement strategy described in Section 2.1 it is necessary to have an estimate of the future prediction error. For procedures that are linear in the responses (4) a variety of estimators (model selection criteria) have been proposed (Akaike, 1970, Mallows, 1973, Craven and Wahba, 1979, Shibata, 1980, Breiman and Freedman, 1983). For a *given* knot placement (fixed set of regression variables) our method is linear in the responses. However, we use the response values to determine where to place the knots. As a result our curve estimator is not linear in the responses ($H(x, x_i)$ depends upon $y_1 \cdots y_n$). There is increased variance in the curve estimates corresponding to the variability associated with the knot placement that is not incorporated into the above criteria. For nonlinear procedures, techniques based on sample reuse (Cross-validation, Stone, 1974, and Bootstrap, Efron, 1983) are appropriate. These require considerable computation, however, and a common practice is simply to ignore the increased variability associated with model selection. If the number of selected variables is not very much smaller than the size of the initial set, the increased variance is not large, and such a strategy may be effective. In our situation, however, this is not the case. We intend to select a few knots usually from a very large number of potential locations.

The basis for our model selection strategy lies in the work of Hinkley (1969, 1970) and Feder (1975). They consider the problem of testing the hypothesis that a two-segment piecewise linear regression function in fact consists of only a single segment, in the presence of normal homoscedastic errors. Specifically, it is assumed that

$$Y_i = a + bX_i + c(X_i - t)^+ + \varepsilon_i \qquad (15)$$

with $\varepsilon_i \sim N(0, \sigma^2)$, and one wishes to test the hypothesis that $c = 0$. If the knot location t is specified in advance then (under the null hypothesis $H_0 : c \equiv 0$) the difference between the (scaled) residual sums of squares from the respective two and three parameter least-squares fits follows a chi-squared distribution on one-degree-of-freedom, χ_1^2. That is, the additional parameter, c, uses one additional degree-of-freedom.

When one adjusts the knot location t, as well as the coefficient c, then this is no longer the case. Furthermore, under the condition $c = 0$ the parameter t is not identifiable, and so we cannot use the usual asymptotic theory and just add a degree-of-freedom for the additional fitted parameter t. Feder (1975) shows that (under $H_0 : c \equiv 0$) the difference between the residual sum-of-squares from the respective two and four parameter fits asymptotically follows the distribution of the maximum of a large number of correlated χ_1^2 and χ_2^2 random variables. Furthermore, the precise correlational structure (and thus the distribution) depends on the spacings of the observations. Such a distribution will give rise to considerably larger test statistic values than χ_1^2 and generally larger values than even χ_2^2. That is, the additional parameter t uses *more* than one additional degree-of-freedom. Hinkley (1969, 1970) reports strong empirical evidence that the distribution closely follows a chi-squared on three degree-of-freedom. Thus, fitting both the additional coefficient, c, and the corresponding knot location, t, uses about *three* additional degrees-of-freedom.

A similar effect was reported by Hastie and Tibshirani (1985) in the context of projection pursuit regression (Friedman and Stuetzle, 1981). Here the model is

$$y_i = g(\sum_{j=1}^{p} \alpha_j x_{ji}) + \varepsilon_i,$$

with $\varepsilon \sim N(0, \sigma^2)$, and g is a smooth function whose argument is a linear combination of the p predictor variables. The objective is to minimize the residual sum of squares jointly with respect to the parameters defining both the function and the linear combination in its argument. The null hypothesis H_0 is that g is a constant function. Hastie and Tibshirani (1985) performed a simulation experiment to obtain the distribution of the scaled difference of the residual sum of squares as a function of the number of parameters associated with the function g, for $p = 5$ and $N = 360$. They found that the expected value of this distribution was always greater than the sum of the number of parameters associated with both the curve and the linear combination (except for the degenerate case - g linear). This effect became more pronounced as more parameters were associated with g. These results, together with those of Hinkley (1969, 1970) and Feder (1975), indicate that the number of degrees-of-freedom associated with nonlinear least-squares regression can be considerably more than the number of parameters involved in the fit.

Our knot placement strategy does not perform an unrestricted minimization, but rather minimizes the ASR over a restricted set of eligible knot locations. In the absence of a large number of ties, however, the solution value for the ASR is not likely to be a great deal different. Thus, following Hinkley (1969, 1970) and associating a loss of three degrees-of-freedom for each knot adaptively placed (with our strategy) seems reasonable, if a bit conservative. We therefore use

$$d(K) = 3K + 1, \tag{16}$$

in conjunction with the generalized cross-validation estimate of future prediction error (6), as a model selection criterion to be minimized.

2.5 Piecewise cubic fitting

Continuous piecewise linear curves provide maximum flexibility for a given (small) number of knots. They also have the advantage of ready interpretation: linear relationship within subintervals of the range of X. Their principal disadvantage is the discontinuity of the first derivative (infinite second derivative) at each knot location. This causes the curve to be cosmetically unappealing to some.

Also, if the true underlying function $f^*(x)$ (5) does not have a locally high second derivative close to a knot location, then a piecewise linear approximation will exhibit a small increased error in the neighborhood near that knot. (This is in contrast to the corresponding first, and especially, the second derivative estimates which contain much larger errors.) If the second derivative of $f^*(x)$ is everywhere slowly varying then (slightly) more accurate curve estimates can be obtained by restricting the variation of the second derivative. This is at the expense of reduced flexibility to fit curves that do have locally rapidly varying second derivatives.

The same considerations (see Section 2.0) that led to the desirability of piecewise linear approximations guide our approach to piecewise cubic fitting. We seek a curve estimate whose function and first derivative values are everywhere continuous. Under that constraint we would like an estimate that closely resembles the corresponding piecewise linear fit. In particular, we do not wish to require, in addition, everywhere continuous second derivatives.

A simple modification of our basis functions (8) (used for piecewise linear fitting) leads to an appropriate basis for the corresponding piecewise cubic approximation:

$$B_k(x) = \begin{cases} 0 & x \leq t_{k-} \\ q_k(x - t_{k-})^2 + r_k(x - t_{k-})^3 & t_{k-} < x < t_{k+} \\ x - t_k & t_{k+} \leq x \end{cases} \tag{17}$$

with $t_{k-} < t_k < t_{k+}$.

Setting the coefficients q_k and r_k to

$$\begin{aligned} q_k &= (2t_{k+} + t_{k-} - 3t_k)/(t_{k+} - t_{k-})^2 \\ r_k &= (2t_k - t_{k+} - t_{k-})/(t_{k+} - t_{k-})^3 \end{aligned} \tag{18}$$

causes $B_k(x)$ (17) to be everywhere continuous and have continuous first derivatives. Outside the interval $t_{k-} < x < t_{k+}$, $B_k(x)$ is identical to the corresponding piecewise linear basis function $b_k(x)$ (8) with a knot at t_k. Inside the interval $B_k(x)$ is a cubic function whose average first and second derivatives (over the interval) match those for the corresponding $b_k(x)$. The second derivatives of of $B_k(x)$ exhibit discontinuities at t_{k+} and t_{k-}. Far from the central knot location t_k, $B_k(x)$ has the same properties as $b_k(x)$, so that both bases will have similar characteristic spans (see Section 2.0). Close to the central knot (inside $[t_{k-}, t_{k+}]$) $B_k(x)$ is an approximation to $b_k(x)$ with continuous first derivative.

Knot placement based on piecewise linear fitting (Sections 2.1 - 2.4) is used to select knot locations for piecewise cubic fits. The resulting knot locations $t_1 \cdots t_K$ are used as the central knots for the cubic basis $B_1(x) \cdots B_K(x)$ (17). The side knots $\{t_{k-}, t_{k+}\}, 1 \le k \le K$, are placed at the midpoints between the central knots. Let $t_{(1)} \cdots t_{(K)}$ be the central knots in ascending abscissa value. Then

$$\begin{aligned} t_{(k)-} &= (t_{(k)} + t_{(k-1)})/2 \\ t_{(k)+} &= (t_{(k)} + t_{(k+1)})/2 \end{aligned} \tag{19}$$

for $2 \le k \le K - 1$. The extreme knot locations, t_{1+} and t_{K-} are defined as in (19). The outer side knots are defined by

$$\begin{aligned} t_{(1)-} &= (t_{(1)} + x_{(1)})/2 \\ t_{(K)+} &= (t_{(K)} + x_{(N)})/2 \end{aligned} \tag{20}$$

where $x_{(1)}$ and $x_{(N)}$ are respectively the lowest and highest sample abscissa values. If the knot placement procedure happens to put a knot at $x_{(1)}$ (pure linear term in the model) then the corresponding basis function is taken to be $B_{(1)}(x) = x - x_{(1)}$.

The piecewise cubic curve estimate

$$f_c(x) = a_0 + \sum_{k=1}^{K} a_k B_k(x) \tag{21}$$

is obtained by minimizing the ASR with respect to the coefficients $a_0 \cdots a_K$. In the interior, $t_{(1)-} < x < t_{(K)+}$, it is piecewise cubic with second derivative discontinuities at the midpoints between the central knots $t_{(k)+} = t_{(k+1)-}$, $1 \le k \le K - 1$. In the outer regions, $x \le t_{(1)-}$ or $x \ge t_{(K)+}$, the curve estimate is taken to be linear. This helps to control the high variance associated with the extremes of the interval ("end effects").

Although the piecewise cubic fit seldom provides a dramatic improvement, it requires very little computation (one additional linear least squares fit) beyond that required for the (piecewise linear) knot placement. One can compare the GCV (6) (16) (equivalently, the ASR) for the piecewise linear and cubic estimates, choosing the one that is best. If a strong prejudice exists for continuous first derivatives, then one might prefer the cubic estimate even if it provides a slightly poorer fit to the data.

3.0 Additive modeling

The simplest extension of smoothing to the case of multiple predictor variables, $X_1 \cdots X_p$, is the additive model (2). Flexible additive regression has been the focus of considerable recent interest. It is a special case of the projection pursuit regression model ("projection selection", Friedman and Stuetzle, 1981). It also represents special cases of the ACE (Breiman and Friedman, 1985) and generalized additive models (Hastie and Tibshirani, 1984, 1986). Stone and Koo (1985) suggest additive modeling based on a central cubic spline approximation, with linear approximation past the extremes, and nonadaptive knot placement.

The smoothing procedure described in the previous section has a natural extension to multiple predictor variables. The piecewise linear basis functions analogous to (8) become

$$b_k(x) = (x_{j(k)} - t_k)^+ \tag{22}$$

where k, $1 \leq k \leq K$, labels the knots and $j(k)$, $1 \leq j(k) \leq p$, labels a predictor variable corresponding to each knot. Each knot location t_k is associated with a particular predictor variable, $j(k)$, and all of the predictor variables provide eligible locations for knot placement. Additive modeling in this context can simply be regarded as a (univariate) smoothing problem with a larger number (pN versus N) of ordinate abscissa pairs. The forward/backward knot placement strategy, minimum span (with pN replacing N), and model selection criteria directly apply, as do the updating formulae derived in Section 2.3 (reinitialized to zero for each new variable). The resulting piecewise linear model

$$f(x) = a_0 + \sum_{k=1}^{K} a_k(x_{j(k)} - t_k)^+ \tag{23}$$

can be cast into the form given by (2) with

$$f_i(x_i) = \sum_{j(k)=i} a_k(x_i - t_k)^+, \quad 1 \leq i \leq p. \tag{24}$$

Note that the means of the individual (predictor) variable functions (24) can be considered arbitrary for purposes of interpretation.

The corresponding piecewise cubic basis (17) is constructed in a manner analogous to that for the smoothing problem ($p = 1$). The only difference is that the side knots $t_{(k)-}, t_{(k)+}$ (19) are positioned at the midpoints between the central knots (t_k) defined on the *same* variable. The end knots (20) are positioned using the corresponding endpoints on the same variable. The resulting basis functions $B_k(x_{j(k)})$ define individual variable functions analogously to (24)

$$f_i(x_i) = \sum_{j(k)=i} a_k B_k(x_i), \quad 1 \leq i \leq p, \tag{25}$$

again with arbitrary means.

Although exceedingly simple, this method of additive modeling has some powerful characteristics. The knot placement strategy considers each potential knot location in conjunction with all existing knots on all the predictor variables - not just those defined on the same variable - when deciding whether to add (or delete) a particular knot. At each point the forward stepwise strategy decides (in a natural way) whether to increase the flexibility of an already existing variable curve (24) (25) or whether to add another variable, either linearly or nonlinearly. Variable subset selection thereby occurs as a natural byproduct of this approach. Note that the smallest abscissa value on each predictor variable *is always made eligible for knot placement* (irrespective of the minimum span value - Section 2.2) so that any predictor variable can potentially enter in a purely linear way.

The additive modeling strategy outlined above places no special emphasis on linearity. A purely linear relationship in any variable is represented by one of the eligible knot locations (the first) on that variable. One can (if desired) place such special emphasis by requiring that the first knot entered for each variable be at its smallest value. The price paid for this is increased variance in estimating some monotone relationships and dramatically increased bias against non-monotone relationships.

Our strategy does, however, place some special emphasis on monotonicity. Monotone trends will enter before somewhat stronger highly nonmonotone relationships. Also, there is a slight preference for certain types of monotone trends, namely those that start with a small slope. These can be approximated with a single knot as can a purely linear trend.

This method of additive modeling is invariant to the locations and individual spreads of the variables. Translating or rescaling each of the variables by a (different) constant factor will, in principle, not affect the solution. If, however, the predictor variables have very large absolute locations (compared to their scales) and/or wildly different scales, there can be undesirable numerical consequences associated with the updating and least-squares fitting. In such cases (as with ordinary linear least-squares regression) it is wise to center and/or rescale the predictor variables to remove the large locations and/or wild scale differences before applying the modeling procedure. The resulting solution is easily transformed back to the original variable locations and scales.

4.0 Confidence intervals

When attempting to interpret the individual predictor variable curve estimates, it is important to have a notion of how far the estimate is likely to deviate from the true underlying (population) conditional expectation. This can be quantified by the expected squared error

$$E[f_i^*(x_i) - f_i(x_i)]^2 = (f_i^*(x_i) - Ef_i(x_i))^2 + \text{Var} f_i(x_i). \tag{26}$$

Here $f_i^*(x_i)$ is the true population curve and $f_i(x_i)$ is the estimate from the sample. The expected values in (26) are over repeated samples of size N drawn from the population distribution. For linear (nonadaptable) procedures (knots fixed in advance) and homoscedastic errors (1), one can estimate the variance term in (26) through standard formulae for the covariances of the a_k appearing in (24) and (25) and an estimate of the true underlying error variance, $\hat{\sigma}^2$. With adaptable procedures such as ours this can be highly overoptimistic because it does not account for the variability associated with the knot placement.

One way to mitigate this effect is to inflate $\hat{\sigma}^2$ to account for the additional degrees-of-freedom used by the adaptive knot placement (total of three for each knot). Even this, however, does not give completely satisfactory results. For example, the (constant) predictor variable curves associated with no knots would be calculated to have zero variance. This is clearly not the case. In addition, there is seldom reason to expect homoscedasticity. Even if one could accurately estimate the variance it is, in any case, only one part of the expected-square-error. There is still the unknown and potentially large bias-squared term in (26).

Bootstrapping (see Efron and Tibshirani, 1986) provides a means of estimating the variance of the curve estimates (assuming only independence) and can give some indication of the bias as well. This is, of course, at the expense of additional computing. However, the additive modeling procedure described here is generally fast enough (see Section 2.3) to permit substantial bootstrapping, and honest uncertainty estimates are usually worth it.

The basic idea underlying the bootstrap is to substitute the sample for the population and study the behavior of estimates under repeated samples of size N drawn from it. In particular, we can estimate the expected squared error (26) by

$$\hat{E}[f_i^*(x_i) - f_i(x_i)]^2 = E_B[f_i(x_i) - f_i^{(B)}(x_i)].^2 \tag{27}$$

Here E_B is the expected value over repeated "bootstrap" samples of size N drawn (with replacement) from the data, and $f_i^{(B)}$ is the (ith) curve estimate for the bootstrap samples. In fact, one can approximate the distribution of $f_i^*(x_i) - f_i(x_i)$ by that of $f_i(x_i) - f_i^{(B)}(x_i)$.

Our goal is to take maximal advantage of the flexibility of the bootstrap to estimate asymmetric intervals about the curve that reflect the potentially asymmetric nature of the distribution of $f_i^*(x_i) - f_i(x_i)$. This can be due to either asymmetric error distribution or biased curve estimates (or both). In addition, we wish our interval estimates to reflect (probable) heteroscedasticity of the errors. To this end we repeatedly draw bootstrap samples (of size N with replacement) from the data. For each such sample we perform the same modeling procedure as was applied to the original data, thereby obtaining a set of curve estimates $f_i^{(B)}(x_i), 1 \le i \le p$. At each (original data) value, x_i, two averages are computed:

$$e_+^2(x_i) = E_B^{(+)}[f_i(x_i) - f_i^{(B)}(x_i)]^2 \tag{28a}$$

$$e_-^2(x_i) = E_B^{(-)}[f_i(x_i) - f_i^{(B)}(x_i)]^2. \tag{28b}$$

The first average (28a) is over those bootstrap replications for which $f_i(x_i) - f_i^{(B)}(x_i) > 0$, and the second (28b) is over those for which $f_i(x_i) - f^{(B)}(x_i) < 0$. The individual averages so obtained at each value of $x_i, e_\pm^2(x_i)$, are then smoothed against x_i using a simple (constant span) running average smoother. The resulting smoothed estimates $\hat{e}_\pm^2(x_i)$ are then used to define confidence intervals about the original data estimate $f_i(x_i)$:

$$f_i^{(\pm)}(x_i) = f_i(x_i) \pm \sqrt{\hat{e}_\pm^2(x_i)}. \tag{29}$$

In addition to assessing the variability of the individual predictor variable curve estimates $f_i(x_i)$, it is important to obtain a realistic estimate of the future prediction error, FPE, of the entire additive model (2),

$$FPE = E[Y - \sum_{i=1}^{p} f_i(x_i)]^2.$$

Here the expected value is over the population joint distribution of the response and predictor variables. Sample reuse techniques such as bootstrapping (Efron, 1983) and cross-validation (Stone, 1974) provide a variety of such estimates. Of these, the so-called "632-bootstrap" has shown superior performance in several simulation studies (Efron, 1983, Gong, 1982, Crawford, 1986). This estimate is a convex combination of two different estimates

$$FPE_{632} = 0.632 FPE_{\setminus B} + 0.368 ASR. \tag{30}$$

The second, ASR, is the average squared residual corresponding to the original data fit. The first estimate, $FPE_{\setminus B}$, is obtained from bootstrap sampling. As a consequence of the random nature of selecting observations for the bootstrap samples, a (different) subset of the observations will fail to be selected to appear at all in a particular bootstrap sample. On average, 0.368 N data observations will not contribute in this way to a bootstrap sample. Each time an observations does not so appear, its prediction error (squared) is computed, based on the model estimated from the corresponding bootstrap sample from which it is absent. The quantity $FPE_{\setminus B}$ is the average of these prediction errors over all such left out observations throughout the entire sequence of bootstrap replications.

The bootstrapping procedure outlined above simulates situations where the response and predictors are both random variables sampled (independently) from some joint distribution. That is, if another sample were to be selected, different values of the predictor variables as well as the response would be realized. Therefore, the resulting confidence interval and FPE estimates are not conditional on the design (realized set of predictor values). This is appropriate in most observational settings. There are situations, however, where the design is presumed to be fixed. That is, every replication of the experiment results in an identical set of values for the predictor variables and only the responses are random. Bootstrapping (as outlined above) will tend to over estimate both the confidence intervals and the FPE in fixed design situations (just as estimates conditioned on the design underestimate them for observational settings). Therefore, if the design is fixed these bootstrap estimates should be regarded as conservative.

5.0 Simulation studies and data examples

In this section we compare the technique outlined in the previous sections (referred to for identification as the "TURBO" smooth/model) to some other methods commonly used for smoothing and additive modeling through a limited simulation study and application to data. The goal is to identify those settings in which this procedure can be expected to provide good performance when compared to existing methodology. For the smoothing problem ($p = 1$) we compare with smoothing splines (Reinsch, 1967), a popular nonadaptive local averaging method, and a recently proposed adaptive span smoother, "SUPER SMOOTHER", (Friedman, 1984). With smoothing splines the roughness penalty was automatically chosen through generalized cross-validation (Craven and Wahba, 1979). For additive modeling we make comparisons with the projection selection/ACE approach using SUPER SMOOTHER. In all examples, the knot placement increment is given by (14) with $\alpha = 0.05$.

5.1 Smoothing pure noise

This is a simulation study to compare how well these three smoothers estimate a *constant* function in the presence of homoscedastic noise. That is, how much structure do they estimate when there is no underlying structure in the population? A set of response-predictor pairs (x_i, y_i), $1 \leq i \leq N$, were generated, with $0 \leq x_i \leq 1$ randomly sampled from a uniform distribution, and the y_i drawn from a standard normal distribution. Figures 1a, 1b, and 1c show a scatter plot of one such sample ($N = 20$) with the corresponding TURBO, smoothing spline, and SUPER smooths, respectively, superimposed. The TURBO curve estimate is seen to be a constant (no knots) equal to the sample response mean. The smoothing spline and SUPER SMOOTHER estimates show a gentle dependence on x.

Since one cannot discern expected performance based on one realization, we study average performance over 100 such realizations, for each of $N = 20$ and $N = 40$. The results are shown in Figures 1d and 1e respectively; for the larger sample size the errors are generally smaller, but the qualitative comparisons are the same. In both cases the average absolute error is plotted as a function of abscissa value. (For the TURBO smoother, the piecewise linear and cubic smooths give almost identical results.) The TURBO smoother (solid line) is seen to give uniformly smaller average error than the other methods, though of course this overall performance is mostly due to the relative amount of smoothing chosen (automatically) by the method rather than to the choice of method itself. Perhaps of more interest is the uniformity of the error across the range of observations; for this problem in particular, TURBO seems not to exhibit large error near the ends of the interval ("end effects") associated with the other methods. The especially poor performance of SUPER SMOOTHER (dashed line) in very high noise environments has been noted before (Breiman and Friedman, 1985). It is also known, as most easily seen by considering the "equivalent kernel" formulation discussed by Silverman (1984), that the smothing spline will have higher variance near the ends. Also, the smoothing spline can be affected by bias effects if the true underlying curve does not satisfy appropriate boundary conditions (see Rice and Rosenblatt, 1983); Agarwal and Studden (1980) showed that these end bias effects are not felt if one uses piecewise polynomial models with fixed knots, but since the underlying model is constant in this case, the bias effects are not relevant. It is clear that further theoretical work will be required to understand TURBO's apparent improvement in boundary behavior over other methods.

5.2 Smoothing a monotonic function

Our next example increases the complexity of the problem slightly. Here $N = 25$ response-predictor pairs (x_i, y_i) were generated according to the prescription

$$y_i = \exp(6x_i) + \varepsilon_i \tag{31}$$

with the x_i randomly drawn from a uniform distribution in the interval $[0, 1]$ and the ε_i are drawn from a (heteroscedastic) normal distribution

$$\varepsilon_i \sim N(0, [100(1-x)]^2). \tag{32}$$

In this example the curvature of the true underlying conditional expectation is increasing with abscissa value and the noise is heteroscedastic with standard deviation decreasing with abscissa value.

Figure 2a shows a scatter plot of such a sample superimposed with both the piecewise linear and piecewise cubic TURBO smooths and the true underlying conditional expectation, $\exp(6x)$. Figure 2b and 2c show the corresponding smoothing spline and SUPER smooths. In this case, the piecewise cubic TURBO estimate gives a slightly better fit than the piecewise linear to the sample (as well as the true underlying curve). The smoothing spline estimate exhibits considerable variability in the high noise region and the SUPER SMOOTHER somewhat less.

In order to study expected performance, 100 replications (25 observations each) were generated according to (31), (32), and fit with the three smoothing methods: piecewise cubic TURBO model, smoothing splines, and SUPER SMOOTHER. Figure 2d plots their average absolute error, $|f(x) - \exp(6x)|$, as a function of abscissa value, x. In the high noise region $x < 0.2$ both the smoothing spline (dotted line) and SUPER SMOOTHER (dashed line) exhibit large error associated with the high variance of their estimates. In the intermediate region $0.2 < x < 0.9$ both the TURBO (solid line) and SUPER smoothers have comparable performance. In the low noise high curvature extreme, $x > 0.9$, all three methods produce considerable increased error (bias) with the SUPER SMOOTHER degrading the least. Over most of the region the (nonadaptable) smoothing spline method gives relatively poor performance. This might be expected since both the curvature and noise level are varying, thereby causing a single span value to be less appropriate.

5.3 A difficult smoothing problem

Our final smoothing example is intended to emulate the motorcycle impact data in Silverman (1985, Fig. 6). A random sample of 50 (x_i, y_i) pairs were generated with the x_i from a uniform distribution in the interval $[-0.2, 1.0]$ and the y_i given by

$$y_i = \begin{cases} \varepsilon_i & x_i \leq 0 \\ \sin[2\pi(1 - x_i)^2] + \varepsilon_i & 0 < x_i \leq 1 \end{cases}$$

with the ε_i randomly generated from

$$\varepsilon_i \sim N[0, \max{}^2(0.05, x_i)].$$

The second derivative of the underlying conditional expectation changes sign four times and is infinite at $x = 0$. The standard deviation of the additive noise is small and constant for $X \leq 0.05$, and then increases linearly with x. Figure 3a shows a scatter plot of such a sample. Figure 3b superimposes the piecewise linear and cubic TURBO smooths along with the true underlying conditional expectation. Figures 3c and 3d show respectively the corresponding smoothing spline and SUPER SMOOTHER smooths. All but the piecewise linear estimate have a downward bias at the derivative discontinuity. Both TURBO smooths have a downward bias at the minimum, whereas the smoothing spline and SUPER smooths have an upward bias. The smoothing spline estimate exhibits considerably more variation in the higher noise regions. The piecewise cubic TURBO smooth again gives a slightly better fit to the data than does the piecewise linear.

As in the previous examples, we compare expected performance of the three methods over 100 replications of 50 observations each. Figure 3e shows the average absolute error (from the true underlying conditional expectation) for the piecewise cubic TURBO smooths, smoothing splines, and SUPER SMOOTHER. In the higher noise regions ($X > 0.25$) the TURBO and SUPER smoothers are seen to have comparable error, but in the lower noise high curvature region ($x < 0.25$) the SUPER SMOOTHER exhibits about 20% higher accuracy. It has considerably less bias at the derivative discontinuity and the minimum points. Smoothing splines exhibit relatively poorer performance over almost the entire interval. Again, this might have been expected since this is a highly heteroscedastic situation with varying curvature. Nonadaptable smoothers must choose a compromise smoothing parameter for the entire region, whereas the adaptable procedures can adjust the span to try to account for such effects.

5.4 Additive modeling with pure noise.

Since it is as important for a method to *not* find predictive structure when it is absent, as it is to find it when present, we first study the performance of our additive modeling procedure when there is no predictive relationship between the response and predictors. Two simulation experiments were performed. In the first, 100 replications of a sample of size $N = 50$ were generated. The responses were drawn from a standard normal distribution. There were $p = 10$ predictor variables each independently drawn from a uniform distribution in the interval [0, 1]. The TURBO modeling procedure was applied to each of these 100 replicated samples. In 67 replications no knots were placed on any of the ten predictors. The estimated response function was taken as the sample response mean. In 24 replications one knot was placed and in 9 cases two knots were used. Thus, two thirds of the time the TURBO model reported no predictive relationship. In the rest of the cases it reported a small one. Table 1 summarizes the distribution of both the sample multiple correlation (R^2) between the response and the estimated model, and the root mean squared distance $(\mathrm{ESE})^{1/2}$ of the estimated model from the truth, $f(x_1 \cdots x_{10}) = 0$.

For comparison we also applied to these data sets the projection selection procedure (Friedman and Stuetzle, 1981), or equivalently, the ACE procedure with the response transformation restricted to be linear (Breiman and Friedman, 1985), using the SUPER SMOOTHER (Friedman, 1984). The corresponding distribution of R^2 and $(\mathrm{ESE})^{1/2}$ are also summarized in Table 1. In contrast to the TURBO model, this method is seen to seriously overfit the data as reflected in the high values of both quantities. The propensity of ACE (based on the SUPER SMOOTHER) to overfit in low signal to noise situations was discussed by Folkes and Kettenring (1985), and Breiman and Friedman (1985).

A second simulation experiment was performed, using the same setting but increasing the sample size of each replication to $N = 100$. The TURBO model placed no knots 63 times. The frequency of one through five knots were, respectively 26, 6, 3, 1, 1. The corresponding distributions for both methods are shown in Table 1. The increased sample size is seen to improve the performance of both methods but the qualitative aspects of their comparison are the same as with the smaller ($N = 50$) sample size. The TURBO modeling procedure is seen to be fairly conservative. It should be noted that the tendency of the ACE method to drastically overfit in low signal to noise small sample settings is not a fundamental property, but is mainly a consequence of its implementation using the highly flexible SUPER SMOOTHER.

5.5 A highly structured additive model

This example is intended to contrast with the previous one. As in the previous example there are $p = 10$ predictor variables each independently generated from a uniform distribution on [0, 1]. Two simulation experiments of 100 replications each were performed with $N = 50$ and $N = 100$. The response variables were generated by

$$y_i = f^*(x_{1i} \cdots x_{10i}) + \varepsilon_i$$

with the ε_i independently drawn from a standard normal distribution. The function f^* was taken to be

$$f^*(X_1 \cdots X_{10}) = 0.1e^{4X_1} + \frac{4}{1 + e^{-(X_2-0.5)/0.05}} + 3X_3 + 2X_4 + X_5.$$

In this case the signal to noise ratio (standard deviation of f^*) is 2.47. The true underlying conditional expectation is additive in the ten predictor variables. The relationship is highly nonlinear in the first two, linear with decreasing strength in the next three, and constant (zero) in the last five.

Figures 4a - 4e show the piecewise linear and cubic curve estimates (24), (25) for the first five variables in the first replication of $N = 50$. Also, superimposed on the figures is the true underlying function for the corresponding variable (solid line), and with the errors ε_i added to it (dots). As can be seen the TURBO model placed one knot on X_1, two on X_2, and one each on variables X_3, X_4, and X_5. No knots were placed on the last five predictor variables. Both the piecewise linear and cubic models fit the data with R^2 values of 0.93. The root mean-squared error of the piecewise linear model from the true $f^*(X_1 \cdots X_{10})$ was 0.45, whereas for the corresponding piecewise cubic it was 0.47.

More important than performance on a single sample is average performance over 100 independent replications of this situation. Table 2 summarizes the results for piecewise cubic fitting. The results shown in Fig. 4 (based on the first replication of the 100) are seen to be somewhat more favorable than those on the average. A second simulation experiment with 100 replications of $N = 100$ observations each was also performed. These results are summarized in Table 2 as well. The ACE/SUPER SMOOTHER procedure was applied to the same sets of replicated data with the results also shown in Table 2.

Comparing the results, the TURBO modeling procedure is seen to exhibit substantially better performance in terms of root mean squared error. The effect is, however, less dramatic than in the pure noise case. On average, ACE/SUPER SMOOTHER fits the data sample 3.7 times more closely than the TURBO model for $N = 50$. For $N = 100$ this factor is 1.8. This overfitting results in an increased median modeling error of 16% for $N = 50$ and 30% for $N = 100$. On the other hand, the TURBO model has a tendency to be conservative and under fit the data, producing estimates that are sometimes overly smooth (too few knots). This has an interpretational advantage and a predictive advantage when the curvature variation of the true underlying conditional expectation is reasonably gentle. This example, however, simulates a situation in which that variation is fairly dramatic and the advantage of TURBO modeling procedure (in terms of expected squared error) is thereby somewhat reduced.

5.6 Molecular quantitative structure - activity relationship.

We illustrate here TURBO modeling on a data set from organic chemistry (Wright and Gambino, 1984). The observations are 36 compounds that were collected to examine the structure activity relationship of 6-anilinouracils as inhibitors of Bacillus subtilis DNA polymeraze III. The four structural variables measured on each compound are summarized in Table 3. The response variable is the logarithm of the inverse concentration of 6-anilinouracil required to achieve 50% inhibition of enzyme activity.

TURBO modeling applied to these data placed four knots: one on the first variable, two on the second, and one on the third. The $e^2 = 1 - R^2$ for the piecewise linear fit was 0.12, while for the piecewise cubic it was 0.11. The corresponding 632-bootstrap estimates (30) were 0.23 and 0.22. Figures 5a-5d show the piecewise cubic curve estimates $\hat{f}_i(x_i), i = 1,4$, along with the bootstrap confidence intervals (29). The data points (dots) on the figures are the scaled residuals from the fit added to the curve at each abscissa value (component plus residual plot). The scale factor is the square root of the ratio of the 632 bootstrap estimate to the resubstitution e^2. The curve estimates on the first three predictors are all seen to be fairly nonlinear, especially the second one.

ACE/super smoother was also applied to these data. The resubstitution e^2 was 0.054 while the 632-bootstrap estimate was 0.29. As in the simulated data example (Section 4.5), ACE/Super smoother is seen to fit the data more closely than the TURBO model, but the resulting overfit results in inferior future prediction error in this case.

5.7 Air pollution data.

This data set consists of daily measurements of ozone concentration and eight meteorological variables for 330 days of 1976 in the Los Angeles basin. Table 4 describes the variables. These data were introduced by Breiman and Friedman (1985) to illustrate the ACE procedure. They were also analyzed by Hastie and Tibshirani (1984) using their Generalized Additive modeling method (see also Hastie and Tibshirani, 1986). In contrast to previous examples this is a large (N=330), complex, and not very noisy data set. One might therefore expect that the simple TURBO modeling procedure would be at a disadvantage when compared to the more sophisticated approaches that have been applied to these data.

Applying the TURBO model resulted in ten knots being placed: one each on variables 1, 4, 5, and 6, and two each on variables 3, 8, and 9. The resulting resubstitution e^2 was 0.20 for both the piecewise linear and cubic fits. The corresponding 632-bootstrap estimates (20 replications) were 0.24 for both. The piecewise cubic individual variable curve estimates, $\hat{f}_i(x_i), 1 \leq i \leq 9$, (25) are shown in Figs. 6a-6i, along with their bootstrap confidence intervals (29) and (scaled) residuals.

Exact comparison with the ACE results in Breiman and Friedman (1985) is not possible since they applied ACE in a mode that estimates an optimal (minimum e^2) response transformation as well. The resulting response estimate was, however, not too far from the identity function so that a rough comparison is possible. They applied a variable based forward stepwise procedure, selecting five variables. Their resubstitution e^2 for the optimal response function was 0.18. The variables that were selected and the corresponding curves are fairly consistent with (but not identical to) the TURBO model results. Generally, the TURBO curves are a bit simpler than the corresponding ACE/SUPER smoother estimates. Since bootstrapping or cross-validating the forward stepwise ACE procedure would be prohibitively expensive, no estimate of (honest) future prediction error could be given.

Hastie and Tibshirani (1984) also analysed these data. Their Generalized Additive Modeling procedure as applied in this setting is equivalent to the ACE method with the response function constrained to linearity. Therefore we can make direct comparison with their results. Hastie and Tibshirani did not employ SUPER SMOOTHER, but rather a nonadaptable local linear smoother with constant span. With all nine predictors in the regression function they obtained an e^2 of 0.20. With the same subset of variables as used by Breiman and Friedman (1985) the e^2 was 0.22. Hastie and Tibshirani (1986) provide a method of estimating the equivalent degrees-of-freedom used by their fitting process. This estimate accounts for the flexibility associated with the resulting smooths but does not account for the (nonlinear) span selection and variable subset selection process. They report 21.8 degrees-of-freedom for their fit with all variables and 12.4 for the five variable subset. The corresponding degree-of-freedom count for the TURBO fit would be 11 (constant term plus coefficients for ten knots).

6.0 Discussion

The examples of Section 5 indicate that the smoothing method outlined in Section 2, and the corresponding additive modeling procedure described in Section 3, are competitive with the techniques to which they were compared. They seem to have substantial advantage in situations with low sample size and high noise, where the underlying functions are fairly simple. In this context a simple function is one that can be reasonably well approximated by a piecewise linear function with few (judiciously placed) knots. This was the case in the examples of Sections 5.1, 5.2, 5.4, 5.5, and 5.6. Our procedures appeared to have similar performance to the corresponding competitors in large sample low noise situations, again with fairly simple underlying functions (Section 5.7).

The example in Section 5.3 represented a moderate sample size situation with both high and low noise regions (strong heteroscedasticity) and a complex underlying function. In this particular case SUPER SMOOTHER appeared to perform somewhat but not dramatically better.

FORTRAN programs implementing the procedures herein described are available from the authors.

Acknowledgment

We thank Ani Adhikari and Leo Breiman for bringing to our attention the motivating work of Smith (1982).

Table 1

Comparison of TURBO and ACE additive modeling of pure noise (Section 5.4). The 5, 50, and 95 percent points are given for the distribution of the multiple correlation R^2 (resubstitution), and the root expected squared error $(\mathrm{ESE})^{1/2}$.

	R^2			$(\mathrm{ESE})^{1/2}$		
	.05	.5	.95	.05	.5	.95
			$N = 50$			
TURBO	0.0	0.0	0.21	0.02	0.18	0.50
ACE	0.74	0.91	0.97	0.68	0.85	1.00
			$N = 100$			
TURBO	0.0	0.0	0.12	0.008	0.12	0.41
ACE	0.49	0.70	0.86	0.55	0.69	0.89

Table 2

Comparison of TURBO and ACE additive modeling in a higher signal to noise situation (Section 5.5). The 5, 50, and 95 percent points are given for the distribution of the multiple correlation R^2 (resubstitution), and the root expected squared error $(\mathrm{ESE})^{1/2}$.

	R^2			$(\mathrm{ESE})^{1/2}$		
	.05	.5	.95	.05	.5	.95
			$N = 50$			
TURBO	0.79	0.86	0.93	0.34	0.75	0.99
ACE	0.97	0.99	1.0	0.68	0.87	1.00
			$N = 100$			
TURBO	0.84	0.87	0.91	0.31	0.48	0.62
ACE	0.93	0.96	0.99	0.60	0.72	0.85

Table 3

Variables associated with molecular quantitative structure-activity data example (Section 5.6).

X_1 - meta substituent hydrophobic constant
X_2 - para substituent hydrophobic constant
X_3 - group size of substituent in meta position
X_4 - group size of substituent in para position
Y - logarithm of the inverse concentrations of 6-anilinouracil required to achieve 50% inhibition of the enzyme.

Table 4

Variables associated with the air pollution data example (Section 5.7).

X_1 - Vandenburg 500 millibar height
X_2 - humidity
X_3 - inversion base temperature
X_4 - Sandburg Air Force Base temperature
X_5 - inversion base height
X_6 - Daggot pressure gradient
X_7 - wind speed
X_8 - visibility
X_9 - day of the year
Y - Upland ozone concentration

References

Agarwal, G.G. and Studden, W.J. (1980). Asymptotic integrated mean square error using least squares and minimizing splines. *Ann. Statist.* **8**, 1307-1325.

Akaike, H. (1970). Statistical predictor identification. *Ann. Statist.* **22**, 203-217.

Belsey, D.A., Kuh, E., and Welsch, R.E. (1980). Regression Diagnostics. New York: John Wiley.

Bradley, J.V. (1968). Distribution-Free Statistical Tests. Englewood Cliffs, N.J.: Prentice-Hall.

Breiman, L. and Freedman, D. (1983). How many variables should be entered in a regression equation? *J. Amer. Statist. Assoc.* **78**, 131-136.

Breiman, L., Friedman, J.H., Olshen, R., and Stone, C.J. (1984). Classification and Regression Trees. Belmont, CA: Wadsworth.

Breiman, L. and Friedman, J.H. (1985). Estimating optimal transformations for multiple regression and correlation (with discussion). *J. Amer. Statist. Assoc.* **80**, 580-619.

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74**, 828-836.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik* **31**, 317-403.

Crawford, S. (1986). Resampling strategies for recursive partitioning classification with the CART^{TM} algorithm. Ph.D. Dissertation, Department of Education, Stanford University.

de Boor, C. (1978). A Practical Guide to Splines. New York: Springer-Verlag.

Efron, B. (1983). Estimating the error rate of a prediction rule: improvement on cross-validation. *J. Amer. Statist., Assoc.* **78**, 316-331.

Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Science* **1**, 54-77.

Feder, P.I. (1975). The log likelihood ratio in segmented regression. *Ann. Statist.* **3**, 84-97.

Folkes, E.B. and Kettenring, J.R. (1985). Discussion on: Breiman, L. and Friedman, J.H., Estimating optimal transformations for multiple regression and correlation. *J. Amer. Statist. Assoc.* **80**, 607- 613.

Friedman, J.H. and Stuetzle, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76**, 817-823.

Friedman, J.H. and Stuetzle, W. (1982). Smoothing of scatterplots. Dept. of Statistics, Stanford University Technical Report ORION 003.

Friedman, J.H. (1984). A variable span smoother. Department of Statistics. Stanford University Technical Report LCS5.

Golub, G.H. and Van Loan, C.F. (1983). Matrix Computations. The Johns Hopkins University Press, Baltimore.

Gong, G. (1982). Cross-validation, the jackknife, and the bootstrap: excess error estimation in forward logistic regression. Ph.D. dissertation, Dept. of Statistics, Stanford University, Technical Report No. 80.

Harrison, D. and Rubinfeld, D.L. (1978). Hedonic housing prices and the demand for clean air. *J. Environ. Econ. Management* **55**, 81-102.

Hastie, T. and Tibshirani, R. (1984). Generalized additive models. Dept. of Statistics, Stanford University, Technical Report LCS2.

Hastie, T. and Tibshirani, R. (1985). Discussion of P. Huber: Projection Pursuit. *Ann Statist.* **13**, 502-508.,

Hastie, T. and Tibshirani, R. (1986). Generalized additive models (with discussion). *Statist. Science* 1 297-318.

Hinkley, D.V. (1969). Inference about the intersection in two-phase regression. *Biometrika* **56**, 495-504.

Hinkley, D.V. (1970). Inference in two-phase regression. *J. Amer. Statist. Assoc.* **66**, 736-743.

Mallows, C.L. (1973). Some comments on Cp. *Technometrics* **15**, 661-675.

Reinsch, C.H. (1967). Smoothing by spline functions. *Numer. Math* **10**, 177-183.

Rice, J. and Rosenblatt, M. (1983). Smoothing splines: regression, derivatives and deconvolution. *Ann. Statist.* **11**, 141-156.

Rosenblatt, M. (1971). Curve estimation. *Ann. Math. Statist.* **42**, 1815-1842.

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Amer. Statistician* **8**, 147-164.

Silverman, B.W. (1984). Spline smoothing: the equivalent variable kernel method. *Ann. Statist.*, **12**, 898-916.

Silverman, B.W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J.R. Statist. Soc. B* **47**, 1-52.

Smith, P.L. (1982). Curve fitting and modeling with splines using statistical variable selection techniques. NASA, Langley Research Center, Hampton, VA, Report NASA 166034.

Stone, C.J. and Koo, Cha-Yong (1985). Additive splines in statistics. Proceedings, Annual meeting of Amer. Statist. Assoc., Statist. Comput. Section, August.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictors (with discussion). *J.R. Statist. Soc. B* **36**, 111-147.

Wright, G.E. and Gambino, J.J. (1984). Quantitative structure- activity relationships of 6-anilinouracils as inhibitors of bacillus subtilis DNA Polymerase III. *J. Med. Chem.* **27**, 181-185.

FIGURE CAPTIONS

1. Smoothing a small sample ($N = 20$) of pure noise.
   - a) TURBO smooth
   - b) Smoothing spline
   - c) SUPER SMOOTHER
   - d) Average absolute error as a function of abscissa value (TURBO smooth : solid, smoothing spline : dots, SUPER smooth : dashed)
   - e) Average absolute error for a larger ($N = 40$) sample.

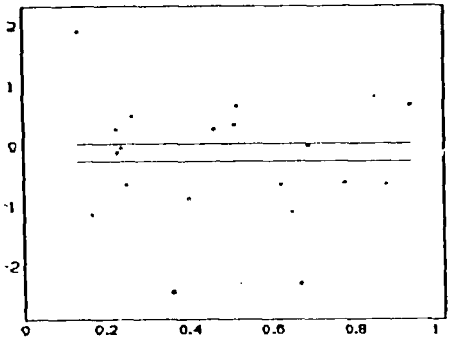
2. Smoothing a *monotonic function* with heteroscedastic noise
   - a) TURBO smooth
   - b) Smoothing spline
   - c) SUPER SMOOTHER
   - d) Average absolute error as a function of abscissa value (TURBO smooth : solid, smoothing spline : dots, SUPER SMOOTHER : dashed)

3. Difficult smoothing problem
   - a) data scatter plot
   - b) TURBO smoother
   - c) smoothing spline
   - d) SUPER SMOOTHER
   - e) Average absolute error as a function of abscissa value (TURBO smooth : solid, smoothing spline : dots, SUPER SMOOTHER : dashed)

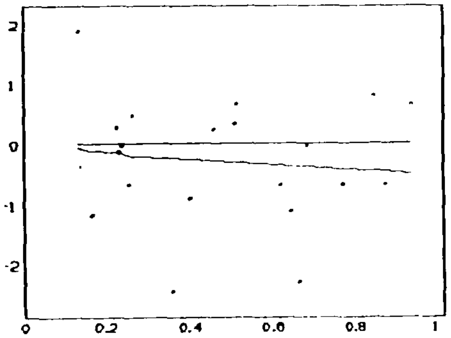
4. Solution predictor variable curves for the simulated additive modeling example.
   - a) $f_1(X_1)$ c) $f_3(X_3)$ e) $f_5(X_5)$
   - b) $f_2(X_2)$ d) $f_4(X_4)$

5. Solution predictor variable curves for the *quantitative structure activity* relationship (see Table 3)
   - a) $f_1(X_1)$ c) $f_3(X_3)$
   - b) $f_2(X_2)$ d) $f_4(X_4)$

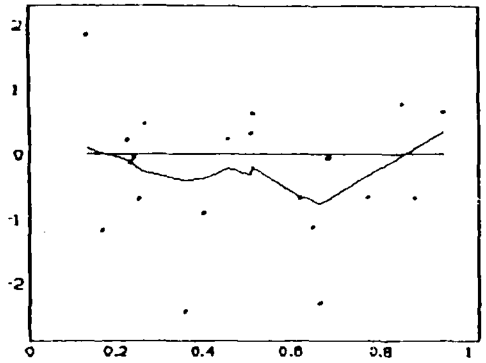
6. *Solution* predictor variable curves for the air pollution data (see Table 4)



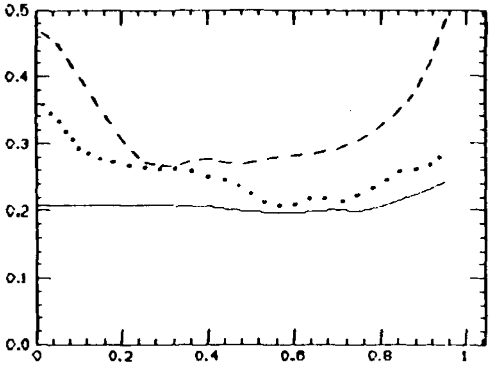
1a



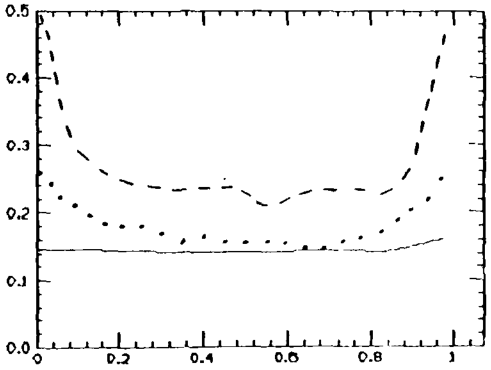
1b


2
1
0
-1
-2
0
0.2
0.4
0.6
0.8
1

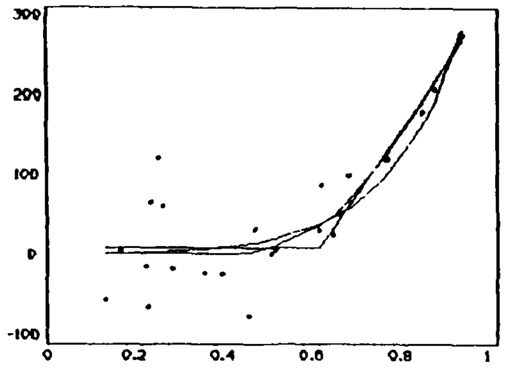
1c


0.5
0.4
0.3
0.2
0.1
0.0
0
0.2
0.4
0.6
0.8
1

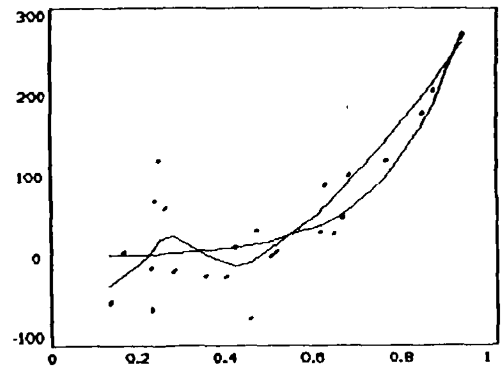
1d


0.5
0.4
0.3
0.2
0.1
0.0
0
0.2
0.4
0.6
0.8
1

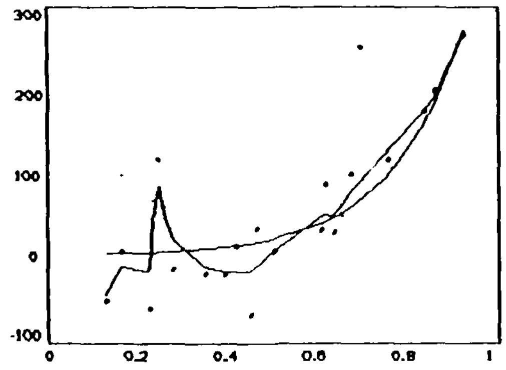
1e


300
200
100
0
-100
0
0.2
0.4
0.6
0.8
1

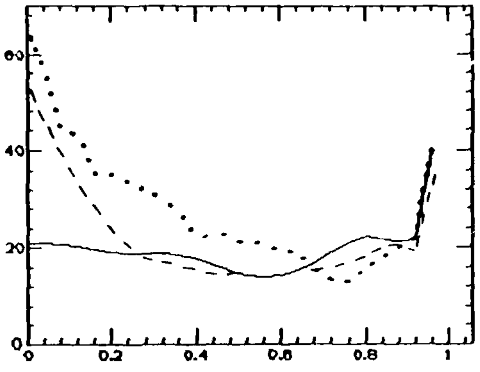
2a


300
200
100
0
-100
0
0.2
0.4
0.6
0.8
1

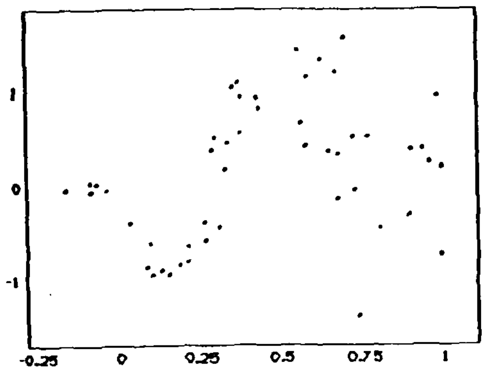
2c


300
200
100
0
-100
0
0.2
0.4
0.6
0.8
1

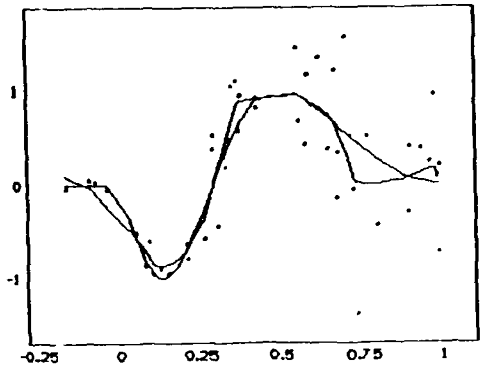
2b


60
40
20
0
0
0.2
0.4
0.6
0.8
1

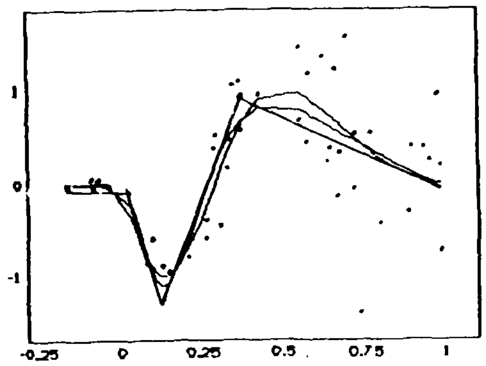
2d


1
0
-1
-0.25
0
0.25
0.5
0.75
1

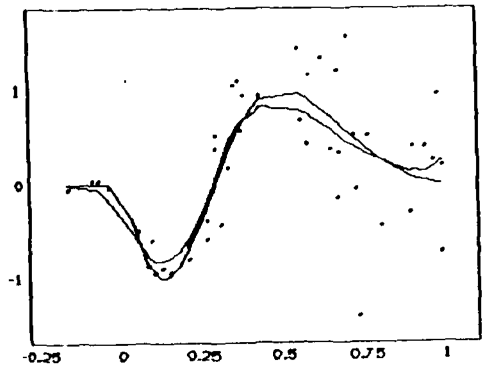
3a


1
0
-1
-0.25
0
0.25
0.5
0.75
1

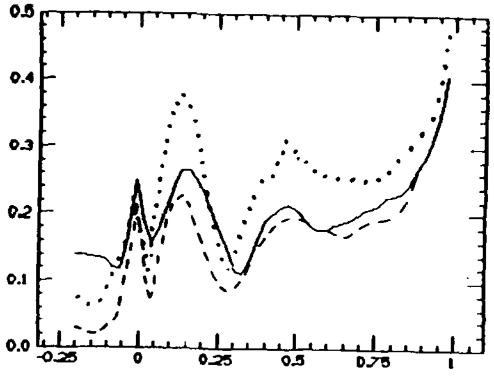
3c


1
0
-1
-0.25
0
0.25
0.5
0.75
1

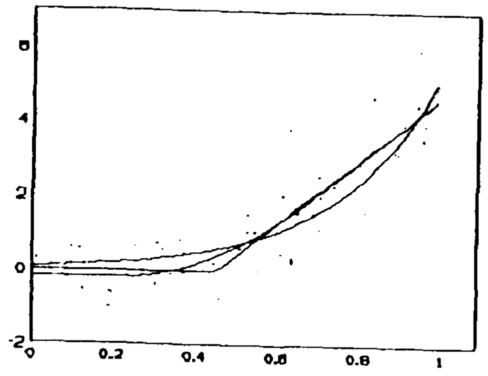
3b


1
0
-1
-0.25
0
0.25
0.5
0.75
1

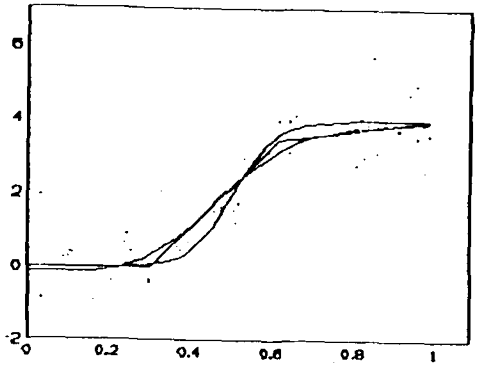
3d


0.5
0.4
0.3
0.2
0.1
0.0
-0.25
0
0.25
0.5
0.75
1

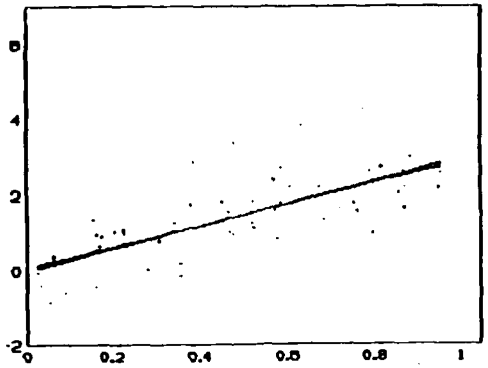
3e


6
4
2
0
-2
0
0.2
0.4
0.6
0.8
1

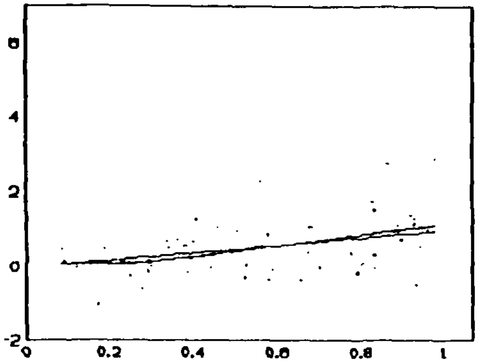
4a


6
4
2
0
-2
0
0.2
0.4
0.6
0.8
1

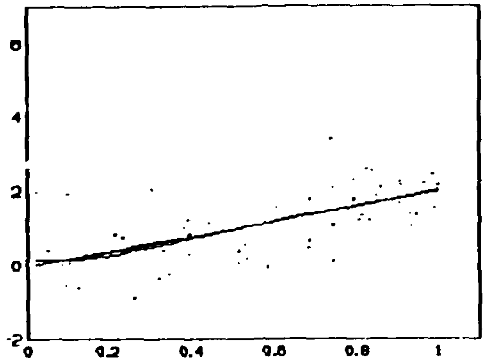
4b



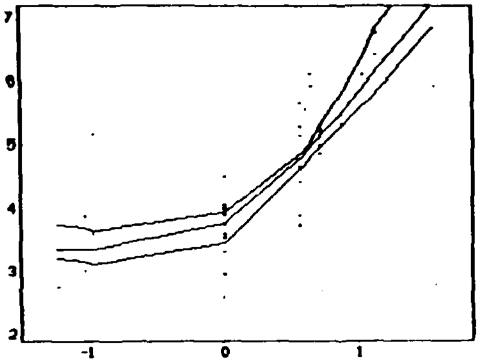
4c



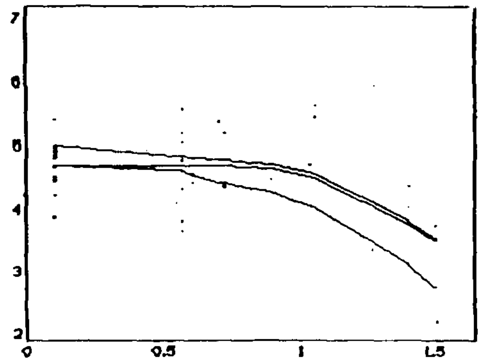
4e



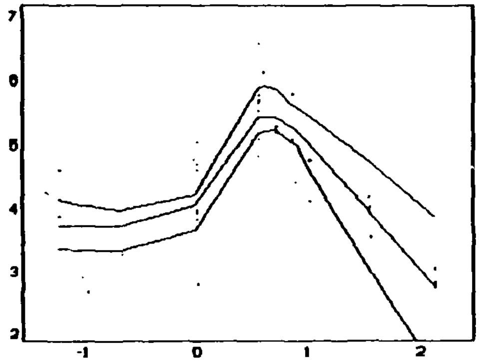
4d



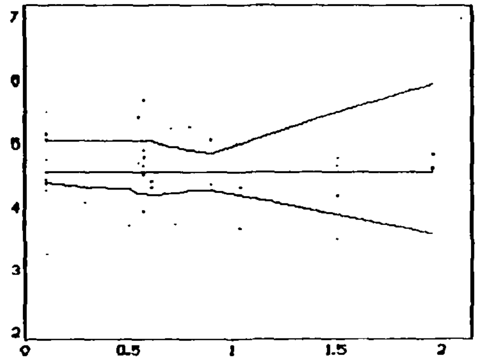
5a



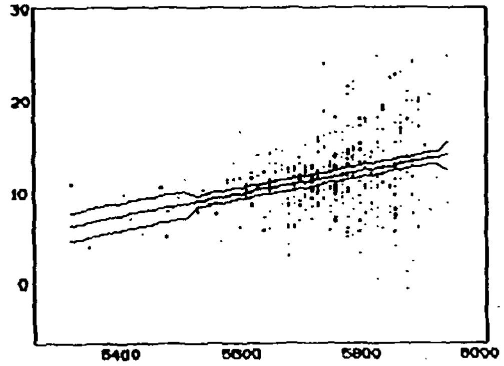
5c



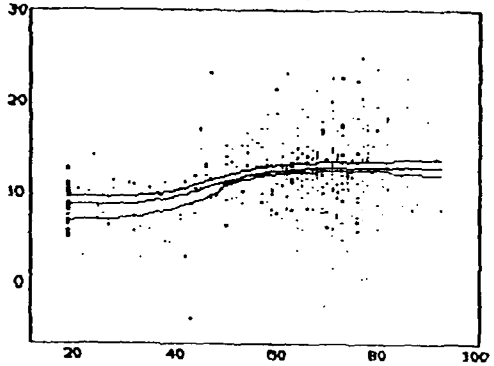
5b



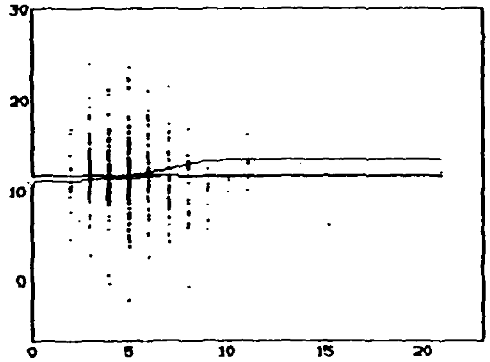
5d


30
20
10
0
5400
5600
5800
6000

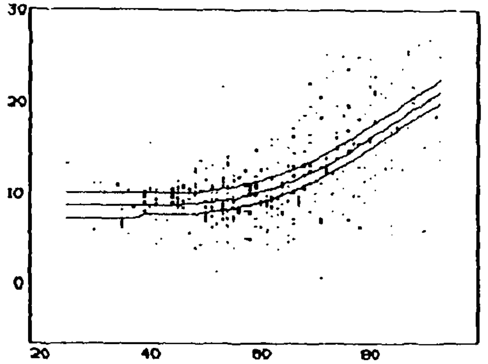
6a


30
20
10
0
20
40
60
80
100

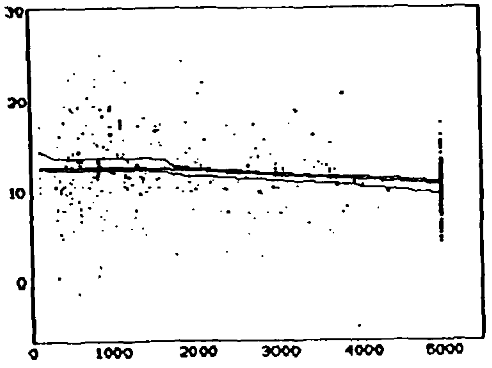
6c


30
20
10
0
0
5
10
15
20

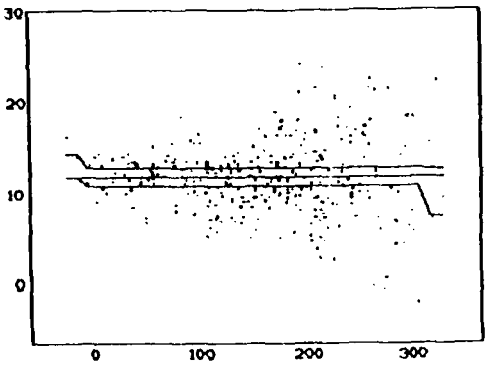
6b


30
20
10
0
20
40
60
80

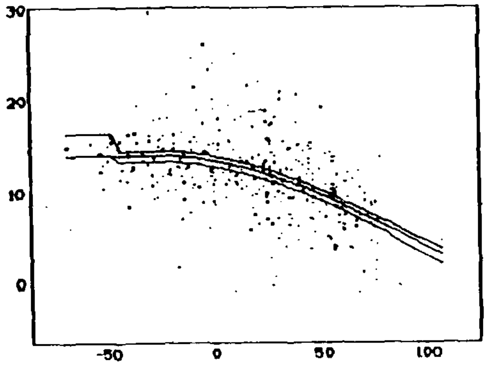
6d


30
20
10
0
0
1000
2000
3000
4000
5000

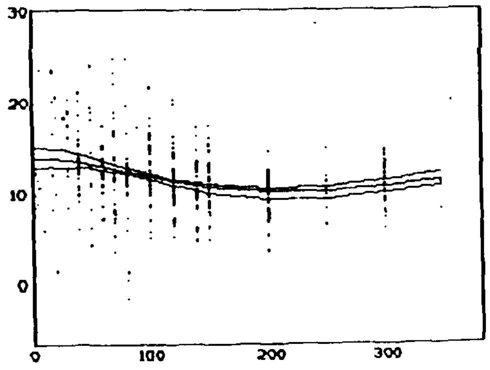
6e


30
20
10
0
0
100
200
300

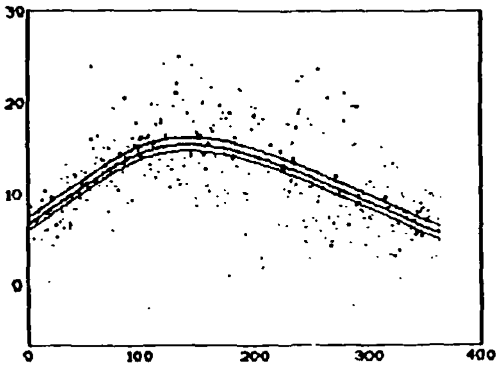
6g


30
20
10
0
-50
0
50
100

6f


30
20
10
0
0
100
200
300

6h



61

Appendix 6

A smoothed EM algorithm for the solution of Wicksell's corpuscle problem.

by

J. D. Wilson

[illegible] Vol. [illegible] pp. [illegible]
Reprints available directly from the publisher
Photocopying permitted by license only

A SMOOTHED EM ALGORITHM FOR THE *SOLUTION OF WICKSELL'S CORPUSCLE* PROBLEM

J. D. WILSON

Department of Mathematical Sciences, University of Bath, Claverton Down
Bath BA2 7AY, UK

(Received 20 April 1988)

In this paper a new method called the EMS algorithm is used to solve Wicksell's corpuscle problem, that is the determination of the distribution of the sphere radii in a medium given the radii of their profiles in a random slice. The EMS algorithm combines the EM algorithm, a procedure for obtaining maximum likelihood estimates of parameters from incomplete data, with simple smoothing. The method is tested on simulated data from three different sphere radii densities, namely a bimodal mixture of Normals, a Weibull and a Normal. The effect of varying the level of smoothing, the number of classes in which the data is binned and the number of classes for which the estimated density is evaluated is investigated. Comparisons are made between these results and those obtained by others in this field.

KEY WORDS: Wicksell's corpuscle problem, EM algorithm, smoothing, simulation, stereology

1 INTRODUCTION

The problem to be discussed in this paper is the well-known "corpuscle problem", first considered by Wicksell (1925). In this problem we have a three-dimensional specimen consisting of opaque spheres in a translucent medium. It is assumed that the centres of the spheres are distributed according to a spatial Poisson process with constant, though unknown intensity, λ and we make the standard assumption that λ is sufficiently small to ensure that the possibility of overlapping spheres can be ignored. A section is taken across the medium and examined under a microscope. A number of circular profiles will be observed which correspond to those spheres cut by the section. Our aim is to determine the density of the radii of the spheres in the medium from this sample of profile radii. Wicksell (1925) demonstrated that the density of sphere radii in the medium, $f(r), 0 < r \leqq R$ and the density of profile radii in a slice, $g(y), 0 < y \leqq R$ are related by the equation

$$g(y) = \frac{1}{\mu}\int_{y}^{R} \frac{y}{\sqrt{r^2 - y^2}}\, f(r)\,dr \qquad 0 < y \leqq R \tag{1.1}$$

where R is an upper bound on the sphere radii (which will be determined by practical considerations) and μ is the mean of the sphere radii distribution given by

$$\mu = \int_0^\infty r f(r) dr$$

This is the simplest case of the corpuscle problem and does not allow for the effects of truncation and overprojection which have more recently been incorporated into the framework of the problem. Both truncation and overprojection occur at the observation of profiles stage and may as a result give a distorted picture of the data. For example, some profiles will be too small to be observed, even under a microscope and thus information will be lost about small profiles and hence small spheres. This is the truncation effect. Overprojection arises because the section we take through the medium has a finite thickness. As the *spheres are* opaque the profile diameter will correspond to the largest cross section of that part of the sphere in the slice. Both these effects and the effect of "capping" (another form of truncation) are considered in more detail in Cruz Orive (1983). It suffices to say that their inclusion considerably complicates the problem. More recent papers have made allowance for all or at least some of these effects in an attempt to produce a more accurate estimate of the sphere density. Cruz Orive (1983) provides a comprehensive review of research in this area. A distinction is drawn between parametric and nonparametric (or distribution *free) methods with* the emphasis on a distribution-free numerical approach, the method of finite differences. Cruz Orive (1983) recommends this method as it satisfies four important criteria. It is simple compute, makes no parametric assumptions about the shape of the sphere density, may be adapted to incorporate the effects of truncation and overprojection, and will accept grouped profile data. However, the disadvantage of the method of finite differences, as proposed by Cruz Orive (1983), is its numerical instability. Furthermore, since positivity constraints are not included in the procedure, negative values of the *sphere density may arise*, an outcome which has no meaning in probability terms.

A more recent paper by Blodner, Muhlig and Nagel (1984) compares by a simulation study the method of Cruz Orive (1983) with other nonparametric methods. The results of this study will be considered in more detail in Section 3.3. One further paper that will be referred to in later sections, is Nychka, Wahba, Goldfarb and Pugh (1984). In that paper cross validated spline methods are used to estimate the size distribution of tumours in mouse livers from two-dimensional cross sections. The method takes account of truncation effects *and in simulations* produces smooth density estimates which are, in general, good reproductions of the original densities. However, the method involves considerable computation and does not include positivity constraints. Thus negative estimates of the density may arise and it is not immediately obvious what to do under such circumstances.

The method recommended in this paper for the solution of the corpuscle problem is a new procedure first appearing in Jones, Silverman and Wilson (1987) and called the EMS algorithm. It is an iterative procedure which combines the

Expectation Maximum Likelihood Steps of the EM algorithm (see Dempster, Laird and Rubin 1977) with simple smoothing. The EM algorithm is used in a similar context in a paper by Vardi Shepp and Kaufman (1985) on the subject of Positron Emission Tomography. There the aim was to determine the metabolic activity of different sections of the brain by introducing a radioactive tracer into the area of interest and measuring the number of positrons emitted, by an array of detectors arranged *around the body. The use of the EMS procedure for this* tomography problem is developed in Jones, Silverman and Wilson (1987).

In this paper the effectiveness of the EMS algorithm in recovering sphere radii densities with both speed and accuracy will be demonstrated by means of a simulation study. The EMS algorithm has many of the advantages of the finite differences method. It is simple to understand and compute, will handle grouped data, assumes nothing about the sphere density other than its smoothness and incorporates a lower resolution threshold to account for the effect of truncation. Moreover, the procedure, by its *definition, produces only non-negative estimates* of the density function. The EMS algorithm is described in full in Section 2. Section 3 includes a description of the method of simulation, together with the results of the simulation study and a section comparing the EMS algorithm with other approaches. Concluding remarks appear in Section 4.

2 THE EMS ALGORITHM

The EMS algorithm is a new algorithm which combines an iterative procedure called the EM algorithm with a smoothing procedure. The EM algorithm, [as described in detail in Dempster, Laird and Rubin (1977)] is an iterative procedure used to compute maximum likelihood estimates of parameters when the observed data is "incomplete". The data is "incomplete" in the sense that it is not the data we would have liked. For example, in this problem the data are only indirect observations on the density to be estimated. The term "incomplete", moreover, implies the existence of an underlying set of "complete" data. Given a set of incomplete data *Y*, the EM algorithm aims to find that set of parameters ϕ which maximises the log likelihood of observing *Y*. To do so, it makes essential use of the log likelihood of the underlying complete data. The EM algorithm will be explained more fully in the context of the current problem, but first it is necessary to explain why the addition of a smoothing step is so important. The EM algorithm is designed to compute maximum likelihood estimates of the parameters. In this case, our parameters are proportional to the values of a density function at an equally spaced grid of points. The maximum likelihood estimates produced by the EM algorithm are not satisfactory as they lead to jagged density estimates. In addition, the EM algorithm converges very slowly and only to a unique solution under certain quite specialised conditions. On the other hand, the EMS algorithm produces more appealing smooth density estimates in a much smaller number of iterations and simulation experience suggests that the estimated density is a unique solution to the problem for that given level of smoothing.

We begin by establishing a notational framework for the Wicksell problem and describing the application of the EM algorithm in the stereology context. Much of this is very similar to the theory developed in Vardi *et al* (1985) for the tomography problem. We shall then extend the theory to include smoothing and hence define the EMS algorithm.

First recall that the centres of the spheres in the medium are distributed according to a spatial Poisson process with an unknown intensity λ and that the radius r of a sphere is a random variable with unkown density $f(r), 0 < r \leqq R$. The circular profiles observed in a slice taken across the medium are counted into N bins according to the size of their radii $y, \varepsilon \leq y \leq R$. The value ε corresponds to a lower threshold on the size of profile radii, below which profiles are too small to be observed. The data therefore consists of a histogram of profile radii "counts", which in effect correspond to a sample taken from the density $g(y)$ (of profile radii in a random slice) truncated at $y = \varepsilon$. We denote by $n_Y(i)$ the number of discs whose radii fall in bin $i, i = 1, \ldots, N$. *Our aim is to estimate, on the basis of these N* data values, $n_Y(1), \ldots, n_Y(N)$, the density function $f(r)$ on the interval $0 < r \leqq R$. However, since we have no information about profiles with radii $y < \varepsilon$, we are *unable to deduce anything about the distribution of spheres with radii* $r < \varepsilon$ and hence when recovering the sphere density f we assume that $f(r) = 0$ for all r not in the interval $\varepsilon \leqq r \leqq R$. Since, moreover, we only have a finite number of data values N from which to determine the continuous function $f(r), \varepsilon \leq r \leqq R$, we shall discretise f over the interval $[\varepsilon, R]$ by dividing $[\varepsilon, R]$ into M equal intervals (or bins) of length h_j and by assuming that f takes a constant value on each interval.

We then define the value of the sphere intensity at a given sphere radius r, namely $\lambda f(r)$, to be

$$\lambda f(r) = \sum_{j=1}^{M} c_j \chi[r \in j\text{th interval}] \qquad \varepsilon \leq r \leq R \tag{2.1}$$

where c_j is the constant value of the sphere intensity λf on the jth interval and $\chi[\;]$ is an indicator function. The constant c_j is defined to be the expected number of spheres whose radii lie in the jth interval $\phi(j)$ divided by the width of the interval. We write

$$c_j = \frac{\phi(j)}{h_j} \qquad j = 1, \ldots, M$$

Now let $n_R(j)$ denote the number of sphere radii falling in interval j. From the original Poisson process assumption, we know that the number of spheres in the *medium is a Poisson random variable and hence it follows that* $n_R(1), \ldots, n_R(M)$ are independent Poisson random variable whose unknown means we have called $\phi(1), \ldots, \phi(M)$. If a sphere has radius r in interval j, then we denote by $p(i, j)$ the *conditional probability that it will be recorded as a disc in bin i*. That is

$$p(i, j) = P(\text{disc in bin } i \mid \text{sphere in bin } j) \qquad i = 1, \ldots, N, \; j = 1, \ldots, M$$

Let $N(i, j), i = 1, \ldots, N, j = 1, \ldots, M$, be the number of spheres with radii in interval j that are recorded as discs with radii in bin i. By definition, these $N(i, j)$, $i = 1, \ldots, N, j = 1, \ldots, M$, are independent random variables and since for each $j, N(i, j), i = 1, \ldots, N$, amounts to a thinning of the Poisson random variable $n_R(j)$, it follows that each $N(i, j)$ is a Poisson random variable, whose mean is given by

$$E(N(i, j)) = \phi(j)p(i, j) \qquad i = 1, \ldots, N, \; j = 1, \ldots, M \tag{2.2}$$

The conditional probabilities $p(i, j)$ are theoretically known and may be approximated in practice by considering the stereological aspects of the problem. An outline of the method used in this paper to calculate the the $p(i, j)$'s is included in the appendix.

The measured data $n_Y(i)$, corresponding to the total number of discs recorded in bin i for $i = 1, \ldots, N$, may then be written

$$n_Y(i) = \sum_{j=1}^{M} N(i, j) \tag{2.3}$$

It follows that each $n_Y(i)$ is a Poisson random variable whose mean is given by

$$E(n_Y(i)) = \sum_{j=1}^{M} \phi(j)p(i, j) \qquad i = 1, \ldots, N \tag{2.4}$$

Our aim is to determine the values of the parameters $\phi(j)$ $j = 1, \ldots, M$ and hence to determine the density of the sphere radii $f(r), \varepsilon \leqq r \leqq R$ by means of the relationship in Eq. 2.1. We shall begin by considering the solution to this problem obtained by the use of the standard EM algorithm and then extend this, through the addition of smoothing, to the EMS algorithm.

The EM algorithm aims to find those values of the paramaters $\phi(j)$ $j = 1, \ldots, M$ which maximise the log likelihood *of observing the data* $n_Y(i), i = 1, \ldots, N$. However our data, (which I shall denote $\mathbf{n}_Y$), is incomplete. The complete data will therefore be defined to be the independent Poisson random variables $N(i, j)$, $i = 1, \ldots, N, j = 1, \ldots, M$. The log likelihood function of the complete data, $\mathbf{N}$ given the parameters $\phi(j), j = 1, \ldots, M$ is then given by

$$l(\mathbf{N}|\phi) = \sum_{i,j} \left(-\phi(j)p(i, j) + N(i, j)\log(\phi(j)p(i, j)) - \log N(i, j)! \right)$$

$$= -\sum_{j=1}^{M} \phi(j) \sum_{i=1}^{N} p(i, j) + \sum_{j=1}^{M} \log \phi(j) \sum_{i=1}^{N} N(i, j) + C \tag{2.5}$$

where C is a constant term independent of $\phi(j)$. It is clear from 2.5 that the sufficient statistic $t(\mathbf{N})$ for the complete data is then given by

$$t(\mathbf{N})_j = \sum_{i=1}^{N} N(i,j) \qquad j = 1, \ldots, M \tag{2.6}$$

The EM algorithm is an iterative procedure where each iteration consists of two steps: an **Expectation Step** (or *E* Step) and a **Maximization Step** (or *M* Step). Let us denote the value of the vector parameter ϕ after p iterations of the algorithm by ϕ_p. In the Expectation Step we estimate the sufficient statistic $t(\mathbf{N})$ for the complete data $\mathbf{N}$, given the observed data n_1 and the current value of ϕ, ϕ_p. That is we estimate $t_p = E(t(\mathbf{N})|n_1, \phi_p)$. The jth element of t_p is given by

$$(t_p)_j = E\left(\sum_{i=1}^{N} N(i,j)\Big|n_1,\phi_p\right) = \sum_{i=1}^{N} E\left(N(i,j)\Big|n_1(i) = \sum_{j=1}^{M} N(i,j),\phi_p\right)$$

But if $X_k, k = 1, \ldots, M$ are independent Poisson random variables with means $\alpha_k, k = 1, \ldots, M$, the conditional distribution of X_k given $\sum_{k=1}^{M} X_k = X$ is Bionomial with mean $X\alpha_k/(\sum_{k=1}^{M}\alpha_k)$. Thus

$$(t_p)_j = \sum_{i=1}^{N} \frac{n_1(i)\phi(j)p(i,j)}{\sum_{j'=1}^{M} \phi(j')p(i,j')} \tag{2.7}$$

Now that we have estimates $(t_p)_j$ for the components $t(\mathbf{N})_j$ of the sufficient statistic and hence for the quantities $\{\sum_{i=1}^{N} N(i,j), j = 1, \ldots, M\}$, we are able to substitute for these hitherto unknown values in the expression for the log likelihood function, in 2.5. It is now possible to find the maximum likelihood estimates of the parameters $\phi(j), j = 1, \ldots, M$ by maximising the expression

$$M(\psi) = -\sum_{j=1}^{M} \phi(j) \sum_{i=1}^{N} p(i,j) + \sum_{j=1}^{M} \log\phi(j)(t_p)_j$$

in the usual manner. This is the *M*-Step. In this case the maximum likelihood estimates of the $\phi(j)$ are given by,

$$\hat{\phi}(j) = \frac{(t_p)_j}{\sum_{i=1}^{N} p(i,j)} \qquad j = 1, 2, \ldots, M \tag{2.8}$$

Combining these two steps the EM algorithm can be defined in the following way. If ϕ_p denotes the current value of ϕ after p iterations then ϕ_{p+1} is given by,

$$\phi_{p+1}(j) = \frac{1}{\sum_{i=1}^{N} p(i,j)} \sum_{i=1}^{N} \frac{n_1(i)\phi_p(j)p(i,j)}{\sum_{j'=1}^{M} \phi_p(j')p(i,j')} \qquad j = 1, \ldots, M \tag{2.9}$$

The procedure is repeated until convergence is reached, that is until $\|\phi_{p+1} - \phi_p\| < \delta$ for some sufficiently small δ and some appropriate norm $\|\ \|$.

Observe that ϕ must be non-negative. One of the advantages of this algorithm is that choosing initial values of $\phi_0(j) > 0, j = 1, \ldots, M$ will ensure that the condition of non-negativity of density estimates is never violated. If this was not the case, considerable additional computation would be required to maximise the log likelihood using constrained optimisation procedure.

The properties of convergence of the general EM algorithm are discussed in Wu (1983). For the convergence properties of this particular EM sequence, the results of Vardi *et al.* (1985) in the tomography situation are directly applicable to this problem and may be carried over with only notational amendments. In essence, it is true that this EM sequence will converge to a global maximum of the log likelihood function but that this global maximum need not be unique except under exceptional circumstances. Where the maximum is not unique, the choice of starting value ϕ_0 will determine the final solution. The jagged nature of the maximum likelihood estimate together with the slowness of convergence and the non-uniqueness of the solution, lead us to consider the addition of a Smoothing Step (or *S*-Step) into the algorithm.

The EMS algorithm consists of three steps each iteration: the Expectation Step; the Maximum Likelihood Step, and the Smoothing Step. The Smoothing Step consists of taking a weighted average of each maximum likelihood estimate $\hat{\phi}(j)$ (obtained from Eqs. 2.7 and 2.8) with its neighbours. For k point smoothing (where k is odd), the smoothed value of $\hat{\phi}(j)$, $\phi^S(j)$, is a weighted average of $\hat{\phi}(j)$ and the $(k-1)/2$ maximum likelihood estimates on either side of it. For example, if $k = 7$,

$$\phi^S(j) = a_1\hat{\phi}(j-3) + a_2\hat{\phi}(j-2) + a_3\hat{\phi}(j-1) + a_4\hat{\phi}(j)$$

$$+ a_5\hat{\phi}(j+1) + a_6\hat{\phi}(j+2) + a_7\hat{\phi}(j+3) \tag{2.10}$$

The weights $a_1, \ldots, a_k$ are Binomial in the sense that,

$$a_r = \frac{b_r}{\sum_{s=1}^{k} b_s} \qquad r = 1, \ldots, k \tag{2.11}$$

where b_r corresponds to the coefficient of the rth term in the Binomial expansion of $(1+x)^{k-1}$. These weights sum to one, by definition, are symmetric about the point to be smoothed $\hat{\phi}(j)$ and decrease in magnitude away from $\hat{\phi}(j)$, so that the more immediate neighbours have more influence than those further away. In the example above where $k = 7$, the values of the a_r's are,

$$a_1 = a_7 = 0.015625 \quad a_2 = a_6 = 0.09375 \quad a_3 = a_5 = 0.234375 \quad a_4 = 0.3125$$

It is important to notice that a larger value of k will produce a greater degree of smoothing, as less weight will be given to the value $\hat{\phi}(j)$ and more to its neighbouring values. The effect of making k too large is to "oversmooth" and to lose resolution. This will be illustrated in Section 3.

Consider now the expression in 2.10. It is clear that for j near to 1 or j near to M, some of the terms in 2.10 will be undefined. For example, if $j = 2$ then $\hat{\phi}(j-3)$ and $\hat{\phi}(j-2)$ *will not be defined. There are a number of ways to handle this* problem. One is to assume that all undefined terms have the value zero. This method will have the effect of pulling the ends of the final solution function f down to zero. This will tend to depress any "asymptotic" behaviour that may exist near the origin and in particular, recovery of densities similar to a Weibull will be impaired. The second method, and the one that I shall use, is to give the weights of undefined terms the value zero and then to increase the weight of the term $\hat{\phi}(j)$ accordingly, so that the sum of the weights is still unity. For example, if $k = 7$ and $j = 2$ the weights in Eq. 2.10 are given by,

$$a_1 = a_2 = 0$$

$$a_3 = 0.234375$$

$$a_4 = 0.3125 + 0.09375 + 0.015625 = 0.421875$$

$$a_5 = 0.234375$$

$$a_6 = 0.09375$$

$$a_7 = 0.015625$$

This method does not presuppose anything about the values of the parameter ϕ outside the range of the M bins and is therefore preferable to the first method proposed. Notice that *k point smoothing can be expressed as the multiplication of* a smoothing matrix S_k with the vector of maximum likelihood estimates ϕ obtained from Eqs 2.7 and 2.8. Here S_k is a banded symmetric matrix of size $M \times M$ with at most k non-zero entries in each row. The entries of the matrix correspond to the smoothing coefficients and the row elements (and hence column elements) sum to one. For example, for $k = 3$ the smoothing matrix S_3 is given by,

$$S_3 = \begin{pmatrix} 0.75 & 0.25 & 0 & \cdots & 0 & 0 \\ 0.25 & 0.50 & 0.25 & \cdots & 0 & 0 \\ 0 & 0.25 & 0.50 & \cdots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & 0.50 & 0.25 \\ 0 & 0 & 0 & \cdots & 0.25 & 0.75 \end{pmatrix}$$

The EMS algorithm with k point smoothing can then be defined.

If ϕ_p denotes the current value of ϕ after p iterations then ϕ_{p+1} is given by,

$$\phi_{p+1} = S_k \hat{\phi}_p \tag{2.12}$$

where S_k is the k point smoothing matrix and $\hat{\phi}_p$ is given by,

$$\hat{\phi}_p(j) = \frac{1}{\sum_{i=1}^{N} p(i,j)} \sum_{i=1}^{N} \frac{n_s(i)\phi_p(j)p(i,j)}{\sum_{j'=1}^{M} \phi_p(j')p(i,j')} \qquad j = 1, \ldots, M$$

(as in Eq. 2.9). The EMS sequence is said to have converged when $\|\phi_{p+1} - \phi_p\| < \delta$ for some sufficiently small δ and some appropriate norm. In all the simulations, δ was taken to be 10^{-6} and

$$\|\phi_{p+1} - \phi_p\| = \frac{\sum_{j=1}^{M} (\phi_{p+1}(j) - \phi_p(j))^2}{\sum_{j=1}^{M} \phi_p(j)^2} \tag{2.13}$$

To prove the convergence of the EMS algorithm is by no means trivial and has not yet been achieved to the satisfaction of the author of this paper. However, in all the simulations conducted the sequence has converged and moreover, to a solution which appears to be unaffected by varying the initial value ϕ_0 of the EMS sequence. This would suggest that convergence to a unique solution is guaranteed by this algorithm (providing $k > 1$). The theoretical proof of this assertion must await further investigation.

The value of the vector parameter ϕ, to which the EMS algorithm converges may then be used to determine the intensity λf by using expression 2.1. Normalising will then produce the required estimate of the sphere density f.

3. THE SIMULATION STUDY

The simulation study is designed principally to test the ability of the EMS algorithm to recover densities of sphere radii from simulated data sets. The advantage of the use of simulations over real data sets being that the true sphere density is known and we are thus able to measure in some way the accuracy of our recovered densities. The simulation study has also been used to investigate the efficiency of the EMS algorithm *with regard to computer processing time* and moreover to investigate the effect on the final density estimate of varying the number of data bins N or the number of bins M in the solution. By choosing densities of sphere radii similar to those chosen by other researchers in this area, it has also been possible to compare the EMS algorithm with other approaches. In particular, comparisons have been made between this method and the cross-validated spline methods used in Nychka *et al* (1984). Reference has also been made to the results and conclusions of Blodner *et al* (1984), a simulation study which compared different methods of solving the Wicksell problem.

3.1 The method of simulation

All the simulation programs have been written in Fortran 77 and run on a Sun 3/260 computer with a floating point accelerator. We begin by choosing a theoretical sphere size density f_0 for spheres in the medium. Three different densities have been considered: a bimodal density corresponding to a mixture of two Normals, chosen to test the ability of the EMS algorithm to resolve distinct peaks; a Weibull density, chosen because its mass is concentrated near zero, the point where the solution is most numerically unstable; and a Normal distribution. This last distribution has been used extensively to investigate the effect of binning on the solution. Together with the density f_0, a maximum value of the sphere radii, R, needs to be chosen. R should be large enough to ensure that there is only a very small probability of choosing a sphere radius $r > R$, while at the same time not so large that undue effort is given to the estimation of tail probabilities at the expense of the shape of the overall density function. A value for ε, the lower threshold for observations must also be selected. In the simulations that follow, ε is taken to be 0.04 mm, an approximation to the value of ε given in Nychka *et al.* (1984), namely 0.03846 mm. This was considered after experiment to be the value of the smallest cross-sectional radius reliably detected by all the researchers identifying cross-sections. Assuming that the probability of obtaining a sphere radius $r > R$ or $r \leq 0$ from the density f_0 is so small as to be neligible, the quantity $f_0(r)\,\delta r, 0 < r \leq R$, then corresponds to the probability that a sphere in the medium has actual radius in the interval $[r, r + \delta r)$ for small δr. If a random slice is taken through the medium, then the probability that a sphere, whose radius falls in the interval $[r, r + \delta r)$ is cut by the slice is given by

$$\frac{1}{\mu} r f_0(r)\,\delta r \qquad 0 < r \leq R \tag{3.1}$$

where $\mu = \int_0^R r f_0(r)\,dr$ is the mean of the sphere size distribution. The truth of this last assertion is demonstrated by Wicksell (1925). In practice, it is intuitive that large spheres are more likely to be cut by a random slice than small spheres and this must be accounted *for in any simulation procedure. Sphere radii must* therefore be generated not from the sphere density f_0 but the size-weighted density f^* given by

$$f^*(r) = \frac{1}{\mu} r f_0(r) \qquad 0 < r \leq R \tag{3.2}$$

For all the densities considered it is straightforward to simulate radii r from f_0 in

the interval $(0, R]$ and then to use an acceptance-rejection method to simulate *radii from the size-biased density* f^*.

However, the lower resolution on the observed radii poses a further problem, namely that we are unable to observe any cut spheres with radii $r < \varepsilon$. Thus it is necessary to simulate from the size-weighted distribution f^* truncated at ε. One *way of doing this is to simulate from f^* and reject any radii less than ε.* This method was used in the casè of the biomodal mixture of Normals density and the Normal density but proves to be very inefficient when stimulating from the Weibull density, as there are such a large number of spheres with radii less than ε. Under these circumstances it is better to stimulate *directly from the truncated* density given by

$$f_\varepsilon(r) = \frac{(r/\mu)\, f_0(r)}{\int_\varepsilon^R (r/\mu)\, f_0(r)\,dr} = \frac{r\, f_0(r)}{\int_\varepsilon^R r\, f_0(r)\,dr} = \frac{r}{\mu_\varepsilon} f_0(r) \qquad \varepsilon \leq r \leq R. \tag{3.3}$$

This involves simulating from the density f_0 truncated at ε and then, as before, using an acceptance-rejection method to obtain radii r from the size-weighted truncated density f_ε.

The radii that we generate in this way correspond to those spheres with radii exceeding ε that are cut by the slice. For each r generated we must therefore generate a corresponding profile radius y. Given that the slice intersects a sphere of radius r_0, the perpendicular distance d from the centre of the sphere to the planar slice will be distributed uniformly on $[0, r_0]$. Thus we generate d from $U(0, r_0)$ and then the required profile radius y using

$$y = \sqrt{r_0^2 - d^2} \tag{3.4}$$

It is possible that spheres with radii $r \geq \varepsilon$ will nevertheless give rise to profiles with radii $y < \varepsilon$. Any such y are ignored and the rest are binned into N data bins to *produce the data* $n_s(1), \ldots, n_s(N)$.

3.2 Results of the simulation study

The EMS algorithm outlined in Section 2 takes the N data values and determines the M values of the parameter ϕ and hence, by normalisation (to remove the unknown constant λ), the M values of our solution function f on the interval $[\varepsilon, R]$. It is desirable to find some way of measuring the accuracy of f as compared with the theoretical density f_0. Since, however, f is only defined on the interval $[\varepsilon, R]$ and is assumed to be zero elsewhere, it is necessary for the purposes of comparison to transform f_0 to a density truncated at ε. We therefore define

$$f_t(r) = \frac{f_0(r)}{\int_t^R f_0(r)\,dr} \qquad t \le r \le R \tag{3.5}$$

The error in estimating f_t by f is then measured by the expression

$$E = h_r \sum_{j=1}^{M} |f_j - f_{t,j}| \tag{3.6}$$

where f_j is the constant value of f on bin j and $f_{t,j}$ is defined to be the value of the truncated density f_t at the upper limit r_j of the jth bin. This expression is an approximation to the integral $\int_t^R |f(r) - f_t(r)|\,dr$, a measure of distance which is always well-defined for densities and has desirable invariance properties (Devroye and Gyorfi, 1985). Since f and f_t are both densities, it can be shown that the value of E must lie in the interval $[0, 2]$, values of E close to zero indicating that the two densities are very similar. With this in mind, we shall proceed to the results of the simulation study.

In the first example the theoretical sphere density is essentially a bimodal mixture of Normals given by

$$f_0(r) = \gamma f_1(r) + (1-\gamma) f_2(r) \qquad 0 < r < \infty \tag{3.7}$$

where $\gamma = 0.7$, $f_1(r)$ is a Normal density with mean 0.15 mm and standard deviation 0.03 mm and $f_2(r)$ is a Normal density with mean 0.275 mm and standard deviation 0.03 mm. The density is truncated at $t = 0.04$ mm. The value of R was fixed at 0.4 mm. As mentioned earlier, this distribution was chosen firstly to test the ability of the EMS algorithm to recover peaks and secondly to correspond as closely as possible to the density used in Experiment 3 of the simulations in Nychka *et al.* (1984) to enable comparisons to be made. Ten different data sets were simulated from this bimodal density. On average, 330 profiles were observed and grouped into $N = 50$ bins. The number of bins M in the solution was also taken to be 50 and three levels of smoothing were used, namely $k = 1$ point, 5 point, and 9 point smoothing. The value $k = 1$ has the effect of not smoothing at all. In this instance, the EMS algorithm is reduced to the EM algorithm.

In our consideration of the results it is necessary to distinguish between that variation in the quality of the solution f due to the difference in data sets and that variation arising out of the degree of smoothing used. With regard to the first source of variation, consider Table 1. Table 1 gives the value of the error function E and the number of iterations of the EMS algorithm to convergence, for each of the 10 data sets with $k = 5$ point smoothing. Figures 1–3 illustrate graphically the recovered density f for three of the data sets in the table, data sets 0, 3, 5 respectively. In each of the figures the solid line represents the theoretical truncated density f_t plotted on a grid of 50 points. The dashed line represent the density f recovered by the EMS algorithm, obtained by normalising the parameters $\phi_j(I)$, $j = 1, \ldots, M$ and plotting the resulting M values. Although the bimodal

Table 1 The value of the error function E for the ten data sets (with $k = 5$ point smoothing)

Data set	*Value of E*	*Number of iterations*
0	0.2353	29
1	0.2438	30
2	0.2490	53
3	0.3241	30
4	0.1198	49
5	0.1255	39
6	0.2339	39
7	0.2314	44
8	0.1632	31
9	0.2783	34

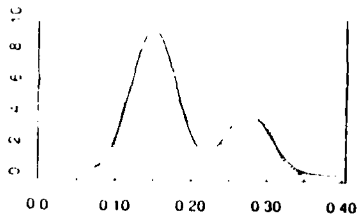


Figure 1 The EMS solution for data set 0 with 5 point smoothing (---) and the truncated bimodal Normal density f_t( )

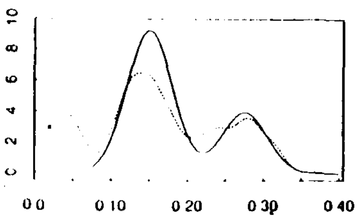


Figure 2 The EMS solution for data set 3 with 5 point smoothing (---) and the truncated bimodal Normal density f_t( )

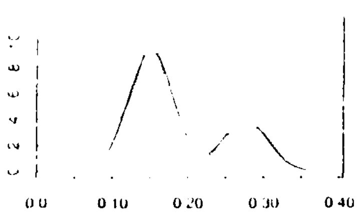


Figure 3 The EMS solution for data set 5 with 5 point smoothing ( ) and the truncated bimodal Normal density f( )

Table 2 The value of the error function for data set 5

Level of smoothing	*Value of E*	*Number of iterations*
None	0.8348	430
5pt	0.1255	39
9pt	0.1523	29

behaviour is apparent in all three of the solutions, there is a noticeable difference in the quality of the recovered densities between data sets (Compare in particular Figures 2 and 3.) This is a symptom of the ill-posed nature of the problem, which means that small random variations in the data can result in big differences in the final solution. Observe also the unstable behaviour of the solution f in Figures 1 and 2 near the point of truncation. This is a feature of this particular problem when very small profile radii y are observed.

Comparison of the results above with those of Experiment 3 of Nychka *et al.* (1984) as given in Figure 9 of their paper, reveals similarities between the solutions obtained by the two approaches. We can observe in Nychka *et al.* (1984) both the considerable variability of solution between data sets and the tendency on some occasions for instability near the resolution threshold ε.

Let us now consider the variation in the solution f that arises out of changing the level of smoothing. To do this we shall consider data set 5. Table 2 records the value of the error function E together with the number of iterations to convergence of the EMS algorithm acting on data set 5, with three levels of smoothing.

The solution functions f obtained using no smoothing, 5 point smoothing and 9 point smoothing are depicted by the dashed lines in Figures 4, 3 and 5 respectively. The improvement between Figures 4 and 3 is dramatic. It is clear that the use of the EM algorithm without smoothing not only takes a considerable number of iterations but tends to a solution which is too "wiggly" and erratic to be realistic. Figure 3 obtained after only a fraction of the former number of iterations is much more plausible and more accurate. Figure 5 could be said to be fractionally "oversmoothed". The effect of making k too large is to sacrifice accuracy for greater smoothness. To increase k further to 11 point or 15 point smoothing would lead to a further deterioration of the solution and possibly one in which all bimodal behaviour has been smoothed away. Clearly, in cases where the true density is unknown, choice of the most appropriate degree of smoothing is very subjective. Any smoothing is clearly better than none at all but it is best *initially to choose a small degree of smoothing, for example, 3 point smoothing.* This will reveal any important features such as bimodality. Increasing the level of smoothing may then be used to eradicate some of the smaller "wiggles" in the solution function, but should not be continued beyond a point where the initial features begin to be eroded.

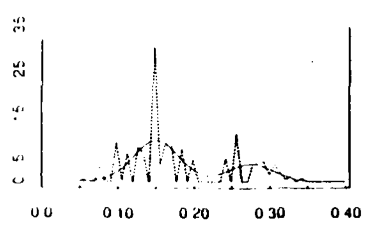


Figure 4 The EM unsmoothed solution for data set 5 (---) and the truncated bimodal Normal density f( )

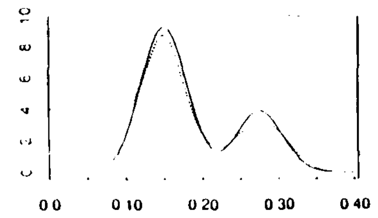


Figure 5 The EMS solution for data set 5 with 3 point smoothing (---) and the truncated bimodal Normal density f( )

The second example chosen is a truncated Weibull distribution. The theoretical underlying distribution is given by,

$$f_0(r)=\alpha\beta r^{\beta-1}e^{-\alpha r^{\beta}} \qquad 0<r<\infty \tag{3.8}$$

where we take $\alpha=12.0$ and $\beta=0.1$. As before, f_ε, the truncated distribution, is obtained by setting ε to be 0.04 mm and R to be 0.4 mm. The Weibull was chosen as it is a distribution whose mass is concentrated near the point of truncation, the point where the solution is most numerically unstable. Once again 10 different

Table 3 The value of the error function E for the ten data sets (with k = 9 point smoothing)

Data set	*Value of E*	*Number of iterations*
0	0.0928	18
1	0.2341	18
2	0.1564	18
3	0.2731	19
4	0.1940	19
5	0.1519	20
6	0.1086	18
7	0.1163	17
8	0.2292	19
9	0.1446	17

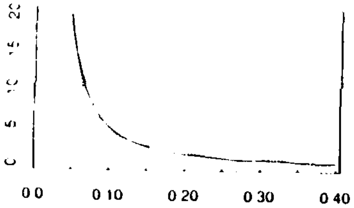


Figure 6 The EMS solution for data set 0 with 9 point smoothing ( ) and the truncated Weibull density f( )

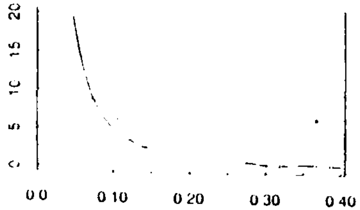


Figure 7 The EMS solution for data set 3 with 9 point smoothing (---) and the truncated Weibull density f( )

data sets were simulated and on average 190 profiles were observed. N and M were taken to be 50. Table 3 records the value of the error function E and the numer of iterations to convergence for each of the 10 data sets, this time with k = 9 point smoothing.

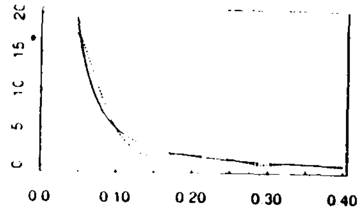


Figure 8 The EMS solution for data set 4 with 9 point smoothing (---) and the truncated Weibull density f( )

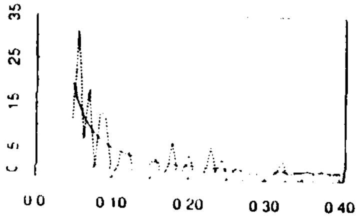


Figure 9 The EM unsmoothed solution for data set 4 (---) and the truncated Weibull density f( )

Table 4 The value of the error function for data set 4

Level of smoothing	*Value of E*	*Number of iterations*
None	0.8791	316
3pt	0.2329	28
9pt	0.1940	19

Figures 6–8 illustrate the density f (dashed line), recovered by the EMS algorithm from data sets 0, 3 and 4 respectively, plotted against the theoretical truncated Weibull f_t (solid line). Considerable variability in the quality of f between data sets is again apparent but overall Table 3 and the graphs shown indicate that the recovered solutions are very good estimates of the original density. If we consider the effec to varying the level of smoothing on data set 4, we see that the unsmoothed EM algorithm (see Figure 9 and Table 4) takes a massive 316 iterations to converge to the maximum likelihood solution, which is in any case very spiky and unrealistic. Introducing 3 point smoothing leads to a considerable improvement while taking k to be 9 point smoothing produces an even better solution (see *Figure 8*) after only *19 iterations*.

Choice of the level of smoothing, as before, is subjective. Some small "wiggles" in the final solution will be inevitable due to the simplicity of the smoothing procedure. To employ more sophisticated techniques to remove them would have a largely cosmetic function and it is debatable whether the improvement in the solution would justify the added computation.

The third example is a Normal distribution. The theoretical underlying distribution is essentially given by

$$f_0(r) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{(r-\mu)^2}{\sigma^2}\right) \qquad 0<r<\infty \qquad (3.9)$$

where $\mu = 0.55$ mm and $\sigma = 0.14$ mm. This is the distribution used by Blodner *et al* (1984) in their simulation study. As before, truncation occurs at $\varepsilon = 0.04$ mm and the upper bound R is fixed at 1.0 mm. This distribution has been used extensively to investigate the effect of binning on the recovery of the solution. To this end, the number of data bins N into which the generated profile radii were binned (on average 300 profiles were generated) was allowed to take the values 10, 20, 50 and 75. The number of bins M in the final solution was also varied, taking values of 10, 20, 50, 75 and 100. For given values of N and M, solutions were recovered using the EMS algorithm for each of the 10 data sets and varying levels of smoothing. Because of the scope of this part of the study, no attempt will be made to describe individual results but a summary of the findings will be presented.

Firstly we observe that due to smoothing we are able to choose the value of M independently of N. In particular, M may be made larger than N, the accuracy of the solution being maintained by increasing the level of smoothing as M increases. For the number of data bins N taking values 10, 20 and 50, it was found that it was optimal to choose M between 20 and 50, greater accuracy being possible as M increases. To choose M less than 20 led to jagged results which could not be smoothed effectively. This will be demonstrated in a later example, when comparisons are made with the work of Blodner *et al* (1984). Choosing M above 50 led to only marginal improvement which did not justify the additional computation required.

Secondly, it is important to assess the effect of binning on the data. It may, for example, be advantageous to regroup a set of data into a smaller number of bins before seeking a solution. To investigate this problem M was fixed at 50 bins and N was allowed to take the values 10, 20, 50 and 75. Although intuitively it would seem that the more data bins we have the better, in reality this is not always the case. Results have indicated that there are occasions when to use $N = 10$ data bins is preferable to taking $N = 20$ or more. In particular, solution functions exhibiting some degree of numerical instability near the point of truncation, for example, the solution in Figure 1, may be improved in this way. This would seem natural since regrouping data into a smaller number of bins should serve to smooth out some of the random fluctuations in the data that could lead to erratic behaviour in the solution function. Consider the bimodal density given in Figure 1. Here the simulated data set 0 is grouped in 50 bins and the solution function f (given by the dashed line) is obtained after 29 interations of the EMS algorithm with 5 point smoothing. The same data set, when grouped in 10 data bins, gives rise to the solution in Figure 10 after 36 iterations with $M = 50$ and 5 point smoothing.

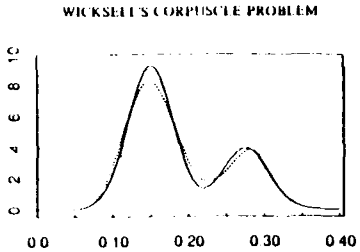


Figure 10 The EMS solution for data set 0 (binned in 10 data bins) with 5 point smoothing (···) the truncated bimodal Normal density f( )

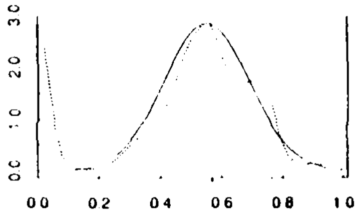


Figure 11 The EMS solution (···) for a data set simulated from a Normal density without truncation, plotted against the original density ( )

The improvement in f is striking. Reducing the number of data bins from 50 to 10 has eliminated the unstable behaviour near ε, thereby improving the overall shape of the density and reducing the value of the error function E from 0.2353 to 0.1003. It should be stressed that this marked improvement is the exception, not the norm. There are instances where reducing N, with the consequent reduction in information, leads to a deterioration in the final solution. As with smoothing, the choice of N has to be chosen subjectively, and in so doing, a satisfactory compromise must be reached between the need for numerical stability and the need to retain enough information to be able to resolve peaks and fluctuations in the solution function.

We have seen, in the results of the simulations, how erratic behaviour is possible near the resolution threshold ε on those occasions when very small profile radii y near to ε are observed. This behaviour is very much more extreme when there is no resolution threshold, that is when ε is taken to be zero. Consider the two graphs given in Figure 11 and Figure 12.

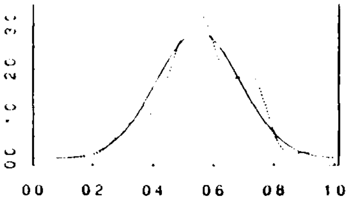


Figure 12 The EMS solution ( ) for the data set in Figure 11 but with truncation at ε = 0.04, plotted against the original density truncated at ε ( )

In the first, the data has been generated from N (0.55, 0.0196) with $\varepsilon = 0$, R taken to be 1.0 and the number of bins $N = M = 50$. The solid line depicts the theoretical truncated Normal density f_ε on the interval $(0, R)$ and the dashed line is the density recovered after 17 iterations of the EMS procedure with 5 point smoothing. In the second, exactly the same data has been generated but it has been grouped in 50 bins on the interval (0.04, 1.0) rather than (0.0, 1.0), and all y generated, with values less than 0.04, have been deleted. The solid line now represents the truncated Normal density f_ε on (0.04, 1.0) and the dashed line, the density recovered after 37 iterations with 5 point smoothing. The example illustrates very graphically the effect of numerical instability near the origin and demonstrates the importance of truncation, which to all practical intents and purposes is inevitable.

To conclude this section we shall briefly discuss the question of the uniqueness of the solution obtained by the EMS algorithm. In Section 2, it was emphasised that although the EM algorithm in the context could be expected to converge to a global maximum of the likelihood function, the global maximum need not be unique, and under these circumstances the choice of the starting value ϕ_0 would determine the final parameter values $\phi(j)$, $j = 1, \ldots M$ and hence the final solution f. Simulation experience suggests that the EMS algorithm will always converge to a solution ϕ. However, what the solution represents in terms of the likelihood function is not as yet clear and it is necessary to establish the uniqueness of the solution. In all the simulations so far, ϕ_0 has been taken to be a discrete approximation to the Uniform distribution on the interval $[\varepsilon, R]$. To address this last question, three different starting values were chosen and for each one a number of different simulations were run with varying data sets, levels of smoothing and theoretical densities. In all cases the solutions obtained by the EMS algorithm for a given problem were the same for all four starting values. Although this by no means constitutes a proof, it does suggest quite strongly that the EMS algorithm will produce unique solutions for any level of smoothing $(k > 1)$.

Table 5 The value of the error function E for data set in Blödner et al. (1984)

Algorithm	*Value of* E	*Number of iterations*
SA	0.2395	
EM	0.3277	119
EMS 3pt	0.3640	19

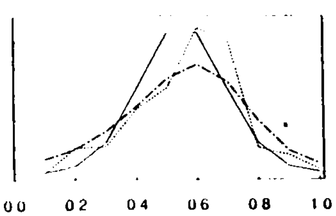


Figure 13 The EM solution (---) and the EMS solution with 3 point smoothing ( ) and the *original density* ( ) from which the data was simulated

COMPARISON OF THE EMS ALGORITHM WITH OTHER APPROACHES

In this section we consider two papers. The first by Blödner *et al.* (1984) and the second by Nychka *et al.* (1984). The paper by Blödner *et al.* (1984) uses a simulation study to compare different approached to the Wicksell corpuscle problem, including the approaches of Scheil Schwartz-Saltykov and Goldsmith-Cruz-Orive. The method that appears to be favoured for the recovery of continuous densities is the Scheil Schwartz Saltykov method (denoted SA for short), a numerical method in which the integral relationship between the sphere density and profile radii density is replaced by a finite sum, producing a system of linear equations to be solved. In the paper (p. 68), a simulation example of a Normal distribution with mean $\mu = 0.55$ and standard deviation $\sigma = 0.14$ is given. For the data given in this particular example and with $M = N = 10$, $\varepsilon = 0.0$ and $R = 1.0$, I used the EM and the EMS algorithms to obtain solutions and then compared these with the solution obtained using SA. (See Table 5).

Consideration of the values of the error function E for the EMS method and the SA method clearly demonstrate the inadequacy of the EMS algorithm when M is small. This point was made earlier when we considered the effect of binning on the solution function and is illustrated in Figure 13, where the jaggedness of both the EM solution and the EMS solution after 3 point smoothing can be seen in the dashed curves and dot-dashed curves respectively. The solid curve depicts the theoretical density f, plotted on 10 points. For the purposes of comparison, Figure 14 gives the solution obtained by the SA method plotted against f_ε.

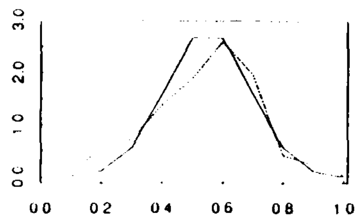


Figure 14 The SA solution (- - -) for the same data as in Figure 13, and the original density (——)

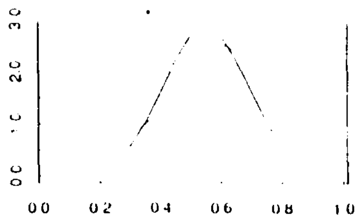


Figure 15 The EMS solution with 3 point smoothing (plotted on $M = 20$ bins) (- - -) for the same data as in Figure 13, and the original density (——)

However, increasing M to 20 bins and using 3 point smoothing reduces the error function E to a value of 0.2265 and produces the solution given by the dashed curve in Figure 15. (The solid curve is the theoretical truncated density f, plotted on 20 points.) This compares favourably with the result obtained by the SA method. Increasing the value of M to 50 and using 5 point smoothing further improves the solution function. This illustrates one of the advantages of the EMS algorithm over the SA method, namely its capacity to use large values of M and N to improve the final solution. In Blödner *et al.* (1984) it is recommended that the SA method be used on data grouped in at most 10 bins to prevent "waves" in the solution. Without smoothing, these "waves" are also a feature of the EMS solution (see Figure 4) for $M > 10$, but the smoothing procedure "smooths" out such fluctuations in the solution to leave only the underlying trends.

I have already made some comparisons between the EMS algorithm and the cross-validated spline methods of Nychka *et al.* (1984) by comparing the results of the simulations. Now we shall consider the effectiveness of the EMS algorithm on some real data that appeared in their paper. The problem is to estimate the tumour-size distribution of spherical microtumours induced in mouselivers following the injection of a carcinogen. The data arising from parallel sections taken through the liver of the mouse appears in 64 bins, and is truncated below at

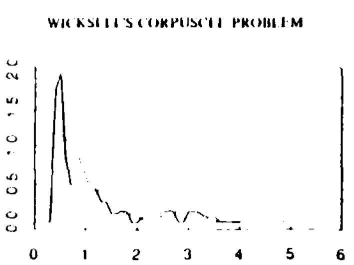


Figure 16 The EMS solution with 3 point smoothing (- - -) for the real mouse liver data, and the true density (——)

$\varepsilon = 0.3846$ mm and above at $R = 5.05$ mm. In this example, the mouse liver has been completely dissected, which will enable comparisons to be made with a "true" realisation from the assumed underlying density.

The solid line in Figure 16 corresponds to the true histogram of the radii of mouse tumours, as given in Figure 6 of the paper by Nychka *et al.* (1984), transformed to a density. The broken line depicts the result of the EMS algorithm after 25 iterations with $M = 50$ and 3 point smoothing. Although underestimating the density near ε, it produces an estimate which is otherwise consistent with the "truth". The result of using cross-validated splines on the data appears in Figure 6 of Nychka *et al.* (1984). Comparison of the two estimates reveal that they are very similar.

To conclude, it can be said that the EMS algorithm is superior to the methods investigated in Blödner *et al.* (1984) since it has the ability to yield much more accurate solutions from the same data, simply by increasing the number of solution bins M and using appropriate smoothing. The SA method and many of the others investigated in Blödner et al. (1984) are limited to small values of M and N. This means that not only does data have to be regrouped into a small number of bins, with the consequent reduction in information, but also the solution functions obtained are very "rough" as they have to be plotted on a coarse grid of points.

Comparison of the EMS algorithm with the method in Nychka *et al.* (1984) for both simulated and real data, facilitated by comparing the graphs in this paper with those given on pages 838–841 of the paper by Nychka, indicate that the results of the two approaches are similar. With this in mind, the EMS is to be recommended for its simplicity and computational efficiency.

4. CONCLUDING REMARKS

The EMS algorithm is a new procedure which utilises the EM algorithm in conjunction with simple smoothing. Simulation studies strongly suggest that, unlike the EM algorithm, the new procedure converges quickly to a unique solution independent of the starting value. The amount of computer processing time used to recover a sphere radii density from a given set of data depends upon the number of bins and the degree of smoothing. However the majority of examples took between 10 and 25 seconds of CPU time on the Sun 3/260 with floating point accelerator. The algorithm is able to recover distinct peaks in densities and to locate concentrations of mass, even near the point of truncation. A significant advantage of the procedure is that non-negativity is preserved automatically.

The choice of the level of smoothing is currently subjective and will depend upon the number of bins required in the solution and the prior belief that is held about the smoothness of the sphere density. Some smoothing is essential whenever the number of solution bins M is greater than 10. Even small degrees of smoothing will significantly improve the rate of convergence and the accuracy of the final solution, over that obtained by the EM algorithm.

Numerical instability may sometimes be a problem near the point of truncation. Regrouping the data into a smaller number of bins may help to alleviate the problem, but such throwing away of information should be done with considerable care.

Many of the advantages of this method have been emphasised throughout the paper and Section 3.3 has demonstrated that it compares favourably with those methods used in Blodner *et al.* (1984) and Nychka *et al.* (1984). However, considerable work still needs to be done to prove in theory what simulation experience has suggested in practice, namely that the EMS algorithm converges and to a unique solution. Consideration also needs to be given to the inclusion of over projection effects and capping angles, as well as lower resolution thresholds. Finally, we observe that other methods, for example Cruz-Orive (1983), produce estimates of the total number of spheres in the medium, as well as estimates of the sphere density. It would be useful to extend the use of the EMS algorithm and provide such an estimate.

Acknowledgements

This work was carried out under the supervision of Prof. B. W. Silverman and with the financial support of the Science and Engineering Research Council. The author would also like to thank Dr G. K. Eagleson and Dr M. C. Jones for their many helpful comments on earlier drafts of this paper.

References

Blodner, R., Muhlig, P. and Nagel, W. (1984). The comparison by simulation of solutions of Wicksell's corpuscle problem. *J. Microsc.* **135**, 61–74.

Cruz-Orive, L. M. (1983). Distribution-free estimation of sphere size distributions from slabs showing overprojection and truncation, with a review of previous methods. *J. Microsc.* **131**, 265–290.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM Algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39**, 1–38.

Devroye, L. and Gyorfi, L. (1985). *Nonparametric Density Estimation*. Wiley, New York.

Jones, M. C., Silverman, B. W. and Wilson, J. D. (1987). A smoothed EM approach to a class of problems in image analysis and integral equations. *In draft*.

Nychka, D., Wahba, G., Goldfarb, S. and Pugh, T. (1984). Cross-validated spline methods for the estimation of three-dimensional tumor size distributions from observations on two-dimensional cross sections. *J. Amer. Statist. Ass.* **79**, 832–846.

Vardi, Y., Shepp, L. A. and Kaufman, L. (1985). A statistical model for positron emission tomography. *J. Amer. Statist. Ass.* **80**, 8–37.

Wicksell, S. D. (1925). The corpuscle problem. Part I. *Biometrika* **17**, 84–99.

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.* **11**, 95–103.

APPENDIX: CALCULATION OF THE CONDITIONAL PROBABILITIES $p(i, j)$

Recall that $p(i, j)$ is the probability that a sphere is recorded as a disc of radius y in bin i given that the sphere has radius r in bin j ($i = 1, \ldots, N$; $j = 1, \ldots, M$).

It is assumed that the truncated domain for the disc radii density, $\varepsilon \leqq y \leqq R$, is divided into N data bins of a predetermined size, $[y_0, y_1), [y_1, y_2), \ldots, [y_{N-1}, y_N]$ where $y_0 = \varepsilon$ and $y_N = R$. We divide the truncated domain for the sphere radii density, $\varepsilon \leqq r \leqq R$, into M solution bins of equal width h_r, which are denoted by $[\varepsilon, r_1), [r_1, r_2), \ldots, [r_{M-1}, R]$.

Let $P_Y(i) = \Pr$ (a sphere is recorded as a disc in bin i | the sphere radius falls in the interval $[\varepsilon, R]$). Then

$$P_Y(i) = \sum_{j=1}^{M} p(i, j) P_R(j) \qquad i = 1, \ldots, N \tag{A.1}$$

where we define $P_R(j)$ to be Pr (a sphere has radius r in the jth interval). It follows that

$$\sum_{i'=1}^{N} P_Y(i') = \Pr \text{ (a sphere will be cut and recorded i.e., } y \geqq \varepsilon \mid \text{sphere radius } r \in [\varepsilon, R]).$$

This quantity may be less than one since (a) spheres need not be cut and (b) spheres that are cut may give rise to discs with radii too small to be recorded. Thus

$$\Pr(\text{a disc in } i \mid \text{a sphere is recorded}) = \frac{P_Y(i)}{\sum_{i'=1}^{N} P_Y(i')} \qquad i = 1, \ldots, N. \tag{A.2}$$

Consider now Eq. 1.1 in Section 1 which relates the density of the radii of the discs in the plane $g(y), 0 < y \leqq R$ to the density of the radii of the spheres in the medium $f(r), 0 < r \leqq R$.

$$g(y) = \frac{1}{\mu}\int_y^R \frac{y}{\sqrt{r^2-y^2}} f(r)\,dr \qquad 0<y \tag{A 3}$$

Then

$$\int_{y_{i-1}}^{y_i} g(y)\,dy = \Pr(\text{a disc in bin } i \mid \text{sphere is cut})$$

and

$$\sum_{i=1}^{N}\int_{y_{i-1}}^{y_i} g(y)\,dy = \Pr(\text{a sphere is recorded as a disc i.e. } y \geq \varepsilon \mid \text{the sphere is cut})$$

Using probability arguments it can be shown that,

$$\Pr(\text{a disc in bin } i \mid \text{sphere is cut and recorded}) = \frac{\Pr(\text{a disc in bin } i \mid \text{sphere is cut})}{\Pr(\text{sphere is recorded} \mid \text{sphere is cut})} = \frac{\int_{y_{i-1}}^{y_i} g(y)\,dy}{\sum_{i=1}^{N}\int_{y_{i-1}}^{y_i} g(y)\,dy} \tag{A 4}$$

It follows that (A 2) and (A 4) are equivalent and hence that $P_Y(i)$ is proportional to $\int_{y_{i-1}}^{y_i} g(y)\,dy$

Approximating $f(r), 0<r\leq R$ by the step function

$$f(r) = f_0\,\chi[r\in(0,\varepsilon)] + \sum_{j=1}^{M} f_j\,\chi[r\in j\text{th interval}] \tag{A 5}$$

where χ is an indicator function, we obtain

$$\int_{y_{i-1}}^{y_i} g(y)\,dy = \sum_{j=1}^{M}\frac{f_j}{\mu}\int_{y_{i-1}}^{y_i}\int_y^R \frac{y}{\sqrt{r^2-y^2}}\,\chi[r\in[r_{j-1},r_j)]\,dr\,dy \tag{A 6}$$

(Notice the term in f_0 may be disregarded as y_{i-1} is never less than ε.)

Denote by L_{ij} the double integral in (A 6). Then we can write

$$P_Y(i) \propto \int_{y_{i-1}}^{y_i} g(y)\,dy \propto \sum_{j=1}^{M} f_j L_{ij} \qquad i=1,\dots,N \tag{A.7}$$

Comparing this with (A.1) and writing $P_R(j) = h_j f_j$, $j=1,\dots,M$, it can be seen that $p(i,j)$ is proportional to the double integral L_{ij}

The quantities L_{ij} are determined by considering the relative positions of the ith data bin and the jth solution bin and evaluating the integral $\int_a^b (y/\sqrt{r^2-y^2})\,dr$ between appropriate limits. For examples, if $y_{i-1}<y_i\leq r_{j-1}<r_j$, then

$$L_{ij} = \int_{y_{i-1}}^{y_i}\int_y^R \frac{y}{\sqrt{r^2-y^2}}\,\chi[r\in[r_{j-1},r_j)]\,dr\,dy$$

$$= \int_{y_{i-1}}^{y_i}\int_{r_{j-1}}^{r_j} \frac{y}{\sqrt{r^2-y^2}}\,dr\,dy$$

$$= \int_{y_{i-1}}^{y_i}\left[y\ln\left(\frac{r+\sqrt{r^2-y^2}}{y}\right)\right]_{r_{j-1}}^{r_j} dy$$

and we approximate this final integral by evaluating the integrand at the midpoint of the interval and multiplying by the width of the interval. A better approximation to this integral was tried using Romberg integration, but it did not significantly improve the solution and so the simpler procedure above was retained. Since $p(i,j)$ is only known to be proportional to L_{ij} with unknown constant of proportionality C^* (the same for all i and j), if we use the L_{ij} as our estimates of the $p(i,j)$ in the EMS algorithm then the resulting estimates of the parameters ϕ increase by a factor of C^* at every iteration. To adjust for this, it is necessary to normalise the $\phi_{p+1}(j)$'s to a density at the end of each iteration and to compare successive densities when testing for convergence.

Appendix 7

An evaluation of the ICM algorithm for image reconstruction.

by

R. H. Glendinning

An Evaluation of the ICM Algorithm for Image Reconstruction

R. H. GLENDINNING

School of Mathematical Sciences,University of Bath, Bath BA2 7AY, U.K.

We examine the properties of Iterated Conditional Modes (ICM) estimation for a number of synthetic binary images using simulation.

KEY WORDS : *Ill-posed problem;image reconstruction;ICM; Simulated Annealing;smoothing parameter; neighbourhood system;Monte-Carlo.*

1. INTRODUCTION

In the last few years considerable interest has been shown in the problems posed by the analysis of images corrupted by random noise. The reconstruction of such images leads to special difficulties as it is an ill-posed problem (in the sense described by O'Sullivan, 1986). Typically the reconstruction of an array of pixels will have as many parameters as observations. A number of techniques have been proposed which *solve* ill-posed problems by restricting the class of admissible solutions,see Marroquin,Mitter & Poggio (1987). This is achieved by introducing a priori knowledge about admissible solutions.

Much interest currently centres on techniques which incorporate knowledge about the underlying image using Bayesian methodology, See Geman & Geman (1984) ; Kashyap & Lapsa (1984). These techniques assume that the underlying scene can be adequately described as a realisation from a prescribed Markov random field. Motivated by this approach Besag (1986) introduced a technique known as **ITERATED CONDITIONAL MODES (ICM)**. This iterative procedure incorporates knowledge about the underlying scene by the choice of a 'neighbourhood system' ,weight function and smoothing parameter. Broadly speaking this method exploits the tendency of adjacent pixels to have the same colour. A similar approach based on spatial auto regression is described in Woods,Dravida & Mediavilla (1987).

In this paper we use simulation to evaluate the performance of ICM in reconstructing binary (black-white) images. The reconstruction of binary images is of considerable practical importance as many problems in object recognition and manipulation fall into this category. For simplicity we suppose that the underlying scene can be partitioned into an array of pixels (picture elements) which are *uniquely* coloured black or white. At each pixel we observe a signal which depends on its colour. We consider the case where each signal is additively corrupted by independent normally distributed noise. These are highly unrealistic assumptions as they ignore the problems associated with mixed pixels, signal spread etc. However we believe that the study of ICM in this simplified setting will give valuable insight into its behaviour in more complex situations.

In section 2 we describe the basic ICM algorithm and recall some basic facts about Markov random fields. The synthetic scenes used in this study are described in section 3. In section 4 we examine the influence of the neighbourhood system and weight function on the quality of our reconstructions. The choice of smoothing parameter is discussed in section 5. We are particularly interested in identifying properties of the underlying scene which influence the value given to β (the smoothing parameter). Some distributional properties of ICM reconstructions are discussed in section 6. The numerical performance of the basic ICM algorithm is discussed in section 7. We describe several modifications of the basic algorithm which enhance its efficiency. Our findings are summarised in section 8.

The problem of restoring corrupted images has a long history in the image processing literature, where a number of techniques of varying sophistication have been suggested, see Bovik,Huang & Munson (1987) or Rosenfeld & Kak (1982). A comparison of ICM with the multitude of competing techniques is not attempted in this paper.

2. THE ICM ALGORITHM AND MARKOV RANDOM FIELDS

Let W be a rectangular window in the plane which is partitioned into an $(m \ x \ n)$ array of rectangular pixels of equal size. We assume that each pixel can be uniquely coloured. The available colours are labelled $(1,2,...,c)$. In this paper we restrict attention to scenes with two colours which we call black and white. The colour of the $(i,j)^{th}$ pixel is denoted by x_{ij}. We refer to (x_{ij}) as the true or underlying scene. Suppose we observe an array of signals (y_{ij}) generat-

$$y_{ij} = \mu(x_{ij}) + \varepsilon_{ij}, \tag{2.1}$$

where (ε_{ij}) are independent and identically distributed random variables and $\mu(.)$ is a function of x_{ij} only. The object of image analysis is to estimate the true or underlying scene (x_{ij}) from (y_{ij}). In this paper we consider real-valued signals only. Models of this form are not canonical in the study of corrupted images and the reader is referred to Besag (1986) for a discussion of alternative models.

At first sight the natural way of estimating (x_{ij}) is by maximum likelihood. In this approach we find (x_{ij}) which maximises

$$l(\ (y_{ij}) \mid (x_{ij})\ ) = \prod_{i=1}^{m}\prod_{j=1}^{n} f(y_{ij} \mid x_{ij}). \tag{2.2}$$

where $f(y_{ij} \mid x_{ij})$ is the fully specified density function of y_{ij} conditional on x_{ij}. The estimates produced by this approach are usually unsatisfactory as (2.1) has as many parameters (x_{ij}) as observations. To improve the situation Geman & Geman (1984) and Besag (1986) introduce information about (x_{ij}) into the estimating procedure. This is achieved by regarding (x_{ij}) as a realisation from a **Markov random field** (MRF) . A detailed account of the salient features of MRF's can be found in Geman & Geman (1984) ; Besag (1974,1986) or Suomela (1976). We briefly outline the main properties of MRF's relevant to the discussions in this paper.

For each pixel (i,j) we associate a set of pixels $F_{(i,j)}$, not including (i,j) called the **neighbourhood** of (i,j). The collection of sets $(F_{(i,j)})$ is called a neighbourhood system and satisfies the condition : $(p,q) \in F_{(i,j)} \Leftrightarrow (i,j) \in F_{(p,q)}$.

Then (x_{ij}) is a MRF if

(1) $P(x_{ij} \mid (x_{pq}, p \neq i, q \neq j)) = P(x_{ij} \mid x_{pq}, (p,q) \in F_{(i,j)})$,

(2) $P((x_{ij})) > 0$,

where $P((x_{ij}))$ is the probability associated with the realisation (x_{ij}). Conditions 1 and 2 impose severe restrictions on $P(.)$. Valid forms of $P(.)$ are given by the Hammersley-Clifford Theorem, see Besag (1974) or Suomela (1976).

We follow Geman & Geman (1984) and adopt a Bayesian approach where we estimate (x_{ij}) from its posterior distribution

$$l((y_{ij}) \mid (x_{ij}))P((x_{ij})). \tag{2.3}$$

A plausible estimate of (x_{ij}) is the value of (x_{ij}) which maximises (2.3). This is the MAP estimate of (x_{ij}). Geman & Geman (1984) use simulated annealing to maximise (2.3). Van Laarhoven & Aaris (1987) give a comprehensive description of simulated annealing and its application to image analysis. Note that Greig,Porteous and Seheult, in the discussion of Besag (1986) show that the MAP estimate of a binary scene can be calculated exactly. It is not known whether the MAP estimator has any desirable properties in this context.

Besag (1986) introduces an alternative estimator of (x_{ij}) known as **ITERATED CONDITIONAL MODES (ICM)**. This algorithm converges to a local maximum of (2.3). Let $(\hat{x}_{ij})$ be the current estimate of (x_{ij}). For each pixel we find the value of x_{ij} which maximises

$$f(y_{ij} \mid x_{ij})P(x_{ij} \mid (\hat{x}_{ij})), \tag{2.4}$$

where $P(x_{ij} \mid (\hat{x}_{ij}))$ depends on the neighbours of (i,j) only. Consider an example. Let (x_{ij}) be a binary scene and (ε_{ij}) an array of independent normally distributed random variables with zero mean and variance σ^2. We represent our knowledge of (x_{ij}) by a MRF with neighbourhood system $(F_{(i,j)}=((i-1,j),(i+1,j),(i,j-1),(i,j+1)))$ and conditional probabilities

$$P(x_{ij}=k \mid (x_{pq}p \neq i, q \neq j)) = \frac{\exp(\beta u_{ij}(k))}{\exp(\beta u_{ij}(0)) + \exp(\beta u_{ij}(1))} \quad , \quad k=0,1 \ , \tag{2.5}$$

where the weight function $u_{ij}(k)$ is the number of neighbours of x_{ij} with colour k. The value of x_{ij} which maximises (2.4) minimises

$$(2\sigma^2)^{-1}(y_{ij} - \mu(x_{ij}))^2 - \beta\hat{u}_{ij}(x_{ij}), \tag{2.6}$$

where $\hat{u}_{ij}(x_{ij})$ is the number of neighbours of (i,j) which have colour x_{ij} under the current estimate $(\hat{x}_{ij})$. We call β the smoothing parameter. The extension of (2.6) to non-gaussian noise is immediate.

Notice that (2.6) is in the form of a penalised likelihood and may be interpreted in this way without recourse to Bayesian arguments. Note that ICM and MAP are not equivalent for most scenes. Typically smaller values of β (relative to ICM) are required for MAP, see Greig,Porteous and Seheult,in the discussion of Besag (1986). The relationship between techniques like ICM and other regularisation procedures is discussed in Titterington (1985).

3. DESCRIPTION OF THE SIMULATION STUDY

Seven scenes of varying complexity were constructed by partitioning the unit square into 10^4 square pixels of equal size. The colour of each pixel was assigned to the colour of its mid-point. In this study we use black and white scenes only.

To identify properties of ICM more easily we restrict attention to simple synthetic scenes which cover a small alphabet of forms rather than use naturally occurring images. Five simple geometric scenes are displayed in figures 1 to 5. The remaining scenes, MRF2 and MRF3 (figures 6 and 7) are realisations from a Markov random field with *prescribed* number of black pixels (approx 50%). MRF2 and MRF3 were constructed using the algorithm described in Cross & Jain (1983). Notice that we are sampling from the *conditional* distribution of the prescribed MRF. We believe that realisations constructed in this way capture much of the local structure of the *unconditional* model. In the next section we describe three Markov random fields, (Models I,II and III) which are commonly used in this context. MRF2 is drawn from Model II with β=0.5 and MRF3 from Model III with β=0.75.

We construct an array of signals (y_{ij}) using (2.1) with $\mu(black)=1$, $\mu(white)=0$ and (ε_{ij}) an array of independent normally distributed random variables with zero mean and variance σ^2. The maximum likelihood reconstruction is calculated and used as the initial state for the ICM algorithm. This iterative procedure is terminated after twelve iterations. Typically convergence occurs after six iterations. This process is repeated fifteen times for each combination of parameter and underlying model. The efficiency of this algorithm is discussed in section 7.

Many criteria can be used to evaluate reconstructions. Essentially its choice depends on the image characteristics of greatest interest. In this paper we use the number of misclassified pixels as an appropriate measure. The suitability of this criteria has been the subject of much recent debate, see the discussion of Besag (1986). We point out the limitations of this criteria where appropriate.

Figs 1-7
here

4. THE CHOICE OF MODEL.

In this section we examine the effect of choosing three different weight functions in (2.6). The choice of β is discussed in section 5. In a Bayesian framework we are modelling our knowledge of the uncorrupted scene by a MRF with prescribed structure. Cross & Jain (1983) show that simple MRF's can generate a wide variety of binary scenes. The problem of choosing suitable MRF's to model specific scenes is not well understood,see Kashyap & Chelappa (1983) , Enting & Welberry (1978) and Pickard (1987). The last two authors discuss parameter estimation for Markov random fields. An additional complication arrises when our knowledge about the underlying scene is imprecise or difficult to model by a MRF. The success of this approach rests on the assumption that only certain modest properties of our 'prior' are important. Some tentative observations on the robustness of ICM reconstruction to model specification are given in sections 4 and 5.

In this section we use three different MRF's to describe our knowledge about the scenes presented in figs 1 to 7. We examine the misclassification rate achieved by ICM using each model and several values of the parameter β. The models used are as follows:

MODEL I : A first order neighbourhood.

$$F_{(i,j)} = (\ (i-1,j),(i+1,j),(i,j+1),(i,j-1)\ ).$$

$$P(x_{ij}=k \mid F_{(i,j)}) = \frac{\exp(\beta u_{ij}(k))}{\exp(\beta u_{ij}(0)) + \exp(\beta u_{ij}(1))},\quad k=0,1. \qquad (4.1)$$

where

$$u_{pq}(k) = 1 \quad , \quad when \quad (p,q)\in F_{(i,j)} \text{ and } x_{pq} = k, \qquad (4.2)$$

and zero otherwise.

MODEL II: A second order neighbourhood.

$$F_{(i,j)} = (\ (i-1,j+1),(i+1,j+1),(i-1,j-1),(i+1,j-1),$$
$$(i-1,j),(i,j+1),(i+1,j),(i,j-1))$$

$P(x_{ij}=k \mid F_{(i,j)})$ is given by (4.1) and (4.2).

MODEL III : As for II with down weighted diagonals. $F_{(i,j)}$ as for the previous model and $P(x_{ij}=k \mid F_{(i,j)})$ given by (4.1) with

$$u_{pq}(k) = 1 \quad , \ (p,q) \in ((i+1,j),(i-1,j),(i,j+1),(i,j-1)) \ \text{ and } \ x_{pq} = k.$$

$$u_{pq}(k) = 2^{-1/2} \ , \ (p,q) \in ((i-1,j+1),(i+1,j-1),(i+1,j+1),(i-1,j-1)) \ \text{ and } \ x_{pq} = k.$$

$$u_{pq}(k) = 0 \ \textit{otherwise}. \tag{4.3}$$

There are conflicting opinions as to whether models should be modified for pixels adjacent to the window,see Ripley (1984). In this study we use the unmodified models I,II and III. The effects of modification appear small relative to the standard errors encountered in this study. Cross & Jain (1983) show that models like I,II and III can be used to construct a wide variety of binary scenes.

TABLE I

Comparison of models I, II and III
Smallest average percentage of misclassified pixels
β taking values in (0.25,0.5,0.75,1.0,1.25,1.5) for Models II and III
β taking values in (0.5,1.0,1.5,2.0,2.5,3.0) for Model I
The standard error of this estimate is given in brackets

$\sigma^2 = 0.5$			ML	15.87		
			Model			
Picture	I		II		III	
BCIR	2.24	(0.07)	0.55	(0.04)	0.60	(0.04)
CROSS	2.66	(0.05)	1.00	(0.07)	0.98	(0.06)
TWO	2.40	(0.09)	1.11	(0.05)	0.97	(0.06)
MANY	3.94	(0.10)	2.41	(0.07)	2.27	(0.08)
VMANY	8.40	(0.13)	7.11	(0.10)	7.24	(0.10)
MRF3	6.81	(0.07)	4.92	(0.09)	4.98	(0.10)
MRF2	9.50	(0.14)	7.85	(0.09)	7.98	(0.09)

$\sigma^2 = 1.0$			ML	30.85		
			Model			
Picture	I		II		III	
BCIR	6.33	(0.16)	1.32	(0.07)	1.32	(0.05)
CROSS	6.85	(0.22)	2.07	(0.12)	2.04	(0.10)
TWO	6.88	(0.13)	2.55	(0.08)	2.41	(0.08)
MANY	8.84	(0.15)	4.52	(0.16)	4.55	(0.12)
VMANY	15.11	(0.22)	13.44	(0.13)	12.92	(0.17)
MRF3	12.13	(0.16)	8.16	(0.19)	8.09	(0.17)
MRF2	14.92	(0.20)	11.40	(0.23)	11.34	(0.22)

Each scene described in figs 1 to 7 is reconstructed using models I,II and III with various values of σ^2 and β. For models II and III we find the value of β in the set (0.25,0.5,0.75,1.0,1.25,1.50) which gives the smallest average misclassification rate. For model I we consider values of β in the set (0.5,1.0,1.5,2.0,2.5,3.0). We choose different values of β for model I as there is strong empirical evidence that the 'optimal' value of β lies in this range for the scenes considered. In Table I we display the smallest average misclassification rate for $\sigma^2 = 0.5$ and 1.0 . Similar results were obtained using different values of σ^2. Notice that ICM is superior to the ML estimate for all scenes. It is readily apparent that model I is vastly inferior to II and III for all scenes considered. Model III is marginally superior to model II in the majority of the scenes. In their study of edge penalties Brown and Silverman (1987) present an argument which supports the use of model III in preference to Model II for the majority of scenes. Recall that MRF2 and MRF3 are realisations from a Markov random field with a fixed number of black pixels. Using the 'correct' neighbourhood system appears to have little effect on the quality of the reconstruction.

As the 'optimal' β will usually be unknown we examine the average misclassification rates for model II and III for several values of β. The average percentage of misclassified pixels is presented in Tables II to VII for various values of β.

In Tables II and III we display the average percentage of misclassified pixels using models II and III for various values of β and σ^2=0.5. Similar results were obtained for other values of σ^2. There is strong evidence to suggest that the 'optimal' value of β using model III is larger than the corresponding value for model II. In figure 15 we compare the average percentage of misclassified pixels when MRF3 is reconstructed using models II and III (σ^2=0.5). We plot the average percentage of misclassified pixels using model II against β. For Model III we plot the corresponding percentage against (1/1.17)β. From this figure we see that a useful first approximation is to multiply the value of β used with model II by 1.17 when using model III. This ensures that the second term in (2.6) has the same value for both models when $\hat{a}_{ij}(x_{ij})=8$.

TABLE II

Average percentage of misclassified pixels using Model II
Standard errors in brackets
Optimal reconstruction is bold faced

$\sigma^2 = 0.5$

β	BCIR	CROSS	TWO	MANY	VMANY	MRF2	MRF3
0.25	4.53	4.75	4.96	5.91	9.78	9.86	7.74
	(0.10)	(0.11)	(0.09)	(0.13)	(0.09)	(0.14)	(0.12)
0.50	0.80	1.02	1.30	**2.41**	**7.11**	**7.85**	**4.92**
	(0.03)	(0.04)	(0.04)	(0.07)	(0.10)	(0.09)	(0.09)
0.75	**0.55**	**1.00**	**1.11**	2.48	8.04	8.44	5.13
	(0.04)	(0.07)	(0.05)	(0.09)	(0.19)	(0.10)	(0.07)
1.00	0.63	1.01	1.20	2.53	9.56	9.01	5.48
	(0.04)	(0.05)	(0.07)	(0.09)	(0.18)	(0.09)	(0.08)
1.25	0.75	1.22	1.44	3.19	11.60	9.83	6.16
	(0.05)	(0.08)	(0.10)	(0.10)	(0.25)	(0.12)	(0.12)
1.50	0.70	1.27	1.78	3.61	13.19	10.40	6.77
	(0.03)	(0.08)	(0.12)	(0.12)	(0.28)	(0.13)	(0.12)

TABLE III

Average percentage of misclassified pixels using model III
Standard errors in brackets. Optimal reconstruction is bold faced

$\sigma^2 = 0.5$

β	BCIR	CROSS	TWO	MANY	VMANY	MRF2	MRF3
0.25	6.31	6.54	6.71	7.58	11.02	11.12	9.20
	(0.12)	(0.13)	(0.12)	(0.13)	(0.09)	(0.14)	(0.12)
0.50	1.18	1.38	1.59	2.78	**7.24**	**7.98**	5.22
	(0.05)	(0.05)	(0.06)	(0.07)	(0.10)	(0.09)	(0.09)
0.75	**0.60**	1.01	1.08	2.38	7.37	8.01	**4.98**
	(0.04)	(0.07)	(0.05)	(0.09)	(0.15)	(0.11)	(0.10)
1.00	0.64	**0.98**	**0.97**	**2.27**	8.26	8.63	5.20
	(0.04)	(0.06)	(0.06)	(0.08)	(0.19)	(0.08)	(0.09)
1.25	7.11	1.08	1.25	2.81	9.76	9.25	5.72
	(0.04)	(0.06)	(0.09)	(0.08)	(0.25)	(0.11)	(0.09)
1.50	6.87	1.08	1.44	3.13	11.24	9.72	6.22
	(0.04)	(0.08)	(0.09)	(0.11)	(0.30)	(0.12)	(0.11)

In Tables IV to VII we present the analogous results for black and white pixels. These results are similar to those in Tables II and III. Notice that the 'optimal' value of β is larger for white pixels than for black in the majority of scenes. This may be due to the higher proportion of boundary pixels for black features in most scenes (see Table IX).

TABLE IV

Average percentage of black pixels classified white using model II
Standard errors in brackets.
Optimal reconstruction is bold faced

$\sigma^2=0.5$

β	BCIR	CROSS	TWO	MANY	VMANY	MRF2	MRF3
0.25	4.43	7.72	7.59	**11.91**	**16.16**	9.51	7.70
	(0.15)	(0.34)	(0.28)	(0.39)	(0.33)	(0.20)	(0.18)
0.50	0.77	**4.30**	**4.87**	12.13	18.46	**7.87**	**4.98**
	(0.06)	(0.41)	(0.21)	(0.44)	(0.48)	(0.16)	(0.12)
0.75	0.42	5.36	5.33	14.80	24.96	8.11	5.04
	(0.06)	(0.37)	(0.33)	(0.67)	(0.54)	(0.17)	(0.14)
1.00	0.37	5.21	5.94	16.98	32.37	8.86	5.36
	(0.05)	(0.34)	(0.58)	(0.65)	(0.74)	(0.16)	(0.20)
1.25	**0.30**	7.37	6.70	22.43	39.91	9.05	5.98
	(0.03)	(0.72)	(0.41)	(1.07)	(0.81)	(0.28)	(0.27)
1.50	0.36	7.23	8.12	25.37	46.71	10.34	6.70
	(0.04)	(0.81)	(0.84)	(0.92)	(1.05)	(0.30)	(0.16)

However the accurate estimation of the 'optimal' value of β is difficult in many cases as the plot of the average misclassification rate against β (see figs 8 to 14) is J-shaped in the area of interest.

TABLE V

Average percentage of black pixels classified white using model III
standard errors in brackets
Optimal reconstruction is bold faced

$\sigma^2=0.5$

β	BCIR	CROSS	TWO	MANY	VMANY	MRF2	MRF3
0.25	6.27	9.37	9.12	12.75	**16.42**	10.91	9.13
	(0.12)	(0.33)	(0.34)	(0.34)	(0.25)	(0.20)	(0.19)
0.50	1.16	4.52	4.81	**11.30**	16.88	8.08	5.33
	(0.06)	(0.37)	(0.19)	(0.37)	(0.36)	(0.21)	(0.16)
0.75	0.49	5.11	4.64	13.05	21.26	**7.74**	**4.98**
	(0.06)	(0.40)	(0.27)	(0.60)	(0.46)	(0.14)	(0.14)
1.00	0.40	**4.82**	**4.48**	14.55	26.60	8.47	5.25
	(0.06)	(0.41)	(0.39)	(0.56)	(0.70)	(0.16)	(0.20)
1.25	**0.35**	6.35	5.63	18.96	32.26	8.62	5.48
	(0.04)	(0.59)	(0.39)	(0.86)	(0.87)	(0.18)	(0.22)
1.50	0.37	6.05	6.05	21.36	39.30	9.67	6.19
	(0.04)	(0.63)	(0.58)	(0.87)	(1.03)	(0.23)	(0.17)

TABLE VI

Average percentage of white pixels classified black using model II
Standard errors in brackets
Optimal reconstruction is bold faced

$\sigma^2=0.5$

β	BCIR	CROSS	TWO	MANY	VMANY	MRF2	MRF3
0.25	4.61	4.45	4.60	5.08	7.58	10.22	7.79
	(0.11)	(0.12)	(0.10)	(0.13)	(0.16)	(0.24)	(0.19)
0.50	0.83	0.69	0.80	1.07	3.21	**7.83**	**4.85**
	(0.05)	(0.04)	(0.05)	(0.05)	(0.12)	(0.13)	(0.19)
0.75	**0.64**	**0.55**	**0.53**	0.77	2.22	8.79	5.23
	(0.04)	(0.06)	(0.06)	(0.05)	(0.11)	(0.16)	(0.15)
1.00	0.82	0.58	0.54	0.53	**1.71**	9.16	5.60
	(0.07)	(0.05)	(0.05)	(0.05)	(0.10)	(0.18)	(0.14)
1.25	1.08	0.60	0.71	**0.53**	1.85	10.63	6.35
	(0.10)	(0.07)	(0.10)	(0.05)	(0.18)	(0.28)	(0.25)
1.50	0.95	0.66	0.90	0.60	1.66	10.46	6.85
	(0.07)	(0.05)	(0.08)	(0.05)	(0.10)	(0.27)	(0.16)

TABLE VII

Average percentage of white pixels classified black using model III
Standard errors in brackets
Optimal reconstruction in bold face

$\sigma^2=0.5$

β	BCIR	CROSS	TWO	MANY	VMANY	MRF2	MRF3
0.25	6.35	6.25	6.38	6.87	9.17	11.33	9.28
	(0.14)	(0.13)	(0.12)	(0.14)	(0.15)	(0.25)	(0.18)
0.50	1.20	1.06	1.14	1.60	3.93	**7.87**	5.10
	(0.06)	(0.05)	(0.06)	(0.05)	(0.14)	(0.15)	(0.18)
0.75	**0.68**	0.59	0.59	0.91	2.60	8.28	**4.97**
	(0.05)	(0.06)	(0.05)	(0.05)	(0.08)	(0.20)	(0.14)
1.00	0.82	0.58	**0.49**	**0.57**	**1.95**	8.80	5.14
	(0.07)	(0.05)	(0.05)	(0.04)	(0.08)	(0.13)	(0.10)
1.25	0.99	**0.55**	0.64	0.57	2.02	9.89	5.96
	(0.08)	(0.06)	(0.08)	(0.07)	(0.17)	(0.22)	(0.20)
1.50	0.93	0.58	0.80	0.60	1.59	9.78	6.25
	(0.07)	(0.05)	(0.08)	(0.05)	(0.10)	(0.26)	(0.17)

The number of misclassified pixels is a crude image summary which takes no account of the spatial characteristics of the scene. To gain further insight into the differences between model II and III we use an image summary which counts the number of misclassified pixels close to the true boundary between black and white areas. A similar procedure was suggested by Owen, in the discussion of Ripley (1986).

TABLE VIII

Average percentage of misclassified boundary pixels
for MRF3. Standard errors in brackets
The optimal reconstruction in bold face
(There are 2712 boundary pixels in MRF3)

σ^2=0.5

		β					
Model		0.25	0.50	0.75	1.0	1.25	1.5
II	Boundary	16.74	**16.02** (0.23)	17.35 (0.17)	18.33 (0.21)	20.30 (0.22)	21.42 (0.25)
II	All	7.74	**4.92** (0.12)	5.13 (0.09)	5.48 (0.07)	6.16 (0.08)	6.77 (0.12)
III	Boundary	17.27	**15.98** (0.25)	16.78 (0.18)	17.51 (0.28)	19.11 (0.25)	20.19 (0.20)
III	All	9.20	5.22 (0.12)	**4.98** (0.09)	5.20 (0.10)	5.72 (0.09)	6.22 (0.09)

We reconstruct MRF3 using models II and III with σ^2=0.5. The average percentage of misclassified boundary pixels are displayed in Table VIII. In this table we call a pixels with at least one neighbour of a different colour (*in the true scene*) a boundary pixel. It is immediately apparent that the majority of misclassified pixels lie near colour boundaries when moderate values of β are used. When MRF3 is reconstructed using model III and β=0.5 there are approximately 433 misclassified boundary pixels and 89 elsewhere. There is some evidence that the optimal reconstruction of boundary pixels require a smaller value of β than the scene as a whole. This is also apparent from the example described by Owen in the discussion of Ripley (1986). There appears to be little observable difference between Model II and III using this image summary.

5. THE CHOICE OF THE SMOOTHING PARAMETER.

In this section we attempt to identify features of the underlying scene and error distribution which influence the choice of β in (2.6). We restrict attention to model II. First we examine the relationship between the 'optimal' value of β and the signal variance σ^2. In figures 8 to 14 we plot the average percentage of misclassified pixels against β for various values of σ^2. Notice that the value of β which gives the smallest average misclassification rate is approximately the same for all values of σ^2 considered. The results for VMANY (fig 12) behave atypically. In this respect the ICM algorithm differs from simple linear regularisation techniques where the 'optimal' smoothing parameter is typically proportional to the noise to signal ratio , Hall & Titterington (1986, p 336). The effect of grossly misspecifying σ^2 can be large as the example given in figure 7 of Ripley (1986) shows. However the relative stability of the misclassification rate to changes in β chose to its 'optimal' value suggests that ICM is robust to modest misspecification of σ^2. We see from figs 8 to 14 that worthwhile gains can be achieved using the 'optimal' value of β.

Figs 8-15 *here*

In the remainder of this section we examine the relationship between the 'optimal' value of β and certain features of the underlying scene. First we consider the relationship between the 'optimal' value of β and its maximum pseudo-likelihood estimate. In this approach we calculate the value of β which maximises the conditional likelihood

$$\prod_{i=1}^{m}\prod_{j=1}^{n} P(x_{ij} \mid F_{(i,j)}, \beta). \tag{5.1}$$

From Table IX we see that the pseudo-likelihood estimates of β using model II (β_{lik}) are usually greater that the value of β giving the smallest average misclassification rate. This behaviour may be due to the fact that the majority of scenes considered are untypical realisations from a MRF. For the scenes constructed by sampling from a *conditional* MRF a different pattern emerges. In this case the 'optimal' β is precisely the value of β used to construct the underlying scene (see Tables II,III and IX), **provided we use the correct model in our reconstruction.** The pseudo-likelihood approach has the disadvantage of indicating an infinite value of β for certain pixel configurations.

Next we introduce two statistics which measure the smoothness of the underlying scene.

DEFINITION : TWO IMAGE SUMMARIES

B : *Total boundary length between black and white pixels (excluding the window).*

Q_{T_i} *: The number of pixels which have at least one neighbour of a different colour using an* i^{th} *order neighbourhood.*

Notice that Switzer (1976) measures the 'smoothness' of a random function by the total arc length of its contour plot at certain levels. Applying this measure to binary random functions gives the statistic B. The image summary Q_{T_i} can be written as the difference between the statistics e_t and d_t defined in Ripley (1986, p 94) where pixels adjacent to the window are neglected. See Ripley (1977) for a discussion of image summaries and their application. Notice that $Q_{T_2} = 2B$ for many scenes (see Table IX for several examples). These statistics differ in their treatment of 'small' features. An isolated black pixel will contribute 4 to the total boundary length and 9 to Q_{T_2}.

There is strong evidence (see Table IX) to suggest that the misclassification rate for a feature is strongly influenced by the percentage of boundary pixels (as measured by Q_{T_2} or boundary length, B). This effect is indicated by the difference in the average percentage of misclassified black and white pixels. The scene BCIR appears to behave in an anomalous way. There is some evidence (see Table IX) that the value of β giving the lowest average proportion of misclassified pixels decreases as the proportion of boundary pixels (as measured by Q_{T_2} or total boundary length) increases. The value of β giving the smallest average percentage of misclassified pixels gives the strongest evidence for this relationship. There appears to be little difference in the descriptive ability of Q_{T_2} and B. In the scenes considered we see that the pseudo-likelihood estimates of β are not closely related to the smoothness measures described above.

TABLE IX

Smallest average percentage of misclassified pixels using model II
and the 'optimal' value of β vs smoothness measures.
(* pseudo likelihood estimate using model III)

$\sigma^2 = 0.5$

Picture	black	white	all	Q_{T_2}	B	β_{lik}
BCIR	0.30	0.64	0.55			
β	1.25	0.75	0.75	344	172	1.85
pixels	4300	5700	10000			
CROSS	4.30	0.55	1.00			
β	0.50	0.75	0.75	516	260	2.09
pixels	926	9074	10000			
TWO	4.87	0.53	1.11			
β	0.5	0.75	0.75	480	240	2.12
pixels	1225	8775	10000			
MANY	11.91	0.53	2.41			
β	0.25	1.25	0.5	1248	624	2.62
pixels	1216	8784	10000			
VMANY	16.16	1.71	7.11			
β	0.25	1.0	0.5	3776	1888	1.98
pixels	2560	7440				
MRF2	7.87	7.83	7.85	4109	2324	0.50
β	0.5	0.5	0.5			
pixels	5065	4935				
MRF3	4.98	4.85	4.92	2712	1453	0.63
β	0.5	0.5	0.5			(* 0.75)
pixels	5065	4935				

A useful indication of the effectiveness of a reconstruction technique can be obtained by considering its properties in reconstructing a one colour scene. In

Table X we display the average percentage of misclassified pixels when a one colour scene is reconstructed using model II. For values of β less that 0.4 appreciable errors are incurred. So for scenes with large monochrome areas we should choose $\beta \geq 0.4$.

TABLE X

Average percentage of misclassified pixels for a one colour scene
(using model II) for various values of σ^2
Standard error in brackets

β	0.2	0.25	0.3	0.35	0.4
$\sigma^2 = 0.25$	4.98 (0.03)	3.35 (0.03)	2.15 (0.02)	1.30 (0.02)	0.80 (0.01)
$\sigma^2 = 0.50$	6.6 (0.06)	3.93 (0.05)	2.26 (0.03)	1.31 (0.03)	0.75 (0.02)
$\sigma^2 = 0.75$	7.14 (0.06)	4.06 (0.06)	2.34 (0.05)	1.40 (0.04)	0.82 (0.03)
$\sigma^2 = 1.0$	7.24 (0.08)	4.25 (0.07)	2.61 (0.06)	1.59 (0.05)	1.06 (0.05)
$\sigma^2 = 1.25$	7.49 (0.08)	4.46 (0.09)	2.68 (0.07)	1.87 (0.07)	1.31 (0.06)
$\sigma^2 = 1.50$	7.68 (0.10)	4.52 (0.09)	3.03 (0.08)	2.11 (0.08)	1.52 (0.05)

To illustrate this point further consider the percentage of misclassified pixels for BCIR with $\sigma^2=0.25$. Recall that the majority of pixels in BCIR are far from the colour boundaries. In Figure XI we compare the percentage of misclassified pixels using ICM with the percentage of misclassified pixels for a one colour scene using the same model.

TABLE XI

A comparison of the average percentage of misclassified pixels of BCIR
and a monochrome scene when reconstructed using model II
Standard errors in brackets (60 realisations for mono scene)
Optimal reconstruction is bold faced

$\sigma^2=0.25$

	β					
	0.25	0.50	0.75	1.00	1.25	1.50
BCIR	4.53	0.80	**0.55**	0.65	0.75	0.70
	(0.10)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)
Monochrome	3.34	0.27	**0.02**	<0.02	<0.02	<0.02
	(0.02)	(0.01)	(0.003)	(<0.001)	(<0.001)	(<0.001)

The optimal reconstruction is obtained with $\beta=0.75$, where the percentage of misclassified pixels is 0.55. The contribution of pixels far from the colour boundary is approximately 0.02%. These result suggest that the errors incurred during the reconstruction of scenes like BCIR occur near the colour boundaries for moderate values of β (see Table VIII).

Consider a black pixel which has k white neighbours when it is updated. The probability of misclassifying this pixel during the *current iteration* can be calculated from (2.6). In Table XII we display this probability for model II with independent normally distributed noise ($\sigma^2=0.5$).

TABLE XII

The probability that a black pixel is classified white
at a *particular iteration* when it has k white neighbours

$\sigma^2=0.5$

	β		
k	0.25	0.50	1.0
8	0.98	1.00	1.00
7	0.92	1.00	1.00
6	0.76	0.98	1.00
5	0.50	0.76	0.98
4	0.16	0.16	0.16
3	0.08	0.02	0.00
2	0.02	0.00	0.00
1	0.00	0.00	0.00
0	0.00	0.00	0.00

These calculations strongly suggest that model II behaves like a simple majority vote when $\beta \geq 1.0$. Table XII can be used to estimate the 'vulnerability' of image features for various values of β. As an example consider the corner pixels (k=5) of a black rectangle. This configuration is highly vulnerable when $\beta \geq 0.5$. As ICM is an iterative procedure this calculation will not give the probability of misclassifying a given pixel. However calculations of this type are useful in visualising the effect of ICM with various values of β and neighbourhood system. Using this approach to choose β is analogous to a method suggested by Ripley (1986) with the important addition, that information is included about the noise distribution.

6. SOME DISTRIBUTIONAL PROPERTIES OF ICM

There appears to be no work in the literature on the distributional properties of the ICM estimator of (x_{ij}) or any functional of interest. The only relevant work is due to Geman and Geman (1984), who describe how to sample from the posterior distribution of (x_{ij}). In this section we examine the variance of the percentage of misclassified pixels. The number of misclassified pixels can be regarded as a functional of the scene formed by a comparison between (x_{ij}) and its reconstruction. In Table XIII we display the average percentage of misclassified pixels with its standard deviation in brackets for $\sigma^2=0.5$ and model II. The figures for the optimal reconstruction are given in bold face. Recall that ICM is a 'local' procedure. This suggests a poisson approximation for the number of misclassified pixels. The coefficient of variation of the percentage of misclassified pixels at the 'optimal' value of β appears to decrease as the misclassification rate (and complexity) increases. This is not consistent with a poisson assumption. In particular we see from Table VIII that misclassified pixels cluster near colour boundaries. The skewness $(b_1^{1/2})$ and kurtosis (b_2) of the percentage of misclassified pixels were calculated and suggest a symmetric distribution with b_2 between two and three. These are tentative conclusions as the number of realisations used in this study is small.

TABLE XIII

The standard deviation (in brackets) and the average percentage
of misclassified pixels using model II
The optimal reconstruction is bold faced

σ^2=0.5

β	BCIR	CROSS	TWO	MANY	VMANY	MRF2	MRF3
0.25	4.53	4.75	4.96	5.91	9.78	9.86	7.74
	(0.40)	(0.42)	(0.33)	(0.49)	(0.34)	(0.54)	(0.48)
0.50	0.80	1.02	1.30	**2.41**	**7.11**	**7.85**	**4.92**
	(0.10)	(0.16)	(0.17)	**(0.29)**	**(0.39)**	**(0.35)**	**(0.33)**
0.75	**0.55**	**1.00**	**1.11**	2.48	8.04	8.44	5.13
	(0.14)	**(0.26)**	**(0.18)**	(0.34)	(0.73)	(0.39)	(0.25)
1.00	0.63	1.01	1.20	2.53	9.56	9.01	5.48
	(0.16)	(0.20)	(0.25)	(0.35)	(0.71)	(0.35)	(0.32)
1.25	0.75	1.22	1.44	3.19	11.60	9.83	6.16
	(0.21)	(0.33)	(0.38)	(0.41)	(0.96)	(0.45)	(0.47)
1.50	0.70	1.27	1.78	3.61	13.19	10.40	6.77
	(0.12)	(0.33)	(0.46)	(0.46)	(1.10)	(0.51)	(0.45)

7. COMPUTATIONAL DETAILS

Pseudo-random deviates distributed uniformly on [0,1] were generated using Wichmann & Hill (1982). We take ix=27631 , iy=5627 and iz=10234. Pseudo-normal deviates with zero mean and unit variance were constructed using the Box-Muller transformation. The first step in our algorithm is to determine the maximum likelihood estimate of (x_{ij}). This colouring is used as the initial state (iteration zero) of our algorithm. Each pixel is visited in raster scan order and the colour of the $(i,j)^{th}$ pixel is updated using (2.6). The cpu time taken by our algorithm is proportional to the size of the neighbourhood system used, the number of pixels and the size of σ^2.

In Table XIV we display the average number of pixels whose colour changes during the k^{th} iteration when MRF3 is reconstructed using model II.

The average percentage of misclassified pixels is also presented. In this table one iteration is equivalent to a complete sweep of the scene (10^4 pixel visits).

Notice that the majority of changes occur during the first iteration (more changes are made as β increases). Typically only one or two pixels change colour during later iterations. This pattern is repeated for each combination of scene, σ^2 and model considered.

TABLE XIV

Average number of changes per iteration and percentage of
misclassified pixels for MRF3 (model II)
Standard errors in brackets

$\sigma^2=0.5$

	$\beta=0.25$		$\beta=0.50$		$\beta=1.0$	
k	changes	% miscl'd	changes	% miscl'd	changes	%miscl'd
1	1587	9.84	2117	6.47	2346	6.58
	(8)	(0.13)	(12)	(0.09)	(10)	(0.08)
2	206	8.18	189	5.31	153	5.93
	(5)	(0.12)	(4)	(0.08)	(3)	(0.08)
3	42	7.87	44	5.07	50.0	5.70
	(2)	(0.12)	(3)	(0.08)	(2)	(0.08)
4	12	7.78	16	4.98	21	5.58
	(1)	(0.12)	(1)	(0.08)	(1)	(0.08)
5	3	7.75	6	4.95	10	5.52
	(1)	(0.12)	(1)	(0.08)	(1)	(0.08)
6	1	7.74	3	4.93	5	5.50
	(0)	(0.12)	(0.6)	(0.08)	(1)	(0.08)
12	0	7.74	0	4.92	0	5.48
		(0.12)		(0.09)		(0.08)

This suggests the following modification of the basic algorithm:

Pixels are only updated when they are flagged as 'active'. The pixel (i,j) is 'active' when the colour of at least one of neighbours has changed during the current iteration. Pixels are visited in raster order. When a pixel's colour *changes its neighbours become active. Pixels are de-activated after they are updated.*

Using this algorithm we would visit (see Table XIV) less than nine hundred pixels on average (using a second order neighbourhood) during the third iteration. We expect the modified algorithm to converge after approximately 3 iterations in general. To obtain further gains in efficiency we might 'switch off' pixels whose colour has a low probability of being changed during the current iteration,see Ripley (1986). For example a pixel which has no neighbours of a different colour can be de-activated.

8. CONCLUSIONS

From the simulation study described in this paper we suggest the following rules of thumb for prospective users of ICM.

1. Should I use ICM ?

Our empirical results suggest that the misclassification rate of a feature increases with the proportion of boundary pixels (see Table IX and compare the misclassification rate for black and white pixels). Typically small feature will be 'erased'. If the aim of an analysis is to find small features then a technique based on masks will probably be preferable to ICM. However it is apparent from Table I that substantial gains over the maximum likelihood estimate, can be achieved by smoothing.

2. Which model should I use?

We suggest that model III should be used in the absence of specific knowledge about the uncorrupted scene. If we know that the underlying scene is non-homogeneous we can exploit this by using a hierarchical model, see Derin & Elliot (1987) or Woods,Dravida & Mediavilla (1987).

3. What value of β should I use?

This is a difficult question to answer in the absence of any information about the underlying scene. The examples considered in this paper suggest that useful gains can be achieved using the 'optimal' value of β rather than a portmanteau value of , say β=1.5. We distinguish between two cases. In the first we assume that the underlying scene is a 'typical' realisation from a MRF. Then the 'optimal' reconstruction is obtained using the neighbourhood system and value of β specified by the underlying MRF. When the underlying scene cannot be regarded as a 'typical' realisation from a MRF we suggest the used of smoothness measures such as the total boundary length in the choice of the 'optimal' value of β. In both cases we see that the 'optimal' value of β does not depend on σ^2. From figs 8 to 14 we see that there is some leeway in choosing the 'optimal' value of β.

4. Is the ICM estimate difficult to calculate?

From the discussions in section 7 we see that a single reconstruction of a binary 10^4 pixel scene can be computed simply. The calculations appear well suited to parallel implementation. The scene VMANY with σ^2=0.5 was reconstructed in around 39 seconds (using model II with β=0.5) on a SUN-3 Work Station with a floating point accelerator.

Acknowledgement

I would like to thank Dr. C. Jennison and Professor B.W. Silverman for their helpful comments during the course of this work. The financial support of the European Research Office of the U.S. Army is gratefully acknowledged.

References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with Discussion). *J.Roy. Statist. Soc. Ser. B* **36**, 192-236.

Besag,J. (1986). On the statistical analysis of dirty pictures (with Discussion). *J. Roy. Statist. Soc. Ser. B* **48** , 259 - 302.

Bovik, A.C. , Huang, T.S. & Munson, D.C. (1987). The effect of median filtering on edge estimation and detection. *I.E.E.E. Trans. Pattern Anal. Machine Intell.* **9** , 181 - 194.

Brown,T.C. & Silverman, B.W. (1987). Edge-process models for regular and irregular pixels. *Technical report no. 267. Stanford University.*

Cross, G.R. & Jain, A.K. (1983). Markov random field texture models. *I.E.E.E. Trans. Pattern Anal. Machine Intell.* **5**, 25 - 39.

Derin,H. & Elliot, H. (1987). Modeling and segmentation of noisy and textured images using Gibbs random fields. *I.E.E.E Trans. Pattern Anal. Machine Anal. Intell.* **6** , 39 - 55.

Enting, I.G. & Welberry, T.R. (1978). Connections between Ising models and various probability distributions. *Suppl. Adv. in. Appl. Probab.* **10**, 65 - 72.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *I.E.E.E. Trans. Pattern Anal. Machine Intell.* **6**, 721 - 741.

Hall, P. & Titterington, D.M. (1986). On some smoothing techniques used in image restoration. *J.R. Statist. Soc. Ser B* **48** , 330 - 343.

Kashyap, R.L. & Chelappa, R. (1983). Estimation and choice of neighbors in spatial interaction models of images. *I.E.E.E. Trans. Inform. Theory* **29** , 60 - 72.

Kashyap, R.L. & Lapsa, P.M. (1984). Synthesis and estimation of random fields using long correlation models. *I.E.E.E. Trans. Pattern Anal. Machine Intell.* **6** , 800 - 809.

Van Laarhoven, P.J.M & Aaris, E.H.L. (1987). *Simulated Annealing: Theory and Applications*. D.Reidel Publishing Company.

Marroquin, J. , Mitter, S. and Poggio, T. (1987). Probabilistic solution of ill-posed problems in computational vision. *J. Amer. Statist. Assoc.* **82** , 76 - 89.

Pickard, D.K. (1987). Inference for discrete markov fields: The simplest non-trivial case. *J. Amer. Statist. Assoc.* **82** , 90 - 96.

Ripley, B.D. (1977). Modelling spatial patterns. *J. Roy. Statist. Soc. Ser. B* **39** , 172 - 212.

Ripley, B.D. (1984). Spatial Statistics.Developments 1980-3. *Internat. Statist. Rev.* **53** , 141-150.

Ripley, B.D. (1986). Statistics,images and pattern recognition. *Canad. J. Statist.* **14** , 83 - 111.

Rosenfeld, A. and Kak, A.C. (1982). *Digital Picture Processing.* Academic Press,Orlando.

Serra, J. (1982). *Image Analysis and Mathematical Morphology.* Academic Press, London.

O'Sullivan, F. (1986). A Statistical perspective on ill-posed problems (with discussion). *Statist. Sci.* **1** , 502 - 527.

Switzer, P. (1976). Geometrical measures of the smoothness of random functions. *J. Appl. Probab.* **13** , 86 - 95.

Suomela, P. (1976). Construction of nearest neighbour systems. *Annales Academiae Scientiarum Fennicae. Series A. Mathematica Dissertationes* **10.**

Titterington, D.M. (1985). Common structure of smoothing techniques in statistics. *Internat. Statist. Rev.* **53** , 141-170.

Wichmann, B.A. & Hill, I.D. (1982). Algorithm AS 183. An efficient and portable pseudo-random number generator. *J. Roy. Statist. Soc. Ser. C* **31** ,188-190.

Wolberg, G. & Pavlidis, T. (1985). Restoration of binary images using stochastic relaxation with annealing. *Pattern Recognition Letters.* **3** , 375 - 388.

Woods,J.W. , Dravida, S. & Mediavilla, R. (1987). Image estimation using doubly stochastic Gaussian random field models. *I.E.E.E. Trans. Pattern Anal. Machine Intell.* **9** , 245 - 253.

CAPTIONS FOR FIGURES 1 TO 15

FIGURE 1 BCIR : Circle centred at (30,30) with radius 40. The origin is at the bottom left hand corner of the window which has dimensions (0,100)x(0,100).

FIGURE 2 CROSS : Two rectangles with corners at {(10,40),(60,20),(70,30),(20,50)} and {(25,20),(30,15),(55,50),(50,55)}

FIGURE 3 TWO : Two rectangles with corners at {(10,40) , (60,40) , (60,50) , (10,50)} and {(20,55) , (65,55) , (65,60) , (20,60)}

FIGURE 4 MANY : Eight circles of radius 6 centred at , (25,20) , (45,20) , (65,20) , (80,20) , (25,80) , (45,80) , (65,80) , (85,80) and ten circles of radius 3 centred at (20,40) , (35,40) , (50,40) , (65,40) , (80,40) , (20,60) , (35,60) , (50,60) , (65,60) , (80,60).

FIGURE 5 VMANY : Eighty circles with radius 3 and centres at $(5+10j, 10k-7)$ for $j=1,\ldots,8$ and $k=1,\ldots,10$.

FIGURE 6 MRF2 : A synthetic realisation from the MRF specified in MODEL II with $\beta=0.5$. This scene was constructed using an algorithm given in Cross and Jain (1983).

FIGURE 7 MRF3 : A synthetic realisation from the MRF specified in Model III with $\beta=0.75$. This scene was constructed using the algorithm given in Cross and Jain (1983).

FIGURE 8 A plot of the average percentage of misclassified pixels against β and σ^2 when BCIR is reconstructed using MODEL II

FIGURE 9 A plot of the average percentage of misclassified pixels against β and σ when CROSS is reconstructed using MODEL II

FIGURE 10 A plot of the average percentage of misclassified pixels against β and σ when TWO is reconstructed using MODEL II

FIGURE 11 A plot of the average percentage of misclassified pixels against β and σ when MANY is reconstructed using MODEL II

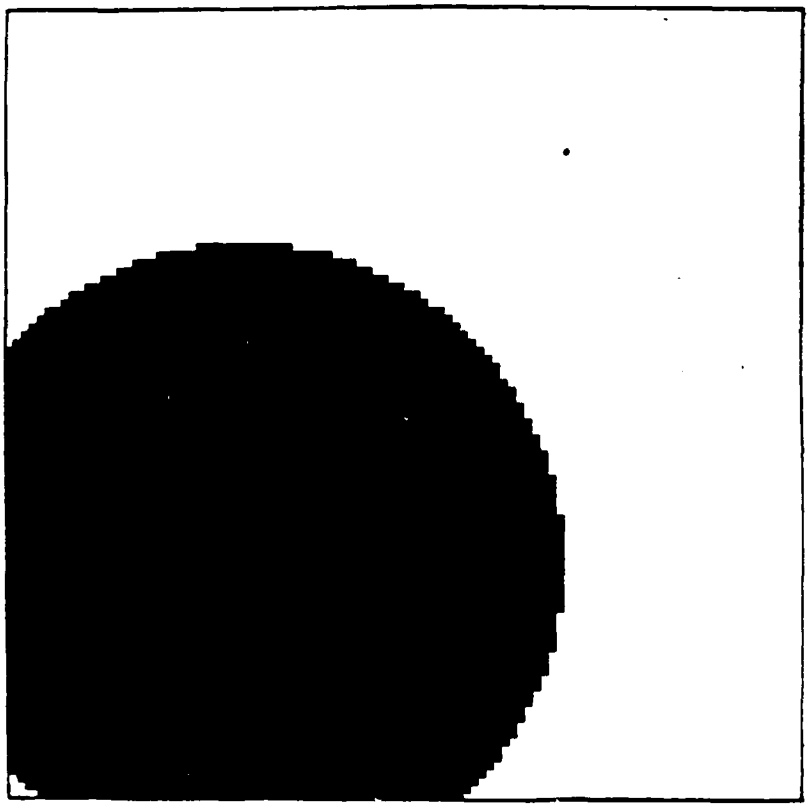
FIGURE 12 A plot of the average percentage of misclassified pixels against β and σ when VMANY is reconstructed using MODEL II

FIGURE 13 A plot of the average percentage of misclassified pixels against β and σ when MRF2 is reconstructed using MODEL II

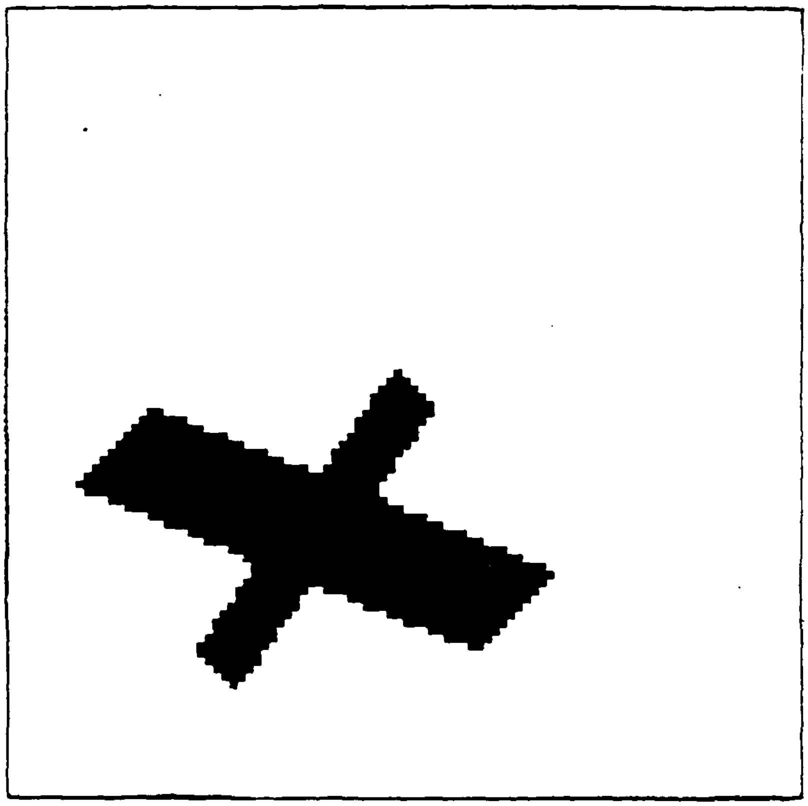
FIGURE 14 A plot of the average percentage of misclassified pixels against β and σ when MRF3 is reconstructed using MODEL II

FIGURE 15 A plot of the average percentage of misclassified pixels against β for model II and (1/1.117)β for model III

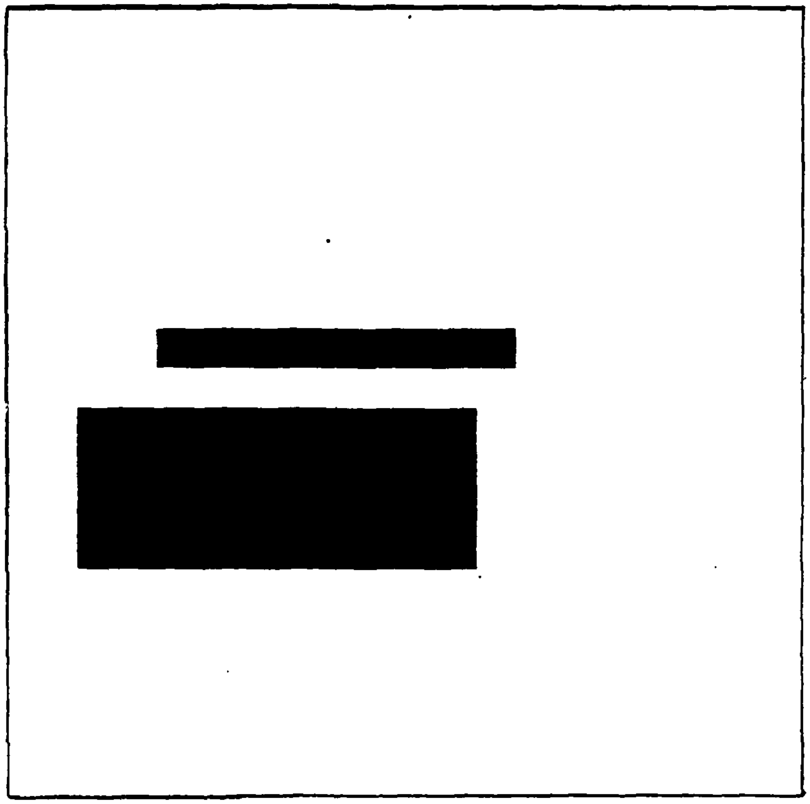
R.H.Glendinning : FIGURE 1



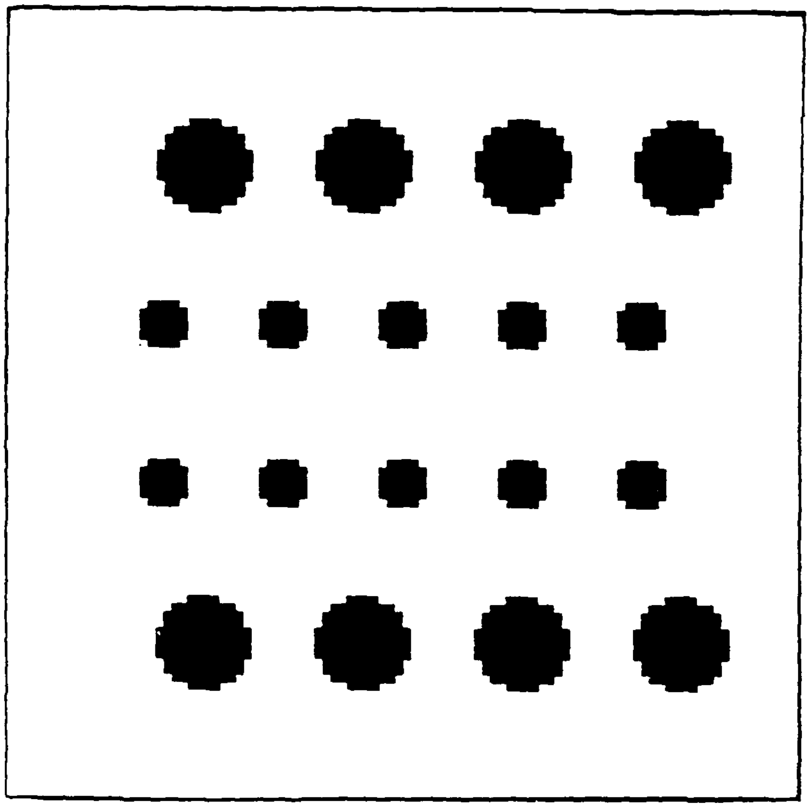
R.H.Glendinning : FIGURE 2



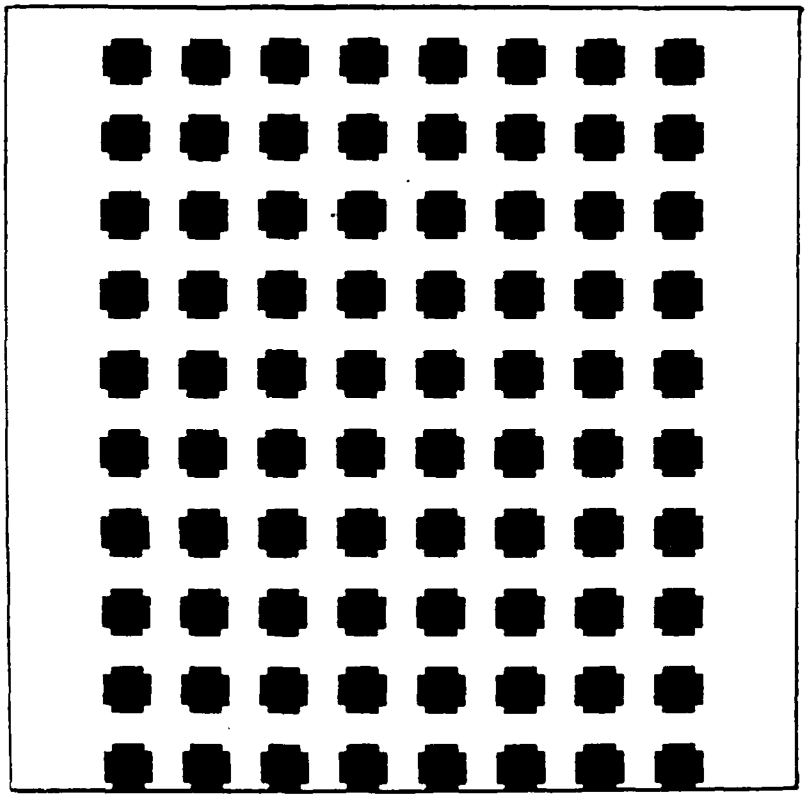
R.H.Glendinning : FIGURE 3



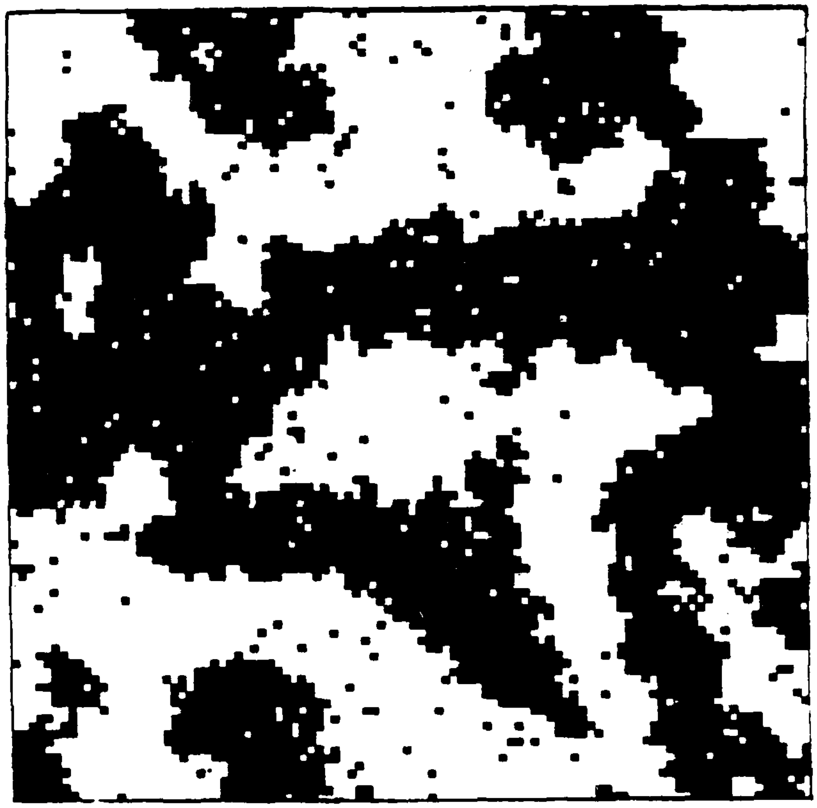
R.H.Glendinning : FIGURE 4



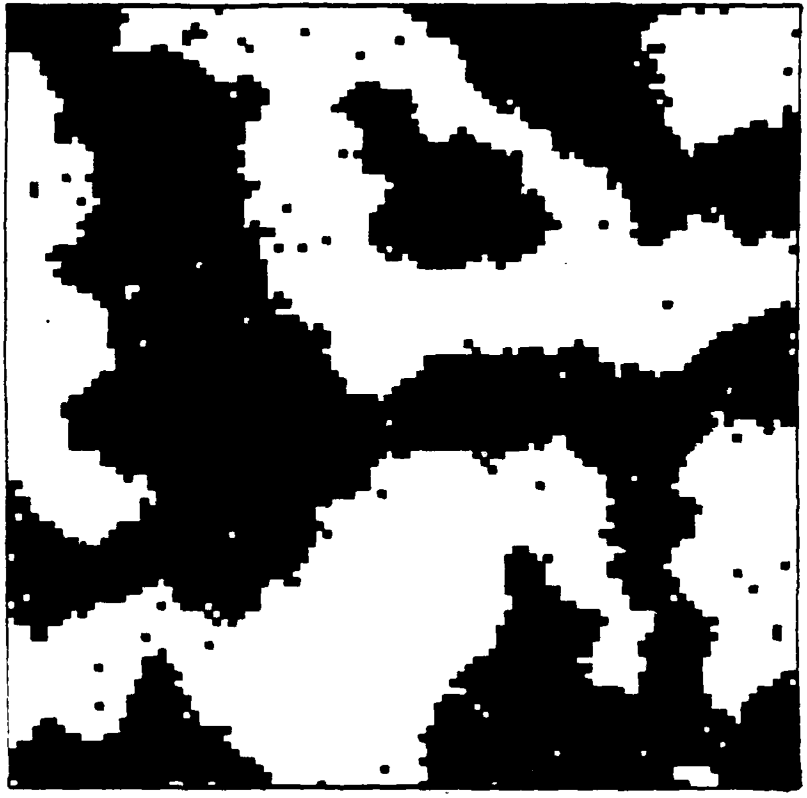
R.H.Glendinning : FIGURE 5



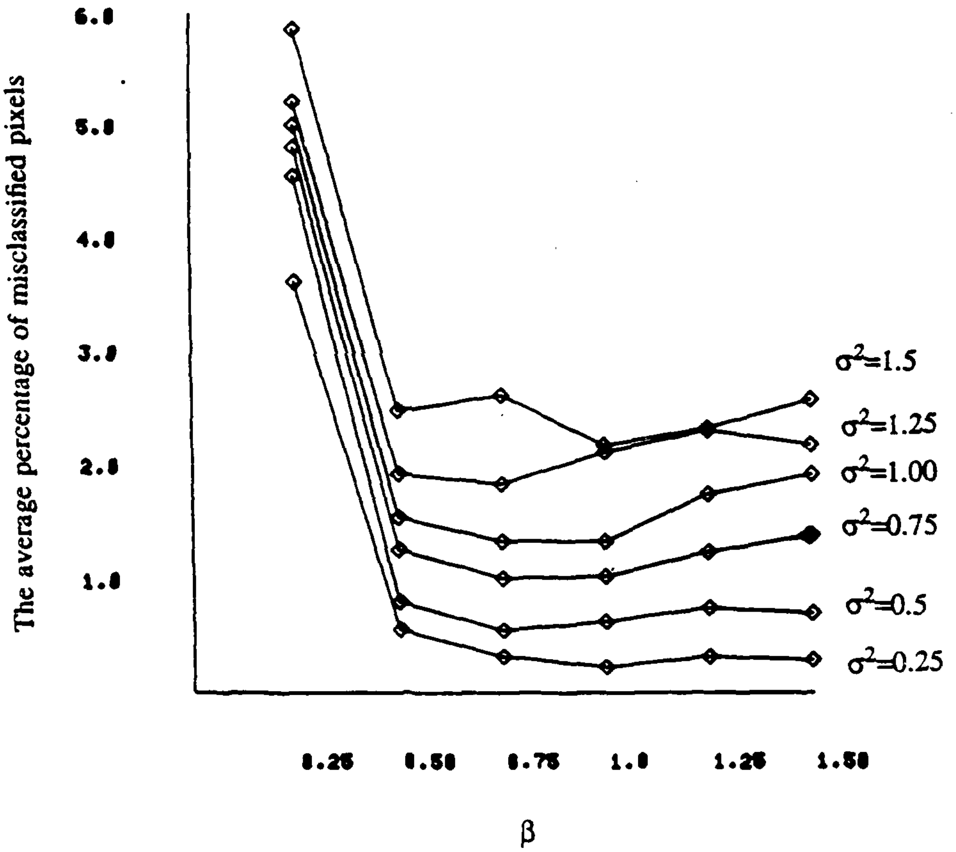
R.H.Glendinning : FIGURE 6



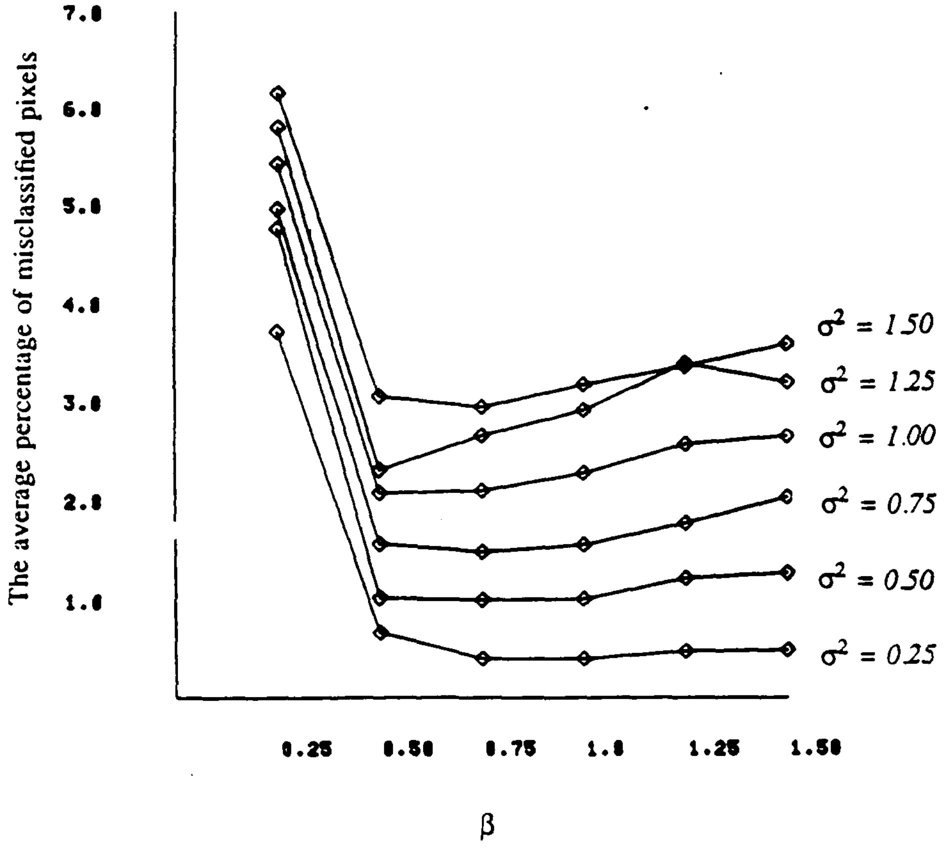
R.H.Glendinning : FIGURE 7



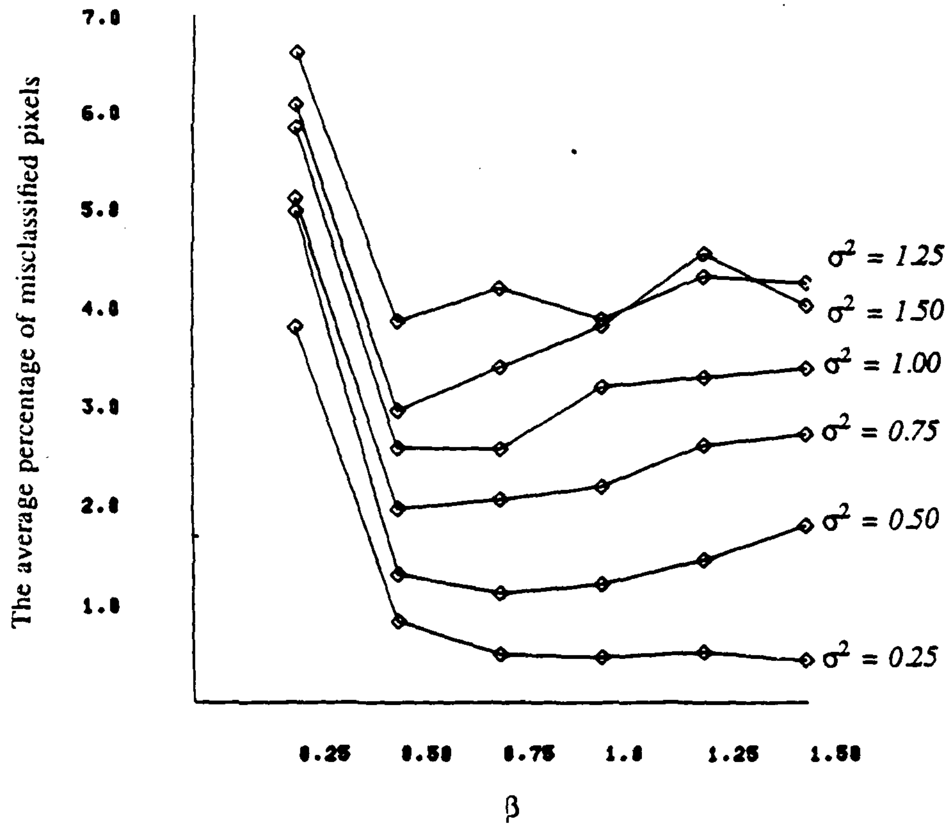
R.H.Glendinning : FIGURE 8



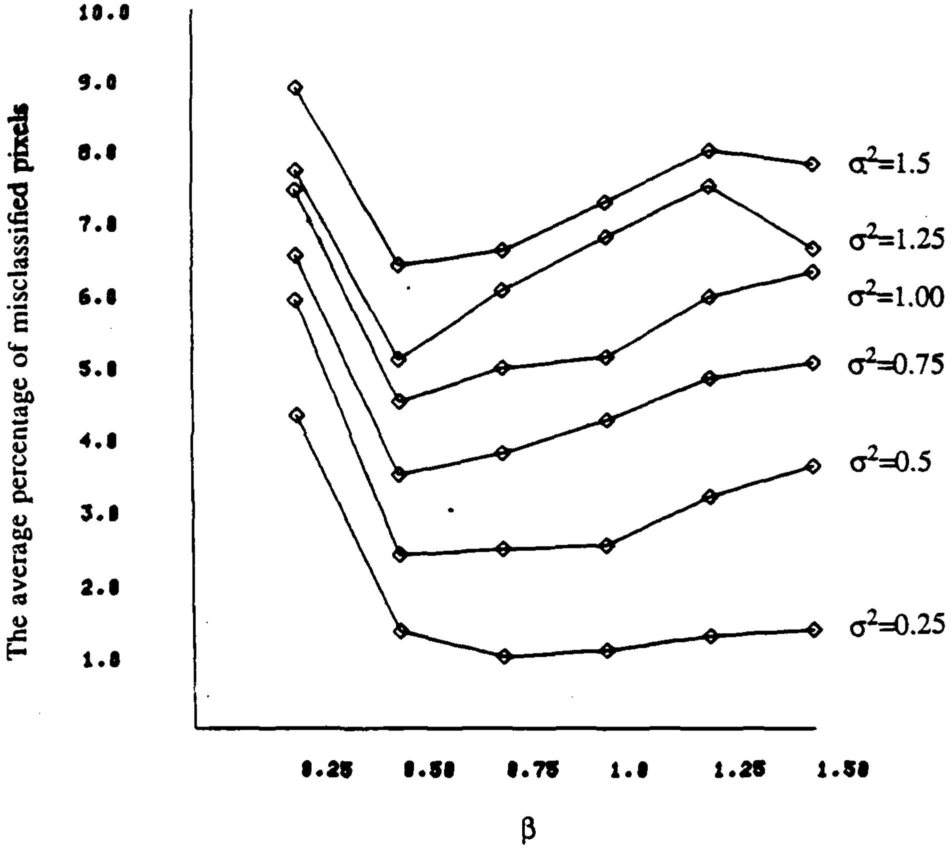
R.H.Glendinning : FIGURE 9



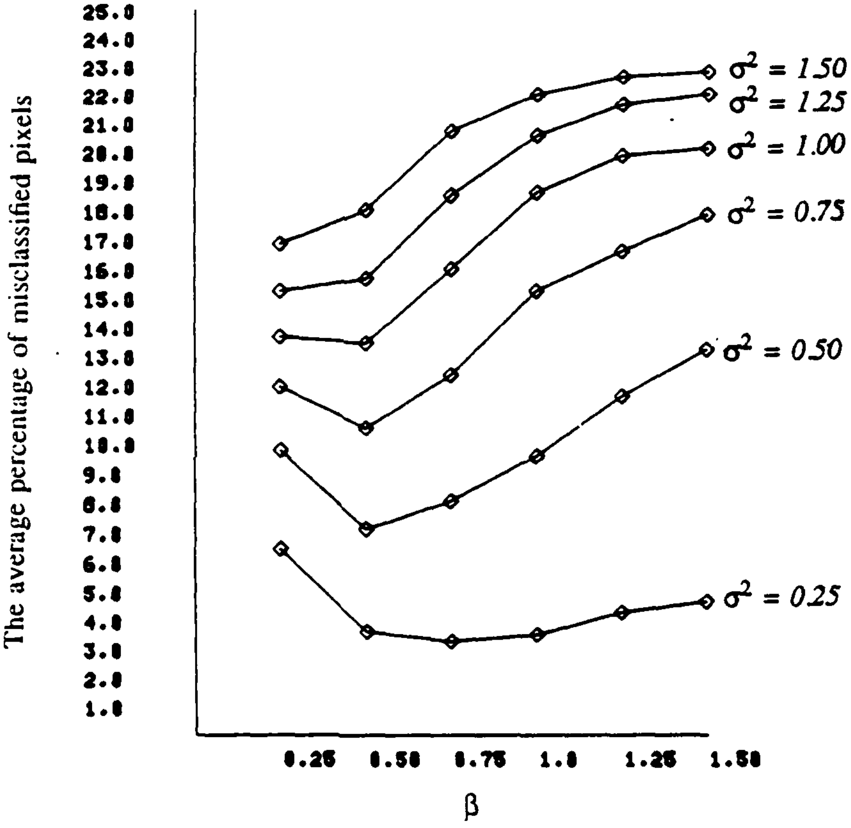
R.H.Glendinning : FIGURE 10



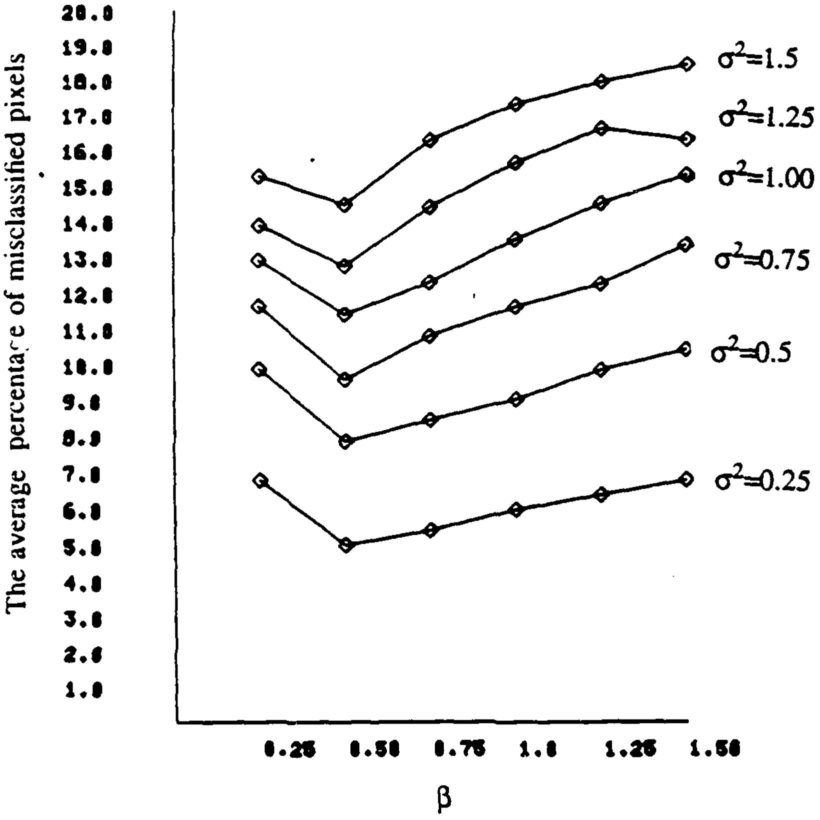
R.H.Glendinning : FIGURE 11



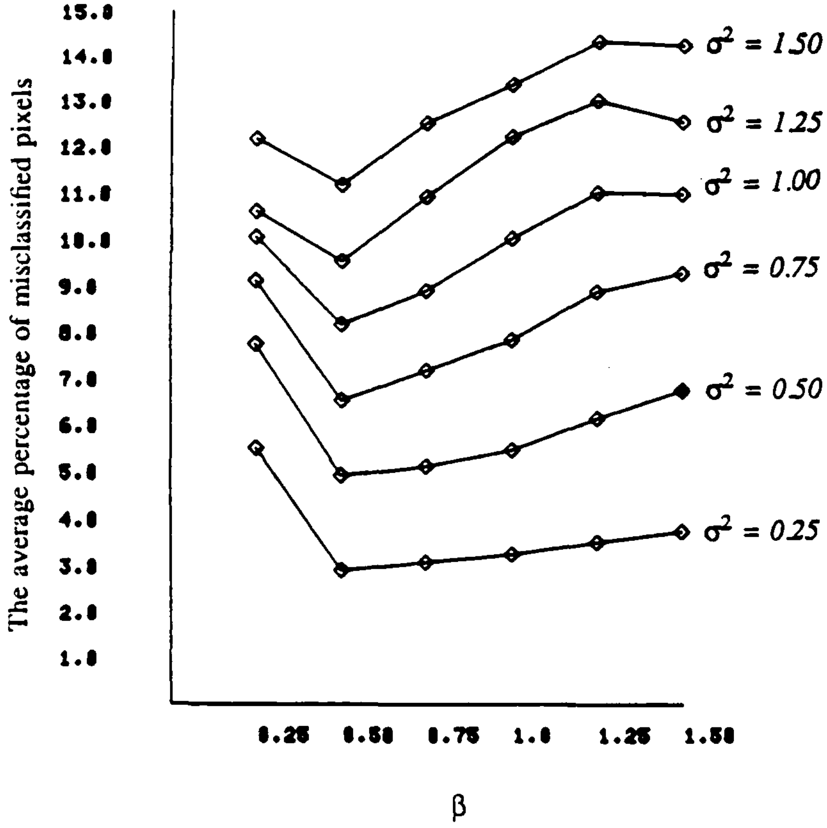
R.H.Glendinning : FIGURE 12



R.H.Glendinning : FIGURE 13



R.H.Glendinning : FIGURE 14



Appendix 9

Speed of estimation in positron emission tomography.

by

I. M. Johnstone & B. W. Silverman

Speed of Estimation in Positron Emission Tomography

by Iain Johnstone and B. W. Silverman

Stanford University and University of Bath

Several algorithms for image reconstruction in positron emission tomography (PET) have been described in the medical and statistical literature. We study a continuous idealisation of the PET reconstruction problem, considered as an example of bivariate density estimation based on indirect observations. Given a large sample of indirect observations, we consider the size of the equivalent sample of observations, whose original exact positions would allow equally accurate estimation of the image of interest. Both for indirect and for direct observations, we establish exact minimax rates of convergence of estimation, for all possible estimators, over suitable smoothness classes of functions. For indirect data and (in practice unobservable) direct data, the rates for mean integrated square error are $n^{-p/(p+2)}$ and $(n/\log n)^{-p/(p+1)}$ respectively, for densities in a class corresponding to bounded square-integrable pth derivatives. We obtain numerical values for equivalent sample sizes for minimax linear estimators using a slightly modified error criterion. Modifications of the model to incorporate attenuation and the third dimension effect do not affect the minimax rates.

Iain M. Johnstone
Department of Statistics
Stanford University
Stanford CA 94305
U. S. A.

B. W. Silverman
School of Mathematical Sciences
University of Bath
Bath BA2 7AY
U. K.

AMS 1980 *subject classifications.* Primary 62G05; secondary 62C20, 65R20, 65U05.
Key words and phrases. Attenuation, biased sampling, density estimation, Fano's lemma, image analysis, information theory, linear inverse problems, minimax, orthogonal series, Radon transform, Sobolev norms, singular value decomposition, tomography.

5 December 1988 (revised 12.3.90)

1 Introduction

Tomography is a non-invasive technique for reconstructing the internal structure of an object of interest, often in a medical context. Positron emission tomography (PET) deals with the estimation of the amount and location of a radioactively labeled metabolite on the basis of particle decays indirectly observed outside the body. Emission tomography in general, and PET in particular, has been the subject of considerable recent research in nuclear medicine, and has attracted the interest of statisticians as an example of a reconstruction problem involving incomplete and noisy data.

The formulation of the PET problem we shall consider is basically that given by Shepp and Vardi (1982) and Vardi, Shepp and Kaufman (1985). Following their convention we shall consider a particular PET experiment, where the brain is scanned by counting radioactive emissions from tagged glucose. The distribution of glucose within the brain corresponds to the glucose uptake mechanism, and so a map of the glucose distribution within the brain gives an indication of the pattern of the brain's metabolic activity. In the idealisation we shall consider, following Vardi et al. (1985), the radioactive tagging of the glucose gives rise to emissions of positrons distributed as a Poisson process in space and time; the spatial intensity of emissions is the same as the distribution of glucose. Each positron that is emitted annihilates with a nearby electron, and yields two photons that fly off in opposite directions along a line with uniformly distributed orientation. One or more rings of sensors placed around the patient's head make it possible to detect the photon pairs and hence, for each emission that is detected, to give a line on which the point of emission must have occurred. However, for equipment of the kind discussed here, it is not possible to detect the position of the emission on the line.

The PET problem is just one of a large number of statistical problems involving *indirect observations* of the phenomenon of interest; in our case the observations are indirect in that the emissions themselves are not observed directly. Such problems arise, for example, in geophysics, in stereology and wherever linear deconvolution with known filter is required. Our aim in the present paper is not just to study the PET problem but also to develop theory that can be applied in many other contexts.

In a typical PET scan, a large number, perhaps one to ten million, radioactive emissions are recorded, and the image of interest, a slice through the patient's brain or body, is reconstructed in some way from this apparently vast data set. But is ten million observations really a large sample in this kind of context? One way of gaining some insight into the problem is to think in terms of *equivalent sample sizes.* We make some smoothness assumptions about the image of interest, and then ask how accurately it could possibly be reconstructed given a particular indirect sample. The equivalent sample size would be the number of emissions whose *original* positions could yield an equally accurate estimate. The equivalent sample size gives, in terms more attuned to usual statistical intuition, a quantification of the information actually available from our sample of ten million indirectly observed emissions, and hence gives an idea of how much is lost by the indirect nature of the observation process.

In Section 2 below, we formulate the reconstruction problem as an example of nonparametric bivariate density estimation based on indirect data, in fact an example of a linear inverse problem in a function space. The function we estimate is the intensity function of emissions in the slice through the brain. A key feature of our treatment is the explicit singular value decomposition of the transform linking the unknown density with that of the observed data. The main conclusions of the paper are summarised in Section 3. In particular we give in Section 3 a table of explicit equivalent sample sizes, admittedly for our mathematical idealisation of the PET problem. In Section 4 we confine attention to linear estimators, and to intensities falling in a suitable smoothness class of functions. We find the exact minimax rates of consistency, that is the rate for the least favourable density and the best linear estimator. We then show, in Section 5, that these rates cannot be improved by extending consideration to all possible estimators, linear or non-linear. Thus we do not consider particular iterative non-linear algorithms proposed elsewhere for practical use, but instead we establish the best possible performance achievable by any estimator.

Section 6 of the paper considers modifications of our mathematical idealisation in order to take account of attenuation of the emitted photons and of the three dimensional nature of the problem. Our broad conclusions carry over when these effects are incorporated. In Section 7, we extend our results to some error measures based on the derivatives as well as the values of the images and their reconstructions. Finally, in Section 8, we make some concluding remarks, and mention some possible issues for future research.

A subsidiary objective of the paper is to illustrate, in a relatively simple and concrete setting, the general approach to deriving lower bounds to estimation risk developed by Le Cam (1985, for example), Ibragimov and Hasminskii (1981), and Birgé (1983). This method relates the best possible speed of estimation (in a given "global" metric) to the metric entropy structure of the parameter space. We need a minor modification to handle the present indirect estimation setting, introducing a form of "modulus of continuity" of the inverse transform. This material is presented mainly in Section 5.

There is a substantial literature on practical algorithms for reconstruction in the PET setting. An extensive survey covering the period up to 1979 is given by Budinger, Gullberg and Heusman (1979); this includes adaptation of methods from X-ray transmission tomography and the orthogonal series method of Marr (1974). Maximum likelihood methods were proposed by Rockmore and Macovski (1977); they were implemented via the EM algorithm by Shepp and Vardi (1982) (see also Vardi, Shepp and Kaufman, 1985) and modified in various ways to incorporate smoothing by Geman and McClure (1985) and Silverman et al. (1988). Some practical illustration of the orthogonal series method introduced in the present paper is given by Jones and Silverman (1989). A recent survey of algorithms is given by Tanaka (1987). Papers considering noise limitations in X-ray and transmission tomography include Chesler et al. (1977) and Tretiak (1978, 1979). The focus of these papers differs from ours in that they consider estimation of a fixed finite number of real-valued functions of a particular unknown intensity, using discrepancies based on variance rather than mean square error.

2. Mathematical model and technical preliminaries

2.1 An idealised problem and the Radon transform

In our idealised version of the PET problem, the ring of detectors defines a slice of the patient's head, and the reconstruction aims to display a picture of the glucose density within that slice. Emissions that give rise to photon pairs, one or both of which miss the detector ring, will go unrecorded. Bearing this in mind, we shall regard the slice as a plane and consider an essentially two-dimensional problem where (see Fig. 2.1) emissions take place in the plane according to some density within a detector circle taken to be the unit circle in the plane. An emission at P gives rise to a photon pair whose directions of flight lie in the plane along a line l through P with random, uniformly distributed, orientation. The finite size of the detectors is ignored and it is assumed that the points Q and R of the intersection of l with the detector circle are observed exactly.

Fig 2.1 near here

Give the name *detector space* to the space D of all possible unordered pairs QR of points on the detector circle, and call *brain space* the original disc B in the plane enclosed by the detector ring. Assume that coordinates are chosen so that B is the unit disc. Brain space is parametrised either by cartesian or standard polar coordinates. To parametrise detector space, let s be the length of the perpendicular from the origin to the detected line QR as in Figure 2.2, and φ the orientation of this perpendicular. Thus D is $\{(s,\varphi):0\leq s\leq 1,\ 0\leq\varphi\leq 2\pi\}$.

Fig 2.2 near here

We now define dominating measures on brain space and on detector space. Define a measure μ on brain space to be $\pi^{-1}\times$ lebesgue measure, so that $d\mu(r,\theta) = \pi^{-1} r dr d\theta$ for $0\leq r\leq 1$ and $0\leq\theta<2\pi$ if polar coordinates are used, and $d\mu(x_1,x_2) = \pi^{-1} dx_1 dx_2$ for $\|x\|\leq 1$ in Cartesian coordinates. On detector space, define a measure λ by $d\lambda(s,\varphi) = 2\pi^{-2}(1-s^2)^{\frac{1}{2}}ds\,d\varphi$. Both μ and λ integrate to 1.

Suppose an emission takes place at a point distributed with probability density $f(x_1,x_2)$ with respect to μ in brain space. Let $g = Pf$ be the probability density in detector space, with respect to λ, of the corresponding detection of a pair of photons, so that the mapping P maps the actual density of emissions to the corresponding observable density in detector space. We shall show below that Pf is given by

$$Pf(s,\varphi) = \tfrac{1}{2}(1-s^2)^{-\frac{1}{2}} \int_{-\sqrt{(1-s^2)}}^{\sqrt{(1-s^2)}} f(s\cos\varphi - t\sin\varphi,\ s\sin\varphi + t\cos\varphi)dt \tag{2.1}$$

The integral in (2.1) is the so-called *Radon transform* (see Marr, 1974; Deans, 1973) of the density f, namely the line integral of f along the line l with co-ordinates (s,φ) in detector space. Since the length of the segment QR is $2(1-s^2)^{\frac{1}{2}}$, it can be seen at once that $Pf(s,\varphi)$ is the average of f over the part of l that intersects the detector disc $\|x\|\leq 1$. If f is the uniform density in brain space, so that $f(x_1,x_2) = 1$ for all $\|x\|\leq 1$, then we will have $Pf(s,\varphi) = 1$ for all s and φ. Thus the probability measure λ in detector space is the detector space distribution corresponding to the uniform measure μ in brain space.

Fig 2.3 near here

It remains to verify (2.1). Suppose an emission takes place at (X_1,X_2) and that the corresponding photon pair has trajectory at angle Ψ as shown in Figure 2.3; taking $0\leq\Psi\leq\pi$ for definiteness, the joint probability density with respect to $dx_1dx_2d\psi$ on $\|x\|\leq 1$ and $0\leq\psi\leq\pi$ is given by

$$f_{X_1,X_2,\Psi}(x_1,x_2,\psi) = \pi^{-2}f(x_1,x_2)$$

using the definition of μ and the fact that Ψ is independent of X_1 and X_2. Now change variables by setting

$$S = |X_1\cos\Psi + X_2\sin\Psi|$$

$$\Phi = \begin{cases} \Psi & \text{if } X_1\cos\Psi + X_2\sin\Psi \geq 0 \\ \Psi+\pi & \text{otherwise} \end{cases}$$

$$T = -X_1\sin\Psi + X_2\cos\Psi\ ;$$

the variables (S,Φ) are the coordinates of the detected photon pair. After making the transformation, which has unit Jacobian, and integrating out the unobserved variable T, we obtain the joint density with respect to $ds\,d\varphi$

$$f_{S,\Phi}(s,\varphi) = \pi^{-2}\int_{-\sqrt{(1-s^2)}}^{\sqrt{(1-s^2)}} f(s\cos\varphi - t\sin\varphi, s\sin\varphi + t\cos\varphi)dt.$$

The density (2.1) with respect to λ follows at once from the definition of λ.

2.2 Estimators and loss functions

In this section, we define various classes of estimator of f that we shall be considering, as well as two measures of the accuracy of estimation of f. The proofs of the three propositions stated in this section are given in the Appendix.

Two particular classes of estimator are of obvious interest. Let $\mathcal{T}_D(n)$ be the class of all possible estimators based on a sample of n independent *direct* observations in brain space from the density f. Let $\mathcal{T}_I(n)$ be the class of all estimators of f based on a sample of n indirect observations, i.e. observations in detector space drawn from the density Pf. It will also be important in some of our work to concentrate attention on those estimators that are *linear estimators*. An estimator $\hat{f}$ based on observations $Z_1,\ldots,Z_n$ is called *linear* if there exists a weight function $w(x,z)$ such that $\int w(x,z)d\mu(x)=1$ for all z in the space of the observations, and

$$\hat{f}(x) = n^{-1}\sum_{i=1}^{n} w(x,Z_i) \quad \text{for all } x \text{ in } B. \tag{2.2}$$

Let $\mathcal{T}_{LD}(n)$ be the set of all linear estimators based on a direct sample of size n subject to the additional condition $\int\int w(x,x')^2 d\mu(x)d\mu(x')<\infty$, and let $\mathcal{T}_{LI}(n)$ be the set of all linear estimators of f based on an indirect sample of size n for which $\int\int w(x,y)^2 d\mu(x)d\lambda(y)<\infty$. The additional square integrability conditions are mild; they ensure that $\hat{f}$ has finite mean integrated square error if f is bounded.

One natural measure of the accuracy of an estimator $\hat{f}$ is the mean integrated square error $M(\hat{f};f) = E_f\int_B(\hat{f}-f)^2 d\mu$. By standard calculations,

$$M(\hat{f};f) = \int[\operatorname{var}_f\hat{f}(x) + \{E_f\hat{f}(x)-f(x)\}^2]d\mu(x)\ , \tag{2.3}$$

where the suffix f indicates that the mean and variance are calculated for data drawn from f in the direct case and Pf in the indirect case. We define the *surrogate mean integrated square error* $M^*(\hat{f};f)$ by replacing the variance term in (2.3) by the corresponding term calculated for the uniform density on brain space

$$M^*(\hat{f};f) = \int[\operatorname{var}_1\hat{f}(x) + \{E_f\hat{f}(x)-f(x)\}^2]d\mu(x)\ , \tag{2.4}$$

where var_1 denotes a variance calculated with respect to data drawn from the probability measure μ in the direct case and λ in the indirect case. An important relation between the surrogate and the true mean integrated square error for *linear* estimators is given by the following lemma.

Proposition 2.1 *Suppose that f is bounded above and below away from zero. Then, for all $\hat{f}$ in $\mathcal{T}_{LD}(n)$ or in $\mathcal{T}_{LI}(n)$*

$$\inf_B f(x) \le M(\hat{f};f)/M^*(\hat{f};f) \le \sup_B f(x).$$

2.3 The singular value decomposition of the Radon transform

The singular value decomposition (SVD) of the normalised Radon transform P defined in (2.1) is the key to our study of the loss of information about f due to indirect observation. To establish notation, let H and K be Hilbert spaces and $P : H \to K$ a bounded linear operator. Under suitable conditions, there exist orthonormal sets of functions $\{\varphi_\nu\}$ in H and $\{\psi_\nu\}$ in K, and positive real numbers $\{b_\nu\}$, the singular values of P, such that the $\{\varphi_\nu\}$ span the orthogonal complement of the kernel of P, the $\{\psi_\nu\}$ span the range of P, and $P\varphi_\nu = b_\nu \psi_\nu$ for all ν.

Thus P is diagonal in the bases $\{\varphi_\nu\}$ and $\{\psi_\nu\}$. If a singular value b_ν is small, then noise encountered in estimation of the component of f along φ_ν will be amplified by a factor of b_ν^{-1}. Some form of regularization method (Tikhonov and Arsenin, 1977) is needed to deal with this instability, and one such method, based on tapered orthogonal series, will be exploited in Section 4 below.

In our PET model, H is the space $L^2(B,\mu)$ of functions on brain space which are square-integrable with respect to the dominating measure μ. Correspondingly, K is the space $L^2(D,\lambda)$ of detector-space functions square-integrable relative to λ. Suppose that $X = (X_1, X_2)$ is drawn at random (according to μ) from brain space B. If a direction φ is specified by $u_\varphi = (\cos\varphi, \sin\varphi)$, then

$$Pf(s,\varphi) = E\{f(X) \mid u_\varphi \cdot X = s\}$$

From this representation it follows at once that P is a bounded operator from $L^2(B,\mu)$ to $L^2(D,\lambda)$ with norm 1 and, by arguments involving characteristic functions, it is one-to-one.

The SVD of the Radon transform in this specific setting appears to have been first derived by workers in optics and tomography; we now review its properties drawing material from Born and Wolf (1975, Chapter 9.2.1 and Appendix VII), Marr (1974) and Deans (1983, Section 7.6). Since the underlying spaces are two dimensional, we need double indices, specifically $\nu \in N = \{(l,m) : m = 0,1,2,\ldots;\ l = m, m-2, \ldots, -m\}$. In brain space, an orthonormal basis for $L^2(B,\mu)$ is given by

$$\varphi_\nu(r,\theta) = (m+1)^{\frac{1}{2}} Z_m^{|l|}(r) e^{il\theta} \qquad \nu = (l,m) \in N,\ (r,\theta) \in B, \tag{2.5}$$

where Z_m^k denotes the *Zernike polynomial* of degree m and order k. Zernike polynomials satisfy the orthogonality relation $\int_0^1 Z_{k+2s}^k(r) Z_{k+2t}^k(r) r\,dr = \frac{1}{2}(k+2s+1)^{-1}\delta_{st}$, and can be expressed in terms of the more general family of Jacobi polynomials. They arise naturally from a study of the action of rotation on $L^2(B,\mu)$.

The corresponding orthonormal functions in $L^2(D,\lambda)$ are

$$\psi_\nu(s,\varphi) = U_m(s) e^{il\varphi} \qquad \nu = (l,m) \in N, \qquad (s,\varphi) \in D \tag{2.6}$$

where $U_m(\cos\theta) = \sin(m+1)\theta / \sin\theta$ are the Chebychev polynomials of the second kind. We have $P\varphi_\nu = b_\nu \psi_\nu$, with the singular values $b_\nu = b_{lm}$ specified by

$$b_\nu = (m+1)^{-\frac{1}{2}} \qquad \nu = (l,m) \in N \, . \tag{2.7}$$

The relatively slow decay of the singular values with degree m (independently of l) suggests that the costs of indirect observation in the PET problem are not inordinately large.

Since we work with *real* densities f, we may identify the complex bases (2.5) and (2.6) with equivalent real orthonormal bases in a standard fashion. For example $f = \sum f_\nu \varphi_\nu = \sum \tilde{f}_\nu \tilde{\varphi}_\nu$ where

$$\tilde{\varphi}_{l,m} = \begin{cases} \sqrt{2}\,\mathrm{Re}(\varphi_{l,m}) & \text{if } l>0 \\ \varphi_{0,m} & \text{if } l=0 \\ \sqrt{2}\,\mathrm{Im}(\varphi_{l,m}) & \text{if } l<0 \end{cases}$$

and similarly for the real coefficients $\tilde{f}_{l,m}$. From now on, we suppress the tildes in the notation and use whichever basis is convenient.

2.4 Smoothness classes

In our subsequent analysis, we place constraints on the unknown density f over brain space by assuming it lies in a particular class $\mathcal{F}$. For reasons of mathematical tractability, this class is taken to be a particular ellipsoid $\mathcal{F}$ in the Hilbert space $H = L^2(B,\mu)$, specified by an array of constants $\{a_\nu\}$ and a threshold c:

$$\mathcal{F} = \{ f = \sum f_\nu \varphi_\nu \quad : \quad \sum a_\nu^2 f_\nu^2 \le c \} . \tag{2.8}$$

Ellipsoid conditions can amount to the imposition of smoothness and integrability requirements. For example in the simple case where $\{\varphi_\nu\}$ is the sequence of trigonometric polynomials on a bounded interval $[0,2\pi]$ in one dimension and $a_\nu \sim \nu^{-p}$, $\sum a_\nu^2 f_\nu^2 < \infty$ if and only if the periodic function f has p square-integrable derivatives on the interval.

To describe specific ellipsoids in the PET problem, it is useful to transform the index set N by the change of variables $j = (m+l)/2$, $k = (m-l)/2$ into the lattice orthant $N' = \{j,k) : j \ge 0, k \ge 0\}$. Using the real version of the basis $\{\varphi_\nu\}$, let

$$\mathcal{F}_{p,C} = \{ f \in H : f_{00} = 1, \ \sum (j+1)^p (k+1)^p f_{jk}^2 \le 1 + C^2 \} . \tag{2.9}$$

This set is characterised by the following proposition.

Proposition 2.2: *The function f in H lies in some $\mathcal{F}_{p,C}$ if and only if f has p weak derivatives that are square integrable on B with respect to the modified dominating measure $d\mu_{p+1}(x) = (p+1)(1-\|x\|^2)^p d\mu(x)$.*

The condition derived in Proposition 2.2 is of course somewhat weaker than requiring square-integrability with respect to μ and the reason for the modification of the dominating measure is discussed in the proof; a similar technical phenomenon occurs in Cox (1988). Nevertheless, $\mathcal{F}_{p,C}$ can be regarded as imposing a set of smoothness and integrability conditions : the higher p is, the smoother are the functions allowed in $\mathcal{F}_{p,C}$.

How smooth are the functions that we are trying to reconstruct? In X-ray transmission tomography, there may be discontinuities, or at least sharp jumps, in tissue density across the boundaries of various regions. As noted by Natterer (1980, 1986), functions that are piecewise smooth with jumps only along smooth curves lie in Sobolev spaces corresponding to $p < \frac{1}{2}$ square integrable (fractional) derivatives. In emission tomography, with its inherently lower resolution, it may perhaps be reasonable to postulate somewhat smoother emission densities of the labelled metabolite. In any case, our theory is presented for arbitrary values of the smoothness $p > 0$ wherever possible.

To ensure that elements of $\mathcal{F}_{p,C}$ are bona fide probability densities, some further restrictions are needed. To have total mass 1, we require $f_{00} = 1$. By restricting the constant C that governs the ellipsoid size, we can ensure that $f(x) \geq 0$. This is a consequence of the following proposition.

Proposition 2.3: *Suppose* $p \geq 1$ *and* $f \in \mathcal{F}_{p,C}$. *Then*

$$\sup_{x \in B} |f(x) - 1| \leq C 2^{(1-p)/2}. \tag{2.10}$$

Equality is attained in (2.10) *if* f *is a linear function of* x.

It follows from the proposition that $\mathcal{F}_{p,C}$ will be a class of nonnegative functions on B if and only if $C \leq 2^{(p-1)/2}$. Note also that if $g = Pf$, then

$$\sup_y |g(y) - 1| \leq \sup_x |f(x) - 1| \tag{2.11}$$

since P is an averaging operator.

3. Main conclusions of the paper

3.1 Arbitrary estimators

We use minimax mean integrated square error as our basic approach to the quantification of the information available in a given sample. The maximum is taken over a smoothness class $\mathcal{F}_{p,C}$ of unknown functions f, and the minimum is then taken over a class of estimators $\mathcal{T}$, whose specification of takes account of whether the sample is "direct" or "indirect". We define the various classes of estimators as in Section 2.2 above, and the smoothness classes $\mathcal{F}_{p,C}$ as in Section 2.4.

Suppose we have a sample from a density f and an estimator $\hat{f}$ of f based on that sample. An assessment of the accuracy of $\hat{f}$ that does not depend on a *particular* unknown f can be obtained by merely restricting f to lie in a fixed class, for example $\mathcal{F}_{p,C}$ for some fixed p and C, and finding the maximum mean integrated square error

$$R(\hat{f}) = \sup_{f \in \mathcal{F}_{p,C}} M(\hat{f};f) \; . \tag{3.1}$$

The maximum risk gives an indication of how well any given estimator will perform, but a large value of $R(\hat{f})$ might indicate either that there is not much information in the sample or that an inefficient estimator is being used. Because we are interested in the experiment itself rather than any particular estimator, we consider the minimum value of $R(\hat{f})$ over suitable classes of estimators $\hat{f}$.

Define

$$r_D(n) = \inf_{\hat{f} \in \mathcal{T}_D(n)} R(\hat{f}) \tag{3.2}$$

and

$$r_I(n) = \inf_{\hat{f} \in \mathcal{T}_I(n)} R(\hat{f}) \; . \tag{3.3}$$

These minimax risks quantify the information about the unknown density inherent in "direct" and "indirect" data sets of size n, in a manner that is independent of the method of estimation. Comparing their relative values gives an indication of how much information is lost because data can only be observed indirectly in practice.

We can now state our first main result, which gives exact orders of magnitude for $r_D(n)$ and $r_I(n)$ for fixed p and C. The condition placed on C is precisely that needed to ensure that all elements of $\mathcal{F}_{p,C}$ are positive probability densities. Here and subsequently we use the notation $a_n \asymp b_n$ to mean that the sequences $\{a_n\}$ and $\{b_n\}$ satisfy $\inf_n(a_n/b_n)>0$ and $\sup_n(a_n/b_n)<\infty$.

Theorem 3.1: *For fixed* $p \geq 1$ *and* $0<C<2^{\frac{1}{2}(p-1)}$, *with the definitions* (3.1) *to* (3.3),

$$r_D(n) \asymp (\log n/n)^{p/(p+1)} . \tag{3.4}$$

and

$$r_I(n) \asymp (1/n)^{p/(p+2)} . \tag{3.5}$$

The proof of Theorem 3.1 is given in Sections 4 and 5 below. It can be seen from (3.4) and (3.5) that the effect of the indirect nature of the observations taken in practice is to reduce somewhat the rate at which the minimax risk converges to zero.

Suppose, for example, $p=1$, corresponding to f having square-integrable first weak derivatives. Then (neglecting the logarithmic term) the rate is reduced from $n^{-1/2}$ to $n^{-1/3}$ by taking indirect rather than direct observations. Note that both these rates are slower than the n^{-1} rate usually obtained for mean square error in parametric statistics; this is because, even with the restriction that f lies in $\mathcal{F}_{p,C}$, the space of possible parameters is infinite dimensional.

Theorem 3.1 also leads to some qualitative conclusions about equivalent sample sizes. Define the *equivalent sample size* $m(n)$ to a given indirect sample size n to be the number of emissions knowledge of whose original positions in the brain would allow us to estimate f with the same minimax accuracy, so that

$$r_D(m(n)) = r_I(n) \ . \tag{3.6}$$

Some simple algebra from (3.4) and (3.5) yields the order of magnitude of the equivalent sample size as

$$m(n) \asymp n^{(p+1)/(p+2)} \log n. \tag{3.7}$$

Perhaps not surprisingly, the order of magnitude of the equivalent sample size depends on the smoothness assumptions made on the density f. The smoother f is assumed to be, the larger will be the index p. Hence for very smooth densities the power in (3.7) will be close to 1 and little will be lost as a result of the indirect nature of the observation process. However, in reality, we ought not to assume that the true emission density necessarily varies very smoothly, since tissue boundaries and/or localised areas of high metabolic activity may lead to discontinuities, certainly in high derivatives of f and possibly in f itself.

3.2 Linear estimators

More precise numerical quantitative conclusions cannot be drawn directly from (3.7), because Theorem 3.1 only gives orders of magnitude for the relevant risks. We are able, however, to give explicit approximate numerical equivalent sample sizes for minimax risks calculated restricting attention to linear estimators and using as a measure of error the surrogate mean integrated square error M^* defined in (2.4). By analogy to (3.2) and (3.3) define *surrogate linear minimax risks* $r^*_{LD}(n)$ and $r^*_{LI}(n)$ by

$$r^*_{LD}(n) = \inf_{\hat{f} \in \mathcal{T}_{LD}(n)} \ \sup_{f \in \mathcal{F}_{p,C}} M^*(\hat{f};f) \tag{3.8}$$

and

$$r^*_{LI}(n) = \inf_{\hat{f} \in \mathcal{T}_{LI}(n)} \ \sup_{f \in \mathcal{F}_{p,C}} M^*(\hat{f};f). \tag{3.9}$$

The second main result gives leading terms of asymptotic expansions for r^*_{LD} and r^*_{LI}. The leading orders of magnitude are exactly the same as those given for the corresponding quantities in Theorem 3.1, and so the restriction to linear estimators does not affect the rates of convergence available. All the constants c_j depend only on the smoothness p and are collected in Table 1. One of our reasons for introducing surrogate mean integrated square error is that we have been able to derive these more precise expressions, and hence obtain numerical results. The other reason is that the result of Theorem 3.2 is a key step in the proof of Theorem 3.1.

Table 1 near here

Theorem 3.2: *For* $x > 1$, *let* $\alpha(x)$ *denote the solution to* $\alpha \log \alpha = x$, *and set*

$$\eta_n^{p+1} = c_1^{p+1} \alpha(c_2 n C^2). \tag{3.10}$$

Then, provided $0 < C \leq 2^{\frac{1}{2}(p-1)}$,

$$r^*_{LD}(n) = c_3 n^{-1} \eta_n (\log \eta_n + c_4) + O(n^{-1} \eta_n^{\frac{1}{2}}) \tag{3.11}$$

$$= c_5 C^{2/(p+1)} (\log n/n)^{p/(p+1)} \{1 + o(1)\} \tag{3.12}$$

and

$$r^*_{LI}(n) = c_6 C^{4/(p+2)} n^{-p/(p+2)} + O(n^{-(p+1)/(p+2)} \log n) . \tag{3.13}$$

The form (3.12) for r^*_{LD} is more transparent, but the error term can be shown to have the same polynomial order as the leading term; the error term in (3.11) is of lower order and so we use (3.11) in numerical computations. Of course, $\alpha(x)$ can be found numerically when required and is asymptotic to $x/\log x$ for large x.

For any particular indirect sample size n, the approximate equivalent sample size $m^*(n)$ can be found: equate the expressions (3.11) for $r^*_{LD}(m^*)$ and (3.13) for $r^*_{LI}(n)$, neglect the lower order terms, and solve numerically for m^*. For definiteness we take $C^2 = 2^{p-1}$, the largest value for which all f in $\mathcal{F}_{p,C}$ are non-negative probability densities, so long as $p \geq 1$ (Proposition 2.3). Some representative cases are given in Table 2. As expected, the equivalent sample size increases as the assumed amount of smoothness rises. If technology allows an order of magnitude increase in the amount of data collected, then the equivalent direct sample sizes increase by a factor of between 5 and 8, this factor itself increasing with assumed smoothness.

Table 2 near here

For the quantity $m^*(n)$ the asymptotic constant of proportionality in the expression corresponding to (3.7) can be found. A simple calculation uses relations (3.12) and (3.13), with the error terms ignored, to conclude that

$$m^*(n) = (p+1)(p+2)^{-1} (c_5/c_6)^{(p+1)/p} C^{-2/(p+2)} n^{(p+1)/(p+2)} \log n \{1 + o(1)\} .$$

In summary, our results confirm intuition that for the PET problem, the amount of information available is still substantial, but it is by no means as great as if a sample of the same number of direct observations were available.

4. Convergence rates for linear estimators

The main aim of this section is to prove Theorem 3.2, which gives the asymptotic behaviour of the surrogate risks (2.4) for linear estimators. It is a consequence of Propositions 2.1 and 2.3 that, provided $C<2^{\frac{1}{2}(p-1)}$, the ratio exact to surrogate mean integrated square error for linear estimates will be bounded above and below away from 0 uniformly over $\mathcal{F}_{p,C}$. Since $\mathcal{T}_{LD}(n)$ and $\mathcal{T}_{LI}(n)$ are subclasses of $\mathcal{T}_D(n)$ and $\mathcal{T}_I(n)$ respectively, it then follows that the orders of magnitude of $r_D(n)$ and $r_I(n)$ are bounded above by those obtained in Theorem 3.2 for surrogate linear minimax risks. Once Theorem 3.2 has been proved, the proof of Theorem 3.1 will be completed in Section 5 by showing that these are also lower bounds.

4.1 Structure of the linear minimax estimator

We consider the indirect case first; the argument we shall use will apply to the direct case also. We start by defining some notation. Suppose that $\hat{f}$ is in $\mathcal{T}_{LI}(n)$. For ν and π in N define $w_{\nu\pi} = \int w(x,y)\varphi_\nu(x)\psi_\pi(y)\,d\mu(x)\,d\lambda(y)$; because of the condition $\int\int w^2 d\mu d\lambda < \infty$, standard functional analysis gives that, in the L^2 sense,

$$w(x,y) = \sum_\nu \sum_\pi w_{\nu\pi}\,\varphi_\nu(x)\psi_\pi(y). \tag{4.1}$$

As in Section 2.3 and 2.4, we expand f as $\Sigma f_\nu \varphi_\nu$. We write W for the infinite matrix $(w_{\nu\pi})$ *and* $\mathbf{f}$ *for the vector* (f_ν). The index set of all vectors and matrices will be the set N; the subscript (0,0) will be written as 0 for simplicity. Since $\int f d\mu = 1$ the coefficient $f_0 = 1$. Write $B = \mathrm{diag}(b_\nu)$, the singular values of the operator P. Let $\mathbf{e}_\nu$ be the vector $(\delta_{\nu\pi} : \pi \in N)$. The first lemma gives a matrix form for the surrogate mean integrated square error of the linear estimator $\hat{f}$.

Lemma 4.1 *With the above definitions,*

$$M^*(\hat{f}\,;f) = n^{-1}\,\mathrm{tr}\,W(I-\mathbf{e}_0\mathbf{e}_0{}^{\mathrm{T}})W + \mathbf{f}^{\mathrm{T}}(I-WB)^{\mathrm{T}}(I-WB)\mathbf{f} \tag{4.2}$$

Proof Write $\hat{f} = \sum \hat{f}_\nu \varphi_\nu$. From (4.1) it follows that $\hat{\mathbf{f}} = W\eta$ where $\eta_\nu = n^{-1}\sum_i \psi_\nu(Y_i)$. Each Y_i has density $g = \sum g_\nu \psi_\nu$ where $\mathbf{g} = B\mathbf{f}$, and for each ν $E_f \eta_\nu = \int \psi_\nu g d\lambda = g_\nu$, so that $E_f\,\eta = B\mathbf{f}$. Hence $E_f \hat{\mathbf{f}} = WB\mathbf{f}$, and the integrated square bias

$$\int (E_f\hat{f}-f)^2\,d\mu = \|E\hat{\mathbf{f}}-\mathbf{f}\|^2 = \|WB\mathbf{f}-\mathbf{f}\|^2 = \|(I-WB)\mathbf{f}\|^2 \tag{4.3}$$

If f is the uniform density, then $\mathbf{f} = \mathbf{e}_0$ and so, writing E_1 for an expectation relative to the uniform density f, $E_1\eta = B\mathbf{e}_0 = \mathbf{e}_0$ since $b_0 = 1$.

By the orthonormality of the ψ_ν, the matrix $E_1\eta\eta^{\mathrm{T}} = n^{-1}I$ and so η has covariance matrix $n^{-1}(I-\mathbf{e}_0\mathbf{e}_0{}^{\mathrm{T}})$ under the uniform distribution. Thus the surrogate variance term

$$\int \mathrm{var}_1 \hat{f} d\mu = E_1 \|\hat{\mathbf{f}} - E_1 \hat{\mathbf{f}}\|^2 = E_1 \| W(\eta - E_1 \eta)\|^2 = n^{-1} \mathrm{tr}\, W(I - \mathrm{e}_0 \mathrm{e}_0{}^{\mathrm{T}}) W^{\mathrm{T}} \quad (4.4)$$

by a standard multivariate calculation. To complete the proof, substitute (4.3) and (4.4) into the definition (2.4) of surrogate mean integrated square error.

□

Our second lemma provides an expression for the surrogate linear minimax risk and gives the general form of the minimax estimator. The smoothness class $\mathcal{F}$ is defined as in (2.8) and (2.9) to be $\mathcal{F} = \{f : f_0 = 1, \mathbf{f}^{\mathrm{T}} A \mathbf{f} \le 1 + C^2\}$ where we write $A = \mathrm{diag}(a_\nu^2)$, and assume that $a_0 = 1, \sup a_\nu^2 = \infty$, and that every f in $\mathcal{F}$ is non-negative.

Lemma 4.2

$$\inf_{\hat{f} \in \mathcal{T}_{LI}(n)} \ \sup_{f \in \mathcal{F}} \ M^*(\hat{f}; f) = n^{-1} \sum b_\nu^{-2} (1 - a_\nu \gamma^{\frac{1}{2}})_+ \quad (4.5)$$

where γ is chosen to ensure that

$$n^{-1} \sum_{\nu \ne 0} b_\nu^{-2} a_\nu^2 (\gamma^{-\frac{1}{2}} a_\nu^{-1} - 1)_+ = C^2. \quad (4.6)$$

The minimax estimator is given by setting, in (4.1),

$$w_{\nu\pi} = \delta_{\nu\pi} \text{ for } \nu = 0 \text{ and } w_{\nu\pi} = \delta_{\nu\pi} b_\nu^{-1} (1 - \gamma^{\frac{1}{2}} a_\nu)_+ \textit{ otherwise}. \quad (4.7)$$

The form of the minimax estimator is worth noting, since it corresponds to a *diagonal* matrix of weights and hence is an estimator of the form $\hat{f}(x) = n^{-1} \sum b_\nu^{-1} u_\nu \psi_\nu(Y_i) \varphi_\nu(x)$. Although the derivation of the estimator has been performed for theoretical reasons, some examples of the use of estimators of this kind are given by Jones & Silverman (1989). Similar results to Lemma 4.2 exist for standard regression (for example Pinsker, 1980, Speckman, 1985) and for other nonparametric problems (for example Buckley et al., 1988). Our proof is an extension of that of Speckman (1985, pp.981-982).

Proof of Lemma 4.2 The condition $\int w(x,y) d\mu(x) = 1$ for all y implies that $w_{00} = 1$ and $w_{0\nu} = 0$ for $\nu \ne 0$. Let $\mathcal{W}$ be the set of matrices W satisfying this condition and for which $\sum \sum w_{\nu\pi}^2 < \infty$; the matrices W in $\mathcal{W}$ correspond precisely to the estimators $\hat{f}$ in $\mathcal{T}_{LI}(n)$. We use Lemma 4.1 and find the minimax value of the expression (4.2) over W in $\mathcal{W}$ and f in $\mathcal{F}$. Let

$$J(W) = \sup_{f \in \mathcal{F}} \left\{ \| (I - WB) \mathbf{f} \|^2 + n^{-1} \mathrm{tr}\, W(I - \mathrm{e}_0 \mathrm{e}_0{}^{\mathrm{T}}) W^{\mathrm{T}} \right\}. \quad (4.8)$$

Let W^0 be the matrix $\mathrm{diag}(w_{\nu\nu})$; we show that $J(W) \ge J(W^0)$, and hence that we may restrict attention to diagonal matrices in $\mathcal{W}$.

For fixed κ in N, $\kappa \neq 0$, let $\mathcal{F}_\kappa$ be the set $\{f = 1 + f_\kappa \varphi_\kappa, a_\kappa^2 f_\kappa^2 \leq C^2\}$. Then

$$\sup_{\mathcal{F}_\kappa} \|(I-WB)\mathbf{f}\|^2 = \sup_{\mathcal{F}_\kappa} \sum_\nu (w_{\nu 0}+w_{\nu\kappa} b_\kappa f_\kappa - f_\nu)^2 \geq \sup_{\mathcal{F}_\kappa} \{w_{\kappa 0}+(w_{\kappa\kappa} b_\kappa - 1) f_\kappa\}^2$$

$$\geq (1-w_{\kappa\kappa} b_\kappa)^2 C^2/a_\kappa^2, \tag{4.9}$$

by picking out the κ term from the summation and performing some elementary algebra. Again by restricting the sum, we have

$$\operatorname{tr} W(I-\mathbf{e}_0\mathbf{e}_0{}^{\mathrm{T}})W^{\mathrm{T}} = \sum_{\nu\neq 0} \sum_{\kappa\neq 0} w_{\nu\kappa}^2 \geq \sum_{\nu\neq 0} w_{\nu\nu}^2. \tag{4.10}$$

Restricting the supremum to f in $\cup\mathcal{F}_\kappa$, and substituting (4.9) and (4.10), we obtain

$$J(W) \geq \sup_{\kappa\neq 0}(1-w_{\kappa\kappa}b_\kappa)^2 C^2/a_\kappa^2 + n^{-1} \sum_{\nu\neq 0} w_{\nu\nu}^2 = J(W^0) \tag{4.11}$$

by checking that every inequality in our argument is an exact equality when W is diagonal.

Let $\gamma = \sup_{\kappa\neq 0}(1-w_{\kappa\kappa}b_\kappa)^2 a_\kappa^{-2}$. Now reason from (4.11) as in Speckman (1985) to obtain (4.7); then substitute into the expression for $J(W^0)$ in (4.11) and minimize over γ to complete the proof.

□

To obtain corresponding results for the direct case, set the operator P to the identity in the whole of the preceding argument. The minimax surrogate risk $r_{LD}^*(n)$ is given by (4.5) and (4.6) with all b_ν set to 1. The minimax estimator $n^{-1} \sum_{i,\nu} w_{\nu\nu} \varphi_\nu(X_i)\varphi_\nu(x)$ is a probability density estimate of tapered orthogonal series form as introduced and studied by Watson (1969).

4.2 Integral approximation of the minimax risks

In this subsection we explicitly approximate the expression (4.5), and the corresponding expression for the direct case, to complete the proof of Theorem 3.2. We set $\mathcal{F}=\mathcal{F}_{p,C}$ as in (2.4) so that $a_{jk}^2=(j+1)^p(k+1)^p$. The key to our treatment is the following approximation lemma, obtained by approximating sums by integrals.

Lemma 4.3 *For any η, let $\sum_{(\eta)}$ denote a sum over $\{(j,k)\colon 1<(j+1)(k+1)\leq\eta\}$. For fixed $r\geq 0$, as $\eta\to\infty$,*

$$\textstyle\sum_{(\eta)} (j+1)^r(k+1)^r = (r+1)^{-1}\eta^{r+1}\{\log\eta+2\gamma_E-(r+1)^{-1}\} + O(\eta^{r+\frac{1}{2}}) \tag{4.12}$$

where γ_E is Euler's constant, and

$$\textstyle\sum_{(\eta)}(j+k+1)(j+1)^r(k+1)^r = \frac{1}{3}\pi^2(r+2)^{-1}\eta^{r+2} + O(\eta^{r+1}\log\eta)\ . \tag{4.13}$$

Proof For the proof, we transform the sums by replacing $j+1$ by j and $k+1$ by k; denote by $\sum_{[\eta]}$ the sum over the transformed range $\{(j,k) : j\geq 1, k\geq 1 \text{ and } 1<jk\leq\eta\}$. By symmetry in (j,k), the sum in (4.12) satisfies

$$1+S = \sum_{[\eta]} j^r k^r = 2\sum_{k=1}^{[\eta^{\frac{1}{2}}]} k^r \sum_{j=1}^{[\eta k^{-1}]} j^r - \sum_{j=1}^{[\eta^{\frac{1}{2}}]}\sum_{k=1}^{[\eta^{\frac{1}{2}}]} j^r k^r .$$

From the relation $\sum_{j=1}^{t} j^r = (r+1)^{-1} t^{r+1} + O(t^r)$, we obtain

$$\begin{aligned} S &= 2(r+1)^{-1}\sum_{k=1}^{[\eta^{\frac{1}{2}}]} k^r\{[\eta k^{-1}]^{r+1} + O(\eta^r k^{-r})\} - \{(r+1)^{-1}[\eta^{\frac{1}{2}}]^{r+1} + O(\eta^{r/2})\}^2 - 1 \\ &= 2(r+1)^{-1}\eta^{r+1}\sum_{k=1}^{[\eta^{\frac{1}{2}}]} k^{-1} - \sum_{k=1}^{[\eta^{\frac{1}{2}}]} k^r\{(\eta k^{-1})^{r+1} - [\eta k^{-1}]^{r+1}\} - (r+1)^{-2}\eta^{r+1} + O(\eta^{r+\frac{1}{2}}) \\ &= 2(r+1)^{-1}\eta^{r+1}\{\tfrac{1}{2}\log\eta + \gamma + O(\eta^{-\frac{1}{2}})\} - (r+1)^{-2}\eta^{r+1} + O(\eta^{r+\frac{1}{2}}), \end{aligned} \tag{4.14}$$

which yields the result of (4.12).

To deal with (4.13), we need an integral approximation, valid for $s\geq 0$ and $x\geq 1$,

$$\sum_{j=1}^{[x]} j^s = (s+1)^{-1}x^{s+1} + c_{sx}x^s, \qquad 0\leq |c_{sx}| \leq c_s, \tag{4.15}$$

which follows from the bounds $(s+1)^{-1}[x]^{s+1} \leq \int_0^{[x]} t^s\,ds \leq \sum_{j=1}^{[x]} j^s \leq \int_1^{[x+1]} t^s\,ds \leq (s+1)^{-1}[x+1]^{s+1}$. Assuming that η is an integer, it then follows that

$$\begin{aligned} 1 + \sum_{[\eta]} j^{r+1}k^r &= \sum_{k=1}^{\eta} k^r \sum_{j=1}^{[\eta k^{-1}]} j^{r+1} = \sum_{k=1}^{\eta} k^r (r+2)^{-1}(\eta k^{-1})^{r+2} + \sum_{k=1}^{\eta} k^r c_{r+1,\eta,k}(\eta k^{-1})^{r+1} \\ &= (r+2)^{-1}\eta^{r+2}\sum_{1}^{\eta} k^{-2} + O(\eta^{r+1}\sum_{1}^{\eta} k^{-1}) \\ &= \frac{1}{6}\pi^2 (r+2)^{-1}\eta^{r+2} + O(\eta^{r+1}\log\eta). \end{aligned} \tag{4.16}$$

To complete the proof of (4.13), transform the sum to $\sum_{[\eta]}(j^{r+1}k^r + j^r k^{r+1} - j^r k^r)$. Then substitute (4.16) for each of the first two terms, and use (4.15) to absorb the third term into the error. ☐

Completion of proof of Theorem 3.2. We will have $\gamma a_\nu^2 \leq 1$ if and only if $(j+1)(k+1) \leq \gamma^{-1/p}$ and so the $(\ )_+$ in (4.6) and (4.7) may be replaced by $(\ )$ if the sums over all ν are replaced by $\sum_{(\eta)}$ with $\eta = \gamma^{-1/p}$. The constants c_r will be defined as in Table 3.2..

In the direct case, we replace $\mathcal{T}_{LI}$ in (4.5) by $\mathcal{T}_{LD}$ and set all b_ν to 1. Applying (4.12), equation (4.6) becomes

$$\begin{aligned} C^2 &= n^{-1} \textstyle\sum_{(\eta)} (\gamma^{-\frac{1}{2}} a_v - a_v^2) \\ &= n^{-1} \textstyle\sum_{(\eta)} \{\gamma^{-\frac{1}{2}} (j+1)^{p/2} (k+1)^{p/2} - (j+1)^p (k+1)^p\} \\ &= n^{-1} \eta^{p+1} (c_7 \log \eta + c_8) + n^{-1} O(\eta^{p+\frac{1}{2}}). \end{aligned} \tag{4.17}$$

The substitution $\eta = c_1 y^{1/(p+1)}$ reduces equation (4.17) with the error term omitted to the form $y \log y = c_2 n C^2$; it follows that η_n as defined in (3.10) is the solution for η of this equation. Apply similar manipulations to (4.5) to obtain

$$r_{LD}^*(n) = n^{-1} \textstyle\sum_{(\eta)} (1 - \gamma^{\frac{1}{2}} a_v) = c_3 n^{-1} \eta_n \{\log \eta_n + c_4\} + n^{-1} O(\eta_n^{\frac{1}{2}}),$$

completing the proof of (3.11). To prove (3.12), substitute the definition of η_n into (3.11), and use the fact that $\alpha(x) = (x/\log x)\{1 + o(1)\}$ for large x.

For the indirect case, we use the values (2.7) for the b_v. Equation (4.6) then becomes

$$\begin{aligned} C^2 &= n^{-1} \textstyle\sum_{(\eta)} (j+k+1)\{\gamma^{-\frac{1}{2}} (j+1)^{p/2} (k+1)^{p/2} - (j+1)^p (k+1)^p\} \\ &= c_9 n^{-1} \eta^{p+2} + n^{-1} O(\eta^{p+1} \log \eta). \end{aligned} \tag{4.18}$$

where $c_9 = (\pi^2/3) p (p+2)^{-1} (p+4)^{-1}$. Set $\tilde{\eta}_n = (nC^2/c_9)^{1/(p+2)}$, the solution to (4.18) with the error omitted. Then the solution to (4.18) with the error included satisfies $\eta = \tilde{\eta}_n + O(\log \tilde{\eta}_n)$. Substitute back into (4.5), apply Lemma 4.3, and perform some elementary algebra to obtain (3.13), and hence to complete the proof of Theorem 3.2.

□

To summarise this section, we have shown that, for linear estimators, the indirect nature of the PET observations reduces the minimax rate of consistency in mean integrated square error from $O\{(n/\log n)^{-p/(p+1)}\}$ to $O(n^{-p/(p+2)})$. It will be shown in the next section that these rates of consistency are both best possible even if we allow the class of estimators to be extended to cover all linear and non-linear estimators.

5. Lower bounds

In this section we establish lower bounds on the rates of consistency of arbitrary estimators based on direct and indirect observations. These lower bounds show that the minimax rates obtained for linear estimators in Section 4 cannot be improved by extending the class of estimators considered. As noted at the beginning of Section 4, this will complete the proof of Theorem 3.1.

5.1 Moduli of continuity and a general lower bound for global norms

Our approach is based on Fano's lemma of information theory, as developed by Ibragimov and Hasminskii (e.g. 1981) and Birgé (1983), although a slight extension of Birgé's formulation is needed for the indirect observation case. Although we continue to focus on the PET example, it will be seen that the methodology applies quite generally to estimation with global norms in linear inverse problems of both density and regression estimation type.

The convergence rate in the indirect problem clearly depends on the operator P^{-1} mapping the observable density g to the target density f. One convenient approach to computing convergence rates has two parts: (i) compute a "modulus of continuity" $\tau(\varepsilon)$ for P^{-1}, and (ii) argue that a lower bound to the minimax convergence rate is given by (essentially) $\tau(n^{-\frac{1}{2}})$. This approach separates stochastics and analysis: step (ii) uses the information theory lemma to bound the estimation error by $\tau(n^{-\frac{1}{2}})$ while step (i) is a concrete optimisation problem for the particular operator in question. This viewpoint was taken recently by Donoho and Liu (1989) in their study of estimation of linear *functionals*. We begin with step (ii), which computes a modulus $\sigma(\delta)$ which is more convenient for the problems at hand. We return to step (i) in Section 5.2 below.

Suppose, in general, there are available n i.i.d. observations $Y^{(n)} = (Y_1, \ldots, Y_n)$ from a density $g(y)d\lambda(y)$, $y \in D$, and that we wish to estimate $f = P^{-1}g$. We assume that $f \in \mathcal{F} \subset H$, and that $\mathcal{F}$ is a translate $f^0 + H_0$ of a set H_0 that is balanced about the origin ($h \in H_0 \Rightarrow -h \in H_0$). Let M be a finite-dimensional subspace of H: we write $|M|$ for the dimension of M and $B_M(\delta)$ for the open ball of radius δ about 0 in M. The norm of the restriction of P to M is defined by $\|P\|_M = \sup\{\|Ph\|/\|h\| : h \in M\}$. Finally, let $\mathcal{M}_\delta = \{M : B_M(\delta) \subset H_0\}$. The modulus $\sigma(\delta)$ may now be defined as

$$\sigma(\delta) = \delta \inf\{\|P\|_M / |M|^{\frac{1}{2}} : M \in \mathcal{M}_\delta\}. \tag{5.1}$$

Loosely speaking, $\sigma(\delta)$ measures the decay of the singular values of P relative to the parameter space H_0 at resolution δ. Since σ is strictly increasing, a left-continuous inverse $\tau(\varepsilon) = \sigma^{-1}(\varepsilon)$ can be defined.

Let $\hat{f} \in \mathcal{T}_l(n)$ be an arbitrary estimator based on $Y^{(n)}$. The significance of the modulus functional is that an (often sharp) lower bound for the rate of convergence of $\|\hat{f} - f\|$ over $\mathcal{F}$ is given by $\tau(n^{-\frac{1}{2}})$. For the proof we need an additional assumption bounding the Kullback-Leibler information divergence $K(g_\alpha, g_\beta) = \int \log(g_\alpha/g_\beta)\, g_\alpha d\lambda$ over $\mathcal{G} = P\mathcal{F}$:

$$\text{For some } A < \infty, \quad K(g_\alpha, g_\beta) \le A\|g_\alpha - g_\beta\|_K^2 \quad \text{for all } g_\alpha, g_\beta \in \mathcal{G}. \tag{5.2}$$

This condition will be satisfied provided the densities g in $\mathcal{G}$ are uniformly bounded above and below away from zero. In the context of Theorem 3.1, this is a consequence of (2.11) and (2.10).

Proposition 5.1 *If condition* (5.2) *holds, there exist constants* d_1, d_2 *such that*

$$\inf_{\hat{f}\in\mathcal{T}_l(n)} \sup_{f\in\mathcal{F}} E_f \|\hat{f}-f\|_H^2 \geq d_1\tau^2(d_2 n^{-\frac{1}{2}}). \tag{5.3}$$

Proof Choose a subset $\mathcal{F}^0 = \{f_1,\dots,f_r\} \subset \mathcal{F}$ that is *2δ-distinguishable* : namely $\|f_\alpha - f_\beta\| > 2\delta$ if $\alpha \neq \beta$. Set $g_\alpha = Pf_\alpha$ and write $K^n(g_\alpha, g_\beta) = n\int \log(g_\alpha/g_\beta)g_\alpha d\lambda$, the Kullback-Leibler discrepancy based on a sample of size n.

Consider the discrimination problem of choosing among the r hypotheses $\mathcal{F}^0$. Given an estimator $\hat{f}\in\mathcal{T}_l(n)$, define a discrimination rule $\varphi(Y^{(n)})$ taking values in $\mathcal{F}^0$ that picks the closest element in $\mathcal{F}^0$ to $\hat{f}$. Then, by elementary probability and analysis,

$$\sup_{f\in\mathcal{F}} E_f\|\hat{f}-f\|^2 \geq \sup_{f\in\mathcal{F}^0} E_f\|\hat{f}-f\|^2 \geq \delta^2 \sup_{f\in\mathcal{F}^0} P_f(\|\hat{f}-f\|>\delta)$$

$$\geq \delta^2 r^{-1} \sum_{\alpha=1}^{r} P_{f_\alpha}(\|\hat{f}-f_\alpha\|>\delta) \geq \delta^2 r^{-1} \sum_{\alpha=1}^{r} P_{f_\alpha}\{\varphi(Y^{(n)})\neq f_\alpha\}. \tag{5.4}$$

since $\varphi(Y^{(n)})\neq f_\alpha$ implies that $\|\hat{f}-f_\alpha\|>\delta$, because of the 2δ-distinguishability.

By Birgé's version (1983, p.196) of Fano's lemma, the average error rate in the discrimination problem can be bounded below as follows:

$$r^{-1}\sum_{\alpha=1}^{r} P_{f_\alpha}(\varphi(Y^{(n)})\neq f_\alpha) \geq 1 - \{\sup_{1\leq\alpha,\beta\leq r} K^n(g_\alpha, g_\beta) + \log 2\}/\log(r-1). \tag{5.5}$$

Combining (5.4) and (5.5), and substituting (5.2), we obtain the lower bound

$$\delta^{-2}\sup_{f\in\mathcal{F}} E_f\|\hat{f}-f\|^2 \geq 1 - \{nA \sup_{1\leq\alpha,\beta\leq r} \|Pf_\alpha - Pf_\beta\|_K^2 + \log 2\}/\log(r-1) . \tag{5.6}$$

To make use of this lower bound, we use the metric dimension properties of $\mathcal{F}$ and the operator P to construct a suitable set $\mathcal{F}^0$ for which r is large and $\sup\|Pf_\alpha - Pf_\beta\|_K^2$ is small. From the definition (5.1) of the modulus σ, choose a subspace M of H for which $B_M(4\delta) \subset H_0$ and $4\delta\|P\|_M/|M|^{\frac{1}{2}} \leq 2\sigma(4\delta)$. A useful lemma of approximation theory (e.g. Lorentz, 1966, p.905) asserts that a k dimensional ball of radius R contains an $R/2$ distinguishable subset of cardinality at least 2^k. Setting $r=2^{|M|}$, use this lemma to choose $h_1,\dots,h_r \in B_M(4\delta)$ such that $\|h_\alpha - h_\beta\| \geq 2\delta$ and define the 2δ-distinguishable set $\mathcal{F}^0$ by f_α as $f^0 + h_\alpha$ for $\alpha=1,\dots,r$. By construction, for any α and β,

$$\|Pf_\alpha - Pf_\beta\|_K^2 \leq \|P\|_M^2 \|f_\alpha - f_\beta\|^2 \leq \tfrac{1}{4}\delta^{-2}|M|\sigma(4\delta)^2 . 64\delta^2 = 16|M|\sigma(4\delta)^2. \tag{5.7}$$

Substituting back into (5.6), and performing some elementary algebra, we have $\sup_{\mathcal{F}} E_f\|\hat{f}-f\|^2 \geq \delta^2[1 - d_3 n\sigma^2(4\delta)]$ where d_3 is an appropriate constant. Now choose δ so that $d_3 n\sigma^2(4\delta) = \frac{1}{2}$ and the proof of Proposition 5.1 is complete. □

The estimation problem we study can be thought of as estimation of Qg , where $g \in \mathcal{G}$ and Q ($=P^{-1}$) is an unbounded operator. The term "modulus of continuity" might be more appropriately applied to a measure of the rate of growth of the singular values of Q relative to $\mathcal{G}$. Indeed it is in this form that the similarity to the modulus of Donoho and Liu (1989) is clearer. Now suppose that $\mathcal{G}$ is a translate g^0+K_0 of a set balanced about the origin in K. We denote finite dimensional balls about 0 in K by U. Define the normalised radius $\rho(U)$ to be the radius of U divided by the square root of the dimension of U.

Define a generalised modulus of continuity of Q over the parameter space K_0 by

$$\tilde{\tau}(\varepsilon) = \sup \inf_{v \in \partial U} \|Qv\|_H, \tag{5.8}$$

where the supremum is taken over the class of finite dimensional balls $U \subset K_0$ for which $\rho(U)=\varepsilon$. Notice that if Q is a linear *functional* (so that $(H,\|\cdot\|_H) = (\mathbf{R},|\cdot|)$), the above definition reduces to

$$\tau(\varepsilon) = \sup \{ |Qv| : \|v\|_K = \varepsilon \text{ and } tv \in K_0 \text{ for } |t| \le 1 \},$$

which is the modulus of continuity studied by Donoho and Liu (1989).

It can be shown that $\tilde{\tau}$ is approximately inversely related to the modulus σ defined at (5.1) in the sense that $\tilde{\tau}(\sigma(\delta)) \le \delta$. Thus $\sigma^{-1}(\varepsilon) \ge \tilde{\tau}(\varepsilon)$, and so the lower rate bounds derived from use of σ are at least as good as those that would follow from $\tilde{\tau}$. It turns out that these rate bounds are in fact equivalent for all the applications discussed in this paper. These results and extensions will be discussed more fully elsewhere.

5.2 Completion of proof of Theorem 3.2

We now return to the PET setting to prove two propositions that complete the Theorem 3.1. Both these are proved by finding reasonable lower bounds to $\tau(\varepsilon)$.

Proposition 5.2 *Subject to the conditions of Theorem* 3.1, *there exists a constant* $d_D(p,C) > 0$ *such that*

$$r_D(n) \ge d_D(\log n/n)^{p/(p+1)}.$$

Proof Set $H = K = L^2(B,\mu)$ and $P = I$. Let f^0 be the uniform density and $H_0 = \mathcal{F}_{p,C} - f^0$. A good upper bound for $\sigma(\delta)$ as defined in (5.1) can be obtained by considering high dimensional subspaces M subject to the constraint that $B_M(\delta) \subset H_0$. For large η, let $M_\eta = \mathrm{span}\{\varphi_\nu : a_\nu^2 \le \eta^p\}$. Then $B_{M_\eta}(\delta) \subset H_0$ when $\eta^p \le C^2/\delta^2$. From the definition of $\sigma(\delta)$, it follows that

$$\sigma^2(\delta) \le \delta^2 / \sup\{ |M_\eta| : \eta^p \le C^2/\delta^2 \}.$$

Using χ to denote the characteristic function of a set, $|M_\eta| = \sum \chi\{1<(j+1)(k+1)\le\eta\} = \eta \log \eta \{1+o(1)\}$ by Lemma 4.3. Hence $\sigma^2(\delta) \le d_4 \delta^{2(p+1)/p} / \log \delta^{-2}$, from which it follows that $\tau^2(\varepsilon) \ge d_5(\varepsilon^2 \log \varepsilon^{-2})^{p/(p+1)}$, so that $\tau^2(cn^{-\frac{1}{2}}) \ge d_6(\log n/n)^{p/(p+1)}$. Substitute back into Theorem 5.1 to complete the proof. $\square$

Proposition 5.3 *Subject to the conditions of Theorem* 3.1, *there exists a constant* $d_l(p,C) > 0$ *such that*

$$r_l(n) \ge d_l(1/n)^{p/(p+2)}.$$

Proof Now take H and H_0 as above, and let K be the Hilbert subspace of $L^2(D,\lambda)$ generated by the orthonormal set of singular functions $\{\psi_v\}$. This time a good bound for $\sigma(\delta)$ must use high dimensional subspaces (with $B_M(\delta) \subset H_0$) for which in addition $\|P\|_M$ is small. For given η, set $M_\eta = \text{span}\{\varphi_{j0} : \frac{1}{2}\eta \le j+1 \le \eta\}$. Then $\|P\|^2_{M_\eta} = \max\{b_v^2 : \varphi_v \in M_\eta\} \le 2\eta^{-1}$, and $|M_\eta| \ge [\frac{1}{2}\eta]$. As in the proof of Proposition 5.2, $B_{M_\eta}(\delta) \subset H_0$ if $\eta^p \le C^2/\delta^2$. Substituting into (5.1), we have, for sufficiently small δ,

$$\sigma^2(\delta) \le \delta^2 \inf \{2\eta^{-1}/[\tfrac{1}{2}\eta] : \eta^p \le C^2/\delta^2\} = d_7 \delta^{2(p+2)/p}.$$

Consequently $\tau^2(\varepsilon) \ge d_8 \varepsilon^{2p/(p+2)}$ and $\tau^2(cn^{-\frac{1}{2}}) \ge d_9 n^{-p/(p+2)}$, which, as above, can be substituted into Proposition 5.1 to complete the proof. $\square$

We close this section by remarking that Ibragimov and Hasminskii (1981) and Stone (1982) have shown that the minimax rate of convergence of global mean integrated square error for *direct* nonparametric density and regression problems is $n^{-2p/(2p+d)}$, where p is the assumed amount of smoothness and d is the dimension, $d=2$ in our case. They consider classes of functions constrained by a Hölder continuity condition of order $\alpha \in (0, 1]$ on the s^{th} derivative, so that $p=s+\alpha$. The extra $\log n$ term in the rate of convergence $(\log n/n)^{2p/(2p+d)}$ obtained in the present paper reflects the slightly reduced smoothness imposed by requiring only square-integrability of the p^{th} weak derivative.

6. Biased sampling and attenuation

In any practical PET scan, not all pairs of emitted photons are detected. We shall show in this section that two of the main reasons for this incompleteness of sampling can be placed within the same mathematical framework, and that our results can, in part, be extended to account for them. Under mild assumptions, the incompleteness of sampling has no effect on the minimax rate of consistency found in Theorem 3.1.

6.1 The effect of the third dimension

Up to now, we have considered the detectors as forming a circle in the plane, and we have assumed that all the paths of emitted photons fall in this plane. Of course, in reality the detectors form a ring of finite thickness $d>0$, and the orientation of the line of flight of the photons is uniformly distributed in $\mathbf{R}^3$. We shall assume that the emission density is constant over the thickness of the cylindrical slab enclosed by the detector ring. Only emissions taking place in this slab will be considered, since only they have any chance of being detected at all.

Given any emission, the photon line-of-flight is now parametrised by three coordinates (s,φ,φ'), where (s,φ) are the coordinates in detector space of the projection of the line onto the detector plane, and the *vertical angle* $\varphi'(-\pi/2<\varphi'\leq\pi/2)$ is the angle between the line and its projection. The assumption that the line has uniformly distributed direction implies that, independently of (s,φ), the vertical angle has probability density $\frac{1}{2}\cos\varphi'\,d\varphi'$. An emission line will only be detected if its vertical angle is such that both photons hit the detector ring. If the emission is detected, only the coordinates s and φ are observed.

Fig 6.1 near here

Condition on a particular s and φ, and let $l = 2(1-s^2)^{\frac{1}{2}}$, the length of the corresponding detector tube. Assume that an emission takes place at distance t from the centre of the tube and at vertical position Z as shown in Figure 6.1. Assume that the projection of the line of flight of the emitted photons has coordinates (s,φ). Let $(-\varphi_1,\varphi_2)$ be the range of vertical angles over which both photons will hit the detectors. For given φ_1 and φ_2 the probability of detection will be

$$\int_{-\varphi_1}^{\varphi_2} \tfrac{1}{2}\cos\varphi'\,d\varphi' = \tfrac{1}{2}(\sin\varphi_1 + \sin\varphi_2).$$

We have (see Figure 6.2)

$$\tan\varphi_2 = \begin{cases} Z/(t+\frac{1}{2}l) & \text{if } Z<(t+\frac{1}{2}l)d/l \\ (d-Z)/(\frac{1}{2}l-t) & \text{otherwise} \end{cases}.$$

By assumption, Z is uniformly distributed over $(0,d)$. By elementary calculus, the expected value of $\sin\varphi_2$ over this distribution of Z is equal to

$$\begin{aligned} & d^{-1}\int_0^{(\frac{1}{2}l+t)d/l}\sin[\tan^{-1}\{z/(\tfrac{1}{2}l+t)\}]\,dz + d^{-1}\int_{(\frac{1}{2}l+t)d/l}^{1}\sin[\tan^{-1}\{(d-z)/(\tfrac{1}{2}l-t)\}]\,dz \\ &= d^{-1}(\tfrac{1}{2}l+t)\int_0^{d/l}\sin(\tan^{-1}u)\,du + d^{-1}(\tfrac{1}{2}l-t)\int_0^{d/l}\sin(\tan^{-1}u)\,du \\ &= d^{-1}l\,\{(1+d^2/l^2)^{\frac{1}{2}}-1\} \qquad (6.1) \end{aligned}$$

By symmetry, the expected value of $\sin\varphi_1$, and hence the expected probability of detection conditional on s,t and φ, will also be equal to the expression in (6.1). Note that this probability is independent of t and only depends on the tube length l. Letting $a_{3D}(s,\varphi)$ be the probability that an emission in tube (s,φ) is actually detected, it follows from (6.1) that

$$a_{3D}(s,\varphi) = \{4(1-s^2)d^{-2}+1\}^{\frac{1}{2}} - 2(1-s^2)^{\frac{1}{2}}d^{-1}.$$

This quantity increases as s increases, reflecting the fact that emissions in shorter tubes (large s) are more likely to be detected. We have, finally,

$$0<(1+4d^{-2})^{\frac{1}{2}}-2d^{-1} \le a(s,\varphi) \le 1 \text{ for all } s\in[0,1].$$

6.2 Attenuation

The other effect we shall consider is *attenuation*, defined as being the loss of a detection caused by the absorption or scattering of one of the photons in flight. Let us model the probability of such loss of a photon as it travels between $\mathbf{x}$ and $\mathbf{x}+d\mathbf{x}$ as $\mu(\mathbf{x})|d\mathbf{x}|$ and assume that $\mu(\mathbf{x})$ is bounded. Suppose an emission occurs at a point $\mathbf{x}_0$ and that y is the line of flight of the emitted photons. Let $y_+(\mathbf{x}_0)$ and $y_-(\mathbf{x}_0)$ be the half-lines of y emanating from $\mathbf{x}_0$, and assume y intersects the detector ring. By standard Poisson process theory, the probability that neither photon will be lost is given by

$$\exp\{-\int_{y_+(\mathbf{x}_0)}\mu(\mathbf{x})d\mathbf{x}\}\exp\{-\int_{y_-(\mathbf{x}_0)}\mu(\mathbf{x})d\mathbf{x}\} = \exp\{-\int_y \mu(\mathbf{x})d\mathbf{x}\} = a_A(s,\varphi), \text{ say }.$$

Just as in Section 6.1, the probability that the emission will be detected depends only on the detector tube (s,φ) and is independent of the emission's position within that tube. In general, if both effects are considered, the probability that any particular detection will not be lost will be $a_{3D}(s,\varphi)a_A(s,\varphi)$. Both effects are important in PET; intensities reconstructed ignoring them can, in practice, be too low by a factor of three in the centre of the image (F. Natterer, personal communication). A common technique for correcting for attenuation is to estimate it separately, for example by a transmission scan.

6.3 A general framework and the extension of our results

The two effects we have discussed can be combined by assuming the existence of a function $a(s,\varphi)$, $0<a(s,\varphi)\le 1$ such that a positron emission at (x_1,x_2) gives rise to a detection at (S,Φ) as defined in Section 2.1 with probability $a(S,\Phi)$ conditional on (S,Φ); with probability $1-a(S,\Phi)$ the detection is lost. It follows from this formulation that the *observed* detections will form a *biased sample* with density in detector space with respect to $d\lambda(s,\varphi)$

$$g_a(s,\varphi) = P_af(s,\varphi) = a(s,\varphi)Pf(s,\varphi)/\int_D Pf(s,\varphi')\, a(s,\varphi')d\lambda(s,\varphi').$$

Let $\mathcal{T}_B(n)$ be the class of all estimators of f based on a sample of size n from P_af, and let $r_B(n)$ be the minimax mean integrated square error over $\hat{f}$ in $\mathcal{T}_B(n)$ and f in $\mathcal{F}_{p,C}$.

Theorem 6.1. *Suppose that* $\inf_D a(s,\varphi) = a_0 > 0$, *and make the assumptions of Theorem* 3.1. *Then*

$$r_B(n) = n^{-p/(p+2)}. \tag{6.2}$$

Proof. The order of magnitude in (6.2) is of course the same as that obtained for unbiased indirect estimation in (3.5).

Suppose, first, that f is the least favourable density in $\mathcal{F}_{p,C}$ for estimation by estimators in $\mathcal{T}_I(n)$. Let $n' = [\frac{1}{2}a_0 n]$. Suppose $Y_1, Y_2, \ldots$ is an i.i.d. sequence drawn from Pf. Construct an i.i.d. sequence $Z_1, \ldots$ from $P_a f$ by including each Y_i in the sequence with probability $a(Y_i) \geq a_0$. Let $\hat{f}$ be the estimator of f based on $Z_1, \ldots, Z_{n'}$ using the minimax estimator in $\mathcal{T}_B(n')$, so that $M(\hat{f};f) \leq r_B(n')$. Now let N be the number of $Y_1, \ldots, Y_n$ that are included in the Z_j sequence, and let $\hat{f}_1$ be equal to $\hat{f}$ if $N \geq n'$ and 1 otherwise. Since $\hat{f}_1$ is based on $Y_1, \ldots, Y_n$, and f is least favourable for $\mathcal{T}_I(n)$, we have $M(\hat{f}_1;f) \geq r_I(n)$. By an elementary argument, $M(\hat{f}_1;f) \leq M(\hat{f};f) + P(N<n')\int(f-1)^2 d\mu$, so that $r_B(n') \geq r_I(n) - P(N<n')$, making use of Proposition 2.3 and the assumption $C < 2^{\frac{1}{2}(p-1)}$ to bound $\int(f-1)^2$ by 1. A crude bound now suffices for $P(N<n')$; since N is stochastically larger than a $Bi(n,a_0)$ random variable, $P(N<n') \leq P\{Bi(n,a_0) \leq \frac{1}{2}na_0\} = O(n^{-1})$ by Chebyshev's inequality. We conclude that $r_B([\frac{1}{2}na_0]) \geq r_I(n) - O(n^{-1})$.

Now reverse the role of biased and unbiased samples throughout the argument. If $Z_1, \ldots$ is an i.i.d. sample from $P_a f$, then a sample $Y_1, \ldots$ from Pf can be constructed by including each Z_i with probability $a_0/a(Z_i)$; this quantity necessarily lies between a_0 and 1. The analogous argument to that used above yields that $r_I([\frac{1}{2}na_0]) \geq r_B(n) - O(n^{-1})$. Applying Theorem 3.1 it now follows that $r_B(n)$ has the same order of magnitude $n^{-p/(p+2)}$ as $r_I(n)$. ☐

There is of course a distinction between a biased sample of n observations drawn from $P_a f$ and a censored sample consisting of all the observations that are detected arising from n emissions in brain space. The censored sample will consist, in the notation of the proof of Theorem 6.1, of N observations from $P_a f$. Implicit in the proof of Theorem 6.1 is a demonstration that the minimax mean integrated square error for estimation based on this censored sample will have the same order of magnitude as $r_I(n)$ under the assumption $a_0 > 0$.

For the third dimension effect, as the detector ring thickness $d \to 0$, we have $a_{3D}(s,\varphi) \sim \frac{1}{2}d(1-s^2)^{-\frac{1}{2}}$ and $a_0 \to 0$. In the limiting case, the biased sample density will be proportional to $(1-s^2)^{-\frac{1}{2}}Pf(s,\varphi)$, whose ratio to $Pf(s,\varphi)$ is unbounded as $s \to 1$. Theorem 6.1 no longer applies, but it can be shown that the biased sampling has at most a logarithmic order effect, in that the order of magnitude of $r_B(n)$ lies between $(n \log n)^{-p/(p+2)}$ and $n^{-p/(p+2)}$. This is a consequence of the following more general result on singular biased sampling, whose proof is omitted.

Theorem 6.2: *Suppose* $p \geq 1$ *and* $0 < C < 2^{\frac{1}{2}(p-1)}$.

(a) *If* $\int_D a(s,\varphi)^{-1}(1-s^2)^{-1} d\lambda(s,\varphi) < \infty$, *then there exists* c_1 *such that* $r_B(n) \leq c_1 n^{-p/(p+2)}$.

(b1) *If* $\int_D a(s,\varphi)(1-s^2)^{-1} d\lambda(s,\varphi) < \infty$, *then there exists* c_2 *such that* $r_B(n) \geq c_2 n^{-p/(p+2)}$.

(b2) *If* $\int_D a(s,\varphi) d\lambda(s,\varphi) < \infty$, *and* $\sup_s (1-s^2)^{\frac{1}{2}} \int_0^{2\pi} a(s,\varphi) d\varphi < \infty$, *then there exists* c_3 *such that* $r_B(n) \geq c_3 (n \log n)^{-p/(p+2)}$.

For $a(s,\varphi) = (1-s^2)^{-\frac{1}{2}}$, the conditions of (a) and (b2) hold but the integral in (b1) is infinite.

7. Alternative error measures

Our results can be extended to some more general measures of the discrepancy between the estimator and the unknown function than mean integrated square error. We can treat a class of losses that takes into account the closeness of derivatives, as well as values, of the estimate to those of the true unknown function; these losses take more account of the "shape" of the function than does ordinary mean integrated square error.

Define measures μ_p as in Proposition 2.2. It is noted in the Appendix that, for integers $q \geq 0$, the squared norm

$$\int |f|^2 d\mu_1 + \sum_{r_1+r_2=q} \int |\partial^{r_1+r_2} f / \partial x_1^{r_1} \partial x_2^{r_2}|^2 d\mu_{q+1} \tag{7.1}$$

is equivalent to

$$\|f\|_q^2 = \sum_{j,k \geq 0} (j+1)^q (k+1)^q f_{jk}^2. \tag{7.2}$$

For non-integer values of q, the norm $\|\cdot\|_q$ will be a more general Sobolev norm (Adams, 1975), although some care will be necessary because of the nonstandard dominating measures μ_q; this is a topic for future investigation.

We can now state and prove a theorem that gives the exact minimax rates in the $\|\cdot\|_q$ norm for both direct and indirect estimation. Theorem 3.1 is the special case $q=0$ and it can be seen that the rates available are both reduced, in a natural way, when higher order norms are used.

Theorem 7.1: *For fixed* $p \geq 1$, $0 < C < 2^{\frac{1}{2}(p-1)}$, *and* $0 \leq q < p$,

$$\inf_{\hat{f} \in \mathcal{T}_D(n)} \sup_{f \in \mathcal{F}_{p,C}} E\|\hat{f}-f\|_q^2 \asymp (\log n/n)^{(p-q)/(p+1)}$$

and

$$\inf_{\hat{f} \in \mathcal{T}_I(n)} \sup_{f \in \mathcal{F}_{p,C}} E\|\hat{f}-f\|_q^2 \asymp (1/n)^{(p-q)/(p+2)}.$$

Proof The proof is analogous to that of Theorem 3.1 and we shall confine ourselves to a brief outline of the necessary changes. Define $c_\nu^2 = c_{jk}^2 = (j+1)^q (k+1)^q$ and $\Gamma = \mathrm{diag}(c_\nu)$. To obtain upper bounds, define the surrogate risk $M_q^*(\hat{f};f) = \sum_\nu c_\nu^2 \{\mathrm{var}_1 \hat{f}_\nu + (E_f \hat{f}_\nu - f_\nu)^2\}$. The result corresponding to Proposition 2.1 is immediate. As in Lemma 4.1, $M_q^*(\hat{f};f) = n^{-1} \mathrm{tr}\, \Gamma W (I - \mathrm{e}_0 \mathrm{e}_0{}^{\mathrm{T}}) W^{\mathrm{T}} \Gamma + \|\Gamma(I - WB)\mathbf{f}\|^2$. As in Lemma 4.2, the minimax surrogate risk for linear estimates over the ellipsoid $\mathcal{F}_{p,C}$ is given by $n^{-1} \sum_{(\eta)} (c_\nu^2 b_\nu^{-2} - c_\nu b_\nu^{-2} a_\nu \gamma^{\frac{1}{2}})$, where γ is chosen to satisfy $n^{-1} \sum_{(\eta)} (a_\nu c_\nu b_\nu^{-2} \gamma^{-\frac{1}{2}} - a_\nu^2 b_\nu^{-2}) = C^2$. So long as $q < p$, we obtain surrogate linear minimax rates of convergence equal to $(n/\log n)^{-(p-q)/(p+1)}$ and $n^{-(p-q)/(p+2)}$ in the direct and indirect cases respectively. Clearly it would be possible to obtain more precise results corresponding to Theorem 3.2, but we shall refrain from doing so.

The methods of Section 5 show that these are in fact the exact minimax rates of convergence for the $\|\cdot\|_q$ norm for general estimators. From Proposition 5.1, it is only necessary to compute the modulus $\sigma(\delta)$ of (5.1), now with respect to the $\|\cdot\|_q$ norm on H. This calculation goes as in Propositions 5.2 and 5.3, even using the same definitions of the subspaces M_η. Since the $\|\cdot\|_q$ norm is now used on H, $\|P\|_{M_\eta}^2 = \sup\{b_\nu^2 c_\nu^{-2} : \varphi_\nu \in M_\eta\}$ and $B_{M_\eta}(\delta) \subset H_0$ if $\eta^{p-q} \leq C^2/\delta^2$. With these changes, the proof is completed. □

8. Some concluding remarks

This paper has focused on lower and upper bounds for one particular bivariate density estimation problem for indirect data. The same formalism applies to many other density and regression estimation problems. The celebrated "unfolding" problem for sphere size distributions is an example involving univariate density estimation from indirect data and the singular value decomposition of the Abel transform. For recent results and further references on this problem, see, for example, Hall and Smith (1988), Nychka and Cox (1984), Silverman et al. (1988) and Wilson (1988).

Noisy integral equations of the form $y_i = (Pf)(t_i) + \varepsilon_i$ can be treated using our methods, at least under appropriate assumptions on the distributions of (t_i, ε_i). For example, if the observation points t_i follow a known distribution $\lambda(dt)$ and the errors

ε_i are independently Gaussian $(0,\sigma^2)$, then the information divergence between the hypotheses f_1 and f_2 is $K(P_{f_1}^{(n)}, P_{f_2}^{(n)}) = \frac{1}{2}n\int\{Pf_1(t)-Pf_2(t)\}^2\lambda(dt)$, so that the lower bound methods of Section 5 immediatedly apply. Upper bound results are given, for example, by Nychka and Cox (1984).

For a generic one-dimensional problem with singular value decomposition $P\varphi_\nu = b_\nu\varphi_\nu$, $b_\nu \sim \nu^{-\beta}$, and with ellipsoid determined by $a_\nu^2 = \nu^{2\alpha}$, corresponding to "α derivatives", the exact minimax rate of convergence of the mean square error in $n^{-2\alpha/(2\alpha+2\beta+1)}$. This should be compared with the exact rate of $n^{-2\alpha/(2\alpha+1)}$ for the corresponding "direct" case. Related calculations for a large class of one-dimensional convolution equation models appear in Wahba and Wang (1987).

One important topic for future attention is the effect of the discretisation of detector space due to the finite size of the detectors. It is clear intuitively that if the number of detectors is sufficiently large relative to the size of the sample collected, then the minimax rates will not be affected, and of course it would be interesting to quantify this notion more precisely. Some PET machines (see, for example, Snyder and Politte, 1983) are able to use time-of-flight information to provide an approximate indication of the place in the detector tube where an emission occurs. This is usually accompanied by a loss in detector efficiency and hence a smaller sample size n. It would be desirable to extend our framework to make a quantitative evaluation of this trade-off. Kaufman, Morgenthaler and Vardi (1983) report some earlier work on this issue.

Another issue that could be explored is the further extension of our results to deal with more general metrics on images. Finally there is very little known about the theoretical performance of algorithms commonly used in practice; our results at least provide a framework and a benchmark against which particular algorithms can be judged.

APPENDIX

Proof of Proposition 2.1: The proof is elementary. Consider the direct case first. Suppose Ξ is a random variable drawn from the uniform density and that X is a random variable with density f. Then $\mathrm{var}_f\hat{f}(x)/\mathrm{var}_1\hat{f}(x) = n^{-1}\mathrm{var}\,w(x,X)/n^{-1}\mathrm{var}\,w(x,\Xi)$. Now

$$\begin{aligned}\mathrm{var}\,w(x,X) &\le E\{w(x,X)-Ew(x,\Xi)\}^2 = \int\{w(x,\xi)-Ew(x,\Xi)\}^2 f(\xi)\,d\mu(\xi)\\ &\le \sup_B f\int\{w(x,\xi)-Ew(x,\Xi)\}^2 d\xi = \sup_B f\,\mathrm{var}\,w(x,\Xi)\ ,\end{aligned}$$

and similarly $\mathrm{var}\,w(x,\Xi) \le \sup_B(1/f)\,\mathrm{var}\,w(x,X) = \mathrm{var}\,w(x,X)\inf_B f$. Thus $\mathrm{var}_f\hat{f}/\mathrm{var}_1\hat{f}$, and hence M/M^*, is bounded between $\inf f$ and $\sup f$. In the indirect case an exactly analogous argument bounds M/M^* between $\inf_D Pf$ and $\sup_D Pf$. It follows from (2.1) that $\inf_D Pf \geq \inf_B f$ and that $\sup_D Pf \leq \sup_B f$, completing the proof.

Proof of Proposition 2.2: We employ Gegenbauer (ultraspherical) polynomials as normalised in Gradshteyn and Ryzhik (1980, p.827). An orthogonal basis for $L^2(B,\mu_\alpha)$ is given by the polynomials

$$\tilde{\varphi}_{jk}^{\alpha}(\mathbf{x}) = (2\pi)^{-1}\int_0^{2\pi} e^{i(j-k)\theta} C_{j+k}^{\alpha}(\mathbf{u}_\theta{}^{\mathrm{T}}\mathbf{x})\,d\theta \qquad j,k\geq 0,\ \mathbf{u}_\theta=(\cos\theta,\sin\theta)^{\mathrm{T}}.$$

Defining the operator $(P_\alpha f)(s,\theta)=E\{f(\mathrm{X})|u_\theta{}^{\mathrm{T}}\mathrm{X}=s\}$ where $\mathrm{X}\sim\mu_\alpha$, the polynomials $\tilde{\varphi}_{jk}^{\alpha}$ are the pullback by the adjoint P_α^* of the singular functions $e^{i(j-k)\theta}C_{j+k}^{\alpha}(s)$ of $P_\alpha P_\alpha^*$. This construction of the SVD of P_α is explained in Johnstone (1988), following Davison (1981, 1983) but with different notation and normalisations. It can be shown that $\tilde{\varphi}_{jk}^{1} = (j+k+1)^{-\frac{1}{2}}\varphi_{jk}$, so that $f=\sum(j+k+1)^{\frac{1}{2}} f_{jk}\tilde{\varphi}_{jk}^{1}$.

Let $D_z = \frac{1}{2}(\partial/\partial x_1 - i\partial/\partial x_2)$ and $D_{\bar{z}} = \frac{1}{2}(\partial/\partial x_1 + i\partial/\partial x_2)$. Since $(d/dt)C_m^{\alpha}=2\alpha C_{m-1}^{\alpha-1}$, we have, setting $\tilde{\varphi}_{jk}=0$ if j or $k<0$, $D_z\,\tilde{\varphi}_{jk}^{\alpha} = \alpha\tilde{\varphi}_{j-1,k}^{\alpha+1}$ and $D_{\bar{z}}\,\tilde{\varphi}_{jk}^{\alpha} = \alpha\tilde{\varphi}_{j,k-1}^{\alpha+1}$. The raising of the index from α to $\alpha+1$ leads us from the original measure μ of Section 2 to the family μ_{p+1}, so that, for example, the family of derivatives $D_z\varphi_v$ and $D_{\bar{z}}\varphi_v$ is orthogonal with respect to μ_2, not μ_1.

It is shown in Johnstone (1989) that if $r+s=p$, then, for certain constants c_{jkrs} all falling in $[(p+1)^{-2p+1},(p+1)p^{2p}]$,

$$\begin{aligned}\int(D_z^r D_{\bar{z}}^s f)^2 d\mu_{p+1} &= p!\sum_{\substack{j\geq r\\ k\geq s}} (j+k+1)f_{jk}^2\int(\tilde{\varphi}_{j-r,k-s}^{p+1})^2 d\mu_{p+1}\\ &= \sum_{\substack{j\geq r\\ k\geq s}} c_{jkrs}^2 (j+1)^p(k+1)^p(p+1)f_{jk}^2 \ .\end{aligned}$$

Standard arguments of analysis complete the proof.

Proof of Proposition 2.3: In the complex form of the basis, we have

$$f(r,\theta) = \sum_{(j,k)\in\mathcal{N}'} f_{jk}\,(j+k+1)^{\frac{1}{2}} e^{i(j-k)\theta} Z_{j+k}^{|j-k|}(r).$$

Zernike polynomials satisfy $\sup_{0\leq r\leq 1}|Z_m^l(r)| = Z_m^l(1) = 1$, as a consequence of the representation in terms of Jacobi polynomials: $Z_{l+2s}^l(r) = r^l P_s^{(0,l)}(2r^2-1)$, together with the results of Szegö (1939, p.163), applied to the polynomials $Q_s^l(t)=P_s^{0,l}(2t-1)$ as s varies. Hence

$$\sup|f-1| \leq \sum_{\mathcal{N}'\setminus(0,0)} (j+k+1)^{\frac{1}{2}}|f_{jk}|$$

The ellipsoid $\mathcal{F}_{p,C}$ has exactly the same description in terms of either the real or the complex form of the basis. Setting $x_{jk} = (j+1)^{p/2}(k+1)^{p/2}|f_{jk}|/C$, we obtain

$$\sup_{\mathcal{F}_{p,C}} \sup|f-1| \le C \sup \{ \sum_{N'\setminus(0,0)} (j+k+1)^{\frac{1}{2}}(j+1)^{-p/2}(k+1)^{-p/2}x_{jk} : \sum_{N'\setminus(0,0)} x_{jk}^2 \le 1 \}$$
$$\le C \sup_{N'\setminus(0,0)} (j+k+1)^{\frac{1}{2}}(j+1)^{-p/2}(k+1)^{-p/2} = C\ 2^{(1-p)/2},$$

provided $p \ge 1$. To complete the proof, note that the linear function $1 + 2^{(1-p)/2} C r \cos\theta$ falls in $\mathcal{F}_{p,C}$ and satisfies $\sup |f-1| \doteq C\ 2^{(1-p)/2}$.

Acknowledgments. Grant support from NSF (MCS 84-51750, 86-00235), NIH (PHS GM 21215-12), SERC and USARO is gratefully acknowledged. For helpful conversations and useful references the authors are delighted to thank Dennis Cox, David Donoho, Peter Hall, Chris Jones, Mark Matthews, Frank Natterer, John Rice, Larry Shepp and Yehuda Vardi. The referees and an associate editor provided extremely stimulating comments on a previous version.

REFERENCES

Adams, R. A. (1975). *Sobolev Spaces.* Academic Press, New York.

Birgé, L. (1983). Approximation clans les espaces metriques et theorie de l'estimation. *Z. Wahrsch. Verw. Gebiete* **65** 181-237.

Born, M. and Wolf, E. (1975). *Principles of Optics*, 5th ed. Pergamon, New York.

Buckley, M. J., Eagleson, G. K. and Silverman, B. W. (1988). The estimation of residual variance in nonparametric regression. *Biometrika* **75** 189-199.

Budinger, T. F., Gullberg, G. T. and Huesman, R. H. (1979). Emission computed tomography. In G. T. Herman (ed.), *Image Reconstruction from Projections.* Springer-Verlag, Berlin, Heidelberg, New York.

Chesler, D. A., Rieder, S. J. and Pelc, N. J. (1977). Noise due to photon counting statistics in computed X-ray tomography. *J. Comput. Assist. Tomog.* **1** 64-74.

Cox, D. D. (1988). Approximations of least squares regression on nested subspaces. *Ann. Statist.* **16** 713-732.

Davison, M. E. (1981). A singular value decomposition for the Radon transform in *n*-dimensional Euclidean space. *Numer. Funct. Anal. and Optimiz.* **3** 321-340.

Davison, M. E. (1983). The ill-conditioned nature of the limited angle tomography problem. *SIAM J. Appl. Math.* **43** 428-448.

Deans, S. R. (1983). *The Radon Transform and some of its Applications.* Wiley, New York.

Donoho, D. and Liu, R. C. (1989). Hardest one-dimensional subproblems. Preprint, Department of Statistics, University of California-Berkeley.

Geman, S. and McClure, D. E. (1985). Bayesian image analysis: an application to single photon emission tomography. *Proc. Amer. Statist. Assoc., Statist. Comp. Section* 12-18.

Gradshteyn, I. S., and Ryzhik, I. M. (1980). *Table of Integrals, Series, and Products.* Academic Press, New York.

Hall, P. G. and Smith. R. L. (1988). The kernel method for unfolding sphere size distributions. *J. Comput. Phys.* **74** 409-421.

Ibragimov, I. A. and Hasminskii, R. Z. (1981). *Statistical Estimation, Asymptotic Theory.* Springer-Verlag, Berlin, Heidelberg, New York.

Johnstone, I. M. (1989). On a singular value decomposition for the Radon transform and smoothness classes of functions. Technical Report 000, Department of Statistics, Stanford University.

Jones, M. C. and Silverman, B. W. (1989). An orthogonal series density estimation approach to reconstructing positron emission tomography images. Submitted for publication.

Kaufman L., Morgenthaler, S. and Vardi, Y. (1983). Maximum likelihood reconstruction in emission tomography with time-of-flight information: a limited study. Technical Report ONR 33, Statistics Center, Massachusetts Institute of Technology.

Le Cam, L. (1985). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York.

Lorentz, G. G. (1966). Metric entropy and approximation. *Bull. Amer. Math. Soc.* **72** 903-937.

Marr, R. B. (1974). On the reconstruction of a function on a circular domain from a sampling of its line integrals. *J. Math. Anal. Appl.* **45** 357-374.

Natterer, F. (1980). A Sobolev space analysis of picture reconstruction. *SIAM J. Appl. Math.* **39** 402-411.

Natterer, F. (1986). *The Mathematics of Computerized Tomography.* Teubner, Stuttgart; Wiley, Chichester.

Nychka, D. W. and Cox, D. D. (1984). Convergence rates for regularized solutions of integral equations from discrete noisy data. Preprint.

Pinsker, M. S. (1980). Optimal filtration of functions from L_2 in Gaussian noise. *Probl. Peredachi Inform.* **16** 52-68.

Rockmore, A. J. and Macovski, A. (1977). A maximum likelihood approach to emission image reconstruction from projections. *IEEE Trans. Nuclear Science* **NS-32** 1428-1432.

Shepp, L. A. and Vardi, Y. (1982). Maximum likelihood reconstruction in positron emission tomography. *IEEE Trans. Med. Imaging* **1** 113-122.

Silverman, B. W., Jones, M. C., Wilson, J. D. and Nychka, D. W. (1988). A smoothed EM approach to a class of problems in image analysis and integral equations. Submitted for publication.

Snyder, D. L. and Politte, D. G. (1983). Image reconstruction from list-mode data in an emission tomography system having time-of-flight measurements. *IEEE Trans. Nuclear Sci.* **NS-30** 1843-1849.

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

Stone, C. J. (1982). Optimal global rates of convergence for non-parametric regression. *Ann. Statist.* **10** 1040-1053.

Szegö, G. (1938). *Orthogonal Polynomials.* Amer. Math. Soc. Colloq. Publications **XXIII**. Providence, RI.

Tretiak, O. J. (1978). Noise limitations in X-ray computed tomography. *J. Comput. Assist. Tomog.* **2** 477-480.

Tretiak, O. J. (1979). Fundamental limitations in radionuclide imaging. In J. Raviv, J. F. Greenleaf and G. T. Herman (eds.), *Computer Aided Tomography and Ultrasonics in Medicine*, pp. 109-122. North Holland, New York.

Tanaka, E. (1987). Recent progress on single photon and positron emission tomography — from detectors to algorithms. *IEEE Trans. Nuclear Science* **NS-34** 313-320.

Tikhonov, A. N. and Arsenin, V. Y. (1977). *Solution of Ill-Posed Problems*. Winston and Sons, Washington DC.

Vardi, Y., Shepp, L. A. and Kaufman L. (1985). A statistical model for positron emission tomography. *J. Amer. Statist. Assoc.* **80** 8-37.

Wahba, G. and Wang, Y. (1987). When is the regularization parameter insensitive to the choice of loss function? Technical Report No. 809, Department of Statistics, University of Wisconsin-Madison.

Watson, G. S. (1969). Density estimates by orthogonal series. *Ann. Math. Statist.* **40** 1496-1498.

Wilson, J. D. (1988). A smoothed EM algorithm for the solution of Wicksell's corpuscle problem. *J. Statist. Comput. Simul.*, in press.

TABLE 1

Constants needed for Theorem 3.2. Euler's constant $\gamma_E = 0.57722\ldots$.

$c_1 = \exp(-c_8/c_7)$

$c_2 = c_7^{-1}(p+1)\exp\{(p+1)c_8/c_7\}$

$c_3 = p(p+2)^{-1}$

$c_4 = 2\gamma_E-(p+4)/(p+2)$

$c_5 = (p+1)^{1/(p+1)}c_7^{p/(p+1)}$

$c_6 = \frac{1}{2}\{\pi^2 p/3(p+4)\}^{p/(p+2)}(p+2)^{2/(p+2)}$

$c_7 = p(p+1)^{-1}(p+2)^{-1}$

$c_8 = 2\gamma_E c_7-\{4(p+2)^{-2}-(p+1)^{-2}\}$

TABLE 2

Equivalent direct sample sizes $m^*(n)$ *to achieve the same surrogate linear minimax risk over smoothness class* p *as for an indirect sample of size* n.

	$n=10^7$	$n=10^8$	ratio $m(10^8)/m(10^7)$
$p=1$	1.93×10^5	1.03×10^6	5.34
$p=2$	4.85×10^5	3.12×10^6	6.44
$p=5$	1.29×10^6	1.05×10^7	8.09

Figure Captions

Fig. 2.1. The patient and the detector circle

Fig. 2.2. Parametrising the line l

Fig. 2.3. Transforming the coordinates

Fig. 6.1. The two cases for φ_2

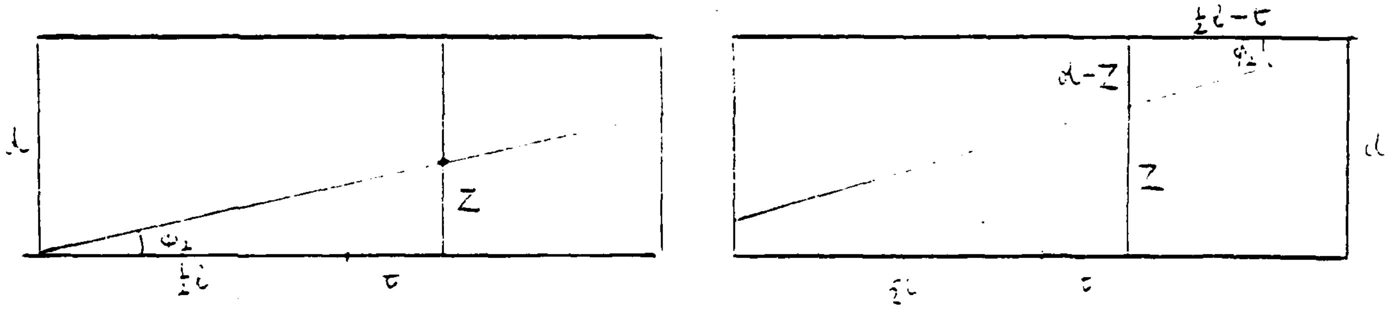


Figure 6.1: The two cases for φ_2 .

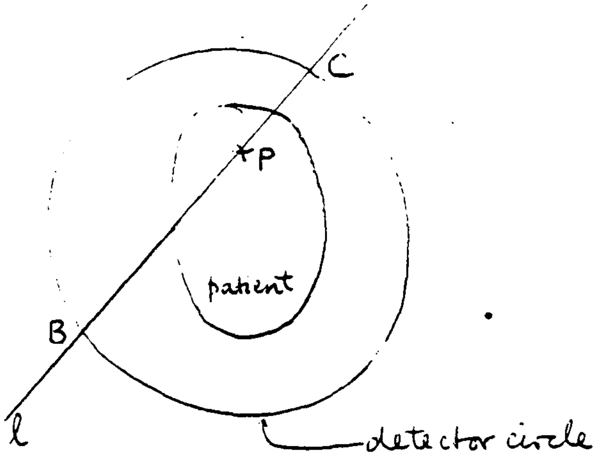


Figure 2.1 : The patient and the detector circle

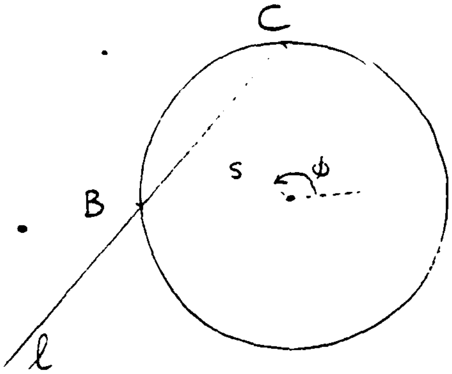


Figure 2.2 : Parametrising the line l

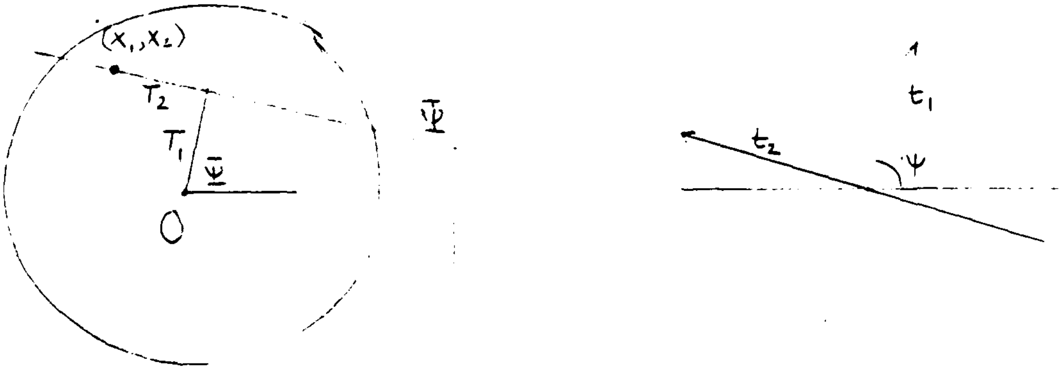


Figure 2.3 : Rotating the coordinates

Appendix 11

A smoothed EM approach to a class of problems in image analysis and integral equations.

by

B. W. Silverman, M. C. Jones, J. D. Wilson & D. W. Nychka

A Smoothed EM Approach to a Class of Problems in Image Analysis and Integral Equations

By B.W. SILVERMAN, M.C. JONES, J.D. WILSON and D.W. NYCHKA
University of Bath, UK *North Carolina State University, USA*

SUMMARY

There are many practical problems where the observed data are not drawn directly from the density g of real interest, but rather from another distribution derived from g by the application of an integral operator. The estimation of g then entails both statistical and numerical difficulties. A natural statistical approach is by maximum likelihood, conveniently implemented using the EM algorithm, but this provides unsatisfactory reconstructions of g. In this paper, we modify the maximum likelihood / EM approach by introducing a simple smoothing step at each EM iteration. In our experience, this algorithm converges in relatively few iterations to good estimates of g that do not depend on the choice of starting configuration. Some theoretical background is given that heuristically relates this smoothed EM algorithm to a maximum penalised likelihood approach. Two applications are considered in detail. The first is the classical stereology problem of determining particle size distributions from data collected on a plane section through a composite medium. The second concerns the recovery of the structure of a section of the human body from external observations obtained by positron emission tomography; for this problem, we also suggest several technical improvements on existing methodology, in particular, a new pixellation of the circular image.

Keywords: ILL POSED PROBLEMS; INDIRECT OBSERVATIONS; INTENSITY ESTIMATION; MAXIMUM LIKELIHOOD; PENALISED LIKELIHOOD; POSITRON EMISSION TOMOGRAPHY; SMOOTHING; STEREOLOGY.

Revised version, 9 Sept 88.

1. INTRODUCTION

A wide variety of practical problems, in fields ranging from medicine to remote sensing, involve *indirect observations*. Suppose that events occur on a space Y according to a nonhomogeneous Poisson process of rate $g(y)$. These events cannot be observed directly; instead, an event at a point y gives rise to an *observable* datum at a point x in some space X. The observed datapoints fall as a nonhomogeneous Poisson process on X with intensity $f(x)$, where f and g are related by the integral equation

$$f(x) = \int K(x,y)\, g(y)\, dy \ . \tag{1.1}$$

Here, $K(x,y)$ is a non-negative *kernel function* which is assumed known. In some contexts X and Y are the same space, but we shall see that this is by no means always the case.

In this paper, we shall introduce and discuss a simple general approach to the estimation of the non-negative function g from such indirect data. The general problem, and our solution to it, will be discussed in two particular contexts. The first, involving univariate functions, is the classical stereology problem of determining particle-size distributions from data collected on plane sections through a composite medium. The second is an interesting image processing problem, the recovery of the structure of a section of a radioactive emission intensity in the human body from external observations obtained by positron emission tomography (PET). Our intention is to contribute concretely to these problems and also methodologically more generally.

Equations of the form (1.1) are called *first kind integral equations* and have been the subject of much study by numerical analysts, mainly from the point of view that the function f is observed accurately. Most of this work does not take account of constraints on g. Another, more statistical, problem that has been studied is the case where values of f itself are observed subject to random error; see, for example, Titterington (1985a) and O'Sullivan (1986). The relationship of this statistical problem to the one we shall discuss is precisely that between regression and density estimation for directly observed data. The problems have some similarities but sufficient differences to make distinct approaches appropriate.

Our work builds on the paper of Vardi, Shepp and Kaufman (1985) who give a good introduction to the statistical aspects of the PET problem, and provide a method for its solution based on the EM algorithm (Dempster, Laird and Rubin, 1977, Little and Rubin, 1987). In general, a natural statistical approach to the estimation of g is by maximum likelihood (ML) and it is to the solution of the ML problem that the EM algorithm is addressed. However, as we shall see, ML reconstructions of g are unsatisfactory. As in nonparametric curve and surface estimation generally, ML yields "noisy" or "spiky" estimates that do not fully reflect knowledge about the structure of the problem under consideration, and some kind of smoothing is necessary to provide good estimates of g; see, for example, Silverman (1985a, 1986) and Besag (1986). This is because of the high or infinite dimensionality of the parameter space. In the problem we are considering, the ill-posed nature of the inversion of the integral equation (1.1) exacerbates this difficulty. The classical mathematical work on integral equations (e.g. Tikhonov and Arsenin, 1977) makes it clear that smoothing would be necessary in numerical reality even if f were observed to arbitrary accuracy, for example, from an infinite number of observations.

The EM algorithm is an iterative approach that increases the likelihood of the estimate of g at each iteration. Each stage of the algorithm consists of an E (for *expectation*) step and an M (for *maximisation*) step. Our proposal is to introduce a third S (for *smoothing*) step at each iteration where the current estimate of g is smoothed in a suitable way. The EMS algorithm maintains various advantages of the EM algorithm but appears to eliminate some of its disadvantages. Using very simple smoothing schemes, we have found that the algorithm converges, in a relatively small number of iterations, to good estimates of g. For the problems we shall discuss, a little smoothing goes a long way.

The general structure of such smoothed EM algorithms for indirect observation problems is set out in Section 2. The stereology example is discussed in detail in Section 3 and the PET example in Section 4. Our treatment of the PET example incorporates some other algorithmic improvements over those of Vardi *et al.* (1985) and others. Some theoretical background is given in Section 5 that heuristically relates the EMS algorithm to a maximum penalised likelihood approach in which the penalty term is quadratic in the square roots of the intensities.

2. THE EM ALGORITHM AND SMOOTHING

2.1 Notation and Preliminaries

For practical reasons, the data with which we are concerned arise in histogram form, so the data space X will be divided into bins. Index the data bins by $t=1,\ldots,T$ and denote the observed counts in these bins by $n(t)$, $t=1,\ldots,T$. To facilitate reconstruction, we also introduce a discretisation of the space Y into bins; let $s=1,\ldots,S$ index these "reconstruction" bins. Note that the discretisation of the data-gathering aspect of both our applications is an irremovable physical constraint, while the reconstruction discretisation arises from algorithmic considerations. We shall seek to reconstruct the discretised version of g in (1.1), denoting expected total occurrences in bin s by $\lambda(s)$, $s=1,\ldots,S$. The discretised version of the kernel function will be denoted by $p(s,t)$, $s=1,\ldots,S$, $t=1,\ldots,T$; assuming that g is constant within each bin, $p(s,t)$ is the integral of $K(x,y)$ over x in bin t and y in bin s, divided by the size of bin s. Write

$q(s) = \sum_t p(s,t)$, $s=1,\ldots,S$. Neglecting the variation of K over bin s, we have the appealing interpretation

$$p(s,t)/q(s) = \text{Prob}\{\text{datum counted in bin } t \mid \text{datum is observed, having arisen from an event in bin } s\}.$$

Define $k(s,t)$ to be the number of events occurring in bin s which contribute to the count in bin t. It is immediately clear that all the $k(s,t)$'s are independent with, for each s and t,

$$k(s,t) \sim \text{Poisson}\,\{\lambda(s)p(s,t)\}.$$

The observed data arise simply from these as $n(t) = \sum_s k(s,t)$ so that, for $t=1,\ldots,T$,

$$n(t) \sim \text{Poisson}\,\{\sum_s \lambda(s)p(s,t)\}, \tag{2.1}$$

independently for each t. On the other hand, an important set of unobservables is the s-bin counts, $m(s) = \sum_t k(s,t)$, $s=1,\ldots,S$. All these m's are also mutually independent and distributed as:

$$m(s) \sim \text{Poisson}\,\{\lambda(s)q(s)\}.$$

Define $\mathbf{m} = (m(1),\ldots,m(S))^T$, $\mathbf{n} = (n(1),\ldots,n(T))^T$, $\boldsymbol{\lambda} = (\lambda(1),\ldots,\lambda(S))^T$ and $\mathbf{k}$ to be the $(S\times T)$ matrix with (s,t)'th element $k(s,t)$.

2.2 The EM Algorithm

Consider the estimation of $\boldsymbol{\lambda}$ by maximising the log likelihood, $l(\mathbf{n}\,|\,\boldsymbol{\lambda})$, based on the data $\mathbf{n}$. These data can be regarded as an incomplete version of the complete data, $\mathbf{k}$, which we would like to have been able to observe. Dempster *et al.*'s (1977) EM algorithm, applied to the PET version of the present context by Vardi *et al.* (1985), then gives a two step iteration for increasing the likelihood of a current estimate $\boldsymbol{\lambda}^{(i)}$ of $\boldsymbol{\lambda}$. In the E step for the current problem, we find the expected value of the complete data, given the incomplete data, under the current estimate of parameter values; in the M step, we find the ML estimate of the parameters using the estimated complete data from the E step. From Vardi *et al.* (1985), this gives:

E STEP

$$\text{compute } \hat{k}(s,t) = n(t)\,\frac{\lambda^{(i)}(s)p(s,t)}{\sum_r \lambda^{(i)}(r)p(r,t)} \text{ for each } s \text{ and } t,$$

M STEP

set $\lambda^{(i+1)}(s) = \sum_t \hat{k}(s,t)/q(s)$ for each s.

The two steps combine to give the simple updating formula

$$\lambda^{(i+1)}(s) = \lambda^{(i)}(s) \sum_t n(t) \frac{p(s,t)/q(s)}{\sum_r \lambda^{(i)}(r)p(r,t)} \tag{2.2}$$

for $s=1,\dots,S$. An even simpler interpretation of EM is possible in this case, most obviously if we treat **m** rather than **k** as the complete data: estimate **m** by *its current* expectation, $\hat{\mathbf{m}}^{(i)}$, given $\lambda^{(i)}$ and **n**, use $\hat{\mathbf{m}}^{(i)}$ as the "new data", then iterate.

As well as its conceptual simplicity, the EM algorithm has other apparently appealing properties. First, it necessarily increases the log likelihood at each iteration (Dempster *et al.*, 1977) and, since the log likelihood at (2.1) is concave, convergence of the algorithm is guaranteed in theory. There is *not* usually, however, a unique ML solution (certainly not when $S > T$), so the EM converges to *one* of the reconstructions maximising the likelihood, that one depending on the choice of the initial values $\lambda^{(0)}$. A second consideration is that each $\lambda^{(i)}(s)$ is automatically non-negative provided the initial image is. Taking account of this non-negativity constraint can be important; see, for example, Bertero and Dovi (1981). In methods other than EM, this constraint needs to be incorporated at considerable cost in computational complexity or else ignored with detrimental repercussions for quality of reconstruction. Thirdly, we note, as do Vardi *et al.* (1985), that the EM updating formula (2.2) also arises directly from the likelihood at (2.1) as an iterative solution to the Kuhn-Tucker conditions for a maximum. The EM is just one possible optimisation algorithm for this problem and the question arises whether there are advantages to be had using an alternative optimisation technique. Kaufman (1987) investigates this in the PET context. Although it is possible to accelerate the optimisation in its early stages, the EM proves to be a sensible approach to the computation of ML estimates. Kaufman (1987) argues that it can be thought of as a "preconditioned" steepest ascent method, having properties similar to steepest ascent in many situations and considerably improving on it in others.

Vardi *et al.* (1985), however, found that in practice the convergence to an ML estimate is exceedingly slow. Furthermore, as the iterations proceeded beyond a certain point, the quality of the reconstructions actually deteriorated, and we shall see, in the computationally simpler problem of Section 3, that an ML estimate itself is unsatisfactory. Their proposed solution was to start with a uniform image and to abandon any attempt to iterate to convergence; instead, they terminate the process after a chosen number of steps (probably a long way from convergence). In this way, Vardi *et al.* (1985) obtained pleasing reconstructions for the PET problem.

We feel it is philosophically more satisfactory to abandon the aim of finding ML estimates altogether and to replace the technique just described by an explicit smoothing procedure. Also, we seek estimates that are the realisable limits of an algorithm that actually converges in a reasonably small number of iterations and that yields results independent of the starting configuration. We must stress, however, that we wish to build on, and not to disparage, the very important work of Vardi *et al.* (1985), without which the present paper would not have been possible. Indeed, Vardi *et al.* themselves suggested that some smoothing might improve PET reconstructions.

2.2.1 The EM$_p$ Algorithm

In order to provide smoother estimates of λ than those given by ML, an appealing approach is regularisation or penalised maximum likelihood (see, for example, Silverman, 1985b, and Titterington, 1985b): instead of maximising $l(\mathbf{n}|\lambda)$, maximise

$$l(\mathbf{n}|\lambda) - R(\lambda). \tag{2.3}$$

Here, $-R(\lambda)$ might be interpreted as a log prior density for λ in a Bayesian framework or as a penalty term which discourages roughness in a penalised likelihood approach. Choosing $\hat{\lambda}$ to maximise (2.3) can in principle be achieved by EM methods too, as noted by Dempster *et al.* (1977), to give, say, an EM$_p$ algorithm:

E STEP as for EM,

M$_p$ STEP

find $\lambda^{(i+1)}$ by maximising $\sum_s \sum_t [\, \hat{k}(s,t) \log\{\lambda(s)p(s,t)\} - \lambda(s)p(s,t)\,] - R(\lambda)$.

Repeating E and M$_p$ steps affords convergence to a maximum penalised likelihood solution as required; what is more, convergence can be expected to be rather quicker than the basic EM.

Computational considerations, however, militate against performing the M$_p$ step at each iteration of an EM$_p$ algorithm; comparison with the trivial M step of EM for Poisson likelihoods emphasises the extra burden. The M$_p$ step involves a full penalised likelihood reconstruction for the case where the data depend on the intensity function of interest through a Poisson likelihood. In any context where λ is a pixel image, the important work of Greig, Porteous and Seheult (1986) casts doubt on the existence of any method at present for actually achieving the penalised likelihood solution, although of course it would be an interesting avenue for investigation to apply the image processing methods of Besag (1986) in an M$_p$ step.

2.3 The EMS Algorithm

Our proposed approach is slightly *ad hoc*, but is very straightforward. We suggest introducing a further step which smooths the result of E and M steps in a simple way. This gives an EMS algorithm:

E STEP as for EM,

M STEP as for EM, except call the output $\mu^{(i+1)}$, say,

S STEP

smooth $\mu^{(i+1)}$ to give $\lambda^{(i+1)}$.

For the problems of interest in this paper, this becomes the iteration: update by (2.2) and smooth. (If, as in the PET application of Section 4, the reconstruction bin sizes $\{a(s)\}$ differ, apply the smoother to the $\mu^{(i+1)}(s)/a(s)$ values then multiply the resulting values by the corresponding a's to get $\lambda^{(i+1)}$.) This EMS approach is the major tool used throughout the rest of the paper.

Choice of appropriate smoothing method is problem dependent and will be considered in Sections 3 and 4 although it turns out that similar methods in both contexts prove useful even though the perception of what constitutes "smooth" is somewhat different in the two problems. Sensible smoothing schemes should retain automatic non-negativity. We no longer have an appealing direct interpretation of a reconstruction obtained by EMS in terms of the solution of a specified optimization problem, although the work of Section 5 yields a heuristic relationship with such an approach. This backs up our empirical evidence which suggests that sensible smoothing regimes allow the EMS algorithm to converge, and that at an increased rate compared with EM, due to the smoothing. Moreover, it seems from our empirical experience that we can expect convergence to a unique solution.

3. A FIRST APPLICATION : STEREOLOGY

3.1 The Problem

A classical problem in stereology is the following. A three-dimensional specimen consists of some translucent material in which are situated a number of opaque non-overlapping spheres. Interest centres on the size distribution of these spheres; in an example considered later, they represent tumours in the liver of a mouse. It is not possible to observe the three-dimensional internal structure directly. Rather, a thin slice is taken through the specimen at some random orientation. When this section is examined, usually under a microscope, a number of circles is observed, each corresponding to a slice through one of those spheres which happen to be cut by the section. Our aim is to recover the intensity of the radii of the spheres in the medium from this sample of circle radii.

We make the standard assumption (that cannot be more than approximately true) that the centres of the spheres are distributed according to a three-dimensional Poisson process with constant intensity. The sphere radii are bounded above by Y, say, a constant determined by the practical context. A further practical constraint introduces a lower bound ε, say; circle radii smaller than ε cannot physically be observed. Thus, we are concerned with a truncated circle radius intensity $f(x)$, $x \in \mathbf{X} = [\varepsilon, Y]$, and seek to reconstruct a similarly truncated version $g(y)$, $y \in \mathbf{Y} = \mathbf{X}$, of the sphere radius intensity. Both ε and Y are assumed known. The relationship between f and g can be written in a form directly comparable with (1.1):

$$f(x) = c_\varepsilon \int_\varepsilon^Y \frac{I_{[x,Y]}(y)\, x}{(y^2-x^2)^{\frac{1}{2}}}\, g(y)\, dy. \tag{3.1}$$

Here, $I_\Phi(y)$ is the indicator function (1 if $y \in \Phi$, 0 otherwise) and c_ε is a constant. This equation was first derived by Wicksell (1925) for the case $\varepsilon = 0$ and extended to $\varepsilon \neq 0$ in many subsequent papers (see Cruz-Orive, 1983). The ill-posedness of the kernel function in (3.1) arises, intuitively, because circles of a given radius can be obtained from sections through spheres of any radius larger than that observed.

Discretisation of (3.1) proceeds exactly as set out in Section 2. All quantities defined there transfer directly to the stereology context and we retain the same notation in this section; s-bin quantities are now to do with sphere radii, t-bins with circle radii. In the real data example treated briefly in Section 3.4, circle radii were, indeed, recorded in binned form only; these bins and our reconstruction bins are all of equal width. The form of the kernel in (3.1) allows exact computation of the $p(s,t)$'s in a straightforward manner.

Alternative approaches to estimating g in (3.1) are reviewed by Cruz-Orive (1983) and Nychka *et al.* (1984). Cruz-Orive (1983) also discusses some other practical difficulties which, for simplicity, we have omitted. Not all that many previously proposed solutions have been statistical in nature and, of these, very few have resulted from a nonparametric approach. A notable exception is the method proposed by Nychka *et al.* (1984) which is discussed in Section 3.3; this paper inspired much of the simulation and practical work reported here. Further details of the application of the EMS algorithm to the stereology problem and more empirical evidence are reported in Wilson (1987).

3.2 EM and EMS Reconstructions for Simulated Data

In this section, we apply EM and EMS algorithms to simulated data from the stereology problem. Following Nychka *et al.* (1984), we chose $\varepsilon = 0.04$ and $Y = 0.4$ and considered two particular choices of g: appropriately truncated versions of a

Weibull density, $g(y)=\alpha\beta y^{\beta-1}\exp(-\alpha y^{\beta})$, with parameters $\alpha=12$ and $\beta=0.1$, and of a mixture of two normals, one with mean 0.15 chosen with probability 0.7, the other with mean 0.275 and both with standard deviation 0.03. This scaling (in millimeter units) and the Weibull density were chosen to imitate theoretically the real data situation described in Section 3.4; this is Nychka *et al.*'s "Experiment 1". The bimodal normal mixture follows "Experiment 3" of Nychka *et al.* and was chosen to test the ability of the reconstruction methods to recover distinct peaks in an intensity. It is not difficult to generate data (from f) by mimicking the physical process: choose a candidate sphere radius from the distribution with density g, decide whether this sphere was cut by a random section using an acceptance/rejection technique (resulting in a sphere radius from the length-biased distribution corresponding to g) and then determine the corresponding circle radius by slicing the sphere at a uniformly distributed distance from its centre. For further details, see Wilson (1987). Again to be roughly comparable with the work of Nychka *et al.* (1984), we generated an average of 190 circle radii in each simulation of the Weibull case and 330 for the normal mixture. These data were grouped into $T=50$ bins.

3.2.1 EM Reconstructions

Typical ML estimates, using $S=50$ reconstruction bins, are shown in Figs 3.1 and 3.2 for the Weibull and normal mixture cases, respectively. In these and all remaining figures in Section 3, g is represented by a broken line and the estimate of g by a solid line. The spiky nature of these EM reconstructions has already been alluded to; Figs 3.1 and 3.2 are genuinely representative of the kind of reconstruction always preferred by ML and are clearly unacceptable as estimates of g. Incidentally, in this instance early termination of the EM algorithm, even though started from a uniform initial allocation, is not an effective remedy.

3.2.2 Local Smoothing

Smoothness of an intensity function can be defined in a number of ways. For current purposes, however, a heuristic notion of smooth as (in binned form) values in neighbouring bins "differing little" will suffice. We propose using a very simple smoother to have this effect; we claim no "optimality" properties for our choice, but appeal to its practical effectiveness and simplicity as justification for its use. The scheme is a weighted average of a bin value and the values of its nearest neighbours, using binomial weighting factors. Recalling the notation used in the definition of the EMS algorithm in Section 2, we set

$$\lambda^{(i+1)}(s) = 2^{-2j} \sum_{r=-j}^{j} \binom{2j}{r+j} \mu^{(i+1)}(s+r).$$

Typically, such J-point smoothing, where $J = 2j+1$, is used for $J = 3,5,7,9$ or perhaps 11; the greater is J, the more smoothing is applied. Various modifications are possible at the ends of the range of bins; Wilson (1987) describes the one used here.

3.2.3 EMS Reconstructions

In this form, EMS maintains the EM property of automatically scaling successive estimates so that $\sum_s \lambda^{(i)}(s) = N$ for $i \geq 1$, where $N = \sum_t n(t)$ is the total number of circles in the section. We then normalise and join the estimated values at bin midpoints by straight lines to obtain a frequency polygon, calling this $\hat{g}$; it is the density estimate displayed in the figures.

Employing the EMS algorithm with 5-point smoothing to the normal mixture data which gave rise to Fig. 3.2 produces Fig. 3.3. The improvement in quality of reconstruction with the introduction of smoothing is strikingly dramatic. Indeed, this EMS reconstruction provides an excellent estimate of g.

Not all EMS estimates, however, provide quite such good reconstructions. To measure the discrepancy between g and $\hat{g}$, we essentially use the L_1 distance $\int_{\varepsilon}^{Y} |g(y) - \hat{g}(y)|\, dy$, approximated by

$$A = b \sum_{s=1}^{S} |g_s - \hat{g}_s|,$$

where g_s and $\hat{g}_s$ are the values of g and $\hat{g}$ at the midpoints of the s-bins, and $b = S^{-1}(Y - \varepsilon)$ is the bin width. In all, ten different datasets (of essentially the same size) were generated from the normal mixture model and EMS reconstructions (with $J = 5$) performed. According to A, $\hat{g}$ of Fig. 3.3 is the second best of the ten ($A = 0.1251$), the best having $A = 0.1198$ and the worst corresponding to $A = 0.3243$. This worst reconstruction is shown in Fig. 3.4. One striking feature in this picture is the poor behaviour of $\hat{g}$ near ε. This effect was observed in a minority of cases and appears to be due to an inherent numerical and statistical instability, possibly connected with the lack of information at small circle radii. Nychka *et al.* (1984) noted the same phenomenon; Wilson (1987) shows that the difficulty sometimes disappears if the data are re-binned. The other disappointing aspect of this $\hat{g}$ is its behaviour where there is a trough in g; having said that, there is certainly still some hint of the underlying bimodality or, at least, of a strong indication that g is not unimodal. Most of the ten simulated datasets resulted in rather better estimates of g, however.

A further important advantage of the EMS algorithm over basic EM is also well illustrated by these simulations, namely, an enormous improvement in the computer time taken to reach the solution. Using the convergence criterion "stop as soon as

$\|\lambda^{(i+1)} - \lambda^{(i)}\|^2 < 10^{-6} \|\lambda^{(i)}\|^2$", the EM reconstruction of Fig 3.2 took 484 iterations to complete; the EMS reconstructions of Figs 3.3 and 3.4 took just 39 and 29 iterations, respectively. Since the binomial smoothing step adds only a very small extra computational burden to each iteration, these savings are impressive. Uniqueness of EMS reconstructions also seems to hold: in experiments with different starting configurations, we have never obtained more than a single solution per dataset.

We have not considered automatic choice of the smoothing parameter J; rather, a more subjective approach has been found to work well. Reconstructions using $J = 3$, then $J = 5$, 7 etc., can be looked at in turn, the process stopping when major features in the estimate start to disappear. In practice, only a very few (at most 4 or 5) such reconstructions need to be calculated; that even this is not computationally overdemanding follows from the excellent convergence rates discussed above.

An EMS reconstruction in the Weibull case is shown in Fig. 3.5; 9-point smoothing turned out to be suitable here. Fig. 3.5 is based on the same dataset as the EM reconstruction of Fig. 3.1; the vast improvement brought about by the smoothing is again impressive. Ten datasets were simulated in this case, too; Fig. 3.5, with $A = 0.1947$, is only the seventh best estimate of these, thus demonstrating that a good correspondence between true and estimated densities is quite typical of our Weibull reconstructions. Even in the worst cases, the estimate of the density's tail is pleasingly accurate and the reconstructions always indicate an increase in density near ε; serious discrepancies arise only in the estimate of the magnitude of this effect. The EM reconstruction of Fig. 3.1 took 328 iterations to arrive at; typically, EMS reconstructions — here with a greater degree of smoothing than in the normal mixture case — took fewer than 20 iterations each to converge.

3.3 Remarks on Nychka et al. *(1984)*

The reconstructions of Section 3.2.3 can be compared with those of Figs 7 and 9 of Nychka *et al.* (1984). The immediate impression is of a broad similarity of the results of the two approaches; that our reconstructions are certainly no worse than Nychka *et al.*'s is important, since we believe that the EMS approach of this paper has several advantages over the cross-validated spline approach of Nychka *et al.* (1984). Nychka *et al.* take a regression approach to what is a density estimation version of the integral equation problem. This is done by treating the data histogram values as if they were values of the intensity function f observed with error. Some justification for this is to appeal to the *asymptotic* result that the "error terms" will have zero mean, be jointly normal and weakly correlated; the latter correlation and unequal error variances were then ignored. The usual penalised least squares approach to such problems could then be applied with the value of the smoothing parameter involved chosen automatically by the well-known method of generalised cross-validation (see Nychka *et al.*, 1984, for references). The advantages we perceive for our EMS algorithm over Nychka *et al.*'s approach are its computational simplicity and speed, its more natural incorporation of the non-negativity constraint, and the fact that it attacks the Poisson likelihood directly.

3.4 A Real Data Example

The result of applying the EMS algorithm (with $J=5$) to some mouse liver data considered by Nychka *et al.* (1984) is given in Fig. 3.6. This reconstruction arises from a section through the liver of a mouse in which there are a number of malignant microtumours induced by injection of a carcinogen. A total of 154 tumour cross-sections were observed; we took $\varepsilon = 0.038$, $Y = 0.51$ (although the plot stops at 0.4; beyond this, $\hat{g} = 0$), $T = 64$ and $S = 50$. Fig. 3.6 is directly comparable with Fig. 6 of Nychka *et al.* (1984). The outstanding feature of this comparison is, once again, a remarkable similarity in reconstructions obtained by the two approaches. We have preferred, perhaps, a little less smoothing of the two; any minor differences can be largely attributed to this.

This particular mouse liver was, in fact, completely dissected and the histogram of sphere radii found is also shown on Fig. 6 of Nychka *et al.* (1984). In one sense, this forms a true distribution; comparing the reconstruction with the histogram reveals a generally good agreement, except for discrepancies in the magnitude and slope of the density near ε. However, this comparison is not entirely fair: we have been estimating a (presumed) smooth density of malignant tumours in mouse livers, of which the histogram is itself another (unsmooth) estimate, albeit based on a much larger sample of directly observed spheres.

4. A SECOND APPLICATION : POSITRON EMISSION TOMOGRAPHY

4.1 The Problem

PET is a medical diagnostic technique that studies the pattern of blood flow and metabolic activity in an organ by producing an indirectly observed image of a planar section through the patient's body. Such pictorial representations of internal structure have considerable appeal as a means of diagnosing certain diseases and in assessing the effectiveness of treatments. Some of the material of this section is a general review of the problem, but we shall suggest several technical improvements on existing methodology in addition to the use of our smoothed EM procedure. One particular advance is a new discretisation of the "body space" Y which affords considerable computational economies; see Section 4.1.2.

PET operates as follows. A radioactive tracer — here, a substance (often glucose) emitting positrons — is introduced into the area of interest and the occurrence of these emissions is recorded by an array of detectors arranged around the body; this apparatus is a *tomograph*. The amount of radiation given off at any point reflects the degree of activity present there, so the overall "emission density" provides the required portrait of internal structure which we estimate. The physics of PET is described by Vardi *et al.* (1985) thus: "When a positron is emitted, it 'finds' a nearby electron and annihilates with it. The annihilation creates two X-ray photons that fly off the point of annihilation, at the speed of light, in (nearly) opposite directions along a line with a completely random (i.e. uniformly distributed in space) orientation. There is an array of discrete detector elements surrounding the [area of interest], and the two photons are detected in coincidence by a pair of detector elements that define . . . a *tube*. Thus the only information acquired when a pair of detectors count a coincidence is that the annihilation occurred somewhere inside the tube defined by the two 'firing' detectors". This is illustrated in Fig. 4.1; see also Fig. 1 of Vardi *et al.* (1985) or Kaufman (1987). Fig. 4.1 is a planar view. It is important, however, to bear in mind the three-dimensional nature of the emission process and, consequently, the finite "depth" of the detectors; the effect of this (not considered by Vardi *et al.*, 1985) is discussed in Section 4.1.4. The tube counts comprise the data $\mathbf{n}$. Note that the tube space $\mathbf{X}$ differs from the body space $\mathbf{Y}$.

For more details on physical aspects of PET, see Vardi *et al.* (1985) and Hoffman and Phelps (1986). PET is a fairly recent innovation, many aspects of which are still at the development stage. Research interest in PET covers several disciplines; see Phelps, Mazziota and Schelbert (1986) for an up-to-date account, including an idea of the scope of medical application. Other kinds of tomography exist. Transmission tomography has had more impact: X-ray transmission tomography and related techniques are well-known, but are mathematically quite distinct from PET so the methods discussed here do not apply. Our methods can be modified for use, however, with another emission technique called single photon emission tomography (SPECT) which is little different, as far as mathematical or statistical analysis goes, from PET; see, for example, Geman and McClure (1985, 1987).

4.1.1 The Detectors

There are a number of detector configurations in current use at PET installations. We follow Vardi *et al.* (1985) in considering a single stationary circle of D detectors, each of equal size and with no gaps between them. Without loss of generality, the circle has unit radius. This is shown in Fig. 4.1; as there we take $D = 128$ in all our illustrations (D is often a power of 2). Practical variations on this set-up include alternative detector array shapes, gaps between detectors, two or more such arrays and movement of detectors. All $T = \frac{1}{2}D(D-1)$ pairs of detectors form the tubes or data bins. Although the PET problem is a bivariate (spatial) analogue of the univariate application of Section 3, it will remain convenient to index tubes by $t = 1,\ldots,T$ in what follows; we note that for computational purposes, however, the spatial location of detector tubes is best described by a polar coordinate system.

4.1.2 Discretising the Disc

The space **Y** is the disc enclosed by the circle of detectors. We require a discretisation of this disc on which to reconstruct and display emission densities; discretised functions are piecewise constant taking a single value across each bin or *pixel*. Many workers, including Vardi *et al.* (1985), simply superimpose a square grid of pixels over **Y**, but this approach suffers from important computational disadvantages compared with discretisations that better take into account the geometry of the situation. By more properly exploiting circular symmetries, it is possible to make substantial savings in both storage and time requirements. In order best to represent an image by a step function, all pixels should be, at least approximately, of equal area and shape.

Suppose we allow $D_1 = 2^k$ divisions of the detector circle into arcs of equal length, for some integer k. Then, our proposal is to use the discretisation shown in Fig. 4.2, constructed as follows. First divide the disc into $R = D_1/4$ equal-width rings by drawing circles of radius i/R, $i=1,2,\ldots,R$; for each i, set $j=[\log_2 i]$, where $[x]$ denotes the largest integer strictly less than x, and divide ring i into 2^{j+3} pixels of equal size and shape. Thus the pixellation is achieved by doubling segmentations of the circle at appropriate stages, at the expense of introducing "seams" between the 2^jth and (2^j+1)st rings, $j=0,1,\ldots,k-3$. Except for the innermost ring of all, each pixel is of the same general shape, while the ratio of maximum to minimum pixel area is strictly less than 2. For $D = 128$, the choice $D_1 = D$ yields what we regard to be too coarse a grid. Rather, we employ $D_1 = 2D$ pixels in the outermost ring and identify pairs of adjacent pixels with detectors. In this application, $s = 1,\ldots,S$ refer to these pixels.

Kearfott (1985) and Kaufman (1987) also recognise the advantages to be gained by using such a "ring grid". Kearfott's (1985) simple discretisation of the disc results in the division of the central area into very many long thin pixels, to the obvious detriment of discretised picture quality. Kaufman (1987) presents a discretisation which overcomes this problem. Kaufman's ring grid is, however, rather less easy to describe than is ours: "The ith ring is divided into n_i sectors so that $n_i = j_i \times k_i$ where j_i is a small integer and k_i is a divisor of D," but there appears to be no simple scheme for choice of these values. Further, Kaufman (1987) uses variable ring widths — although the widths "vary no more than about 25 percent" — to obtain pixels of equal area; it is not clear that this complication is worthwhile when pixels vary correspondingly more in shape. A full description of Kaufman's pixellation of the disc thus requires specifying values of j, k and width individually for each ring. For $D = 128$, Kaufman's (1987) discretisation results in 12,884 pixels, while the comparable grid in Fig. 4.2 has rather fewer — 10,924; each has $R = 64$. The value of these circularly symmetric discretisations for computational purposes is best reflected in the number P of different possible relationships of pixels to tubes, modulo rotations. For Kaufman's (1987) grid, $P = 200$; for ours, $P = R = 64$ — just one per ring. These numbers compare with $P = 2,080$ for a comparable square grid discussed by Kaufman (1987).

4.1.3 More on the Problem

We have no real data from a PET installation, but rather seek to reconstruct a relevant mathematical model (or mathematical *phantom*) of an image using simulated data. The phantom we use is (essentially) that of Vardi *et al.* (1985); in Fig. 4.3, we present a grey-level picture of that phantom, using 64 grey-levels to reflect emission intensity in the obvious way. This image is designed "as a simplified imitation of the brain's metabolic activity" with different areas representing the skull, grey matter, tumours and so on. Note that a property of this picture is that the emission density consists of areas of constant intensity with fairly large contrast between different areas. Fig. 4.3 is, of course, a discretised version of the ideal image (Fig. 2 of Vardi *et al.*, 1985), pixels overlapping area boundaries being regarded as having a weighted average of values present, in (approximate) proportion to area of pixel covered. Note that we actually aim to reconstruct this discretised emission density, and denote total emissions in pixels by $\lambda(s)$, $s=1,\dots,S$. Also, differing pixel areas must be taken into account in the smoothing and plotting; the intensities we plot are $\varphi_s = \lambda(s)/a(s)$, where $a(s)$ is the area of pixel s.

Positron emissions are assumed to occur uniformly at random over homogeneous regions, but at appropriately differing rates between areas of dissimilar material i.e. they occur according to a nonhomogeneous spatial Poisson point process with intensity function the emission density. The unobserved pixel counts are $\mathbf{m}$ and $k(s,t)$ is the number of emissions occurring in pixel s which are detected in tube t. Vardi *et al.* (1985) state that the Poisson process assumption "seems beyond challenge and requires no justification" in the PET problem.

The discretised kernel function becomes the probability that a uniformly orientated line through y in pixel s intersects the two detectors defining the tube t, averaged over all y in pixel s, for $s=1,\dots,S$, $t=1,\dots,T$. The geometrical problem of evaluating the p's exactly is non-trivial; we propose using a simple approximation. If the problem were the strictly two-dimensional one suggested by Fig. 4.2, the basic idea is to use

$$p(s,t) = \begin{cases} D^{-1} & \text{if the centre of pixel } s \text{ falls in tube } t, \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

We thus gloss over small variations in p's due to the planar geometry, but note that there is an important effect due to the third dimension which we shall discuss in Section 4.1.4. The computational advantage of concentrating on the centres of pixels in our disc discretisation is great. For $D = 128$ and $D_1 = 256$, we need only store the locations of 8,192 nonzero p's — compared with 27,378 reported for Kaufman's (1987) setup — and save considerably on computer time by addressing only those terms corresponding to nonzero p's in the EM update (2.2).

In practice, real PET apparatus involves numerous further important practical aspects including, for example, time-of-flight considerations (non-coincident arrival at detectors), attenuation problems ("soaking up" of X-ray photons) and scattering (non-axial photons); see Vardi *et al.* (1985), Kearfott (1985) and Hoffman and Phelps (1986). Some of these effects, such as time-of-flight and scattering, involve alterations only to the $p(s,t)$'s, so our general methodology would carry over unchanged. Non-linear effects like attenuation, where the $p(s,t)$'s depend on the unknown image, would require more substantial modifications.

4.1.4 Accounting for the Third Dimension

Photon lines are emitted in directions distributed uniformly in 3-dimensional space and the detectors have a finite depth, d, say; this has not yet been taken into account. Consider a tube of length l, say, where l is large relative to d, and condition on the emission being in the direction of that tube. Suppose the annihilation takes place at a distance l_1 from the left hand detector at a height x and write $l_2 = l - l_1$ (take $x < \frac{1}{2}d$ and $l_1 < l_2$, without loss of generality); see the cross-sectional view of Fig. 4.4. It is natural to assume that x is uniformly distributed on $[0, \frac{1}{2}d]$; this reflects an assumption that d is small enough for there to be negligible change in intensity over that distance. The contribution to $p(s,t)$ due to this third dimension is what we consider here, namely, the probability that a particular emission yields a photon pair that hits both detectors.

Suppose that ψ is the angle that the photon line makes with the plane of the detectors. Since $l \gg d$, only small ψ's can occur, so that $\psi \approx \tan\psi$ is approximately uniformly distributed on its range of admissible values. Evaluating this range is not difficult. First, if $l^{-1} l_1 d < x \leq \frac{1}{2}d$, any line hitting the right hand detector automatically also hits the left hand one; the range of appropriate ψ's is thus approximately d/l_2. Otherwise, $0 \leq x \leq l^{-1} l_1 d$ and the range, which is governed by the angles allowed by the bottom edges of both detectors, is approximately $xl/(l_1 l_2)$. Averaging over the distribution of x yields an average range of ψ which corresponds to the required probability. A simple calculation shows that, to the degree of approximation used above, this probability is d / l.

Now, this quantity does not depend on where in the tube the annihilation takes place, but only on the length $l = l(t)$ of the tube. Thus, $p(s,t)$ is modified by a factor, inversely proportional to $l(t)$, depending only on t, and not s. The effect on the data is clear: a smaller proportion of emissions occurring towards the centre of the disc will be detected than of those happening near the edge, with a consequent degradation of reconstruction quality to be expected in the (important) central area. That this third dimension effect remains important while an apparently similar effect in the planar case — imagine Fig. 4.4 as the view looking down on a tube in Fig. 4.2 — does not, is due to short tubes in the plane also becoming thin tubes (d decreases with l), but retaining their depth in the third dimension.

The real importance of the third dimension lies in the fact that the 3-dimensional problem does not tend, in the limit as $d \rightarrow 0$, to the 2-dimensional one. To see this, note that any $d>0$ results, after proper normalisation, in an identical set of $p(s,t)$'s; these include $l(t)$, the 2-dimensional ones do not. Since the real PET problem is 3-dimensional, our approximating to that case is much preferable to approximating the planar situation only. Since the change to $p(s,t)$ depends only on t, the extra computational burden imposed by taking account of the third dimension is virtually nil.

4.2 EM Reconstruction

We are now in a position to apply the EM algorithm for ML estimation to the PET problem, exactly as described in Section 2. Shepp and Vardi (1982) were the first to do so; Vardi *et al.* (1985) and Kaufman (1987) follow up this work (see also Lange and Carson, 1984). The uniform initialisation / early termination version of EM which is actually employed is widely regarded as being among the best PET reconstruction procedures currently available; see, for example, Shepp *et al.* (1984), Mintun *et al.* (1985) and Vardi *et al.* (1985), the last named outlining several competing reconstruction methods. Most commercial PET installations persist in using other techniques (especially "convolution back projection", see Shepp and Kruskal, 1978) because of the computational advantage such approaches afford (Kaufman, 1987).

A dataset arising from the phantom of Fig. 4.3 was simulated; all reconstruction attempts to be portrayed in succeeding figures are based on these data. In line with many other studies, a total number of emissions, N, of 10^6 was chosen (this is, however, rather fewer than the number employed by Vardi *et al.*, 1985, and Kaufman, 1987). Data simulation was again performed by mimicking the physical process: obtain points from the Poisson process with the intensity displayed in Fig. 4.3 by the obvious acceptance/rejection method, obtain randomly oriented lines through these points by choosing uniformly distributed angles and finally perform a further acceptance/rejection step with acceptance probabilities inversely proportional to the lengths of these lines to take the third dimension into account.

A reconstruction of the phantom obtained by allowing the EM algorithm to run for some considerable time — here, 200 iterations — is shown in Fig. 4.5. The result is typical of the unacceptability of "ML reconstructions" in this context. The image obtained is itself very noisy: putative flat areas are estimated to be extremely rough. As well as the lack of aesthetic appeal, the effect of this is that only the very strongest features — here, the large circle and ellipse, both with very different intensity from the background — survive for inspection; this is clearly unsatisfactory. The speckled nature of Fig. 4.5 reflects the roughness of the reconstructed surface in plan view; neighbouring pixels are estimated to have widely differing intensities. It is important to note that the EM algorithm has not yet converged and the roughness described here *gets worse* if we allow the algorithm to run further. This is because ML is trying to suggest a "spiky" solution to the problem, much as in Figs 3.1 and 3.2 for the univariate analogue, this effect being mitigated here by the smoothing due to the discretisation of the disc. The same grey scaling is used on all reconstructions. The great variability in Fig. 4.5 implies that in the darker areas, some estimated pixel intensities lie above the highest grey level and have been redefined to be black; some of the speckled nature of the picture, especially on the largest circle, has thus been concealed.

Of course, in practice, application of the EM algorithm is not allowed to reach a state like that of Fig. 4.5. Rather, the iterations are terminated early: Fig. 4.6 displays the reconstruction obtained by stopping after just 24 EM iterations. Calling this (erroneously) "the ML reconstruction" accounts for the good performance attributed to the method: in Fig. 4.6, large "objects" are well reconstructed and roughness, compared with Fig. 4.5, is considerably reduced. (Small scale features present in the phantom are hinted at, if not reproduced convincingly.) Veklerov and Llacer (1987) propose a data-dependent rule for selecting the point at which early termination of the EM algorithm should occur. The use of a constant initial configuration is important here; it is a smoothing influence which persists through the early stages of the EM algorithm. Roughly speaking, early iterations quickly make manifest approximate shapes and intensities of objects, while the later iterations are responsible for roughening the image. The uniform starting point is the ultimate in smooth images in the sense appropriate to PET. So, rather than the common approach of smoothing a rough image towards such a smooth one, the EM iterations are used to "roughen" away from the ultrasmooth.

Different choices of initial estimate result in different EM reconstructions; a vivid illustration of how properties of initial reconstruction can persist to appear in iterated reconstructions is given in Fig. 5 of Kaufman (1987). Extensive recent work on accelerating convergence of the EM algorithm (Lewitt and Muehllehner, 1986, Kaufman, 1987, Lange, Bahn and Little, 1987) seems gratuitous since, as we have argued, the ML optimisation is inappropriate to the problem at hand.

All our reconstructions incorporate the third dimension effect described in Section 4.1.4. These turn out to be slightly smoother than comparable reconstructions of the purely two-dimensional version; the length bias has the effect of making the problem less ill-posed. As anticipated in Section 4.1.4, there is a slight deterioration in quality of reconstruction towards the centre of the image; perhaps it would be more realistic to suppose the area of interest filled a smaller portion of the tomograph disc, whence such an effect might become more important. Note also that we might expect the incorporation of more physical considerations into the $p(s,t)$'s to result in less smooth reconstructions than here, since most would have a smoothing effect on the kernel and a consequent worsening of the ill-posed nature of the problem.

4.3 Smoothed EM Reconstruction

4.3.1 Local Smoothing

We utilise perhaps the most natural (and common) approach to smoothing values on a spatial grid: replace the value at each pixel by some function of that value and those of its nearest neighbours. Examples of useful smoothing functions follow in Section 4.3.2. A little care needs to be taken over the definition of neighbours in our circular discretisation scheme. For a rectangular discretisation, Besag (1986, p.262), for example, identifies nearest neighbours of a pixel in a natural way: first-order neighbours are those pixels adjacent vertically and laterally to the one of interest, while a second-order neighbourhood additionally includes diagonal adjacencies. The effect of a finite window is to modify these definitions (in an obvious way) for boundary and corner pixels. It is not difficult to translate these notions to the circular grid although, because of the seaming, we need deal with 8 (rather than 3) different pixel types. First and second order neighbours of pixels of each type are identified in Fig. 4.7 (using $D_1 = 64$ for clarity). These definitions retain the desirable property of symmetry of neighbour pairs: if s_1 is a neighbour of s_2, s_2 is a neighbour of s_1. Following Besag (1986), we view the second order scheme as the most useful one (and use it throughout). A further argument here for going beyond first-order neighbourhoods is that alternate pixels on the outside of a seam have different types of adjacency on the inner edge; this leads to an undesirable "castellation" effect on reconstructions using first-order neighbours only.

Unlike the basic EM algorithm, boundary and seam effects mean that EMS algorithms do not automatically scale so that $\sum_s \lambda^{(i)}(s) = N$, all $i \geq 1$. Operationally, we rescale smoothed images to have this property; although this has no effect on successive reconstructions, it is useful in making successive values of the (log) likelihood comparable.

4.3.2 Smoothed Reconstructions

In this section, we present some examples of applying versions of the smoothed EM algorithm to reconstructing the phantom of Fig. 4.3. Because the phantom, as well as images likely to occur in practice, is not everywhere smooth but contains discontinuities, we have experimented with both simple linear smoothers and with local non-linear ones. Smoothers that purport to preserve edges are necessarily non-linear in the values at the pixel of interest and its neighbours; Scher *et al.* (1980) and Chin and Yeh (1983) describe a number of methods that have been proposed in the literature for use in cleaning up noisy images containing discontinuities. However, the performance of non-linear smoothers within EMS has been disappointing. We report reconstructions based on just two of these smoothing schemes out of the several we have tried. In Fig. 4.8, we exhibit the result of using the EMS algorithm with local median smoother i.e. replacing a pixel value by the median of it and its neighbours' values. In Fig. 4.9, a slightly more sophisticated non-linear smoother — the best we have used in this context — was deployed. This is the mean of the central pixel value and of the two neighbouring values closest to the central one; in this way, we try to average only over pixels on "the right side" of an edge (this is a special case of KAVE of Chin and Yeh, 1983). Neither of these, nor any others that we have investigated, yields a good reconstruction. As well as eradication of the smaller features of the phantom, significant distortions are introduced as artefacts of the methods used. Both Figs 4.8 and 4.9 are pictures produced after 200 EMS iterations. It is important to note that neither of these non-linear methods converged.

Returning to linear local smoothers — and thus relaxing our concern for trying to avoid blurring feature edges — we get better results. A simple scheme, which works well, is this: take a weighted average of the form weight 1 for the central value and equal weights W^{-1}, say, for each neighbouring value, normalised appropriately (other linear smoothing possibilities are in Russ and Russ, 1984). This is closely related to the way we smoothed in the stereology context. It turns out that we need only a small amount of smoothing (W large) for good effect. Reconstructions for the ongoing example are given using $W = 200$, 50 and 10 in Figs 4.10, 4.11 and 4.12, respectively. The first of these reflects the effect of (slightly) undersmoothing: background roughness remains too high, although objects are fairly clearly visible. The last is a little oversmoothed: better background but loss of resolution in object reconstruction. The choice $W=50$ in Fig. 4.11 seems to be about as good a compromise as can be obtained by this method. We have not considered automatic choice of smoothing parameter, but are encouraged by the fact that "best" choosing W might not be critical: reconstructions (not shown) using W between say 100 and 25 are not substantially different from that of Fig. 4.11. Note that in virtually all of our reconstructions a minor effect due to the seam in our discretisation with radius one half that of the disc is faintly visible. In particular, this artefact has had a slightly detrimental effect on the quality of reconstruction of the pair of small ellipses towards the bottom of the phantom which lie near to this seam.

Quantifying reconstruction quality in image analysis is not easy. We briefly report L_1 discrepancies between phantom and reconstructions; this quantity is

$$B = \sum_{s=1}^{S} a(s) \, |\Lambda(s) - \hat{\Lambda}(s)|$$

where Λ and $\hat{\Lambda}$ are grey scale values corresponding to λ and $\hat{\lambda}$ respectively. Now, $B=6.191$ for our $W=50$ reconstruction, although smaller values of B are achieved for smoother pictures: $B=5.728$ for $W=25$ is the best achieved. Our visual preferences are better reflected in other L_1 values, though: Fig. 4.5 yields the very large value $B=20.844$, Fig. 4.9 is just preferable ($B=9.825$) to Fig. 4.8 ($B=10.222$) and is much preferable to other non-linear EMS solutions, and the reconstruction of Fig. 4.6 after 24 EM iterations is quite good with $B=6.910$ (this is comparable with EMS reconstructions with W between 75 and 100). It is noticeable that B displays a preference for oversmoothed images but is otherwise satisfactory. In any case, it is widely recognised that this type of measure does not really give a good reflection of the human observer's sense of image fidelity especially when, as here, the true image contains features with distinct edges. Indeed, the provision of image metrics that properly reflect visual perception remains a difficult question: see Baddeley (1987), for example.

Finally, the EMS algorithm using local linear smoothing has always converged in a reasonable number of iterations. Indeed, using a convergence criterion essentially corresponding to that in Section 3.2.3, the numbers of iterations required for convergence of EMS with $W=200$, 50 and 10 were 62, 43 and 32, respectively. Moreover, simulation experience suggests that the local linear EMS reconstruction is unique.

4.3.3 Other Smoothed EM Methods in Emission Tomography

To the best of our knowledge, this is the first time our simple EMS algorithm has been proposed. There are, however, some other suggestions for incorporating smoothing into the EM algorithm in PET and/or SPECT in the recent literature. We have already mentioned a penalised likelihood approach — an EM_p algorithm — in Section 2. In the PET context, the M_p step is essentially the same problem as that arising in image processing problems which are approached by maximum *a posteriori* estimation (see Geman and Geman, 1984, and Besag, 1986). Even when an appropriate penalty function or prior distribution has been decided upon — locally dependent Markov random fields form a class of priors capable of quantifying notions of local smoothness (Besag, 1986) — the computational problem of locating the global maximum of the penalized likelihood is immense and not yet satisfactorily solved (Greig *et al.*, 1986). Obtaining a local maximum at the M_p stage is more reasonable. A successful method for finding a "good" local maximum in image processing is Besag's (1986) iterated conditional modes (ICM) algorithm. Roughly speaking, ICM is not all that different from our simple local smoothing: it performs a few iterations of a sequential local smooth (i.e. "current" pixel values include those already updated, not just the originals) using a local smoother based on maximising a penalized marginal likelihood. We would not be surprised to find that the ICM approach yields good reconstructions; we wonder, though, if even its level of sophistication will ultimately prove to be worthwhile. Indeed, Geman and McClure (1985, 1987), considering the application of such methods in the context of SPECT, decided to fall short of a full implementation of such an M_p algorithm. Rather, they obtained a reconstruction by some other method to act as initial estimate and then applied a single M_p step of the above sort. Perhaps a better perspective on Geman and McClure's approach is as the application of popular image processing techniques to cleaning up reconstructions obtained in other ways. Note too that Geman and McClure (1987) consider posterior mean reconstructions (their penalised likelihoods are posterior distributions) as alternatives to posterior modes. Less appealing to the current authors are other EM_p approaches utilising pixel-by-pixel priors designed to encourage smoothing towards prespecified, or estimated, images. Examples are given by Hart and Liang (1987), Lange *et al.* (1987) and Levitan and Herman (1987). Other regularisation procedures, based, we believe, on inappropriate roughness penalty functions, are considered by Girard (1987) and Miller and Snyder (1987).

A rather different approach to smoothed EM algorithms for PET is taken by Snyder and Miller (1985) (also Miller, Snyder and Moore, 1986, and Snyder *et al.*, 1987). These authors force their emission density estimate to have kernel convolution form i.e.

$$\hat{\lambda}(s) = \int \Theta(s,z)\theta(z)dz \qquad s=1,\ldots,S,$$

for Θ a known kernel, and θ is estimated by ML — this is a kernel convolution sieve (Geman and Hwang, 1982). This is identical with replacing the point-spread function by a kernel-smoothed version of it,

$$p^{\Theta}(s,t) = \sum_r \Theta(s,r)p(r,t),$$

say, and proceeding by the EM algorithm. As a last step, this ML solution is smoothed once by application of Θ. This approach, while certainly yielding smooth images, requires the (spiky) ML solution to what is an even more ill-posed problem (caused by the smoothing effect of Θ) and so it retains and perhaps even exacerbates all the numerical convergence problems of obtaining true ML reconstructions by the EM method.

4.3.4 Closing Remarks on the PET Application

A first striking feature of the reconstructions shown in this paper is the similarity between that obtained by the uniform start / early termination modification of basic EM, in Fig. 4.6, and the "best" weighted local mean EMS reconstruction shown in Fig. 4.11. We have certainly shown that nothing need be lost in terms of reconstruction quality by the introduction of our explicit smoothing procedure and would argue that the latter image is indeed a slight improvement over the former. Moreover, the EMS formulation offers prospects of further improvement: other local smoothing schemes can be fitted into the same framework and might work better, while the benefits of the provision of an apparently uniquely convergent algorithm include scope for further computational improvement such as accelerating that convergence.

5. SOME THEORETICAL BACKGROUND

The clear empirical success of the EMS algorithm immediately asks several theoretical questions. It has been observed in practice that the EMS algorithm employing linear smoothers converges relatively quickly and that its limit point is apparently unique. Obviously it would be of interest to prove these properties rigorously. Unfortunately, we have not been able to do so, but in this section we provide a heuristic discussion that relates the EMS procedure to an EM_p approach where the likelihood is penalised by a term that is quadratic in the vector of *square roots* of the intensities. This relationship gives some insight into the good properties of EMS and it is our hope that it will be a useful starting point for future theoretical work.

5.1 A Lemma

The first step in our development is a simple algebraic lemma.

Lemma. Suppose that W is a diagonal matrix of weights and that S is a matrix for which $S_{rs} \geq 0$ for all r and s and $\sum_s S_{rs} = 1$ for all r. Suppose that for some $\delta > 0$

$$|\, w_r^{-1} w_s - 1 \,| \leq \delta$$

for all (r,s) such that $S_{rs} \neq 0$; here, the w's are the diagonal elements of $\mathbf{W}$. Define a matrix $\mathbf{T}$ by $\mathbf{T} = \mathbf{W}^{-1}\mathbf{S}\mathbf{W}$. Then for any vector $\mathbf{x}$

$$|(\mathbf{T}\mathbf{x} - \mathbf{S}\mathbf{x})_r| \leq \delta \sup_u |x_u|. \tag{5.1}$$

Proof. $|(\mathbf{T}\mathbf{x} - \mathbf{S}\mathbf{x})_r| = |\sum_s S_{rs}(w_r^{-1}w_s - 1)\, x_s| \leq \delta \sup_u |x_u| \sum_s S_{rs} = \delta \sup_u |x_u|.$ □

The implication of the lemma in the current context is as follows. Suppose that $\mathbf{x}$ is indexed by our reconstruction bins and that $\mathbf{S}$ is a local smoothing operator so that $S_{rs} = 0$ unless r and s are neighbouring bins. Suppose that $\mathbf{W}$ is an array of weights that vary continuously over the space $\mathbf{Y}$, that is, $w_r \approx w_s$ whenever r and s are neighbours; then δ can be chosen to be small. The operator $\mathbf{T}$ corresponds to weighting an intensity by the w weights, smoothing by the operator $\mathbf{S}$ and then unweighting. The lemma therefore quantifies the intuitive notion that $\mathbf{S}$ and $\mathbf{T}$ will have approximately equal effects.

5.2. A Relationship Between Local Smoothing and Maximum Penalised Likelihood for Poisson Random Variables

Write $\boldsymbol{\varphi}$ for the vector of φ_s's where $\varphi_s = \lambda(s)/a(s)$ as in Section 4.1.3 and define $\pi_s = \sqrt{\varphi_s}$ for all s with $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_S)$. Also write Ψ as the diagonal matrix with diagonal elements $\psi_s = q(s)a(s)$. Let $\mathbf{S}$ be a smoothing matrix all of whose eigenvalues lie in $(0,1]$ and define $\mathbf{R} = \Psi^{\frac{1}{2}}(\mathbf{S}^{-1} - \mathbf{I})\Psi^{\frac{1}{2}}$. Suppose that observations $m(s)$ of independent Poisson$(\varphi_s\psi_s)$ random variables are available and let $l_p(\boldsymbol{\pi})$ be the log likelihood penalised by $\boldsymbol{\pi}^T\mathbf{R}\boldsymbol{\pi}$ i.e.

$$l_p(\boldsymbol{\pi}) = \sum_s m(s)\log(\pi_s^2\psi_s) - \sum_s \pi_s^2\psi_s - \boldsymbol{\pi}^T\mathbf{R}\boldsymbol{\pi}.$$

To see that $\boldsymbol{\pi}^T\mathbf{R}\boldsymbol{\pi}$ has the effect of being a roughness penalty, note that the eigenvectors corresponding to large eigenvalues of $\mathbf{R}$ will be those corresponding to small eigenvalues of $\mathbf{S}$, and so will consist, loosely speaking, of high frequency oscillations. The following theorem demonstrates a connection between the penalised ML estimate of $\boldsymbol{\pi}$, and the estimate obtained by direct smoothing of the ML estimate of $\boldsymbol{\varphi}$ and by taking square roots.

Define $\hat{\boldsymbol{\pi}}$ to be the maximiser of $l_p(\boldsymbol{\pi})$ in $\{\pi_s > 0\}$. Set $\mathbf{W} = \Psi^{\frac{1}{2}}\hat{\Pi}^{-1}$, where $\hat{\Pi} = \mathrm{diag}(\hat{\boldsymbol{\pi}})$. The ML estimate of $\boldsymbol{\varphi}$ is $\Psi^{-1}\mathbf{m}$. Define $\mathbf{T} = \mathbf{W}^{-1}\mathbf{S}\mathbf{W}$ as in Section 5.1 and let $\tilde{\boldsymbol{\pi}} = (\tilde{\pi}_1, \ldots, \tilde{\pi}_S)$ where $\tilde{\pi}_s = (\mathbf{T}\Psi^{-1}\mathbf{m})_s^{\frac{1}{2}}$ for each s.

Theorem. With the above definitions, $\hat{\boldsymbol{\pi}} = \tilde{\boldsymbol{\pi}}$.

Proof. Write

$$l_p(\pi) = -\pi^T(\Psi + \mathrm{R})\pi + 2\sum_s m(s)\log \pi_s + \sum_s m(s)\log \psi_s .$$

Hence the Hessian matrix of l_p is $-2(\Psi + \mathrm{R} + \mathrm{diag}\{m(s)/\pi_s^2\})$, which is strictly negative definite in $\{\pi_s > 0\}$, and so $\hat{\pi}$ will be uniquely defined by $\nabla l_p(\hat{\pi}) = 0$. This is true if and only if

$$\{(\Psi + \mathrm{R})\hat{\pi}\}_s = m(s)\,\hat{\pi}_s^{-1} \text{ for each } s.$$

It is easy to see that $\Psi + \mathrm{R} = \Psi^{\frac{1}{2}}\mathrm{S}^{-1}\Psi^{\frac{1}{2}}$ and that, if $\hat{\varphi}$ is the vector of $\hat{\pi}_s^2$'s, $\hat{\varphi} = \Psi^{\frac{1}{2}}\mathrm{W}^{-1}\hat{\pi}$. Therefore, the vector with components $\hat{\pi}_s\{(\Psi + \mathrm{R})\hat{\pi}\}_s$ is equal to

$$\hat{\Pi}(\Psi + \mathrm{R})\hat{\pi} = (\mathrm{W}^{-1}\Psi^{\frac{1}{2}})(\Psi^{\frac{1}{2}}\mathrm{S}^{-1}\Psi^{\frac{1}{2}})\Psi^{-\frac{1}{2}}\mathrm{W}\hat{\varphi} = \Psi\mathrm{W}^{-1}\mathrm{S}^{-1}\mathrm{W}\hat{\varphi} = \Psi\mathrm{T}^{-1}\hat{\varphi}.$$

Thus, $\mathbf{m} = \Psi\mathrm{T}^{-1}\hat{\varphi}$ so that $\hat{\varphi} = \mathrm{T}\Psi^{-1}\mathbf{m}$ and therefore $\hat{\pi}_s = (\mathrm{T}\Psi^{-1}\mathbf{m})_s^{\frac{1}{2}} = \tilde{\pi}_s$ for all s. ☐

Of course, the smoothing matrix T depends, through the weights w_s, on $\hat{\pi}$ and so the expression of $\hat{\varphi}$ as a smoothed version of $\Psi^{-1}\mathbf{m}$ is not immediately of practical use. However, a heuristic argument based on the lemma relates the smoothed ML estimate $\varphi^* = \mathrm{S}\Psi^{-1}\mathbf{m}$ to $\hat{\varphi}$ as follows. Since the penalty $\pi^T\mathrm{R}\pi$ can be expected to penalise for roughness in π, the penalised ML estimates $\{\hat{\pi}_s\}$ will vary continuously. Provided the ψ_s's also vary continuously, so will the weights $\{w_s\}$ and hence, by the lemma, the effects of smoothing by the operators S and T will be almost identical. Thus $\varphi^* = \mathrm{S}\Psi^{-1}\mathbf{m} \simeq \mathrm{T}\Psi^{-1}\mathbf{m} = \hat{\varphi}$. Note that in the PET context, the ψ_s's do not vary continuously across the seams in our discretisation of Y but this does not appear in practice to have an important effect. In the stereology example, there is no such discontinuity.

5.3 Smoothed EM and Penalised EM

Return now to the EM algorithm and consider the construction of an EM_p algorithm to maximise $l(\mathbf{n}\,|\,\lambda)$ penalised by a term $\pi^T\mathrm{R}\pi$ as above. Recall that, at each iteration, in the notation of Section 2, $\hat{m}(s) = \sum_t \hat{k}(s,t)$. If the M_p step is then approximated by finding the smoothed ML estimate $\lambda^*(s) = \varphi_s^* a(s)$, where $\varphi^* = \mathrm{S}\Psi^{-1}\hat{\mathbf{m}}$, then the effect is precisely an iteration of the EMS algorithm using the smoothing operator S. Thus, each EMS iteration corresponds approximately to an iteration of the EM_p-algorithm with the penalty on the square roots of the intensities; this is the point we aimed to make.

This heuristic equivalence may account for the rapid convergence of the EMS algorithm; see the remarks of Dempster *et al.* (1977) about the EM_p algorithm. We have been unable to prove that the penalised likelihood has a unique maximum but our empirical experience suggests that this is so. Certainly it will be the case in general that at any maximum of the penalised likelihood the Hessian matrix is positive definite so the maximum will be strict.

6. SUGGESTIONS FOR FURTHER WORK

We have introduced a simple algorithm that is widely applicable to a large class of problems involving indirect observations. Clearly, there is much scope to supplement our fruitful empirical studies by further theoretical and practical work. In particular, the work of Section 5 might be carried further. Once this is done, it would then be of interest to study the theoretical properties of the EMS solution, both for their own sake and in order, hopefully, to give insight into the choice of smoothing parameter. Alternative smoothing schemes are also of interest.

The whole area of statistical methods for indirect data is not enormously well understood. One interesting question is that of quantifying the information loss inherent in the indirectness of the data. Johnstone and Silverman (1988) have made progress for the PET problem towards finding an equivalent "direct sample size", i.e. the number of emissions whose exact position would have to be observed to give the same accuracy of estimation as the given sample of indirect observations.

ACKNOWLEDGEMENTS

The authors warmly thank Richard Glendinning, Chris Jennison, John Rice, Robin Sibson, Glenn Stone and several anonymous referees for their helpful comments and suggestions, and Adrian Baddeley and Stuart Geman for providing prepublication copies of their work. The financial support of the Science and Engineering Research Council and the U.S. Army European Research Office are gratefully acknowledged.

REFERENCES

Baddeley, A.J. (1987) A class of image metrics. To appear.

Bertero, M. and Dovi, V. (1981) Regularized and positive-constrained inverse methods in the problem of object restoration. *Optica Acta*, **28**, 1635-1649.

Besag, J.E. (1986) On the statistical analysis of dirty pictures (with discussion). *J. Roy. Statist. Soc. Ser. B*, **48**, 259-302.

Chin, R.T. and Yeh, C.-L. (1983) Quantitative evaluation of some edge-preserving noise-smoothing techniques. *Comput. Vision, Graphics, Image Processing*, **23**, 67-91.

Cruz-Orive, L.M. (1983) Distribution-free estimation of sphere size distributions from slabs showing overprojection and truncation, with a review of previous methods. *J. Microsc.*, **131**, 265-290.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the *EM* algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B*, **39**, 1-38.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intelligence*, **PAMI-6**, 721-741.

Geman, S. and Hwang, C.R. (1982) Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.*, **10**, 401-414.

Geman, S. and McClure, D.E. (1985) Bayesian image analysis: an application to single photon emission tomography. *Proc. Amer. Statist. Assoc., Statist. Comput. Section*, 12-18.

——— (1987) Statistical methods for tomographic image reconstruction. *Bull. ISI*, **52**, to appear.

Girard, D.A. (1987) Optimal regularized reconstruction in computerized tomography. *SIAM J. Sci. Statist. Comput.*, **8**, 934-950.

Greig, D.M., Porteous, B.T. and Seheult, A.H. (1986) Contribution to the discussion of Besag (1986). *J. Roy. Statist. Soc. Ser. B*, **48**, 282-284.

Hart, H. and Liang, Z. (1987) Bayesian image processing in two dimensions. *IEEE Trans. Med. Imaging*, **MI-6**, 201-208.

Hoffman, E.J. and Phelps, M.E. (1986) Positron emission tomography: principles and quantitation. In Phelps *et al.* (1986), 237-286.

Johnstone, I.M. and Silverman, B.W. (1988) Speed of estimation in positron emission tomography. *Ann. Statist.*, tentatively accepted for publication.

Kaufman, L. (1987) Implementing and accelerating the EM algorithm for positron emission tomography. *IEEE Trans. Med. Imaging*, **MI-6**, 37-51.

Kearfott, K.J. (1985) Comment on Vardi *et al.* (1985). *J. Amer. Statist. Assoc.*, **80**, 26-28.

Lange, K., Bahn, M. and Little, R. (1987) A theoretical study of some maximum likelihood algorithms for emission and transmission tomography. *IEEE Trans. Med. Imaging*, **MI-6**, 106-114.

Lange, K. and Carson, R. (1984) EM reconstruction algorithms for emission and transmission tomography. *J. Comput. Assisted Tomography*, **8**, 306-316.

Levitan, E. and Herman, G.T. (1987) A maximum *a posteriori* probability expectation maximization algorithm for image reconstruction in emission tomography. *IEEE Trans. Med. Imaging*, **MI-6**, 185-192.

Lewitt, R.M. and Muehllehner, G. (1986) Accelerated iterative reconstruction for positron emission tomography based on the EM algorithm for maximum likelihood estimation. *IEEE Trans. Med. Imaging*, **MI-5**, 16-22.

Little, R.J.A. and Rubin, D.B. (1987) *Statistical Analysis with Missing Data.* New York: Wiley.

Miller, M.I. and Snyder, D.L. (1987) The role of likelihood and entropy in incomplete-data problems: applications to estimating point-process intensities and Toeplitz constrained covariances. *Proc. IEEE*, **75**, 892-907.

Miller, M.I., Snyder, D.L. and Moore, S.M. (1986) An evaluation of the use of sieves for producing estimates of radioactivity distributions with the EM algorithm for PET. *IEEE Trans. Nucl. Sci.*, **NS-33**, 492-495.

Mintun, M.A., Gorman, J., Swift, A.G. and Snyder, D.L. (1985) Evaluation of the maximum-likelihood method for reconstruction images in positron emission tomography. Abstract in *J. Nucl. Med.*, **26**, 720.

Nychka, D., Wahba, G., Goldfarb, S. and Pugh, T. (1984) Cross-validated spline methods for the estimation of three-dimensional tumor size distributions from observations on two-dimensional cross sections. *J. Amer. Statist. Assoc.*, **79**, 832-846.

O'Sullivan, F. (1986) A statistical perspective on ill-posed inverse problems (with comments). *Statist. Sci.*, **1**, 502-527.

Phelps, M.E., Mazziota, J.C. and Schelbert, H.R. (eds) (1986) *Positron Emission Tomography and Autoradiography: Principles and Applications for the Brain and Heart.* New York: Raven Press.

Russ, J.C. and Russ, J.C. (1984) Image processing in a general purpose microcomputer. *J. Microsc.*, **135**, 89-102.

Scher, A., Velasco, F.R.D. and Rosenfeld, A. (1980) Some new image smoothing techniques. *IEEE Trans. Systems Man Cybernet.*, **SMC-10**, 153-158.

Shepp, L.A. and Kruskal, J.B. (1978) Computerized tomography: the new medical X-ray technology. *Amer. Math. Monthly*, **85**, 420-439.

Shepp, L.A. and Vardi, Y. (1982) Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging*, **MI-1**, 113-122.

Shepp, L.A., Vardi, Y., Ra, J.B., Hilal, S.K. and Cho, Z.H. (1984) Maximum likelihood PET with real data. *IEEE Trans. Nucl. Sci.*, **NS-31**, 910-912.

Silverman, B.W. (1985a) Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion). *J. Roy. Statist. Soc. Ser. B*, **47**, 1-52.

——— (1985b) Penalized maximum likelihood estimation. In *Encyclopedia of Statistical Sciences*, vol. 6 (eds: S. Kotz & N.L. Johnson). New York: Wiley, 664-667.

——— (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Snyder, D.L. and Miller, M.I. (1985) The use of sieves to stabilize images produced with the EM algorithm for emission tomography. *IEEE Trans. Nucl. Sci.*, NS-32, 3864-3872.

Snyder, D.L., Miller, M.I., Thomas, L.J. and Politte, D.G. (1987) Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography. *IEEE Trans. Med. Imaging*, MI-6, 228-238.

Tikhonov, A.N. and Arsenin, V.Y. (1977) *Solutions of Ill-Posed Problems*. Washington, D.C.: Winston.

Titterington, D.M. (1985a) General structure of regularization procedures in image reconstruction. *Astronom. and Astrophys.*, 144, 381-387.

——— (1985b) Common structure of smoothing techniques in statistics. *Internat. Statist. Rev.*, 53, 141-170.

Vardi, Y., Shepp, L.A. and Kaufman, L. (1985) A statistical model for positron emission tomography (with comments). *J. Amer. Statist. Assoc.*, **80**, 8-37.

Veklerov, E. and Llacer, J. (1987) Stopping rule for the MLE algorithm based on statistical hypothesis testing. *IEEE Trans. Med. Imaging*, MI-6, 313-319.

Wicksell, S.D. (1925) The corpuscle problem. A mathematical study of a biometric problem. *Biometrika*, 17, 84-99.

Wilson, J.D. (1987) A smoothed EM algorithm for the solution of Wicksell's corpuscle problem. Submitted for publication.

FIGURE LEGENDS

Fig. 3.1. An EM reconstruction (——) of the truncated Weibull density (- - -).

Fig. 3.2. An EM reconstruction (——) of the truncated normal mixture density (- - -).

Fig. 3.3. An EMS reconstruction using $J=5$ (——) of the truncated normal mixture density (- - -). The reconstruction is based on the same data as Fig. 3.2.

Fig. 3.4. Another EMS reconstruction using $J=5$ (——) of the truncated normal mixture density (- - -). The reconstruction is based on a different dataset.

Fig. 3.5. An EMS reconstruction using $J=9$ (——) of the truncated Weibull density (- - -). The reconstruction is based on the same data as Fig. 3.1.

Fig. 3.6. The EMS reconstruction using $J=5$ (——) of the sphere radius intensity for the mouse liver data.

Fig. 4.1. A planar section through an elliptical "body" within a circular detector set; edges of individual detectors are marked. An emission, at ∗, yields a randomly orientated line in 3-space. Two such possible lines are shown.

Fig. 4.2. Our discretisation of the disc.

Fig. 4.3. The phantom.

Fig. 4.4. A cross-section through a tube showing the distances defined in the text. The annihilation spot is marked ∗.

Fig. 4.5. Reconstruction after 200 EM iterations.

Fig. 4.6. Reconstruction after 24 EM iterations.

Fig. 4.7. Neighbours in the circular discretisation scheme: O = pixel of interest, ∗ = first order neighbour, × = second order neighbour. There are eight different pixel types in all; two of these are illustrated on separate insets.

Fig. 4.8. EMS reconstruction using local median smoother.

Fig. 4.9. EMS reconstruction using 2AVE smoother.

Fig. 4.10. EMS reconstruction using weighted local mean smoother with $W=200$.

Fig. 4.11. EMS reconstruction using weighted local mean smoother with $W=50$.

Fig. 4.12. EMS reconstruction using weighted local mean smoother with $W=10$.

Fig. 3.1

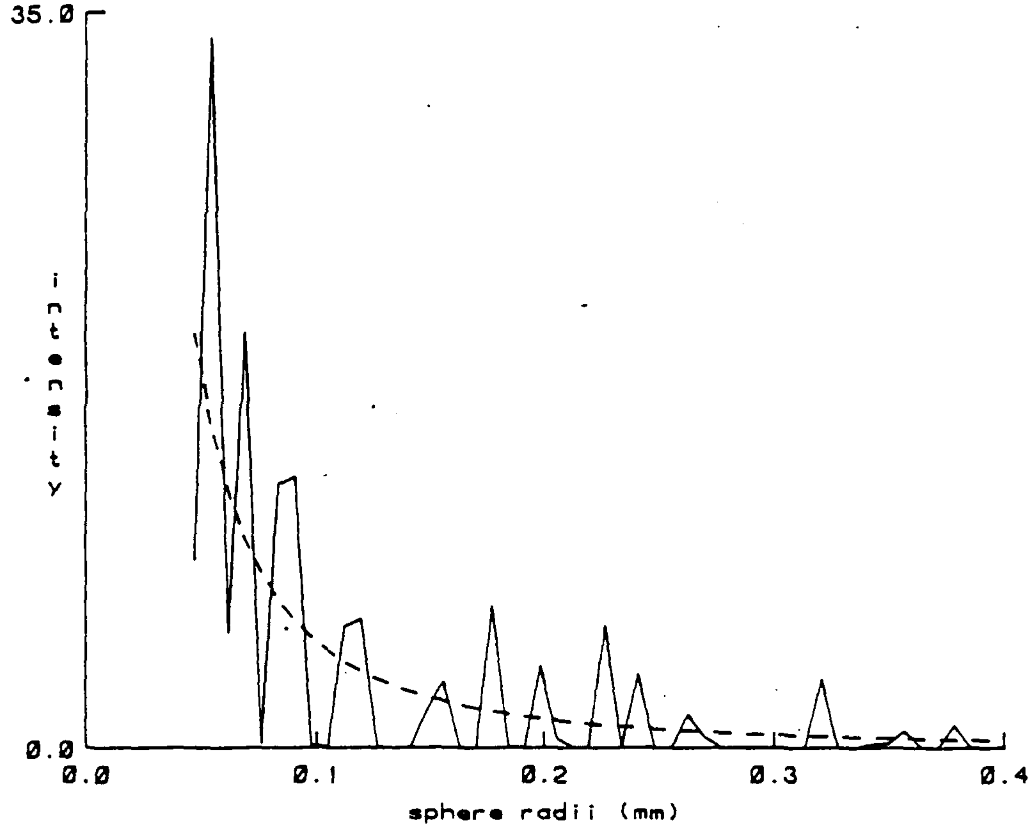

35.0
0.0
intensity
0.0
0.1
0.2
0.3
0.4
sphere radii (mm)

Fig. 3.2

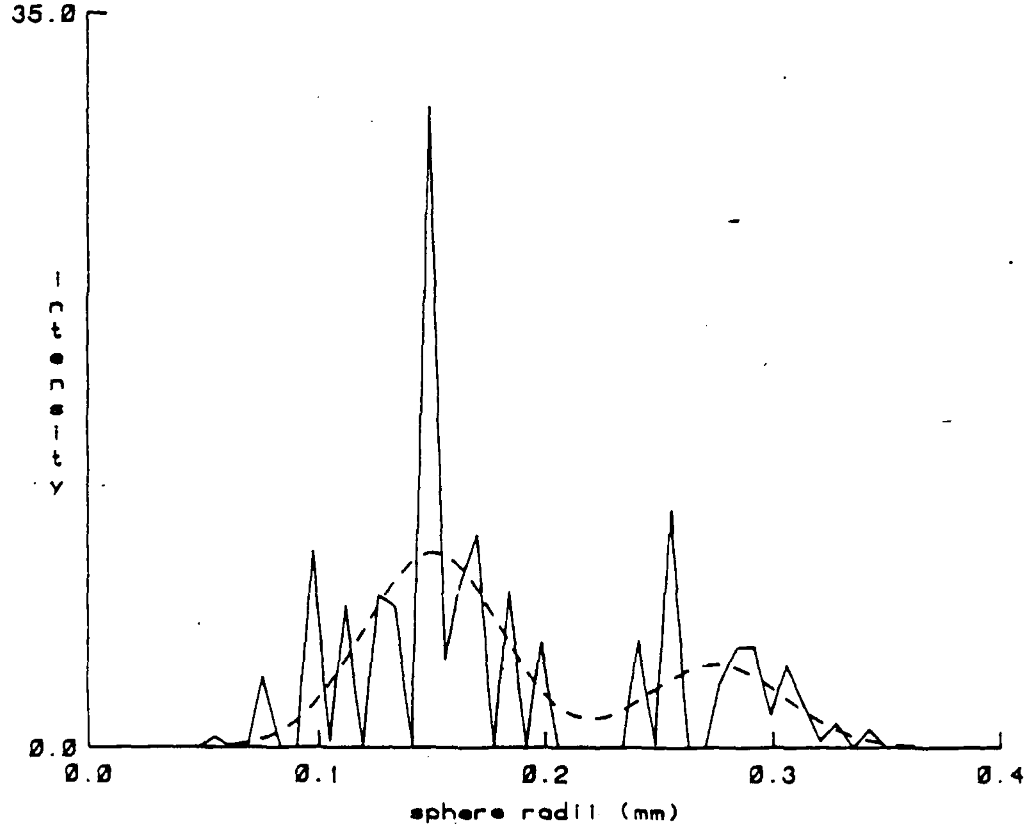

35.0
0.0
intensity
0.0
0.1
0.2
0.3
0.4
sphere radii (mm)

Fig. 3.3

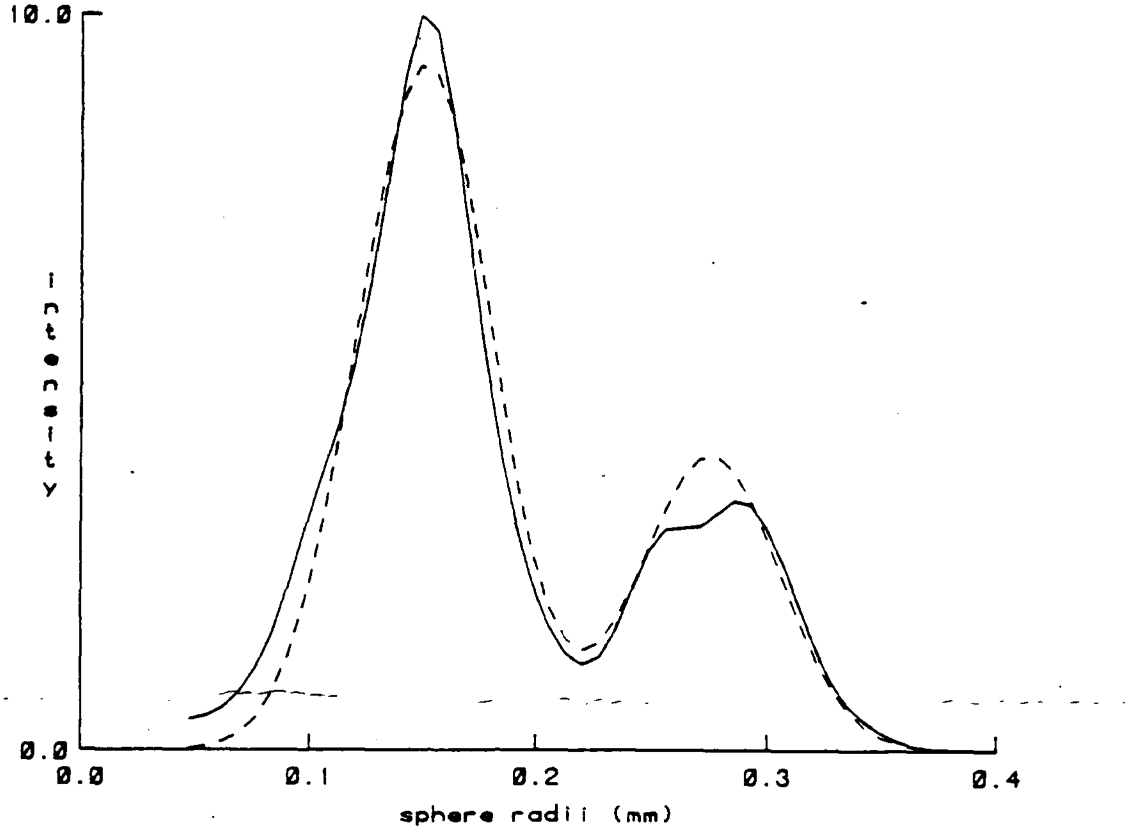

10.0
0.0
intensity
0.0
0.1
0.2
0.3
0.4
sphere radii (mm)

Fig. 3.4

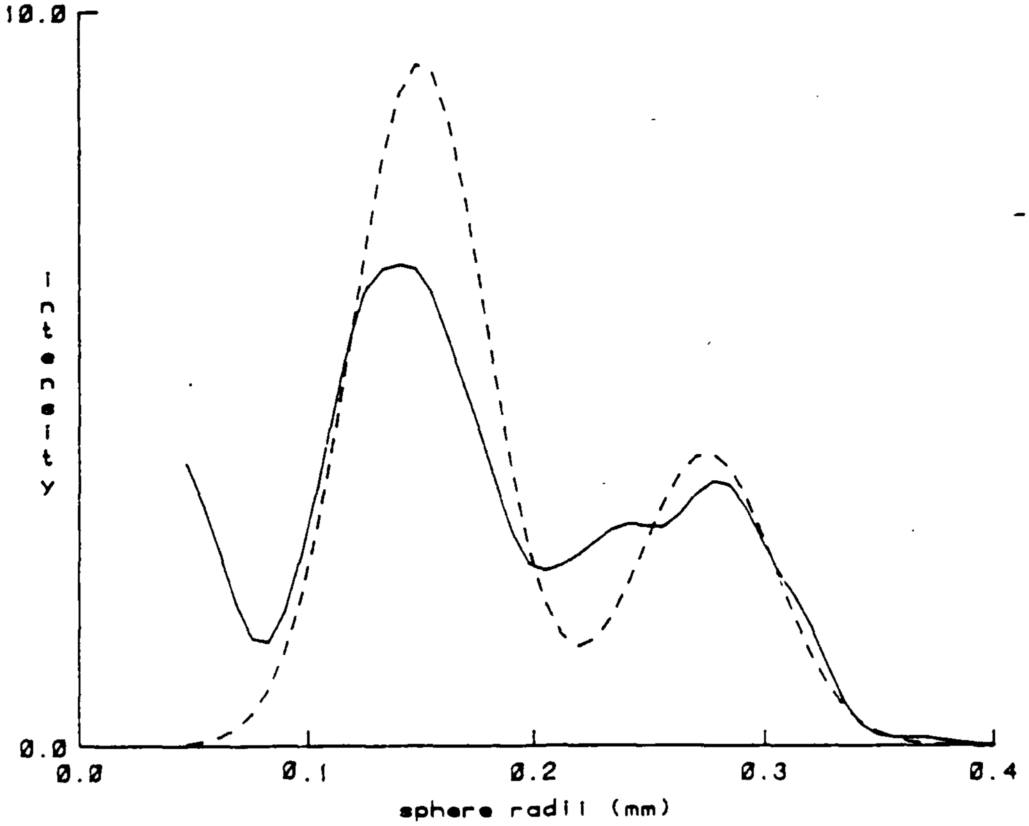

10.0
0.0
intensity
0.0
0.1
0.2
0.3
0.4
sphere radii (mm)

Fig. 3.5

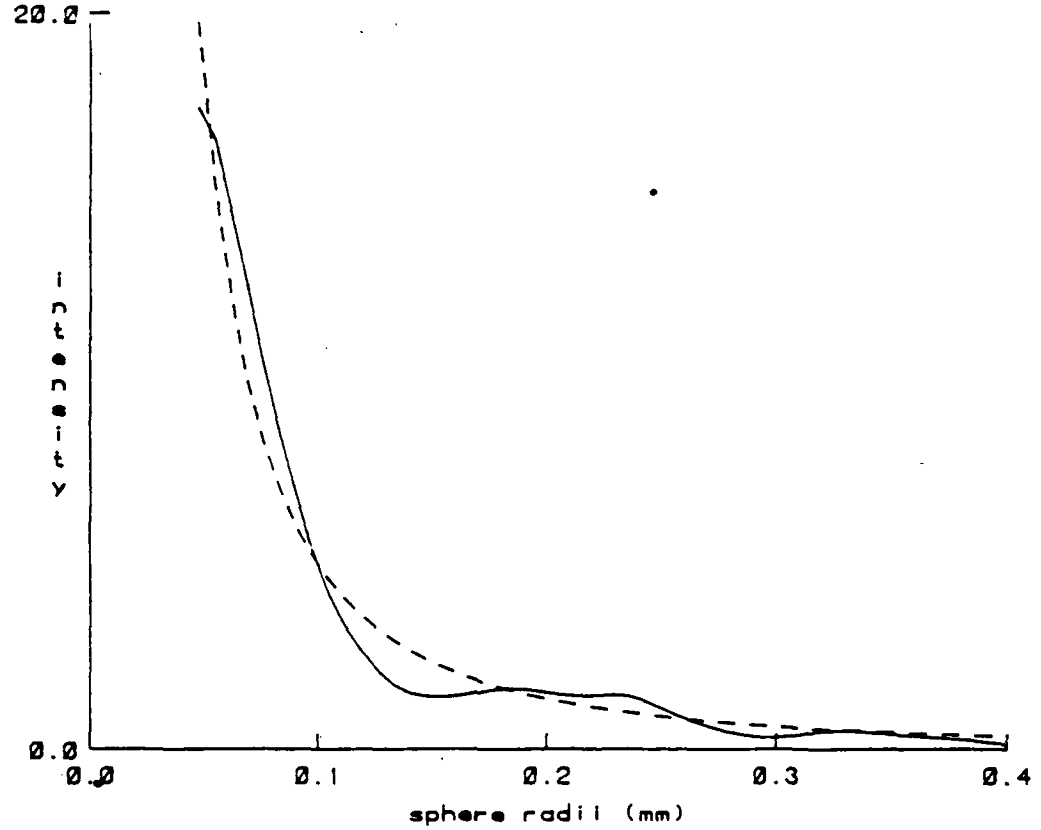

20.0
0.0
intensity
0.0
0.1
0.2
0.3
0.4
sphere radii (mm)

Fig. 3.6

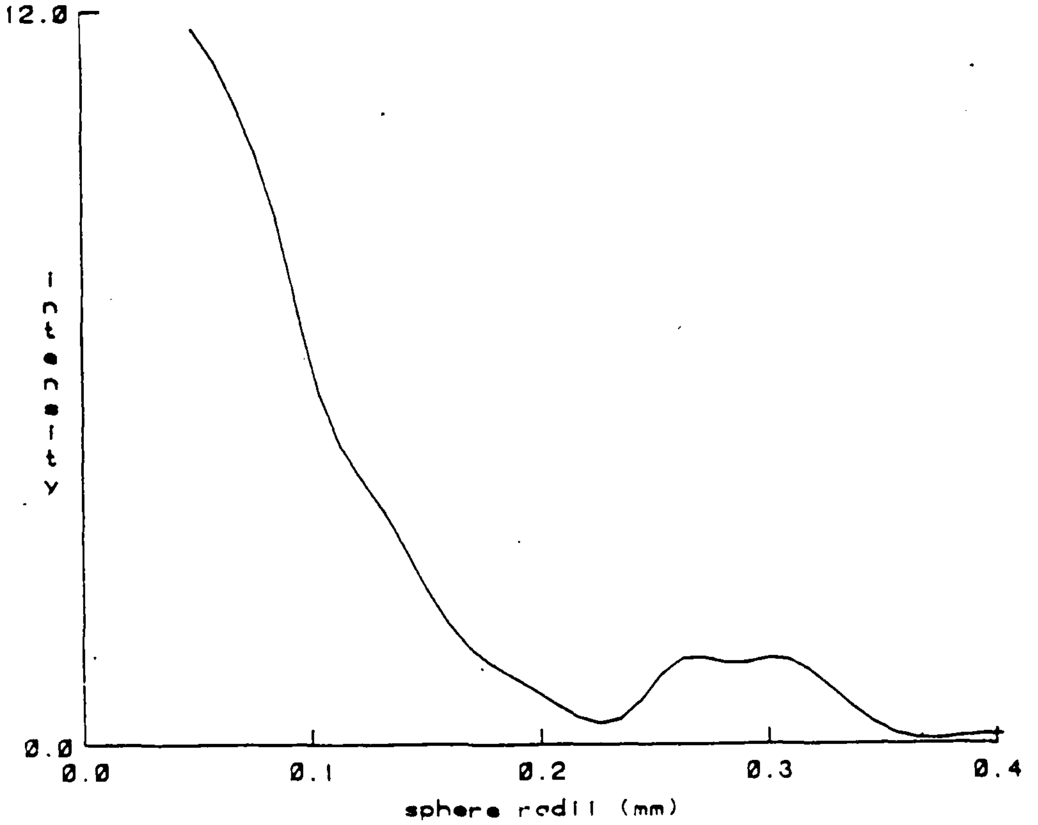

12.0
0.0
intensity
0.0
0.1
0.2
0.3
0.4
sphere radii (mm)

Fig. 4.1

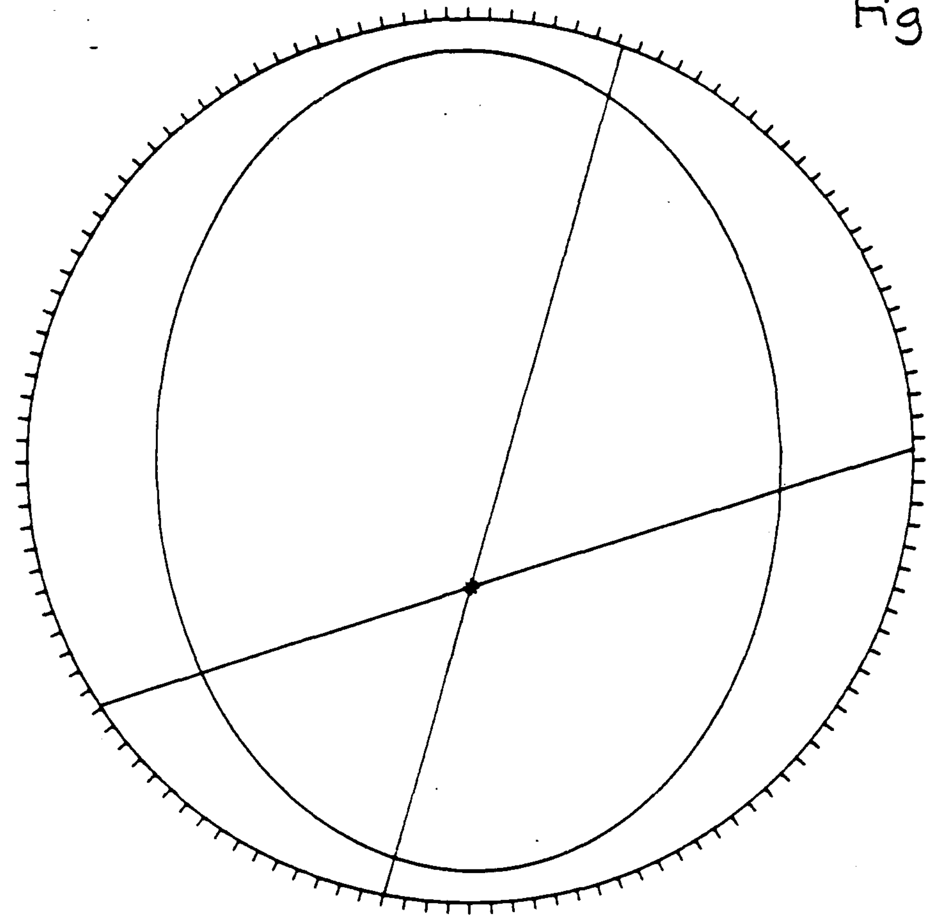


Fig. 4.2

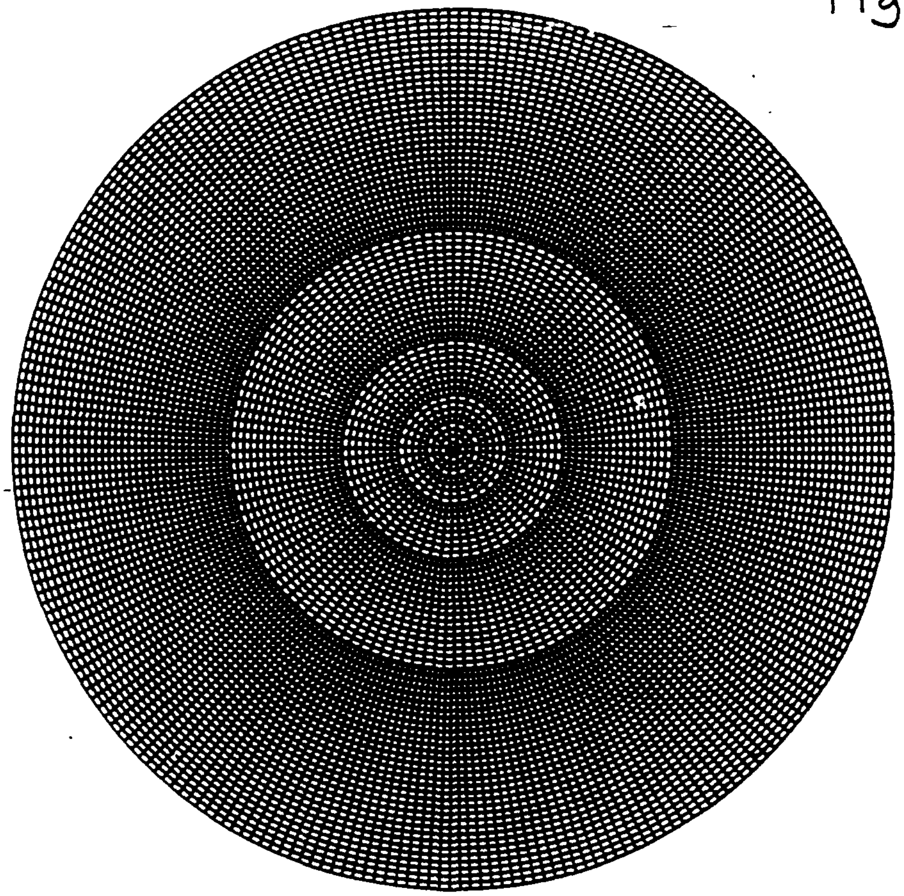
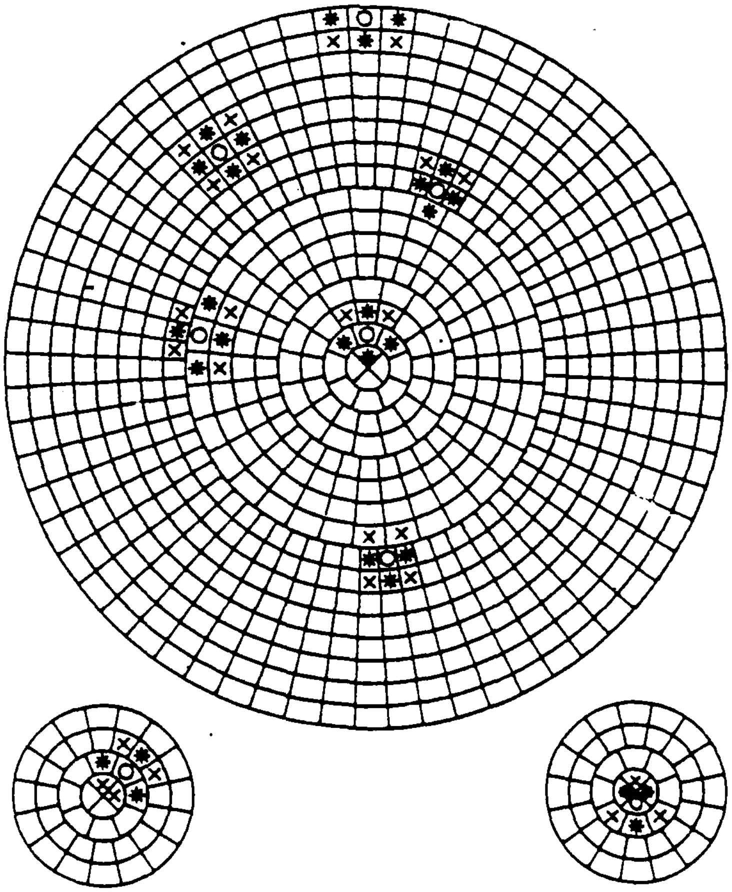
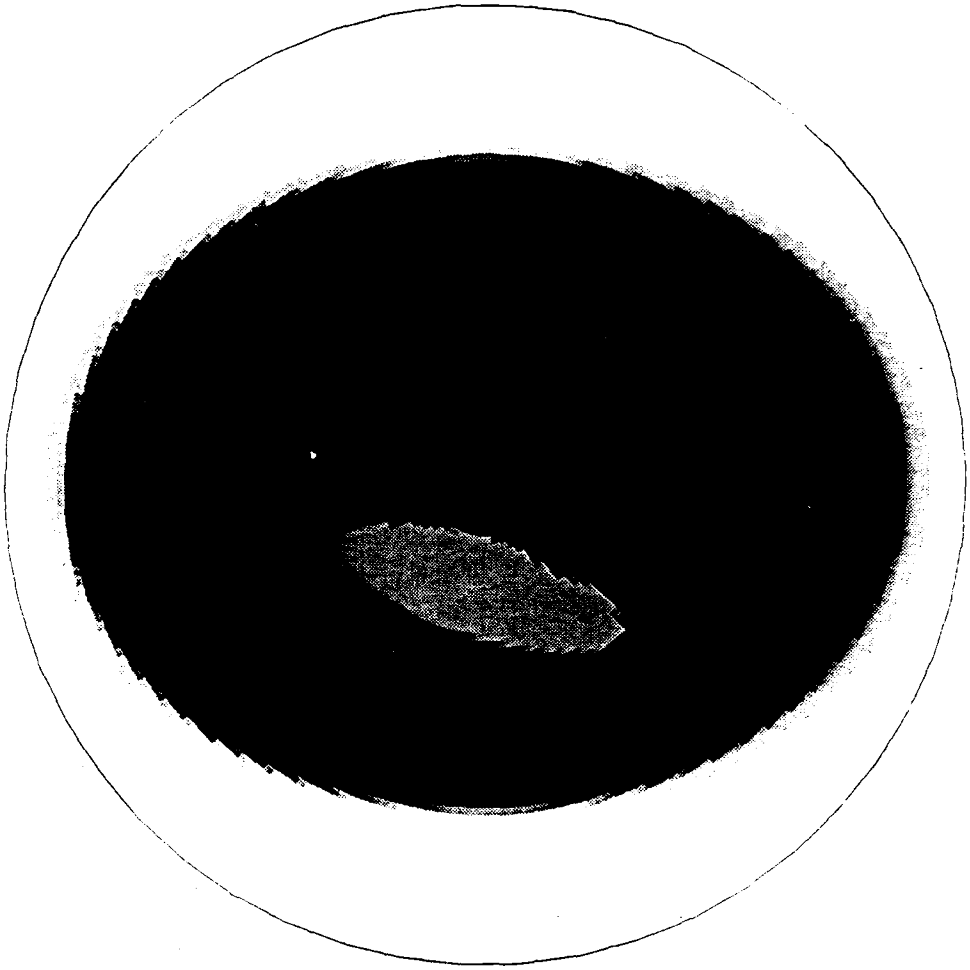
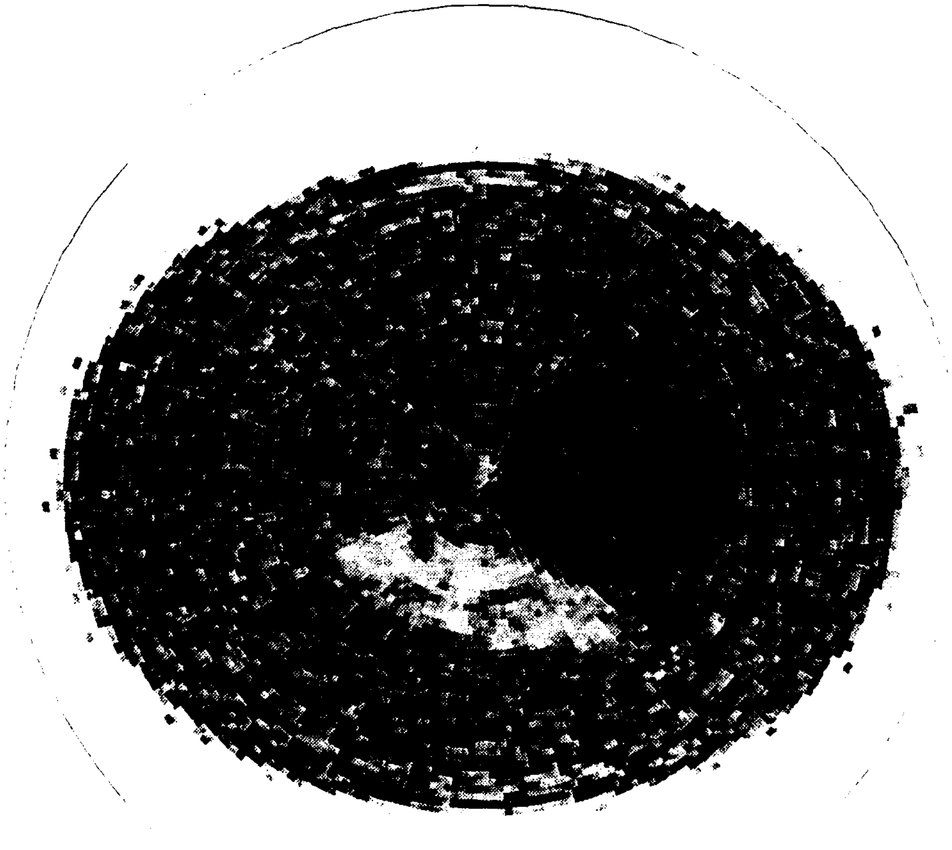


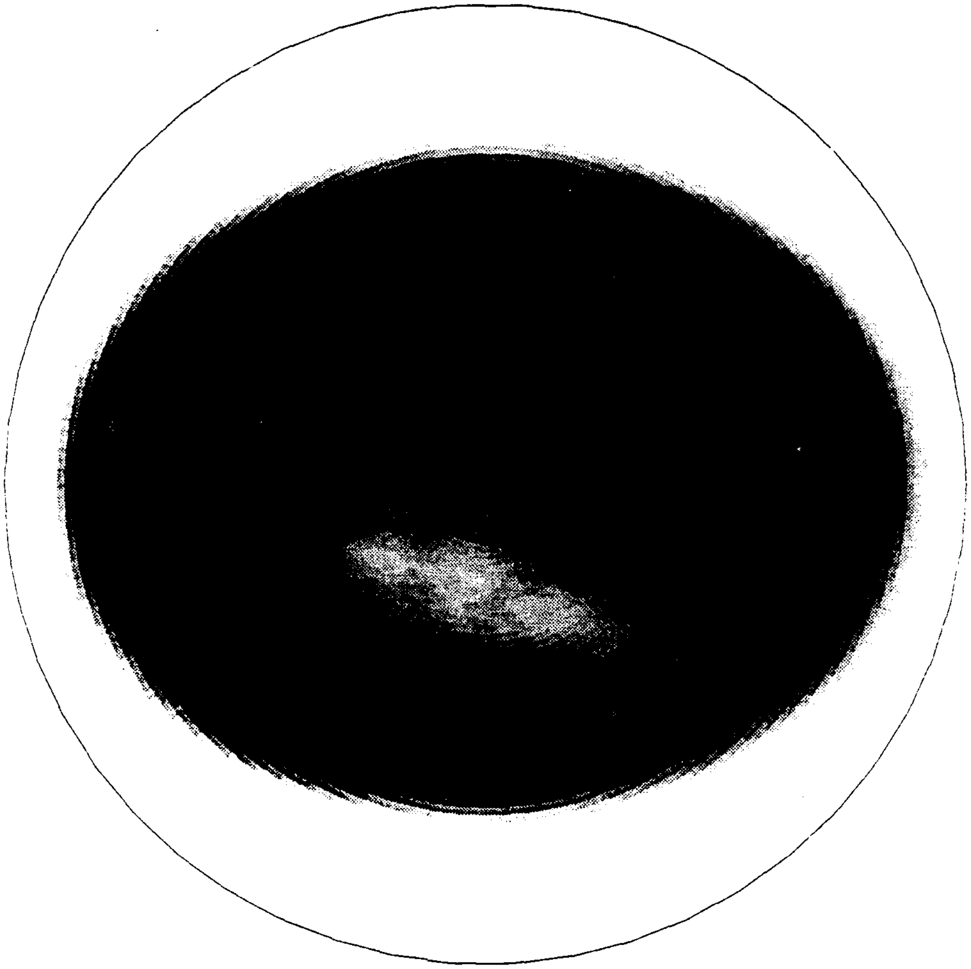
Fig. 4.4

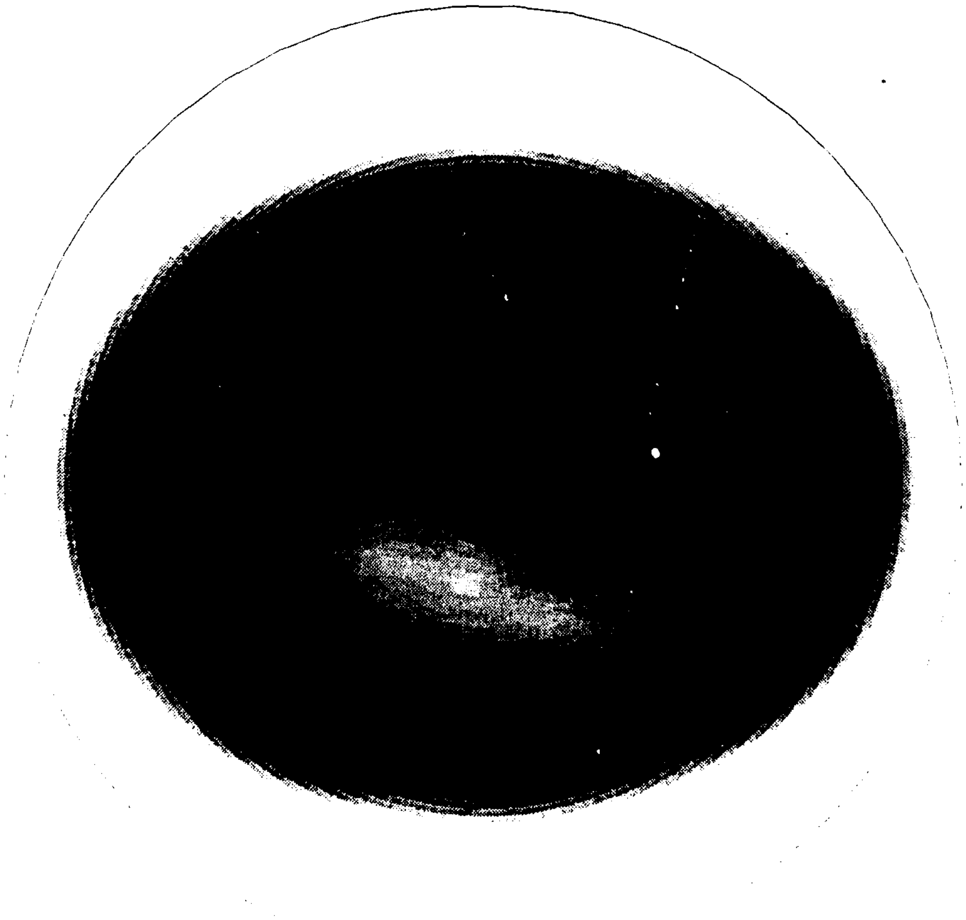
Fig. 4.7

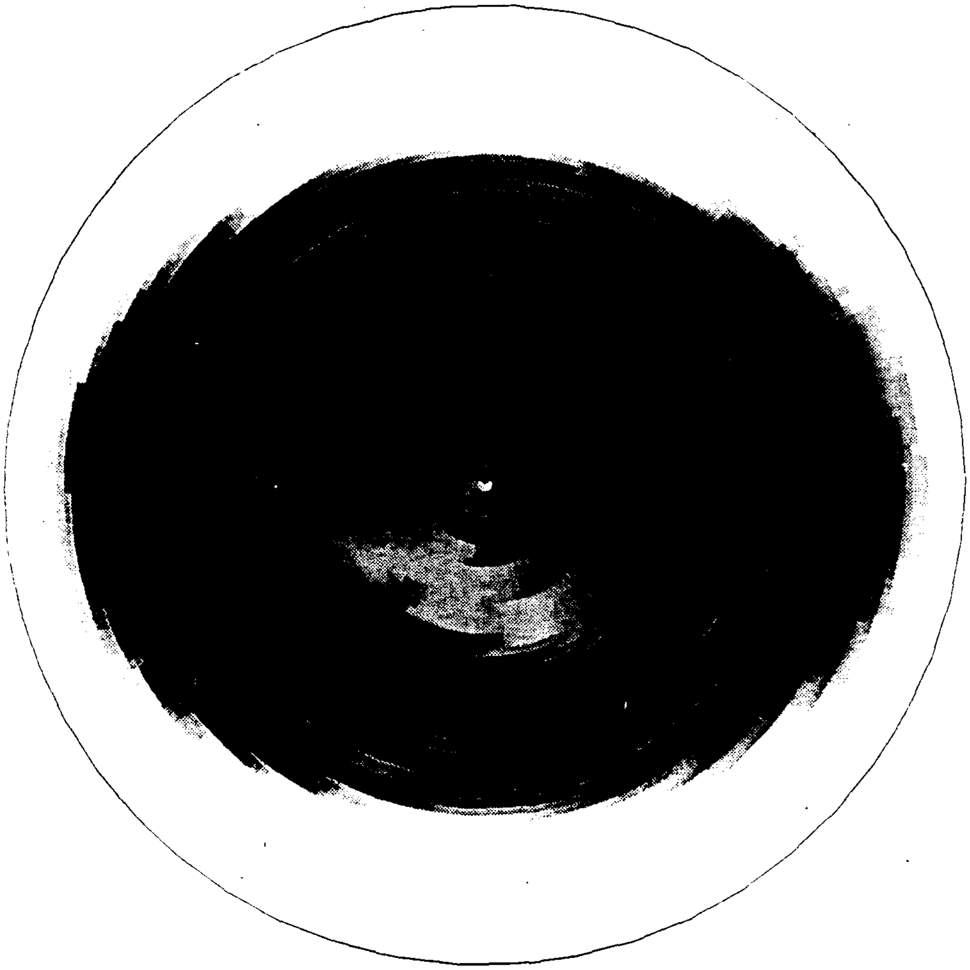


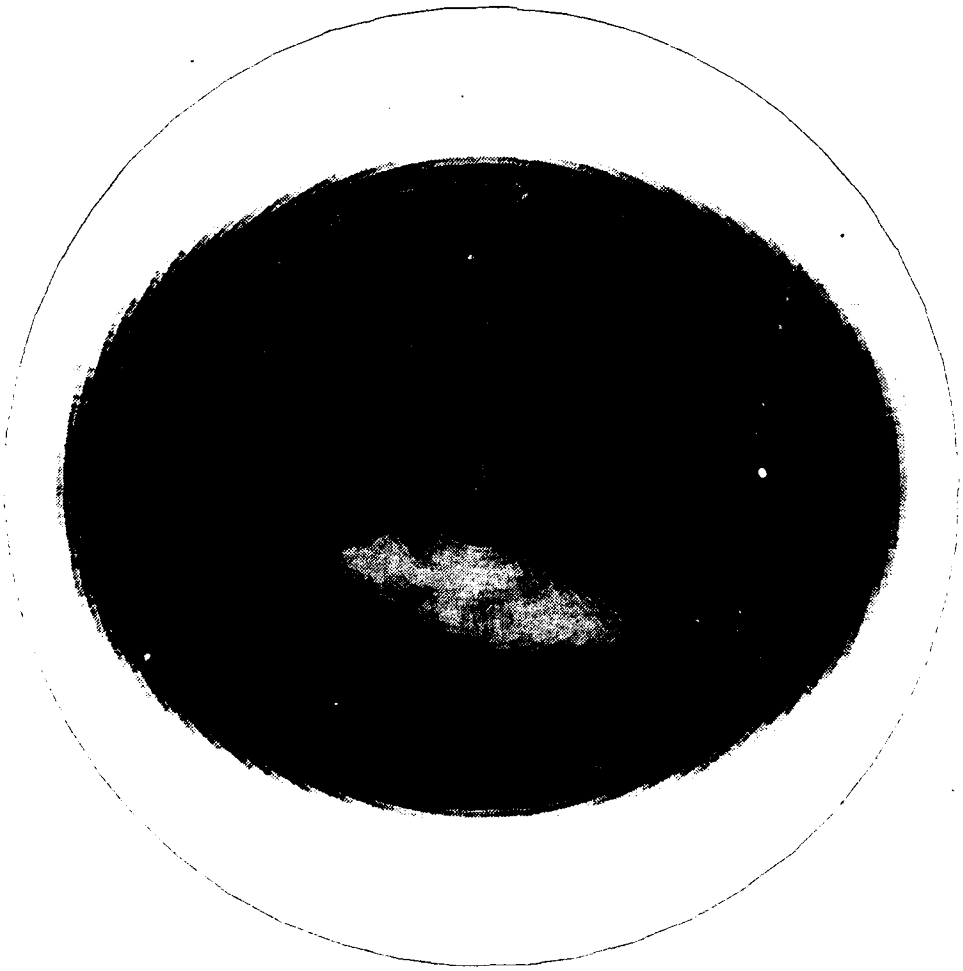












Appendix 12

An orthogonal series density estimation approach to reconstructing positron emission tomography images.

by

M. C. Jones & B. W. Silverman

An Orthogonal Series Density Estimation Approach to Reconstructing Positron Emission Tomography Images

M.C. JONES & B.W. SILVERMAN, *University of Bath*

SUMMARY *Positron emission tomography (PET) is an important medical imaging technique. Statistically, the PET image reconstruction problem comprises estimating the intensity function of a nonhomogeneous Poisson process from a set of indirectly observed data (an integral transform is involved). In this paper, we investigate a new reconstruction method consisting in the adaptation of orthogonal series density estimation techniques to use with an idealised form of the PET problem. The method provides reasonable reconstructions quickly; its computational speed is its major advantage. It has further advantages (e.g. no pixellation required) and various disadvantages (e.g. difficulties with object boundaries, non-negativity not guaranteed) which are discussed. Its major disadvantage, however, is the difficulty associated with generalising the approach to cope with more realistic versions of the PET model.*

1 Introduction

It is often desired to infer something about the internal structure of an object when to look at that structure directly is impossible. Instead, we may be able to obtain measurements external to the object which are, in some way, derived from the internal structure of interest and from which we might hope to be able to estimate that structure. This scenario occurs frequently in medicine. Suppose, for concreteness, interest centres on a patient's brain and, especially, in the metabolic activity in a particular slice through the brain. An idealised image illustrating the kind of pattern of activity we might expect to obtain is shown in Fig. 1. Here, grey levels are used to represent different levels of activity. How do we get at such a useful portrait of unobservable features?

The particular technique for this kind of investigation with which we are concerned in this paper is positron emission tomography (PET). In PET, radioactive material is introduced into the area of interest — often tagged glucose in the brain — with the idea that it distributes itself around in direct proportion to the property (metabolic activity) of interest. The radioactive tracer emits positrons, each of which in turn creates (in conjunction with a nearby electron) a pair of X-ray photons which fly off in opposite directions and which can be detected externally; the point of photon generation corresponding to a typical emission is marked on Fig. 1 by a circle, together with two lines through the point representing two potential photon paths which in fact occur at a uniformly distributed random angle. An array of detectors positioned around the patient — in Fig. 1, they form a division of the outer circle into $D = 128$ arcs of equal length — registers coincident photon arrivals. Thus, our data are the counts amassed in the $T = D(D-1)/2 = 8128$ "tubes" defined by all pairs of detectors. In typical PET applications, the total emission count numbers from several hundreds of thousands upwards. The present and potential usefulness of PET and other medical tomographic techniques is considerable. Research interests in the many stages that make up a complete PET system cover a wide variety of disciplines. There is a number of important statistical questions concerned with PET of which just the most obvious one of best reconstructing the internal image from the external observations is considered here. For a general introduction to PET, see Phelps, Mazziota & Schelbert (1986); for more discussion of the idealised PET setup in which we work here, see Section 2.

PET therefore provides a challenging image analysis problem which differs from many image analysis problems in two important ways. The first of these lies in the indirect nature of the image observation process described above. Many other problems, such as those discussed in Besag (1986) for example, concern noisy direct observation, in the sense that what is observed in each pixel depends only on the true scene's value in that pixel, and not elsewhere, together with some modifying noise process. Here, emissions from completely different areas of the brain contribute to the same data values since all that each datum registration means is that an emission occurred somewhere in the given tube. In fact, observation intensity and image intensity functions are related by an integral transform given in Section 2. The second difference between PET and many other superficially similar problems is that the image of interest is the intensity function of a nonhomogeneous Poisson process — emissions occur uniformly throughout areas of constant activity in Fig. 1 but with rates differing between areas in direct proportion to the respective activity levels — and direct data, if available, would be a realisation of that Poisson process; this contrasts with data which are values of some true regression-type function observed with error.

There are several popular techniques for nonparametric estimation of an intensity function, or equivalently of a probability density function, available in the literature (see Silverman, 1986) for the case of directly observed data. Here, we investigate the application of one of these — orthogonal series intensity estimation — to the PET problem concerning indirect observations. It turns out that the orthogonal series approach extends easily and naturally to the indirect case, at least for one particular idealisation of the PET reconstruction problem; details are given in Section 3.

The current work provides a practically oriented companion paper to the theoretical investigation of Johnstone & Silverman (1988). Johnstone & Silverman were concerned with quantifying the ill-posedness of the PET problem. In particular, they calculated theoretically the order of magnitude of the size of a sample of directly observed positron emissions that would be required to be equivalent to a given sample size of the indirectly observed data which is available in practice, in the sense of yielding equally accurate image reconstructions. They conclude that their "results confirm intuition that for the PET problem, the amount of information available is still substantial, but it is by no means as great as if a sample of . . . direct observations were available". Johnstone & Silverman introduce the orthogonal series intensity estimation method as a purely theoretical device to aid their investigation. They mention that it "might be used as the basis for practical reconstructions". Here, we follow up this suggestion.

Various properties of the orthogonal series intensity estimation approach to PET image reconstruction are investigated in later sections of the paper. In Section 5, the method is applied to a simulated example. It is possible to understand how the orthogonal series smoothing works by displaying pictures of the "equivalent weight function" which a weight function estimate based on direct observations would need to employ to obtain the same answers; this is done in Section 6. In Section 7, a proposal for the automatic choice of the smoothing parameter associated with this method is made.

Broadly speaking, techniques for image reconstruction in PET fall into two categories. On the one hand, the best quality estimates thus far available derive from iterative algorithms which are costly in terms of computer time. One such class of methods is based on the EM algorithm, as developed by Vardi, Shepp & Kaufman (1985), for which much recent interest has centred on incorporating some kind of smoothing — see Silverman *et al.* (1988) for our own contribution to this area and further references. On the other hand, practical PET implementations tend to use different algorithms which are much quicker to compute but sacrifice something in terms of image accuracy. A favourite example of this type is the "convolution backprojection" method described in, for example, Shepp & Kruskal (1978). We see the orthogonal series intensity estimation approach as fitting more into the latter category although the quality of the resulting reconstructions remains fairly good. A major disadvantage of the proposed method, however, is the difficulty associated with generalising the approach to cope with more realistic versions of the PET problem. Further discussion of the pros and cons of the orthogonal series approach is given in the closing Section 9.

2 More on PET

The idealised PET setup that we have briefly described in Section 1 is the one discussed by Vardi *et al.* (1985) in a paper that provides an excellent introduction to the topic for the statistician. In practice, there are a number of potentially important factors — such as time-of-flight considerations, attenuation problems, scattering and so on — which are ignored in this model; they serve to modify the integral transform linking points of emission and the data at hand and should, if possible, be incorporated in practical situations. An effect due to the nonzero thickness of the detectors, first included in the model by Silverman *et al.* (1988), is also omitted in this paper, but see Section 8.

Our notation follows that of Johnstone & Silverman (1988, Section 2.2) and is briefly reviewed here. We first consider an entirely continuous version of the PET model: as well as the naturally continuous "brain space" (the unit disc), parametrised by the usual polar coordinates (r,θ), $0 \le r \le 1$, $0 \le \theta < 2\pi$, suppose the "detector space" consists not of the T tubes of reality, but is the space of all possible unordered pairs of points on the unit circle. Parametrise detector space in a polar fashion too: elements of this space are given by (s,φ), $0 \le s \le 1$, $0 \le \varphi < 2\pi$ where s is the length of the perpendicular from the origin to the detected line and φ is the orientation of that perpendicular (see Fig. 2.2 of Johnstone & Silverman, 1988). It is convenient to renormalise the emission intensity to be a probability density function $f(r,\theta)$, say, with respect to normalised Lebesgue measure μ, where $d\mu(r,\theta) = \pi^{-1} r\, dr\, d\theta$. Write $g(s,\varphi) = (Pf)(s,\varphi)$ for the probability density in detector space with respect to the transformed measure λ given by $d\lambda(s,\varphi) = 2\pi^{-2}(1-s^2)^{\frac{1}{2}} ds\, d\varphi$. The mapping P is the well-known Radon transform of the density f given by

$$(Pf)(s,\varphi) = \tfrac{1}{2}(1-s^2)^{-\frac{1}{2}} \int_{-\sqrt{(1-s^2)}}^{\sqrt{(1-s^2)}} f(s\cos\varphi - t\sin\varphi,\, s\sin\varphi + t\cos\varphi)\, dt. \qquad (1)$$

As is intuitively clear, the Radon transform represents the average value of f over the line connecting the pair of points on the circle. See Johnstone & Silverman (1988) for more details of the above and Deans (1983) for a good introduction to the Radon transform in general.

In reconstructing PET images in this paper, we maintain the continuous nature of brain space but are forced to discretise detector space. The former continuity contrasts with many other reconstruction methods (including those of Vardi *et al.*, 1985, and Silverman *et al.*, 1988) which work with a discrete pixellation of the disc. The latter discretisation of detector space is an irremovable constraint due to the physical setup. We denote the corresponding discrete tubecounts by n_t, $t = 1, \ldots, T$ where the order of indexing tubes by t is immaterial.

3 Appropriate Orthogonal Series Estimation

We wish to estimate the emission intensity f. If direct observations drawn from f were available, the usual orthogonal series estimation paradigm is as follows. Firstly, expand f in terms of orthonormal functions $\{\eta_v\}$ i.e. write

$$f(r,\theta) = \sum_v f_v \eta_v(r,\theta);$$

Secondly, estimate the coefficients $\{f_\nu\}$ by the average of η_ν over the sample of (r,θ)'s; call these $\{\hat{f}_\nu\}$. Finally, introduce some smoothing either by a collection of tapering weights or, as here, by cutting off the potentially infinite sum after some finite number, K, of terms. Our estimate is then

$$\hat{f}(r,\theta) = \sum_{\nu \le K} \hat{f}_\nu \eta_\nu(r,\theta). \tag{2}$$

See Section 2.7 of Silverman (1986) for an account of this approach to density estimation and Section 7 of Izenman (1988) for more references. Since our f is a bivariate function, ν represents a double subscript.

We can equally well expand g as

$$g(s,\varphi) = \sum_\nu g_\nu \psi_\nu(s,\varphi)$$

for appropriate functions $\{\psi_\nu\}$ and use a similar procedure to estimate g. Note that the $\hat{g}_\nu$'s are practically calculable from our indirect data. Now, provided that the orthonormal sets $\{\eta_\nu\}$ and $\{\psi_\nu\}$ are such that there exists a set $\{b_\nu\}$ of positive real numbers with

$$(P\eta_\nu)(s,\varphi) = b_\nu \psi_\nu(s,\varphi), \tag{3}$$

we can write $g_\nu = b_\nu f_\nu$ so that

$$f(r,\theta) = \sum_\nu b_\nu^{-1} g_\nu \eta_\nu(r,\theta).$$

The natural orthogonal series estimate of f based on indirect observations is therefore

$$\hat{f}(r,\theta) = \sum_{\nu \le K} b_\nu^{-1} \hat{g}_\nu \eta_\nu(r,\theta). \tag{4}$$

The fact that a set of quantities with the above properties — a singular value decomposition — exists for the Radon transform (see Deans, 1983, Section 7.6) is what makes the orthogonal series intensity estimation approach applicable to our idealised PET model. In brain space, the appropriate orthonormal functions are

$$\eta_\nu(r,\theta) = (m+1)^{\frac{1}{2}} Z_m^l(r)\, e^{il\theta}. \tag{5}$$

We have written ν as (l,m): $m = 0, 1, 2, \dots$ is what becomes truncated at K, while l varies from $-m$ to m in steps of 2. The functions $Z_m^l(r)$ are the *Zernike polynomials* of degree m and order l which have a history of application in optics (Born & Wolf, 1980). See Deans (1983, Section 7.6) for their properties. In detector space, we take

$$\psi_\nu(s,\varphi) = U_m(s)\, e^{il\varphi} \tag{6}$$

where $U_m(s)$ is a *Chebyshev polynomial of the second kind* (see Deans, 1983, Appendix C). The singular values $\{b_\nu\}$ are given very simply by

$$b_\nu = (m+1)^{-\frac{1}{2}}. \tag{7}$$

Of course, the orthonormal functions in (5) and (6) are real-valued. In each, the quantity of the form $e^{il(\text{angle})}$ is simply a useful shorthand for coping with sine and cosine terms; in appropriate combination, all imaginary terms disappear. More details on the above development can be found in Sections 5 and 6 of Johnstone & Silverman (1988).

The discrete nature of the tubecount data affects the estimates $\{\hat{g}_\nu\}$ of $\{g_\nu\}$. Suppose the line parallel to the sides of tube t but located at its centre has coordinates (s_t, θ_t). Then we use the natural sample average of $\bar{\psi}_\nu$ based on the grouped data, namely

$$\hat{g}_\nu = N^{-1} \sum_{t=1}^{T} n_t \bar{\psi}_\nu (s_t, \varphi_t), \tag{8}$$

the bar denoting complex conjugation. Here, $N = \sum n_t$ is the total number of emissions. Plugging (7), (8) and the definitions (5) and (6) into (4) yields a complete description of $\hat{f}$ (for fixed K). Recursions involved in calculating both types of orthogonal polynomial help to keep the computational burden down.

4 Presentation of Figures

Figs 2 to 8 are all grey level images of PET image reconstructions and related quantities. Each uses 32 grey levels scaled in a rather arbitrary way, increases in darkness representing increases in (estimates of) metabolic activity. Orthogonal series intensity estimation results in (high order) polynomial surfaces defined at all points of the disc. Representing such smooth functions is a task well suited to the application of a high quality contouring package; in our figures, we have used the excellent CONICON3 programs of Sibson (1987). The grey level images result from suppressing drawing of the contours themselves and filling in the areas between successive contours with appropriate shades of grey. CONICON3 requires value and gradient information on the function to be contoured only at a regular grid of values — a 20 × 20 square grid usually sufficed here. Computation and presentation of the images given in this paper were performed on a SUN 3/160 workstation, copies of the pictures being produced by an Apple LaserWriter II printer.

5 A *Simulated Example*

We illustrate use of the orthogonal series intensity estimation algorithm on data simulated from the image — the "phantom" — shown in Fig. 1. This phantom is a piecewise constant function made up of elliptical areas of constant intensity (representing ventricles, tumours and so on) on a large background ellipse (the head). The key property of this idealisation which, we believe, transfers to real images is the presence of edges of features at which there may be a considerable jump in intensity; ideally, we would like to estimate such edges well. The constancy property (within objects) may prove less realistic than some kind of smooth variation, but this is less of an issue. This phantom is essentially the same as that of Fig. 2 of Vardi *et al.* (1985) and was also used in Silverman *et al.* (1988). Fig. 1 has something of a discretised look about it, having been obtained by using CONICON3 on a fine 100 × 100 grid; this comes about since we are applying CONICON3 to an entirely inappropriate piecewise constant function! Nonetheless, Fig. 1 bears comparison with the discretised version of the phantom given as Fig. 4.3 of Silverman *et al.* (1988), giving a good impression of the features present in the image and serving as a kind of bound on how well the true phantom could be reconstructed using the representation tools at hand. A total of $N = 10^6$ emissions — commensurate with real applications — was generated from this intensity function using the acceptance/rejection method in the obvious way. The corresponding tubecounts form the data for this experiment.

Figs 2 to 4 are three reconstructions obtained from these data; they correspond to $K = 10$, 36 and 50, respectively. The first (Fig. 2) is clearly oversmooth. It is encouraging that even here large features present in the phantom are reproduced to some extent but the total disappearance of the smaller objects gives cause for concern. Figs 3 and 4 are progressively less smooth. By the time $K = 50$ (Fig. 4) it can be argued, given knowledge of the true image, that even the smaller features are indicated fairly well but, of course, that (practically unobtainable) knowledge is required to differentiate the small objects on the reconstruction that should be there from the others such an undersmoothed reconstruction gives that should not. On balance, the choice $K = 36$ (Fig. 3) seems to be about as good as we can get. Large features are well represented; there can be rather less confidence, though, in the smaller structure. Of course, the smooth polynomial nature of our reconstruction method is a drawback when, as here, piecewise smooth areas with considerable discontinuities in value at feature boundaries make up the true image. The reader is left to append his or her own adjectives to the goodness or otherwise of Fig. 3 as an approximation to Fig. 1!

White areas in Figs 2 to 4 are below the zero contour. The presence of such negativity in our reconstructions is a property of the method that may be felt to be undesirable; we note, at least, that negativity occurs in these figures only outside the head region where there are no emissions in reality. Towards the outside of brain space, some increase in estimated intensity levels is an edge effect which should be ignored.

6 Equivalent Weight Functions

Many density/intensity estimation methods can be written in the form of general weight function estimators (e.g. Silverman, 1986, Section 2.9). In the usual case where $\hat{f}$ is obtained from direct observations $\{(r_i, \theta_i)\}$ as in (2), we can write

$$\hat{f}(r,\theta) = N^{-1} \sum_{i=1}^{N} w((r_i,\theta_i),(r,\theta)) \tag{9}$$

where the weight function w is given by

$$w((R,\Theta),(r,\theta)) = \sum_{\nu \le K} \bar{\eta}_\nu(R,\Theta)\, \eta_\nu(r,\theta). \tag{10}$$

When such direct observations are available from the density of interest, the weight function expresses how a particular observation is smoothed out in making its contribution to the overall estimate and hence gives insight into the nature of the smoothing process; see, for example, Silverman (1984) for another relevant context.

Since the PET observation process is an indirect one, some modification of the above discussion is necessary. An appropriate alternative definition of the weight function, equivalent in the case of direct sampling, is as an "impulse response function". That is, suppose that the true image consisted of a point mass at (R,Θ) and that N indirect observations from this image were taken. Then, ignoring the tubecount discretisation and with the degree of smoothing held fixed, it is easily shown that $\hat{f}(r,\theta)$ based on these data approaches w in (10) as $N \rightarrow \infty$. Thus, w remains the appropriate weight function to study in the case of indirectly observed data too. As w and an alternative version of w which properly takes the data discretisation into account are virtually indistinguishable, we have not incorporated the data discretisation modification here.

To make more of the above we present some pictorial illustrations. Fig. 5 shows w for $R = 0$ and $\Theta = 0$. Again, grey scale images are used in an obvious way (although the overall scaling of the pictures in this section differs from that in Section 5). White areas again define regions of negativity. The main features of Fig. 5 are the spherical symmetry of the weight function and its concentration about the point (0,0). As in the familiar kernel estimation approach, w has a mode at the point of interest and falls smoothly away, resulting in an averaging over neighbouring values whose influence becomes less as their distance from the centre increases. Beyond this central area, w is small but not always positive; rather, there is a smooth fluctuation about zero, resulting in a series of low positive peaks and shallow negative troughs. In Fig. 5, we have taken $K = 10$. Larger values of K smooth less by narrowing the scope of the main part of the weight function and thus averaging significantly over fewer neighbouring points.

The choice $\Theta = 0$ in Fig. 5 is quite general; w is rotation equivariant so other Θ's result simply in rotations of the $\Theta = 0$ picture. Different R's are worthy of further consideration, though: in Fig. 6, we take $R = 0.5$ and in Fig. 7, $R = 0.9$. The general pattern of a peak at the point of interest, a smooth falling away of w in a neighbourhood of the point and the small positive/negative fluctuations in the tails persist. The weight function is, of course, no longer essentially spherically symmetric but rather is distorted somewhat in a way consistent with fitting w appropriately into the disc. What is important about Figs 5 to 7 is that the amount of smoothing (essentially the extent and shape of the area in which w is significantly nonzero) does not differ greatly at different points in brain space. Varying degrees of smoothing in response to properties of f is an option (not considered here) that may well be desirable; varying degrees of smoothing purely as a geometric function is not.

7 Automatic Choice of Smoothing Parameter

We saw in Section 5 how the parameter K controls the level of smoothing applied to the data. Subjective choice of smoothing parameter, as there, is sufficient in many applications of smoothing techniques but, in PET imaging, a fully automatic procedure, and thus an automatic method for choosing K appropriately, might well be thought desirable. In this section, we illustrate how a rather natural approach to choosing the smoothing parameter in orthogonal series density estimation in general adapts to the PET case.

Suppose we consider the mean integrated squared error (MISE) to be an appropriate measure of discrepancy between f and $\hat{f}$. The following development is entirely analogous to the Fourier series density estimation case worked out in Hart (1985) and references therein. First note that

$$\iint |\hat{f}(r,\theta) - f(r,\theta)|^2 \, d\mu(r,\theta) = \sum_{\nu \le K} b_\nu^{-2} |\hat{g}_\nu - g_\nu|^2 + \sum_{\nu > K} b_\nu^{-2} |g_\nu|^2 .$$

Now ignore the tube discretisation for the moment (i.e. define $\hat{g}_\nu$ like $\hat{f}_\nu$ in (2) rather than by using (8)), so that $\mathrm{E}(\hat{g}_\nu) = g_\nu$ and $\mathrm{Var}(\hat{g}_\nu) = N^{-1}\sigma_\nu^2$ where $\sigma_\nu^2 = \mathrm{Var}(\psi_\nu(S,\Phi))$. Taking expectations in the above expression, we get

$$\mathrm{MISE} = \sum_{\nu \le K} b_\nu^{-2}(N^{-1}\sigma_\nu^2 - |g_\nu|^2) + \iint |f(r,\theta)|^2 d\mu(r,\theta). \tag{11}$$

The value of K that minimises this MISE is a candidate for being a good choice of K for the PET problem. Of course, we do not know MISE or its optimal K. Rather, we drop the second term in (11) from further consideration because it is independent of K and estimate the first term, I, say, as best we can; choosing K to minimise this estimate ($\hat{I}$) yields a practical procedure which, it is hoped, comes close to using the truly optimal value of K. It is not difficult to show that

$$\hat{I} = \sum_{\nu \le K} b_\nu^{-2}(N-1)^{-1} \{2\hat{s}_\nu - (N+1)\,|\hat{g}_\nu|^2\} \tag{12}$$

is an unbiased estimate of I, so it is this formula that we minimise. In (12), $\hat{s}_v$ is the sample average of the $|\psi_v|^2$'s. In practice, we are stuck with the discretisation of detector space, so we use

$$\hat{s}_v = N^{-1} \sum_{t=1}^{T} n_t \, |\psi_v(s_t, \varphi_t)|^2$$

and $\hat{g}_v$ as in (8) in $\hat{I}$.

Since K is an integer, it is straightforward to minimise (12) by evaluating it over a range of values of K; $(D-1)$ is an upper bound to this range due to an aliasing effect although the optimal K is most likely to be much less than this anyway. Doing this for the simulated example yields $K = 36$ and the corresponding figure is Fig. 3. This is the image we preferred in Section 5 on subjective grounds. Of course, on the basis of this one example only we make no great claims for the supremacy of our automatic procedure. For one thing, there is always scope for wayward choices due to errors in estimating the MISE-optimal K. More importantly, the propriety or otherwise of MISE as risk function is in question. It is widely recognised that this type of measure does not give a good reflection of the human observer's sense of image fidelity especially when, as here, the true image contains features with distinct edges. The provision of image metrics that properly reflect visual perception remains a difficult question; see Baddeley (1987) for some ideas. We persevered with the MISE development above largely on grounds of tractability but are encouraged by the results: it is to be hoped that alternative image metrics would also be open to a similar kind of approach.

Replacing the simple cutoff K in (4) by a sequence of weights $\{w_v\}$ remains an alternative option but is one with similar problems of smoothing parameter choice. Johnstone & Silverman (1988, Section 7) discuss optimal weight sequences for MISE; these are, as is to be expected, not immediately practicable because they depend on the true f. Otherwise, we might experiment with *ad hoc* weight sequences; the formulae in Wahba (1981) become one possibility. These have not been pursued here.

8 The Third Dimension Effect

Photon lines are in reality distributed uniformly in 3-dimensional space, not just in the plane, and detectors have a finite depth, d. This effect of the third dimension is not incorporated into the reconstruction scheme above although it is important because it persists even when $d \to 0$ i.e. our 2-dimensional model differs from the limit of the 3-dimensional one. It turns out that, to a good approximation, the third dimension effect results in a weighted Radon transform, the weight factor being inversely proportional to the length of the detector tube (or, at least, its continuous analogue); see Section 4.1.4 of Silverman *et al.* (1988) and Section 10.2 of Johnstone & Silverman (1988) for details.

We have not yet managed to modify the orthogonal series estimation approach to cope with this. Rather, here we demonstrate the considerable effect that failure to do so has on quality of image reconstruction. We can easily simulate data from the phantom of Fig. 1 taking the third dimension into account by adding a further acceptance/rejection step to deal with the inverse length bias; in fact, the resulting dataset is precisely that used by Silverman *et al.* (1988) in their simulation example. Applying the current (2-dimensional) reconstruction algorithm (here with $K = 36$) to these (3-dimensional) data gives Fig. 8; compare this with Fig. 3 in particular. The third dimension effect on the data is clear: a smaller proportion of emissions occurring towards the centre of the brain space will be detected than of those occurring nearer to the edge. The consequences for the 2-dimensional reconstruction are equally clear: greater intensities are attributed to outer regions than should be the case, while central areas suffer the reverse mistake.

9 Discussion

That the orthogonal series intensity estimation approach to PET image reconstruction is quick compared with iterative procedures is borne out by the approximately 30-fold improvement in computer time we have observed in comparison with the best EMS procedure of Silverman *et al.* (1988). That it also suffers in comparison in terms of important image quality criteria is also evident in at least three major ways:

(I) The smoothness of images made up of polynomials is not consistent with the presence of edges which, we argued in Section 5, are most likely to be an important feature of the real images we set out to reconstruct.

(II) There cannot be areas of the brain emitting negative numbers of positrons! EM (Vardi *et al.*, 1985) and EMS (Silverman *et al.*, 1988) algorithms naturally result in non-negative reconstructions; as we have seen, the orthogonal series approach does not.

(III) The Zernike/Chebyshev polynomial based approach is appropriate only to direct and indirect observation spaces being linked by the basic Radon transform. Early in Section 2 we noted the many modifications to this transform that are needed to properly model the practical situation. The major obstacle to use of the orthogonal series approach in more realistic circumstances is the need to obtain the singular value decomposition associated with the correct integral transform. Note that for EM-based approaches, it is only necessary (for many modifications) to identify the right transform and to discretise it to get the $p(b,d)$'s of Vardi *et al.* (1985).

The orthogonal series approach has further advantages as well as disadvantages.

(i) It is straightforward to understand in the sense that it is a fairly direct application of a well-known technique.

(ii) That there is no need to discretise brain space to facilitate reconstruction is particularly nice; the truly continuous nature of orthogonal series reconstruction is, conceptually, most appealing.

Acknowledgements

We are most grateful to Iain Johnstone for his continuing interest in this subject and to Robin Sibson and Glenn Stone for much assistance with computational matters. Financial support was provided by the Science and Engineering Research Council and the U.S. Army European Research Office.

Correspondence: Dr M.C. Jones, Statistics Group, Mathematical Sciences Department, IBM Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598, U.S.A.; Prof. B.W. Silverman, School of Mathematical Sciences, University of Bath, Bath BA2 7AY, U.K.

REFERENCES

BADDELEY, A.J. (1987) A class of image metrics. To appear.

BESAG, J.E. (1986) On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society* Ser. B, *48*, 259-302.

BORN, M. & WOLF, E. (1980) *Principles of Optics* (sixth edition), Pergamon Press, Oxford.

DEANS, S.R. (1983) *The Radon Transform and Some of its Applications*, Wiley, New York.

HART, J.D. (1985) On the choice of a truncation point in Fourier series density estimation. *Journal of Statistical Computation and Simulation*, *21*, 95-116.

IZENMAN, A.J. (1988) A review of nonparametric density estimation. *Journal of the American Statistical Association*, to appear.

JOHNSTONE, I.M. & SILVERMAN, B.W. (1988) Speed of estimation in positron emission tomography. To appear.

PHELPS, M.E., MAZZIOTA, J.C. & SCHELBERT, H.R. (eds) (1986) *Positron Emission Tomography and Autoradiography. Principles and Applications for the Brain and Heart*, Raven Press, New York.

SHEPP, L.A. & KRUSKAL, J.B. (1978) Computerized tomography: the new medical X-ray technology. *American Mathematical Monthly*, *85*, 420-439.

SIBSON, R. (1987) CONICON3 handbook. University of Bath.

SILVERMAN, B.W. (1984) Spline smoothing: the equivalent variable kernel method. *Annals of Statistics*, *12*, 898-916.

SILVERMAN, B.W. (1986) *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

SILVERMAN, B.W., JONES, M.C., WILSON, J.D. & NYCHKA, D.W. (1988) A smoothed EM approach to a class of problems in image analysis and integral equations. To appear.

VARDI, Y., SHEPP, L.A. & KAUFMAN, L. (1985) A statistical model for positron emission tomography (with comments). *Journal of the American Statistical Association*, *80*, 8-37.

WAHBA, G. (1981) Data-based optimal smoothing of orthogonal series density estimates. *Annals of Statistics*, *9*, 146-156.

FIGURE LEGENDS

Fig. 1. An idealised PET image within a circular array of detectors. Two possible photon lines arising from an emission at O are superimposed.

Fig. 2. Reconstruction with $K = 10$.

Fig. 3. Reconstruction with $K = 36$.

Fig. 4. Reconstruction with $K = 50$.

Fig. 5. An equivalent weight function corresponding to $R = 0$.

Fig. 6. An equivalent weight function corresponding to $R = 0.5$.

Fig. 7. An equivalent weight function corresponding to $R = 0.9$.

Fig. 8. Reconstruction ($K = 36$) arising from data incorporating the third dimension effect.

Fig 1

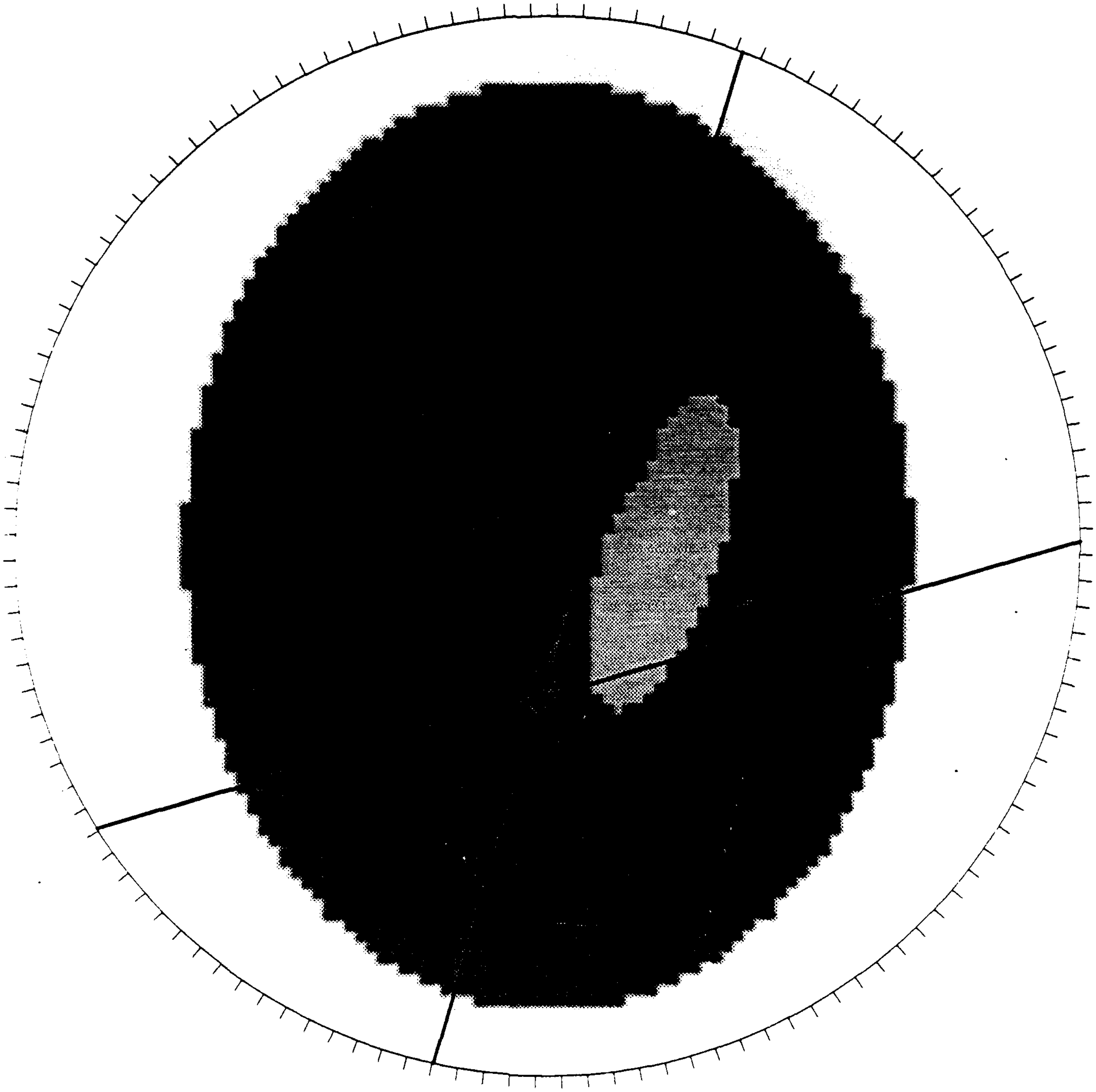


Fig 2

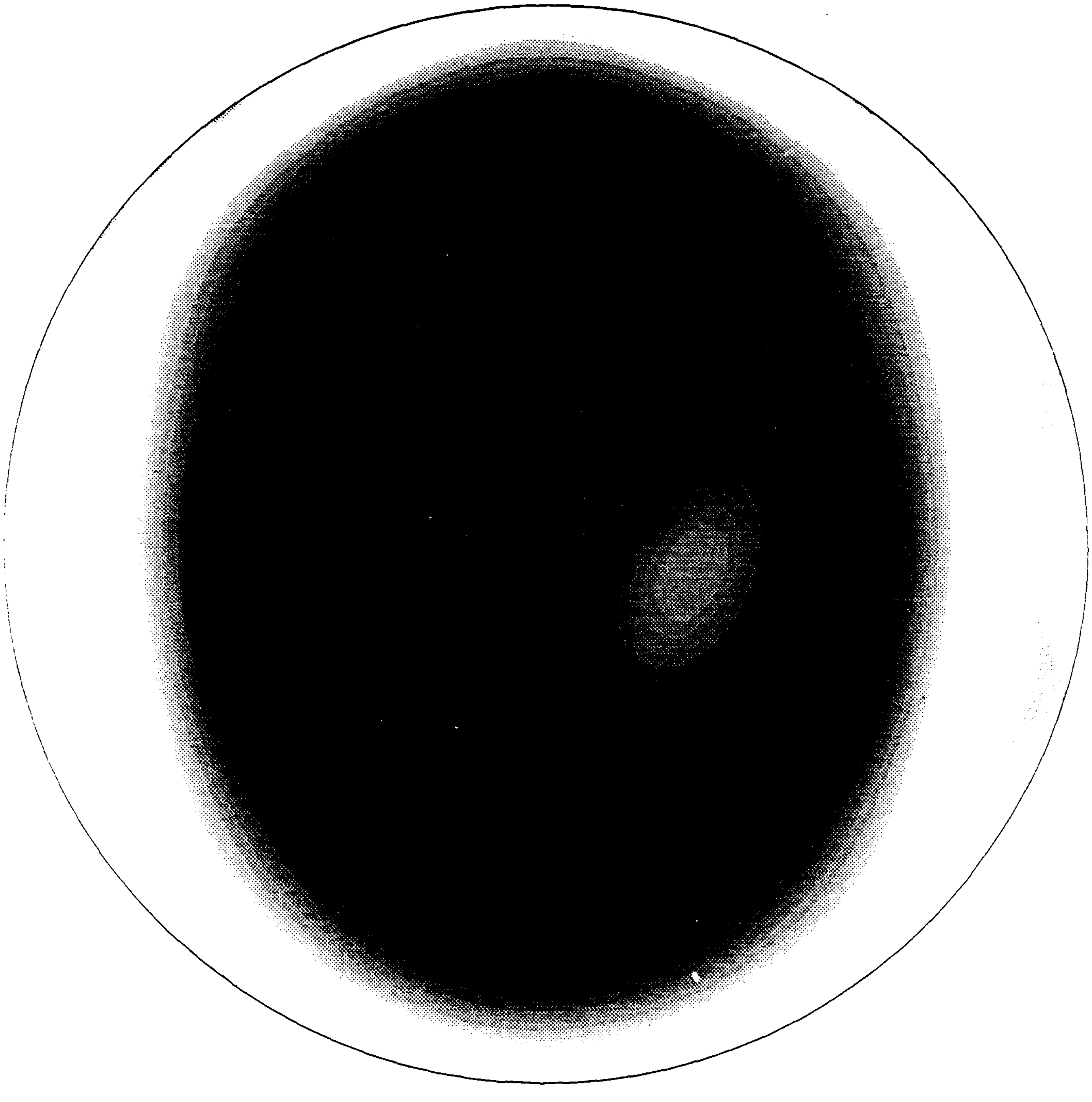


Fig 3

Fig 4

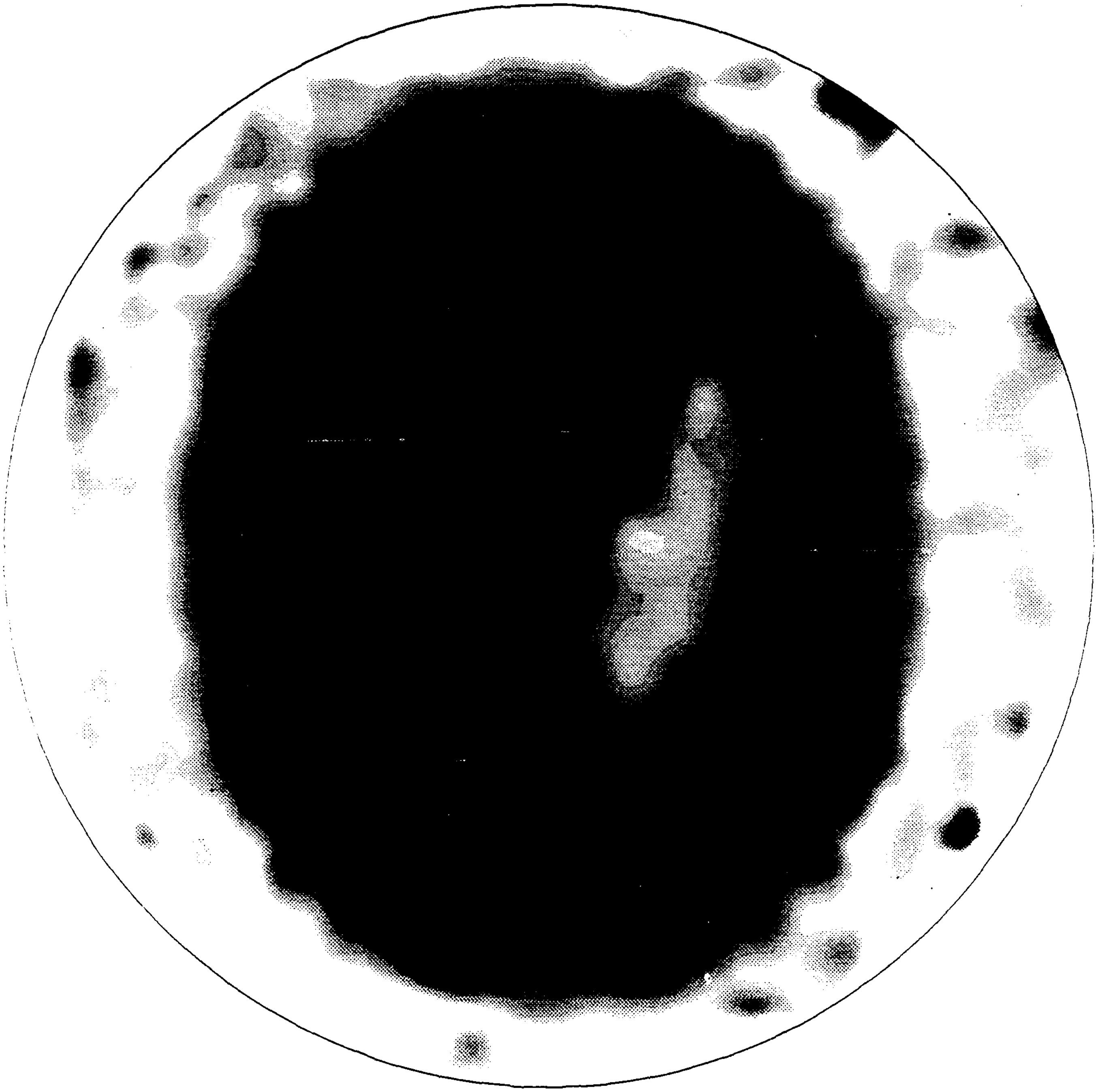


Fig 5

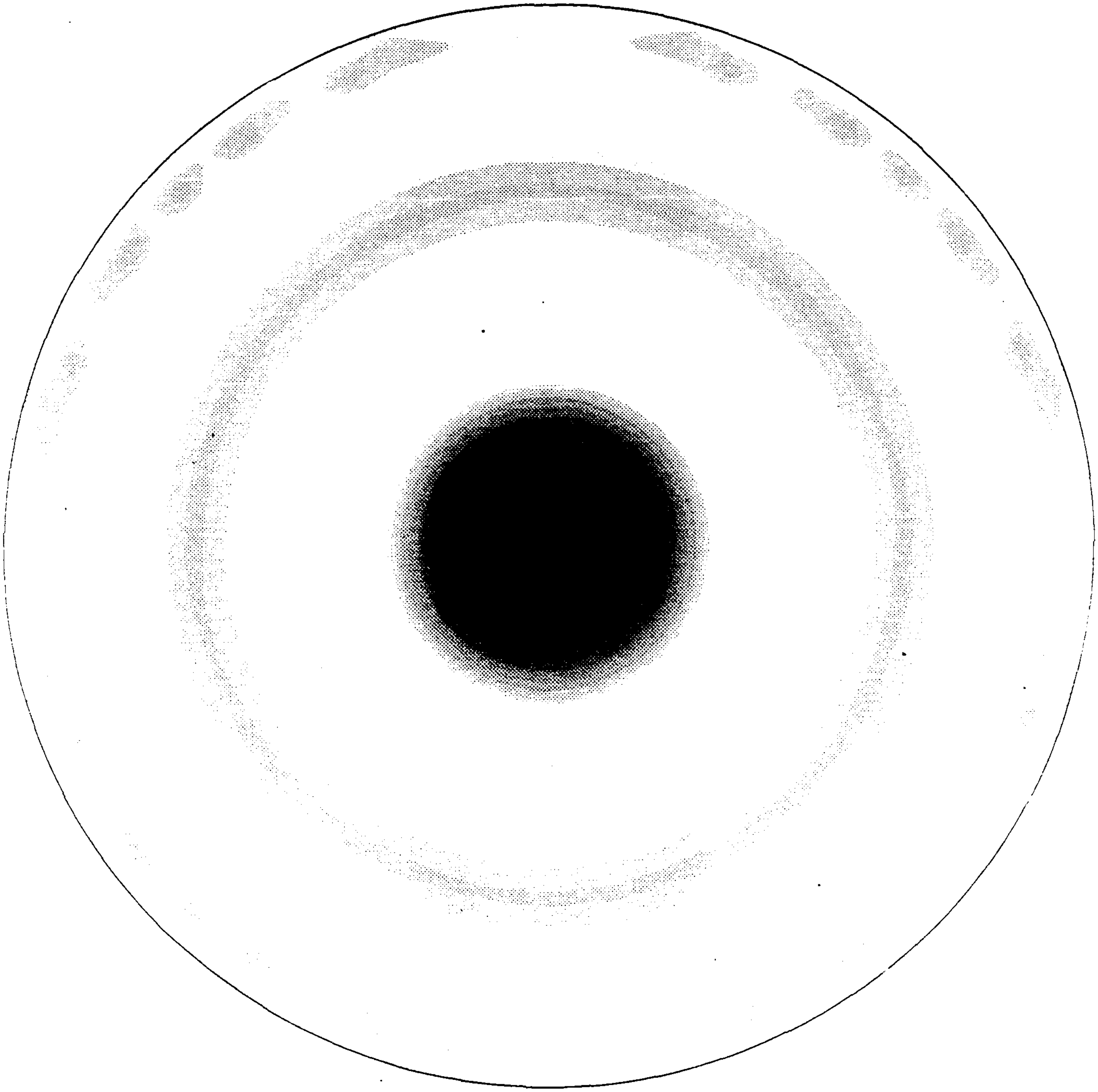


Fig 6

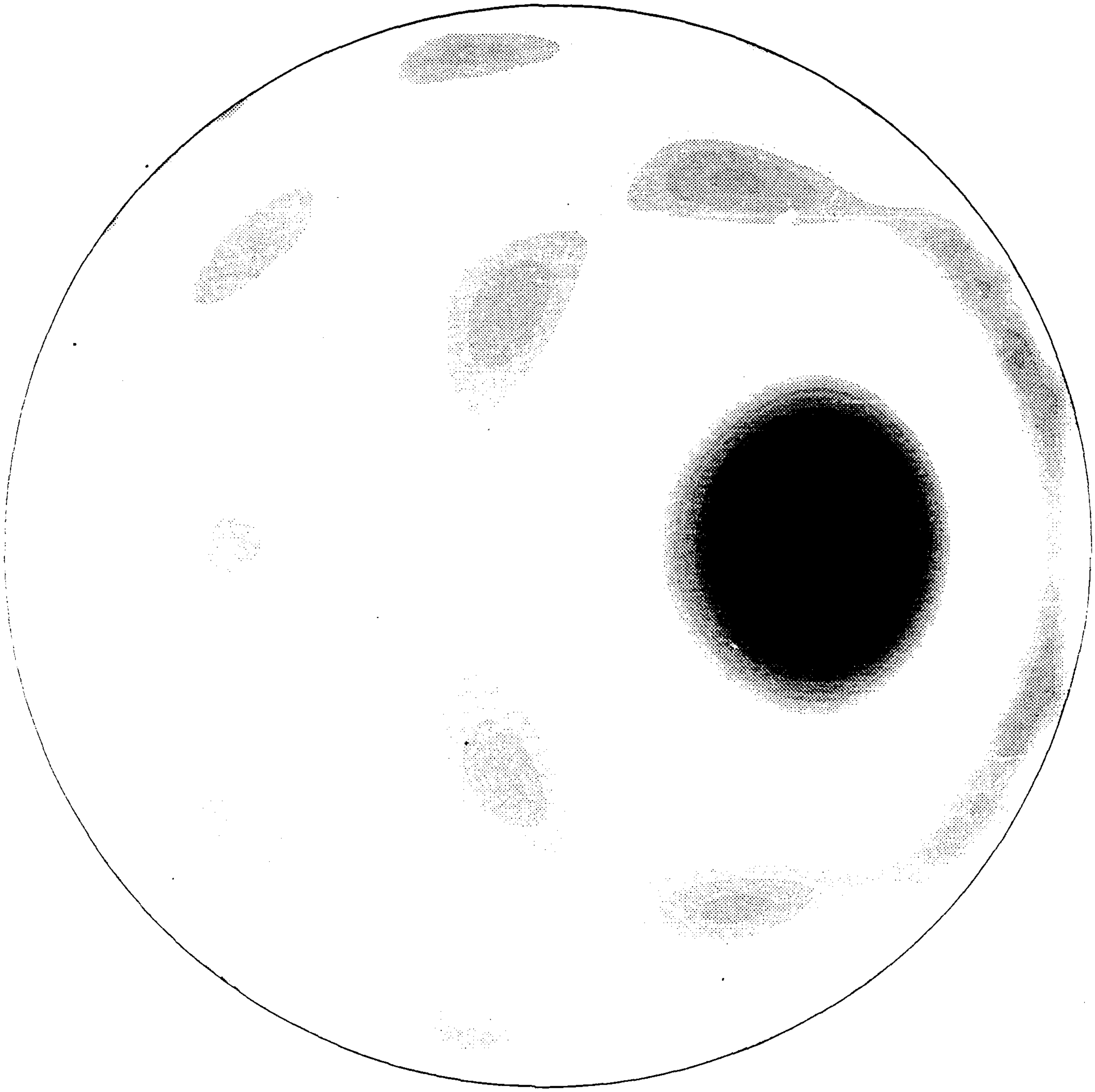


Fig 7

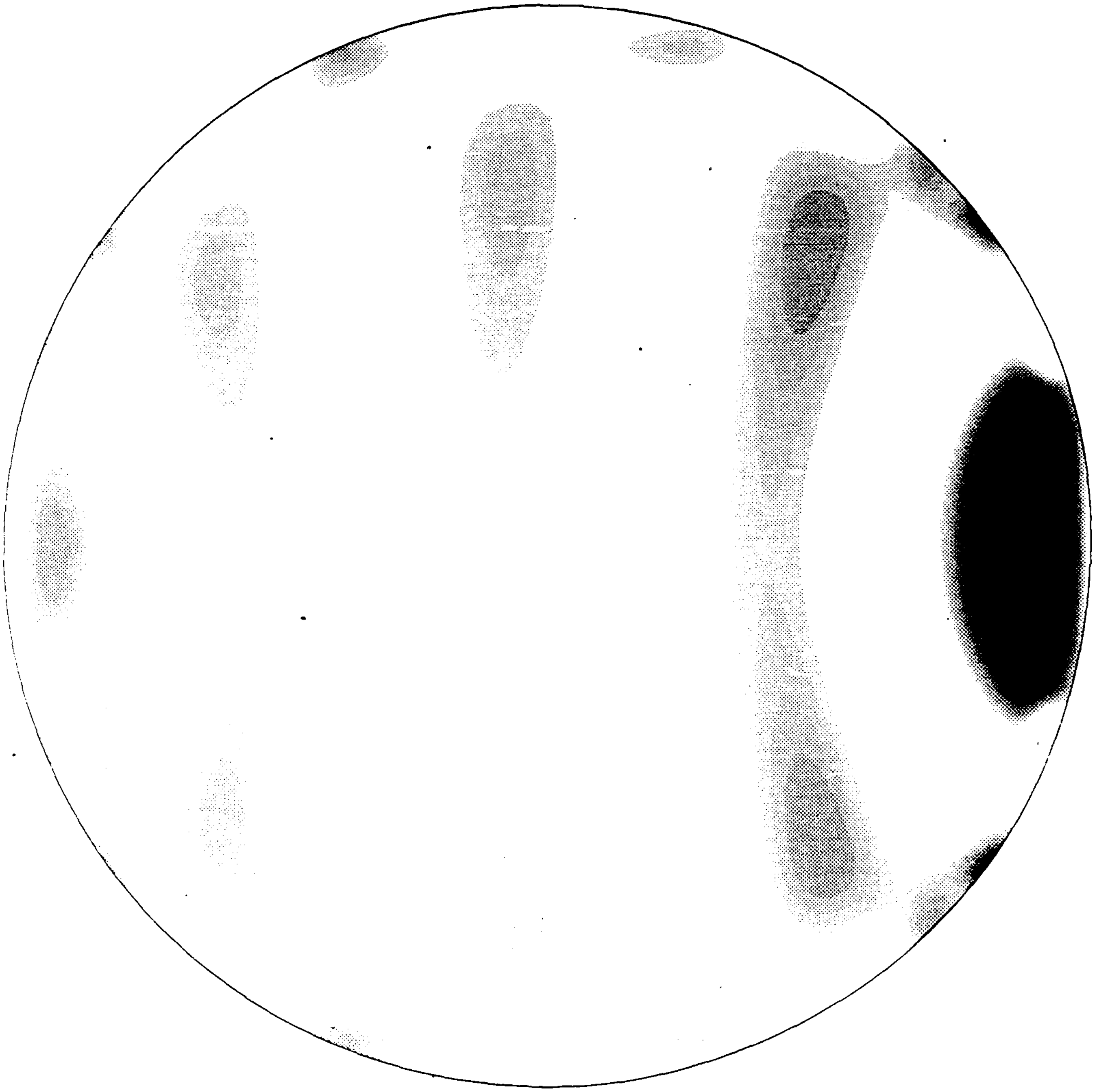


Fig 8

Appendix 14

Aggregation and refinement in binary image restoration.

by

M. Jubb and C. Jennison

1. Introduction

Recent developments in statistical image restoration use a Bayesian approach. One observes a degraded version of a true scene after the addition of noise and, possibly, blurring. If the degradation process and noise distribution are known, the likelihood of the record can be combined with a prior probability model to produce a posterior distribution for the true scene. A common approach is then to seek the *maximum a posteriori* (MAP) estimate of the scene and present this as the restored image.

For computational purposes it is extremely convenient to work with Markov random field (MRF) models. Under a MRF model the scene is divided into pixels, each of which can take a single colour or grey level, a neighbourhood structure for the pixels is specified and the key property of the model is that the distribution of the colouring of any pixel is conditionally independent of all other pixels, given the colouring of its neighbours.

There are two main approaches to searching for the MAP estimate. Geman & Geman (1984) proposed the method of simulated annealing. They have shown this to be a versatile and effective method although the amount of computation involved is often high. Besag (1986) suggested a computationally simpler method which he refers to as the method of iterated conditional modes (ICM). This method will normally converge to a local rather than global maximum of the *a posteriori* likelihood; however, convergence is rapid and, given the approximate nature of the MRF model, failure to find the global maximum may not be a serious drawback.

Jennison (1986) and Jennison & Jubb (1987) have shown that the same form of MRF model can be used to obtain restorations of an image with detail at a finer level than the pixel grid on which records are observed. In their original examples the noise level was very low. The work reported in this paper grew out of an investigation into the use of "refinement" methods in the presence of greater noise: the main problem in this case is to find a good starting point for the refinement algorithm. In some of our exploratory examples we discovered that the ICM method itself experienced serious difficulties at very high noise levels. One solution to this problem is to increase the signal to noise ratio by aggregating the records of, say, each 2 by 2 block of pixels into a single record: satisfactory results were obtained by applying ICM to the aggregated signal and then using the resulting restoration as the starting point for ICM on the original pixel grid. A natural extension of this idea is a "cascade" algorithm, similar to that of Gidas (1989), which produces restorations on successively finer pixel grids, starting with a single large pixel and ending with the original grid. We have found that this approach provides a simple and efficient way of adapting the ICM method to very noisy data. It also solves the refinement problem, since the end product of this algorithm, or even a restoration based on aggregated data, will provide a good starting point for the refinement process.

Our intention in this paper it to follow the ICM approach as much as possible. There are several places where simulated annealing might be incorporated but it would require substantially more computing, and there is no guarantee that it would provide better results. The main advantage of simulated annealing is that it allows one to escape from a local maximum of the posterior likelihood by a process of trial and error, however, use of the cascade algorithm to choose a good starting point for the deterministic ICM algorithm may be just as effective. We do introduce a version of simulated annealing to implement the refinement method of Section 5. Although this provides a very convenient way of exploring a larger set of restorations, its impact on the final restored image for our example is slight.

Some comment on the role of the prior model for the true scene is called for. Gidas (1989) goes to great lengths to ensure that, in his cascade algorithm, the models at different pixel sizes are mutually consistent. We are not committed to a single model and will be happy as long as the final restoration is a good one. It should also be remembered that all that we require of the end product of one stage of the cascade algorithm is that it should provide a good starting point for the next. We do not assume that we have a global MAP estimate at any stage, nor do we try to make use of such a property.

We shall use a single illustrative example throughout the paper. In the original image the boundaries of objects are smooth in parts but irregular in other places and certain features are extremely difficult to restore given the level of noise in the data. Thus, the example shows both the power of the proposed method and its limitations.

2. Model and notation

We first consider a rectangular region partitioned into pixels labelled $1,2,\ldots,n$. Each pixel is coloured black or white and the colour of pixel i is denoted by x_i^* which takes the value 0 for white and 1 for black. The x_i^* are unobserved. It is assumed that the conditional density function $f(y_i|x_i^*)$ is known and for the remainder of this paper we shall assume that the records y_i are independently distributed as Gaussian with mean x_i^* and variance σ^2. The set of records is denoted by $y = \{y_i;\ i=1,\ldots,n\}$. A colouring of pixel i (not necessarily the true colouring, x_i^*) is denoted by x_i and a specific colouring of the whole region is denoted by $x = \{x_i;\ i=1,\ldots,n\}$.

In the MRF model for the true scene we shall use a neighbourhood system in which pixels are considered to be first order neighbours if they are horizontally or vertically adjacent to each other and second order neighbours if they are diagonally adjacent. In our model, the prior distribution for the true scene, $p(x)$, is

$$p(x) \propto \exp[-\{\beta_1 Z_1(x)+\beta_2 Z_2(x)\}], \tag{2.1}$$

where $Z_1(x)$ is the number of discrepant first order pairs in the scene x, i.e. the number of pairs of first order neighbours which are of opposite colour, $Z_2(x)$ is the number of discrepant second order pairs and β_1 and β_2 are fixed positive constants.

The MAP estimate of the true scene is the value of x which maximises $P(x|y)$, the conditional probability of x given the record y. By Bayes' theorem

$$P(x|y) \propto l(y|x)p(x), \tag{2.2}$$

where $l(y|x)$ is the conditional likelihood of the observed record y, given the true colouring, x, and $p(x)$ is the prior probability of x. Thus, the maximisation of $P(x|y)$ corresponds to the minimisation of

$$\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-x_i)^2 + \left\{\beta_1 Z_1(x)+\beta_2 Z_2(x)\right\}, \tag{2.3}$$

over values of $x = \{x_i; i=1,\ldots,n\}$.

Besag's (1986) method of iterated conditional modes updates each pixel in turn, choosing for it the most likely colour based on its record and the current colouring of its neighbours, i.e., minimising (2.3) with respect to x_i with all the other pixel colourings fixed. The expression in (2.3) must decrease or remain constant at each updating but convergence will usually be to a local minimum. We shall see later in this paper that the choice of the initial colouring can have a great influence on the accuracy of the final restoration. Throughout this paper, when ICM is applied, a second order neighbourhood system will be used with $\beta_2=\beta_1/\sqrt{2}$; this ratio of β_1 to β_2 minimises the rotational variance of the second term of (2.3) with respect to the positioning of the pixel grid on a given scene (see Brown, Jennison and Silverman, 1987).

In the above model for the true scene it is assumed that each pixel is coloured wholly black or white. This is at best an approximation: more generally, one might expect pixels on the boundary of an object to contain areas of each colour, in which case the record y_i will be distributed as Gaussian with variance σ^2 and mean equal to the proportion of pixel i coloured black. Although we shall consider problems in which there is a general true scene, we start by considering restorations based on a discrete MRF model in which each pixel has a single colour. The refinement method described in Section 5 does, however, allow boundary pixels to be coloured partly black and partly white.

3. An example

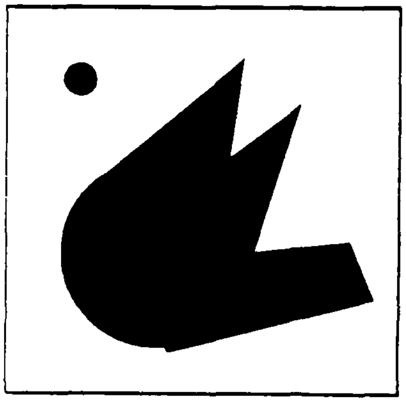


Figure 1. The true scene.

An example of a binary scene containing two separate objects is shown in Figure 1. A 256 by 256 pixel grid was superimposed on this scene and the proportion, p_i, of black in pixel i was calculated for each pixel. The record y_i was obtained by adding Gaussian noise with variance 4 to this proportion, p_i . Figure 2 shows the closest mean classifier for this record, in which a pixel is coloured black if its record is greater than 0.5 and white otherwise. One would not normally hope to recover an image which has been exposed to such a large amount of noise and Figure 3 shows the rather unsatisfactory restoration obtained by applying ICM with $\beta_1=4$. The value $\beta_1=4$ is unusually high but we found this to give the best results. (Note that even if $\beta_1\rightarrow\infty$ certain configurations of pixels remain unsmoothed.)

The major problem in our example is the low signal to noise ratio. This ratio may be improved by aggregating the record, i.e., by replacing sets of 2 by 2 pixels by a single large pixel with record equal to the average of the original four. This also

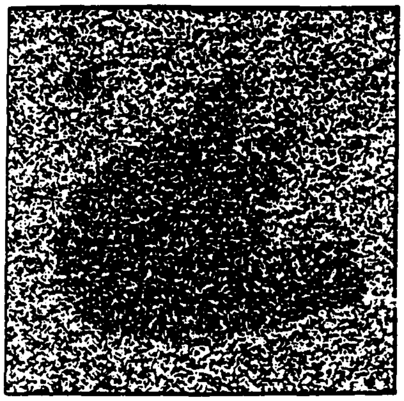


Figure 2.

Figure 3.

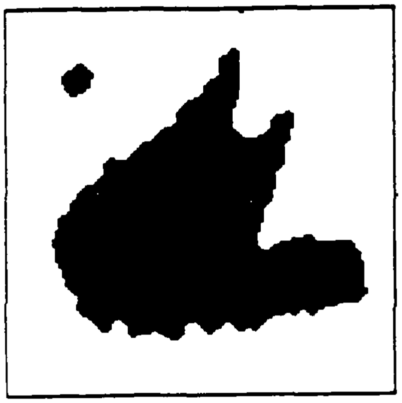


Figure 4.

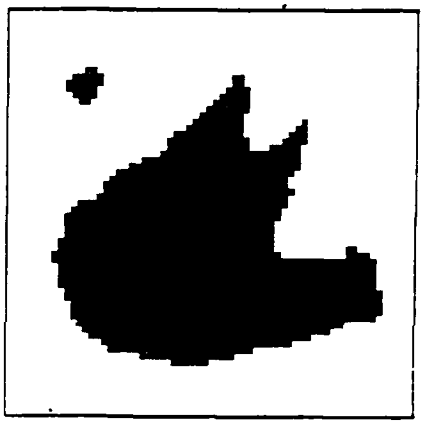


Figure 5.

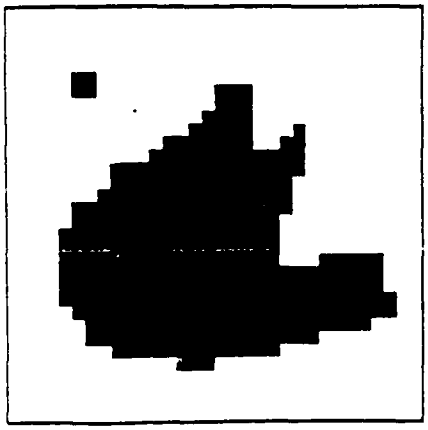


Figure 6.

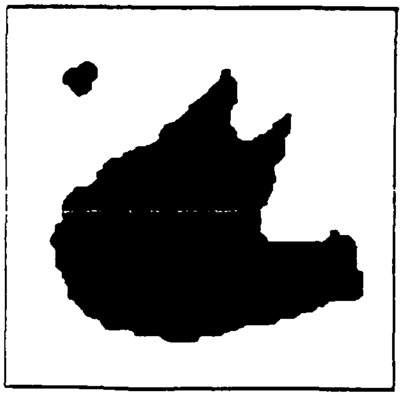


Figure 7.

corresponds to viewing the original image on a coarser grid. The variance of the new record is one quarter that of the original but the range of the p_i's is still $[0,1]$; thus there is a substantial increase in the signal to noise ratio. The restoration shown in Figure 4 was obtained by applying ICM to the aggregated record; the prior model for the true scene had the same form as (1.1) but was applied to larger pixels, the value $\beta_1 = 4$ was also used here as it was found to give the best results. The clear superiority of this restoration to that shown in Figure 3 demonstrates the advantage of working with the aggregated record. One explanation of the success of this restoration process is that it allows the ICM algorithm to look further afield when gathering neighbour information: ICM on the original pixel grid can easily be trapped in a local maximum of the *a posteriori* likelihood when only one pixel is allowed to change at a time.

Repeating the aggregation process gives the restorations shown in Figures 5 and 6, which are the restorations at two and three levels of aggregation respectively. These restorations were obtained using $\beta_1 = 1$, a more typical value, which we have found gives good results in cases where the signal to noise ratio is moderate. Note that the computational time and storage requirements for the processing of a 32 by 32 image are approximately $\frac{1}{64}$ times those needed to process a 256 by 256 image.

So far, we have followed Besag's method and used the closest mean classifier as our initial colouring for the 256 by 256 case and this is partly responsible for the poor quality of the restoration in Figure 3. A better initial colouring might be the final restoration obtained from an aggregated record. Figure 7 shows the result of using Figure 5 as the initial colouring for ICM on the 256 by 256 grid with $\beta_1 = 4$; a similar result is obtained with $\beta_1 = 1$. The superiority of this restoration to that of Figure 3 demonstrates the influence of the initial colouring on the resulting image.

The method of simulated annealing is less dependent on the initial colouring, since it can progress from one local minimum of (2.3) to another whilst passing through higher intermediate values. Thus, simulated annealing is able to search at least a little further afield than the myopic ICM strategy. An advantage of using an aggregation procedure is that it allows the ICM approach to use more distant neighbour information whilst maintaining its computational speed.

4. The Cascade Algorithm

In the previous section we introduced the idea of using the restoration obtained from an aggregated record as the initial colouring for restoration on a finer scale. We now extend this idea to define a "cascade" algorithm in which restorations obtained from 2^m by 2^m grids are used as the initial colourings for restorations on 2^{m+1} by 2^{m+1} grids. A single pixel restoration is obtained by aggregating the record until it is one pixel in size: this is then used as the initial colouring for the ICM method on the 2 by 2 grid. This restoration is in turn used as the initial colouring for ICM on the 4 by 4 grid and we continue in this way, obtaining restorations right up to the level of the original record. The last six in the series of restorations for our example are shown in Figures 8-13: the value $\beta_1 = 1$ was used at each level, though it is interesting to note that using higher values at the 128 and 256 levels made virtually no difference to the image obtained.

The method of Gidas (1989) is very similar to the procedure we have just described. However, Gidas uses a single MRF model defined on the finest pixel grid and employs the "renormalization group" approach to compute the models implied for coarser grids. Both the complexity of the models at the aggregated levels and the use

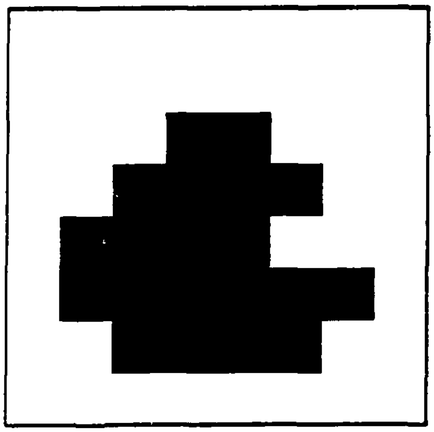


Figure 8.

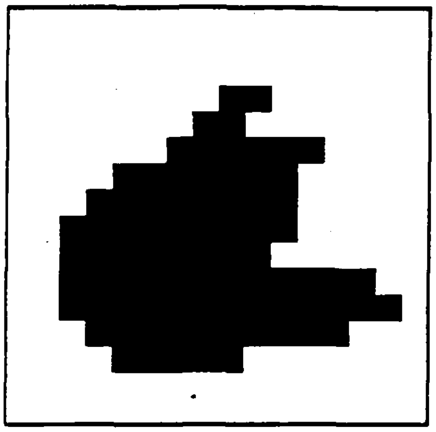


Figure 9.

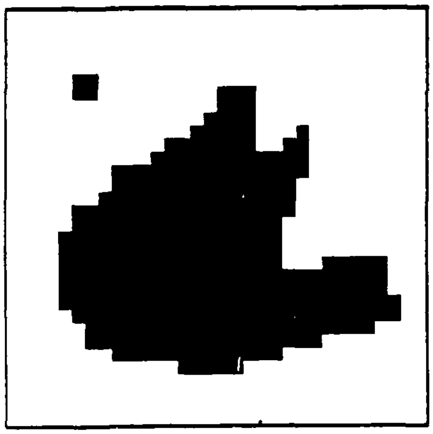


Figure 10.

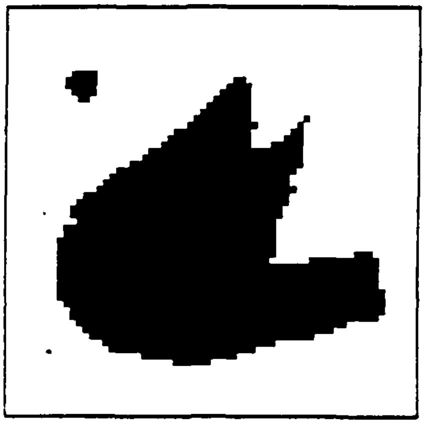


Figure 11.

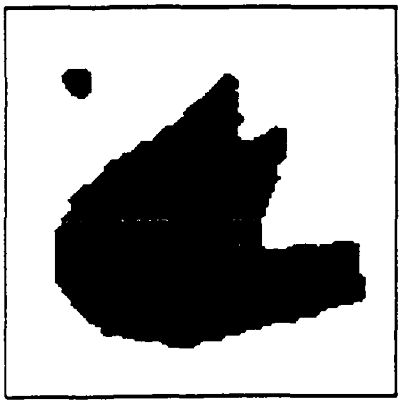


Figure 12.

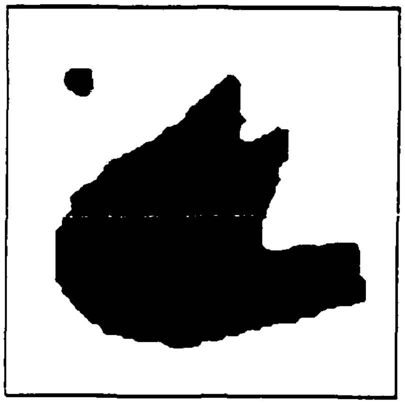


Figure 13.

of simulated annealing at each stage makes this a computationally demanding method. We have tried to keep computation to a minimum at the expense of a less rigorous treatment of the prior model: given the approximate nature of this model, we would argue that this is not unreasonable.

One might at least try to develop theoretical arguments to produce a "correct" sequence of values of β_1 for use at different stages of the cascade algorithm. Brown, Jennison and Silverman (1987) interpret the second term of (2.3) as a penalty and suggest that it should be chosen to be approximately independent of the pixel grid superimposed. They suggest that this penalty should approximate a constant multiple of the total boundary length in the image. In our application this would imply that the parameter β_1 be halved as the pixel sizes are quartered but we have not found this to be very successful in practice. Using the same value of β_1 at each stage produced substantially better results.

When processing the larger images we avoid unnecessary computations by storing the coordinates of pixels whose colourings have changed in the current iteration. If the number of these is small, only pixels whose neighbours have changed colour in the last iteration are considered for updating in the next iteration. For each of the images shown in Figures 8-13 one complete iteration plus some minor changes was all that was required. Summing a geometric series, we see that the total computation required is approximately equivalent to $1\frac{1}{3}$ iterations of ICM on the finest pixel grid.

We have seen that the restorations obtained on the finer grids have been insensitive to the choice of β_1. This is partly attributable to the high noise level (updating is essentially by the "majority vote rule" at quite low values of β_1) but also suggests that, for a given image, restoration at too fine a pixel level is unnecessary, adding only computation and superfluous detail to what is already a satisfactory restoration. We are able to make a direct comparison of restorations obtained at different levels of aggregation by superimposing the finer grid on the coarser image and calculating penalties for both, based on the finer record and the MRF model at that level. The coarser image is disadvantaged, since it was chosen when searching for the minimum of a different penalty. We measure the benefit of restoring at the finer level by the percentage decrease in the penalty. The values are tabulated below.

Grid size of coarse restoration	Grid size of fine restoration	percentage reduction in penalty
2 × 2	4 × 4	68.1
4 × 4	8 × 8	75.8
8 × 8	16 × 16	49.6
16 × 16	32 × 32	21.5
32 × 32	64 × 64	5.2
64 × 64	128 × 128	1.6
128 × 128	256 × 256	0.6

Analysis of these values is purely subjective but appears to suggest that the 64 by 64 level is satisfactory. Inspection of Figures 8-13 also leads to the same conclusions.

5. Subpixel refinement.

So far the restoration techniques we have used have coloured each pixel wholly one colour, even though pixels on the edges of objects in the true scene may be partly black and partly white. We now consider techniques which allow both colours to appear in a single pixel. Jennison (1986) used a modification of the ICM method to obtain a restoration in which each pixel was divided into 4 subpixel quarters and a separate colour allocated to each subpixel. His method used the ICM restoration at full pixel size as a starting point for restoration at the subpixel level. The success of this technique prompted Jennison and Jubb (1987) to consider the further refinement of pixels.

Since the number of different colourings of a pixel grows exponentially with the number of subpixels, the extension of Jennison's method to a finer subdivision of each pixel is computationally prohibitive. However, the limit of this process, in which an arbitrary colouring of each pixel is allowed, can be made tractable. Rather than specify a MRF model for the true scene we interpret the minimisation of (2.3) as a form of penalised maximum likelihood. The second term of (2.3) is, approximately, a multiple of the total boundary length in the image, x. Thus, an analogous penalty for a general restoration, x, is

$$\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - p_i(x))^2 + \beta L(x), \tag{5.1}$$

where $p_i(x)$ denotes the proportion of black in pixel i, $L(x)$ is the total edge length in scene x and β is a fixed constant. For computational simplicity we restrict attention to restorations in which pixels are either of a single colour or are separated into areas of different colour by a single straight line with the line segments defining such areas in adjacent pixels meeting at a point.

A black and white image can be regarded as a series of line segments separating the two colours. Jennison and Jubb (1987) use the restoration obtained from Jennison's quarter pixel method is used an initial representation for the line segments. The updating process treats pixels in pairs, selecting the best place for two edges to meet, given the current restoration of neighbouring pixels. We repeat the details for completeness.

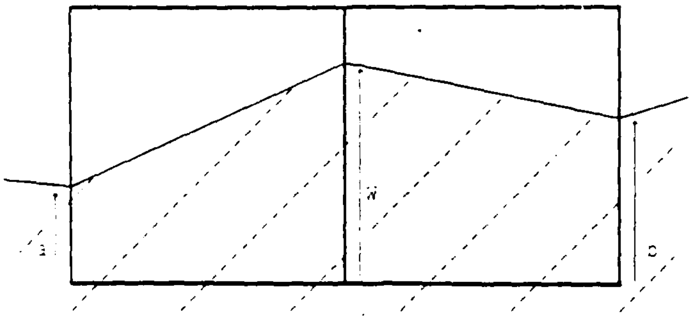


Figure 14. Updating the position of edges in pixels i and j.

As an example, consider the configuration at pixels i and j shown in Figure 14. The distances a and b are determined by the current colouring of neighbouring pixels and treated as constant for the moment. The distance W is chosen to minimise the contribution from pixels i and j to the total penalty (5.1), i.e.

$$g(W) = \frac{1}{2\sigma^2} \sum_{k=i,j} (y_k - p_{kW})^2 + \beta(e_{iW} + e_{jW}), \tag{5.2}$$

where e_{kW} is the length of edge in pixel k when the join is at W and p_{kW} is the proportion of black in pixel k when the join is at W.

For the case shown in Figure 14, this penalty is

$$g_1(W) = \frac{1}{2\sigma^2}\{(y_i - a - \tfrac{1}{2}(W-a))^2 + (y_j - b - \tfrac{1}{2}(W-b))^2\} + \beta\{\sqrt{1+(W-a)^2} + \sqrt{1+(W-b)^2}\}.$$

This can not be minimised directly but the form of

$$\frac{dg_1(W)}{dW} = \frac{1}{4\sigma^2}(2W+a-2y_i+b-2y_j) + \beta\left[\frac{(W-a)}{\sqrt{1+(W-a)^2}} + \frac{(W-b)}{\sqrt{1+(W-b)^2}}\right]$$

suggests an iterative approach. Given an approximate solution W_{s-1} we solve

$$\frac{1}{4\sigma^2}(2W_s+a-2y_i+b-2y_j) + \beta\left[\frac{(W_s-a)}{\sqrt{1+(W_{s-1}-a)^2}} + \frac{(W_s-b)}{\sqrt{1+(W_{s-1}-b)^2}}\right] = 0$$

to obtain

$$W_s = \frac{4\sigma^2\beta\left[\frac{a}{\sqrt{1+(W_{s-1}-a)^2}} + \frac{b}{\sqrt{1+(W_{s-1}-b)^2}}\right] + (2y_i-a+2y_j-b)}{2+4\sigma^2\beta\left[\frac{1}{\sqrt{1+(W_{s-1}-a)^2}} + \frac{1}{\sqrt{1+(W_{s-1}-b)^2}}\right]}.$$

Starting from any sensible initial value, W_0, accuracy to 3 decimal places was achieved after at most four iterations. In practice we take W_0 to be the value of W prior to this update.

Different forms of (5.2) are possible depending on which neighbours of pixels i and j contain both colours. There are only four distinct cases that may arise and these are shown in Figure 15.

We have shown the method of solution for case (i); cases (ii) - (iv) are solved in a similar way. All other cases can be reduced to one of the above by means of exchanging and/or inverting the pixels and their colours. The edge pixels are updated in turn, following an edge around, completing circuits of the edge until convergence.

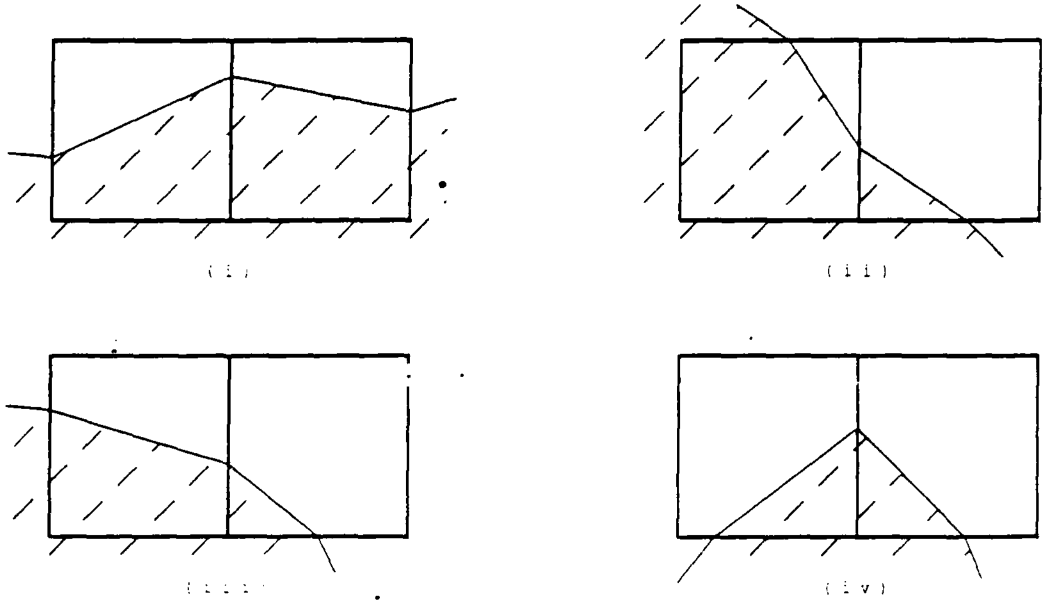


Fig. 15. Possible configurations of edges in two neighbouring pixels.

The complete restoration algorithm

We can now combine both aggregation and refinement into a three stage algorithm:

Stage 1:	Apply the cascade algorithm using ICM on the aggregated records up to a suitable point. The record is now fixed at this level and no further use will be made of the original record. (If the record is still aggregated at this level substantial savings in computation will result.) avoiding unnecessary computation.
Stage 2:	Iterate Jennison's quarter pixel refinement to convergence. This is very quick and supplies a good starting point for the line fitting process.
Stage 3:	Apply the line fitting algorithm to convergence.

A development in the line fitting algorithm

In the line fitting algorithm described by Jennison and Jubb (1987) the route that the lines take through pixel edges is determined once and for all by the restoration obtained at the quarter pixel level.

We have now extended the algorithm to allow changes in this route. Each time the point at which the edge crosses a pixel boundary is updated an alternative route is compared. A number of cases have to be treated separately: three qualitatively different configurations are shown in Figure 16.

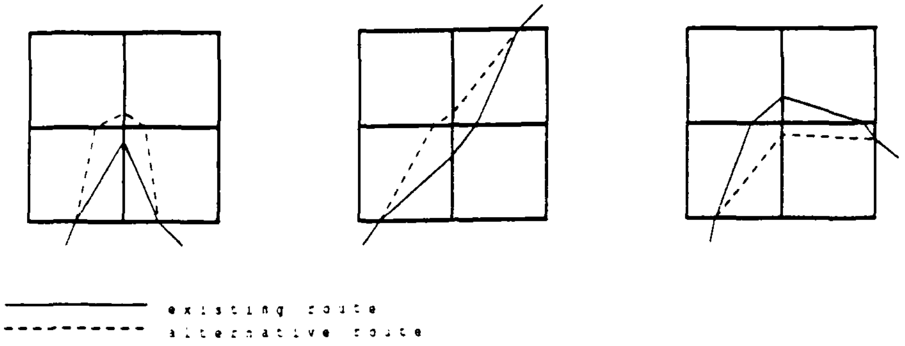


Fig. 16. Examples of configurations at which alternative routes are considered.

The contribution to the total penalty from all four pixels is calculated for each of the two routes with line edges chosen optimally for that route. In the basic method, the route which has smallest penalty is then chosen.

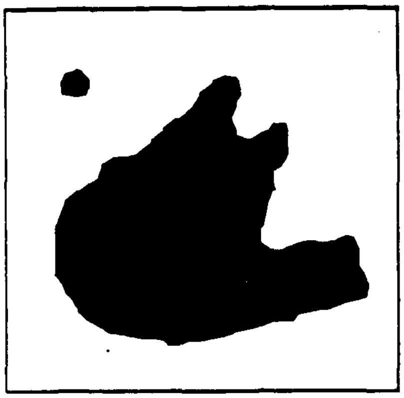


Figure 17.

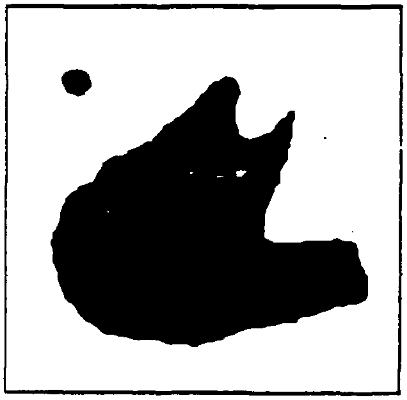


Figure 18.

Figures 17 and 18 show the restorations obtained from applying the line fitting method to the aggregated record in the example. In Figure 17 the grid size is 32 by 32 and in Figure 18 it is 64 by 64. In the previous section we suggested that a grid size of 64 by 64 would be sufficient and the restoration shown in Figure 18 is indeed satisfactory. In both cases we used $\beta_1=1$ at the ICM and quarter pixel levels of restoration and $\beta=4$ for the line fitting.

The updating process in the above line fitting procedure has the general characteristics of an ICM method: the penalty (5.1) is minimised with respect to one component of the boundary whilst everything else is held fixed. This method will generally yield a local minimum of (5.1) and it is possible that the final restoration could be improved further by making a number of route changes simultaneously. For example, the penalty (5.1) might be reduced by moving a long vertical edge one pixel to the left whereas it would increase initially if only one route change were made at a time.

To allow further exploration of alternative routes we have implemented a form of simulated annealing. This method retains the property that for a given route the point on a pixel edge at which two line segments meet is chosen optimally. However, when comparing the minimum penalties for *different* routes we allow the route with the larger penalty to be chosen with non-zero probability. Suppose two routes, A and B, have minimum penalties pen_A and pen_B, then, when the annealing process is at temperature T we select route A and its optimal edges with probability

$$\frac{e^{(-pen_A/T)}}{e^{(-pen_A/T)}+e^{(-pen_B/T)}}$$

otherwise we choose route B. Of course, only the contribution to the total penalty from the four pixels concerned need actually be calculated.

By restricting the random choice to the route alone, we ensure that, effectively, the annealing process is applied to a fairly low dimension problem, the number of variables being of the order of the number of boundary pixels. Theorem B of Geman and Geman (1984) demonstrates the convergence of their simulated annealing method. In its stated form, this theorem does not apply to our hybrid procedure whose iterative steps combine a random choice of route with a deterministic choice of edges given that route and currently fixed end points. Perhaps a sufficiently general result could be proved but this would, presumably, still only apply for gentle cooling schedules. However, we prefer to think of the annealing method simply as a convenient numerical procedure which searches a little further afield than the ICM approach.

We have experimented with a variety of cooling schedules for our example using the aggregated record at both the 32 by 32 and 64 by 64 grid levels. The best results were obtained using a cooling schedule in which T decreased logarithmically from 3.5 to 0.5 over several hundred sweeps and linearly from 0.5 to zero over several hundred more. We then continued to update using T=0 until convergence, which usually required only a few sweeps. Although simulated annealing often produced a lower penalty, the restoration produced was never visually superior to that obtained using the local maximisation procedure.

Our conclusion is that the starting point provided by the cascade algorithm was sufficiently good that the deterministic line fitting algorithm was very nearly optimal.

6. Concluding Remarks.

Combining the line fitting procedure with the cascade algorithm has produced a fast and effective method for obtaining a high quality restoration from noisy data. Further work is required to provide an automatic choice of suitable values of β_1 at different grid levels and a criterion for terminating the cascade algorithm at the most appropriate level of aggregation. Although we have considered only two-colour images in this paper, it is clear that the basic ideas are more generally applicable: we hope to continue work on the development of an aggregation and refinement algorithm for grey level images.

References

Besag, J. E. (1986) On the statistical analysis of dirty pictures. *J. Royal Statist. Soc. B*, 148, 259-302.

Brown, T. C., Jennison, C. and Silverman, B. W. (1987) Edge process models for regular and irregular pixels. Submitted for publication.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, **6**, 721-741.

Gidas, B. (1989) A renormalization approach to image processing problems. *IEEE Trans. Pattern Anal. Machine Intell., to appear.*

Jennison, C. (1986) Contribution to discussion of Besag (1986) *J. Royal Statist. Soc. B*, **148**, 288-289.

Jennison, C. and Jubb, M. (1987) Statistical image restoration and refinement. *Proc. XII Inf. Proc. in Med. Imaging Conf., to appear.*