

Maximum Likelihood Reconstruction for Emission Tomography

L. A. SHEPP AND Y. VARDI

Abstract—Previous models for emission tomography (ET) do not distinguish the physics of ET from that of transmission tomography. We give a more accurate general mathematical model for ET where an unknown emission density $\lambda = \lambda(x, y, z)$ generates, and is to be reconstructed from, the number of counts $n^*(d)$ in each of D detector units d . Within the model, we give an algorithm for determining an estimate $\hat{\lambda}$ of λ which maximizes the probability $p(n^*|\lambda)$ of observing the actual detector count data n^* over all possible densities λ .

Let independent Poisson variables $n(b)$ with unknown means $\lambda(b)$, $b = 1, \dots, B$ represent the number of unobserved emissions in each of B boxes (pixels) partitioning an object containing an emitter. Suppose each emission in box b is detected in detector unit d with probability $p(b, d)$, $d = 1, \dots, D$ with $p(b, d)$ a one-step transition matrix, assumed known. We observe the total number $n^* = n^*(d)$ of emissions in each detector unit d and want to estimate the unknown $\lambda = \lambda(b)$, $b = 1, \dots, B$. For each λ , the observed data n^* has probability or likelihood $p(n^*|\lambda)$. The EM algorithm of mathematical statistics starts with an initial estimate λ^0 and gives the following simple iterative procedure for obtaining a new estimate $\hat{\lambda}^{\text{new}}$ from an old estimate $\hat{\lambda}^{\text{old}}$, to obtain $\hat{\lambda}^k$, $k = 1, 2, \dots$,

$$\hat{\lambda}^{\text{new}}(b) = \hat{\lambda}^{\text{old}}(b) \sum_{d=1}^D \frac{n^*(d)p(b, d)}{\sum_{b'=1}^B \hat{\lambda}^{\text{old}}(b')p(b', d)}, \quad b = 1, \dots, B.$$

We show that the likelihood strictly increases at each step (unless it is already a maximum), $p(n^*|\lambda^k) \geq p(n^*|\lambda^{k-1})$, $k \geq 1$; the total number $\sum \hat{\lambda}(b)$ of estimated counts is equal to the total number $\sum n^*(d)$ of observed counts at each step $\hat{\lambda}^k$, $k \geq 1$; $\hat{\lambda}^k$ converges as $k \rightarrow \infty$ to an estimate $\hat{\lambda}^\infty$ which has maximum likelihood; $\hat{\lambda}^k(b) \geq 0$; $l(\lambda) = \log p(n^*|\lambda)$ is concave.

We show by simulation that this algorithm reduces the statistical noise artifact over conventional convolution backprojection algorithms without introducing excessive smoothing. This is important because statistical noise is a major limiting factor in emission tomography. The method is generally applicable to any design geometry (single or multi-ring), to single or double photon emission tomography, and can incor-

porate timing information and correct for positron range and angle effects in a simple way. We discuss methods to speed convergence of the computation.

I. INTRODUCTION

IN emission tomography (ET) [1]–[3], a compound containing a radioactive isotope is introduced into the body and forms an unknown emitter density $\lambda(x, y, z) \geq 0$ under the body's metabolism. Emissions then occur according to a Poisson process with rate $\lambda(x, y, z)$. In the usual case of positron emitter, the positron emitted finds a nearby electron and annihilates with it. A pair of X-ray photons fly off from the point of annihilation at the speed of light in opposite directions along a line which is uniformly distributed in angle. There is an array of discrete detector elements surrounding the body and the two photons are detected in coincidence by a pair of detector elements defining a detector unit, or detector tube, d . It is known only that somewhere along the length of the tube d the annihilation took place. The case of single X-ray photon emitter is similar, but here collimation is used to define the detector tube.

The measured data are then $n^*(1), \dots, n^*(D)$, where $n^*(d)$ is the total number of coincidences counted in the d th detector tube, and the problem is to estimate the emission density $\lambda(x, y, z)$ from the data n^* . For purpose of display and for purpose of machine computations, we discretize the density λ into boxes $b = 1, \dots, B$, at the outset. Thus, in each box b there is an unknown count $n(b)$ with mean $\lambda(b) = En(b)$, $b = 1, \dots, B$ and our problem is to estimate $\lambda(b)$, or roughly, to guess the true unobserved count $n(b)$ in each box, from the observed data $n^*(d)$, $d = 1, \dots, D$.

A (nearly perfect) mathematical model of the above physics is the following one. Denote by $\lambda(b)$ the integral of $\lambda(x, y, z)$ over box b and let a Poisson distributed number $n(b)$ with

Manscript received May 14, 1982.

The authors are with Bell Laboratories, Murray Hill, NJ 07974.

mean $\lambda(b)$ be generated independently in each box,

$$P(n(b) = k) = e^{-\lambda(b)} \frac{\lambda(b)^k}{k!}, \quad k = 0, 1, \dots \quad (1.1)$$

Suppose that each emission in box b is detected in tube d with known probability

$$p(b, d) = P(\text{detected in } d \mid \text{emitted in } b) \quad (1.2)$$

so that $p(b, d) \geq 0$. Thus, the probability of an emission in b being detected at all is given by

$$p(b) = \sum_{d=1}^D p(b, d) \leq 1. \quad (1.3)$$

The transition matrix $p(b, d)$ is assumed exactly known from the detector array geometry. After each emission moves to some detector tube, or is missed and undetected, there is a known total number $n^*(d)$ of counts in each tube $d = 1, \dots, D$. The probability $P(n^* | \lambda)$ of observing $n^*(d)$, $d = 1, \dots, D$ is a function $L(\lambda)$ of the unknown rate $\lambda = \lambda(b)$, $b = 1, \dots, B$. We want to choose an estimate $\hat{\lambda}$ of λ to maximize $L(\lambda)$. Such an estimate $\hat{\lambda}$ is called a maximum likelihood estimate of λ given n^* .

We begin by observing that it is much simpler and involves no loss of generality to assume that equality holds in (1.3). Indeed, if we let $\lambda'(b)$ represent the mean of $n'(b)$ = the number of emissions in b which are actually detected in some detector, then $p'(b, d) = p(b, d)/p(b)$ is the (conditional) probability that a detected photon emitted in b is detected in d and satisfies

$$\sum_{d=1}^D p'(b, d) = 1. \quad (1.3')$$

Since a Poisson process with rate $\lambda'(b)$, $b = 1, \dots, B$ is obtained from the original process $\lambda(b)$, $b = 1, \dots, B$ by thinning according to detection, it is clear ([4] and Section II) that $\lambda'(b) = \lambda(b)p(b)$ and $p(n^* | \lambda) = p(n^* | \lambda')$. Thus, if $\hat{\lambda}$ is a maximum likelihood estimate of λ then $\hat{\lambda}'$ is a maximum likelihood estimate of $\lambda' = \lambda'(b) = \lambda(b)p(b)$ and conversely. Henceforth, we will assume that equality holds in (1.3). This is equivalent to thinking of $\lambda(x, y, z)$ as the density of emitted counts which are detected.

The reader may have observed that the above model is only nearly exact. The assumption that emissions follow Poisson statistics (1.1) seems beyond challenge, but there is ambiguity in the discretization into boxes and how $p(b, d)$ is to be determined. This ambiguity seems to be unavoidable and is certainly less crude than the approximation required to force the model of transmission tomography to fit the emission case as we will see. We will discuss in Sections II and III how $p(b, d)$ is to be determined.

We also neglected two assumptions in describing the physics. First, the positron has nonzero (a few mm) range before annihilation and the angle between the paths of the two photons is (a few degrees) less than 180° unless the positron is exactly at rest. However, we show in Section III that these assumptions can readily be incorporated into the choice of $p(b, d)$ and so may not be a serious problem. The second assumption was to

neglect the fact that either or both photons can be deflected (a scattering error), or two nearly simultaneous annihilations can have exactly one photon detected (an accidental error). In each case the wrong detector tube gets incremented. Scattered and accidental count errors are well-known [1]–[3], and depend on the absorption density of the body and so cannot be incorporated into $p(b, d)$ in an exact way. Ideally, these errors should be eliminated in the physical measurement procedure, e.g., by energy discrimination of scattered counts and by timing discrimination of accidental counts. We shall simply be forced to ignore these problems and to assume that the only source of difficulty is in the statistical fluctuations in the counting statistics n^* .

We begin the mathematical discussion by noting that the variables $n^*(d)$ are independent and Poisson with expectation $\lambda^*(d)$,

$$\lambda^*(d) = E n^*(d) = \sum_{b=1}^B \lambda(b) p(b, d). \quad (1.4)$$

This fact, pointed out by Snyder [4], is seen immediately by noting that

$$n^*(d) = \sum_{b=1}^B n(b, d) \quad (1.5)$$

where $n(b, d)$ is the number of emissions in b detected in d . Indeed, the variables $n(b, d)$, $b = 1, \dots, B$, $d = 1, \dots, D$ are all mutually independent Poisson variables and so the disjoint sums in (1.5) are also independent and Poisson and (1.4) follows from (1.5) by taking expectations. The maximum value of $\exp(-\lambda) \lambda^n / n!$ over λ occurs at $\lambda = n$ so that the maximum likelihood estimator of $\lambda^*(d)$ given $n^*(d)$ is $\hat{\lambda}^*(d) = n^*(d)$, $d = 1, \dots, D$. Now one might attempt to simply solve the simultaneous equations (1.4) for $\lambda = \lambda(b)$ with $n^*(d)$ substituted for $\lambda^*(d)$, and $p(b, d)$ known, to obtain an estimate of λ . Because it is approximately true that, for a single ring planar detector array,

$$\lambda^*(d) = \iint_{T(d)} \lambda(x, y) dx dy \quad (1.6)$$

where $T(d)$ is the d th tube and $p(b, d)$ is approximately constant along $T(d)$, (1.4) are more or less the same equations as in transmission tomography and that mathematical apparatus is usually used in ET [1]–[3]. (See also the remark at the end of Appendix I in this connection.)

We emphasize that there are two problems with this approach. The first, and less serious, problem is that $\lambda^*(d)$ is actually given by

$$\lambda^*(d) = \iint_{T(d)} \lambda(x, y) p((x, y), d) dx dy \quad (1.6')$$

where $p((x, y), d)$ is the probability that a line through (x, y) , at a uniformly distributed random angle, will lie inside tube $T(d)$. The function $p((x, y), d)$ is a complicated function of (x, y) inside $T(d)$, vanishing along the sides of the tube

$T(d)$ and not constant. Thus, it is not quite correct to use an algorithm developed for strip integrals (1.6) to invert (1.6)'. Indeed, in the experiments of Section IV where ET data are generated instead of transmission data, as described there, subtle artifacts are apparent which seem to be explainable on the basis of the discrepancy between (1.6) and (1.6)'. Nevertheless, because of the low resolution (≈ 1 cm) obtained to date in ET, this problem hardly justifies much effort to correct, since the artifacts introduced are small. The second problem with using transmission algorithms to invert (1.4) for λ is that at the low count rates intrinsic to ET, this procedure amplifies the statistical noise in the estimate of λ over that obtained by the maximum likelihood procedure. We show in Section IV by using simulation, that the maximum likelihood procedure results in a reconstruction $\hat{\lambda}$ which has evidently less noise without excessive smoothing. Our explanation for this increased noise effectiveness is that $n^*(d) \approx 10^2$ or 10^3 and (1.4) are nearly singular.

II. MAXIMUM LIKELIHOOD ESTIMATION OF λ GIVEN n^*

Since the variables $n(b, d)$ = the number of emissions in box b detected in tube d are independent Poisson variables with mean

$$En(b, d) = \lambda(b, d) = \lambda(b)p(b, d) \quad (2.1)$$

the likelihood function is given by

$$L(\lambda) = P(n^* | \lambda) = \sum_A \prod_{d=1, \dots, D} \prod_{b=1, \dots, B} e^{-\lambda(b, d)} \frac{\lambda(b, d)^{n(b, d)}}{n(b, d)!} \quad (2.2)$$

where the sum is over all arrays A of $n(b, d)$'s with

$$n^*(d) = n(\cdot, d) \stackrel{\text{def}}{=} \sum_{b=1}^B n(b, d), \quad d = 1, \dots, D \quad (2.3)$$

observed counts in the d th detector tube. Note also that

$$n(b) = n(b, \cdot) \stackrel{\text{def}}{=} \sum_{d=1}^D n(b, d) \quad (2.4)$$

is the (unobservable) number of emissions in the b th box. The sum in (2.2) is very complicated, but we will show that

$$l(\lambda) = \log L(\lambda) \quad (2.5)$$

is concave as a function of $\lambda = \lambda(1), \dots, \lambda(b)$ because of special properties of the Poisson model. We first show that the partial derivative of $l(\lambda)$ with respect to $\lambda(b_0)$ is

$$\frac{\partial l(\lambda)}{\partial \lambda(b_0)} = -1 + \sum_{d=1}^D \frac{n^*(d)p(b_0, d)}{\sum_{b'=1}^B \lambda(b')p(b', d)}. \quad (2.6)$$

By direct differentiation in (2.2),

$$\begin{aligned} \frac{\partial L(\lambda)}{\partial \lambda(b_0)} &= \sum_A \left\{ \prod_{d=1, \dots, D} \prod_{b=1, \dots, B} e^{-\lambda(b, d)} \frac{\lambda(b, d)^{n(b, d)}}{n(b, d)!} \right\} \\ &\quad \cdot \left\{ -1 + \frac{1}{\lambda(b_0)} \sum_{d=1}^D n(b_0, d) \right\}. \end{aligned} \quad (2.7)$$

By definition of A on (2.2), the conditional expectation

$$E[n(b_0) | n^*, \lambda] = \frac{1}{P(n^* | \lambda)} \sum_A \prod_{d=1, \dots, D} \prod_{b=1, \dots, B} e^{-\lambda(b, d)} \frac{\lambda(b, d)^{n(b, d)}}{n(b, d)!} n(b_0). \quad (2.8)$$

Comparing the second term in (2.7) with (2.8) and using (2.4) we get

$$\begin{aligned} \frac{\partial l(\lambda)}{\partial \lambda(b_0)} &= \frac{1}{L(\lambda)} \frac{\partial L(\lambda)}{\partial \lambda(b_0)} = \frac{1}{P(n^* | \lambda)} \\ &\quad \cdot \left\{ -P(n^* | \lambda) + \frac{P(n^* | \lambda)}{\lambda(b_0)} E[n(b_0) | n^*, \lambda] \right\} \\ &= -1 + \frac{1}{\lambda(b_0)} \sum_{d=1}^D E[n(b_0, d) | n^*, \lambda]. \end{aligned} \quad (2.9)$$

Now because of (2.3) and the fact that $n(b, d)$ are all mutually independent, the d th term in the sum in (2.9) is

$$\begin{aligned} E[n(b_0, d) | n^*, \lambda] &= E[n(b_0, d) | n^*(d), \lambda] \\ &= \frac{n^*(d)\lambda(b_0, d)}{\sum_{b'=1}^B \lambda(b', d)} \end{aligned} \quad (2.10)$$

since for independent Poisson variables X, Y with means λ_X, λ_Y , $E[X | X + Y] = (X + Y)\lambda_X / (\lambda_X + \lambda_Y)$. Equation (2.6) now follows from (2.1) and (2.10). From (2.6) we see that

$$\frac{\partial^2 l(\lambda)}{\partial \lambda(b_0) \partial \lambda(b_1)} = - \sum_{d=1}^D \frac{n^*(d)p(b_0, d)p(b_1, d)}{\left[\sum_{b'=1}^B \lambda(b')p(b', d) \right]^2}. \quad (2.11)$$

It follows immediately from (2.11) that for any $z(1), \dots, z(B)$

$$\begin{aligned} \sum_{b_0=1}^B \sum_{b_1=1}^B z(b_0)z(b_1) \frac{\partial^2 l(\lambda)}{\partial \lambda(b_0) \partial \lambda(b_1)} &= - \sum_{d=1}^D n^*(d)c^2(d) \\ c(d) &= \frac{\sum_{b=1}^B z(b)p(b, d)}{\sum_{b'=1}^B \lambda(b')p(b', d)} \end{aligned} \quad (2.12)$$

and since $n^*(d)c^2(d) \geq 0$, the quadratic form is negative semi-definite so that $l(\lambda)$ is concave.

We now describe an iterative scheme, based on the EM algorithm [5, Appendix I] for maximizing l . We start the algorithm with an initial guess $\lambda^0(b)$ satisfying $\lambda^0(b) > 0$, $b = 1, \dots, B$, and then in each iteration, if $\hat{\lambda}^{\text{old}}(b)$ denotes the current estimate of $\lambda(b)$, the new estimate is defined as

$$\hat{\lambda}^{\text{new}}(b) = \hat{\lambda}^{\text{old}}(b) \sum_{d=1}^D \frac{n^*(d)p(b, d)}{\sum_{b'=1}^B \hat{\lambda}^{\text{old}}(b')p(b', d)} \quad b = 1, \dots, B. \quad (2.13)$$

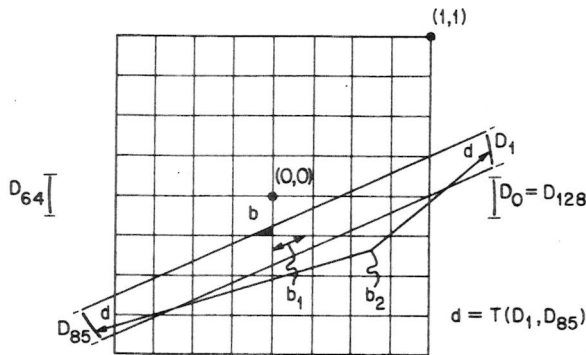


Fig. 1. There are $B \approx \pi/4 \times 128^2$ boxes b inside the circle of radius 1 (patient circle); each box has side $2/128$. The detector circle of radius $\sqrt{2}$ has 128 equidivisions into detectors. A pair of detectors, such as D_1, D_{85} , defines a detector unit, or a tube, d .

(The algorithm could never lead to a quotient of a positive numerator divided by a zero denominator; zero divided by zero is defined as zero.) In the implementation of (2.13), discussed in Section III, we usually use a uniform constant as the initial estimate $\hat{\lambda}^0$. As a point of philosophical interest, the choice of the initial estimate $\hat{\lambda}^0$ is somewhat akin to the choice of a Bayes prior but there is actually no prior measure. For continuity we defer to Appendix I the proof, familiar from the EM algorithm [5], that in each step the new estimate $\hat{\lambda}^{\text{new}}$ is an improvement over the old estimate $\hat{\lambda}^{\text{old}}$:

$$L(\hat{\lambda}^{\text{new}}) \geq L(\hat{\lambda}^{\text{old}}), \quad (2.14)$$

with equality if and only if $L(\hat{\lambda}^{\text{old}}) = \max_{\lambda} L(\lambda)$. It then follows from the concavity of $\log L(\lambda)$ that all maxima of $L(\lambda)$ are global maxima and so the iteration scheme (2.13) gives a sequence

$$\hat{\lambda}^0, \hat{\lambda}^1, \dots \quad (2.15)$$

which converges to a global maximum likelihood estimator $\hat{\lambda}^{\infty}$. Of course if the maximum of $L(\lambda)$ is not unique then by the concavity of $\log L(\lambda)$ any convex combination of maxima is a maximum and $\hat{\lambda}^{\infty}$ depends on the initial choice $\hat{\lambda}^0$.

It is immediate from (2.13) that

$$\sum_{b=1}^B \hat{\lambda}^{\text{new}}(b) = \sum_{d=1}^D n^*(d) \quad \text{and} \quad \hat{\lambda}^{\text{new}}(b) \geq 0 \quad (2.16)$$

so that the true total number of counts is automatically preserved in the estimate $\hat{\lambda}^{\text{new}}$, and $\hat{\lambda}^{\text{new}}(b) \geq 0$ at each stage. This does not hold for the conventional algorithms of transmission tomography adapted to ET.

III. CHOICE OF $p(b, d)$

How should we actually choose $p(b, d)$, the probability that an emission in b is detected in d (see Fig. 1)? Detector unit d may be thought of as a tube defined by the two detectors D_1 and D_{85} at opposite ends of the strip $T = T(D_1, D_{85})$ in Fig. 1 and so $p(b, d)$ is a measure of the angle-of-view of box b into tube d and is the weighted average value of $p((x, y), d)$ [defined after (1.6)'] over the box b , the weight function being $\lambda(x, y)$. Unfortunately, $\lambda(x, y)$ is unknown and so $p(b, d)$ is impossible to compute exactly. Various approximations sug-

gest themselves; what must be made is a choice of the weighting distribution in computing the average of $p((x, y), d)$. Fortunately, we find in Section IV that two such choices lead to reconstructions which do not differ significantly in our simulation.

We show next how it is easy to incorporate known facts or assumptions about positron range and angle to modify $p(b, d)$ to correct for these effects. Indeed suppose it is assumed that positrons range uniformly over a circle of fixed radius ρ (which may be varied) after emission and then annihilate. Then in Fig. 1 there is a certain calculable probability $p(b_1, d)$ that a positron emitted say uniformly in b_1 will range into and cause a count in d . More generally, a positron emitted in b_2 in Fig. 1 may have nonzero velocity and due to conservation of momentum have angle uniformly distributed over some interval depending on the range location, leading at least in principle to an estimate for $p(b_2, d)$. We have not tried this but just want to point out its feasibility and easy incorporation into the Poisson model. It is also worth pointing out that, if as predicted by Ter-Pogossian [3], it soon becomes possible by time-of-flight measurements to know to within say 6 cm where along $T(d)$ the firing took place then this information can also be used. Indeed we could partition each tube d into subtubes d_j , $j = 1, \dots, m$ and compute probabilities $p(b, d_j)$ that an emission in b be detected in detector tube d_j . No essential change in the algorithm (2.13) need be used. It should be mentioned, however, that each of these corrections or modifications costs additional computational effort because the array of nonzero $p(b, d)$ values is made greater in each case and the computational effort depends on how many of the $p(b, d)$ constants are nonzero.

How should the computation (2.13) proceed for maximum efficiency? Ideally the constants $p(b, d)$ should be computed once and stored but since there are BD of these and each of B and D is about 10^4 , this is somewhat inconvenient. Storing the nonzero $p(b, d)$ values is more convenient but the programming demands more effort. We decided to compute $p(b, d)$ anew, on the fly, each time it is needed, as is clear in the program supplied in Appendix II. For this purpose we used $p(b, d)$'s which were easy to calculate, for most of our runs. We chose $p(b, d) = (2nR)^{-1} \times (\text{width of the intersection of the circle of radius } R \text{ about the center of } b \text{ and the strip defined by the tube } d)$, where R is a parameter which we usually take to be the radius of the inscribed circle to b , and n is the number of detector elements around the detector circle. Note $\sum_{d=1}^D p(b, d) = 1$ and $p(b, d)$ is exact for the following imaginary positron range physics. Suppose a positron emitted in b immediately runs to the center of b then runs a distance ρ with probability density $P(\rho) = \rho/(R\sqrt{R^2 - \rho^2})$, $0 < \rho < R$, in a random direction, to reach a point Q in the disk of radius R about the center of b . Upon reaching Q the positron annihilates and the photon pair chooses a line through Q , not at random, but uniformly over the n tube directions. As we see in the simulations of Section IV this choice of $p(b, d)$ produces remarkably accurate reconstructions when compared to the actual box-counts $n(b)$, and are not much different than if $p(b, d)$ are calculated by the more accurate method using the angle-of-view from the center of b into d . Due to cost of

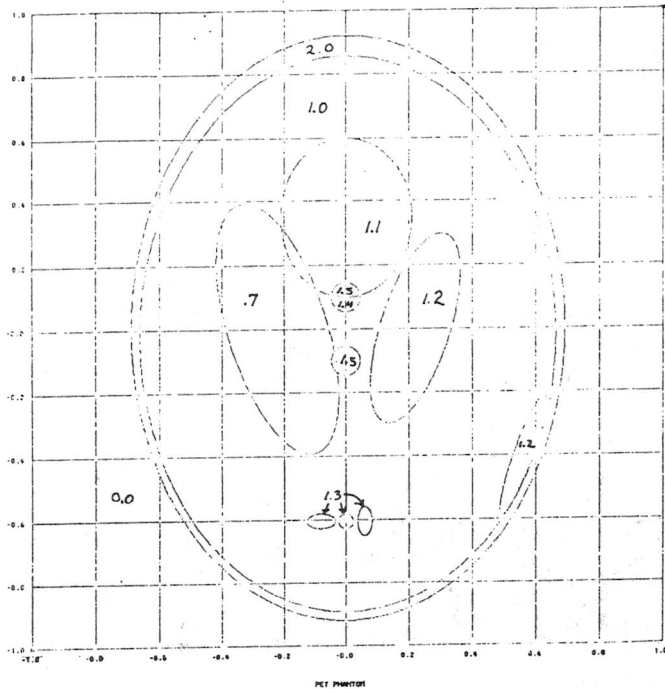


Fig. 2. ET phantom used in the experiments in Section IV. We have modified the phantom of [8] in gray levels to have more contrast and to represent a possible emission density $\lambda(x, y)$. Note, however, that the skull would ordinarily absorb little emitter and has been added simply to define a boundary and for comparison with [8]. C. L. Mallows points out, correctly, that adding the skull as we have done is inadvertently slightly unfair to the convolution-backprojection approach because the artifacts in Figs. 8 and 9 exterior to the skull would not appear for a more relevant phantom.

computation we were unable to use a still more accurate choice for $p(b, d)$ where the emission point is distributed uniformly over b rather than just at the center. However, we are encouraged by the results above that the reconstruction does not depend too critically on the choice of $p(b, d)$ assuming that it is reasonable. We hope to study other more accurate choices for $p(b, d)$. The choice of $p(b, d)$ seems to be less critical than for the ART algorithm of transmission tomography [8] where the corresponding weights are seen to play an important role. It may be that this is because $p(x, y, d)$ vanishes along the sides of the tubes and so even center-weighting does not produce large errors as in [8]. Of course for a more complex or a three-dimensional array of detectors this choice of $p(b, d)$ would be computed once and stored. We feel that our results show that similar quality reconstructions could be expected in any such case.

Of course, (2.13) is not the only way to find a maximum likelihood estimator but is the EM algorithm in this case [5] as we see in Appendix I. We plan to study alternate methods to maximize $l(\lambda)$ based on the gradient since this is available from (2.6) and steepest ascent methods would probably be faster [9]. However, properties (2.14) and (2.16) should not be given up lightly.

The present algorithm (2.13) is in several ways reminiscent of the so-called multiplicative ART [6]. However, our choice of constants $p(b, d)$ is certainly different and related to the actual physics of the problem, as we have seen in Section I. Further, our algorithm itself can be derived from the Poisson

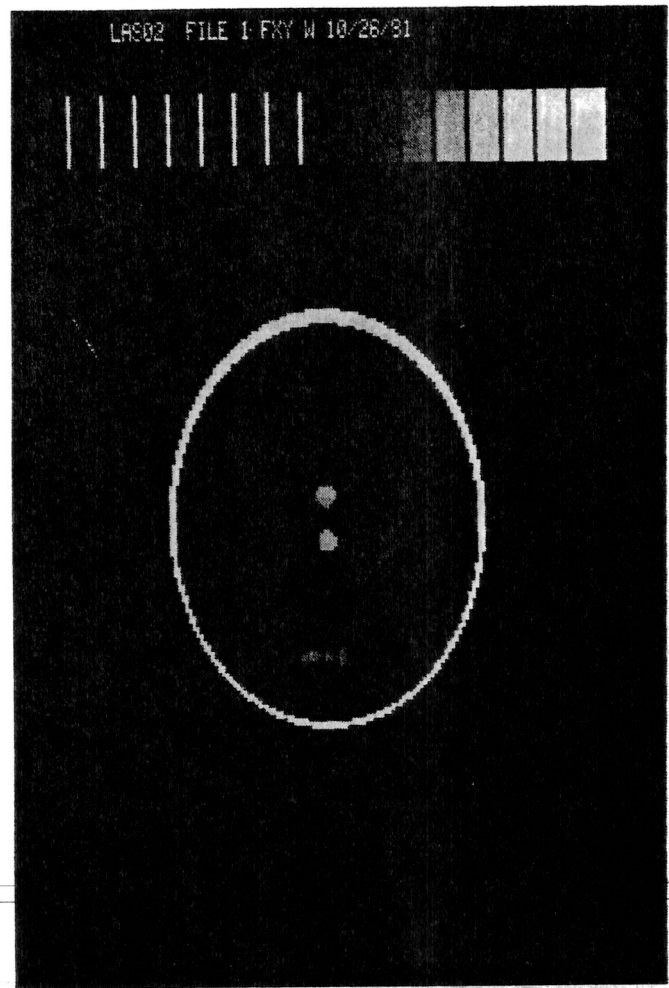


Fig. 3. Photodisplay of Fig. 2.

model as is done in Section II. The algorithm (2.13) has no advantage in transmission tomography, since the use of (1.6)' over (1.6) is then incorrect. It is likely to have advantage over transmission algorithms for emission tomography only at low count rates, or when the geometry of the detector array is so complex that transmission algorithms are hard to obtain.

IV. SIMULATION EXPERIMENTS

We study only two-dimensional, or single-ring ET and assume that there are $B = 128 \times 128$ boxes surrounding the unit circle (the patient circle). We assume 128 discrete detectors equally spaced around the circle of radius $\sqrt{2}$ circumscribing the display boxes. We use the phantom in Figs. 2 and 3 with density $\lambda(x, y)$ and choose 10^7 random points from the density $\lambda(x, y)/\iint \lambda$ independently. Each point is chosen by taking a uniform random point in the square $|x|, |y| \leq 1$ and accepting it with probability $\lambda(x, y)$. This gives the emission points as if $\lambda(x, y)$ were the true emission density. The histogram of the number of emissions in each box is displayed in Fig. 4. After choosing each point we choose a random line L through the point and increment the count for the detector unit corresponding to the two detector intervals that L passes through. This gives the numbers $n^*(d)$, $d = 1, \dots, D \approx 64 \times 65$ since there are 65 detector intervals opposite each one. Note that

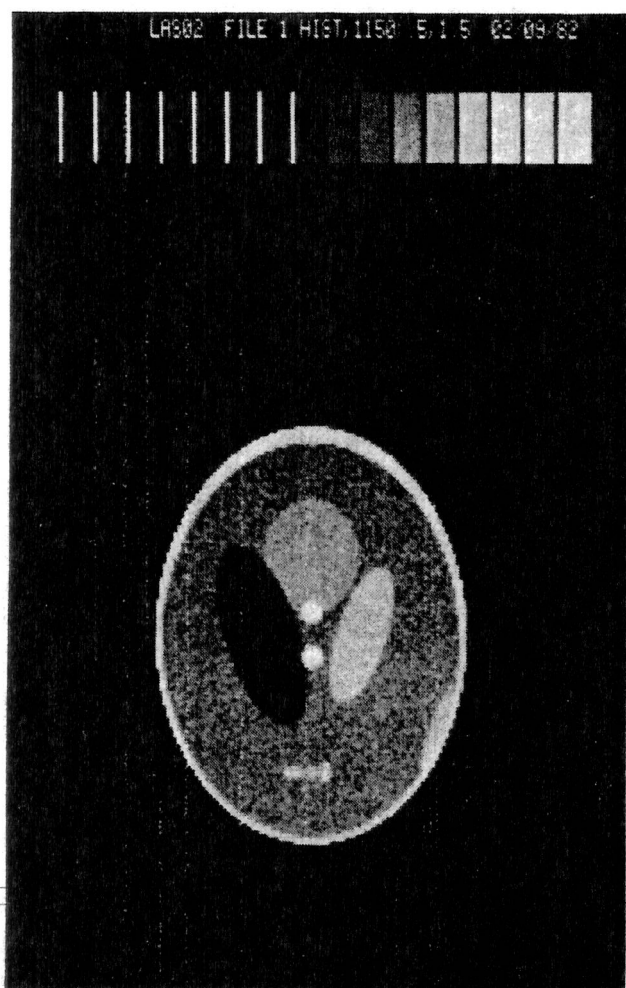


Fig. 4. Histogram $n(b)$ of the 10^7 counts drawn from the phantom of Fig. 2 at a rate proportional to $\lambda(x, y, z)$ at each point. It is felt that 10^7 would be about the right value for single ring positron ET in practice. Of course, without measuring exact time-of-flight, $n(b_0)$ is not observable except in a simulation.

$B \approx 128^2 \pi/4$ since boxes outside the unit circle have known emitter density zero.

The data $n^*(d)$ are fed into the computer program in Appendix II which is included for repeatability purposes and the reconstruction (2.13) with 32 iterations is displayed in Fig. 5. It is remarkably accurate as is seen by comparing the numbers on the line plots, Figs. 6 and 7, of the original histogram (Fig. 6) and the reconstruction (2.13) with the constants of the program in Appendix II along the line $y = 0$. The same data $n^*(d)$ are then fed into the usual convolution-backprojection reconstruction in fan-beam configuration [7] with the Lakshminarayanan filter [7]. The reconstruction is displayed in Figs. 8 and 9 and is clearly noisier than that of (2.13) (Figs. 5 and 7). It also has a cupping artifact (the dip close to the skull) which is perhaps due to the error of using (1.6) instead of (1.6)' as discussed in Section I. Although smoothing with other filters could be used to reduce noise in Figs. 8 and 9 to the level of that in Figs. 5 and 7, we feel that this might result in excessive smoothing. It is interesting that (2.13) is a nonlinear algorithm and smoothing is a linear procedure. The maximum

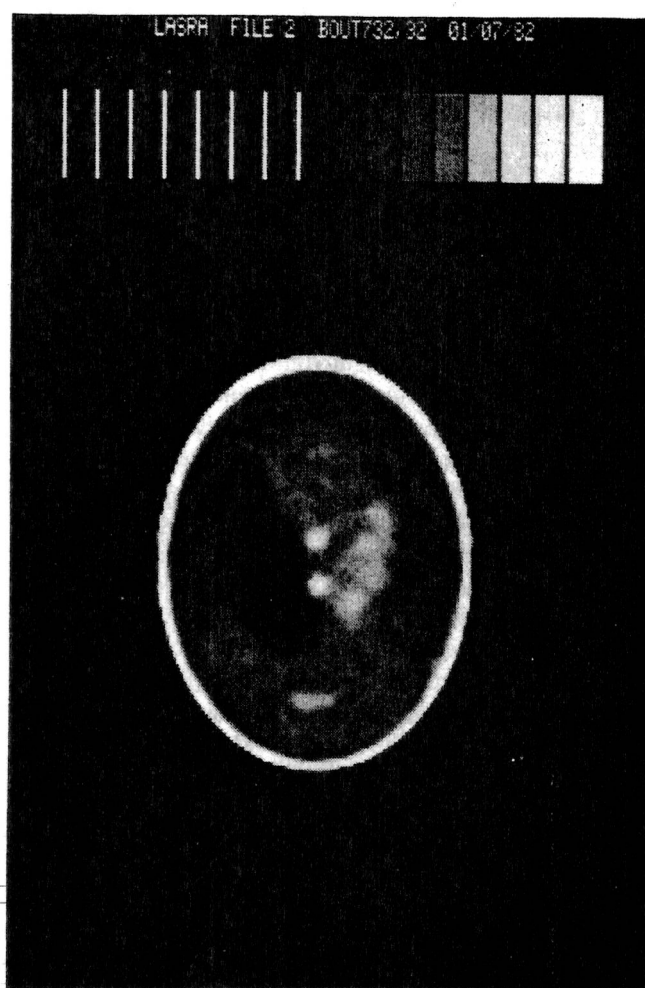


Fig. 5. Reconstruction by (2.13) with $p(b, d)$ as in Section III by the program of Appendix II after 32 iterations.

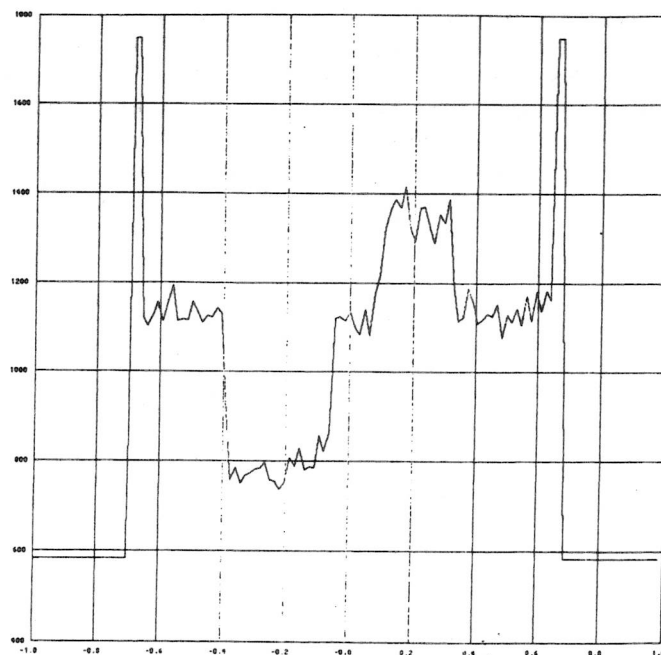


Fig. 6. Line plot of the histogram Fig. 4 through $y = 0$ (the x-axis).

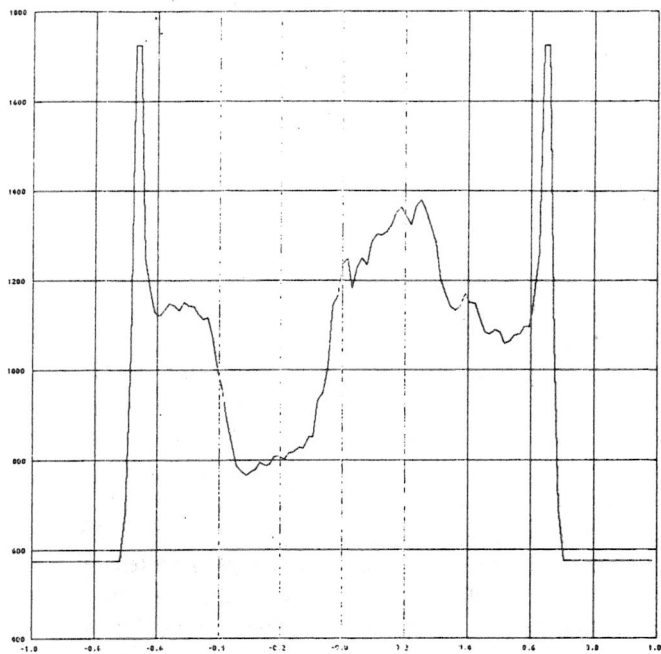


Fig. 7. Line plot of Fig. 5 along $y = 0$ (the x -axis). Note the quantitative agreement between Figs. 6 and 7.

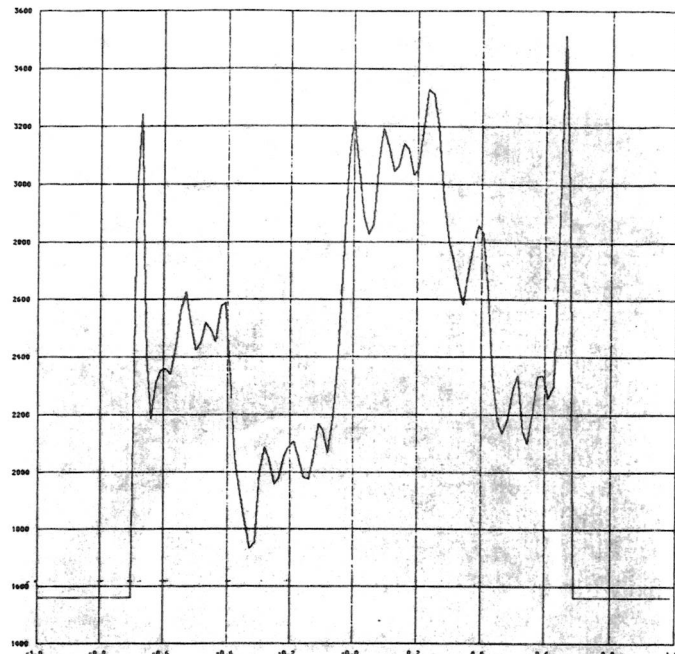


Fig. 9. Line plot of Fig. 8. Here the dip inside the skull is easy to see. Note that the actual numerical values have no significance and some are negative although they have been clipped in this display.

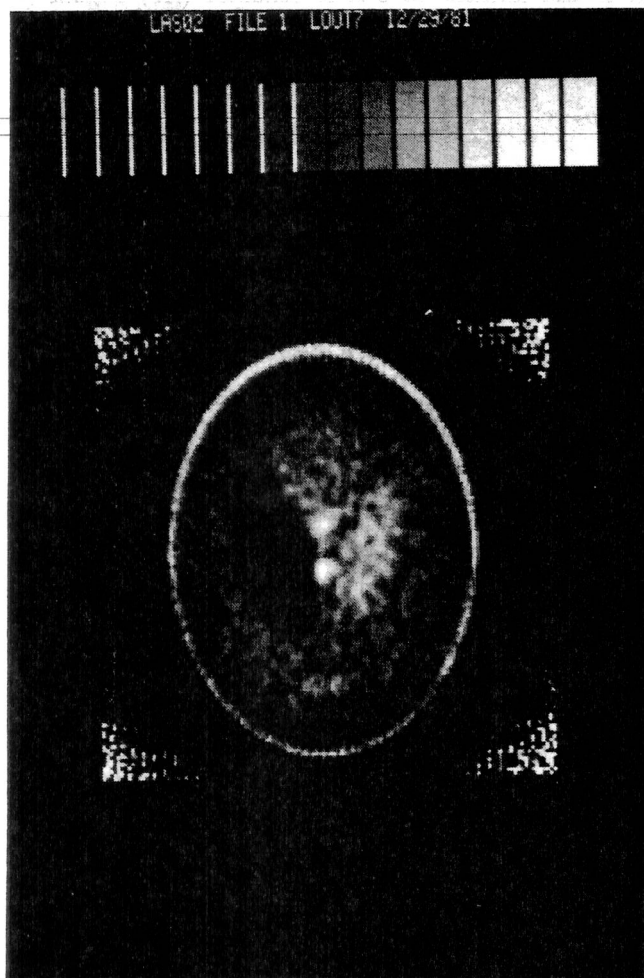


Fig. 8. Reconstruction by convolution backprojection from same data n^* . Note the increased noise and cupping artifact inside the skull.

likelihood estimator thus seems to have good noise suppression over linear algorithms.

In order to test how critical the choice of constants $p(b, d)$ is in (2.13), we took an extreme case where $p(b, d)$ is $1/\pi \times$ (angle of view into detector d from the center of b). This involved an enormous amount of computation because $p(b, d)$ had to be calculated from the relatively complex subroutine determining angle-of-view each time it was needed. The 14th iteration is displayed in Figs. 10 and 11 and are seen to be nearly identical to Figs. 5 and 7. Of course, if $p(b, d)$ could be computed and stored to avoid the need for recomputation, we would expect this choice to be preferable and perhaps uniform weighting in b as discussed in Section III to be still better.

APPENDIX I

We discuss here the relations between (2.13), the EM algorithm [5], and the Kuhn-Tucker (KT) conditions for optimality [9]. Since $l(\lambda)$ of (2.5) is concave, it follows from [9, Theorem 2.19(e)] that sufficient conditions for $\hat{\lambda}$ to be a maximizer of l are the KT conditions which, for our case, turn out to be: for each $b = 1, \dots, B$,

$$0 = \lambda(b) \frac{\partial l(\lambda)}{\partial \lambda(b)} \Big|_{\hat{\lambda}} = -\hat{\lambda}(b) + \sum_{d=1}^D \frac{n^*(d) \hat{\lambda}(b) p(b, d)}{\sum_{b'=1}^B \hat{\lambda}(b') p(b', d)} \quad (A1)$$

and

$$\frac{\partial l(\lambda)}{\partial \lambda(b)} \Big|_{\hat{\lambda}} \leq 0 \quad \text{if } \hat{\lambda}(b) = 0. \quad (A2)$$

Having such a relatively simple expression for the right side of (A1) [which is derived from (2.6)], one could think of many

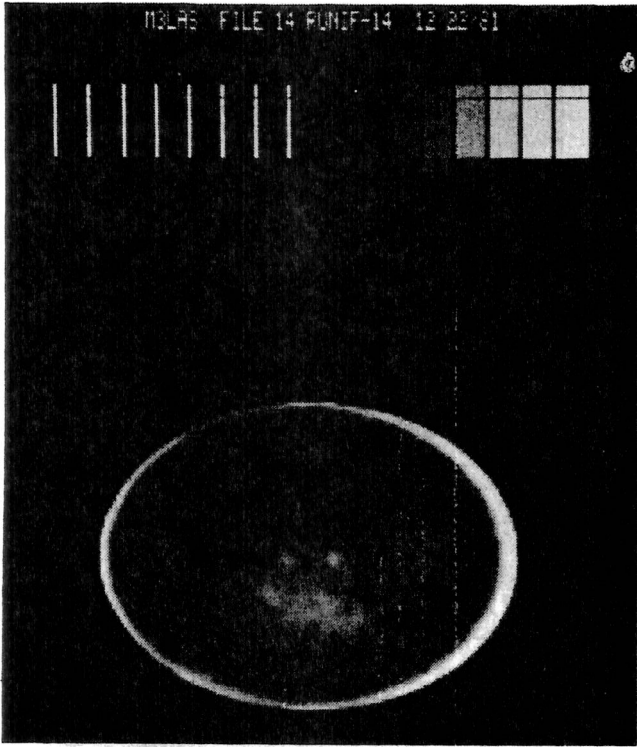


Fig. 10. Reconstruction by (2.13) where $p(b, d) = 1/\pi \times$ angle of view from the center of b into tube d . Almost the same as Fig. 5.

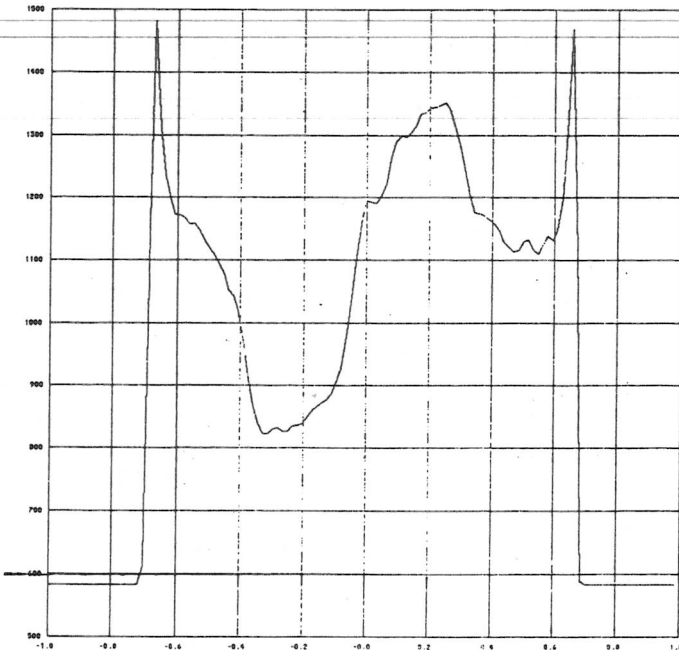


Fig. 11. Line plot of Fig. 10 along $y = 0$ (the x-axis). Almost the same as Fig. 7.

iterative schemes that would converge to a maximum of l . A particularly appealing one is (2.13) because it is a special case of the EM algorithm (E stands for expectation, M for maximization). Here is the rationale behind the EM algorithm, tailored to our example. Had we believed that $\hat{\lambda}^{\text{old}}$ is the true λ , we would estimate the number of annihilations in box b to be $\hat{n}(b) = E[n(b) | \hat{\lambda}^{\text{old}}, n^*]$; this is the expectation step. But, if

$\hat{n}(b)$ is our estimate for the emission count in box b we should, in order to be consistent, estimate the emission density $\lambda(b)$ by $\hat{n}(b)$ (because this is the maximum likelihood estimate if indeed $\hat{n}(b)$ is the emission count in box b); this is the maximization step. This gives the iterative scheme

$$\hat{\lambda}^{\text{new}}(b) = E[n(b) | \hat{\lambda}^{\text{old}}, n^*] \quad b = 1, \dots, B. \quad (\text{A3})$$

As we have seen from (2.10), the right side of (A3) is the right side of (2.13) and so (2.13) is an EM algorithm. Being an EM algorithm, it follows from [5, Theorem 1] that if in (2.13) $\hat{\lambda}^{\text{old}} \neq \hat{\lambda}^{\text{new}}$, then $l(\hat{\lambda}^{\text{old}}) < l(\hat{\lambda}^{\text{new}})$; that is, the algorithm is monotone (the likelihood increases in each step of (2.13), unless $\hat{\lambda}^{\text{old}} = \hat{\lambda}^{\text{new}}$ in which case we have converged). To establish the convergence of (2.13) to a point of maximum we rely on 1) the monotonicity of the algorithm and 2) if $\hat{\lambda}$ is a point of convergence of (2.13) it is a point of maximum (to be proved below). Since the algorithmic map (2.13) is continuous, it follows from convergence [9, Theorem A, p. 91] that indeed the algorithm (2.13) converges to a point of maximum likelihood. (The above argument is similar to the one used in Vardi [10] in another application of the EM algorithm.) To prove 2), we first note that if $\hat{\lambda}$ is a point of convergence then necessarily it satisfies the first KT condition (A1). To see that it also satisfies the second condition (A2), assume (by negation) that there exists a b such that $\hat{\lambda}(b) = 0$ but

$$\left. \frac{\partial l(\lambda)}{\partial \lambda(b)} \right|_{\hat{\lambda}} = -1 + \sum_{d=1}^D \frac{n^*(d)p(b, d)}{\sum_{b'=1}^B \hat{\lambda}(b')p(b', d)} > 0.$$

Now, since $\hat{\lambda}^{(k)} \rightarrow \hat{\lambda}$ we get (because of the continuity of $\partial l / \partial \lambda$) that there exists an $\epsilon > 0$ and k_0 such that for all $k > k_0$

$$\sum_{d=1}^D \frac{n^*(d)p(b, d)}{\sum_{b'=1}^B \hat{\lambda}^{(k)}(b')p(b', d)} > 1 + \epsilon$$

and hence, from (2.13),

$$\hat{\lambda}^{(m+k_0)}(b) \geq \hat{\lambda}^{(k_0)}(b)(1 + \epsilon)^m \rightarrow \infty \quad \text{as } (m \rightarrow \infty)$$

which is a contradiction to (2.16). Note that in the last assertion we assumed $\hat{\lambda}^{(k_0)}(b) > 0$. If this is not the case, so that $\hat{\lambda}^{(k_0)}(b) = 0$, then necessarily b is such that $n^*(d)p(b, d) = 0$ for $d = 1, \dots, D$, and we get

$$\left. \frac{\partial l(\lambda)}{\partial \lambda(b)} \right|_{\hat{\lambda}} = -1 < 0$$

as desired.

Remark: If $\lambda = \lambda^{\text{old}}$ is a solution of (1.14) with n^* substituted for λ^* , i.e.,

$$n^*(d) = \sum_{b=1}^B \lambda^{\text{old}}(b)p(b, d) \quad d = 1, \dots, D \quad (\text{A4})$$

then λ^{old} is a stationary point of (2.13). Nevertheless, because of low count rate variations, (A4) would typically be inconsistent, and hence attempting to solve (A4) would be a futile way of deriving a stationary point of (2.13).

APPENDIX II

The following program reconstructs $B = \lambda$ from $\text{PNP} = n^*$ from (2.13) iteratively. The array PH keeps the denominator in (2.13) and F is used to hold the numerator before updating B . The constants $p(b, d)$ are calculated as W when needed on the fly. The summation over boxes b is by row or column depending on tube direction being lesser or greater than 45° . The calculation of $p(b, d) = W$ is required in calculating the denominator PH and again in calculating the numerator sum to update $B = \lambda$ in each iteration. The $p(b, d)$ have been chosen so that W is calculable by a few addition and maximization operations as described in Section III. The input count data are read in from file 10, the input λ^0 is read in from file 9, and the reconstruction is written out onto file 11.

```

C POSITRON EMISSION RECON WITH B-H ALGORITHM
C I=(I1,I2) INDEXES THE ITH BOX WITH CENTER AT -(I1-2I1)/(2B-1) I=1,2
C B(1)=LAMBDA(1);F(I)=SUM(H(.,K)*C(I,K))/PM(K)
C K=(K1,K2) INDEXES THE KTH TUBE WITH CENTERS AT -(K-1)2PI/RD
C PHP(I)=NO. COUNTS IN TUBE K: PM(K)=SUM(B(J)(C(I,K)))
C NC=COUNTS IN BOX I2, ND=NO. DETECTORS, 5*ND*(ND-1)=NO. TUBES
C T1K,T2K= DISTANCES FROM 0. TO EDGES OF TUBE K, -RD/2<K1<K2<RD
C U=WIDTH OF THE INTERSECTION OF THE KTH TUBE AND THE ITH SPHERE
C RD=RAD. OF DETECTOR CIRCLE DIV. BY RAD. OF PATIENT CIRCLE(=1.)
PARAMETER (N8=128,ND=128,NC=100000,ND2=2*ND,HIER=16)
DIMENSION B(NB,NB),F(100,NB),PHP(NB,ND),PH(ND,NB)
DEFINITION (C(NB),ST(NB2),CT(ND2)
INTGER INH(NB),INX(NB)
C DEFINE CONSTANTS
RHO=1./RB
CC=1./(12.*RHO*ND)
ND=70*NT(2.)
PI=3.14159265
PI2=2.*PI
RND2=.5*RB
RNDPI2=ND/PI2
PI2*ND=12./ND
DO 10 I=1,NB
X(I)=-1.*(I-.5)*2./ND
DO 11 K=1,ND2
TH=5*(K-1)*PI/RD
ST(K)=ST(INH(I),1)
CT(K)=COS(TH)
DO 12 I=1,NB
R2=2*BT(1.-X(I))*X(I)
INX(I)=(1-B2)*.5*RB+.5
INH(I)=NAT(1-((INX(I),1)
2EX(I)=N1+2*TH*.5*RB+.5
JN(I)=SIN(1-((X(I),ND)
12 CONTINUE
TIME=.7*CC.11
N1=55
N2=103
C READ THE RAW "URE DATA
DO 13 I=1,NB
READ(1,102) (PHP(K1,K2),K1=1,ND)
13 CONTINUE
PRINT 1,102,(PHP(1,K2),K1=1,ND)
1002 FORMAT(10F12.1)
CCCCCCCC
C INITIALIZE I2 TO I2ND FILE
DO 229 JR=1,2
DO 331 I2=NR,1,-1
READ(9,1) (I1,I2),(I1,I2),I1=1,NB)
IF(I1,I2)N1,OR,I2=2,N2)PRINT 1,9,(I1,I2),I1=1,NB)
231 CONTINUE
229 CONTINUE
1039 FORMAT(13P12.4)
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
DO 511 ITP=1,NITPP
C UPDATE H(K)=SUM(B(I)*C(I,K))
DO 301 K1=1,ND-1
DO 302 K2=K1+1,ND
PH(K1,K2)=0.
C CK=COS(PHK),SK=SIN(PHK),THK=(ALF+BET)/2 WHERE K*CK+Y*SK=T1K,T2K DEFINE
C THE EDGE LINE OF TUBE K WITH CENTERS AT RDCOSALF,RDSINHALF & AT
C (RDCOSBET,RDSINBET), OK=ALF<BET<2 PI.
CK=V(K1,K2-1)
SK=ST(K1,K2-1)
T1K=RD*CT(K2-K1)
T2K=RD*CT(K2-K1+2)
DK=CK*2./RB
C SKIP IF (X(I1),X(I2)) TOO FAR FROM TUBE K
IF(ABS(CK).LT.ABS(SK))GO TO 305
C NOW RUN THROUGH BOXES BY ROWS
DK=CK*2./RB
DO 303 I2=1,NB
IA=(1.+T2K-RND-X(I2)*SK)/CK)*NR+.1)*.5
IB=(1.+T1K-RND-X(I2)*SK)/CK)*NR+.1)*.5
I1=MAX(0,INH(I2),1+INX(IA,IB))
I2=MIN(INH(I2),NAT(IA,IB))
TIP=CT(I1)*CK+V(I2)*SK-DK*RHO
TIM=TIP-2.*RHO
IF(I1,CT,I2)GO TO 333
DO 304 I1=I1,I2
TIM=TIP+DK
TIE=TIE+DE
U=ANX(I*(TIP-T2K,0.))*ANX(I*(TIP-T1K,0.))-ANX(I*(TIP-T1K,0.))
U=U-ANX(I*(TIP-T2K,.))
PH(K1,K2)=PH(K1,K2)+U*B(1,I2)
PRINT 1003,K1,I2,I1,I2,I1,I2,PH(K1,K2),U,B(1,I2),CK,SK
&11K,T2K,TIP,TIE
&1003 FORMAT(4E9F10.4)
CONTINUE
CONTINUE
GO TO 302
CONTINUE
RUN THROUGH ROWS BY COLS
DK=SK*2./RB
DO 306 I1=1,NB
IA=(1.+T2K-RND-X(I1)*CK)/SK)*NR+.1)*.5
IB=(1.+T1K-RND-X(I1)*CK)/SK)*NR+.1)*.5
I1=MAX(0,INH(I1),1+INX(IA,IB))
I2=MIN(INH(I1),NAT(IA,IB))
TIP=CT(I1)*CK+V(I2)*SK-DK*RHO
TIM=TIP-2.*RHO
IF(I1,CT,I2)GO TO 336
DO 307 I2=I1,I2
TIM=TIP+DK
TIE=TIE+DE

```

```

1010 TIP=TIP+DA
1011 W=ANAL1(TIP-T2K,0.)+ANAL1(TIP-T1K,0.)-ANAL1(TIP-T1K,0.)
1012 W=W-ANAL1(TIP-T2K,0.)
1013 PH(K1,K2)=PH(K1,K2)+W*B(I1,I2)
1014 PRINT 1004,K1,K2,I1,I2,I2,I22,PH(K1,K2),B,5(I1,I2),CK,SK
1015 C 6,T1K,T2K,TIP,T1H
1016 FORMAT(10,9F10.4,' Q')
1017 C
1018 3/7 CONTINUE
1019 3/6 CONTINUE
1020 3/5 CONTINUE
1021 3/4 CONTINUE
1022 DO 321 K2=1,ND
1023 C DO 321 K2=1,ND
1024 C 321 PRINT 1005,(PH(K1,K2),K1=1,ND)
1025 C 1005 FORMAT(10F10.4)
1026 C NOW SET PH(K)=H(.,K)/PH(K)
1027 DO 331 K1=1,ND-1
1028 DO 332 K2=K1+1,ND
1029 IF(PH(K1,K2).GT.TINY)GO TO 334
1030 PRINT 333,PH(K1,K2),K1,K2
1031 C 333 FORMAT(F10.4,2I10,' TINY')
1032 PH(K1,K2)=0.
1033 GO TO 332
1034 PH(K1,K2)=PH(K1,K2)/(PH(K1,K2)/CC)
1035 CONTINUE
1036 CONTINUE
1037 DO 322 K2=1,ND
1038 C DO 322 K2=1,ND
1039 C 322 PRINT 1005,(PH(K1,K2),K1=1,ND)
1040 C FIND F(I)=SUM(H(K)*C(I,K)/PH(K))
1041 DO 400 I2=1,NB
1042 DO 400 I1=INH(I2),INH(I2)
1043 4002 F(I1,I2)=0.
1044 4001 CONTINUE
1045 DO 401 K1=1,ND-1
1046 DO 402 K2=K1+1,ND
1047 C DO 402 K2=K1+1,ND
1048 C 402 SIN(I2K),SIN(I2K)=SIN(I2K)/(1+PH(I2K)/2) WHERE I2K=I*SK+T*SK-T2K DEFINE
1049 C THE HEDGES AT THE END OF THE LIST. THE HEDGES AT NDCOSALF,NDSINALF & AT
1050 C (NDCOSKT,NDSINAKT) ARE <NDCOSKT/2 PI.
1051 CK=T(K1-K2-1)
1052 SK=T(K1-K2-1)
1053 T1K=BD*CY(K2-K1)
1054 T2K=BD*CY(K2-K1+2)
1055 DK=C*K2/ND
1056 C SKIP IF (K(I1),I2) TOO FAR FROM TURN K
1057 IF(ABS(CK).LT.NBS(SK))GO TO 405
1058 C NOW RUN THROUGH BOXES BY ROWS
1059 DK=C*K2/ND
1060 DO 403 I2=1,NB
1061 IA=((1.-T2K-RND-H(I2)*SK)/CK)*NB+1.)*.5
1062 IB=((1.-T1K-RND-H(I2)*SK)/CK)*NB+1.)*.5
1063 I1=HAXO(INH(I2),1+HINO(IA,IB))
1064 I2=HINO(INH(I2),HAXO(IA,IB))
1065 T1P=HINO(CY+H(I2)*SK-DK-RND)
1066 T1H=TIP-2.*RND
1067 IF(I1.GT.I2)GO TO 403
1068 DO 404 I1=I1,I2
1069 T1P=TIP+DK
1070 T1H=TIP+DK
1071 W=ANAL1(TIP-T2K,0.)+ANAL1(T1H-T1K,0.)-ANAL1(TIP-T1K,0.)
1072 W=W-ANAL1(TIP-T2K,0.)
1073 F(I1,I2)=F(I1,I2)+W*PH(K1,K2)
1074 C 1074 PRINT 1003,K1,K2,I1,I2,I2,I22,PH(K1,K2),W,B,5(I1,I2),CK,SK
1075 C 6,T1K,T2K,TIP,T1H
1076 CONTINUE
1077 GO TO 402
1078 405 CONTINUE
1079 C NOW RUN THROUGH BOXES BY COLS
1080 DK=C*K2/ND
1081 DO 406 I1=1,NB
1082 IA=((1.-T2K-RND-H(I1)*CK)/SK)*NB+1.)*.5
1083 IB=((1.-T1K-RND-H(I1)*CK)/SK)*NB+1.)*.5
1084 I2=HAXO(INH(I1),1+HINO(IA,IB))
1085 I22=HINO(INH(I1),HAXO(IA,IB))
1086 T1P=HINO(CY+H(I1)*CK-DK-RND)
1087 T1H=TIP-2.*RND
1088 IF(I2.GT.I22)GO TO 406
1089 DO 407 I2=I2,I22
1090 T1P=TIP+DK
1091 T1H=TIP+DK
1092 W=ANAL1(TIP-T2K,0.)+ANAL1(T1H-T1K,0.)-ANAL1(TIP-T1K,0.)
1093 W=W-ANAL1(TIP-T2K,0.)
1094 F(I1,I22)=F(I1,I22)+W*PH(K1,K2)
1095 C 1095 PRINT 1004,K1,K2,I1,I2,I2,I22,PH(K1,K2),W,B,5(I1,I2),CK,SK
1096 C 6,T1K,T2K,TIP,T1H
1097 CONTINUE
1098 CONTINUE
1099 CONTINUE
1100 CONTINUE
1101 C NOW SET B
1102 SUM=0.
1103 DO 331 I2=NB,1,-1
1104 DO 332 I1=INH(I2),INH(I2)
1105 B(I1,I2)=H(I1,I2)+F(I1,I2)/CC
1106 SUM=SUM+B(I1,I2)
1107 CONTINUE
1108 IF(ITER.NE.8.AND.ITER.NE.NITER)GO TO 431
1109 WRITE(11,1039) (B(I1,I2),I1=1,NB)
1110 IF(C1.DD.N1.0B.I2.DD.42)PRINT 1009,(B(I1,I2),I1=1,NB)
1111 CONTINUE
1112 PRINT 1101,ITER,SUM
1113 FORMAT(I10,F16.5)
1114 CONTINUE
1115 STOP

```

ACKNOWLEDGMENT

L. A. Shepp is grateful to B. Efron for bringing the EM algorithm to his attention in the context of a discussion on ET. Y. Vardi is grateful to D. Schapira for help in programming.

REFERENCES

- [1] T. F. Budinger, S. E. Derenzo, G. T. Gullberg, W. L. Greenberg, and R. H. Huesman, "Emission computer assisted tomography with single-photon and positron annihilation photon emitters," *J. Comput. Assisted Tomogr.*, vol. 1, pp. 131-145, 1977.
- [2] Z. H. Cho, J. K. Chan, and L. Eriksson, "Circular ring transverse axial positron camera for 3-dimensional reconstruction of radionuclides distribution," *IEEE Trans. Nucl. Sci.*, vol. NS-23, pp. 613-622, 1976.
- [3] M. M. Ter-Pogossian, M. E. Raichle, and B. E. Sobel, "Positron

- emission tomography," *Scientific Amer.*, vol. 243, pp. 170-181, Oct. 1980.
- [4] D. L. Snyder, J. T. Lewis, Jr., and M. M. Ter-Pogossian, "A mathematical model for positron-emission tomography systems having time-of-flight measurements," *IEEE Trans. Nucl. Sci.*, vol. NS-28, pp. 3575-3583, 1981.
- [5] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *JRSS*, vol. 39, pp. 1-38, 1977.
- [6] G. T. Herman, A. Lent, and P. H. Lutz, "Relaxation methods for image reconstruction," *Commun. Ass. Comput. Mach.*, vol. 21, pp. 153-158, 1978.
- [7] A. V. Lakshminarayanan, "Reconstruction from divergent ray data," Dep. Comput. Sci., SUNY, Buffalo, NY, Tech. Rep. 92, 1975.
- [8] L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section," *IEEE Trans. Nucl. Sci.*, vol. NS-21, pp. 21-43, 1974.
- [9] W. I. Zangwill, *Nonlinear Programming*. Englewood Cliffs, NJ: Prentice-Hall, 1969.
- [10] Y. Vardi, "Nonparametric estimation in renewal processes," *Ann. Statist.*, vol. 10, no. 3, pp. 772-776, 1982.
-