LBL-27074
Preprint

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

Engineering Division

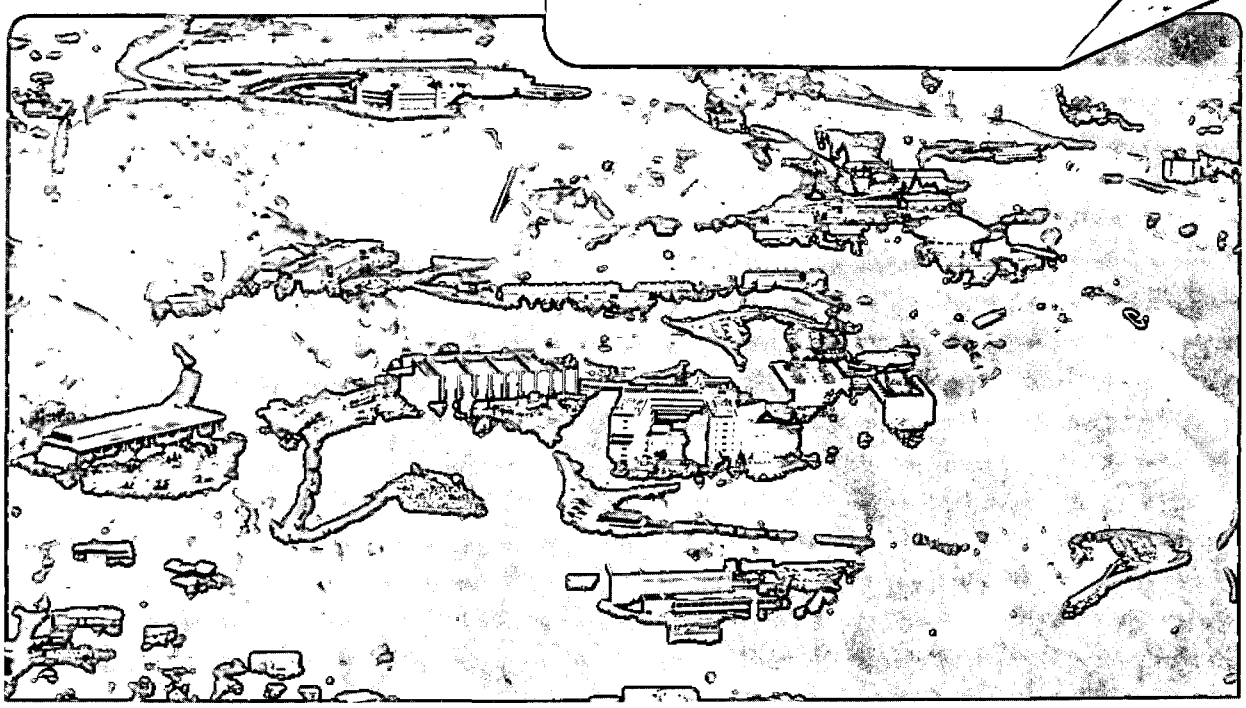
Submitted to IEEE Transactions on Medical Imaging

A Fast Bayesian Reconstruction Algorithm for Emission Tomography with Entropy Prior Converging to Feasible Images

J. Nunez and J. Llacer

January 1989

TWO-WEEK LOAN COPY
This is a Library Circulating Copy
which may be borrowed for two w



Bldg. 50 Library.
Copy 2
LBL-27074

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

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, 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 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 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.

LBL-27074

A fast Bayesian reconstruction algorithm for emission tomography with entropy prior converging to feasible images

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

* On leave from the Department of Physics, University of Barcelona, Spain

This work has been supported, in part, by a grant from the National Institutes of Health, CA-39501 and by the U.S. Department of Energy under Contract No. DE-AC03-76SF00098. Dr. Nunez was supported, in part, by the Center of Catalan Studies, Generalitat de Catalunya, Spain, and University of California, Berkeley.

Abstract - This paper reports on the development and tests of a new iterative reconstruction algorithm for Emission Tomography based on Bayesian statistical concepts. It uses the entropy of the generated image as a "prior" distribution, it can be accelerated by the choice of an exponent and it converges uniformly to feasible images by the choice of one adjustable parameter. A feasible image has been defined as one that is consistent with the initial data, i.e., it is an image that, if it were a true radiation source in a patient, could have generated the initial data by the Poisson process that governs radioactive disintegration. The paper discusses the fundamental ideas of Bayesian reconstruction, the use of an entropy prior with an adjustable "contrast parameter", the use of likelihood with "data increment" parameters as conditional probability and the development of the new Fast Maximum a Posteriori with Entropy (FMAPE) algorithm by the successive substitution method. It is shown that both in the Maximum Likelihood Estimator (MLE) and FMAPE algorithms, the only correct choice of initial image for the iterative procedure in the absence of *a priori* knowledge about the image configuration is a uniform field. Results of reconstruction with mathematical phantoms and with real data from the Hoffman brain phantom are shown to converge to excellent images that do not exhibit the characteristic deterioration of MLE reconstructions at high number of iterations.

Keywords: Bayesian reconstruction, entropy prior, emission tomography, feasible images

A fast Bayesian reconstruction algorithm for emission tomography with entropy prior converging to feasible images

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

I. Introduction

Recent work with the Maximum Likelihood Estimator (MLE) method of image reconstruction [1,2,3,4] has shown that it is possible to obtain visually excellent images from Emission Tomography (ET) data in a reliable manner by stopping the MLE iterative algorithm according to a stopping rule. That rule is based on testing whether the resulting image is "feasible" in a statistical sense. A feasible image has been defined as one that could have generated the original projection data by the statistical process that governs the measurement. Detailed definitions for feasibility have been given in [3,4,5] and the stopping rule has been described in detail in [1] for simulated data and in [3] for real data from a tomograph.

We have postulated that feasible images are important because they are consistent with the data. In the case of MLE reconstructions, images obtained at iterations before the region of feasibility do not have all the detail that could be expected. Past the feasibility region, images become excessively noisy in the regions of high activity. In the feasibility region, reconstructed images are found to present a balance between sharpness and noise in the high activity regions that makes them good representations of the original radioisotope field and may compete favorably in diagnostic value with the best filtered backprojection images.

*On leave from the Department of Physics, University of Barcelona, Spain

Upon further investigation, we have found in [5] a number of ways of obtaining feasible images, in all cases starting the iterative procedure from a uniform image field: 1) through the use of the stopping rule just indicated, 2) by continuing the MLE process past the stopping point by an undetermined number of iterations and post-filtering with a Gaussian kernel, and 3) by reconstructing with a Bayesian method with an entropy prior.[4] We failed to obtain feasible images with a Maximum Entropy reconstruction with constraints, using a simple noise model.[5]

Feasible images from the same set of data are similar to each other in appearance, but they are not identical. Those images obtained by iterating past the stopping rule and slight post-filtering (with kernels of 0.4 to 0.7 pixels std. deviation) have sharpness and noise behavior comparable to the images obtained at the stopping rule, but have a more accurate representation of the activity levels in narrow structures. The same favorable characteristics have been observed in our Bayesian reconstructions with entropy prior. Considering the fact that the latter reconstructions can be made to converge to a feasible image with excellent stability by the selection of only one adjustable parameter, we decided to pursue the improvement of the initial Bayesian algorithm of [4] to make it more reliable and faster. We also needed to better understand its convergence characteristics and the choice of the adjustable parameter.

This paper describes the ideas leading to the Bayesian method with entropy prior for emission tomography, describes the development of the new, improved algorithm, discusses the process of speeding it up and maintenance of non-negativity, the selection of the only adjustable parameter that affects the resulting image, and presents the results of reconstructions with two mathematical phantoms and with real data from the Hoffman brain phantom, using the ECAT-III geometry of UCLA.[6]

II. Bayesian reconstruction with an entropy prior.

A. Fundamental considerations.

There are two fundamentally different ways to solve statistical problems: the classical and the Bayesian approaches. In the classical approach, a set of data generated in accordance with some unknown probability law will only use the given data, without making any assumption about the unknown law. In the Bayesian approach, the use of any reasonable prior knowledge about that law is permitted. When the data are complete and of good quality, the *a priori* knowledge may be useless, but when the data are noisy, sparse and/or incomplete, the *a priori* knowledge can carry as much weight as the data. Frieden discusses the two approaches in [7]. We shall review here the fundamental basis of Bayesian statistics in the context of image reconstruction.

If we denote P(image | data) the conditional probability that the image be true given the data, Bayes' theorem gives the desired P(image | data) from the usually computable P(data | image) and from the probability of the image, P(image), with a normalization constant P(data). The P(image) is called "prior probability" because it is known (in some form) prior to obtaining the data and P(image | data) is called the "a posteriori" probability because it is what we hope to obtain after considering all facts. Bayes's theorem is:

$$P(\text{image} \mid \text{data}) = \frac{P(\text{data} \mid \text{image})\, P(\text{image})}{P(\text{data})} \tag{1}$$

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:

$$\mathbf{f}\ \mathbf{a} + \mathbf{n} = \mathbf{p} \tag{2}$$

where **a** represents the object to be computed, **p** is the given data or projections of the object, **n** is the noise inherent in 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 points in the projections, the discrete version of the imaging equation is:

$$\sum_{i=1}^{B} f_{ji}\, a_i + n_j = p_j \qquad j = 1,\ldots,D \tag{3}$$

in which a_i and p_j are non-negative vectors ($a_i \geq 0$, $i=1,\ldots,B$; $p_j \geq 0$, $j=1,\ldots,D$) and f_{ji} are the elements of a (D x B) matrix that satisfy the normalization condition

$$\sum_{j=1}^{D} f_{ji} = 1 \qquad i=1,\ldots,B\,. \tag{4}$$

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

$$P(\mathbf{a} \mid \mathbf{p}) = P(\mathbf{p} \mid \mathbf{a})\ P(\mathbf{a}) / P(\mathbf{p}). \tag{5}$$

The most probable image, given the data **p**, is obtained by maximizing $P(\mathbf{a} \mid \mathbf{p})$, or the product $P(\mathbf{p} \mid \mathbf{a})\ P(\mathbf{a})$, since $P(\mathbf{p})$ is a constant.

There have been a number of relatively recent applications of Bayesian reconstruction to medical imaging. Without claiming completeness, we shall describe the kinds of prior information that several workers have used in the formulation of the problem. Geman and McClure [8,9] have applied Gibbs energy functions as prior distributions in Single Photon Emission Tomography (SPET). These functions penalize the solution if certain smoothness conditions are not met, except at edges. Levitan and Herman [10] have used a Gaussian multivariate prior probability centered around the values of a smoothed image obtained by a Filtered Backprojection (FPB) reconstruction method in the general problem of Emission Tomography (ET). Hart and Liang [11] have used a

prior distribution in the problem of improving Anger camera images that is based on some actual knowledge of what the source activity shape and activity should be. They warn, however, that the use of a wrong prior can result in erroneous results. Lange, Bahn and Little have proposed in [12] the use of a gamma function prior with mean values based on some independently known parameters. Johnson [13] has been studying the problem of applying Gibbs energy functions to ET reconstructions with refinements over the work indicated above by Geman and McClure. Fundamentally, those approaches make use of some "reasonable" *a priori* information in order to prevent the image deterioration that occurs when maximizing $p(\mathbf{p} \mid \mathbf{a})$ alone in an unconstrained reconstruction, as in the MLE method. More recently, Liang and co-workers [14,15,16] have proposed uniform and non-uniform probability distributions as prior information, leading to Bayesian functions to be maximized which are very closely related to the ones presented in this paper and to which we will refer further in the following sections.

B. Entropy as a prior probability.

The role of entropy in defining the prior probability function has been in discussion for three decades.[17,18] Frieden first used the Shannon form of entropy in the context of image reconstruction.[19] Several papers (see Skilling and Gull [20] and Jaynes [21]) have argued 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 [22] have shown that, if one accepts three fundamental axioms, the prior $P(\mathbf{a})$ has to be a measure of the entropy of the image. As discussed by Skilling and Bryan [31], the practical interest of using the entropy is that the resulting reconstruction "has minimum configurational information consistent with the data, 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 entropy is not defined for negative values.

In this paper we use entropy to define the prior probability $p(\mathbf{a})$. We first describe entropy in terms of multiplicity in the way proposed by Frieden in [19], called the random-grain model by

Andrews and Hunt [23], although we will not limit our interpretation of the entropy expression and of its adjustable parameters only to that model.

Suppose that N is the total energy (usually the 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. Then the values $N/\Delta a$ and $a_i/\Delta a$, $i=1,\ldots,B$ are dimensionless numbers, with Δa having the units of radiance. Assuming a complete lack of prior information regarding the statistical makeup of the object, we make the simple but non-trivial 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 that case, the number of ways that an object **a** can occur is given in [24] as

$$W(a) = \frac{(N/\Delta a)\,!}{\prod_{i=1}^{B} (a_i/\Delta a)\,!} \quad . \tag{6}$$

We take the prior probability P(**a**) of the object to be proportional to its multiplicity, as given by (6). Thus, we have

$$P(\mathbf{a}) \propto W(\mathbf{a}). \tag{7}$$

It is questionable whether the multiplicity based prior (7) is the correct one for image reconstruction. In particular, the interpretation of Δa in terms of intensity jumps is open to question. The existing literature appears to reflect a controversy regarding this point. An alternative interpretation, used by Gull and Skilling in [37],will be described in Section 3E. In that work, Δa is a Lagrange multiplier calculated to obtain a feasible reconstructed image. An example will be given that supports our preference for the latter interpretation. It should be pointed out that, in spite of the

large volume of literature justifying the choice of entropy as a prior distribution, the full proof of why it works is still an open problem. Trussell [38] has found an interesting relationship between the use of entropy as a prior in a Maximum a Posteriori (MAP) reconstruction and a simple Maximum Entropy reconstruction with constraints, for the Gaussian noise case. Hanson [39] has also found a strong similarity between the behavior of a MAP reconstruction with entropy prior and a MAP reconstruction with Gaussian prior. Those similarities may explain why entropy works as a prior distribution for, at least, some specific cases.

C. The likelihood.

The conditional probability $P(\mathbf{p} \mid \mathbf{a})$ describes the noise in the projections and its possible object dependence. It is fully specified in the problem by the likelihood function. In ET, where the detectors count energetic quanta, the Poisson distribution defines the statistical characteristics of the data since radioactive disintegration obeys that statistical distribution. Following Frieden and Wells [25], we use the concept of "data increments" Δp_j as the smallest received intensity increment needed to trigger one count in detector j. Frieden and Wells used a constant value Δp, independent of the detector. We have chosen to make the increments detector dependent because in ET there are γ-ray absorption and detector gain corrections that are customarily applied to the data before reconstruction. These corrections are different for each detector pair or "tube" and correspond to the data increments Δp_j: suppose that the attenuation and gain correction factor for a particular detector pair in Positron Emission Tomography (PET) is 6.0, i.e., for every six emitted γ-pairs in directions that could be detected by the detector ring from one pixel, only one, on the average, will result in one count in the detector system. The rest are absorbed or scattered by the medium on their way to the detector. The value of Δp_j has to be set to 6.0 in this example. In all the work that follows, we define p_j as a fully corrected data set for absorption and detector normalization.

The conditional probability is then:

$$P(p \mid a) = \prod_{j=1}^{D} \exp(-h'_j) \quad [\, h'^{\,p_j/\Delta p_j}_j / (p_j/\Delta p_j)\,!\,] \tag{8}$$

where

$$h'_j = \left(\sum_{i=1}^{B} f_{ji}\, a_i \right) / \Delta p_j, \qquad j = 1,\ldots,D \tag{9}$$

is a modified projection. The data increments Δp_j play an interesting role in (8) and (9). In ET, in general, when the data are corrected multiplicatively for absorption and detector gain variations, the Poisson nature of the data is destroyed. Dividing the p_j by Δp_j in (8) restores the Poisson nature of the data during the reconstruction process. For consistency, Δp_j must also appear in the modified projection (9). If the corrections are applied directly to the matrix elements f_{ji}, all Δp_j could be set to unity.

D. The Maximum A Posteriori Probability

The Maximum a Posteriori (MAP) probability will be obtained, as indicated above, by maximizing the product $P(\mathbf{a} \mid \mathbf{p}) = P(\mathbf{p} \mid \mathbf{a})\, P(\mathbf{a})$, or equivalently, the logarithm of that probability. By using Stirling's approximation and with a constraint for the conservation of number of counts, we obtain, from (6) and (8) the Bayesian function to be maximized:

$$BY = -\sum_{i=1}^{B} (a_i/\Delta a) \log(a_i/\Delta a) + \sum_{j=1}^{D} [\, -h'_j + (p_j/\Delta p_j) \log(h'_j)] - \mu \left(\sum_{i=1}^{B} a_i - N \right) \tag{10}$$

where μ is a Lagrange multiplier for conservation of counts. It is to be noted that the relative weights of the two terms of (10) are regulated by parameter Δa, since we consider the values Δp_j to be specified by the nature of the absorption and detector gain corrections. Expression (10) is equivalent to the uniform probability Bayesian function of Liang et al [16], except that the data increments Δp_j have not been considered by those authors. When $\Delta p_j = 1$ for all j, parameter Δa becomes the equivalent of the arbitrary adjustable parameter ξ in [16].

III. The Successive Substitutions algorithm.

A. General Recursive Relation.

The non-linear nature of the reconstruction problem described by (10) suggests an iterative algorithm for the solution. An Expectation-Maximization (EM) algorithm based on the work of Dempster, Laird and Rubin [26] could be devised, as Shepp and Vardi did for their MLE solution in [27] and Liang et al have done in [16]. We have found, however, that the method of successive substitutions described by Hildebrand [28] and used by Meinel [29] for obtaining recursive formulas for a variety of linear and non-linear restoration problems affords us greater flexibility and results in rapidly converging algorithms for the tomographic image reconstruction problem.

The technique is the following: given a series of equations in the unknowns $a_1,\ldots, a_B$, that can be written in the form

$$a_i = K\,F(\{a_m\}) \qquad i = 1,\ldots,B \tag{11}$$

where F is some function, $\{a_m\}$ is the complete set of variables $a_1,\ldots, a_B$, and K is a normalization constant, then (11) can be transformed into a recursive relation by

$$a_i^{k+1} = K\,F(\{a_m^k\}) \qquad i = 1,\ldots B \tag{12}$$

where k denotes the iteration number. Each of the new values of a_i^{k+1} is calculated from all the known B values of a_m^k from the previous iteration and the complete set $\{a_m^k\}$ is updated at once.

The constant K is obtained by requiring conservation of energy: if N is the total number of counts in the object,

$$\sum_{i=1}^{B} a_i^{k+1} = \sum_{i=1}^{B} K\, F(\{a_m^k\}) = N. \tag{13}$$

The convergence of the iterative scheme defined by (12) is by no means assured for all problems that can be written in the form (11). Hildebrand discusses in [28] conditions for convergence for single or sets of two nonlinear equations, but indicates that for sets of three or more equations the results are not generalizable. Isaacson and Keller show in [30] that there are at least some conditions under which the reformulated general form of the solution

$$a_i^{k+1} = (1 - \alpha)\, a_i^k + \alpha\, K\, F(\{a_m^k\}) \quad i = 1,\ldots,B \tag{14}$$

where α is an acceleration or relaxation constant, will converge for some choice of that parameter α. An analysis of the conditions for convergence involves comparisons of the partial derivatives of the right hand side of (11) with respect to the variables a_i and, for the large size problems that we are dealing with, the analysis does not appear promising. As we shall show below, however, we have found the method to lead to converging solutions for the three phantoms studied, to some extent typical of ET image reconstructions.

B. The preliminary solution.

The maximization of (10) using the method of (14) has been reported earlier by Nunez and Llacer.[4] Since the new algorithm reported in this paper is based on the earlier solution, we shall derive the earlier version first.

In order to obtain the maximum of (10), we first set $\frac{\partial BY}{\partial a_i} = 0$ for i=1,...,B, as follows

$$\frac{\partial BY}{\partial a_i} = -(1/\Delta a)\log(a_i/\Delta a) - 1/\Delta a + \sum_{j=1}^{D}\left\{ f_{ji}(1/\Delta p_j)\left[(p_j / \sum_{i=1}^{B} f_{ji}\, a_i) - 1 \right]\right\} - \mu = 0 \qquad (15)$$

Solving for a_i from the first term, we obtain

$$a_i = K_0 \exp\left\{ \Delta a \sum_{j=1}^{D}\left(f_{ji}(1/\Delta p_j)\left[(p_j / \sum_{i=1}^{B} f_{ji}\, a_i) - 1 \right]\right)\right\} \qquad (16)$$

with the constant K_0 given by

$$K_0 = \Delta a \exp(-1 - \Delta a\, \mu). \qquad (17)$$

Equation (16) is of type (12) and we can, therefore, apply (14) to obtain a recursive relation. The result is then

$$a_i^{k+1} = (1 - \alpha)\, a_i^k + \alpha\, K_0 \exp\left\{ \Delta a \sum_{j=1}^{D} \left(f_{ji}\,(1/\Delta p_j) \left[\left(p_j \Big/ \sum_{i=1}^{B} f_{ji}\, a_i^k \right) - 1 \right] \right) \right\} \tag{18}$$

The normalization constant K_0 can be computed from (13) at the end of each iteration and is equivalent to determining the Lagrange multiplier μ of (10), as seen from relation (17).

The use of the algorithm of (18) has yielded excellent reconstructions for mathematical and real (Hoffman) brain phantoms. Moreover, those reconstructions were feasible if the initial choice of Δa was obtained from our knowledge of the number of activity levels that existed in the object source. Nevertheless, the presence of the exponential in (18) requires the use of only a small fraction of the computed correction at each iteration, in order to maintain convergence. The fraction is controlled by parameter α. In addition, in the case of large number of counts in the data, the magnitude of the exponents that have to be dealt with cannot be handled even in double precision floating point arithmetic. For those reasons, we decided to explore other possibilities afforded by the successive substitution method.

C. The new FMAPE recursive relation.

We shall now derive the new algorithm, which we call FMAPE, for Fast Maximum A Posteriori with Entropy. We start by setting the partial derivatives of BY with respect to a_i's equal to zero, as in (15). Then, instead of solving for a_i from the first term which would lead to an exponential solution, we first separate the constant terms to one side of the equation, then multiply by Δa, then add a constant C to both sides of the equation, raise both sides to the power n and finally multiply both sides by a_i. After some standard algebraic manipulations, the result is

$$a_i \left\{ 1 + \Delta a\, \mu - \log(\Delta a) + C \right\}^n =$$

$$a_i \left\{ \Delta a \sum_{j=1}^{D} \left(f_{ji}\, (1/\Delta p_j) \left[(p_j / \sum_{i=1}^{B} f_{ji}\, a_i) - 1 \right] \right) - \log(a_i) + C \right\}^n \qquad (19)$$

All the terms on the left hand side except a_i are constants that can be replaced by a single constant K_1^{-1}. The result is then an equation of type (12) that leads to the recursive relation

$$a_i^{k+1} = K_1\, a_i^k \left\{ \Delta a \sum_{j=1}^{D} \left(f_{ji}\, (1/\Delta p_j) \left[(p_j / \sum_{i=1}^{B} f_{ji}\, a_i^k) - 1 \right] \right) - \log(a_i^k) + C \right\}^n \quad i = 1,\ldots,B$$

(20)

with K_1 calculated at each iteration to conserve the number of counts. Equation (20) is the main result of this paper. Obtaining K_1 is now equivalent to calculating the Lagrange multiplier μ from

$$K_1 = 1 / \left\{ 1 + \Delta a\, \mu - \log(\Delta a) + C \right\}^n \qquad (21)$$

This algorithm does not involve exponentials and our experience shows that, for values $1 \leq n \leq 3$, we obtain convergence by using the simple formulation of (12) which does not require the use of the parameter α. There are two arbitrary constants, n and C, whose meaning will be discussed below.

It is interesting to note that the FMAPE solution contains the MLE as a special case in the limit when $\Delta a \rightarrow \infty$. This can be seen by taking the limit

$$\lim_{\Delta a \rightarrow \infty} \sum_{i=1}^{B} (a_i/\Delta a) \log(a_i/\Delta a) = 0 \tag{22}$$

in (10). What is left is the likelihood function to be maximized for the MLE solution, along with a requirement for conservation of number of counts. On the other hand, when Δa is very small, the first summation of (10) becomes dominant and the reconstruction will become independent of the projection data, yielding flat images. Thus, the FMAPE includes solutions that range from uniform grey images to the very noisy MLE images at convergence, controllable by the value of the parameter Δa.

The term -1 in the square bracket part of (20) can be removed by a modification of the definition of K_1 in (21), with minor effects on the acceptable range of C and on the speed of convergence. The same could be done in (16), except that, in that case, the size of the exponents needed for the calculations can grow excessively for good accuracy.

D. Speed of convergence and positivity.

Both (18) and (20) solve the same problem of maximizing the Bayesian expression (10), although they have a different form. The constants n and C introduced in deriving (20) are arbitrary and the effect of changing their values will only affect the calculated values of K after each iteration (see (21)). The point of convergence of (20) is, then, solely dependent on Δa, as was the case in the solutions by (18) reported in [4]. Constants n and C serve specific functions, though.

Constant n will clearly affect the rate of convergence. One can expect a range of values of n over which the iterative process is stable. We have empirically observed that for the first of the

data sets to be discussed below and for $n > 1$, the convergence rate improves by a factor of approximately n with respect to the rate for $n = 1$ and that the algorithm is stable up to $n \approx 3$. For $n > 3$, the instability can be corrected by introducing the constant α in the context of the formulation of (14), although no net gain in convergence speed appears to exist in the more complex procedure.

Constant C is introduced to insure the positivity of the solution. The presence of the two negative terms (-1 and $-\log(a_i^k)$) in the right hand side of (20) makes the positivity of the solution not automatically guaranteed, unless an appropriate constant is introduced. Positivity is not only desired in the solution, but it is necessary in order to compute the log terms in (20). As indicated above, different choices of C will not affect the convergence point. In practice, we find that changing C can affect the speed of convergence substantially. Increasing C by a factor of 2 over the approximate smallest value necessary to maintain non-negativity results in slowing down convergence by approximately that same factor. The use of $C = \Delta a$ has usually resulted in adequate convergence speed and non-negativity of the pixels. The choice of too small a value of C results in a computer run time error when trying to find the logarithm of a negative number.

E. The parameter Δa.

The parameter Δa has been described in Section II-B in terms of the number of counts in the finest known intensity jumps that are possible in the object. From the theory of Frieden, it should be set to the maximum common divisor of the number of counts in the pixels of the image source. This knowledge is available in reconstruction experiments with phantoms, but will not generally be available in real medical measurements. The Frieden interpretation of Δa poses an additional problem, however: Consider a phantom with regions of 25% and 100% radioisotope activity levels. The finest intensity jump is 25% and Δa should be set to that fraction of the number of counts that are detected from the pixels in the 100% region. Next, consider the addition of a small lesion with 17% activity in the phantom, for example. Although the small perturbation

small lesion with 17% activity in the phantom, for example. Although the small perturbation cannot be expected to change the conditions of convergence, the theory prescribes that Δa should now become 1% of maximum. Such a drastic change in Δa affects the reconstruction severely and, in our experience, is not desirable.

The Frieden interpretation of Δa follows from the acceptance of the expression for multiplicity (6) as the correct one for entropy in the case of image reconstruction. As indicated above and a survey of the literature amply shows, this acceptance appears difficult to justify fully. An alternative interpretation of the meaning of the adjustable parameters appears in Gull and Skilling [37], leading to a different criterion for the choice of their values: In the first term of (10), Δa appears first in the denominator of the linear factor and second in the denominator of the logarithmic factor. The first Δa acts as a weight that determines the relative influence on the solution of the entropy term with respect to the likelihood term. The second Δa changes the position of the maximum of entropy in the B dimensional space containing the a_i.[38] Thus, if we give the first Δa its new interpretation, we have to define a new nomenclature for the second one, $\Delta a'$, and consider it also an adjustable parameter. Gull and Skilling call this value the "default level" and, for the case with additional prior information on the image, it may be a variable. For the medical tomography case with no additional prior information, we should set $\Delta a'$ = constant. If we now go back to the development of the FMAPE algorithm of (20), starting from equation (10), we observe that $\Delta a'$ becomes part of the constant K_1 which is recalculated at the end of each iteration for the conservation of counts. Its choice is, therefore, immaterial and we shall not concern ourselves with it any longer in this paper. The independence of the solution from $\Delta a'$ was also noted by Gull and Skilling.[37]

According to the second interpretation of Δa, its value will be chosen so that the resulting reconstruction converges to feasible images. It will be shown below, however, that the Frieden interpretation can give a good initial estimate of the value of Δa, and in some cases, that estimate

leads directly to feasible images. The procedure that gives best results when dealing with a new class of images is to start with a value of Δa which is too low and increase it just until the iterative procedure converges to feasible images. Within a class of images, relatively small variations in the data do not affect the optimum value of Δa. In order to impose the effect of the prior smoothly, Liang et al, in [16], use a value of their adjustable parameter that is a function of the iteration number and of four other constants. We have used a fixed value of Δa with good results.

F. The choice of the initial image.

It has been proven by Shepp and Vardi in [27] that MLE reconstructions converge towards a unique solution. Therefore, the solution at convergence is independent of the initial guess used to initiate the reconstruction process. However, when the MLE is stopped before reaching convergence at the point of feasibility, the resulting image does depend on the starting point. We have described in [5] how the choice of a four quadrant "checkerboard" initial image results in feasible reconstructions in which the boundary lines between quadrants are still visible. Iterating towards convergence makes the lines disappear, but, at that point, the reconstructions have suffered strong deterioration.

In this section we shall show that, for the MLE and FMAPE iterative methods, a flat field is the only image to be used as the initial guess consistent with having no *a priori* information about what the image looks like. If we have an *a priori* estimate, that estimate can be used as a starting point, but it will clearly influence the MLE solution at the feasibility point and that of the FMAPE before full convergence. Thus, selecting an unlikely initial distribution, like a checkerboard pattern, is interesting as an experiment, but must not be used unless we know that the reconstructed image should have that pattern. Experimental results with the FMAPE method confirm the idea that using an entropy prior results in reconstructions in which no structures will be observed that are not supported by the data. In particular, even the choice of the checkerboard initial image results in an unflawed reconstruction at convergence in which the boundary lines have

disappeared. Thus, the FMAPE behaves like the MLE at convergence, but without the noise deterioration of the latter. We do not have a mathematical proof of uniqueness of the FMAPE solution at this time.

In order to understand the role of the initial guess in MLE and FMAPE reconstructions, we bring into the discussion the concept of cross-entropy. Frieden has given a discussion of the concept in [32]. When some *a priori* knowledge exists of what an image contains, that knowedge can be introduced in the formulation of a Bayesian reconstruction algorithm by means of the cross-entropy, defined as

$$H' = \sum_{i=1}^{B} \frac{a_i}{\Delta a} \log \frac{a_i}{\Delta a \, Q_i} \tag{23}$$

where Q_i represent our *a priori* image pixel information. This information would be, typically, the expected mean of the pixel values. In the absence of any other data, the maximization of the cross-entropy results in the solution

$$a_i = [\Delta a \; e^{-1}] \; Q_i \tag{24}$$

i.e., the solution is proportional to our prior knowledge. When the cross-entropy of (23) is introduced into the Bayesian formulation of the reconstruction problem instead of the entropy of (6) and (7), the function to be maximized becomes

$$BY = -\sum_{i=1}^{B} (a_i/\Delta a) \log(a_i/\Delta a\, Q_i) + \sum_{j=1}^{D} [-h'_j + (p_j/\Delta p_j) \log(h'_j)]$$

$$- \mu (\sum_{i=1}^{B} a_i - N). \qquad (25)$$

This function generalizes BY in (10) and is equivalent to the case of non-uniform prior in [16].

The maximization of (25) by the successive substitutions method, assuming all Δp_j to be identical, for simplicity, and with the definition of $\rho = \Delta a/\Delta p$, yields the recursive formula

$$a_i^{k+1} = K_2\, Q_i \exp\left\{ \rho \sum_{j=1}^{D} \left[f_{ji}\, p_j / \sum_{i=1}^{B} f_{ji}\, a_i^k \right] \right\} \qquad i = 1,...,B. \qquad (26)$$

Consider next the MLE method. If we maximize the likelihood (8) by the successive substitutions method we obtain

$$a_i^{k+1} = K_2\, a_i^k \exp\left\{ \rho \sum_{j=1}^{D} \left[f_{ji}\, p_j / \sum_{i=1}^{B} f_{ji}\, a_i^k \right] \right\} \qquad i = 1,...,B. \qquad (27)$$

This solution to the MLE problem is computationally different from the algorithm of Shepp and Vardi of [27], which was based on the EM algorithm of Demptster et al [26], but it solves the same

same problem. We observe that the only difference between (26) and (27) is that a_i replaces Q_i in front of the exponential. Thus, solving the MLE problem consists in a succession of Bayesian reconstruction steps with cross-entropy prior in which the result of the previous iteration is used as the prior Q_i values for the next iteration. Consider now the first iteration of an MLE reconstruction: Choosing a_i^0 for an MLE reconstruction is equivalent to defining the *a priori* knowledge about the image to be reconstructed. If we have no information of what the image is, we have to choose a_i^0 = constant to correspond to that lack of knowledge. Evidently, if we do have some knowledge about the image source and we want the resulting reconstruction to reflect it during the reconstruction process, in particular at the feasibility point, a_i^0 should be set to pixel values that reflect that knowledge. At convergence, the MLE will be insensitive to the initial starting point because of the uniqueness of the solution, but when using a stopping criterion before convergence, the resulting image will depend on the choice of a_i^0, which should therefore be set to a constant in the absence of prior information.

In order to analyze the role of a_i^0 in the FMAPE case, let us consider the iterative form of the preliminary solution given in (18). For the case of $\alpha = 1$ and all Δp_j being identical, (18) becomes

$$a_i^{k+1} = K_2 \exp\left\{ \rho \sum_{j=1}^{D} \left[f_{ji}\, p_j \Big/ \sum_{i=1}^{B} f_{ji}\, a_i^k \right] \right\} \qquad i = 1,\ldots,B \qquad (28)$$

which is identical to (26) if Q_i = constant for all i. This result is obvious, of course, if we look at the origin of both equations, i.e., comparing (10) with (25). The implication is that the use of the Bayesian reconstruction method of (18), or its equivalent FMAPE of (20), corresponds to having

no *a priori* knowledge about the image that we are reconstructing. For consistency, the values a_i^0 that appear in the right hand side of (18), (20) or (28) for the calculation of the first iteration, have to be a constant. We have consistently used the average number of counts per pixel as the initial guess for all our reconstructions.

G. Image post-filtering.

While investigating diverse methods of obtaining feasible images from MLE reconstructions in [5], we found that carrying out the iterative process past the point of feasiblity, when the images begin to deteriorate, can still result in feasible images if the noisy images are filtered with a Gaussian kernel of some appropriate width. In addition to being feasible, the post-filtered images present an excellent appearance and, in fact, they are some of the best images that we have obtained, judged visually. Snyder et al have shown in [34] that post-filtering of an image generated by the MLE method at convergence is equivalent to reconstructing by the use of a sieve and a resolution kernel of the same size and equal to the post-filtering kernel. Although we apply post-filtering to MLE images before full convergence, or to FMAPE images, our post-filtering operations are consistent with the idea expressed by Snyder et al that we do not want to see structures in the reconstructed image that could not have been observed by the finite size detectors. We have applied post-filtering to some of the images that will be shown in this paper with excellent visual results. In all cases when this is done, the size of the Gaussian filter kernel is indicated.

IV. Reconstruction results and discussion.

A. Mathematical brain phantom.

Figure 1 shows the source image for a phantom that mimics, in some manner, the complex structures of a brain. The darker sections inside the phantom have 25% relative activity, while the bright sections correspond to 100%. The image shown is the result of allocating 1 million (1M)

counts in a random manner to the different object structures. The projection data were obtained by taking each count in each pixel of the source image and placing it randomly in a projection angle and bin in accordance to the set of probabilities defined by a matrix f_{ji}. The matrix was obtained by the prescription of Shepp and Vardi given in [27], applied to the geometry of the ECAT-III tomograph of UCLA, as described in [6], and was also used subsequently in the reconstruction. The detector geometry used for the reconstruction was one single ring of 512 Bismuth Germanate detectors with a detector center-to-center distance of 6.05 mm. The reconstructions are for a 128 x 128 pixel array and a pixel size of 2.01667 mm on the side.

The progress of the reconstruction was monitored by observing the χ^2 / D statistic, described in [3] as a weak causality test. The function

$$\chi^2 / D = \frac{1}{D} \sum_{j=1}^{D} \frac{(p_j/\Delta p_j - h'_j)^2}{h'_j} \tag{29}$$

should yield a value very close to 1.0 for a feasible reconstruction with Poisson distributed data. As discussed in [31] by Skilling and Bryan, the upper limit for χ^2 at 99% confidence level should be set at $D + 3.29\, D^{1/2}$. A lower limit for a "two tailed" test can then be set at the symmetric point about D. In all the tests that we will report in this paper, D has included all the projection data with $p_j > 0$. In the case of the mathematical brain phantom, D had a value D = 17347, so that the acceptance range for feasibility is $0.975 < \chi^2 / D < 1.025$. In all the cases with computer generated phantoms we have found the application of test (29) to be equivalent to applying the more demanding hypothesis test of Veklerov and Llacer (VL) reported in [1].

Figure 2 compares the values of the χ^2 / D statistic as a function of iteration number for an MLE reconstruction (curve a), and for the FMAPE reconstructions with parameter $\Delta a = 10, 25, 50$ and 100 (curves b, c, d and e, respectively). The region of feasibility is entered by the MLE reconstruction at approximately iteration 28, in agreement with previous work with the VL test reported in [1,2,5]. Reconstructions with $\Delta a = 10$ and 25 result in images that have not developed their full contrast at convergence. A value of $\Delta a = 50$ yields a feasible image and the choice of $\Delta a = 100$ converges to a non-feasible image, characterized by excessive noise in the regions of high activity. All the reconstructions were carried out with n = 1, i.e., not accelerated. In this particular case, the optimum choice of $\Delta a = 50$ could have been obtained from the Frieden interpretation of that parameter: the pixels with highest intensity in the source image have approximately 200 counts and the smallest activity jumps are 25%, so that a value of $\Delta a = 50$ would be indicated. This method of estimating Δa is, however, not reliable.

Figure 3 shows the effect on the χ^2/D statistic of accelerating the reconstructions by changing the value of n. Curves a, b and c correspond to n=1, 2 and 3, respectively. In all cases Δa was set to 50.

Figure 4a shows the MLE reconstruction results near the entrance of the feasibility region, at iteration 30. Figures 4b and 4c show the FMAPE results at convergence with $\Delta a = 10$ and 50, respectively, both for n =1 (not accelerated). Figure 4d shows the results with n=3, indistinguishable from Fig. 4c, and Fig. 4e shows the results of post-filtering the image of Fig. 4d with a Gaussian kernel of $\sigma = 0.6$ pixels. The resulting image is still feasible.

Several profiles of pixel values for feasible images have been examined in detail. The effects of accelerating the reconstruction by the choice of parameter n results in images that do not differ by more than 1% at convergence. Also, stopping the iterative procedure in the FMAPE as soon as the feasible region is entered results in a few per cent differences in pixel values with

respect to full convergence. No significant features are lost in the process. A comparison between MLE at stopping point and FMAPE at convergence shows some differences in intensities, with the FMAPE showing generally slightly higher contrast.

B. Mathematical liver and heart phantom.

Figure 5 shows a computer simulated phantom that has some relationship to a human liver and heart section. Snyder et al have used a similar phantom for their studies of properties of the MLE and regularization of the reconstruction process with sieves in [33,34]. The relative activity levels are shown in Fig. 5. There is a background of 1% and a small region of 5% activity near the large liver section which we expected would be hard to reconstruct accurately.

The reconstruction experiments have been carried out with a source image containing a total of 1 million counts. We have found that the region $30 \leq \Delta a \leq 50$ is the one of interest for the goal of obtaining feasible reconstructions, values not obtainable by the Frieden interpretation of Δa. The values of the χ^2 / D statistic for reconstructions with $\Delta a = 40$, $n = 3$, $C = 50$ are shown in Fig. 6. Figure 7, curves a, b and c, show profiles including the small region of 5% activity for the source image and the reconstructions at iterations 10 and 20, respectively. In spite of the proximity of the small 5% region to the high activity region, it has been reconstructed reasonably well, with an average activity of approximately 3.1%. Better reconstructed activity level in that 5% region was achieved with a "weighted" MLE method discussed in [35].

C. Hoffman brain phantom with real ECAT-III data.

Figure 8 shows an FMAPE reconstruction of the Hoffman brain phantom for a data set with 30 million counts. With the high number of counts, that reconstruction is virtually identical to a filtered backprojection result with the Shepp-Logan filter with a frequency cutoff near its maximum, selected to show high sharpness and acceptably low overall noise. Figure 8 is shown as a possible standard that reconstructions from a data set of 1 million counts should try to

approach. The ratio of activities in the hot vs. cold regions was set to approximately 4.5 : 1. The data were obtained from one ring of the UCLA ECAT-III tomograph, as described in [6]. For the 1 million count data set, the maximum number of counts in one pixel, obtained from other reconstructions, is approximately 3000 counts. If we consider the number of levels to be approximately 5, a value of $\Delta a = 600$ could be used as an initial setting. We found, in fact, that this value is adequate for obtaining feasible images.

For this case with real data, two different kinds of corrections have to be applied to the data set before reconstruction:

1. - Random coincidences have to be subtracted. During the ECAT-III measurement, one delayed coincidence gate is opened every time that a coincidence event is detected by the logic circuits. The latter events are called "prompt" coincidences. The prompt data contain true coincidences plus randoms, while the delayed coincidences contain a sampling of the random coincidences with the same mean values as a function of position as those of the prompt measurement. Standard practice consists in subtracting the delayed from the prompt data, resulting in a first order correction for random coincidences. Only the noise due to the differences between the prompt and delayed samplings remains in the data to be reconstructed. Since the prompt data are known to be Poisson distributed (in the absence of pile-up problems in the instrumentation electronics) and the delayed data are also Poisson distributed, but the difference is not, we have carried out the reconstructions of one set of 1 million counts by two methods:

a. Reconstruct the prompt and the delayed data separately and subtract the resulting images,

b. Subtract first and reconstruct the difference data.

For the data set from the Hoffman brain phantom, with only approximately 6.5% of random coincidences, the images resulting from the two methods of correction are indistinguishable from

each other. To the extent of this test, then, the FMAPE appears to be robust to non-Poisson projection data. The results that will be shown here are for case b.

2. γ-ray absorption and detector gain corrections have to be applied to the data. The product of the absorption factors along the detector coincidence paths and the relative gain of the detector pairs will be taken as the values Δp_j described in Section II.C.

The use of the χ^2 / D statistic or the VL test of [1] for feasibility fails with real data from a tomograph. Llacer and Veklerov ascertained in [3] that the problem lies with the lack of accuracy with which the matrix f_{ji} is known. They proposed a more relaxed test which has one adjustable parameter, ε, that corresponds to the uncertainty with which the matrix elements f_{ji} are known. The test, once ε has been determined empirically, signals the entering of the feasibility region unequivocally, although it is insensitive to going too far in the MLE iterative procedure. For the feasibility tests to be reported in this paper, we have used a value of $\varepsilon = 6.5\%$ which was found in [3] to signal the entering of MLE reconstruction into the feasibility region for the Hoffman brain phantom with 1 million counts.

Figure 9 shows values for the hypothesis test function H for the reconstructions of the Hoffman brain phantom, as a function of iteration number. The threshold for acceptance of an image as feasible is shown as a straight line. The reconstruction parameters were $\Delta a = 600$, $n = 2$ and $C = 450$. Figure 10a shows the reconstruction results of the FMAPE procedure at iteration 30. Figure 10b shows the results at it. 50 where our approximate feasibility criterion is fulfilled. An image with higher contrast is shown in Fig. 10c, at 100 it. Continuation of the iterative procedure with the FMAPE and the indicated parameters results in some deterioration of the image, indicating that the value of $\Delta a = 600$ may be too high for optimum reconstruction. Figure 10d shows the

results at iteration 300 and Fig. 10e shows the results of post-filtering the results of 10d with a Gaussian kernel of $\sigma = 0.6$ pixels. The latter appears visually to be the better image of the set.

V. Conclusions

From the results reported with three phantoms, it appears that the FMAPE reconstruction procedure yields images that are similar to, but somewhat different from the MLE reconstructions at the stopping point. The differences observed are, however, small and not immediately characterizable. An analysis of the resulting images in terms of bias and variances of many instances of the same source image may provide for understanding those differences.

The FMAPE has one characteristic that makes it more appealing than the standard MLE process: it converges to visually good images with excellent stability and, moreover, the resulting images can be made feasible by choice of one contrast parameter. The FMAPE can be accelerated by factors of up to three with respect to the unaccelerated form ($n = 1$), but a number of methods for acceleration of the MLE have already been found, including an equivalent form of (27) that we have investigated, in which the exponential function disappears and an accelerating exponent is introduced, as in the development of the FMAPE. We find the process of reconstruction from real data from the Hoffman brain phantom slower in converging than in the case of computer generated data, as seen by comparing Fig. 2 for $\Delta a = 50$ with Fig. 9. This may be due to the nature of the calculated matrix elements f_{ji} used in the reconstruction, which are less accurate a represen-tation of the true transition matrix elements of the tomograph than in the computer simulation case. It will be interesting to explore the use of a Monte Carlo calculated matrix that represents the ECAT-III tomograph more closely, as was done in [36].

A possible drawback of the FMAPE is the somewhat lengthened computation procedure by having to calculate one logarithm and one power for each pixel in each iteration. The additional

A possible drawback of the FMAPE is the somewhat lengthened computation procedure by having to calculate one logarithm and one power for each pixel in each iteration. The additional length of time is not particularly significant when compared to the time needed for a projection and backprojection in each iteration.

Both the FMAPE and the MLE images stopped at feasibility, or past feasibility with post-filtering, have very low backgrounds and low noise in the low activity areas when compared to filtered backprojection (FBP) images. They exhibit contrast and sharpness at least as good as FBP images with filters and bandwidths which have been selected for a good compromise between sharpness and overall noise. Ascertaining the diagnostic usefulness of the images obtained from MLE and FMAPE awaits, however, the results of a thorough Receiver Operating Characteristics (ROC) study with real data and observers (see, for example, [40]). We feel that the development of the present Bayesian method of solution and the acquired understanding of the MLE method and their relationship to feasible images makes an ROC study a natural next step in the search for improved reconstruction methods for emission tomography.

VI. Acknowledgments

The authors would like to thank E. Veklerov for frequent discussions during the course of the work reported and the inspiring lectures and discussions with John Skilling and Stephen Gull in the June 1988 Joint Conference of the AMS-IMS-SIAM, who set us in the path to Bayesian reconstructions and entropy. This work has been supported, in part, by a grant from the National Institutes of Health, CA-39501 and by the U.S. Department of Energy under Contract No. DE-AC03-76SF00098. Dr. Nunez was supported, in part, by the Center of Catalan Studies, Generalitat de Catalunya, Spain, and University of California, Berkeley.

VII. References.

[1] E. Veklerov and J. Llacer, "Stopping Rule for the MLE Algorithm Based on Statistical Hypothesis Testing," IEEE Trans. Med. Imaging, MI-6, No. 4, 313-319 (1987).

[2] J. Llacer and E. Veklerov, "The Use of a Stopping Rule in Iterative Image Reconstruction," Joint Conf. of the AMS-IMS, SIAM on Spatial Statistics and Imaging, Brunswick, Maine, (1988); to be publ. in Conf. Proc. (1989).

[3] J. Llacer and E. Veklerov, "Feasible Images and Practical Stopping Rules in Iterative Image Reconstruction," IEEE Trans. Med. Imaging, Vol. 8, No. 2, 186-193 (1989).

[4] J. Nunez and J. Llacer, "Maximum Entropy and the Concept of Feasibility in Tomographic Image Reconstruction," Proc. of the SPIE Conf. on Medical Imaging,Vol. 1090, 359-372, (1989).

[5] J. Llacer, E. Veklerov and J. Nunez, "The Concept of Causality (Feasibility) in Image Reconstruction," pres. NATO Advanced Study Institute, Povoa, Portugal, (Sept. 1988); to be publ. in Institute Proc. (1989).

[6] E.J. Hoffman, A.R.Ricci, L.M.A.M. van der Stee and M.E. Phelps, "ECAT-III - Basic Design Considerations," IEEE Trans. Nucl. Sci., NS-30, No. 1, 729-733 (1983).

[7] B. R. Frieden, *Probability, Statistical Optics and Data Testing*, Springer Verlag, Berlin-Heidelberg (1983).

[8] S. Geman and D.E. McClure, "Bayesian image analysis: An Application to single photon emission tomography," Proc. Statist. Comput. Sect., Washington, DC, Americal Statistical Assoc., 12-18 (1985).

[9] S. Geman and D.E. McClure, "Statistical Methods for Tomographic Image Reconstruction," Proc. 46th Sect. of the ISI, Bull. of the ISI, 52 (1987).

[10] E. Levitan and G.T. Herman, "A Maximum A Posteriori Probability Expectation Maximization Algorithm for Image Reconstruction in Emission Tomography," IEEE Trans. Med. Imaging, MI-6, No. 3, 185-192 (1987).

[11] H. Hart and Z. Liang, "Bayesian Image Processing in Two Dimensions," IEEE Trans. Med. Imaging, MI-6, No. 3, 201-208 (1987).

[12] K. Lange, M. Bahn and R. Little, "A Theoretical Study of Some Maximum Likelihood Algorithms for Emission and Transmission Tomography," IEEE Trans. Med. Imaging, MI-6, No. 3, 106-114 (1987).

[13] V. Johnson, W.A. Wong, X. Hu and C.T.Chen, "Data Augmentation Schemes Applied to Image Restoration," pres. NATO Advanced Study Institute, Povoa, Portugal, (Sept. 1988); to be publ. in Institute Proc. (1989).

[14] Z. Liang, "Statistical models of a priori information for image processing," SPIE 914, Medical Imaging II, 677-683 (1988).

[15] Z. Liang, R. Jaszczak and H. Hart, "Study and performance evaluation of statistical methods in image processing," Comput. Biol. Med. 18, No. 6, 395-408 (1988).

[16] Z. Liang, R. Jaszczak and K. Greer, "On Bayesian image reconstruction from projections: Uniform and non-uniform a priori source information," IEEE. Trans. Med. Imaging,Vol. 8, No. 3, 227-235 (1989).

[17] E. T. Jaynes, Phys. Rev., 106, 620-630, also 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).

[18] J.T. Jaynes, "On the Rationale of Maximum Entropy Methods," Proc. IEEE, 70, 939-952 (1982).

[19] B. R. Frieden, "Restoring with Maximum Likelihood and Maximum Entropy," J. Opt. Soc. Amer., 62, 511-518 (1972).

[20] J. Skilling and S.F.Gull, "The Entropy of an Image," SIAM Amer. Math. Soc. Proc., 14, 167-189 (1984).

[21] E.T. Jaynes, "Monkeys, Kangaroos and N," in Maximum Entropy and Bayesian Methods in Applied Statistics, Proc. 4th Maximum Entropy Workshop, Ed. J.H. Justice, U. of Calgary, 26-58 (1984).

[22] J. Skilling and S.F. Gull, pres. AMS-IMS SIAM Joint Conf. on Spatial Statistics and Imaging, Brunswick, Maine (1988); publ. in progress.

[23] H.C. Andrews and B.R. Hunt, *Digital Image Restoration,* Prentice Hall, Inc. (1977).

[24] E. T. Jaynes, "Prior Probabilities," IEEE Trans. Syst. Sci. Cybern., SSC-4, 227-241 (1968).

[25] B.R. Frieden and D.C. Wells, "Restoring with maximum entropy. III: Poisson Sources and backgrounds," J. Opt. Soc. Am., 68, No. 1, 93-103 (1978).

[26] A.D. Dempster, N.M. Laird and D.B. Rubin, "Maximum Likelihood from Incomplete Data Via the EM Algorithm," J. Roy. Stat. Soc., B39, 1-37 (1977).

[27] L.A. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography," IEEE Trans. Med. Imaging, MI-1, No. 2, 113-121 (1982).

[28] F. B.Hildebrand, *Introduction to Numerical Analysis,* 2nd. Ed., McGraw-Hill, N.Y. (1974).

[29] E.S. Meinel, "Origins of linear and nonlinear recursive restoration algorithms," J. Opt. Soc. Am., 3, No. 6, 787-799 (1986).

[30] E. Isaacson and H.B. Keller, *Analysis of Numerical Methods,* John Wiley and Sons (1966).

[31] J. Skilling and R.K. Bryan, "Maximum Entropy reconstruction: general algorithm," Monthly Notices of the Royal Astronomical Soc., 211, 111-124 (1984).

[32] B.R.Frieden, "Unified theory for estimating frequency-of-occurrence laws and optical objects," J. Opt. Soc. Am., 68, No. 1, 93-103 (1978).

[33] D.L. Snyder and M.I.Miller, "The use of sieves to stabilize images produced with the EM algorithm for emission tomography," IEEE Trans. Nucl. Sci., NS-32, 3864-3872 (1985).

[34] D.L.Snyder, M.I. Miller, L.J. Thomas and D.A. Politte, "Noise and edge artifacts in Maximum-Likelihood reconstructions for emission tomography," IEEE Trans. Med. Imaging, MI-6, No. 3, 228-238 (1987).

[35] J. Llacer and E. Veklerov, "The high sensitivity of the Maximum Likelihood Estimator method of tomographic image reconstruction," Proc.Intl. Symp. on Computed Assisted Radiology (CAR'87). Springer, Heidelberg (1987), Lemke, Rhodes, Jaffee and Felix, eds.

[36] E. Veklerov, J. Llacer and E. Hoffman, "MLE reconstruction of a brain phantom using a Monte Carlo transition matrix and a statistical stopping rule," IEEE Trans. Nucl. Sci., NS-35, No. 1, 603-607 (1988).

[37] S.F.Gull and J. Skilling, "Maximum entropy method in image processing", IEE Proceedings, 131, Pt. F., No. 6, 646-659 (1984).

[38] H.J. Trussell, "The relationship between image reconstruction by the maximum a posteriori method and a maximum entropy method.", IEEE Trans. Acoust., Speech, Signal Processing, ASSP-28, 114-117 (1980).

[39] K.M. Hanson, in *Image Recovery: Theory and Applications,* Chapter 3, Academic Press, 1987.

[40]C.E. Metz, "ROC Methodology in Radiologic Imaging", Invest. Radiology, 24, No. 9, 720-733, 1986.

VIII. Figure Captions.

1. - Mathematical brain phantom. Image contains 1 million counts and originated the projection data source used in the reconstructions by assigning each count in each pixel to a projection bin by a statistical process. A matrix of transition probabilities for the ECAT-III tomograph of UCLA was used for the latter process.

2. - Comparison of the χ^2/D statistic as a function of iteration number for an MLE reconstruction of the mathematical brain phantom (curve a), and for the FMAPE reconstruction with $\Delta a = 10, 25, 50$ and 100 (curves b, c, d and e, respectively). The acceleration parameter n was set to unity, i.e., the reconstruction was not accelerated.

3. - Comparison of the χ^2/D statistic for the mathematical brain phantom as a function of iteration number when the reconstruction is accelerated by changing n from 1 to 3. In all cases $\Delta a = 50$.

4. - Results of reconstructing the mathematical brain phantom. a) MLE at iteration 30. b) and c) FMAPE at convergence (it. 50) with $\Delta a = 10$ and 50, respectively. d) FMAPE accelerated reconstruction at convergence (it. 15) with $n = 3$, $\Delta a = 50$. e) accelerated image after post-filtering with a Gaussian kernel of $\sigma = 0.6$ pixels.

5. - Perspective view of a "liver and heart" mathematical phantom showing the relative activities in the different regions.

6. - χ^2 / D values for the "liver and heart" phantom reconstruction, $\Delta a = 40$, $n = 3$, $C = 50$.

7. - Profiles through the 5% activity region of the "liver and heart" phantom for a) the source image of Fig. 5, b) FMAPE at it. 10, c) FMAPE at it. 20. Reconstruction parameters as indicated in Fig. 6. Activity level is in arbitrary units.

8. - Reconstruction of a 30 million count data set for the Hoffman brain phantom, with real data from the ECAT-III tomograph of UCLA.

9. - Values of the hypothesis test function H during the reconstruction of the Hoffman brain phantom data set with 1 million counts. $\Delta a = 600$, $n = 2$ and $C = 450$.

10. - Results of reconstructing the Hoffman brain phantom with the FMAPE algorithm, a) at iteration 30, b) at it. 50, c) at it. 100, d) at it. 300, e) at iteration 300 after post-filtering with a Gaussian kernel of $\sigma = 0.6$ pixels.

LAWRENCE BERKELEY LABORATORY
TECHNICAL INFORMATION DEPARTMENT
1 CYCLOTRON ROAD
BERKELEY, CALIFORNIA 94720

AAF331
LBL Libraries