Title
STOPPING RULE FOR THE MLE ALGORITHM BASED ON STATISTICAL HYPOTHESIS TESTING

Authors
Veklerov, E.
Llacer, J.

Publication Date
1987-02-01

LBL-22963
Preprint

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

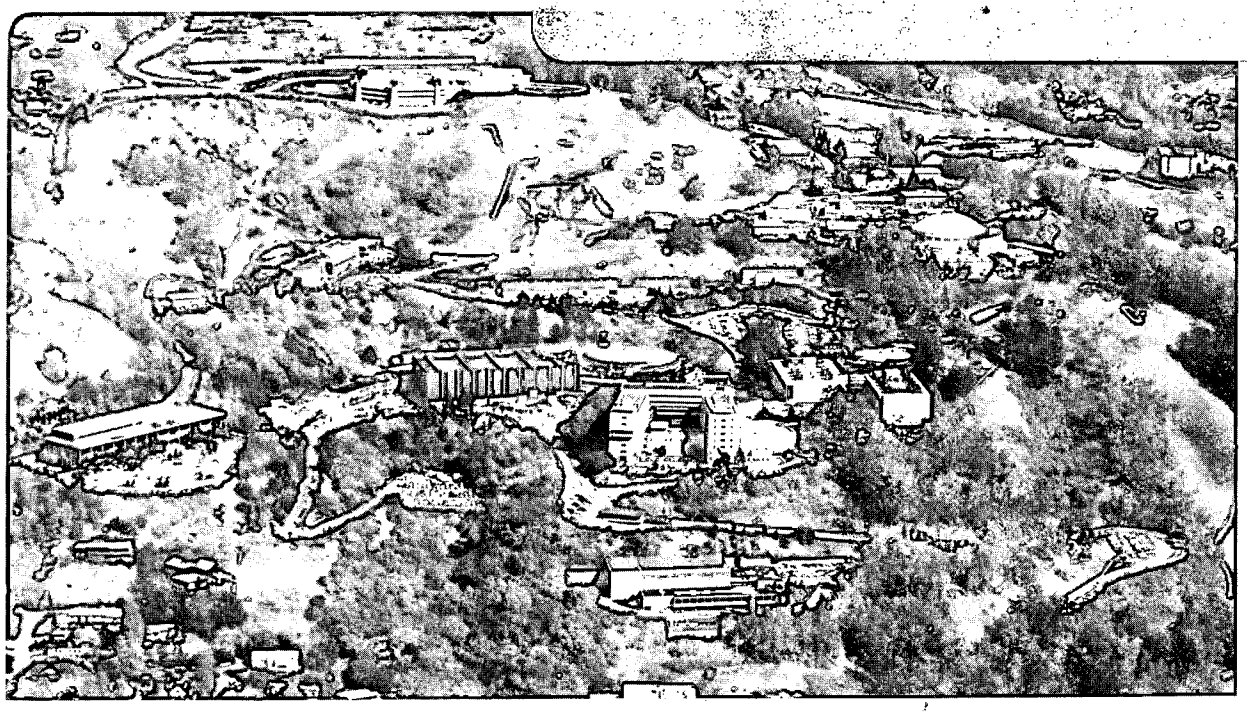
Engineering Division

Submitted to IEEE Transactions on Medical Imaging

Stopping Rule for the MLE Algorithm Based on Statistical Hypothesis Testing

E. Veklerov and J. Llacer

February 1987



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

DISCLAIMER

LBL-22963

Stopping Rule for the MLE Algorithm Based on Statistical Hypothesis Testing

EUGENE VEKLEROV, JORGE LLACER

Lawrence Berkeley Laboratory

University of California, Berkeley, CA 94720[1]

Abstract— **It is known that when the Maximum Likelihood Estimator (MLE) algorithm passes a certain point, it produces images that begin to deteriorate. We propose a quantitative criterion with a simple probabilistic interpretation that allows the user to stop the algorithm just before this effect begins.**

The MLE algorithm searches for the image that has the maximum probability to generate the projection data. The underlying assumption of the algorithm is a Poisson distribution of the data. Therefore, the best image, according to the MLE algorithm, is the one that results in projection means which are as close to the data as possible. It is shown that this goal conflicts with the assumption that the data is Poisson-distributed.

We test a statistical hypothesis whereby the projection data could have been generated by the image produced after each iteration. The acceptance or rejection of the hypothesis is based on a parameter that decreases as the images improve and increases as they deteriorate. We show that the 'best' MLE images, which pass the test, result in somewhat lower noise in regions of high activity than the filtered back-projection results and much

[1]This work was supported in part by the National Cancer Institute (CA-39501) and the U.S. Department of Energy under Contract No. DE-AC03-76SF00098

improved images in low activity regions.

The applicability of the proposed stopping rule to other iterative schemes is discussed.

1. INTRODUCTION

The Maximum Likelihood Estimator (MLE) algorithm has attracted considerable interest in the field of positron emission tomography. It produces superior images in addition to being flexible, simple and allowing a physical interpretation. The use of the maximum likelihood criterion in emission tomography was originally proposed by Rockmore and Macovski [1]. Shepp and Vardi [2] developed and investigated the MLE algorithm. They offered an approximate but efficient algorithm for computing the transition matrix, proposed an algorithm for maximizing the likelihood, and proved some properties of the algorithm. It is a gradient-type algorithm, although [2] provided a physical interpretation of its steps.

Our notation is based on [2]. Let $n^*(d), (d = 1, ..., D)$ be the projection data or the number of coincidences detected in tube d. The problem is to estimate the emission density $\lambda(b), (b = 1, ..., B)$, where B is the number of pixels making up the image. We will assume that the following B by D matrix known as the transition matrix is precomputed:

$$p(b,d) = \text{Prob(detected in } d \mid \text{emitted in } b)$$

and

$$\sum_{d=1}^{D} p(b,d) = 1$$

Let $\lambda^*(d)$ be

$$\lambda^*(d) = \sum_{b=1}^{B} \lambda(b) p(b,d)$$

The goal of the algorithm is to maximize the likelihood function $L(\lambda)$ which is the probability that the image λ generates the given projection data and is defined as follows:

$$L(\lambda) = \prod_{d=1}^{D} e^{-\lambda^*(d)} \frac{\lambda^*(d)^{n^*(d)}}{n^*(d)!} \tag{1}$$

In order to maximize (1), an iterative algorithm was proposed in [2]. It increases $L(\lambda)$ at each step and converges to its maximum value. However, it has been reported that when the algorithm passes a certain point, it produces images that begin to deteriorate. This has been observed and explained by Snyder, et al [3]. This phenomenon was further investigated in [4] and [5]. In particular, it was noticed there that the process of image deterioration starts at a point which depends on the total number of counts in the projection data – the more counts, the later it begins. Once this process has begun, it progresses more rapidly when the total number of counts in the projection data is small.

Some authors (see e.g. [3] and [6]) have proposed to tackle this problem by replacing the likelihood criterion by other criteria designed in such a way that they prevent the deterioration, or

by maximizing the likelihood criteria subject to some constraints. We feel that although their approaches are able to prevent the deterioration, the authors had to include some new elements (functions, parameters, matrices or a priori distributions) which affect the reconstructed images, while selection of specific values of these elements has yet to be justified.

That is why we have opted for another approach. We analyze the original MLE algorithm without introducing any new elements and find the stopping point on the basis of intrinsic properties of the algorithm. It is shown that maximization of (1) and the underlying assumption that the data is Poisson-distributed sooner or later become contradictory to each other. Then we propose a quantitative criterion that allows the user to catch the moment when they are least contradictory.

We have chosen to work on the non-accelerated MLE algorithm first. The applicability of the stopping rule to accelerated schemes (see e.g. Kaufman [7], Lewitt and Muehllehner [8], Tanaka [9]) will be studied in the future.

2. RATIONALE

It was noticed by Snyder, Lewis and Ter-Pogossian [10] that $n^*(d)$ are independent and Poisson random variables and $\lambda^*(d)$ are their means. Therefore, the image maximizing (1) must be such that $\lambda^*(d)$ are as close to $n^*(d)$ as possible. (This is also seen from (1) directly). Although it cannot be guaranteed that all $\lambda^*(d)$ may be made equal to corresponding $n^*(d)$, one can expect that if B is large, the two may be made very close for almost all d. This was observed in [4] and [5].

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. However, if the two get too close for all or almost all projections d, it becomes statistically unlikely that the image could have generated the data. The chance that a source distribution emits a number of γ-rays very close to its mean for all directions where detectors are placed is very small.

After each iteration of the MLE algorithm, we propose to test the hypothesis that $n^*(d)$ for all $d = 1, ..., D$ are jointly statistically valid realizations of Poisson-distributed random variables with the means $\lambda^*(d)$ for corresponding arguments. In other words, we test the hypothesis that $n^*(1)$ is a realization of a Poisson variable with the mean $\lambda^*(1)$, $n^*(2)$ was drawn from a Poisson distribution with the mean $\lambda^*(2)$, etc. It is reasonable to require that only

images passing the test (for which the hypothesis is not rejected) be declared acceptable.

3. SOLUTION

We want to test the hypothesis whereby $n^*(1), n^*(2), ..., n^*(D)$ are realizations of Poisson random variables with the means $\lambda^*(1)$, $\lambda^*(2)$,...,$\lambda^*(D)$, respectively. This is done by deriving a version of Pearson's χ^2 statistic (see e.g. [11]) suitable for our case.

Let us imagine for a moment that all $\lambda^*(d)$ are identical. Then, according to Pearson's scheme, we would break all possible values of $n^*(d)$ (that is, 0, 1, 2, etc.) into arbitrary N classes (say [0, 1], [2, 3, 4], etc.) and compare the observed number of projections falling into the i-th class with the expected number. The expected number would simply be the sum of the Poisson terms corresponding to the i-th class, p_i, times D. (In our example, $p_1 = e^{-\lambda} + e^{-\lambda} \cdot \lambda$, $p_2 = e^{-\lambda} \cdot \lambda^2/2! + e^{-\lambda} \cdot \lambda^3/3! + e^{-\lambda} \cdot \lambda^4/4!$, etc.). Pearson's test allows the user to judge whether the observed and expected numbers are close enough or not.

However, $\lambda^*(d)$ are not identical. For any given λ^*, we can again break all possible values of n^* into N classes. Let $p_1, p_2, ..., p_N$ again be the theoretical probabilities of the N classes. Then, let us take another value of λ^* and try to come up with the *same*

probabilities. If we dealt with a continuous distribution, we would simply scale $n^*(d)$. In our case, unfortunately, we may be unable to break all possible values of n^* into classes with the same theoretical probabilities, whatever classes we choose, because of the discrete nature of the Poisson distribution. That is why we have to resort to the following procedure which, in effect, breaks the discreteness of the Poisson distribution.

We shall do everything in the reverse order. First we shall choose the theoretical probabilities $p_1, p_2, ..., p_N$ arbitrarily. We may assume, for simplicity, that all $p_i = 1/N$. Then we shall define N classes in such a way that, *regardless of the value of* $\lambda^*(d)$, any observed number $n^*(d)$ will be assigned to the i-th class with the probability $1/N$ if the Poisson assumption is correct. The following algorithm will do the job.

For every projection d, compute two probabilities,

$$p_1 = \sum_{i=0}^{n^*(d)-1} e^{-\lambda^*(d)} \frac{\lambda^*(d)^i}{i!}$$

and

$$p_2 = \sum_{i=0}^{n^*(d)} e^{-\lambda^*(d)} \frac{\lambda^*(d)^i}{i!}$$

Then generate a random number, x, uniformly distributed between p_1 and p_2. It is easy to prove that, if $n^*(d)$ is Poisson-distributed with the mean $\lambda^*(d)$, x is uniformly distributed between 0 and 1. (This is the discrete equivalent of the well-known fact that if x is a continuous random variable with an arbitrary distribution

function, F, then $F(x)$ is uniformly distributed between 0 and 1.) Let j be the smallest integer that is equal to or greater than xN. Then this projection is assigned to the j-th class.

Therefore, testing the original hypothesis is reduced to testing that x is uniformly distributed between 0 and 1. Then we can easily follow Pearson's procedure. Let $h_i, (i = 1, ..., N)$ be the number of projections falling into the i-th class of the histogram. If the hypothesis is true and D is sufficiently large, which is always the case in imaging, then

$$H = \sum_{i=1}^{N} \frac{(h_i - D/N)^2}{D/N} \tag{2}$$

is approximately χ^2 distributed with $N - 1$ degrees of freedom. This follows from the fact that the distribution of each

$$\frac{h_i - D/N}{\sqrt{(D/N)(1 - 1/N)}} \tag{3}$$

is approximately the standard normal. Therefore, the hypothesis should be rejected if the value of H exceeds the critical value based on the number of degrees of freedom $(N - 1)$ and the significance level.

The significance level is the probability of rejecting the hypothesis, given that it is true. In our examples we generated histograms with 20 classes and used the significance levels of 0.2, 0.1, 0.05 and 0.01 which are common. The corresponding critical values taken from a table of quantile values of the χ^2 distribution (see e.g. [11])

are 23.9, 27.2, 30.1 and 36.2.

Viewing the histogram of x is rather instructive in itself. After the first iterations, most of the projections concentrate in the extremely left and extremely right bins and the histogram is convex. This is another way of saying that $n^*(d)$ are too far away from corresponding $\lambda^*(d)$. Gradually, the projections spread out and the histogram becomes more and more uniform. Then, however, the projections tend to move on towards the central bins of the histogram and the histogram becomes concave, which is a way of saying that now $n^*(d)$ are too close to $\lambda^*(d)$.

4. EXPERIMENTS

We used a source image with 2 million counts obtained by a random process described in [4], based on the activity distribution shown in Fig. 1. The high and low activity regions of the source image are shown in Fig. 2, in which 16 levels of grey have been compressed into the upper quarter of the color scale and the same number of levels have been compressed into the lower eighth the color scale. The parameter H of the χ^2 test obtained at the end of each MLE iteration is plotted as curve a in Fig. 3. It reaches its minimum around iteration 30 after which it gradually increases. The two horizontal lines depict the critical values corresponding

to significance levels of 0.1 and 0.01. It follows that the window of acceptable images lies around iteration 30 and has a width of $\pm$ 5 to 10 iterations depending on the significance level. It is important that the window is not too sensitive to the significance level within a reasonable range of the latter.

The image obtained after iteration 30 is shown in Fig 4. Fig. 5 shows the image reconstructed by the filtered back-projection (FBP) method, using the Shepp-Logan filter, after setting negative values to 0. All images are separated into high activity and low activity parts, as was the source image. Fig. 6, a) through e), depicts cuts through the source image, through the MLE images (after 15, 30 and 60 iterations) and through the FBP image.

Examination of the images shows that the 'best' MLE image (after 30 iterations, according the criterion H) has somewhat lower noise in regions of high activity than the FBP image and is by far less noisy than the FBP image in regions of low activity.

Comparison of the three MLE images shows that while the 'best' one already manifests some signs of deterioration in the high activity region, it depicts the small circular low activity region (beside the big ellipse) with higher sharpness than after 15 iterations. The image after 60 iterations is even sharper in the low activity region but is excessively noisy in the high activity region. In other words, the 'best' MLE image is a reasonable compromise between two contradictory trends.

It is interesting to observe the behavior of H when the total number of counts increases. Curves b and c in Fig. 3 show the behavior of H with 8 million and 32 million counts, respectively. It is evident that the higher the total number of counts, the more iterations it takes to reach the 'best' image. Additionally, the 'best' images in cases of higher numbers of counts are even more compatible with the hypothesis (the curves dip lower) and the windows of acceptable images become wider (the curves are flatter around the minima). These results are consistent with our earlier observations in [4] and [5].

5. DISCUSSION OF RESULTS AND APPLICABILITY TO OTHER ITERATIVE SCHEMES

There are many objective functions which can be used to reconstruct images in emission tomography. The likelihood L defined in equation (1) is one of them, another is the sum of the squares of the differences between $n^*(d)$ and $\lambda^*(d)$. The parameter H proposed in this paper is yet another possible candidate. What is the best objective function? We realize that there is no single answer to this question.

We have demonstrated that the goals of maximizing L and minimizing H may contradict each other. As usual, one can try to combine the two contradictory goals by using a weighted sum of L

and H. Another standard approach is to maximize L provided that H does not exceed a given value or to minimize H provided that L is greater than a given value. We intend to continue investigating these approaches in the future.

However, we believe the approach described here is a viable alternative. It reconciles the two goals rather well. As the MLE progresses, it passes though a narrow window of acceptable images, provided that we restrict ourselves to a commonly used range of values of α, and we have observed that the window is not very sensitive within this range. In all our experiments, this window corresponded to the visually best images produced by the MLE and these images were superior to those produced by the filtered back-projection methods.

We believe the reason for this success can be expressed as a general principle: the best images are those for which $n^*(d)$ and $\lambda^*(d)$ are 'close but not too close'. There are several ways to quantify the notion of being 'close', L and the quadratic criterion being just two examples, while H quantifies the notion of being 'not too close'. For this reason, H can be combined with other objective functions as well.

One of them is a 'weighted' L introduced in [5]:

$$WL(\lambda) = \prod_{d=1}^{D} \left[e^{-\lambda^*(d)} \frac{\lambda^*(d)^{n^*(d)}}{n^*(d)!} \right]^{sn^*(d)+t} \tag{4}$$

where s and t are constants. The objective function (2) allows the

user to emphasize projections with large numbers of counts at the expense of those with small numbers or vice versa. An iterative scheme maximizing $WL(\lambda)$ was presented in [5].

We should also note that although the parameter H quantifies the notion of being 'not too close', it is not the only possible parameter suitable for this purpose. We began this investigation by experimenting with the following stopping rule: 'stop as soon as a certain part of the projections, say 95%, are reconciled'. A projection is said to be reconciled if the difference between $n^*(d)$ and $\lambda^*(d)$ is less than $C\sqrt{\lambda^*(d)}$, where C is a constant. We believe that this stopping rule is rather empirical, whereas the rule based on the hypothesis testing is more objective, easier to justify and it more accurately takes into account all available information.

The added computation time to generate the parameter H after each iteration is small compared to the iterative process itself.

6. APPENDIX

This algorithm has been implemented in the following three functions in the C language.

```
#include <stdio.h>
#include <math.h>
#define N_COINC 30000
#define N_BINS 20

histogram (n,lambda,hist)        /*generates histogram h*/
  double n[], lambda[];
  int hist[];
{
  int i, j, bin;
  double sum;              /*accumulated sum of Poisson terms*/
  double term;             /*Poisson terms*/
  double upper;            /* 2 to the power of 31 minus 1*/
  float r, y, h_value, chi();

  srand(1);                /*initialize rand generator*/
  upper = pow((double) 2.0, (double) 31.0) - 1.0;
  for (i = 0; i < N_BINS; i++)
    hist[i] = 0;                    /*initialize histogram*/
  for (i = 0; i < N_COINC; i++) /*main loop for each projection*/
  {
    if (lambda[i] != 0.0)
    {
      if (lambda[i] < 300.0)     /*direct computations*/
      {
        sum = 0.0;
        term = exp(-lambda[i]);
        for (j=1; j<=n[i] && sum<1-1/N_BINS; j++)
        {
          sum = sum + term;
          term = term * lambda[i] / j;
        }
        r = rand() / upper;      /*get random number in [0,1]*/
        bin = (r * term + sum) * N_BINS;
      }
      else                       /*use normal approximation*/
      {
        y = (n[i] - lambda[i]) / sqrt(lambda[i]);
        bin = normal(y);
      }
      hist[bin] = hist[bin] + 1;
    }
  }
  h_value = chi(hist);
  printf("\n The value of H is %f",h_value);
}
```

```
float chi(hist) /*computes the value of H*/
  int hist[];
{
  int j, sum;
  float answer;

  sum = 0;
  for (j = 0; j < N_BINS; j++)
    sum = sum + hist[j];
  answer = 0.0;
  for (j = 0; j < N_BINS; j++)
    answer = answer + pow((hist[j] - sum / N_BINS), 2);
  return(answer * N_BINS /sum);
}


normal(y)   /*area under standard normal curve is divided into
            20 equal regions; normal returns region number of y*/
float y;
{
static float x[N_BINS] = {-1.645, -1.282, -1.037, -0.841, -0.675,
                          -0.525, -0.385, -0.253, -0.125, -0.000,
                           0.125,  0.253,  0.385,  0.525,  0.675,
                           0.841,  1.037,  1.282,  1.645};
int i;

for (i = 0; i < N_BINS - 1; i++)
  if (y < x[i])
    return(i);
return(N_BINS - 1);
}
```

REFERENCES

1. A. J. Rockmore, A. Macovski, "A maximum likelihood approach to emission image reconstruction from projection," *IEEE Trans. Nucl. Sci.*, NS-23, pp. 1428-1432, 1976.

2. L. A. Shepp, Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imaging*, vol. MI-1, pp. 113-122, 1982.

3. D. L. Snyder, M. T. Miller, "The use of sieves to stabilize images produced with the EM algorithm for emission tomography," *IEEE Transactions on Nuclear Science*, vol. NS-32, pp. 3864-3872, 1985.

4. J. Llacer, E. Veklerov, E. J. Hoffman, "On the convergence of the maximum likelihood estimator method of tomographic image reconstruction," *Proc. of Conference on Medical Imaging*, Newport Beach, California, Feb. 1987. SPIE Vol. 767, 1987.

5. J. Llacer, E. Veklerov, "The high sensitivity of the maximum likelihood estimator method of tomographic image reconstruction," *Proc. of Conference on Computer Assisted Radiology*, Berlin, July, 1987.

6. E. Levitan, G. T. Herman, "A maximum a posteriori expectation maximization algorithm for image reconstruction in emis-

sion tomography," Report. Medical Image Processing Group, Department of Radiology, University of Pennsylvania.

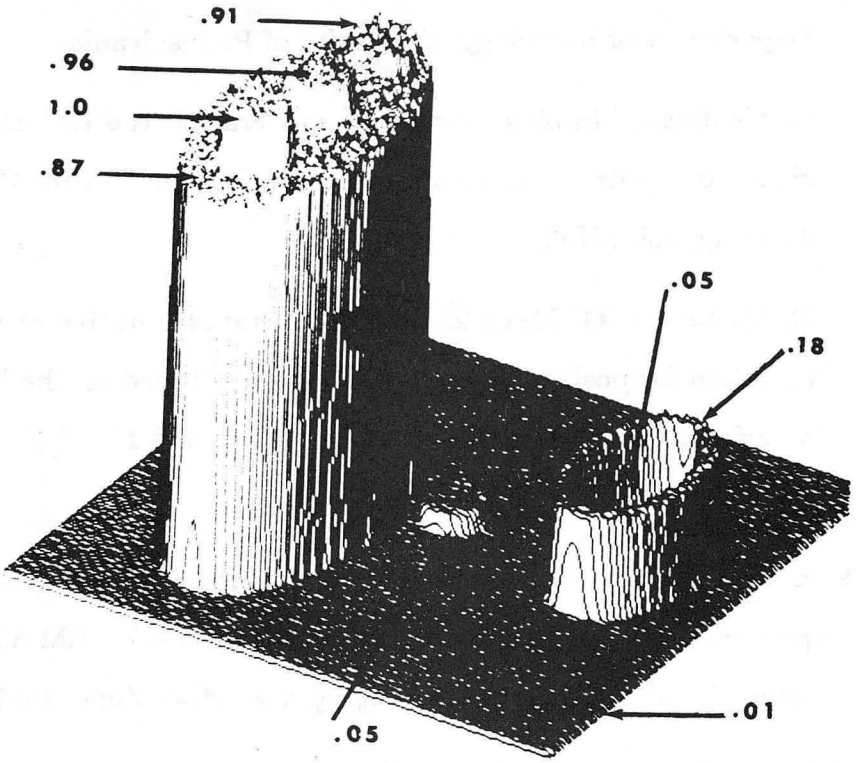
7. L. Kaufman, "Implementing and accelerating the EM algorithm for positron emission tomography," *IEEE Trans. Med. Imaging*, vol. MI-6, March, 1987.

8. R. M. Lewitt, G. Muehllehner, "Accelerated iterative reconstruction for positron emission tomography based on the EM algorithm for maximum likelihood estimation," *IEEE Trans. Med. Imaging*, vol. MI-5, March, 1986.

9. E. Tanaka, "A fast reconstruction algorithm for stationary positron emission tomography based on a modified EM algorithm," *IEEE Trans. Med. Imaging*, vol. MI-6, June, 1987.

10. D. L. Snyder, J. T. Lewis, Jr., M. M. Ter-Pogossian, "A mathematical model for positron-emission tomography systems having time-of-flight measurements," *IEEE Trans. Nucl. Sci.*, NS-28, pp. 3575-3583, 1981.

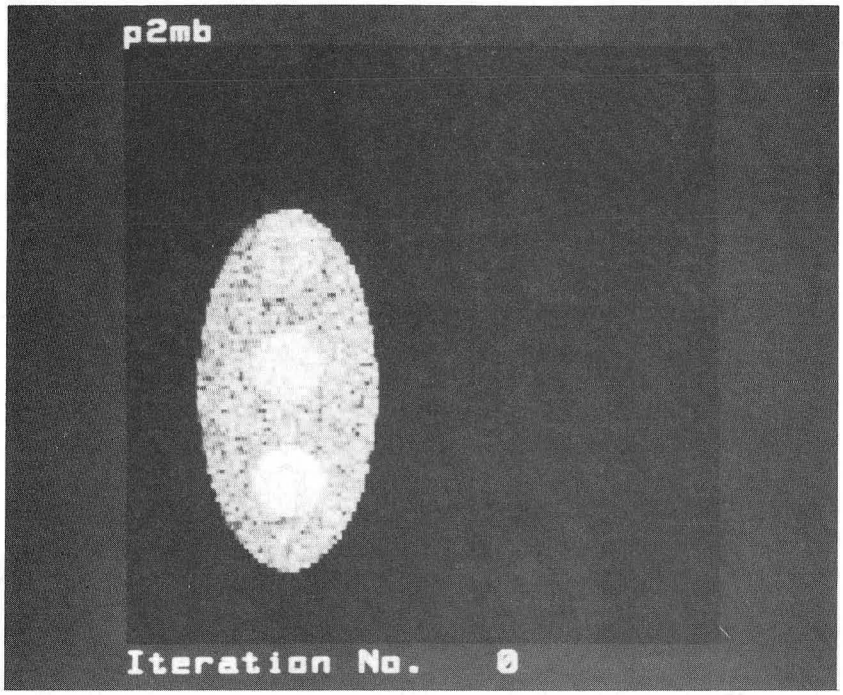
11. G. Canavos, *Applied probability and statistical methods,* Little, Brown and Co., Boston, Toronto, 1984.



XBB 860-10391-A

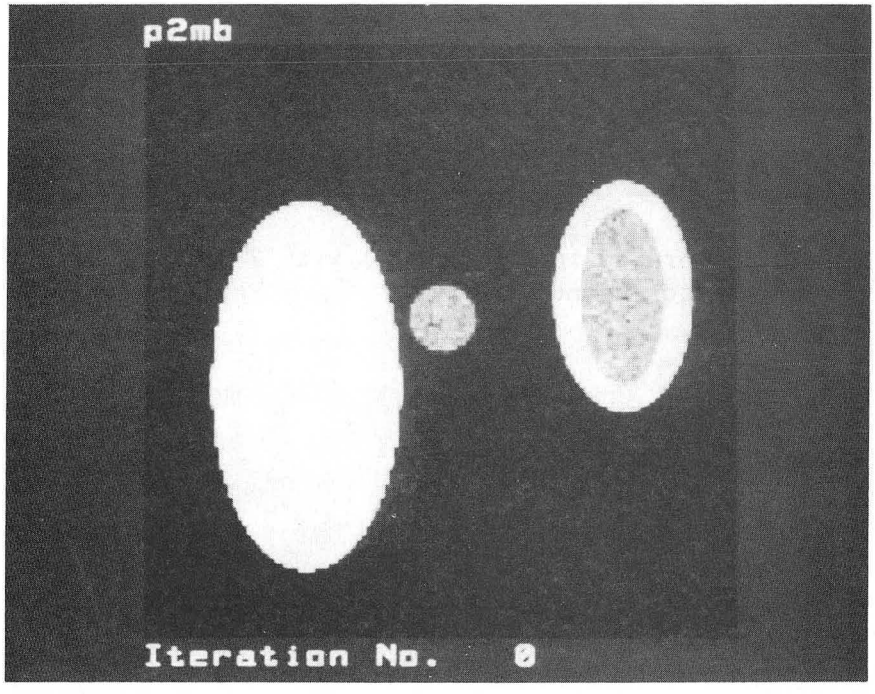
Fig. 1 Activity distribution for the generation of the source image.

Fig. 2: Top



XBC 872-1147

Fig. 2: Bottom



XBC 872-1150

Fig. 2 High (top) and low (bottom) regions of the source image with 2 million counts.

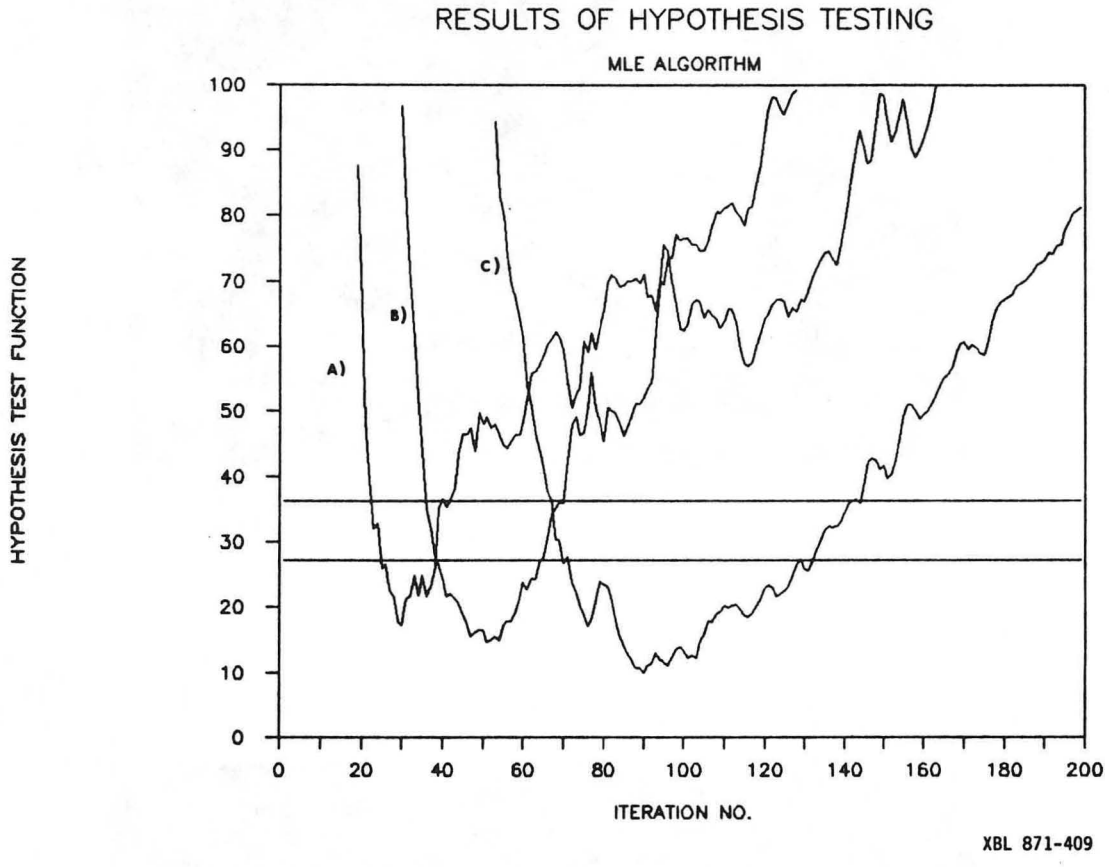
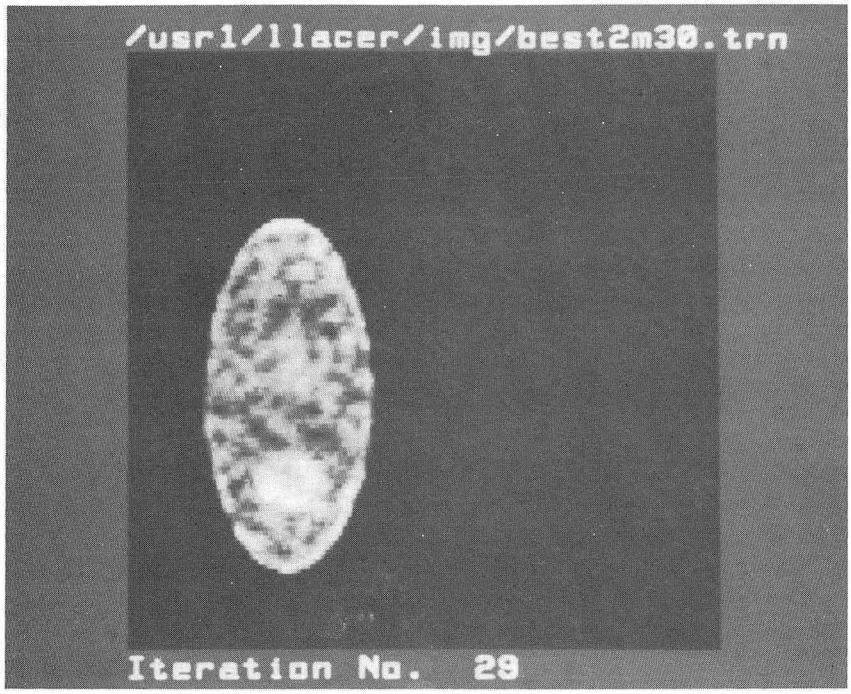


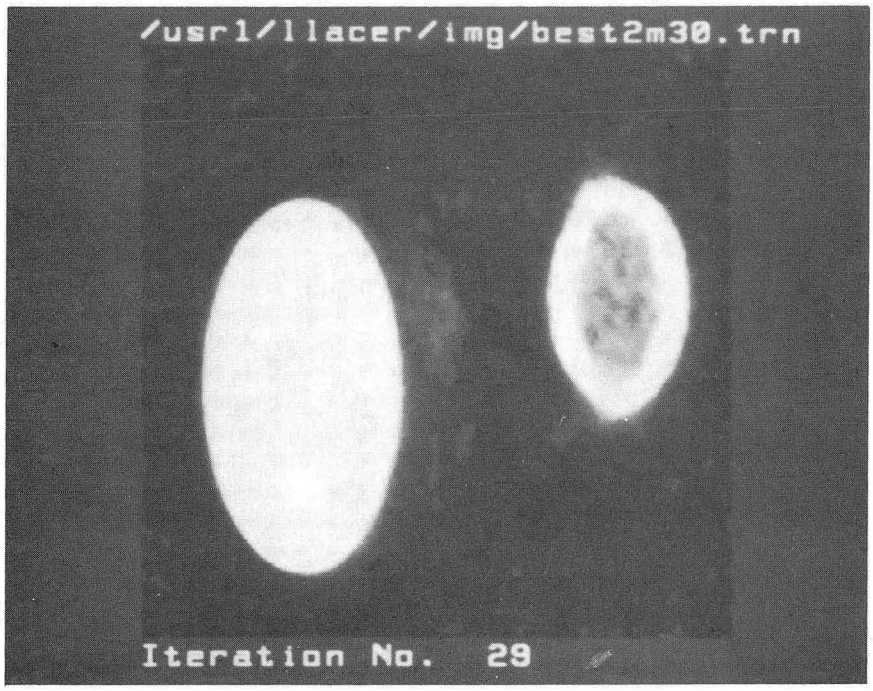
Fig. 3 Parameter H for projection data : (a) with 2 million counts, (b) for 8 million counts, (c) for 32 million counts

Fig. 4: Top



XBC 872-1148

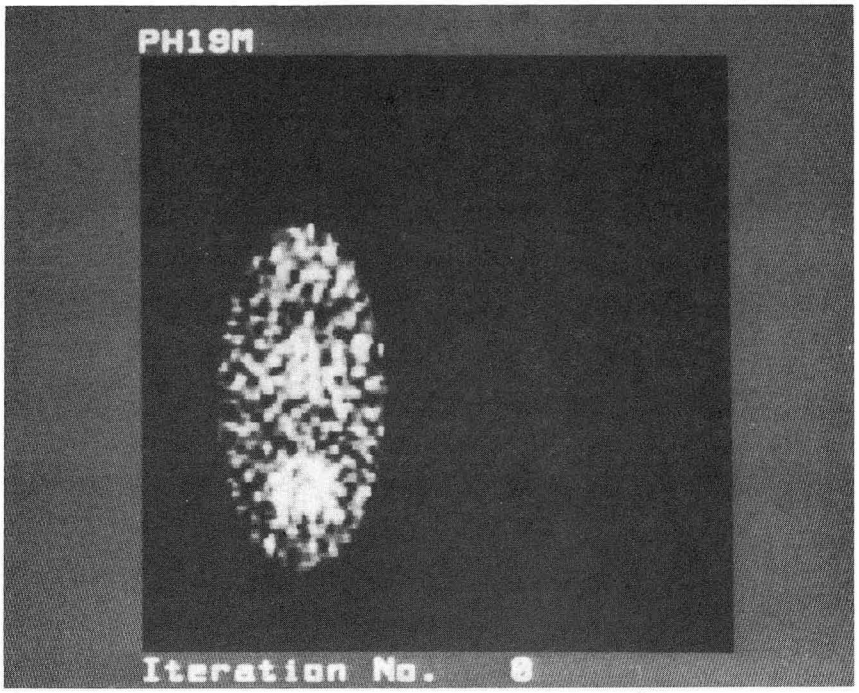
Fig. 4: Bottom



XBC 872-1151

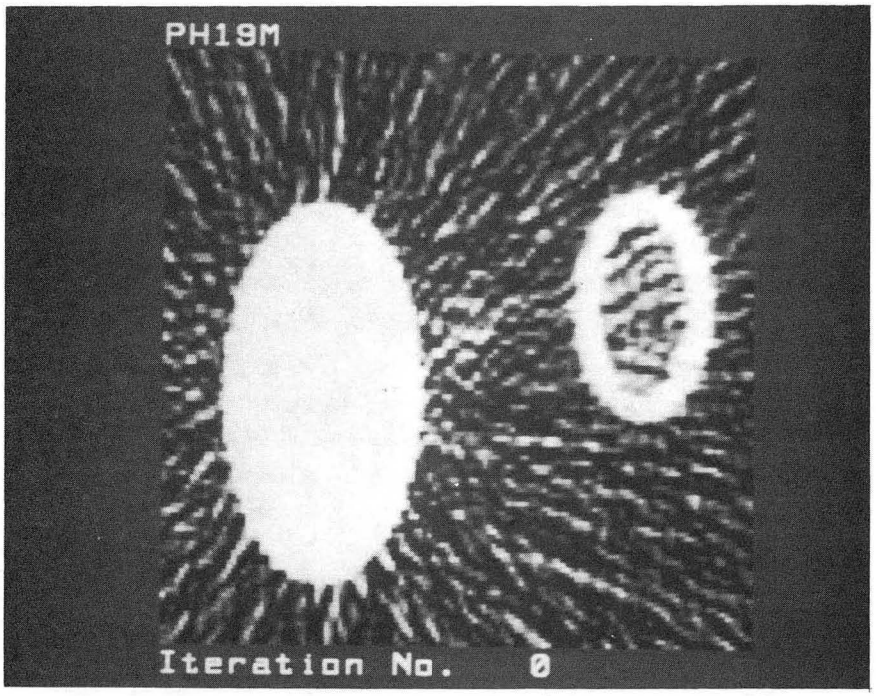
Fig. 4 MLE image after 30 iterations: high activity part (top) and low activity part (bottom)

Fig. 5: Top



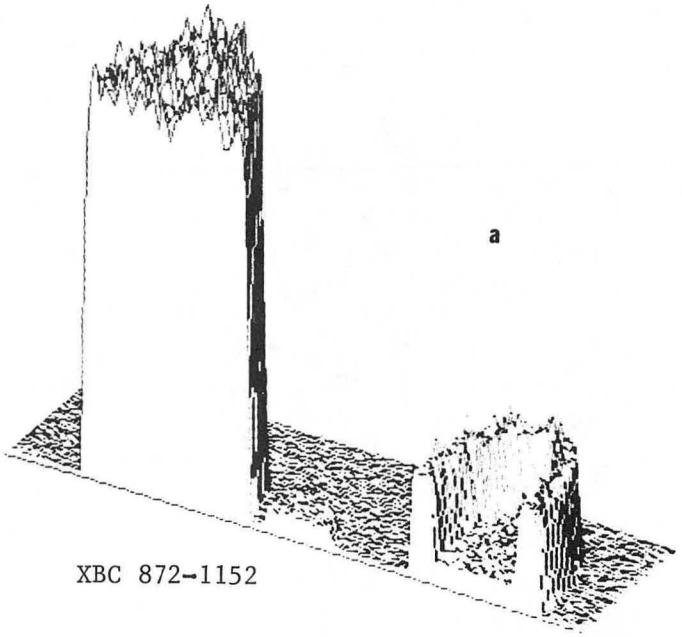
XBC 872-1146

Fig. 5: Bottom

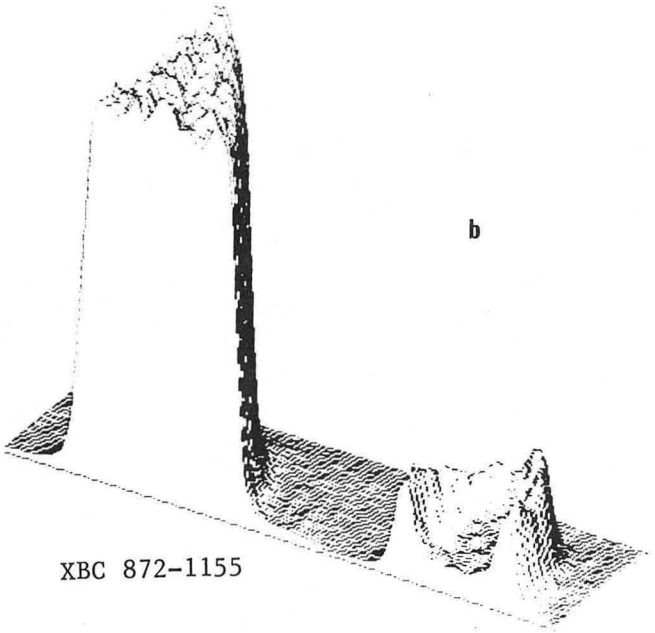


XBC 872-1149

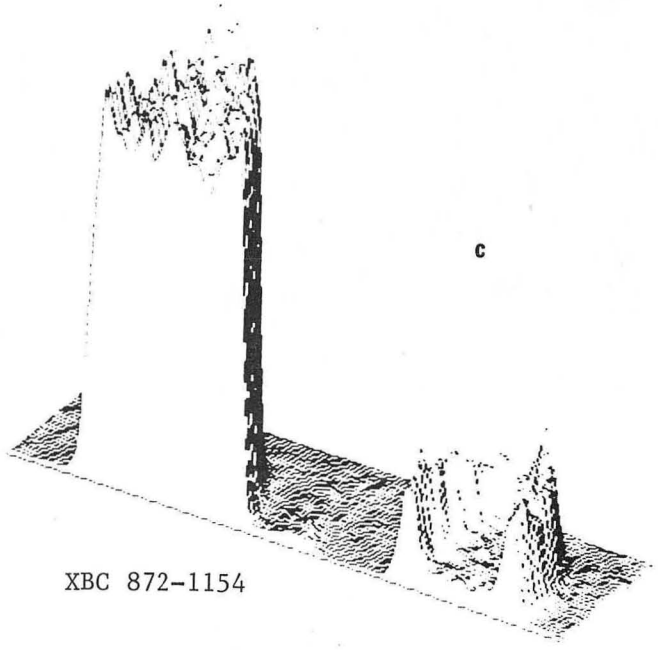
Fig. 5 Image reconstructed by the FBP method: high (top) and low (bottom) activity parts



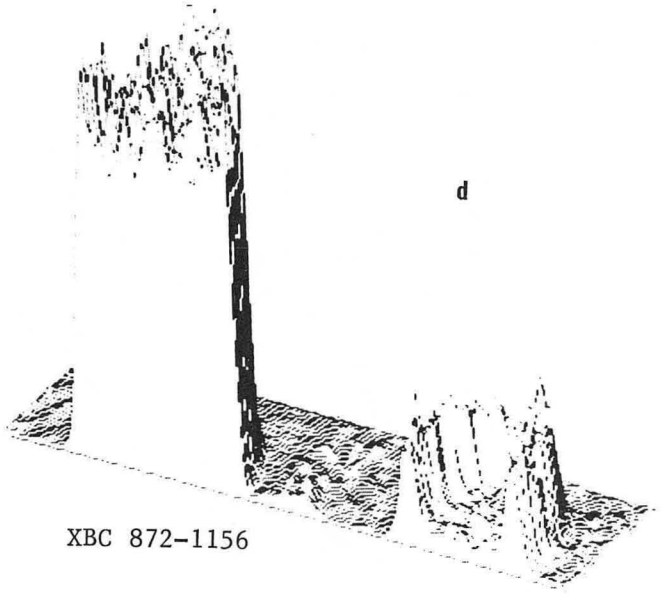
XBC 872-1152



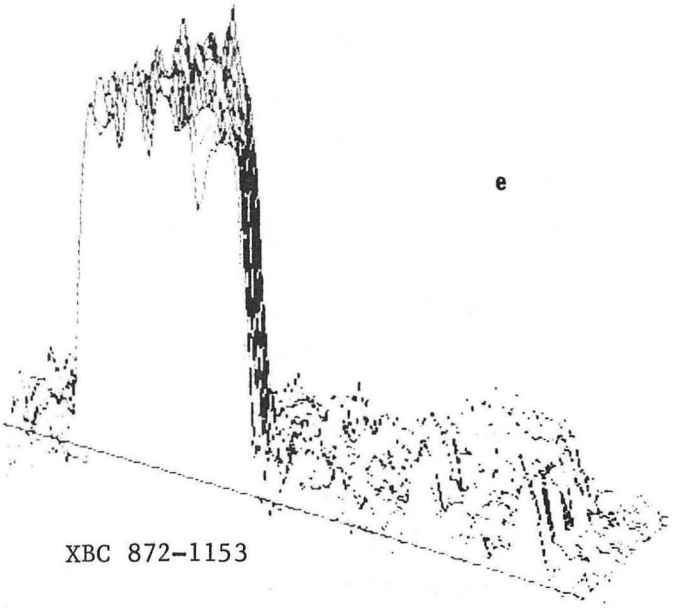
XBC 872-1155



XBC 872-1154



XBC 872-1156



XBC 872-1153

Fig. 6 Cuts through: (a) the source image, (b) MLE-15 image, (c) MLE-30 image (d) MLE-60 image, (e) FBP image

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