Title
THE MAXIMUM LIKELIHOOD ESTIMATOR METHOD OF IMAGE RECONSTRUCTION: ITS FUNDAMENTAL CHARACTERISTICS AND THEIR ORIGIN

Authors
Llacer, J.
Veklerov, E.

Publication Date
1987-05-01

Lawrence Berkeley Laboratory

UNIVERSITY OF CALIFORNIA

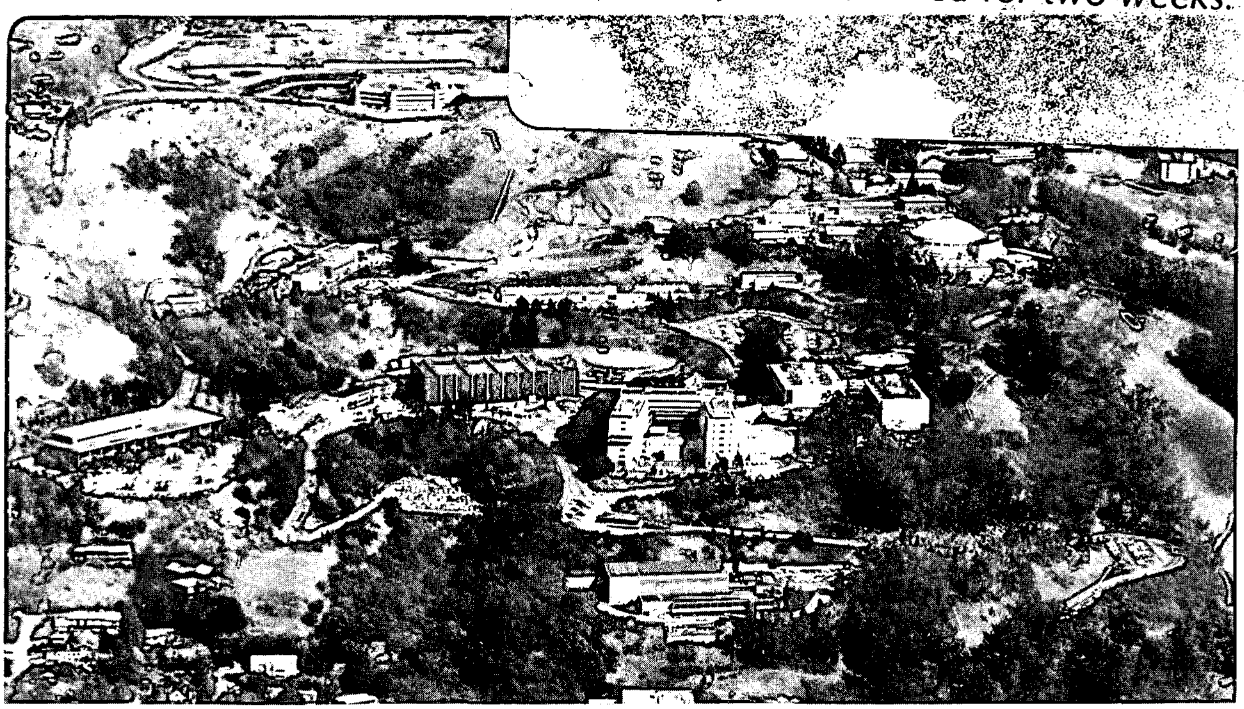
Engineering Division

Presented at the Xth Information Processing in Medical Imaging (IMPI) International Conference, Utrecht, The Netherlands, June 22–26, 1987, and to be published in the Proceedings

The Maximum Likelihood Estimator Method of Image Reconstruction: Its Fundamental Characteristics and Their Origin

J. Llacer and E. Veklerov

May 1987



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

DISCLAIMER

LBL-22417

THE MAXIMUM LIKELIHOOD ESTIMATOR METHOD OF IMAGE RECONSTRUCTION: ITS FUNDAMENTAL CHARACTERISTICS AND THEIR ORIGIN

Jorge Llacer and Eugene Veklerov

Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720 U.S.A.

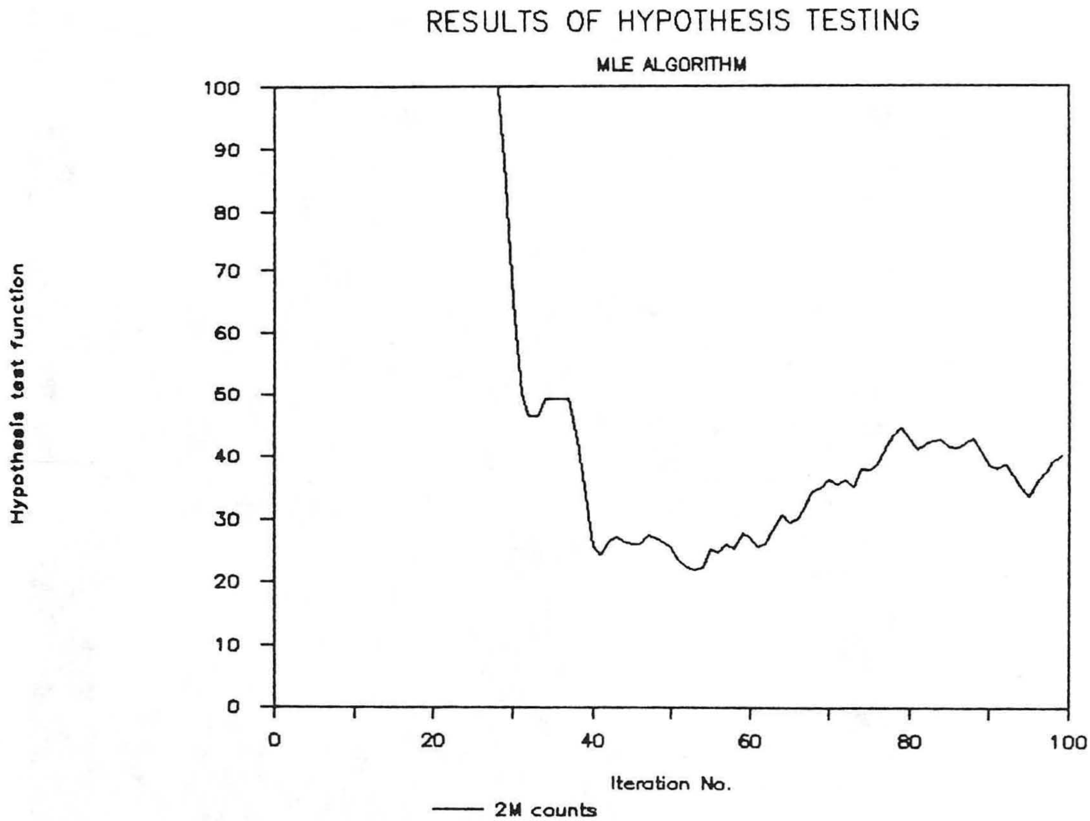
ABSTRACT

In this paper we review our recent work in characterizing the image reconstruction properties of the MLE algorithm. We have studied its convergence properties and confirmed the onset of image deterioration, which is a function of the number of counts in the source. By modulating the weight given to projection tubes with high numbers of counts with respect to those with low numbers of counts in the reconstruction process, we have confirmed that image deterioration is due to an attempt by the algorithm to match projection data tubes with high numbers of counts too closely to the iterative image projections. We have also developed a stopping rule for the algorithm that tests the hypothesis that a reconstructed image could have given the initial projection data in a manner consistent with the underlying assumption of Poisson distributed variables. The rule has been applied to two mathematically generated phantoms with success and to a third phantom with "exact" (no statistical fluctuations) projection data with results which confirm our understanding of the fundamental process of iterative image reconstruction. We conclude that the behavior of the target functions whose extrema are sought in iterative schemes is more important in the early stages of the reconstruction than in the later stages, when the extrema are being approached but the results are in contradiction with the Poisson nature of the measurement.

INTRODUCTION

The non-linear iterative Maximum Likelihood Estimator (MLE) method of image reconstruction for tomography has received increasing attention in the last several years because it promises reconstructions with very low noise and sharp edges, even from incomplete sets of projections. In the last two years, instabilities or image deterioration with increasing number of iterations have been reported in the reconstruction of smooth activity fields. This could make the method rather useless for medical applications. Snyder and Miller (1985) have proposed stabilizing the solution by filtering with sieves along with the iterative process. Levitan and Herman (1986) have used a maximum a posteriori probability expectation maximization algorithm to constrain the MLE algorithm to solutions that do not deviate more than a certain measure from a pre-determined ex-

a) backgrounds in MLE reconstructions are very clean, with very little noise magnification.
b) the meaningful matching of tubes with high counts starts to occur after the low count tubes are already rather well matched.
c) contrast and sharpness in the periphery of objects of high activity are very good in MLE reconstructions.
d) ringing at edges is usually observed in the "best" reconstructions and it is a consequence of the finite size of pixels being traversed by tubes with low and high counts (from different projection angles) in the vicinity of edges.
e) contrast inside regions of high activity is somewhat difficult to attain, as already pointed out by Tanaka (1987), since all the projections involved may have high counts.
f) accurate reconstruction of low activity regions in the neighborhood of high activity regions may be hard to obtain since it involves tubes with both high counts and low counts. It appears then that the low activity region is properly reconstructed when the high activity region is getting very well matched, with image deterioration. In the case of Region 2 of Fig. 6, the proper average activity of 0.05 was only attained past the "best" image point, but could be modulated by using the WLE target function with $s < 0$.



XBL 875-2405

Fig. 11: Hypothesis testing function for the reconstructions of the phantom of Fig. 10, with 2 million counts in the source.

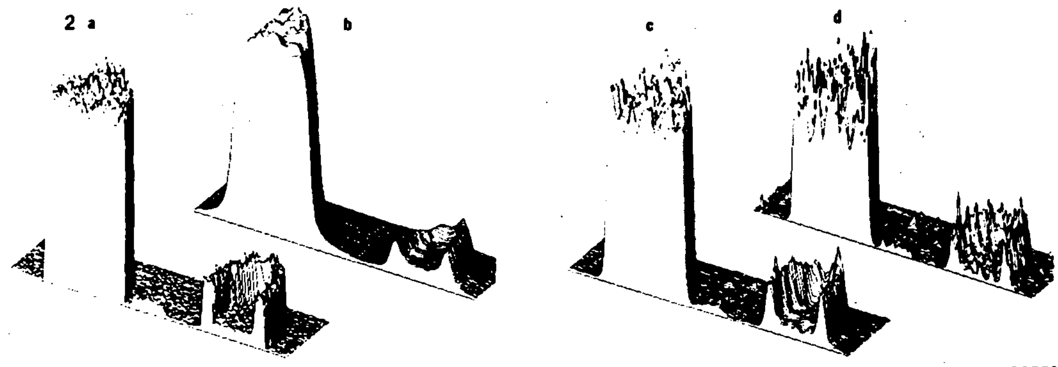


Fig. 2: a) Cut through a line in the source image of 2 million counts. b) ditto for a corresponding MLE reconstruction after 9 iterations. c) ditto after 45 iterations. d) ditto after 200 iterations.

unit area are shown in the figure. The interior of the elliptical shell has a relative activity of 0.05. Random background was simulated by a relative activity of 0.01 over the entire surface of the image. The image plane has been discretized into 128 by 128 pixels. The matrix of p(b,d) values used for projection generation and image reconstruction was obtained following the algorithm described by Shepp and Vardi (1982) and adapted to one ring of the UCLA ECAT-III, 512 detector tomograph, in Hoffman et al (1983).

For a source image with 2 million counts (2M), Fig. 2a shows a cut through a line in the source image and Figs. 2b - d show the MLE reconstructions with 9, 45 and 200 iterations, respectively. The same experiment was repeated with the source image containing 8 million (8M) and 32 million (32M) counts. The results are shown in Figs. 3 and 4.

An examination of the reconstructed images shows that:

a) reconstructions improve towards a reasonable representation of the source images up to 30 - 50 iterations;
b) the image quality remains virtually the same for a number iterations; and
c) after that point, images begin to deteriorate.

It was also observed that images from sources with a higher number of counts require more iterations to reach their best appearance and they remain in that condition for more iterations.

We also carried out an analysis of the probability that the source images could have generated their own projection data and compared the results with the corresponding probabilities that the MLE reconstructed images would have generated such data. This was done by a straightforward application of the Poisson function to the projection of the images as means and the projection data as instances of the distributions. Figure 5 shows the results of the calculations. We see that, as expected, the probability that the MLE images would have generated the projection data increases with iteration number. After all, that is what the MLE algorithm is supposed to do. We see, however, that the source image which actually generated the projection data is substantially removed from having that maximum probability. As the number of counts in the source increases, the distance to that maximum decreases.

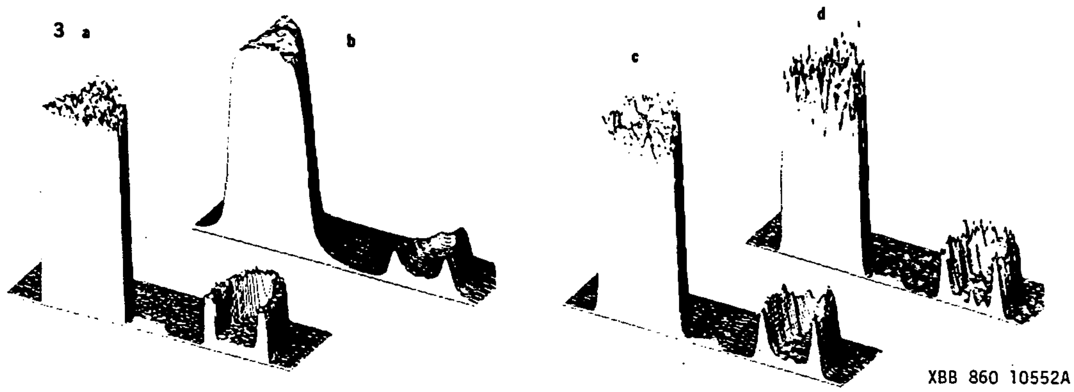


Fig. 3: a) Cut through a line in the source image of 8 million counts. b) ditto for a corresponding MLE reconstruction after 9 iterations. c) ditto after 45 iterations. d) ditto after 200 iterations.

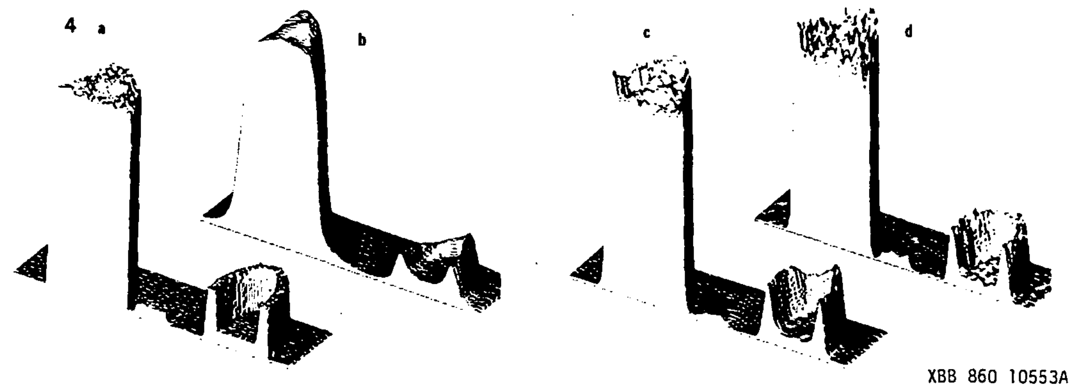


Fig. 4: a) Cut through a line in the source image of 32 million counts. b) ditto for a corresponding MLE reconstruction after 9 iterations. c) ditto after 45 iterations. d) ditto after 200 iterations.

This last set of observations gave us the first indications of the origin of the image deterioration observed in the MLE solutions: because of statistical uncertainty in the process of going from disintegrations in the pixels to detection in projection tubes, the projections present an inconsistent set of data to the matrix of p(b,d) values. As the iterative process continues beyond a certain point, the algorithm tries to find an image that reconciles those inconsistencies as well as possible in the manner precribed by the target function chosen (the likelihood function L of Eq. 1). In fact, it goes beyond the value of L corresponding to the source image.

We have obtained further confirmation of our hypothesis by developing a modified target function, which we call the Weighted Likelihood Estimator (WLE). The function that we seek to maximize is:

$$WL(\lambda) = WP(n^*|\lambda) = \prod_{d=1}^{D} \left[e^{-\lambda^*(d)} \lambda^*(d)^{n^*(d)} / n^*(d)! \right]^{s \cdot n^*(d)+t} \qquad (2)$$

With $s = 0$ and $t = 1$, the function WL is identical to the likelihood function L of Eq. 1. Keeping $t = 1$, $s > 0$ will give higher weight during the process of maximization to those tubes that have higher number of counts, while making $s < 0$ will decrease their weight. The iterative formula for the maximization of Eq. 4 is:

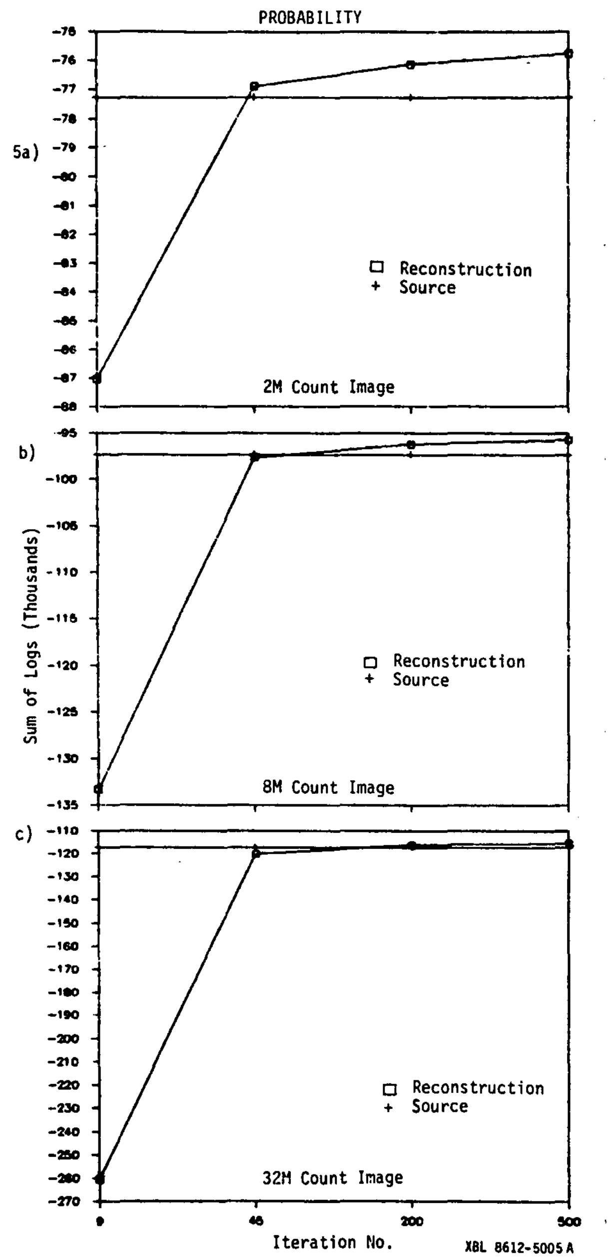
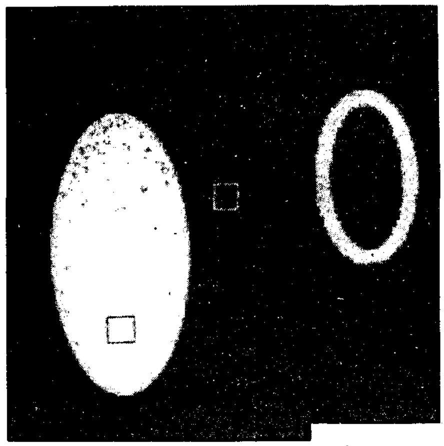


Fig. 5: Probability that the image obtained at a certain number of iterations has yielded the particular set of projection data used as input for the reconstruction; horizontal line indicates the probability for the source image that truly generated the input data. a) case with 2 million counts in the source image. b) case with 8 million counts. c) case with 32 million counts.

$$\lambda^{new}(b) = \lambda^{old}(b) \left[1 + \sum_{d=1}^{D} [s \cdot n^*(d)+t]\, p(b,d) \frac{n^*(d) - 1}{\sum_{b'=1}^{B} \lambda^{old}(b')p(b',d)} \right] \quad (3)$$

We have carried out reconstructions with the WLE target function with the source image of Fig. 1 with 2M counts, with values s = 0.0025, 0 and -0.0015. It is observed that the onset of image deterioration in regions of high activity comes early in the first case and is delayed in the last case, with respect to the s = 0 case. No substantial differences are



XBB 871-482

Fig. 6: Source distribution showing Reg. 1 and 2 for noise evaluation.

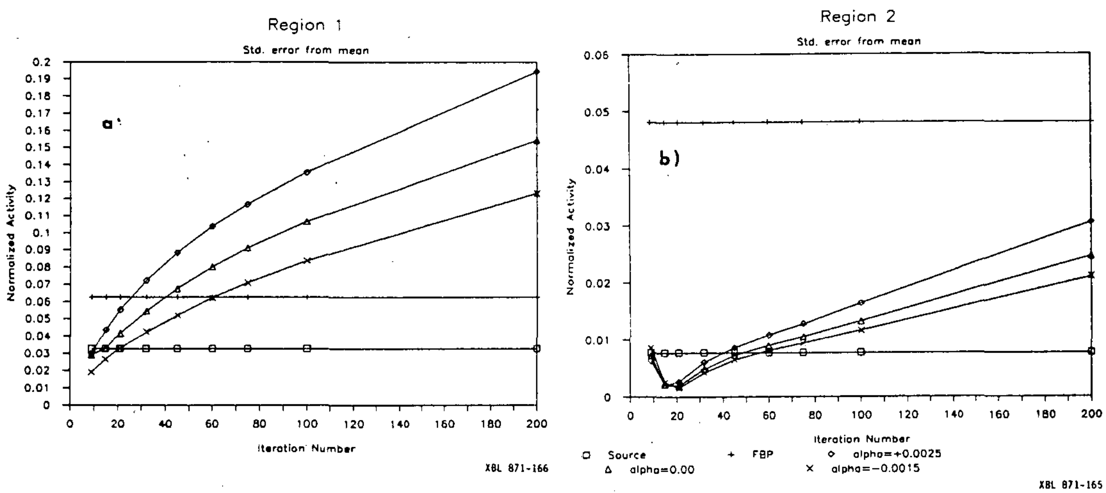


Fig. 7: Plots of standard error σ from the mean as a function of iteration number for different values of parameter s in Eqs. 2 and 3. a) Reg. 1, with high counts (1.0 relative). b) Reg. 2, with low counts (0.05).

observed in regions of low activity, which have its corresponding tubes with low counts reconciled with the projection data very early in the iterative process and suffer little modification in the later stages, seen in Llacer et al (1986). Figure 6 shows two regions (small squares) in the source image chosen for investigation, one in the region of 100% counts, the other in the region of 5% counts. Figure 7 shows plots of standard error from the mean σ, in each of two zones as a function of iteration number for the WLE reconstructions. Also shown are the standard errors for a filtered backprojection (FBP) reconstruction with the Shepp-Logan filter and for the source image. For the region with high counts, Region 1, we observe that σ is a factor of 2 higher than the source image error. For the WLE reconstructions, we see a substantial influence of the parameter s on the iteration number at which σ is equal to that of the FBP method. In the region of low counts, Region 2, the error of the FBP is 0.05, of the same magnitude as the signal, while the WLE results remain under 0.01 (near the source noise) up to iterations 40 to 60.

The observation that modulating the weight of tubes with high counts on the likelihood function changes substantially the onset of image deterioration supports the above argument that image deterioration occurs by an attempt by the MLE algorithm to match the projection tubes with high numbers of counts to the statistically deficient projection data too well. The question then arises naturally: is there a way to determine an optimum stopping point that is image and number of count independent? In the following a concise statement of the stopping rule which we have devised in Veklerov and Llacer (1987) is presented.

STOPPING RULE FOR ITERATIVE ALGORITHMS

We shall first show that the maximization of Eq. 1 and the underlying assumption that the data are Poisson-distributed become contradictory to each other beyond a certain number of iterations. We propose a quantitative criterion that allows the user to catch the moment when they are least contradictory.

It was noticed by Snyder, Lewis and Ter-Pogossian (1981) that $n^*(d)$ are independent and Poisson random variables and $\lambda^*(d)$ are their means. Therefore, the image maximizing Eq. 1 must be such that $\lambda^*(d)$ are as close to $n^*(d)$ as possible. (This is also seen from Eq. 1 directly.)

This fact leads to a paradoxical situation. The closer $\lambda^*(d)$ become to $n^*(d)$, the higher the probability that the image generates the projection data as measured by the likelihood function L. However, for Poisson distributed data, if the two get too close for all or almost all tubes d, it becomes statistically unlikely that the image could have generated the data. This paradox can be illustrated with the aid of the following simplified example. Let us assume that each tube in the initial projection data has exactly 100 counts and the algorithm is able to assign values to the image pixels so that each image projection tube has a value between 99 and 101. These values are means of Poisson distributed variables of which the 100 numbers are realizations. The condition just described would be a very good match as measured by the L criterion or by other criteria, such as the minimum least squares, for example. We assert, however, that the means are too close to the data to be consistent with the underlying assumption that the data are Poisson distributed. Indeed, now we have a situation with a large number of independent measurements each of which differs from the mean by less than 1. However, since the underlying distribution is Poisson, the standard deviation should be approximately 10. This contradiction demonstrates our point.

In order to avoid arriving at this contradictory condition during image reconstruction, we propose to test, at the end of each iteration, the hypothesis that the values of $n^*(d)$ for each $d=1,\ldots,D$ are jointly statistically valid realizations of Poisson-distributed random variables corresponding to their means $\lambda^*(d)$. In other words, we test the hypothesis that $n^*(1)$ is a realization of a Poisson variable with the mean $\lambda^*(1)$ and $n^*(2)$ is a realization of another such distribution with mean $\lambda^*(2)$ and so on. It is reasonable to require that only images passing the test (for which the hypothesis is not rejected) be declared acceptable. The development of the testing procedure is described in Veklerov and Llacer (1987), to which we refer the reader for details. The test is based on an application of Pearson's chi-square test as shown in Canavos (1984) adapted to our case. It results in:

a) A histogram of projection tubes with a number N of bins of equal probability which is flat if the projection data $n^*(d)$ are valid realizations of Poisson distributed variables with means $\lambda^*(d)$. If the histogram is relatively empty at the center, the iterative process has not yet gone far enough and maximization of the likelihood function (or another suitable target function) can continue. If bunching occurs in the central bins, the iterative process has gone too far; the projection data are too close to the projection of the image whose validity is being tested.

b) A hypothesis testing function H which is obtained from the above tube histograms by Pearson's test applied to the results of successive iterations. This function should go through a minimum with values below 36.2 and 27.2, for an example with N = 20 bins, corresponding to significance levels of 0.1 and 0.01, respectively. The significance levels are the probabilities of rejecting the hypothesis, given that it is true.

EXAMPLES OF THE APPLICATION OF THE STOPPING RULE

The above stopping rule has been applied first to the MLE reconstructions of the activity distribution of Fig. 1, as reported in Veklerov and Llacer (1987). The results of the hypothesis testing function H for 2M, 8M and 32M counts in the source image are shown in Fig. 8, curves a), b) and c), respectively. They indicate that the images at iterations near 30, 50 and 90 are "best" in terms of consistency of the images with the Poisson nature of the PET process. This result agrees with our observations about image deterioration as a function of iteration number and counts in the source image, as discussed above. Figure 9 shows a comparison of the source image, an FBP image (Shepp-Logan filter) and MLE recontruction images at 9, 32 and 200 iterations for the 2M count case. At the three count levels tested, the "best" image represents compromises between some deterioration at the regions of high counts (Region 1 of Fig. 6, for example) and the attaining of reasonable sharpness and detail in the regions of low counts (Region 2).

A second test with a source distribution containing substantial detail, mimicking a brain phantom in some fashion, has also been carried out with 2M counts. The source image and projection data have been generated by the same computational method used for the phantom of Fig. 1. Figure 10, left to right, top to bottom, shows the source image, an FBP (Shepp-Logan filter), and MLE reconstructions from 21, 32, 45 and 100 iterations, presented with 16 levels of gray. The "best" image is near iteration 45, determined by the value of the hypothesis testing function H, of Fig. 11. Figure 12 shows the tube histograms for the same set of iterations; the contents of the histogram bins is flattest at 45 iterations.

We have also carried out another experiment to further confirm the correctness of our interpretation of the tube histograms and their relationship to the consistency of the Poisson distribution. Starting with an image of the Hoffman brain phantom in Hoffman et al (1983), supplied to us by the UCLA group, we have generated an exact set of projection data by multiplying the image pixels by the elements of a p(b,d) transition matrix which describes the ECAT-III tomograph in the Shepp-Vardi approximation (1982). The projection data are exact in the sense that no statistical fluctuations have been introduced in going from the pixel counts to the projections. The multiplications have been made in double precision and all fractional counts have been kept. The reconstructions were carried out in double precision using the same transition matrix.

In reconstructing the Hoffman brain phantom from such a set of data, we would expect the MLE process to be able to proceed without image deterioration for a very large number of iterations, since there are no statistical fluctuations in the projections; i.e., the projection data are fully consistent with the transition matrix (matrix of p(b,d)'s). In fact, the projection data contain the statistical accuracy of a nearly infinite number of counts in the source image, although in reality, most of the significant tubes have counts in the order of 100. We would also expect that the tube histogram obtained during the iterative process to

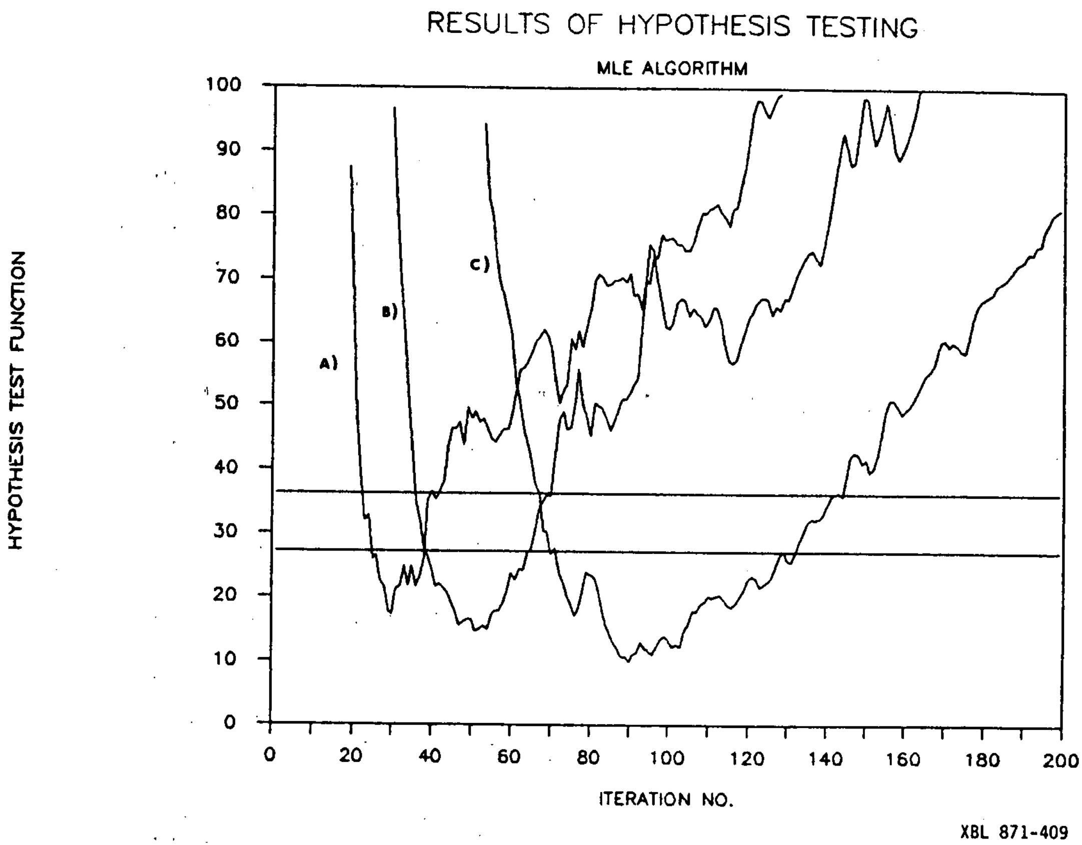


Fig. 8: Value of the hypothesis test function vs. iteration number for MLE reconstructions of the source distribution of Fig. 1. Curve a) is for 2 million counts in the source image, b) for 8 million and c) for 32 million counts. The line at a value of 36.2 is the limit below which the probability of accepting an image when it should be rejected is 0.1. The line at 27.2 is the corresponding limit for a a probability of 0.01.

be badly behaved. For tubes with approximately 100 counts, consistency with a Poisson distribution would require, like in our previous simplified example, that the projection data are all realizations of Poisson distributions with standard deviations of approximately 10. Instead, the reconstruction has an infinitesimally small standard deviation and the tube histogram can be expected to be bunched up in the middle. The results of our reconstructions confirm the above expectations: Fig. 13 shows the source image and the results of reconstruction at iteration 150, showing no deterioration. Figure 14 shows the tube histograms at several iterations. The hypothesis testing function never reaches a value low enough to be considered acceptable.

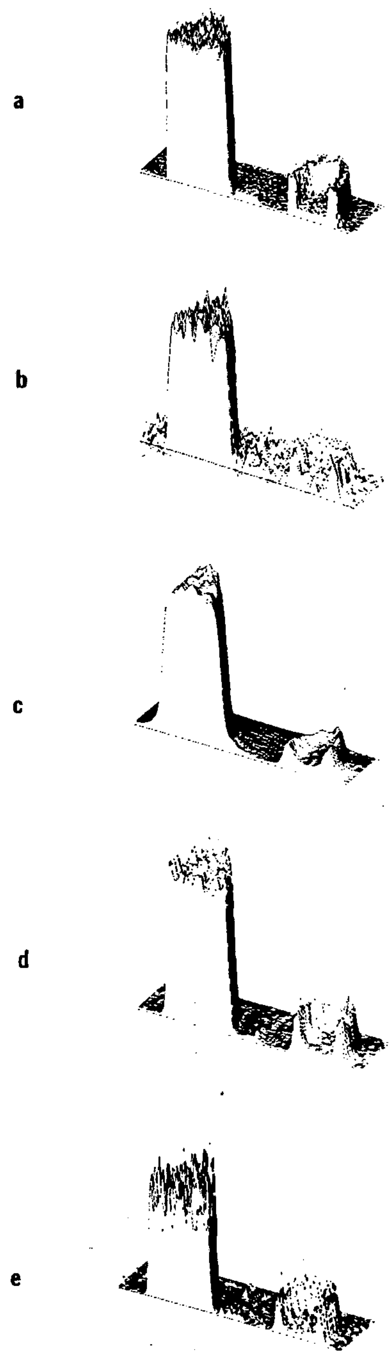
We intend to start applying our stopping rule to reconstructions from real data obtained from tomographs in the near future. We expect some difficulties due to:

a) the existence of background counts from random coincidences which are inconsistent with the transition matrix,
b) detector pair non-uniformities that will probably require applying corrections to the transition matrices directly so as not to modify the statistics of the raw data by a normalization,
c) absorption corrections that we also expect will have to be applied to the transition matrices, and
d) the transition matrix must represent the tomographic instrument with an as yet undetermined level of accuracy.

DISCUSSION AND CONCLUSIONS

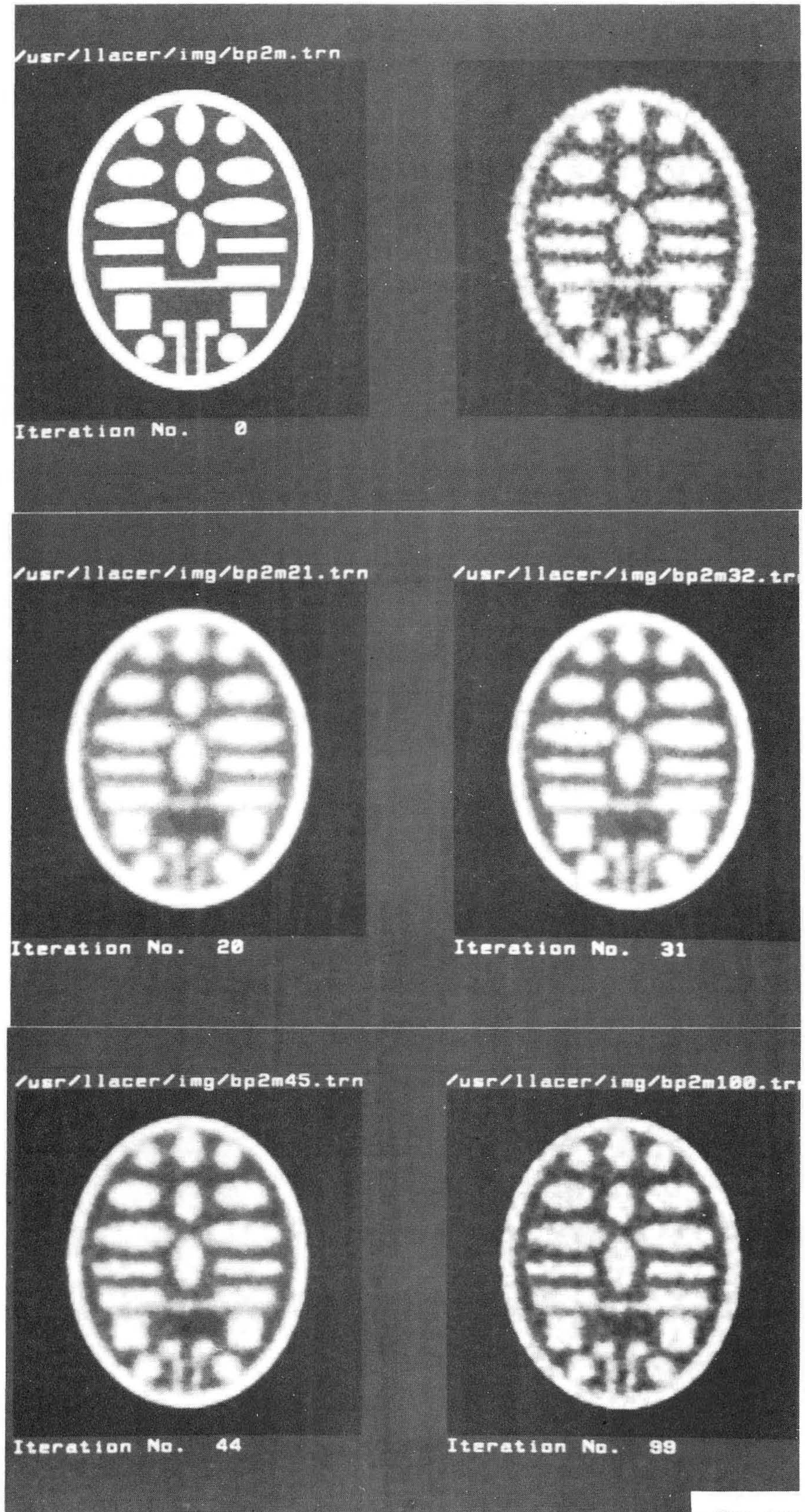
We feel that the most important result that we have obtained in the research reviewed in this paper is the realization that an iterative algorithm that seeks an extremum of some target function may lead to a contradiction with the underlying assumption that the generation of projection data follows Poisson statistics. This contradiction will clearly occur when the target function is one that has an extremum when the projections of the obtained images approach the projection data as the number of iterations increases, as is the case with the likelihood function L of Eq. 1, or a minimum least squares error function. In order not to violate the Poisson nature of the process, we have shown that it is necessary to stop the iterative process at some point and the hypothesis testing method described appears to be a suitable method of doing so.

Stopping the iterative process before reaching an extremum of a target function places the choice of such a function in a different light from the customary one, since we would seldom expect to attain the goal for which it was designed. What is most important is the behavior of the target function in the earlier stages of the iterative process, while it follows a path towards its extremum. In that respect, we have obtained a substantial amount of information about the MLE algorithm. In its useful range of iterations its behavior is determined principally by the fact that the magnitude of the derivative of the Poisson function with respect to its mean in the vicinity of the mean is higher for low means that for high means. During the iterative process, the likelihood function L of Eq. 1 gains more by matching $\lambda^*(d)$ to $n^*(d)$ for tubes with low numbers of counts than for tubes with high counts. It appears that this explains most of the observed behavior of the algorithm:



XBB 871-535

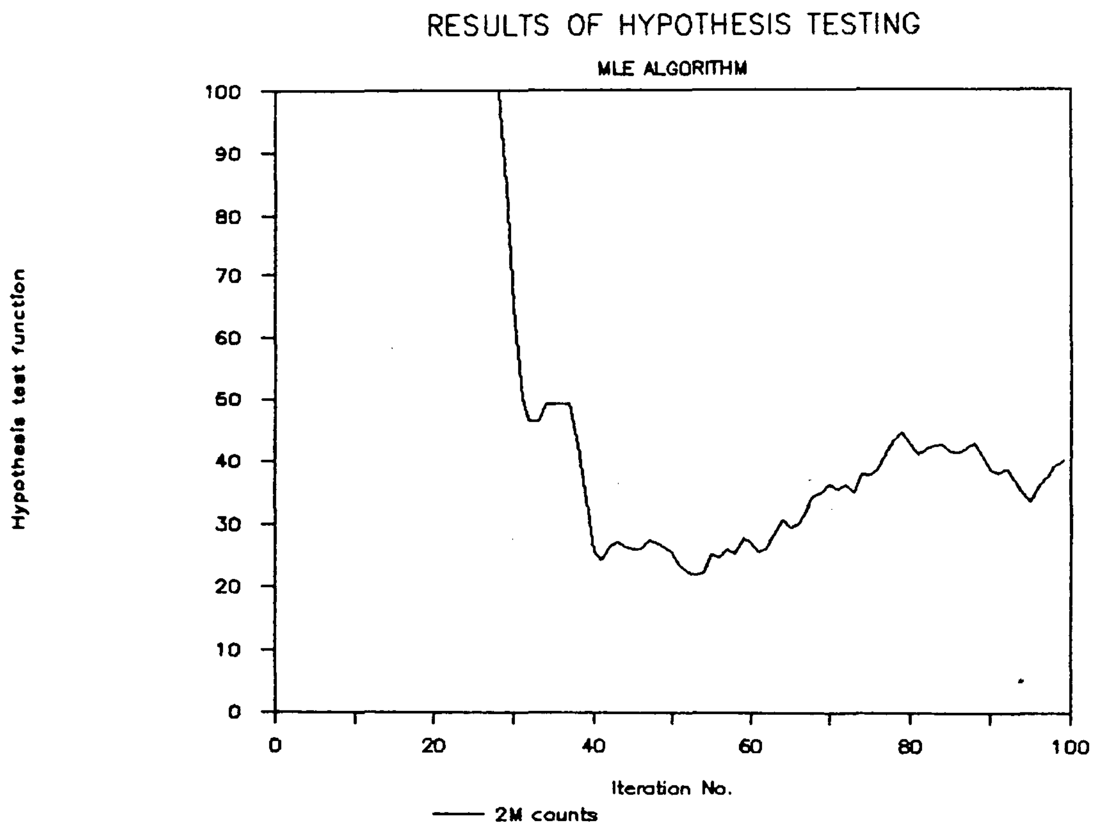
Fig. 9: Cuts through source and reconstructed images. a) source image with 2 million counts. b) reconstruction by filtered back-projection (Shepp-Logan filter). c) MLE reconstruction, 9 iterations. d) ditto, 32 it. ("best" image). e) 200 iterations.



XBB 878-6851

Fig. 10: Left to right, top to bottom: source image, FBP (Shepp-Logan filter) and MLE reconstructions for 21, 32, 45 and 100 iterations, presented with 16 levels of gray. The "best" image is near iteration 45, determined by the value of the hypothesis testing function H, of Fig. 11.

a) backgrounds in MLE reconstructions are very clean, with very little noise magnification.
b) the meaningful matching of tubes with high counts starts to occur after the low count tubes are already rather well matched.
c) contrast and sharpness in the periphery of objects of high activity are very good in MLE reconstructions.
d) ringing at edges is usually observed in the "best" reconstructions and it is a consequence of the finite size of pixels being traversed by tubes with low and high counts (from different projection angles) in the vicinity of edges.
e) contrast inside regions of high activity is somewhat difficult to attain, as already pointed out by Tanaka (1987), since all the projections involved may have high counts.
f) accurate reconstruction of low activity regions in the neighborhood of high activity regions may be hard to obtain since it involves tubes with both high counts and low counts. It appears then that the low activity region is properly reconstructed when the high activity region is getting very well matched, with image deterioration. In the case of Region 2 of Fig. 6, the proper average activity of 0.05 was only attained past the "best" image point, but could be modulated by using the WLE target function with $s < 0$.



XBL 875-2405

Fig. 11: Hypothesis testing function for the reconstructions of the phantom of Fig. 10, with 2 million counts in the source.

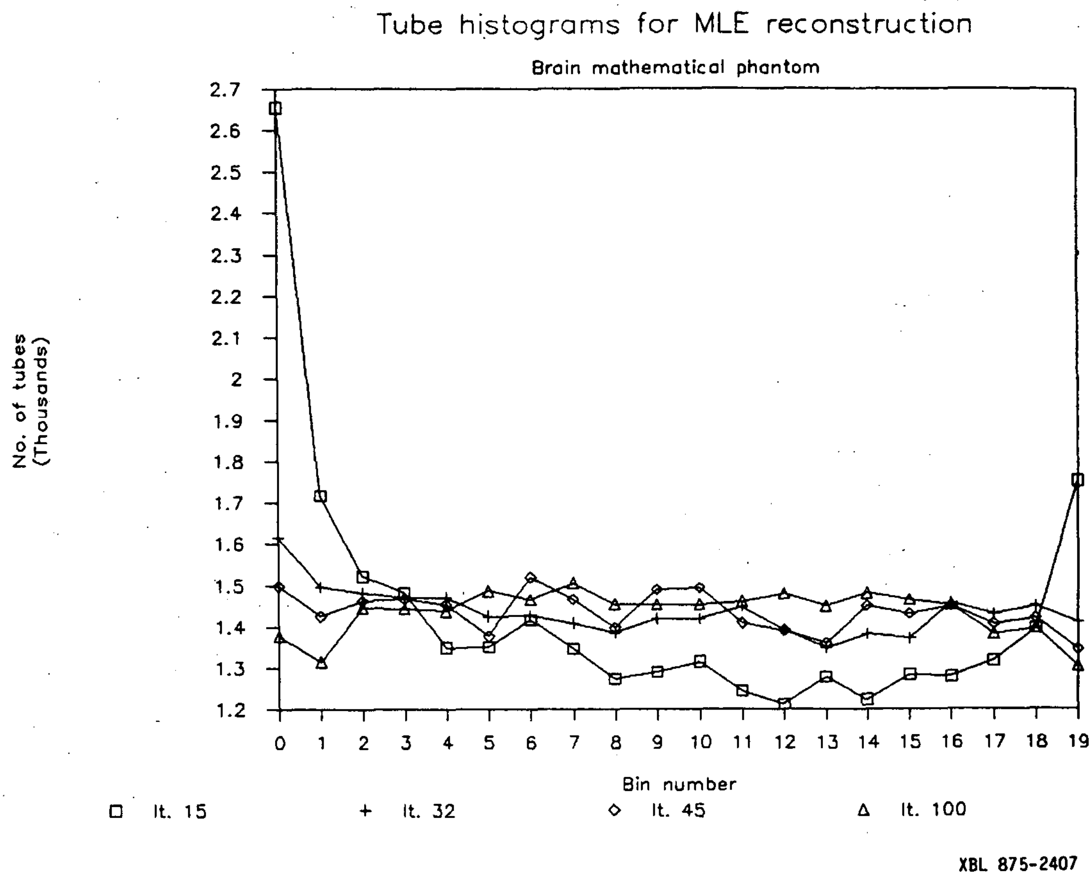


Fig. 12: Tube histograms for the MLE reconstructions with 15, 32 and 100 iterations. The flat distribution is attained near the "best" reconstruction.

In conclusion, we see the likelihood target function of Eq. 1 as a very interesting one that may have usefulness in medical reconstructions, but it is not clear to these authors, at this time, that its virtues are important in clinical diagnosis. Along with a continued study of a broader spectrum of target functions, we intend to test the usefulness of the MLE algorithm by a series of ROC studies in collaboration with the UCLA PET group.

Fig. 13: Source image consisting of a reconstructed Hoffman brain phantom and MLE reconstruction at iteration 150. An exact generation of projection data (without statistical fluctuations) results in no image deterioration.

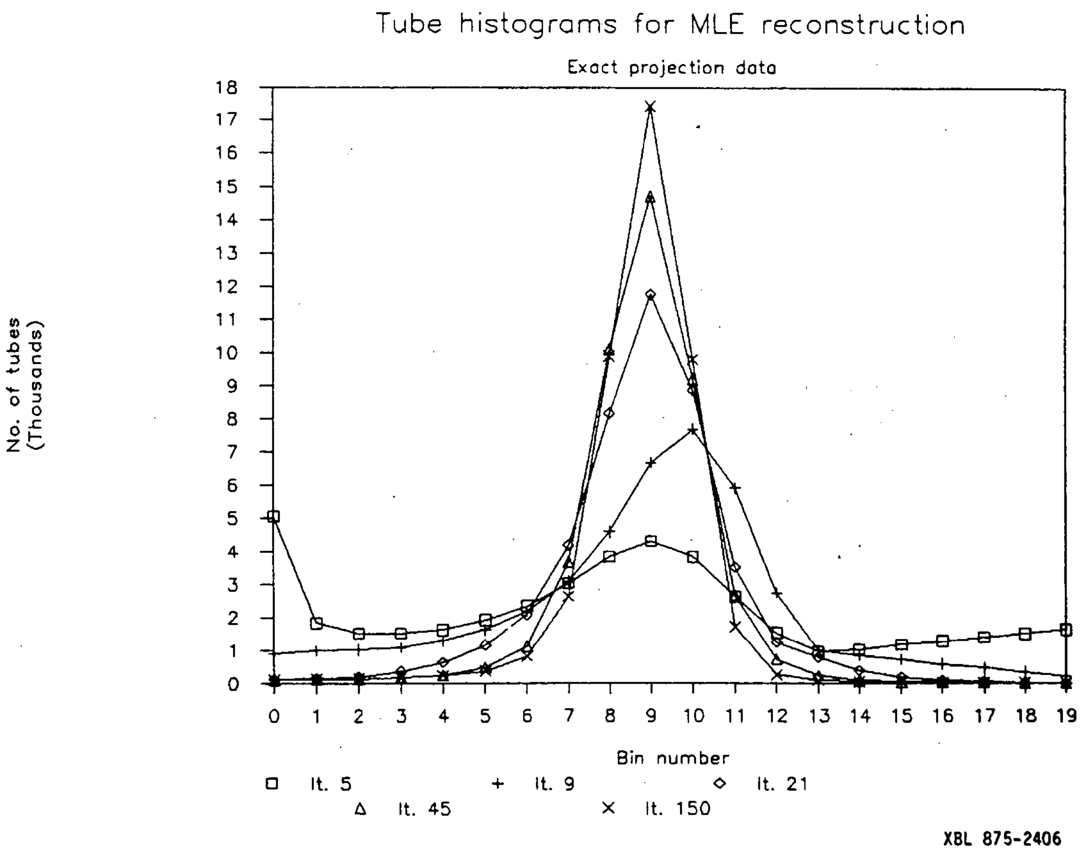


Fig. 14: Tube histograms for the reconstruction of Fig. 14 at 5, 32 and 200 iterations. A good flat histogram is never obtained for the "exact" (non-Poisson) projection data.

ACKNOWLEDGMENTS

The active collaboration of Edward J. Hoffman, UCLA Dept. of Radiological Sciences, in several aspects of our work is gratefully acknowledged. This work was supported in part by the National Cancer Institute (CA-39501) and by the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

REFERENCES

Canavos, G. (1984). Applied Probability and Statistical Methods, Little, Brown and Co., Boston, Toronto.

Hoffman, E. J., Ricci, A. R., Var der Stee, L. M. A. M. and Phelps, M. E. (1983). ECAT-III, basic design considerations, IEEE Trans. Nucl Sci., NS-30, No. 1, 729.

Levitan, E. and Herman, G. T. (1986). A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography, Medical Image Proc. Group Tech. Report No. MIPG115, Department of Radiology, U. of Pennsylvania.

Llacer, J., Andreae, S. and Chatterjee, A. (1986). On the applicability of the maximum likelihood estimator algorithm for image recovery in accelerated positron emitter beam injection, SPIE, 671, 269.

Llacer, J., Veklerov, E. and Hoffman, E. J. (1987). On the convergence of the maximum likelihood estimator method of tomographic image reconstructions, Proc. Conf. on Medical Imaging, Newport Beach, CA SPIE, 767.

Llacer, J. and Veklerov, E. (1987). The high sensitivity of the maximum likelihood estimator method of tomographic image reconstruction, Proc. Int. Symposium on Computed Assisted Radiology (CAR '87), Berlin.

Shepp, L. A. and Vardi, Y. (1982). Maximum likelihood reconstruction for emission tomography, IEEE Trans. Med. Imaging, MI-1, No. 2, 113-121.

Snyder, D. L., Lewis, J. T. and Ter-Pogossian, M. M. (1981). A mathematical model for positron-emission tomography systems having time-of-flight measurements, IEEE Trans. Nucl. Sci., NS-28, 3575.

Snyder, D. L. and Miller, M. I. (1985). The use of sieves to stabilize images produced with the EM algorithm for emission tomography, IEEE Trans. Nucl. Sci., NS-32, No. 5, 3864.

Tanaka, E. (1987). A Fast Reconstruction Algorithm for Stationary Positron Emission Tomography Based on a Modified EM Algorithm, to be publ. in IEEE Trans. on Med. Imaging.

Veklerov, E. and Llacer, J. (1987). Stopping rule for the MLE algorithm based on statistical hypothesis testing, submitted to IEEE Trans. Medical Imaging for publication.

LAWRENCE BERKELEY LABORATORY
TECHNICAL INFORMATION DEPARTMENT
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720