10/9-5-90 88①

LBL-26633
UC-408

Lawrence Berkeley Laboratory

UNIVERSITY OF CALIFORNIA

Engineering Division

Bayesian Image Reconstruction. Application to Emission Tomography

J. Nunez and J. Llacer

February 1989

DO NOT MICROFILM
COVER

Prepared for the U.S. Department of Energy under Contract Number DE-AC03-76SF00098.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. Neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial products process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California and shall not be used for advertising or product endorsement purposes.

Available to DOE and DOE Contractors from the
Office of Scientific and Technical Information
P.O. Box 62, Oak Ridge, TN 37831
Prices available from (615) 576-8401, FTS 626-8401

Available to the public from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road, Springfield, VA 22161
Price: Printed Copy A04, Microfiche A01

Lawrence Berkeley Laboratory is an equal opportunity employer.

LBL-26633

LBL--26633

DE90 016621

Bayesian Image Reconstruction.
Application to Emission Tomography

J. Nunez and J. Llacer

Engineering Division
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

This work was supported, in part, by a grant from the National Institutes of Health (CA-39501), in part by the Catalan Studies Program (University of California and the Generalitat de Catalunya), and by the Director, Office of Energy Research, Office of Health and Environmental Research, Physical and Technological Division of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

BAYESIAN IMAGE RECONSTRUCTION. APPLICATION TO EMISSION TOMOGRAPHY

Jorge Nunez* and Jorge Llacer
Engineering Division
Lawrence Berkeley Laboratory
1 Cyclotron Road
Berkeley, CA 94720

ABSTRACT

In this paper we propose a Maximum a Posteriori (MAP) method of image reconstruction in the Bayesian framework for the Poisson noise case. We use entropy to define the prior probability and likelihood to define the conditional probability. The method uses sharpness parameters which can be theoretically computed or adjusted, allowing us to obtain MAP reconstructions without the problem of the "grey" reconstructions associated with the pure Bayesian reconstructions. We have developed several ways to solve the reconstruction problem and propose a new iterative algorithm which is stable, maintains positivity and converges to feasible images faster than the MLE method. We have successfully applied the new method to the case of Emission Tomography, both with simulated and real data.

I. INTRODUCTION

Maximum entropy and full Bayesian solutions are being increasingly used for reconstructing images from noisy and incomplete data in fields as diverse as astronomy (Frieden, 1972; Frieden and Wells, 1978; Skilling and Bryan; 1984, Narayan and Nytyananda, 1987), molecular biology (Bryan et al 1983), economics (Akaike, 1986), medical tomography (Minerbo, 1979; Levitan and Herman, 1987; Liang and Hart, 1988) and also in seismology, spectroscopy, information theory and communications. A review of the methods is given in Jaynes (1982, 1984a).

*On leave from the Dept. of Physics, University of Barcelona, Spain.

In this paper describe the rationale for a new algorithm based on Bayesian considerations for computing Maximum a Posteriori (MAP) images. We present the application of the method to the problem of tomographic image reconstruction. We also present a comparison of the method with the Maximum Likelihood Estimator Method (MLE) introduced by Shepp and Vardi (1982).

II. THE BAYESIAN STRATEGY

As stated by Frieden (1983a), a mathematical science ordinarily admits several possible routes to a solution. This is true also in statistics (on which image reconstruction theory is based), with the particularity that there are two fundamentally different approaches to solving certain statistical problems. They are the Bayesian and the classical approaches. Both give correct answers within their frameworks.

Given a certain set of data generated by means of an unknown probability law, the classical view is that nothing should be assumed about that law. Any analysis must be based exclusively upon the given data (data come first). There is nothing subjective and, if a solution can be found in such a way, nobody doubts that it is the correct one. The problem arises when the formulation of the solution requires the use of the unknown probability law. In that case there is no classical answer and the problem is often changed to a less ambitious but soluble one. By contrast the bayesian strategy permits some subjective estimate of the unknown probability law. Any "reasonable" estimate of the probability law, called hereafter "prior probability" is permitted. The maximum entropy estimates given in the above cited bibliography and the stopping criteria developed for the MLE algorithm (Veklerov and Llacer 1987) are based in such prior knowledge, for example.

In fact, when the data are complete and of very good quality, the "a priori" probability is useless, but when we the data are noisy, sparse and/or incomplete, the "a priori" knowledge can carry as much weight as the data.

Bayes' rule was first given by Thomas Bayes in a post-mortem

publication in 1763 and was rediscovered by Laplace in 1774, who used it for more than 40 years. If we denote $P(\text{hypotesis}|\text{data})$ the conditional probability that the hypotesis be true given the data, Bayes' theorem gives the desired $P(\text{hypotesis}|\text{data})$ from the usually computable $P(\text{data}|\text{hypotesis})$, along with the probability of the hypotesis $P(\text{hypotesis})$ and the normalization constant $P(\text{data})$. The $P(\text{hypotesis})$ is called "prior probability" because it is known (in some form) prior to obtaining the data and $P(\text{hypotesis}|\text{data})$ is called the "a posteriori" probability because is what we hope to obtain after considering all facts. Bayes' theorem is:

$$P(\text{hypotesis}|\text{data}) = \frac{P(\text{data}|\text{hypotesis}) P(\text{hypotesis})}{P(\text{data})}$$

In the image reconstruction problem, the object or its image, the projection of the object and the noise are assumed connected by a linear imaging equation:

$$f a + n = p$$

where a represents the object to be computed, p is the given data about the projection of the object, n is the noise inherent to the process of emission, detection or both and f is the point spread function assumed to be known. If we consider the object discretized in B pixels and we have obtained D data about the projections, the discrete version of the imaging equation is:

$$\sum_{i=1}^B f_{ji} a_i + n_j = p_j \quad j = 1, \dots, D \quad (2.1)$$

in which a_i and p_j are positive vectors ($a_i \geq 0 \ i=1, \dots, B$; $p_j \geq 0 \ j=1, \dots, D$) and f_{ji} are the elements of a $(D \times B)$ matrix. We use one-dimensional notation for simplicity. Throughout this paper, the subscript i characterizes the object space (intensities in the object or in its reconstructions and columns in the point spread function matrix f_{ji}) and the subscript j characterizes the projection space (data, projection of the object or of its reconstructions and rows of the point spread function matrix f_{ji}). Sometimes the projection space is called image space but we will use the term "image" in the usual sense to describe any representation of the object or of its reconstructions. We assume also the following normalization condition:

$$\sum_{j=1}^D f_{ji} = 1 \quad i = 1, \dots, B \quad (2.2)$$

The application of Bayes' theorem to the image reconstruction problem gives:

$$P(a|p) = P(p|a) P(a) / p(p)$$

where $P(a|p)$ is the "a posteriori" probability of an image a given the data p , $P(p|a)$ is the conditional probability of obtaining the measurements p for a given image a , $p(a)$ is the "a priori" distribution of the object and the denominator $P(p)$ is independent of a and serves merely to insure the correct normalization.

Given the data p the most probable image is obtained by maximizing $P(a|p)$, or equivalently by maximizing the product $P(p|a) P(a)$. This maximization is called "maximum a posteriori" or, in short, MAP.

III. THE LIKELIHOOD

The conditional probability $P(p|a)$ describes the noise in the projections and its possible object dependence. It is fully specified in the statement of the problem as the Likelihood.

We assume that each data element obeys Poisson statistics and that all data elements are statistically independent. The Poisson distribution is widely used for the conditional probability in fields in which the fluctuations in photon arrivals are dominant over the noise statistics of the detectors. In other cases, different choices of the form of the likelihood function should be used, as in Gaussian uncorrelated noise (see, for example, Skilling 1984, Narayan and Nityananda 1987).

The Poisson distribution is the usual one in the application presented in this paper, Positron Emission Tomography (PET) and in medical optics in general, where the detectors count energetic quanta.

Following Frieden and Wells (1978), we discuss next the concepts of image and data increments. A Poisson projection (detection of an object) consists of a finite number of counts at each detector unit. This fact implies the existence of a smallest received intensity increment Δp_j needed to trigger one count in a detector j . Frieden and Wells use a constant value Δp independent of the detector, but we introduce the detector dependence because in tomography the intensity increment can depend on the detector. This is the case when detectors have different gains, for example.

Under the previous assumptions of Poisson noise, statistical independence and intensity increments, the conditional probability is:

$$P(p|a) = \prod_{j=1}^D \exp(-h_j) \frac{(h_j)^{p_j/\Delta p_j}}{(p_j/\Delta p_j)!} \quad (3.1)$$

where

$$h_j = \left(\sum_{i=1}^B f_{ji} a_i \right) / \Delta p_j \quad j = 1, \dots, D \quad (3.2)$$

If all the Δp_j are equal to unity, equation (3.1) becomes the classical likelihood equation. The Maximum Likelihood Estimator method (MLE) consists in the direct maximization of equation (3.1) with all $\Delta p_j=1$ (or other constant value) by some appropriate method. In Positron Emission Tomography (PET), maximum likelihood is obtained by the EM iterative algorithm (Shepp and Vardi, 1982) but the iterations must be stopped before an image deterioration effect sets in. This effect reflects the unsuitability of maximizing the conditional probability alone when solving the image reconstruction problem. There are two possible ways to avoid this effect, one of them is to consider a stopping criterium (Veklerov and Llacer, 1987) and the other is to consider a prior probability in the Bayesian framework.

IV. PRIOR PROBABILITY AND ENTROPY

In image reconstruction, the data alone, whether noisy or not, cannot point to any particular scene because, in words of Jaynes (1984b), "the domain R of the maximum likelihood, where $\chi^2 = 0$, is not a point but a manifold of high dimensionality, every point of which is in the feasible set R' (which we may think of as R enlarged by adding all points at which χ^2 is less than some specified value)". So an informative prior must be used in addition to Bayes' theorem to pick out one point in R' .

While the conditional probability $P(p|a)$ has been fully specified in the statement of the problem, the prior probability $P(a)$ is still unspecified. Several priors have been proposed, some of them depending on the specific problem to be solved. In particular, in the case of PET, we will mention here four attempts at defining $p(a)$, without claiming completeness. Miller and Snyder (1987) have used sieves as a method of regularizing the MLE iterative procedure, with resulting images that do not exhibit the noisy artifact and with adequate sharpness. Geman and McClure (1987) use a prior that requires a certain degree of smoothness in image subregions. Levitan and Herman (1987) have used a smoothed, non-negative filtered backprojection reconstructed image to require that the MLE reconstruction does not deviate excessively from it. Smith, Barret and Paxman (1983) use a method of simulated annealing to, in effect, define a certain prior that allows for a number of pixel counts that do not maximize the likelihood function.

On the other hand, the role of the entropy to define the prior probability function has been discussed for three decades (Jaynes, 1957; 1982). Frieden (1972) first used the Shannon form of entropy in the context of image reconstruction. Several papers (see Skilling and Gull 1984, Jaynes 1984b for details) have shown that maximizing the entropy as a prior is the only consistent method of selecting a single image from the many images which fit the data. In addition, Skilling and Gull (1988) have shown that, if one accepts three fundamental axioms, the prior $P(a)$ has to be a measure of the entropy of the image.

The practical interest of using the entropy as a prior (and maximizing it in the Bayes' framework) is that the resulting reconstruction has minimum configurational information, so there must be evidence in the data for any structure which is seen. Also the constraint of positivity of the solution is automatically invoked since the entropy is not defined for negative values.

In this paper we use entropy to define the prior probability $P(a)$ in the way proposed by Frieden (1972) and called the random-grain model by Andrews and Hunt (1977).

Suppose that N is the total energy (usually number of counts or photons) in the object. Suppose, in addition, that there is an elemental intensity increment Δa describing the finest known intensity jumps that are possible in the object. Now the values $N/\Delta a$ and $a_i/\Delta a$ $i=1, \dots, B$ are dimensionless numbers since Δa contains the unit of radiance. Assuming a complete lack of prior information regarding the statistical makeup of the object, we make the simple assumption that the occurrence of one object unit Δa in a pixel does not affect the possible location of any other object unit. Thus, the object units are statistically independent within each pixel and from pixel to pixel.

In this situation the number of ways that an object a can occur is (Jaynes, 1968):

$$W(a) = \frac{(N/\Delta a)!}{\prod_{i=1}^B (a_i/\Delta a)!} \quad (4.1)$$

We take the prior probability $P(a)$ of the object proportional to its multiplicity given by (4.1). Thus if K is a normalization constant, the prior probability is:

$$P(a) = K W(a) \quad (4.2)$$

V. THE MAXIMUM A POSTERIORI

The maximum a posteriori probability will be obtained, as indicated above, by maximizing the product $P(a|p) = p(a) P(p|a)$ or, equivalently, the natural logarithm of that probability. By Using Stirling's approximation, we obtain, from (4.1) and (4.2):

$$\log P(a) = \log(K) + (N/\Delta a) \log(N/\Delta a) - \sum_{i=1}^B (a_i/\Delta a) \log(a_i/\Delta a) \quad (5.1)$$

the two first terms of the sum being constants that do not affect the maximization. Note that the non-constant part of equation (5.1) is the entropy of the object . If we let $\Delta a = 1$,

$$\log P(a) = N H + \text{const.}$$

where

$$H = - \sum_{i=1}^B (a_i/N) \log(a_i/N) ,$$

the Shanon expresion of the entropy.

Similarly, taking the logarithm of (3.1) we obtain:

$$\log P(p|a) = \sum_{j=1}^D (-h_j + (p_j/\Delta p_j) \log(h_j) - (p_j/\Delta p_j, \log(p_j/\Delta p_j))) \quad (5.2)$$

with h_j defined by (3.2). Note that the last term of the sum is a constant since p_j are the given data and all the Δp_j are constant.

The maximum a posteriori is obtained by maximizing the product of equations (5.1) and (5.2) with the constraint that the total energy N is conserved. Throughout this paper we assume that the number of counts in the object and in its projection is the same). The problem is:

Maximize $P(a) P(P|a)$

with the constraint $\sum_{i=1}^B a_i = N$

We use now the Lagrange method to solve the constrained problem. Assigning a Lagrange multiplier μ to the constraint, using (5.1) and (5.2), and suppressing constants not affecting the maximization problem, we obtain the final formulation of the problem:

Maximize

$$BY = - \sum_{i=1}^B (a_i/\Delta a) \log(a_i/\Delta a) + \sum_{j=1}^D (-h_j + (p_j/\Delta p_j) \log(h_j)) - \mu (\sum_{i=1}^B a_i - N) \quad (5.3)$$

where

$$h_j = (\sum_{i=1}^B f_{ji} a_i) / \Delta p_j \quad j = 1, \dots, D$$

It is to be noted that, in the above maximization problem, the relative weight of the two terms of expression (5.3) is regulated by parameters Δa and Δp_j . We will comment further on the point below. Some important properties of the Bayesian solution and the differences between our work and some of the previously cited ones will also be discussed.

Since the work of Ables (1974) and Gull and Daniell (1978), several workers (see, for example, Narayan and Nytyananda (1987)) have used a method consisting in maximizing the Entropy with the constraint that sets a χ^2 term equal to its expected value for the number of measurements, so that the predictions of the model deviate from the data by the expected noise. For the case of Gaussian noise and using the Lagrange method for the constraint the problem is:

Maximize

$$BY = - \sum_{i=1}^B a_i \log(a_i) + \lambda (\sum_{j=1}^D (\sum_{i=1}^B f_{ji} a_i - p_j)^2 / \sigma_j^2 - D) - \mu (\sum_{i=1}^B a_i - N) \quad (5.4)$$

where σ_j are the standard error on datum j and λ and μ are Lagrange multipliers.

Applying the theory developed in the present work to the case of Gaussian noise, and disregarding the Δa and Δp_j parameters, we obtain the problem:

Maximize

$$BY = - \sum_{i=1}^B a_i \log(a_i) + \left(\sum_{j=1}^D \left(\sum_{i=1}^B f_{ji} a_i - p_j \right)^2 / 2\sigma_j^2 \right) - \mu \left(\sum_{i=1}^B a_i - N \right) \quad (5.5)$$

The problems represented by expressions (5.4) and (5.5) appear very similar. There is, however, a crucial difference, namely, that the relative weight of the two terms in the Bayesian case (5.5), corresponding to the Entropy and to the Likelihood, is not adjustable but instead depends on N . There is no guarantee that the residuals will have the right magnitude. It has been pointed out that when N becomes large, the Entropy term in (5.5) will peak so sharply at the grey reconstruction ($a_i = \text{constant}$) that the data is completely disregarded, leaving very large residuals. This difficulty has led maximum entropy users to use equation (5.4) (or similar ones) and to adjust the Lagrange multiplier λ so as to obtain the expected level of residuals.

As noted before, this problem is not present in our proposed Bayesian solution since parameters Δa and Δp_j introduce a weight system in equation (5.3). Parameters Δp_j are, in general, constants given implicitly as part of the data and parameter Δa is related to the finest known intensity jumps that are possible in the object. So the larger N is the larger is also Δa and the relative weight of the Entropy is smaller. Note that when Δa tends to infinity the Maximum a Posteriori solution tends to the Maximum Likelihood one.

In order to control the relative weight of the Entropy and the Likelihood in our MAP solution, we can choose between estimating the Δa and Δp_j parameters in some adequate manner or adjusting the Δa value to obtain the desired level of residuals in the full Bayesian solution. This point and the relation between Δa and sharpness of the reconstruction will be discussed below.

Frieden (1972) and Frieden and Wells (1978) introduced the Δ_a and Δ_p parameters. They make a similar statement of the reconstruction problem, although they do not invoke Bayes' theorem by name. They make a strong simplification by assuming that the object is dominated by a smooth background. They also use one Lagrangian multiplier for each of the data and solve the problem using the Newton-Raphson method. The use of a Newton-Raphson method is computationally prohibitive in real images since it is necessary to handle $N^2 \times N^2$ matrices, too big even for the most powerful computers.

It is interesting to note (since strong controversy is always present on this subject) that the Bayesian solution of the reconstruction problem is one of the examples of Bayesian parameter estimation. As pointed out very clearly by Jaynes (1984a), "in Bayesian parameter estimation, both the prior and posterior distribution represent, not any measurable property of the parameter, but only our own state of knowledge about it. The width of the distribution is not intended to indicate the range of variability of the true values of the parameter. It indicates the range of values that are consistent with our prior information and data, and which honesty therefore compels us to admit as possible values. What is 'distributed' is not the parameter but the probability. In other words, in the Bayesian scenario we are estimating, from our prior information and data, the unknown constant value that the parameter had when the data was taken".

This means, in the case of image reconstruction, that if the object was emitting photons (following a Poisson law, for example) with certain means (true value of the parameters), but during the real observation certain realization of those means was present, the Bayesian method estimates the value of the emissions that actually happened when the data was taken but not the means.

VI. THE MAXIMIZATION ALGORITHMS

Next, we want to solve the reconstruction problem described by expression (5.3). The nonlinear nature of the problem forces the algorithms to be iterative. Several iterative algorithms are possible

and we propose in this work three of them. A: an algorithm to maximize directly expression (5.3), B: an algorithm (two in fact) based on gradient techniques and C: a new iterative algorithm developed by us to avoid some of the problems presented by the other two.

The direct maximization algorithm (A) and the newly developed algorithm (C) are expressed as recursive formulas using a standard technique. Meinel (1986) has applied this technique to several linear and nonlinear restoration methods to obtain the corresponding recursive formulas. The technique is the following: given a serie of equations in the unknowns, we write them in the form:

$$a_i = K F(\{a_1\}) \quad i = 1, \dots, B \quad (6.1)$$

where F is some function, $\{a_1\}$ is the set of variables $a_1, \dots, a_B$ and K is a normalization constant chosen so that conservation of energy is satisfied. Equations (6.1) are then transformed into a recursive relation. After k iterates (denoted by the superscript), the $k+1$ iterate is computed by:

$$a_i^{k+1} = K F(\{a_1^k\}) \quad i = 1, \dots, B \quad (6.2)$$

Constant K is obtained by the conservation of energy: if N is the total number of counts in the object:

$$\sum_{i=1}^B a_i^{k+1} = \sum_{i=1}^B K F(\{a_1^k\}) = N$$

thus

$$K = N / \left[\sum_{i=1}^B F(\{a_1^k\}) \right] \quad (6.3)$$

It is usefull to consider a modification of equation (6.2) that improves convergence in certain cases. If α is a constant $0 < \alpha \leq 1$, the alternative recursive relation is:

$$a_i^{k+1} = (1 - \alpha) a_i^k + \alpha K F(\{a_1^k\}) \quad i = 1, \dots, B \quad (6.4)$$

Recursive relations (6.2) and (6.4) converge to the same solution.

The proof is as follows: let a_i^{lim} be the limit of the iterative procedure (6.2), then a_i^{lim} is the solution of the equation:

$$a_i^{\text{lim}} = K F((a_1^{\text{lim}})) \quad i = 1, \dots, B.$$

The limit of (6.4) is:

$$\begin{aligned} \lim [(1 - \alpha) a_i^k + \alpha K F((a_1^k))] &= (1 - \alpha) a_i^{\text{lim}} + \alpha K F((a_1^{\text{lim}})) \\ &= (1 - \alpha) a_i^{\text{lim}} + \alpha a_i^{\text{lim}} = a_i^{\text{lim}}. \end{aligned}$$

Thus, the limits of both procedures (6.2) and (6.4) are the same.

A. The direct maximization algorithm

The direct maximization of (5.3) has been reported previously by Nunez and Llacer (1989). The method has several theoretical advantages since it is directly comparable with the MLE and other reconstruction techniques. We have observed that, for the application presented here (tomography), direct maximization is a good computational tool that has no problems with negativity, unlike gradient techniques. The direct maximization algorithm was used by Gull and Daniel (1978).

To obtain the maximum of expression (5.3) the simplest way is to compute:

$$\frac{\partial BY}{\partial a_i} = 0 \quad i=1, \dots, B$$

and solve for a_i .

From equation (5.3) and taking into account that

$$\frac{\partial h_j}{\partial a_i} = \frac{\partial}{\partial a_i} [(\sum_{i=1}^B f_{ji} a_i) / \Delta p_j] = f_{ji} / \Delta p_j$$

we obtain for each of the pixels ($i = 1, \dots, B$):

$$\frac{\partial BY}{\partial a_i} = -(1/\Delta a) \log(a_i/\Delta a) - 1/\Delta a + \sum_{j=1}^D [-f_{ji}/\Delta p_j + (p_j/\Delta p_j) (1/h_j) (f_{ji}/\Delta p_j)] - \mu = 0.$$

From the definition of h_j :

$$h_j \Delta p_j = \sum_{i=1}^B f_{ji} a_i \quad j = 1, \dots, D,$$

we obtain:

$$\frac{\partial BY}{\partial a_i} = -(1/\Delta a) \log(a_i/\Delta a) - 1/\Delta a + \sum_{j=1}^D [f_{ji} (1/\Delta p_j) (p_j / \sum_{i=1}^B f_{ji} a_i - 1)] - \mu = 0 \quad (6.5)$$

and solving for a_i in the first term of the sum in (6.5) we obtain, for each pixel ($i = 1, \dots, B$):

$$a_i = K \exp \left(\Delta a \sum_{j=1}^D [f_{ji} (1/\Delta p_j) (p_j / \sum_{i=1}^B f_{ji} a_i - 1)] \right) \quad (6.6)$$

with constant K given by:

$$K = \Delta a \exp (-1 - \Delta a \mu)$$

Equations (6.6) are of the type of (6.1) and we can apply now the iterative expressions (6.2) or (6.4) to obtain the final recursive formula for our algorithm. After k iterates, (the superscript denotes the iteration number) the next iterate is computed by:

$$a_i^{k+1} = K \exp \left(\Delta a \sum_{j=1}^D [f_{ji} (1/\Delta p_j) (p_j / \sum_{i=1}^B f_{ji} a_i^k - 1)] \right) \quad (6.7)$$

or:

$$a_i^{k+1} = (1 - \alpha) a_i^k + \alpha K \exp \left(\Delta a \sum_{j=1}^D [f_{ji} (1/\Delta p_j) (p_j / \sum_{i=1}^B f_{ji} a_i^k - 1)] \right) \quad (6.8)$$

with the normalization constant K (which contains the Lagrange multiplier μ) being computed by expression (6.3). Note that the

determination of constant K by (6.3) is equivalent to determining the Lagrange multiplier μ and guarantees energy conservation.

Expresion (6.7) is our first restoring algorithm. The algorithm has the important feature that succesive iterates are all positive because the exponential function is always positive. In addition the presence of the exponential in the algorithm allows high values of a_i to develop in relatively few iterations. This is of considerable importance when the dynamic range of the restored object is high. The presence of the exponential results, however, in a problem which is the introduction of instability in the iterative process. For that reason we have used the equivalent expression (6.8) for the computations. Nevertheless, we will keep expresion (6.7) for theoretical purposes because it is simpler.

Note that in each of the iterations it is necessary to perform one projection into data space (sum over i) and one backprojection (sum over j) and, finally, one exponentiation for each pixel to obtain the updated values of the pixel intensities. Since most of the computation time is taken by performing the projection and the backprojection, this algorithm is not more expensive in computer time in each iteration that the MLE algorithm.

We have used the algorithm of (6.8) succesfully in our application to Positron Emission Tomography. When Δa is small, the algorithm of 6.7 works well and developes high values in few iterations. When Δa is bigger, expression (6.8) should be used.

The algorithm proposed in equation (6.7) has an important particular case when all the Δp_j have the same value. That means that the smallest intensity increment associated with each count does not depend on the particular detector. If we use Δp (without subscript) for such an increment, expression (6.7) becomes:

$$a_i^{k+1} = K \exp \left(\rho \sum_{j=1}^D [f_{ji} (p_j / \sum_{i=1}^B f_{ji} a_i^k - 1)] \right), \quad (6.9a)$$

where

$$\rho = \Delta a / \Delta p.$$

Taking into account the normalization condition (2.2) and grouping all constants together we obtain:

$$a_i^{k+1} = K \exp \left(\rho \frac{\sum_{j=1}^D [f_{ji} p_j]}{\sum_{i=1}^B f_{ji} a_i^k} \right) \quad (6.9b)$$

where, now, the constant K is:

$$K = \Delta a \exp (-1 - \Delta a \mu - \rho)$$

Note that in this case we do not need to know both the Δa and Δp parameters since only their ratio ρ appears in the algorithm.

Meinel (1986) obtained a similar restoring algorithm based on maximizing the likelihood in the Poisson noise case smoothed with the Shannon form of the entropy in his formula 48. He did not use Bayes' theorem but obtained the same formula (6.9a) except for the absence of the ρ parameter. This difference is of crucial importance because in (6.9a) ρ is the only parameter that can be used to adjust the relative weight between the entropy and the likelihood and avoid the problem of obtaining too smooth a distribution (even flat) when N is large. We will discuss this effect more extensively in section VII.

As indicated above, besides the desirable properties of the direct maximization algorithm (positivity, high dynamic range and reasonable computational cost), the most important problem of the algorithm is the instability in the iteration caused by the presence of the exponential function. This instability is greater for high values of the Δa or ρ parameters and it is necessary to average the sucesive iterates by using expression (6.8) instead of (6.7). Even so, when Δa becomes rather large, the fraction α has to be reduced drastically. For example, in the reconstructions of the Brain Phantom presented later, for $\Delta a = 50$, $\Delta p = 1$, $\alpha = 0.05$ was used and 50 iterations were needed to reach the point of convergence. That means that only 5 per cent of each new iteration was used. Even then, the number of iterates needed to obtain convergence was not particularly high. For comparison, the best reconstructions (for the

same problem) with the MLE algorithm are obtained stopping the iteration process at iteration 30.

To avoid the instability problem we have tested two gradient type algorithms and developed a new one.

B. The gradient algorithms

B1. The Steepest Ascent.

The simplest method to solve the problem of maximizing expression (5.3) by gradient techniques is to use the Steepest Ascent algorithm. After k iterates, the $k+1$ iterate is:

$$a_i^{k+1} = a_i^k + \omega_k \beta_k \left[\frac{\partial BY}{\partial a_i} \right] \text{evaluated at } \{a_i^k\} \quad i=1, \dots, B \quad (6.10)$$

or, in vector form

$$a^{k+1} = a^k + \omega_k \beta_k (\nabla BY)^k \quad i=1, \dots, B \quad (6.11)$$

where $(\nabla BY)^k$ is the gradient vector at k -th iteration, β_k is the step length and ω_k is an underrelaxation parameter. Expressions (6.10) or (6.11), with the derivatives given by equation (6.5), are formally the restoring algorithm by steepest ascent.

The first problem to solve is the presence of the Lagrange multiplier μ (introduced to conserve the total counts in the reconstruction) as an additive term in the derivative given by (6.5). In the direct algorithm (6.7) μ appeared in multiplicative form and was included in the constant. Since $\beta_k (\nabla BY)^k$ is in additive form, the introduction of a normalization factor K is not recommended. Applying the gradient correction will take the solution out of the $\sum a_i = N$ hyperplane and if we multiply by a factor in order to return to it we would be out of the direction of the gradient. To solve the problem, one can choose μ to fit some given value of the total energy (counts) or

modify slightly the function to be maximized. The second way is simpler because it does not introduce any computational problem and eliminates the additive presence of the Lagrange multiplier μ .

To modify the Bayesian function to be maximized, the easiest way is to modify the entropy part. If we consider the entropy alone (in absence of any data except the energy conservation) it is easy to prove that the solution is the "grey" map:

Maximize

$$BY = - \sum_{i=1}^B (a_i/\Delta a) \log(a_i/\Delta a) - \mu \left(\sum_{i=1}^B a_i - N \right)$$

gives the expected solution:

$$\mu = 1/\Delta a (\log(B \Delta a/N) - 1); \quad a_i = N/B \quad i=1, \dots, B$$

We want then to modify the entropy to obtain the same solution without the use of μ . Formally, the step is equivalent to introducing more "a priori" knowledge by means of the "Cross Entropy". The "Cross Entropy" is defined (Frieden, 1983b) as:

$$H' = - \sum_{i=1}^B (a_i/\Delta a) \log(a_i/\Delta a Q_i) \quad (6.12)$$

where Q_i represent our prior information. In absence of any other data, the maximization of the "Cross Entropy" gives the solution:

$$a_i = \Delta a \exp(-1) Q_i$$

i.e., the solution is proportional to the prior knowledge. If we want that this solution be the grey map as before, we choose as prior knowledge:

$$a_i = \Delta a \exp(-1) Q_i = N/B \quad \text{for all } i$$

or

$$Q_i = (\exp(1) N) / (B \Delta a) = C \quad \text{for all } i$$

The new function to be maximized by gradient techniques and the formulation of the problem equivalent to (5.3) is then:

Maximize

$$BY = - \sum_{i=1}^B (a_i / \Delta a) \log(a_i / (\Delta a C)) + \sum_{j=1}^D (-h_j + (p_j / \Delta p_j) \log(h_j)) \quad (6.13)$$

where

$$C = (\exp(1) N) / (B \Delta a) \quad \text{and} \quad h_j = (\sum_{i=1}^B f_{ji} a_i) / \Delta p_j \quad j = 1, \dots, D$$

The steepest ascent algorithm is then formulated by expressions (6.10) and (6.13) with the derivatives given by

$$\begin{aligned} \frac{\partial BY}{\partial a_i} = & - (1/\Delta a) \log(a_i / (\Delta a C)) - 1/\Delta a + \\ & + \sum_{j=1}^D [f_{ji} (1/\Delta p_j) (p_j / \sum_{i=1}^B f_{ji} a_i - 1)] \quad i=1, \dots, B \end{aligned} \quad (6.14)$$

It is important to choose the step length β_k properly because we want to obtain the maximum of BY along the direction of the gradient. There are several methods to compute the step length (Ortega and Rheinboldt, 1970) but the more accurate the method, the greater the computation time. A common method consists in using a quadratic interpolation polynomial which agrees with the function to be maximized at three points. The maximizer of the polynomial is taken as β_k .

The method is the following: at iteration k compute the gradient by (6.14), chose three values $\beta_1 \beta_2 \beta_3$ (usually $\beta_1 = 0$), compute three new points by (6.10) (use $\omega_k = 1$ in this step), compute the value of the Bayes' function by (6.13) at each of the three points, adjust an osculating parabola to the three values. The maximum of the parabola is chosen to be the step length. It is easy to show that if $B_1 B_2 B_3$ are the values of the Bayesian function given by (6.13), the step length is:

$$\beta_k = \frac{(\beta_2^2 - \beta_3^2)B_1 + (\beta_3^2 - \beta_1^2)B_2 + (\beta_1^2 - \beta_2^2)B_3}{(\beta_2 - \beta_3)B_1 + (\beta_3 - \beta_1)B_2 + (\beta_1 - \beta_2)B_3} \quad (6.15)$$

Since this method to compute the step length is only approximate, we must check that the use of the obtained β_k leads to an increment of the Bayes' function. If it is not the case, it is always possible to choose the underrelaxation parameter $\omega_k < 1$ to obtain the increment.

Computationally, the algorithm needs one projection and one backprojection to compute the gradient by (6.14) at each iteration and at least (using $\beta_1=0$) two more projections by (6.13) to compute the step length. The algorithm needs, therefore, approximately twice the CPU time as the MLE algorithm or the direct maximization discussed in section VI A.

The use of the steepest ascent method has some advantages and some disadvantages with respect to the direct maximization of section VI A.

The main advantage of the method is that the instability introduced in the direct maximization method by the exponential is not present now. This property allows very high values for the parameters Δa or ρ to be used without difficulty.

A non-severe disadvantage of the method is that each iteration needs about double the computer time than the previous algorithm. In addition, convergence is slow. A severe disadvantage of this method (often called a "catastrophic" disadvantage) is that expression (6.10) (basis of the method) does not guarantee positivity. In almost every iteration in which β_k is sufficiently large to enable significant development of the intensities a_i in reasonable time, there are pixels with negative $\partial BY / \partial a_i$ in which a_i becomes significantly negative. At the next iteration, it is not possible to compute the logarithms appearing in the gradient (6.14) (in fact, the entropy is not defined for negative values) and after each iteration any negative intensity must be reset to a small positive value. By doing this, we are deviating the iteration point from the search direction. The consequence is that neither the increase of the function (without underrelaxation parameter) nor the convergence are guaranteed. If many pixels become negative, the effect is to stop the algorithm progressing towards a maximum.

In our computations (see section X for details) for the example of the Brain Phantom in which the direct maximization needed 50 iterations (with $\Delta a = 50$ and $\alpha = 0.05$), the steepest ascent needs at least 250 iterations or about 10 times more computer time.

B2. The Conjugate Gradient

The best way to improve the steepest ascent is to use the conjugate gradient technique (see, p.e. Tarantola, 1987). The conjugate gradient starts like the steepest ascent (indeed, the first iteration is exactly the same) but while in each iteration the steepest ascent method seeks the maximization of the function along the direction of the steepest ascent (defined by a vector proportional to the gradient), the conjugate gradient maximize not in such steepest direction but in the vectorial subspace generated by the steepest ascent vector corresponding to the iteration and the steepest ascent vectors of all the previous iterations. The conjugate gradient was first applied to image reconstruction by Wernecke and D'Addario (1977).

The advantage of the method is that, by keeping all the previous information about the gradient at different points, we are introducing information about the curvature of the surface instead of using only the gradient at the point. We have, thus some information about the second derivatives and, hence, of the Hessian matrix. The convergence rate is quadratic while in the steepest ascent is linear. Also the conjugate gradient converges in a finite number of steps.

From the definition, one could expect that an algorithm based on such a method will converge much more rapidly than the steepest ascent. It is less expected, however, that the operations needed per iteration are not much more complex than those needed in the steepest ascent method.

The algorithm is the following: After k iterations, the $k+1$ iteration is, in vector form:

$$a^{k+1} = a^k + \omega_k \beta_k e^k \quad (6.16)$$

where:

$$e^0 = (\nabla BY)^0$$

$$e^k = (\nabla BY)^k + \gamma e^{k-1}$$

$$\gamma = |(\nabla BY)^k|^2 / |(\nabla BY)^{k-1}|^2.$$

....The function BY and its gradient are computed as before by (6.13) and (6.14), β_k is the step length computed by (6.15) as before and ω_k is the underrelaxation parameter to secure the increase of the function. The coefficient γ can be substituted by other equivalent ones (see, for example. Tarantola 1987).

Note that the operations needed in the conjugate gradient (6.16) are identical to those in the steepest ascent with the exception that during the $k+1$ iteration we must keep the gradient of the k iteration and the modulus of the $k-1$ gradient. This represents a negligible extra computational cost.

A consequence of the design of the algorithm is that when one is well away from the quadratic region near the maximum of the function, the technique becomes ineffective after a few iterations, since the partial information about the second derivatives is obtained at different points and thus quickly gets out of date. It is therefore more efficient to restart the conjugate gradient algorithm every few cycles.

In our computations to maximize the Bayesian function by the conjugate gradient, we have found (for the same example of Brain Phantom used with the steepest ascent) that the conjugate gradient is a considerable improvement over the steepest ascent. For comparison, the conjugate gradient reaches a good solution for the Brain Phantom ($\Delta a = 50$) in about 50 iterations instead of the 250 needed by the steepest ascent. This result is comparable in number of iterations (but worst in CPU time) to the result obtained with the direct maximization algorithm for this moderate value of Δa .

Nevertheless, the problem of negative values for the intensities is

not solved and the algorithm remains plagued by them. Almost all the background pixels become negative at one iteration or another and must be reset. The effect of that reset is to displace the new point from the direction of search. We have noticed that the effect of such displacement is a slowing of the convergence rate and a lack (although not very severe) of energy conservation.

A variant of the conjugate gradient has been developed by Skilling and Bryan (1984). They use the method represented by equation (5.4) and maximize the Entropy with a certain value of the χ^2 as a constraint for the case of Gaussian additive noise. In their method they search in a three dimensional subspace and use an entropy metric. The method helps to keep all the values positive although protection against stray negative values is still needed.

In our work, instead of attempting to modify the algorithm of Skilling and Bryan to fit our Bayesian maximization for the case of Poisson noise, we have preferred to develop a new iterative algorithm.

C. A New Proposed algorithm

The algorithms presented above for the maximization of the Bayesian function (5.3) have some problems that should be solved. These problems are the instability of the direct maximization algorithm and the non-positivity of the gradient techniques. The aim is to obtain a simple algorithm to maximize expression (5.3) with the following properties: 1) does not compute exponentials 2) positivity is guaranteed and 3) is convergent and competitively fast (at least as fast as the MLE or conjugate gradient).

The idea of the algorithm is the following: consider again the iterative procedure described by expressions (6.2), (6.3) and (6.4). Consider also equation (6.5) as the basic equation (as in the direct maximization algorithm). For each of the pixels ($i = 1, \dots, B$)

$$\frac{\partial BY}{\partial a_i} = -(1/\Delta a) \log(a_i/\Delta a) - 1/\Delta a + \\ + \sum_{j=1}^D [f_{ji} (1/\Delta p_j) (p_j / \sum_{i=1}^B f_{ji} a_i - 1)] - \mu = 0$$

Now, instead of solving for a_i from the first term of the sum, leading to an exponential function, we follow the procedure: first multiply by Δa and obtain

$$1 + \Delta a \mu - \log(\Delta a) = \Delta a \sum_{j=1}^D [f_{ji} (1/\Delta p_j) (p_j / \sum_{i=1}^B f_{ji} a_i - 1)] - \log(a_i).$$

Next, add an additive constant C to both sides, raise to the power p and multiply both sides by a_i

$$a_i \left[1 + \Delta a \mu - \log(\Delta a) + C \right]^p = \\ = a_i \left[\Delta a \sum_{j=1}^D [f_{ji} (1/\Delta p_j) (p_j / \sum_{i=1}^B f_{ji} a_i - 1)] - \log(a_i) + C \right]^p.$$

All the terms on the left hand side except a_i are constants that can be replaced by a single constant, leading to

$$a_i = K a_i \left[\Delta a \sum_{j=1}^D [f_{ji} (1/\Delta p_j) (p_j / \sum_{i=1}^B f_{ji} a_i - 1)] - \log(a_i) + C \right]^p \\ i = 1, \dots, B \quad (6.17)$$

Equations (6.17) are of the type of (6.1). Note that both (6.17) and (6.6) are two different alternatives to obtain a recursive formula of the type (6.1) from the basic equation (6.5). We can apply now the iterative expressions (6.2) or (6.4) to obtain the expressions of our new algorithm. After k iterations, (the superscript denotes the iteration number) the next one is computed by:

$$a_i^{k+1} = K a_i^k \left[\Delta a \sum_{j=1}^D [f_{ji} (1/\Delta p_j) (p_j / \sum_{i=1}^B f_{ji} a_i^k - 1)] - \log(a_i^k) + c \right]^p$$

$$i = 1, \dots, B \quad (6.18)$$

or:

$$a_i^{k+1} = (1 - \alpha) a_i^k + \alpha K a_i^k \left[\Delta a \sum_{j=1}^D [f_{ji} (1/\Delta p_j) (p_j / \sum_{i=1}^B f_{ji} a_i^k - 1)] - \log(a_i^k) + c \right]^p$$

$$i = 1, \dots, B \quad (6.19)$$

Equations (6.18) or (6.19) are the most general expressions of the new iterative algorithm to obtain the Maximum a Posteriori distribution defined by expression (5.3). The new algorithm will be referred to it as BMAPE (Bayesian Maximum A Posteriori with Entropy). Although the previous algorithms (by direct maximization and by gradient techniques) are also seeking the maximization of (5.3), we will reserve the abbreviation BMAPE for the specific expressions (6.18) and (6.19) of the new iterative algorithm. The algorithm can be simplified by dropping some of the constants that we have introduced. The meaning of the constants is explained in the next paragraphs.

First, it is important to note that the new algorithm does not compute exponentials and thus the instability of the direct maximization algorithm is no longer present. Also, at each iteration the algorithm computes one projection and one backprojection and, finally, one logarithm for pixel. The speed of the algorithm to compute one iteration is thus the same as the MLE or the direct maximization algorithm. It is also important to point out that all the new constants introduced (α , c and p) do not affect the convergence point and they are only acting on the stability of the process, the convergence rate and the guaranteeing of positivity. The convergence point is determined exclusively by the data and the parameters Δa and Δp_j .

The normalization constant K (which contains the Lagrange multiplier μ) is computed by expression (6.3) and has the same sense as in section VI A. Energy conservation is thus guaranteed.

The constant α ($0 < \alpha \leq 1$) introduced in (6.4) is used as before to correct possible instabilities during the iteration process. Since the algorithm does not compute exponentials, that constant is not necessary now. We have always used expression (6.18). Nevertheless, we keep expression (6.19) for generality.

The constant p has the property of improving the convergence rate. We have empirically observed that for $p > 1$ the convergence rate improves by a factor of approximately p and that for moderate values ($p \leq 3$) the algorithm is stable. For larger values ($p \geq 3$) convergence is not totally stable. The instability can be corrected by using expression (6.19) with a suitable value of α . The process is not very useful because, in that case, more iterations are needed and one loses the desired improvement in convergence rate. We have observed that the best procedure is to use the maximum value of p that does not affect stability and allows the use of expression (6.18). Note that by using $p = 3$ we can save up to 66 per cent of the CPU time needed for a given reconstruction.

The constant C is introduced to secure the positivity of the solution. The presence of two negative signs (-1 and $-\log(a_i^k)$) in the right hand side of (6.18) makes the positivity of the solution not automatically guaranteed unless an appropriate constant is introduced. Positivity is required because of the logarithm computation in (6.18). The effect of introducing the constant C is to lead all the pixels to positive values. Fortunately, the introduction of the constant does not affect the convergence point. The only side effect is a slight slowing of the convergence. This slowing can be practically eliminated by choosing the constant adequately. Since the most important source of possible negative values is the -1 on the right hand side of (6.18), a good guess for C is $C \approx \Delta a$. In our computations, we have always used $C = \Delta a$ without any problem in the convergence rate nor in the positivity of the pixels.

In our computations with the new algorithm (see section X), we have obtained very good results both with simulated and with real data. In particular, for the Brain Phantom example (one million counts, $\Delta a = 50$, $\Delta p = 1$) we have obtained the full Bayesian reconstruction without any positivity problem in 50 iterations with $C = 50$ and $p = 1$ and in 15 iterations (accelerating the algorithm) with $p = 3$ (for $P > 3$ some stability problems arise). This means that with this algorithm it is possible to obtain full Bayesian reconstructions seven times faster than with the Conjugate gradient algorithm and three times faster than with the MLE of the direct maximization presented in VI A.

It is interesting to consider, as in VI A, the particular case in which all the Δp_j have the same value Δp . Taking into account the definition of the parameter $\rho = \Delta a / \Delta p$ and the normalization condition (2.2), expression (6.18) becomes:

$$a_i^{k+1} = K a_i^k \left[\rho \sum_{j=1}^D [f_{ji} p_j / \sum_{i=1}^B f_{ji} a_i^k] - \rho - \log(a_i^k) + c \right]^p$$

$i = 1, \dots, B$

and including the $-\rho$ of the right hand side into the constant C :

$$a_i^{k+1} = K a_i^k \left[\rho \sum_{j=1}^D [f_{ji} p_j / \sum_{i=1}^B f_{ji} a_i^k] - \log(a_i^k) + c \right]^p \quad (6.20)$$

$i = 1, \dots, B$

Equation (6.20) is, thus, the expression of the new algorithm if all Δp_j are the same. Moreover, this expression has the particularity that the -1 that appears at the right hand side of (6.18) is not present. Taking into account that the constant C was introduced to avoid negative values, and considering that for ρ values that are not very small the possible negativities are due mainly to the -1 (the $-\log(a_i^k)$ have little influence on the negativity for ρ not very small), the right hand side of (6.20) is rarely negative and a value of $C = 0$ is often adequate.

The simplest form of the new algorithm is obtained by choosing the constants $C = 0$ and $p = 1$ in (6.20), i.e., not accelerating the convergence:

$$a_i^{k+1} = K a_i^k \left[\rho \sum_{j=1}^D [f_{ji} p_j / \sum_{i=1}^B f_{ji} a_i^k] - \log(a_i^k) \right] \quad (6.21)$$

Equation (6.21) works very well for values of ρ that are not very small. For small ρ , negativity problems could arise and (6.18) should be chosen.

For comparison, the expression of the Maximum Likelihood Estimator algorithm (MLE) introduced by Shepp and Vardi (1982) is:

$$a_i^{k+1} = a_i^k \left[\sum_{j=1}^D [f_{ji} p_j / \sum_{i=1}^B f_{ji} a_i^k] \right] \quad (6.22)$$

Note that the differences between the MLE algorithm (6.22) and the simplest version of the new iterative algorithm (6.21) are the presence of ρ and $\log(a_i^k)$ in (6.21). As we will discuss in section IX, these differences are important.

We have found that the new algorithm, in its complete version (6.18) or in its simplifications (6.20) or (6.21), is a very useful tool to obtain the Maximum a Posteriori full Bayesian reconstructions stated by expression (5.3). The algorithm works well both with simulated and real data, has a high convergence rate and no problems of negativity or stability.

VII. THE SHARPNESS PARAMETERS

By their definitions, parameters Δa , and Δp_j $j=1, \dots, D$ (or ρ in the case of $\Delta p_j = \text{constant}$ for all j) have an important role in the process of reconstruction. They are principal descriptors of the overall process of object emission, projection formation and reconstruction operations. Since they are related to the increments in the object and in the projections, they control the sharpness or smoothness of the reconstruction.

In order to compute what values should be used for the sharpness parameters we have two different approaches. The first is a direct estimation of the parameters and the second is based on the analysis of the residuals. We have used both methods successfully.

The Δp_j parameters (or Δp if there is no distinction between detectors) are very closely related to the data because they represent the smallest intensity increment associated with each detector. In the simplest case, we know that these increments have a value of unity if the j -th detector measures p_j counts with a resolution of one count. A different case is when a detector has counted P_j counts with a resolution of one count but due to the different gains of the detectors, this number should be multiplied by a factor f_j before using the data for a reconstruction. In that case, the result $p_j' = p_j f_j$ is actually p_j "paquets" of f_j counts and $\Delta p_j = f_j$ should be used. Note also that in that case, P_j is one realization of a Poisson variable, but p_j' is no longer a Poisson realization of that same variable. By choosing $\Delta p_j = f_j$ we are restoring to the quotient $p_j' / \Delta p_j = p_j$ tho its original Poisson character in accordance with equation (3.1).

Another method of estimating the Δp_j is based in the fact that, in the Poisson case, Δp_j is precisely linked to the signal-to-noise ratio (SNR). Suppose for simplicity that there is no detector dependence and we want to compute Δp . Suppose, also, that we know the SNR at the maximum intensity $p_{\max}$. Then, for Poisson statistics:

$$\text{SNR} = \sqrt{p_{\max} / \Delta p}$$

thus

$$\Delta p = p_{\max} / \text{SNR}^2$$

and

$$\rho = (\Delta a / p_{\max}) \text{SNR}^2 \quad (7.1)$$

One can see from equation (7.1) that the sharpness parameter ρ is large (and the reconstruction can be made sharp) when we know that Δa is large or when the SNR is large.

We will assume hereafter that the Δp_j parameters are known and we proceed to discuss the Δa parameter. That parameter is a representation of the finest intensity jumps that are possible in the object. This means that if Δa is large, we are allowing high jumps in the object and the reconstruction can be expected to contain a few sharp levels. In the opposite case, if Δa is small, we are allowing more levels in the reconstruction and, hence, we can expect it to be smooth. This basic interpretation coincides with our reconstruction findings, in which the use of a larger Δa results in increased sharpness in the reconstruction.

The main problem is how to estimate the value of Δa since the finest jumps in the object is usually unknown. We have several ways to do it.

Suppose that we want a certain number of levels in the reconstruction because, we know, for example, that our object has only a certain number of different levels, or our image display equipment allows some maximum number of different colour levels. Suppose, also, that by studying the data and/or from other information sources, we have certain knowledge about the maximum intensity in the object. In that case, a first guess for Δa can be obtained by assuming that the difference between all the levels is the same and:

$$\Delta a = \text{maximum in the object} / \text{number of levels}$$

A better estimation of both the maximum and the number of levels can be obtained by studying the results of a reconstruction made with the parameters computed from the first guess indicated above or by studying the reconstructions made with different reconstruction methods like the Filtered Backprojection or the MLE.

A completely different method for obtaining a correct value for the Δa parameter (assumed that Δp_j are known) is to study the residuals of the reconstruction and impose certain new constraints that a reconstruction must satisfy. This method, however, implies some theoretical deviation from the Bayesian solution developed so far in the present work. This is an important point that deserves a new section.

VII. THE CONCEPT OF FEASIBILITY OF A RECONSTRUCTION

The concept of feasibility as presented here has been developed by our group for the Poisson noise case as a result of the study of the image deterioration during the MLE reconstruction, see Llacer and Veklerov (1987a, 1987b, 1988), Llacer, Veklerov and Hoffmann (1988), Llacer, Veklerov and Nunez (1988), Nunez and Llacer (1989), Veklerov and Llacer (1987). In these papers, the concept of feasibility and the tests based on it can be found in detail and in this work we will only summarize the method and its application to our case. Several variations of the same concept for Gaussian noise have been reported by other authors working mainly in Astronomy, see Ables (1974), Gull and Daniell (1978), Reiter and Pfeleiderer (1986), Narayan and Nityananda (1987).

The idea of a feasible reconstruction is based in the size and distribution of the residuals. Suppose that one has made a reconstruction and after projecting the solution one realizes that all the residuals (differences between the projections of the solution and the true data) are very large. One will reject the solution because it is not coherent with the data. Suppose now that all your residuals are positive and all are equal to two standard deviations. Individually considered, each of the residuals is acceptable in a Poisson or Gaussian noise model but obtaining that result for the ensemble of all residuals is so unlikely, however, that we must also reject the solution. If we have 15000 data points, for example, the probability of obtaining such identical residuals is much smaller than obtaining 15000 "heads" consecutively when tossing 15000 fair coins at the same time. The same thing can be said of finding that all the residuals (or many of them) are very small.

The basic idea of the feasibility concept is that the residuals should be distributed in accordance of the Poisson nature of the process. Two definitions have been presented (Llacer, Veklerov and Nunez, 1988) and we present here a slightly modified version of them to include the "sharpness" parameters. They are:

Definition 1: The reconstruction a_i $i=1, \dots, B$ is said to be strongly feasible respect to data p_j $j=1, \dots, D$, if and only if we can accept (not reject) the statistical hypothesis that $p_j/\Delta p_j$ $j=1, \dots, D$ is a Poisson sample with respect to the projections h_j $j=1, \dots, D$ (see definition in 3.2).

Definition 1 is difficult to implement as a constraint to be tested and that is way we introduce another definition:

Definition 2: The reconstruction a_i $i=1, \dots, B$ is said to be k -feasible with respect to data p_j $j=1, \dots, D$, if and only if the first k moments of $p_j/\Delta p_j$ $j=1, \dots, D$ are consistent with the Poisson hypothesis, namely:

$$\sum_{j=1}^D \frac{(p_j/\Delta p_j - h_j)^n}{f_n(h_j)} \approx D \quad n = 1, \dots, k \quad (8.1)$$

where $f_n(x)$; $n=1, \dots, k$ is the expression for the n -th moment of the Poisson distribution with mean x .

The first moment meets this requirement automatically for any unbiased image. Here are the firsts several $f_n(x)$: $f_2(x) = x$, $f_3(x) = x$, $f_4(x) = 3x^2 + x$, $f_5(x) = 10x^2 + x$.

Note that for the case of $n = 2$ (only the second moment is considered) the left-hand side of (8.1) is identical to the expression of the chi-squared (χ^2) statistic and, thus, equation (8.1) is reduced to $\chi^2 = D$. As said in section V, this equation is precisely the constraint used by several authors in the maximum entropy method and we have used it to develop equation (5.4) for the Gaussian noise case. Equation (5.4) is, in our terminology, a maximization of the entropy with the constraint of 2-feasibility. The case of $n=2$ is called in this terminology weak feasibility. Note also that if $k \rightarrow \infty$ a n -feasible reconstruction becomes strongly feasible.

Several tests can be derived from the above definitions. The easiest one is to test for k -feasible images. In order to do so, we consider definition 2 and use expression (8.1) as a constraint that should be

satisfied. In practice, we have obtained high quality reconstructions by testing only for weak feasibility, i.e., testing for the condition $\chi^2 = D$. It is important to note that this condition is not a chi-square test of significance but a constraint. A problem arises when we want to determine an upper and a lower limit to the constraint in order to accept a reconstruction whose χ^2 has been calculated. This can be done by making use of the statistical theory for the χ^2 test of significance, but a new assumption should be made. This will be discussed below.

A test for strong feasibility test has been developed by Veklerov and Llacer (1987) by means of an algorithm that, instead of computing the different moments of the distribution, analyses the shape of the histogram of the residuals. They divide such histogram in N bins of equal probability and compute a function called H that is used to test the goodness of fit of the residuals to the Poisson hypothesis. A critical discussion of its validity is also given in Llacer, Veklerov and Nunez (1988) and in Nunez and Llacer (1989).

A fundamental objection affecting not only the above procedure but also the computation of the upper and lower limits for the weak feasibility test (and, in general, any k -feasibility test) has been pointed out to us by several workers. The objection stems from the fact that the image against which we are testing has been generated by the same set of data that we are testing. Under those conditions, the question implied by testing for any of the definitions of feasibility cannot even be posed. In this case we are only allowed to test the hypothesis that the data were obtained by a Poisson distribution with some values h_j unknown to us (see, p.e. Cramer, 1946). This limited test is unnecessary since we know that the data were really generated by a Poisson process. To avoid this problem a new assumption has been proposed (Llacer, Veklerov and Nunez 1988, Nunez and Llacer, 1989):

We have an image a_i $i=1, \dots, B$ and its projection h_j $j=1, \dots, D$. We do not know if it was obtained during our reconstruction process or if it was generated randomly by the proverbial team of monkeys making images by strewing white dots at random, a_i being the number that land in the i -th pixel. We have also a set of data p_j $j=1, \dots, D$ and we are asked to test the hypothesis that that image, as if completely specified, could

have been the source that generated the data by a Poisson process. We would then proceed with testing the hypothesis by the methods developed or implied by the definitions of feasibility given above.

This assumption gives us also the way to compute the upper and lower limits for the 2-feasibility test ($\chi^2 = D$) stated above. If we are testing the image as if completely specified, the lefthand side of equation (2.2) for $n=2$ obeys a χ^2 distribution with $D - 1$ degrees of freedom (Cramer, 1946). Statistical analysis indicates some upper and lower limits which χ^2 can plausibly take in a "two tail" chi-square test. The region between the upper and lower limits is the weak-feasibility region. The margins depend on the desired confidence levels. For χ^2 the largest acceptable value at 99 per cent confidence is about $D + 3.29 (D)^{1/2}$ (Skilling and Bryan, 1984). For $D=20000$ data points that means an upper limit for χ^2 of 20465. If normalized to the number of data points, the upper limit for χ^2/D is 1.023. The corresponding lower limit in a "two tail" chi-square test is about 0.977. Hebert et al. (1988) have developed their stopping criterion for the MLE algorithm based also on the χ^2 test of significance. It is important to keep in mind, however, that the expression $\chi^2=D$ is a constraint and that the statistical theory has been used only to define some confidence interval around the desired value D .

The application of the feasibility theory to the computation of the sharpness parameters could be carried out by starting with a guess value for Δa (we suppose Δp_j known but the method can be used also in the opposite case in which we know Δa and want to estimate Δp), carrying out the reconstruction using the proposed algorithm and, when it has converged, computing (or monitoring during the iterative process) the moments of the distribution of the residuals or the parameter H of Veklerov and Llacer (1987). If the statistics are out of the margins, the Δa parameter should be modified.

In practice we have obtained high quality reconstructions by monitoring only the χ^2 statistic (second moment). The parameter H was also correct in all of our reconstructions when the value of the χ^2 was between the allowed margins. The region between the upper and lower margins is called 2-feasibility region.

It is important to note that, as discussed by Skilling and Bryan (1984), with real data it is often impossible to reach the theoretical feasibility region since the set of the χ^2 values corresponding to all possible reconstructions compatible with the data has a minimum value and one should not attempt to aim below this minimum.

With regard to the theoretical deviation from a pure Bayesian approach when Δa is obtained from the additional constraint represented by a feasibility test, we know that we are placing an additional requirement (new prior knowledge) for the maximization of expression (5.3) which is not in a form that can be included in the prior probability $P(a)$. We are, therefore, deviating from the "pure" Bayesian reconstruction. In the results reported in this paper, we have estimated Δa by the theoretical method described in section VII, obtaining, therefore, true Bayesian solutions. The resulting images, it turned out, also pass the feasibility tests.

IX. PROPERTIES OF THE ALGORITHM. COMPARISON WITH THE MLE ALGORITHM

To sum up what has been found thus far: We have proposed a Maximum a Posteriori method of reconstruction for the Poisson noise case using entropy to define the prior probability and likelihood to define the conditional probability. We have introduced sharpness parameters Δa and Δp_j , $j=1, \dots, D$ that allow us to obtain a solution thorough expression (5.3) without the problem of the "flat" reconstructions associated with the pure Bayesian reconstructions. We have, also, proposed a new algorithm (6.18) to obtain the reconstruction. The algorithm presents the properties of good convergence rate, stability, and positivity. In the next paragraphs we will comment on the convergence point for different parameters and will make a comparison with the MLE algorithm.

Note that all the algorithms (6.7), (6.10), (6.16) and (6.18) proposed to solve the fundamental problem (5.3) are mathematically equivalent, although computationally different. For the comparison between the different values of the parameters and with the MLE algorithm

we have chosen, for simplicity, the case in which all the $\Delta p_j = \Delta p$ and $\rho = \Delta a / \Delta p$, i.e., we shall use iterative formulas (6.9b) and (6.20). The numerical examples presented here correspond to the Brain Phantom example with one million total counts (see section X for details).

We want to show, first, the differences between the MLE algorithm and the Maximum a Posteriori developed in this work.

As discussed above, the Maximum Likelihood in the case of Poisson noise is obtained by direct maximization of the conditional probability (3.1) with all $\Delta p_j = 1$. Shepp and Vardi (1982), using the EM algorithm, developed the Maximum Likelihood Estimator (MLE) method described by equation (6.22). Consider, now, the simplest form of our Maximum a Posteriori Algorithm (BMAPE) of (6.21) and consider the limit when the sharpness parameter ρ tends to infinity:

$$\lim_{\rho \rightarrow \infty} (\text{BMAPE}) = \lim_{\rho \rightarrow \infty} a_i^{k+1} = \lim_{\rho \rightarrow \infty} K a_i^k \left[\rho \sum_{j=1}^D [f_{ji} p_j / \sum_{i=1}^B f_{ji} a_i^k] - \log(a_i^k) \right]$$

Since the total energy is constant, the values of a_i do not change much when $\rho \rightarrow \infty$ and, thus, the term $-\log(a_i)$ is negligible compared with the other term in the sum containing the ρ parameter. Therefore:

$$\lim_{\rho \rightarrow \infty} a_i^{k+1} = \lim_{\rho \rightarrow \infty} K a_i^k \rho \sum_{j=1}^D [f_{ji} p_j / \sum_{i=1}^B f_{ji} a_i^k].$$

It is now possible to incorporate the parameter ρ into the constant K , thus:

$$\begin{aligned} \lim_{\rho \rightarrow \infty} (\text{BMAPE}) &= \lim_{\rho \rightarrow \infty} a_i^{k+1} = \lim_{\rho \rightarrow \infty} K a_i^k \sum_{j=1}^D [f_{ji} p_j / \sum_{i=1}^B f_{ji} a_i^k] \\ &= K a_i^k \sum_{j=1}^D [f_{ji} p_j / \sum_{i=1}^B f_{ji} a_i^k] = \text{MLE Algorithm} \end{aligned} \quad (9.1)$$

Since K is a constant to guarantee energy conservation, expression (9.1) is identical to the MLE algorithm (6.22).

The conclusion is that our Maximum a Posteriori algorithm is equivalent to the Maximum Likelihood when the sharpness parameter ρ tends to infinity. This fact, theoretically shown by (9.1), is also observed in our computations. We have, indeed, found that the MLE and BMAPE solutions are numerically identical when ρ is very large. This means that, since $\rho = \Delta a / \Delta p$ and Δp is given by the data, the MLE algorithm seeks the reconstruction with maximum sharpness, while in the (BMAPE) the sharpness is given by the parameter Δa (or ρ) which represents the finest known intensity jumps that are possible in the object. In other words, it appears that the MLE algorithm allows the reconstruction to have jumps of infinite height in the object while the BMAPE only allows jumps in accordance with Δa . This fact could explain the deteriorated "checkerboard" appearance of the MLE reconstructions when the reconstruction proceeds to convergence.

Note also that in the opposite case, when $\rho \rightarrow 0$, the BMAPE solution discards all the data P_j and the solution is the flat "grey" map expected. This means that both the flat map and the MLE solution are contained in our full Bayesian solution as extreme cases.

A different formulation of the MLE restoration gives us a different point of view. Consider the introduction of more "a priori" knowledge in our Maximum a Posteriori. This is done, as in (6.12), by means of the "Cross-Entropy". In this case, the first sum of the statement of the problem (5.3) has to be changed by (6.12) to give the problem:

Maximize

$$BY = - \sum_{i=1}^B (a_i / \Delta a) \log(a_i / \Delta a Q_i) + \sum_{j=1}^D (-h_j + (p_j / \Delta p_j) \log(h_j)) - \mu \left(\sum_{i=1}^B a_i - N \right) \quad (9.2)$$

By performing operations similar to the ones leading to equation (6.9b), the solution is found to be:

$$a_i^{k+1} = K Q_i \exp \left(\rho \sum_{j=1}^D [f_{ji} p_j / \sum_{i=1}^B f_{ji} a_i^k] \right) \quad (9.3)$$

Equation (9.3) is, thus, one of the mathematically equivalent expressions of our Maximum a Posteriori method (in the particular case $\Delta p_j=1$) when certain "a priori" knowledge Q_i is available. Note that if the prior information Q_i $i=1, \dots, B$ is unity for all the pixels (or any other constant), the new problem is identical to the fundamental one (5.3) and (9.3) is exactly (6.9b).

Consider now the Maximum Likelihood method. Instead of the formulation of Shepp and Vardi (1982), based on the EM algorithm, consider a different expression for the Maximum Likelihood obtained in the following manner: As indicated before, the Maximum Likelihood method consists in the direct maximization of the conditional probability (3.1). In this section we are considering the particular case in which all the $\Delta p_j = \Delta p$, thus the problem consists in maximizing:

$$L = \log \left(\prod_{j=1}^D \exp(-h_j) \frac{(h_j)^{p_j/\Delta p}}{(p_j/\Delta p)!} \right)$$

where

$$h_j = \left(\sum_{i=1}^B f_{ji} a_i \right) / \Delta p \quad j = 1, \dots, D$$

The maximum Likelihood solution is found by taking the partial derivative of L with respect to a_i for fixed $\{p_j\}$ and setting it equal to zero. Taking into account the normalization condition (2.2), it is easy to show:

$$\frac{\partial L}{\partial a_i} = -1/\Delta p + 1/\Delta p \sum_{j=1}^D [f_{ji} p_j / \sum_{i=1}^B f_{ji} a_i] = 0 \quad (9.4)$$

Multiplying both sides of (9.4) by ρ , exponentiating and multiplying by a_i we obtain:

$$a_i \exp(\rho) = a_i \exp \left(\rho \sum_{j=1}^D [f_{ji} p_j / \sum_{i=1}^B f_{ji} a_i] \right) \quad i=1, \dots, B \quad (9.5)$$

Expression (9.5) is of the type of (6.1), thus we can apply the iterative procedure (6.2) to obtain:

$$a_i^{k+1} = K a_i^k \exp \left(\rho \sum_{j=1}^D [f_{ji} p_j / \sum_{i=1}^B f_{ji} a_i^k] \right) \quad i=1, \dots, B \quad (9.6)$$

Expression (9.6) with K computed via (6.3) is, then, an iterative algorithm to compute the Maximum Likelihood reconstructions in the Poisson noise case. This algorithm is mathematically equivalent to the MLE algorithm (6.22) (but computationally different). In our computations we have obtained identical results at convergence by using the MLE expression (6.22) or the new expression (9.6). We have observed that the effect of the parameter ρ in (9.6) is only to modify the convergence rate.

We can now compare equation (9.3), corresponding to our Maximum a Posteriori, with (9.6), corresponding to Maximum Likelihood. They are similar but there is an important difference. That difference is that the "prior" knowledge Q_i in (9.3) is replaced by the values of the previous iteration a_i^k in (9.6).

The solution of Maximum Likelihood is, thus, equivalent to our Bayesian Maximum a Posteriori if we consider the result of each iteration as "prior" knowledge for the next one. The use of this method in a practical application appears not to be too attractive. Any error or instability is used as "prior" in the next iteration and thus propagated. This relationship illustrates, in our opinion, how the Maximum a Posteriori developed in this work is more cautious than the "pure" Maximum Likelihood method in restoring noisy images.

We conclude that the Maximum Likelihood method in the case of Poisson noise is equivalent to the Bayesian Maximum a Posteriori developed in this work only in the case in which we allow infinite sharpness in our reconstructions or in the case in which we update

continuously the "prior" information using the result of one iteration as a prior for the next.

Another important point is to study the path of the BMAPE algorithm during the iteration process towards the convergence point. It is also important to compare the different behaviour between the MLE and the BMAPE algorithms during the process.

There are several ways to monitor the path of the algorithms. We have chosen the value of the modified χ^2 statistic:

$$\chi^2/D = \frac{1}{D} \sum_{j=1}^D \frac{(p_j/\Delta p_j - h_j)^2}{h_j} \quad (9.7)$$

One of the fundamental differences between the MLE and our Bayesian Maximum a Posteriori is that, as is well known, the Maximum Likelihood method seeks a $\chi^2/D = 0$ solution while the BMAPE seeks a certain non-zero value which depends on the Δa and Δp_j (or ρ) parameters chosen. As indicated in section VIII, a "feasible" reconstruction should have a χ^2/D value between certain margins but never a zero (or very small) value (see section VII for details).

Figure 9.1 shows the comparison between the MLE algorithm and the BMAPE algorithm for different values of the sharpness parameter ρ . The figure plots the χ^2/D statistic against the iteration number for the example of the Brain Phantom with one million counts (see section X). Curve a) is the path of the MLE algorithm; curves b), c) and d) are the paths of the BMAPE algorithm for $\rho = 10$, $\rho = 25$ and $\rho = 50$, respectively. The two horizontal lines limit the feasibility region, as discussed in section VIII.

The BMAPE curves in the case of $\Delta p_j = \text{constant}$ can be computed directly by means of (6.20), but for consistency we have used expression (6.18) with $\Delta p_j = 1$ for all j and hence $\rho = \Delta a$. Values of $C = 20, 30$ and 50 were used for the different curves.

The different behaviour of the BMAPE and MLE algorithms is evident in Fig. 9.1. The path of the MLE algorithm (curve a) crosses the feasibility region from above in relatively few iterations and decreases continuously below it. By contrast, all the BMAPE curves reach a flat path. Curves b) ($\rho = 10$) and c) ($\rho = 25$) reach convergence above the feasibility region, while curve d) ($\rho = 50$) reaches convergence inside the feasibility region. If ρ is made larger ($\rho = 100$, for example) convergence will be below the feasibility region and if $\rho \rightarrow \infty$ the path of the BMAPE will be the MLE path.

The fact that the MLE path crosses the feasibility region and continues below it is the reason for stopping the algorithm after a certain number of iterations. In fact, all the stopping criteria developed for the MLE algorithm try to stop the process when it is inside the feasibility region or near it.

Figure 9.2 shows a comparison between the paths of the BMAPE algorithm in the non accelerated and accelerated cases. The figure plots the χ^2/D statistic as a function of iteration number for the Brain Phantom (one million counts). Curves a), b) and c) correspond to BMAPE reconstructions computed with $\rho = 50$. Curve a) is the non-accelerated case, $p = 1$; curves b) and c) are the BMAPE accelerated, $p = 2$ and $p = 3$ respectively. Curve d) is the MLE reconstruction for comparison.

Figure 9.2 demonstrates the acceleration of the BMAPE algorithm without changing the convergence point. The MLE algorithm (curve a) enters the feasibility region at approximately iteration 27 and leaves it at iteration 40. The non-accelerated ($p = 1$) BMAPE algorithm (curve b) enters the feasibility region at iteration 50 (nearly at convergence) and reaches full convergence at approximately iteration 65. With the higher acceleration obtained with $p = 3$ (curve d), the path of the algorithm enters the feasibility region at iteration 15 and reaches full convergence at iteration 25.

Figures 9.1 and 9.2 show that the new Bayesian Maximum a Posteriori algorithm, in addition to the favorable properties summarized at the beginning of this section, converges inside the feasibility region and

can be made to be faster than the MLE algorithm.

X. APPLICATION TO POSITRON EMISSION TOMOGRAPHY

We have applied the theory and algorithms developed in this work to Positron Emission Tomography (PET). In PET, a radioactive isotope is injected into the body. The radioactive isotope emits positrons following the Poisson process. The annihilation of the positron with an electron forms a pair of γ -rays which are emitted in opposite directions, along a randomly oriented line. A ring of detector elements surrounds the body. The elements detect the photon pairs in time coincidence. Each pair of detectors define a detector unit known as a "tube".

If we detect only the coincidence, we know only that the emission occurred somewhere in the tube. If, in addition, we detect, the difference of time (time of flight) between the arrival of the two γ rays, we can define, with some error, the region the tube in which the emission took place. In our computations we have considered only the first case in which we have no time of flight measurements.

In this application, the elements of the fundamental imaging equation (2.1)

$$\sum_{i=1}^B f_{ji} a_i + n_j = p_j \quad j = 1, \dots, D$$

have the following meanings: a_i $i=1, \dots, B$ is the unknown emitter intensity at the i -th pixel; p_j $j=1, \dots, D$ is the known total number of coincidences counted in the j -th tube; n_j is the noise (Poisson) at j -th tube and, finally, the point spread function element f_{ji} is the probability of an emission in pixel i being detected at tube j .

We have carried out several groups of reconstructions both with computed generated data and with real data from the ECAT-III tomograph of U.C.L.A. The tomograph has 512 detectors defining 49664 tubes. In practice, only the central region of 28720 tubes is used. The reconstructions have been carried out by defining the image as a square

of 128x128 pixels. Thus, in the results presented here, $D = 28720$ and $B = 16384$. The generation of the point spread function (transition matrix), f_{ji} , used in the reconstruction process (and also in the generation of simulated data), was carried out with the geometric parameters of the ECAT-III tomograph following a simple prescription by Shepp and Vardi (1982).

A. Simulated data

We have used a computer generated Brain Phantom with one million counts mimicking complex brain structures. The generation of the projection data was carried out in two steps: 1) generation of the source image by a random process based on the probabilities indicated by the desired intensities in the phantom with a finite number of counts, and 2) generation of the projection data by taking each count in the source image and using a second random process based on the transition probabilities f_{ji} to determine which detector tube would receive the count. Figure 10.1 a) shows the source image.

We computed the sharpness parameters from first principles, as described in section VII. Parameters Δp_j , $j=1, \dots, D$ are known for this case with simulated data. The smallest intensity increment associated which each detector is one count, thus:

$$\Delta p_j = \Delta p = 1 \quad j=1, \dots, D$$

In order to compute parameter Δa , the finest known intensity jump in the object, we have used two methods, as described in section VII :

First method: from previous reconstructions made by stopping the MLE algorithm and by filtered backprojection we know that the maximum of the object is about 200 counts. We know, also, that the number of levels in the object is four (the phantom was designed with relative activities of 25% in the low and 100% in the high radioactivity areas). Then, $\Delta a = \text{maximum in the object} / \text{number of levels} = 200/4 = 50$

Second method: By using the feasibility criterion described in section VIII, we can adjust the Δa parameter to obtain a convergence point inside the feasibility region. Figures 9.1 and 9.2 show that the convergence inside the feasibility region is obtained when Δa is approximately 50. Note that, in this case at least, the agreement between the two is excellent.

Table 10.1 summarizes the main characteristics of the different feasible reconstructions we have carried out with this phantom. The

Table 10.1. Main characteristics of the different reconstructions made with the Brain Phantom (128x128 pixels, 1 million counts, $\rho = \Delta a = 50$, $\Delta p_j = 1$).

Method	Equation	It. Feasibility ^{a)}	CPU time units ^{b)}	Notes
M.L.E.	6.22	30	30	1)
Direct Max.	6.8	50	50	2)
Steepest Asc.	6.11	>250	>500	3)
Conjugate grad.	6.16	35	70	4)
BMAPE (p=1)	6.18	50	50	5)
BMAPE (p=2)	6.18	25	25	6)
BMAPE (p=3)	6.18	15	15	7)

Notes:

a) Iteration at which feasibility is reached. In all cases except MLE the reconstructions at full convergence are nearly identical to the feasibility point ones.

b) CPU time unit is the time required to perform one projection and one backprojection.

1) MLE stopped before convergence point by feasibility criterium.

2) $\alpha = 0.05$ used. Good convergence.

3) Very slow. Convergent with negativity problems.

4) Restarted each 10 iterations. Convergent with negativity problems.

5) $p = 1$ and $C = 50$ used. Very stable. Good convergence.

6) $p = 2$ and $C = 50$ used. Fast and stable. Good convergence

7) $p = 3$ and $C = 50$ used. Very fast, still stable. Good convergence.

first column indicates the method used in the reconstruction; the second column indicates the corresponding equation used; the third column indicates the iteration number at which the reconstruction becomes feasible; the fourth column indicates the relative CPU time requirements to obtain the reconstruction. Figures 9.1 and 9.2 complement the results showed in the table. The changes in the reconstructed images in going from the first feasible iteration up to convergence are so small that they can be considered identical.

Figure 10.1 shows the results of reconstructions for the Brain Phantom with one million counts. a) is the source image; b) and c) are the MLE reconstructions at the stopping point (iteration 30) and at iteration 100 respectively; d) is the BMAPE reconstruction with a low $\Delta a = \rho = 10$. e) is the non-accelerated BMAPE reconstruction computed with the correct $\Delta a = \rho = 50$ at convergence (iteration 50). f) is the accelerated ($p=3$) BMAPE reconstruction, also $\Delta a = \rho = 50$ at convergence (iteration 15).

Note the degraded appearance of the MLE reconstruction at iteration 100 and the fuzzy Bayesian reconstruction d) when a too small a parameter $\rho = 10$ is used. Note also the correct reconstruction e) obtained at convergence when an adequate $\rho = 50$ is used. Figure 10.1 also demonstrates a full Bayesian reconstruction f) in as few as 15 iterations using the new BMAPE algorithm.

B. Real data

In order to test the algorithm with real data we have used the results of scans from the UCLA ECAT-III tomograph, with the Hoffmann Brain Phantom as subject, both with one and 30 million counts. The case with one million counts is usefull for testing the behaviour of the algorithm in a typical real case with a limited number of counts. The case with 30 million counts is useful to test the applicability of the algorithm in a real case with a high number of counts and, therefore, a high value for the Δa parameter (remember the flateness problems of the "pure" Bayesian when the number of counts is high).

In order to compute the Δa and Δp_j parameters we apply the theory developed in section VII. From previous computations (MLE and Filtered Backprojection, see, p.e. Veklerov, Llacer and Hoffman, 1988) we know that, for one million counts, the maximum of the image is approximately 3000 counts in a pixel and that the number of activity levels is approximately 5 (the approximate ratio of activities between high and low activity regions is 4.5 to 1). Parameter Δa was, therefore, set to $\Delta a = 3000/5 = 600$. By similar considerations, for 30 millions counts, the Δa parameter was set to $\Delta a = 16500$.

Since this is a case with real data, there are absorption and detector gain corrections to be applied. The data P_j were multiplied by factors g_j before the reconstruction. One count received by a detector pair (tube) j corresponds to an increment of g_j counts. Thus parameters Δp_j should be set to the values $\Delta p_j = g_j$.

Figure 10.2 shows the results of the reconstructions. Fig. 10.2a) is the outline of the Hoffman brain phantom used. Fig. 10.2b) shows the BMAPE reconstruction for the 30 million counts image with $\Delta a = 16500$, at convergence ($p=2$, $c=12000$). Fig 10.2c) is the BMAPE reconstruction of the one million counts image with $\Delta a = 600$, at convergence ($p = 2$, $c = 450$). The resulting reconstructions pass the feasibility test described by Llacer and Veklerov (1989) for real data. Note that in the case of 30 million counts, the high value of Δa makes it impossible to use the direct maximization algorithm described in section VI A. It would be necessary to handle exponentials of the order of $\exp(12000)$ in equation (6.7), and that is not practical computationally.

The quality of the images of Figs. 10.2b) and 10.2c) is comparable to the best of the MLE solutions and they are superior to the Filtered Backprojection images that we have used for comparison. A statistical analysis of the value of images obtained by the new algorithm will be carried out in the near future.

In all the reconstructions for simulated or real data presented here, parameters Δa and Δp_j have been computed from the pure Bayesian

theory, following section VII. We have obtained, therefore, true Bayesian solutions. The results also fit the feasibility theory presented in section VIII. We do not have at this time, however, the theoretically or practical knowledge to insure that this will always be the case. A practical adjustment of Δa for a given class of images can be carried out by testing the reconstructed images for feasibility, as described in section VIII.

XI. CONCLUSIONS

In this paper we have proposed a Maximum a Posteriori method of image reconstruction in the Bayesian framework for the Poisson noise case. We have used entropy to define the prior probability and likelihood to define the conditional probability. The introduction of the sharpness parameters Δa and Δp_j , $j=1, \dots, D$, which can be theoretically computed or adjusted, allows us to obtain a Maximum a Posteriori solution without the problem of the reconstructions associated with "pure" Bayesian reconstructions.

We have developed several ways to solve the fundamental problem (5.3) and proposed a new algorithm (6.18) to compute the reconstruction through an iterative formulation. The algorithm presents the desirable properties of good convergence rate, stability, and positivity, and converges into the feasibility region faster than the MLE method.

We have also shown that the Maximum Likelihood method in the case of Poisson noise is equivalent to the Bayesian Maximum a Posteriori developed in this work when we allow infinite sharpness in the reconstruction or in the case in which we update continuously the "prior" information by using the result of the previous iteration.

Table 10.1 and Figures 9.1, 9.2, 10.1 and 10.2 show the results of applying the proposed Bayesian Maximum a Posteriori method and the behaviour of the different algorithms in Emission Tomography both with simulated and real data. The quality of the reconstructions obtained allow us to conclude that the method shows promise for a new methodology

for the successful reconstruction of for Emission Tomography images from Bayesian functions.

ACKNOWLEDGEMENTS

The authors would like to acknowledge valuable discussions with E. Veklerov on different aspects of the work presented. This work has been supported, in part, by a grant from the National Institutes of Health (CA-39501) and the U.S. Department of Energy under Contract No. DE-AC03-76SF00098. The work of J. Nunez has been supported in part by the Catalan Studies Program (U. of California and the Generalitat de Catalunya).

REFERENCES

- Ables, J.G., "Maximum Entropy spectral analysis", *Astron. Astrophys. Suppl. Ser.* vol. 15, 383-393, (1974).
- Akaike, H., "The selection of Smoothness priors for distributed lag estimation", in *Bayesian Inference and decision Techniques*, P. Zellner and P. Goel ed. North-holland, Amsterdam 109-118, (1986).
- Andrews, H.C. and Hunt, B.R., *Digital Image Restoration*, Prentice-Hall Inc. (1977).
- Bryan, R.K., Bansal, M., Folkhard, W, Nave, C., and Marvin, D.A. *Proc. Nat. Acad. Sci. U.S.A.* vol. 80, 4728 (1983).
- Cramer, H., *Mathematical Methods of Statistics*. Princeton Univ. Press. Princeton, (1946).
- Frieden, B.R., "Restoring with Maximum Likelihood and Maximum Entropy", *J. Opt. Soc. Amer.*, vol. 62, 511-518, (1972).
- Frieden, B.R., *Probability, Statistical Optics and Data Testing*, Springer-Verlag, Berlin-Heidelberg (1983a).
- Frieden, B.R., "Unufied Theory for Estimating Frequency-of-occurrence laws and optical objects", *J. Opt. Soc. Amer.* vol. 73. 927-938, (1983b).
- Frieden, B.R. and Wells, D.C., "Restoring with maximum entropy. III Poisson sources and backgrounds", *J. Opt. Soc. Am.*, Vol 68, No. 1, 93-103, (1978).

- Geman, S. and McClure, D.E., "Statistical Methods for Tomographic Image Reconstruction". Proc. 46th session of ISI. Bull. of the ISI., vol. 52, (1987).
- Gull, S.F. and Daniel, G.J., "Image Reconstruction from Incomplete and Noisy Data", *Nature*, vol. 272, 686-690, (1978).
- Hebert, T., Leahy, R. and Syngh, M., "Fast MLE for SPECT using an intermediate polar representation and stopping criterion", *IEEE Trans. Nucl. Sci.* Feb., (1988).
- Jaynes, E.T. *Phys. Rev.* vol. 106, 620-630, also vol. 108, 171-190, (1957). (These papers are reprinted in *Papers on Probability, Statistics and Statistical Physics*, a reprint collection, Dordrecht, The Netherlands, D. Reidel, 1982).
- Jaynes, E.T., "Prior Probabilities", *IEEE Trans.* vol. SSC-4, 227-241, (1968).
- Jaynes, E.T., "On the Rationale of the Maximum-Entropy Methods", *Proc. IEEE*, vol. 70, 939-952, (1982).
- Jaynes, E.T., "Bayesian Methods: General Background. An Introductory Tutorial". In *Maximum Entropy and Bayesian Methods in Applied Statistics*. Proc. of the Fourth Maximum Entropy Workshop. Ed. J. H. Justice. Univ. of Calgary. pp. 1-25, (1984a).
- Jaynes, E.T. "Monekys, Kangaroos and N", in *Maximum Entropy and Bayesian Methods in Applied Statistics*, Proc. of the Fourth Maximum Entropy Workshop, Ed. J.H. Justice, Univ. of Calgary, 26-58, (1984b).
- Levitan, E. and Herman, G.T., "A Maximum a Posteriori Probability Expectation Maximization Algorithm for Image Reconstruction in Emission Tomography", *IEEE Trans. Med. Imaging*, vol. MI-6, no.3, 185-192, (1987).
- Liang, Z. and Hart, H. "Bayesian Reconstruction in Emission Computerized Tomography", *IEEE Trans. Nuc. Sci.*, vol 35 no. 1 788-792, (1988).
- Llacer, J. and Veklerov, E., "The high sensitivity of the Maximum Likelihood Estimator method of Tomographic Image Reconstruction". *SPIE Proc., Conf. on Med. Imag.*, vol. 767, 70-76, (1987a).
- Llacer, J. and Veklerov, E. "The Maximum Likelihood Estimator Method of Image Reconstruction: Its Fundamental Characteristics and Their Origin", *Proc. of the Xth Information Processing in Medical Imaging (IMPI), International Conference, Utrecht, The Netherlands*, (published by Plenum Pub. Co.), 201-215 (1987b).
- Llacer, J. and Veklerov, E. "The Use of a Stopping Rule in Iterative Image Reconstruction", invited paper, June 1988 American Mathematical Society-IMS-SIAM Joint Conference on Spatial Statistics and Imaging, Brunswick, Maine, to be published in the Conference Proceedings, (1988).

- Llacer, J. and Veklerov, E. "Feasible Images and Practical Stopping Rules in Iterative Image Reconstruction", accepted for publication, *IEEE Trans. Med. Imaging*, (1989).
- Llacer, J. Veklerov, E. and Hoffman, E.J., "On the convergence of the Maximum Likelihood Estimator Method of Tomographic Image Reconstruction", presented in the SPIE Conference on Medical Imaging, Newport Beach, CA, 1987. *SPIE Proceedings*, vol. 767, 70-76 (1987).
- Llacer, J. Veklerov, E. and Nunez, J., "The Concept of Causality in Image Reconstruction", presented at the Sept 1988 NATO Advanced Study Institute, Povoas, Portugal, to be published in the Institute Proceedings, (1988).
- Meinel, E.S., "Origins of linear and nonlinear recursive restoration algorithms", *J. Opt. Soc. Am. A*, Vol.3, No. 6, 787-799, (1986).
- Miller, M.I. and Snyder, D.L., "The Role of Likelihood and Entropy in Incomplete-data Problems: Application to Estimating Point-process Intensities and Toeplitz Constrained Covariances", *Proc. IEEE*, vol. 75, no. 7, 892-907, (1987).
- Minerbo, G., "MENT : A Maximum Entropy Algorithm for Reconstructing a Source from Projection Data" *Computer Graphics and Image Processing*, vol 10, 48-68 (1979).
- Narayan, R. and Nityananda, R. "Maximum Entropy Image Restoration in Astronomy", *Ann. Rev. Astron. Astrophys.* vol. 24, 127-170, (1986).
- Nunez, J. and Llacer, "Maximum Entropy and the Concept of Feasibility in Tomographic Image Reconstruction", presented in the SPIE Conference on Medical Imaging, Newport Beach, CA, 1989. *SPIE Proceedings*, (1989).
- Ortega, J.M. and Rheinboldt, W.C., *Iterative solutions of nonlinear equations in several variables*. Academic Press. New York. (1970).
- Reiter. J. and Pfeleiderer, J., "Improvement of MEM-deconvolution by an additional constraint", *Astron. Astrophys.* vol. 166, 381-392, (1986).
- Shepp, L.A. and Vardi, Y., "Maximum Likelihood Reconstruction for Emission Tomography", *IEEE Trans. Med. Imaging*, vol. MI-1, No. 2, 113-121, (1982).
- Skilling, J., "Theory of Maximum Entropy Image Reconstruction" In *Maximum Entropy and Bayesian Methods in Applied Statistics*. Proc. of the Fourth Maximum Entropy Workshop. Ed. J. H. Justice. Univ. of Calgary. pp. 156-178, (1984).
- Skilling, J. and Bryan, R.K. "Maximum Entropy Reconstruction: general algorithm", *Monthly Notices of the Royal Astronomical Soc.*, vol. 211, 111-124, (1984).

- Skilling J. and Gull, S.F., "The Entropy of an Image". *SIAM Amer. Math. Soc. Proc.*, vol. 14, pp. 167-189. (1984).
- Skilling, J. and Gull, S.F., personal communication, publication in progress (1988).
- Smith, W.E., Barrett, H.H. and Paxman, R.G., "Reconstruction of Objects from Coded Images by Simulated Annealing", *Optical Letters*, vol. 8 199-201 (1983).
- Tarantola, A. *Inverse Problem Theory*. Elsevier, (1987).
- Veklerov, E. and Llacer, J., "Stopping Rule for the MLE Algorithm Based on Statistical Hypothesis Testing", *IEEE Trans. Med. Imaging*, vol. MI-6, 313-319, (1987).
- Veklerov, E., Llacer, J. and Hoffman, E.J., "MLE reconstruction of a brain phantom using a Monte Carlo transition matrix and a statistical stopping rule", *IEEE Trans. Nucl. Sci.*, vol. NS-35 no. 1, 603-607, (1988).
- Wernecke, S.J. and D'Addario, L.R., "Maximum Entropy Image Reconstruction", *IEEE Trans. on Computers*, vol. C-26, no. 4, 351-364, (1977).

FIGURE CAPTIONS.

Fig. 9.1 - Comparison of the behavior of residuals between MLE and BMAPE algorithms for the mathematical brain phantom with 1 million counts, as a function of iteration number. a) MLE algorithm, b) BMAPE with $\rho = 10$, c) BMAPE with $\rho = 25$, and d) BMAPE with $\rho = 50$.

Fig. 9.2 - Comparison of the behavior of residuals between the BMAPE algorithm with different degrees of acceleration and MLE, for the mathematical brain phantom with 1 million counts, as a function of iteration number. Parameter $\rho = 50$. a) not accelerated, $p = 1$, b) accelerated with $p = 2$, c) accelerated with $p = 3$, and d) MLE algorithm.

Fig. 10.1 Results of reconstructions with simulated data. Mathematical brain phantom with 1 million counts. a) source image, b) MLE reconstruction at stopping point (it. 30), c) MLE reconstruction at iteration 100. d) BMAPF reconstruction with a low $\Delta a = \rho = 10$. e) non-accelerated BMAPE reconstruction with the correct $\Delta a = \rho = 50$ at convergence (it. 50). f) accelerated BMAPF reconstruction ($p = 3$) and the correct $\Delta a = 50$, at convergence (it. 15).

Fig. 10.2 Results of reconstructions with real data. Reconstructions of the Hoffman brain phantom with data from the UCLA ECAT-III tomograph. a) Outline of the phantom, b) BMAPE reconstruction of the 30 million count image computed with $\Delta a = 16500$, at convergence ($p = 2$, $C = 12000$). c) BMAPE reconstruction of the 1 million image, $\Delta a = 600$, at convergence ($p = 2$, $C = 450$), and d) same image as c) with slight Gaussian filtering to improve its appearance, $\sigma = 0.5$ pixels.

Fig. 9.1 - Comparison between MLE and BMAPE

Fig. 9.2 - Acceleration of BMAPE algorithm

Fig. 10.1 - Results of reconstructions with simulated data

5
4
3
2
1

e

XBC 897-5684

f

XBC 897-5685

Fig. 10.2 - Results of reconstruction with real data

ucla1meq200_1.15500

b

XBC 897-5686

ucla1meq300_0.5

c

XBC 897-5687