

Medical Images: Formation, Handling and Evaluation

Edited by

Andrew E. Todd-Pokropek

University College London, Department of Medical Physics
Gower Street, London WC1E 6BT, UK

Max A. Viergever

Utrecht University, Computer Vision Research Group
University Hospital Utrecht E02.222
Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

Science
and
scientific

in conjunction with

Historical
contributions
SI Series.
Publication

Galileo

software

Publishers of



Springer-Verlag
Berlin Heidelberg New York London Paris Tokyo
Hong Kong Barcelona Budapest
Published in cooperation with NATO Scientific Affairs Division

NEW INSIGHTS INTO EMISSION TOMOGRAPHY VIA LINEAR PROGRAMMING

L.A. Shepp
R.J. Vanderbei

AT&T Bell Laboratories
Murray Hill, New Jersey 07974, U.S.A.

ABSTRACT

Suppose that each detector unit, d , of an emission scanner measures a count $n^*(d)$ which represents the number of emissions into d of an unknown emission density λ . The likelihood, $P(n^*|\lambda)$, is the (Poisson) probability of observing n^* under λ . The well-known EM algorithm starts with an estimate λ^0 of λ and produces a sequence λ^k , $k=1,2,\dots$, of estimators having increasing likelihood and which converge to an estimate λ^* with maximum likelihood as $k \rightarrow \infty$. It is well-known that for large k , λ^k becomes noisy or noisy in appearance, and various methods have been proposed to effectively smooth λ^k , but these all give up likelihood to get smoothness. We give a method using linear programming of smoothing λ^k which produces a smoother estimate $\tilde{\lambda}^k$ which nevertheless has the same likelihood as λ^k .

If λ^* is nearly unique among estimators with its likelihood then $\tilde{\lambda}^k$ cannot differ much from λ^k , while if there are many λ with the likelihood of λ^k , then $\tilde{\lambda}^k$ should be smooth. Experiments described here indicate that for large k , $\tilde{\lambda}^k$ is not much smoother than λ^k , so that λ^k seems to be nearly unique. This is surprising in view of the fact the problem considered is severely undetermined in the sense that there are typically 3 or more times as many unknowns as equations (measurements), but of course it is the inequality constraints that cause the uniqueness.

We survey the maximum likelihood approach to emission tomography using the iterative EM algorithm and discuss the known difficulty with the algorithm at high iteration numbers.

1. Introduction

In a typical but especially exciting new application of emission tomography (ET), Petersen et al., [P], assert that the reconstructed

brain ET image varies reproducibly with the word or concepts being considered by the subject's brain during the experiment. Oxygen atoms in radioactive isotope form (^{15}O with half-life = 123sec.) are intravenously injected as H_2O into volunteers who words on a TV screen during an ET scan. This water is metabolically taken up in the active region of the brain where oxygen is needed at that moment. While it is there it decays and emits radioactivity which is detected in a bank of detectors. An image of the emission density is formed by an algorithm (the so-called EM algorithm) for ET reconstruction. A subtraction is made between two images, with and without the visual word stimulus, and, in order to increase signal/noise, several such images from different volunteers are added together after a registration shift and scaling. The resulting image shows [P] that different regions of the brain are active depending on the word flashed. It is difficult to perform this experiment on a single volunteer because of ET's inherently low signal/noise. It is indicated in [P]^{*} that it would be very useful to improve the understanding of ET algorithms so that the need for several volunteers could be avoided and such experiments could be made more efficient. Indeed if these results can be repeated by other groups the technique raises many possible directions for ET applications.

The purpose of this paper is to survey ET and the currently most important algorithm for ET, namely the EM (expectation-maximization) algorithm [SV], [VSK] which was used by [P] to obtain high likelihood estimates of the emission density. Here we focus on one

of the shortcomings of EM in ET and the techniques suggested to avoid this problem. The problem, as shown in §3, is that estimators of very high likelihood tend to be too snowy to use. For newcomers to the field, we point out that ET is different from CAT scanning or transmission tomography [SK] which has much higher resolution (and counts and dose) than ET, but cannot be used for metabolic studies of this type. ET is used mainly to study function while CAT is used clinically.

In ET, there are typically (but not always, see below) a finite set D of detector units (we say units to cover the case of double photons or PET, cf. [SV]). For each $d \in D$ there is a received count $n^*(d) = 0, 1, 2, \dots$ which is the number of counts of radioactivity recorded by unit d during the experiment. The geometric arrangement of the units d and their efficiencies determine the transition probability $p(d|x)$ that an event or emission at x which is detected will be detected in unit d . Thus $p(d|x)$ is known and given and

$$(1.1) \quad \sum_{d \in D} p(d|x) = 1 \quad \text{for each } x.$$

If $\lambda(x)dx$ is the density of, say ^{15}O , emitter at x then the mean number of counts in unit d is

$$(1.2) \quad E n^*(d) = \int \lambda(x)p(d|x)dx \triangleq \mu(d)$$

where the integral is over all of 2 or 3 dimensional space where detected counts can be emitted. Of course $\lambda(x)$ is unknown but for each λ , $n^*(d)$, $d \in D$ are independent Poisson distributed variates whose means are given by $\mu(d)$ defined in (1.2). In [SV] we denoted $\mu(d)$ by $\lambda^*(d)$. The reader should prove the assertion that $n^*(d)$ are independent and Poisson provided that the emission process is a Poisson process with variable rate $\lambda(x)$.

Hint: If x -space is broken into small regions b_1, b_2, \dots the number, $n(b_i)$, of emissions from each b_i are not observable, but are independent and Poisson. With each emission a

line is chosen independently and uniformly through x and so each emission independently reaches d from x with probability $p(d|x)$ and the emission from each b_i into each d are independent and Poisson. Finally, sums and limits of independent Poisson variables have the same property. For more details see [S] or [SV].

Thus for each λ there is a probability $\Lambda(\lambda)$ of observing the actual recorded counts $n^*(d)$ for $d \in D$ if λ is the true emission density. Indeed by independence and the Poisson formula, this probability of observing $n^*(d)$, $d \in D$ under λ is precisely

$$(1.3) \quad \Lambda(\lambda) = \prod_{d \in D} e^{-\mu(d)} \frac{\mu(d)^{n^*(d)}}{n^*(d)!}$$

where $\mu(d)$ is given by (1.2).

The maximum likelihood (ML) viewpoint is now very natural. Since (1.3) is an exact model for the physics (because the Poisson assumption that $n(b_i)$ are independent and Poisson for disjoint b_i is a perfect model of radioactive decay) it seems reasonable to estimate the unknown λ as any λ which maximizes the likelihood of the observed counts, i.e. the product (1.3). Such a λ is called a maximum likelihood estimate (MLE). Note that λ is an arbitrary density so that one would not expect $\Lambda(\lambda)$ to have a unique maximum. We will return to this point in §2, where we give evidence which indicates that *in typical cases the maximum is essentially unique* (of course it is not always unique).

Now we return to the question of the discreteness of D , the set of detector units. There are position-sensitive detectors which measure (up to an error) the point on the detector where the radioactivity strikes. For these detectors it becomes unnatural to think of detection as discrete. Also in time-of-flight (TOF) ET scanners [SV], [STT] a central location is measured for each detected count so that x itself is observed up to an error, usually taken to be a Gaussian displacement. Position-sensitive and TOF detection can be

modelled by a finite D (in the case of TOF one often uses subtubes [SV] and assigns a count proportionally to more than one subtube although this seems to be a somewhat inelegant model. In the non-TOF case with error-free position detection, one can imagine that for each count arising at x we cannot observe x itself, but instead a random line uniformly distributed in angle is chosen through x and observed. It would then be desired to estimate λ given a finite set of observed or detected lines. It can be shown that an MLE must be concentrated on the set of intersections of the given lines. But such is not a density since this is a set of zero measure and so there is no maximum likelihood density. Of course exact detection of lines is not real, but we raise this limiting case to point out some difficulties with the ML criterion. Note Λ is bounded (by 1) since $\Lambda(\lambda)$ is a probability but the maximizing λ is achieved by a measure with no density.

Returning to the simpler discrete detector case, we may also discretize x into pixels or boxes, $b \in B$, and this was done in [SV]. Let $p(b, d) = p(d | b)$ be the probability that a detected emission in box b will be detected in unit d . From (1.2) we have

$$(1.4) \quad p(b, d) = \int_b \lambda(x) p(d | x) dx / \int_b \lambda(x) dx$$

so that $p(b, d)$ depends on λ , which is unknown. But if b has small diameter so that $p(d | x)$ is nearly constant for $x \in b$ then we may assume that $p(b, d)$ is known and say is approximately given by (1.4) where λ is say uniform over $x \in b$. For more on this see [SV].

The EM algorithm gives a simple update rule which starts with $\lambda^0(b)$, $b \in B$ usually taken to be flat, $\lambda^0(b) = 1$, and assigns a new $\lambda^1(b)$, $\lambda^2(b)$, ... in order by the rule

$$(1.5) \quad \lambda^{k+1}(b) = \lambda^k(b) \sum_{d \in D} \frac{n^*(d)p(b, d)}{\mu^k(d)}$$

where

$$(1.6) \quad \mu^k(d) = \sum_{b \in B} \lambda^k(b)p(b, d).$$

As observed in [SV] the update rule (1.5) has elegant properties. First, the likelihood always increases,

$$(1.7) \quad \Lambda(\lambda^{k+1}) \geq \Lambda(\lambda^k)$$

with equality only if λ^k is a maximum of Λ . Moreover, it is clear from (1.5) that $\lambda^k(b) \geq 0$ and that for $k \geq 1$

$$(1.8) \quad \sum_b \lambda^k(b) = \sum_{d \in D} n^*(d)$$

so that λ^k has the same total number of recorded counts. Moreover $\lim_{k \rightarrow \infty} \lambda^k(b)$ always exists [CT, VSK] and this maximizes $\Lambda(\lambda)$.

The algorithm (1.5) is interesting in itself and can be used to approximately nonlinearly "deconvolve" or invert a matrix equation $n^* = \lambda P$ for vector λ given vector n^* and matrix P ,

$$(1.9) \quad n^*(d) = \sum_{b \in B} \lambda(b)p(b, d), \quad d \in D$$

where $\lambda(b) \geq 0$ is sought from measurements of the left side where $p(b, d)$ is a known nonnegative matrix satisfying

$$(1.10) \quad \sum_{d \in D} p(b, d) = 1, \quad b \in B.$$

Our experience with EM in [SV], [VSK] shows that good numerical results can be expected from such a nonlinear inversion of positive matrices. It also suggests that

methods from linear programming [K], [V] can be used to study ET and this will be done in theory (§2) and practice (§3).

Lemma. *If (1.9) has a solution $\lambda(b) \geq 0$ then it must maximize Λ .*

To prove this, call $\lambda^k = \lambda$, and note from (1.9) that $\mu^k(d) = n^*(d)$ and then from (1.5) and (1.10) that $\lambda^{k+1} = \lambda^k$. But (1.7) then holds with equality and so $\lambda = \lambda^k$ is a maximum of Λ . Q.E.D.

With low counts (small $n^*(d)$) there is lots of noise in (1.9) and there is typically no (nonnegative) solution λ in (1.9) as we will see from our experiments in §3. A theoretical analysis of noise as a function of total count is given in [JS].

Where did the EM rule (1.5) come from? Are there general principles underlying (1.5)? Are they useful elsewhere? The answers are yes and the history is as follows: L. Baum [B] studied a Markov chain model for breaking secrecy codes which led him to alternately: (a) use conditional expectation (*E* step) to obtain missing data and (b) use maximum likelihood (*M* step) to estimate the Markov chain parameters given this missing data, then go back to an *E* step to better estimate the missing data, and so on. This was generalized by Baum himself and further by [DLR], who coined the term EM for the two steps; the reader should consult [DLR] for more details. The EM algorithm is particularly elegant in this Poisson model because $\Lambda(\lambda)$ is logconcave and the above desirable properties of positivity, count preservation, and convergence can be obtained. This was first pointed out for ET in [SV]. To see how this *E* and *M* step works in the Poisson case see [VSK, p. 12]; but in a sense one does not really learn anything by "deriving" (1.5) in this way rather than by heuristics because the properties (1.7), (1.8) must then be proved anyway.

A major difficulty [SV, VSK] with the EM algorithm is that for large k , λ^k gets very noisy or snowy, i.e. has large oscillations. Although the likelihood continues to increase with k , at $k \approx 500$, one no longer likes the image. At $k \approx 5000$, one no longer even likes likelihood as a criterion, since as we show in §2, in a typical case, the image as given by λ^{5000} is ridiculous and unrecognizable. Why is this and what to do? If likelihood is a good thing why isn't more likelihood even better?

One explanation [VL, LV] is that likelihood as a criterion reflects noise in the data. Thus if one had direct observations, $n(b)$ of $\lambda(b)$, instead of indirect observations, $n^*(d)$, one would use $\lambda(b) = n(b)$ as the most likely value of $\lambda(b)$ since

$$e^{-\lambda} \frac{\lambda^n}{n!}$$

is maximized for a fixed n at $\lambda = n$. But this can be very noisy if $\lambda(b)$ is small, which is the case if the pixels b are too small.

As a way out of this problem, it has been suggested to smooth λ [SM], or to impose a penalty on λ for lack of smoothness, or equivalently to specify a prior distribution for λ and maximize posterior likelihood rather than absolute likelihood ([LiH], [GM], [LeH], [Rubin's comments to VSK]). Another way out was suggested by Veklerov and Llacer [VL], [LV] who propose to simply stop iterating EM using "a quantitative criterion with a simple probabilistic interpretation that allows the user to stop the algorithm just before this (bad) effect begins" [VL].

We prefer to use a Veklerov-Llacer stopping criterion rather than to arbitrarily smooth or to use an arbitrary questionable prior. Note that using (1.3) one easily computes the incremental likelihood in going from λ^k to λ^{k+1} and one could presumably design a quantitative stopping rule based on this alone (although the likelihood continues to increase with k). However most users of EM prefer other fixes for the problem.

Geman and McClure's method [GM], based on a carefully chosen Gibbs random field as prior, gives amazingly good reconstruction of the original λ in the simulation experiment they describe [GM]. We are however concerned that their choice of Gibbs prior is unfair since it places high probability on λ 's which only take a few different values (piecewise constant) which was true for the original "phantom" used in [SV]. This seems likely to lead to errors in reconstruction when the true phantom is actually smoothly varying – which may indeed be the case for real emission densities. We have not checked it but have been assured by McClure that this is not a problem, and if this is so, their method of smoothing seems very useful.

Another way out of this likelihood paradox is to use a least squares [LeH] or other non-likelihood based criterion to reconstruct λ . However, starting from any λ and letting λ' be the new λ obtained by an iteration of (1.5), will increase the likelihood because of (1.7) and so long as this does not ruin smoothness it seems desirable to continue using EM iterations which leads us back to EM anyway.

In the discussion to [VSK], Gabor Herman claims that EM is similar to earlier algorithms called multiplicative ART. Although this is true and may be useful in suggesting variants of EM, we think it is nice that likelihood and EM is not nearly as arbitrary as the ART algorithms in that EM maximizes the meaningful quantity of likelihood. It does give reasonable reconstructions if one does not iterate too far.

It should be pointed out that EM has other troubles, mainly in slow execution times. However, this can and is being gotten around by using highly parallel computing [MBM]. Another problem, new to us, of "edge overshoot" is described in [PS]. We have never seen this phenomenon. Can it be due to a poor choice of $p(b, d)$'s which do not match the generated counts physically?

In §2 we study the question of uniqueness of λ maximizing $\Lambda(\lambda)$ and describe a linear programming approach to ET which exploits recent developments in linear programming.

In §3 we discuss the methodology used in doing computer simulation experiments to study the performance of algorithms for ET in general, and those experiments related to linear programming suggested by the observations in §2, in particular.

§2. Linear Programming and EM

It has long been believed [VSK, p. 13] that in typical cases of ET where the number of pixels greatly exceeds the number of detector units that $\lambda^\infty = \lim \lambda^k$, the maximum likelihood estimator obtained from λ^0 , is far from unique and different $\lambda^0 > 0$ will produce different λ^∞ 's. In the (absurd) special case when $p(b, d) \equiv 1/D$ for all $b \in B$, $d \in D$, there are indeed many maximizers, since if λ^∞ is any maximum of Λ then any other λ with

$$(2.1) \quad \sum_{b \in B} \lambda(b)/D = \mu^\infty(D)$$

has the same value of $\mu(d) = \sum \lambda(b)p(b, d) = \mu^\infty(d)$, $d \in D$ and since Λ in (1.3) only depends on λ through μ , $\Lambda(\lambda) = \Lambda(\lambda^\infty)$. Of course $p(b, d) = 1/D$ would make a very poor tomograph indeed since there is no information in the counts except for the total number! In realistic cases we will see that even if $|B| \gg |D|$, λ^∞ should be expected to be unique, or at least essentially unique, in low count cases. The essential uniqueness despite the underdeterminedness of the problem in the sense that the number of unknowns $\lambda(b)$ greatly (factor of 3) exceeds the number of equations is surprising to us but is not totally unfamiliar [Bu].

The reason for this uniqueness is the following: Suppose instead that there are many $\lambda = \lambda(b)$, $b \in B$ say $\lambda \in L$ with

$$(2.2) \quad \Lambda(\lambda) = \Lambda(\lambda^\infty) \quad \text{for all } \lambda \in L.$$

Then we may take a convex combination, or weighted mean

$$\bar{\lambda}(b) = \sum_{\lambda \in L} \alpha(\lambda) \lambda(b)$$

where $\alpha(\lambda) > 0$ for $\lambda \in L$ and $\sum \alpha(\lambda) = 1$, of $\lambda \in L$. Since $\Lambda(\lambda)$ is logconcave in λ , [SV], $\Lambda(\bar{\lambda}) = \Lambda(\lambda^\infty)$ as well and so $\bar{\lambda}$ is an MLE but if L is large, $\bar{\lambda}$ will be smooth for some α weighting. That is, we can find a smooth estimate $\bar{\lambda}$ with exactly the same likelihood as λ^∞ , an MLE. Thus either λ^∞ must be unique (essentially unique) or there are smooth MLE's, which is not the case for low count cases. We remark that $\mu(d) = \lambda^*(d)$ is unique among all maximum likelihood estimators by a simple convexity argument.

We can carry the above observation a bit further in two ways. The first way is to note that by the lemma of §1, if we can find a solution of (1.9) then it is an MLE. But (1.9) is a linear programming (LP) problem since it is linear and $\lambda(b) \geq 0$ are the constraints. Recent developments in linear programming (see e.g. [K],[V]) have yielded efficient routines for deciding feasibility and for solving even such large LP's. We have tried the routine on a typical ET case of $p(b, d)$ and $n^*(d)$ in §3. Unfortunately, as was guessed in [SV], (1.9) fails to be feasible in typical cases (at least for the case of the phantom and $p(b, d)$ given in [SV]) as we will see in §3, unless the number of counts is enormously large.

There is another, potentially more fruitful way, to use LP in ET. Namely we will show that given any λ , not necessarily an MLE, say $\lambda = \lambda^k$ for some λ^0 and some k , we can use LP to find a new λ' with the same likelihood as λ and which is as smooth as possible in a Manhattan or L_1 metric norm. So suppose we have iterated EM, $k = 5000$ times, and obtained $\lambda = \lambda^k$. Now we form

$$(2.3) \quad \mu^k(d) = \sum \lambda^k(b) p(b, d), \quad d \in D$$

and regard $\mu^k(d)$ as $n^*(d)$ in (1.9). Now we know that (1.9) has a solution since $\lambda = \lambda^k$ is such. Thus the problem with constraints (1.9) and inequalities $\lambda(b) \geq 0$, $b \in B$, and objective, say

$$(2.4) \quad \sum_{b \in B} \sum_{b' \in N(b)} |\lambda(b) - \lambda(b')| \quad \text{is a minimum}$$

where $N(b)$ is a local neighborhood of b , i.e. the set of nearest neighbors of b , is a feasible LP. Note that objective of the form (2.4) can be viewed as a linear objective despite its nonlinear appearance because one can introduce variables $\epsilon(b, b')$ for each pair $b \in B$ and $b' \in N(b)$ which satisfy the constraints

$$(2.5) \quad \epsilon(b, b') \geq \lambda(b) - \lambda(b') \quad \text{and} \quad \epsilon(b, b') \geq \lambda(b') - \lambda(b)$$

and use as objective

$$(2.6) \quad \text{minimize} \quad \sum_{b \in B} \sum_{b' \in N(b)} \epsilon(b, b')$$

subject to ϵ and λ satisfying (2.5) and (1.9) with $\lambda(b) \geq 0$. Unfortunately even the present LP package we have is not able to easily implement this LP because the introduction of the variables ϵ in (2.5) greatly increases the number of independent variables. Instead in §3 and Figure 3 we describe the solution to the LP with the objective

$$(2.7) \quad \text{minimize the maximum of } \lambda(b), \quad b \in B.$$

This is not exactly a smoothness criterion but should have the same effect since (as we shall see in §3) $\lambda^{5000}(b)$ has enormous values for some b .

In principle, and in practice too, LP can be used in conjunction with EM to give maximally smoothed (in the sense of (2.6) or (2.7)) estimators with the same likelihood as λ^k for any (even small) k . Whether this is enough of an advantage to consider

implementing LP in an EM laboratory is another question that can only be decided by experimentation.

Another way to use LP to improve ET is somewhat ad hoc as follows. Suppose one wants to distort the true value $n^*(d)$ as little as possible so that the LP problem becomes feasible. Thus suppose that $\epsilon_d \geq 0$ are variables and we wish to find the least value of

$$(2.8) \quad \epsilon = \sum_d \epsilon_d$$

for which the problem

$$(2.9) \quad n^*(d) - \epsilon_d \sqrt{n^*(d)} \leq \sum_b \lambda(b) p(b, d) \leq n^*(d) + \epsilon_d \sqrt{n^*(d)}$$

is feasible, i.e. has a solution $\lambda(b) \geq 0$. Such a λ would then be maximum likelihood for some $n^*(d)$ problem (not necessarily integer counts) which differs as little as possible from $n^*(d)$. The factors $\sqrt{n^*(d)}$ were suggested by Y. Vardi. Unfortunately we shall see in §3 that this approach is also unsuccessful.

§3. Experiments

In each of our experiments we use the same approach as in [SK]. Namely we take an anthropomorphic phantom (perhaps one more realistic for ET can be chosen if a particular class of real densities can be envisioned), i.e. a particular $\lambda(x)$. As discussed in [SV] and in [VSK] the true counts $n(b)$ and $n^*(d)$ can be simulated perfectly (neglecting scatter and randoms [SV]) as described in [SV]. Thus Figure 1 represents $n(b)$ for 128×128 pixels b where 10^6 counts were generated. It is very important to understand that in generating $n^*(d)$ no use is made of the pixel boundaries or the shape of the boxes b . Instead each of the 10^6 points is chosen from $\lambda(x)$ by choosing a point x uniformly in the unit square and accepting it with probability proportional to $\lambda(x)$. If it is accepted, a random line l through x is chosen and the tube of l is incremented. It is easy to see (in principle) how to

study effects of range, angulation, scatter, and random with this model (although we have not done so). Note that there are $\binom{128}{2} = 8128$ detector units in [SV], but that half of these have zero counts because they are not sufficiently opposing. Thus there are about $64^2 = 4096$ tubes and about $128^2 \cdot \pi/4 \doteq 12868$ pixels, so that there are more than 3 times as many unknowns $\lambda(b)$ as measurements, $\mu^*(d)$. Appendix 1 gives the program we used to generate the count data.

For $p(b, d)$ we use the value of $p(d | x)$ where x is the exact center of b . It is easy to see that the $p(b, d)$'s used in [SV] lead to infeasibility in (1.9). This is because the $p(b, d)$'s in [SV] are specially chosen for rapid computation and satisfy for example tubes in any (say vertical) direction,

$$(3.1) \quad \sum_{\text{vertical } d} p(b, d) = c = \text{constant in } b.$$

But then for (1.9) to have a solution we must have

$$(3.2) \quad \sum_{\text{vertical } d} n^*(d) = c \sum_b \lambda(b)$$

which is an event of probability zero since $n^*(d)$ are independent. However, for $p(b, d) = p(d | \text{center of } b)$ as above there is no degeneracy as in (3.1) and (1.9) has a chance of being feasible. If there is less noise in the measurements (i.e. large $n^*(d)$) then (1.9) is more likely to be feasible. We determined roughly by experiment that if there are $nb \times nb$ pixels and $\binom{nd}{2}$ tubes then the range of the number nc of counts at which (1.9) has a feasible solution with reasonably positive probability is given in Table 1.

Table 1

nb	nd	nc
5	5	[150, 300]
10	5	[150, 300]
10	10	10^5
20	10	$[2 \times 10^4, 2.5 \times 10^4]$
30	10	$[2 \times 10^4, 2.5 \times 10^4]$
45	15	$[10^5, 3 \times 10^5]$
128	20	$\gg 10^7$

The evidence in Table 1 makes it very likely that for $nb = nd = 128$ as in [SV] the problem (1.9) is infeasible unless $nc \gg 10^{10}$, i.e. it is not going to be feasible under typical circumstances. This was guessed (without real evidence) in [SV], but now that we have a package to do large LP's we have empirically verified that (1.9) is typically not feasible at reasonable count rates.

Figure 2a-d shows the EM reconstruction after 20, 60, 500, 5000 iterations. The log likelihood values $V_k = C - \log \Lambda(\lambda^k)$ are approximately given in Table 2 where C is an arbitrary constant. The true value of the higher likelihood reconstructions is small. Smoothing the 5000th iteration with a crude average on 9 points restores a reasonable image (Figure 2e).

How can we use LP to get better estimates of $\lambda(b)$ without loss of likelihood? One way was described in (2.3) where $n^*(d)$ is replaced by $\mu^k(d)$. Figure 3a shows the result of using (2.7) to maximize smoothness under (2.3) with $k=5000$. Unfortunately there is no improvement over Figure 2d (the large values if $\lambda(b)$ are in the same place) and we conclude that the method fails because the maximum likelihood estimator is essentially unique. We tried to start the EM algorithm off at a nonuniform λ^0 and again we got back to Figure 3a or 2d. Figures 3b, 3c show 3a smoothed once and twice.

The result of minimizing (2.8) subject to the feasibility of (2.9) yields the reconstruction in Figure 4a, again of the type of Figures 2d, 3a, with 10^6 counts. With 10^8 counts we obtain Figure 4b which is slightly smoother. Smoothing Figure 4b gives 4c. Single and double smoothing of 4a gives 4d and 4e.

Our conclusions from these experiments are that the likelihood criterion is flawed. More likelihood cannot be achieved without giving up smoothness. The true maximum likelihood estimator is essentially unique.

Table 2

Loglikelihood as a function of iteration number.

Iteration number k	$V_k = C - \log \Lambda(\lambda^k)$
10	5760
20	273
30	-512
40	-771
50	-901
60	-981
100	-1142
200	-1277
300	-1328
400	-1357
500	-1376
1000	-1430
2000	-1482
3000	-1512
4000	-1531
5000	-1545

c positr
c i=(i1,
c b(i) =
c pnp(k)
c nc=tot
c t1k,t2
c w=widt
c rd=rad

c define

1234

10

11

997

c ck=cos
c the ec
c (rdcos

c

```

c positron emission recon with e-m algorithm
c i=(i1,i2) indexes the ith box with center at -1+(2i-1)/(2b-1) i1=i1,i2
c b(i) = lambda(i); f(i)=sum(n(.,k)c(i,k)/ph(k))
c pnp(k)=no. counts in tube k; ph(k)=sum(b(j)c(j,k))
c nc=total n0. of counts, nd=no. detectors, .5nd(nd-1)=no. tubes
c t1k,t2k= distances from 0. to edges of tube k, -rd<t2k<t1k<rd
c w=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.)
  implicit real*8 (a-h,o-z)
  dimension x(128),st(-1:256),ct(-1:256),a(4)
c define constants
  open(9,file='nbnd')
  rewind(9)
  read(9,1234)nb,nd
1234  format(2i4)
  nd2 = 2*nd
  if (nb .gt. 128 .or. nd .gt. 128) then
    print *, 'ERROR: problem too big'
    stop
  endif
  rho=1./nb
  rd=sqrt(2.)
  pi=3.14159265d+0
  pi2=2*pi
  cc=1./pi
  rnb2=.5*nb
  rndpi2=nd/pi2
  pi2nd=pi2/nd
  do 10 i=1,nb
    x(i)=-1.+(i-.5)*2./nb
10  continue
  do 11 k=-1,nd2
    th=.5*k*pi2nd
    st(k)=sin(th)
    ct(k)=cos(th)
11  continue
  icnt = 1
  write(14,997) icnt
997  format(i7)

  do 301 k1=0,nd-2
    do 302 k2=k1+1,nd-1
c ck=cos(thk),sk=sin(thk),thk=(alf+bet)/2 where x*ck+y*sk-t1k,t2k define
c the edge line of tube with centers at rdcosalf,rdsinalf & at
c (rdcosbet,rdsinbet), 0<alf<bet<2 pi.
      ck=ct(k1+k2)
      sk=st(k1+k2)
      t1k=rd*ct(k2-k1-1)
      t2k=rd*ct(k2-k1+1)
      if(abs(ck).ge.abs(sk)) then
c
        now run through boxes by rows
        do 303 i2=1,nb
          ia=((1.+(t2k-rho-x(i2)*sk)/ck)*nb+1.)*.5
          ib=((1.+(t1k+rho-x(i2)*sk)/ck)*nb+1.)*.5
          i11=max0(1,1+min0(ia,ib))
          i12=min0(nb,max0(ia,ib))

```

```

do 304 il=i1,i12
  tip=x(il)*ck+x(i2)*sk+rho
  tim=tip-2.*rho
  w=amin1(amax1(tip,t2k),t1k)
  -amin1(amax1(tim,t2k),t1k)
  a(1)=atan2(rd*st(2*k1-1)-x(i2),
             rd*ct(2*k1-1)-x(i1))
  a(2)=atan2(rd*st(2*k1+1)-x(i2),
             rd*ct(2*k1+1)-x(i1))
  a(3)=atan2(rd*st(2*k2-1)-x(i2),
             rd*ct(2*k2-1)-x(i1))
  a(4)=atan2(rd*st(2*k2+1)-x(i2),
             rd*ct(2*k2+1)-x(i1))
  ichk = 0
  do 350 i=2,4
    if ( a(i) .lt. a(i-1) ) then
      ichk = ichk+1
      do 351 j=i,4
        a(j) = a(j) + pi2
      continue
    endif
  continue
  if ( ichk .gt. 1 ) then
    print *, 'ERROR: ichk .gt. 1',
    * ' x k1',k1,'k2',k2,'i1',i1,'i2',i2,'a', (a(i),i=1,4)
    stop
  endif
  a(3) = a(3) - pi
  a(4) = a(4) - pi
  w = max(min(a(4),a(2))
          -max(a(3),a(1)),0.0d0)
  if ( w .ne. 0.0 ) then
    write(15,998) w*cc
    write(16,999) (i1-1)*nb+i2
    icnt = icnt + 1
    format(f6.4)
    format(i7)
  endif
  continue
  continue
else
  now run through boxes by cols
  dk=sk*2./nb
  do 306 il=1,nb
    ia=((1.+(t2k-rho-x(il)*ck)/sk)*nb+1.)*.5
    ib=((1.+(t1k+rho-x(il)*ck)/sk)*nb+1.)*.5
    i21=max0(1,1+min0(ia,ib))
    i22=min0(nb,max0(ia,ib))
    do 307 i2=i21,i22
      tip=x(il)*ck+x(i2)*sk+rho
      tim=tip-2.*rho
      w=amin1(amax1(tip,t2k),t1k)
      -amin1(amax1(tim,t2k),t1k)
      a(1)=atan2(rd*st(2*k1-1)-x(i2),
                rd*ct(2*k1-1)-x(i1))
      a(2)=atan2(rd*st(2*k1+1)-x(i2),
                rd*ct(2*k1+1)-x(i1))

```

35

35

307

301

302

301

REFERENCES

- [B]. Baum, L. E. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. In *Inequalities, III: Proceedings of a Symposium*, Oved Shisha ed., Academic Press, New York.
- [Bu] Buchta, C. On nonnegative solutions of random systems of linear inequalities, *Discrete Comput. Geom.* 2, 1987, pp. 85-95.
- [CT]. Csiszar, I. and Tusnady, G. Information geometry and alternating minimization procedures, *Math. Inst. of Hungarian Acad. Sci.*
- [DLR]. Dempster, A. P., Laird, N. M. and Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm, *J. Roy. Stat. Soc. B*, 39, 1977, pp. 1-38.
- [GM]. Geman, S. and McClure, D. E. Statistical methods for tomographic image reconstruction, *Proc. of 46th Session of ISI, Bull. ISI*, vol. 52, 1987.
- [JS] Johnstone, I. and Silverman, B. W. Speed of estimation in PET, Tech. report 290, March '88, Statistics Dept., Stanford University.
- [K]. Karmarkar, N. K. A new polynomial-time algorithm for linear programming, *Combinatorica*, 4, 1981, 373-395
- [LeH]. Levitan, E. and Herman, G. T. A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography, *IEEE Trans. Med. Imag.*, 1987.
- [LiH]. Liang, Z. and Hart, H. Bayesian image processing of data from constrained source distributions I, II, *Bull. Math. Biol.* 49, 1987, 51-91.

- [LV]. Llacer, J. and Veklerov, E. The maximum likelihood estimation method of image reconstruction: its fundamental characteristics and their origin, ms., May 1987.
- [MBM]. McCarthy, A., Barrett, R. and Miller, M. I. Systolic implementation of the EM algorithm for emission tomography on mesh-connected processors, Proc. 22nd Conf. on Inf. Sci. and Systems, Princeton, University, 1988.
- [P]. Petersen, S. E., Fox, P. T., Posner, M. I., Mintum, M. and Raichle, M. E. Position emission tomographic studies of the cortical anatomy of single-word processing. *Nature* 331, pp. 585-589, 1988, Feb.
- [PS]. Polite D. G. and Synder, D. L. The use of constraints to eliminate artifacts in maximum-likelihood image estimation for emission tomography, monograph 489, October 1987, Washington University School of Medicine, St. Louis, MO.
- [S]. Synder, D. L. *Random Point Processes*, Wiley, New York, 1975.
- [SK]. Shepp, L. A. and Kruskal, J. B. Computerized tomography: The new medical X-ray technology, *Amer. Math. Monthly*, 85, 420-439.
- [SM]. Synder, D. L. and Miller, M. I. The use of sieves to stabilize images produced with the EM algorithm for emission tomography, *IEEE Trans. Nucl. Sci.*, NS-32, 1985, pp. 3864-3872.
- [STT]. Synder, D. L., Thomas, L. J., Ter-Pogossian, M. M. A mathematical model for positron-emission tomography systems having time-of-flight measurements, *IEEE Trans. Nucl. Sci.*, NS-28, 1981, pp. 3575-3583.
- [SV]. Shepp, L. A. and Vardi, Y. Maximum-likelihood reconstruction for emission tomography, *IEEE Trans. Med. Imag.*, vol. MI-1, pp. 113-122, 1982.

- [V]. Vanderbei, R. J., Meketon, M., Freedman, B. A modification of Karmarkar's linear programming algorithm, *Algorithmica*, 1, 1986, pp. 395-407.
- [VL]. Veklerov, E. and Llacer, J. Stopping rule for the MLE algorithm based on statistical hypothesis testing, ms. February 1987.
- [VSK]. Vardi, Y., Shepp, L. A., and Kaufman, L. A statistical model for positron emission tomography, *J. Am. Stat. Soc.* 80, 1985, pp. 8-37.