EM Reconstruction Algorithms for Emission and Transmission Tomography

Article *in* Journal of Computer Assisted Tomography · May 1984

CITATIONS
1,341

READS
486

2 authors, including:

Richard Ellis Carson
Yale University
482 PUBLICATIONS **19,052** CITATIONS

SEE PROFILE

Journal of Computer Assisted Tomography
8(2):306–316, April

EM Reconstruction Algorithms for Emission and Transmission Tomography

Kenneth Lange and Richard Carson

Abstract: Two proposed likelihood models for emission and transmission image reconstruction accurately incorporate the Poisson nature of photon counting noise and a number of other relevant physical features. As in most algebraic schemes, the region to be reconstructed is divided into small pixels. For each pixel a concentration or attenuation coefficient must be estimated. In the maximum likelihood approach these parameters are estimated by maximizing the likelihood (probability of the observations). EM algorithms are iterative techniques for finding maximum likelihood estimates. In this paper we discuss the general principles behind all EM algorithms and derive in detail the specific algorithms for emission and transmission tomography. The virtues of the EM algorithms include (a) accurate incorporation of a good physical model, (b) automatic inclusion of non-negativity constraints on all parameters, (c) an excellent measure of the quality of a reconstruction, and (d) global convergence to a single vector of parameter estimates. We discuss the specification of necessary physical features such as source and detector geometries. Actual reconstructions are deferred to a later time. **Index Terms:** Image reconstruction—Emission computed tomography—Maximum likelihood—EM algorithm.

Image reconstruction methods fall broadly into one of two categories (1,2). Fourier methods approximate explicit deterministic inversion formulas for reconstructing a function from its line integrals. On the other hand, algebraic reconstruction techniques can capture stochastic variation in photon counts and, in theory, yield more accurate reconstructions or provide equivalent reconstructions with lower patient radiation dose. Because algebraic techniques are usually iterative, and therefore slower, the single step Fourier-based methods are generally preferred in practice.

With the advent of more powerful computers, the arguments favoring algebraic reconstruction become more compelling. In this paper we present two new reconstruction algorithms based on maximum likelihood techniques. [One of these, the emission algorithm, was derived independently by Shepp and Vardi (3).] Rockmore and Macovski (4), among others, recognized the value of the maximum likelihood approach. What they lacked was a feasible method for computing maximum likelihood estimates for problems with thousands of parameters to estimate. The EM algorithms discussed here supply the missing link in the maximum likelihood approach.

Our optimism about maximum likelihood methods stems from a number of considerations. First, maximum likelihood has excellent theoretical properties (5) and an excellent track record in diverse application areas. Second, the maximum likelihood approach allows modeling of the physics of emission and transmission in a realistic way. A key modeling assumption concerning the Poisson nature of photon counting fits readily into the maximum likelihood framework. This is not true for most algebraic methods. Furthermore, the physical differences between transmission and emission can be carefully maintained. Some emission reconstruction methods blur these differences in an attempt to carry over wholesale the techniques of transmission tomography. The combination of a good statistical model and a superior physical model should produce superior reconstructions. The emission simu-

From the Department of Biomathematics, University of California at Los Angeles, School of Medicine, Los Angeles, CA 90024. Address correspondence and reprint requests to Dr. Lange.

lation studies of Shepp and Vardi (3) support this contention. Equally impressive are the further studies of Snyder and Politte (6), who incorporate time-of-flight information based on reference 3.

In this paper, we first summarize the theory behind the EM algorithm. We then mention briefly some relevant properties of the Poisson distribution. These ideas are applied in deriving the EM algorithm for emission computed tomography (CT). We proceed to the EM algorithm for transmission tomography. This second algorithm is mathematically more difficult to derive; in fact, an exact analytical solution cannot be specified. However, we propose some highly accurate approximations that should overcome all practical obstacles. Throughout the paper, we try to maintain as much commonality of notation as possible to aid in understanding the many similarities between the two algorithms.

The discussion spells out in detail the characteristics of the EM algorithms, with particular emphasis on some of the modeling possibilities inherent in their application to emission (positron and single photon) and transmission tomography. The appendix contains a unified treatment of some convergence results for both algorithms. At a later time we plan to provide detailed simulation studies of both algorithms as well as reconstructions from actual scan data.

Description of the EM Algorithm

The EM algorithm is an iterative technique for computing maximum likelihood estimates. Dempster et al. (7) were the first to state the algorithm in its full generality, although particular examples were known many years previously. Baum et al. (8), Hartley and Hocking (9), Orchard and Woodbury (10), and Sundberg (11) all anticipated various parts of the theory.

The basic idea is fairly simple. Suppose the observed data in some experiment or sequence of experiments is a random vector Y. Let Y have density function $g(Y,\theta)$, where θ is some vector of parameters to be estimated. (For purposes of this paper, we made no distinction between density functions and likelihoods.) In general, it may be difficult to maximize $g(Y,\theta)$ with respect to θ. A possible solution to this predicament is to perform the thought experiment of embedding the sample space for Y in a richer or larger sample space where optimization problems are easier to solve. Thus, the EM algorithm postulates a "complete data" random vector X such that Y is a function $h(X)$ of X. Furthermore, X is supposed to have a density function $f(X,\theta)$ with respect to some measure $\mu(X)$. Under these assumptions, $g(Y,\theta)$ can be recovered by the integration

$$g(Y,\theta) = \int_{\{X:h(X)=Y\}} f(X,\theta)d\mu(X). \quad (1)$$

In the present context, all density functions are discrete and Eq. 1 reduces to the summation

$$g(Y,\theta) = \sum_{\{X:h(X)=Y\}} f(X,\theta).$$

Each iteration of the EM algorithm consists of two steps. In the E-step, one forms the conditional expectation

$$E(\ln f(X,\theta)|Y,\theta^n),$$

where ln denotes natural logarithm, and θ^n denotes the current vector of parameter estimates. In the M-step, this conditional expectation is maximized with respect to θ, θ^n held fixed, to give the new parameter estimates θ^{n+1}. When $f(X,\theta)$ belongs to a regular exponential family, this description can be recast using the sufficient statistic for X (7,11).

The essence of the EM algorithm can be explained by considering the function

$$\theta \to H(\theta|\theta^n) = E(\ln f(X,\theta)|Y,\theta^n) - \ln g(Y,\theta).$$

Regardless of the value of θ^n, $H(\theta|\theta^n)$ attains its maximum when $\theta = \theta^n$. This follows from a variant of Jensen's inequality (5). The EM algorithm exploits this fact by choosing $\theta = \theta^{n+1}$ to maximize

$$E(\ln f(X,\theta)|Y,\theta^n).$$

It follows readily that

$$\ln g(Y,\theta^{n+1}) \geq \ln g(Y,\theta^n),$$

with strict inequality in most circumstances. In other words, the EM algorithm is designed to increase the log likelihood at each iteration. We note for later use that

$$\frac{\partial}{\partial\theta_j} E(\ln f(X,\theta)|Y,\theta^n)|_{\theta=\theta^n} = \frac{\partial}{\partial\theta_j}\ln g(Y,\theta^n)$$

for any component θ_j of θ when θ_j^n is not on a boundary.

Unfortunately, the convergence results in (7) are seriously flawed. Boyles (12) and Wu (13) attempt to remedy the situation. Although rigorous, their arguments do not cover the present problems. These problems have boundary features that complicate the analysis of convergence. We show in the appendix that convergence occurs regardless of the choice of the initial estimate θ^0. Our proof depends strongly on the fact that $\theta \to \ln g(Y,\theta)$ is strictly concave for both reconstruction models.

The art in using the EM algorithm lies in choosing an appropriate complete data specification X. Although there can be many ways of embedding Y in a larger sample space, often physical considerations suggest a natural definition of X. This happens to be the case for both image reconstruction algorithms developed here.

Some Properties of the Poisson Distribution

A non-negative, integer-valued random variable Z follows a Poisson distribution if

$$P(Z = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

for some $\lambda > 0$. The Poisson distribution has some very useful properties.

1. Z has mean $E(Z) = \lambda$.
2. If $Z_1, \ldots, Z_m$ are independent Poisson variates, then

$$Z = \sum_{k=1}^{m} Z_k$$

also is a Poisson variate (14).

3. Suppose that a random number Z of particles is generated according to a Poisson distribution with mean λ. Let each particle be independently distributed to one of m categories, the kth of which occurs with probability p_k. If we let Z_k represent the random number of particles falling in category k, Z_k follows a Poisson distribution with mean λp_k. Furthermore, the collection of random variables $Z_1, \ldots, Z_m$ is independent (14). Since

$$Z = \sum_{k=1}^{m} Z_k,$$

note the consistency between 2 and 3.

4. In the same vein as 3, suppose Z is given. Then the conditional distribution of $(Z_1, \ldots, Z_m)$ given Z is multinomial with cell probabilities $p_1, \ldots, p_m$ and total Z (14).

Examples of 3: Suppose $m = 2$. Category 1 particles are detected by some device, and category 2 particles are lost. The number of detected particles Z_1 is Poisson with mean λp_1. As another example, suppose the particles are generated over a time interval (t_1, t_2). The categories can correspond to a partition of (t_1, t_2) into disjoint subintervals. Yet another example is furnished by particles leaving at a random angle from a fixed point of generation. The categories can correspond to a subdivision of a unit circle centered at the point of generation.

The relevance of these examples to image reconstruction should be obvious. For instance, the last two examples together imply that the numbers of particles detected for the various projections are independent random variables; these projections either occur over disjoint time intervals or involve different angular sectors for each point source of particles (photons). Because of this independence property, the likelihood over all projections reduces to the product of the separate projection likelihoods.

Emission Computed Tomography

In emission CT the goal is to estimate the local intensities of photon emission in the object of interest. The emitted photons can be captured or deflected en route to a detector. Thus, there must be some correction for photon attenuation. With positron emitters, two photons are generated on the annihilation of an emitted positron and a neighboring electron. These two photons retreat from each other at approximately a 180° angle. For an emission event to be counted, both photons must register nearly simultaneously at two opposite detectors. The principal difference between positron emitters and single photon emitters is that the attenuation path for positron emitters amounts to the whole line segment between the pair of detectors. For single photon emitters the attenuation path is just the line segment between the source and the single detector. With this distinction clearly in mind, it is possible to formulate a single reconstruction algorithm that covers both kinds of emitters.

As with all algebraic reconstruction algorithms, it is necessary to divide the plane around and including the object of interest into pixels. These pixels typically are small squares defined by a rectangular grid.

In Table 1 we list necessary notation for both the emission and transmission models. Some of the assumptions behind these definitions deserve comment. First, λ_j and μ_j represent averages within a pixel. Second, although we allow for the Poisson process of photon emission to be nonhomogeneous in space, we require it to be homogeneous in time, excluding the easily corrected process of radioactive decay for the source. Third, the probabilities b_{ij} summarize all remaining physical features such as detector geometry and attenuation. Further specification of these constants is deferred to the discussion section.

With the notation of Table 1, it is clear that the mean number of detected photons (pairs of photons) originating from pixel j during the ith projection is

$$\Delta t_i \lambda_j b_{ij} = c_{ij} \lambda_j. \tag{2}$$

To accommodate radioactive decay of the source material this expression should be modified to

$$\beta^{-1}(e^{-\beta t_{i1}} - e^{-\beta t_{i2}})\lambda_j b_{ij} = c_{ij} \lambda_j,$$

where β is the decay rate and (t_{i1}, t_{i2}) is the time interval for the ith projection. It is important to note that the c_{ij} are considered known constants. The intensities λ_j are parameters to be estimated. For notational convenience, we will arrange λ_j into a single vector λ.

Now let X_{ij}, $j \in I_i$, be the random number of photons (pairs of photons) that are emitted by pixel j and contribute to projection i. As mentioned above, the mean of X_{ij} is $c_{ij}\lambda_j$. Also let Y_i be the total number recorded for projection i. It is obvious that

$$Y_i = \sum_{j \in I_i} X_{ij}.$$

In view of our earlier remarks about Poisson distributions, X_{ij} and Y_i have Poisson distributions. Y_i are the observed data for projection i. By definition, the X_{ij} constitute the unobserved but complete data.

Before specifying the EM algorithm let us check the strict concavity of the log likelihood $\ln g(Y,\lambda)$. Since the Y_i are independent Poisson variates,

$$\ln g(Y,\lambda) = \sum_i \left\{ -\sum_{j \in I_i} c_{ij}\lambda_j + Y_i \ln\left(\sum_{j \in I_i} c_{ij}\lambda_j\right) - \ln Y_i! \right\}. \tag{3}$$

A typical second partial derivative of $\ln g(Y,\lambda)$ is

TABLE 1. *Notation for the models*

Common to both emission and transmission
i, projection subscript
j, pixel subscript
I_i, set of pixels contributing to projection i
J_j, set of projections to which pixel j contributes
Δt_i, length of time over which the ith projection is collected
Unique to the emission case
λ_j, source intensity of pixel j
b_{ij}, probability that a photon (pair of photons) leaving pixel j reaches the ith projection detector (pair of detectors)
Unique to the transmission case
α_i, source intensity for projection i
μ_j, attenuation coefficient of pixel j (probability of photon capture per unit length)
l_{ik}, length of projection line i that intersects pixel j

$$\frac{\partial^2}{\partial\lambda_k\,\partial\lambda_l}\ln g(Y,\lambda) = -\sum_i Y_i \frac{a_{ik}\,a_{il}}{\left(\sum_{j\in I_i} c_{ij}\,\lambda_j\right)^2},$$

where

$$a_{ik} = \begin{cases} c_{ik} & k\in I_i \\ 0 & k\notin I_i. \end{cases}$$

The associated quadratic form

$$\sum_{k,l} v_k \frac{\partial^2}{\partial\lambda_k\,\partial\lambda_l}\ln g(Y,\lambda)\,v_l$$

is strictly negative definite whenever the matrix $A = (a_{ik})$ has at least as many rows (projections) as columns (pixels) and is of full rank. Under these conditions $\ln g(Y,\mu)$ is strictly concave.

To specify the E-step of the EM algorithm we need the complete data log likelihood

$$\ln f(X,\lambda) = \sum_i \sum_{j\in I_i} \{-c_{ij}\lambda_j + X_{ij}\ln(c_{ij}\lambda_j) - \ln X_{ij}!\}.$$

Taking the conditional expectation of X_{ij} with respect to Y_i and the current vector of parameter estimates λ^n yields

$$E(\ln f(X,\lambda)|Y,\lambda^n) = \sum_i \sum_{j\in I_i}\{-c_{ij}\,\lambda_j + N_{ij}\ln(c_{ij}\,\lambda_j)\} + R, \tag{4}$$

where

$$N_{ij} = E(X_{ij}|Y_i,\,\lambda^n) = \frac{c_{ij}\,\lambda_j^n\,Y_i}{\sum_{k\in I_i} c_{ik}\,\lambda_k^n},$$

and R does not depend on the new λ. Note that we have used the fact that, for fixed i, the X_{ij} are multinomial given Y_i.

The M-step is now trivial. Simply take partial derivatives,

$$\frac{\partial}{\partial\lambda_j}E(\ln f(X,\lambda)|Y,\lambda^n) = -\sum_{i\in J_j} c_{ij} + \sum_{i\in J_j} N_{ij}\,\lambda_j^{-1}, \tag{5}$$

set them to zero, and solve. The result is the very pleasing method of moments estimate

$$\lambda_j^{n+1} = \frac{\sum_{i\in J_j} N_{ij}}{\sum_{i\in J_j} c_{ij}} = \frac{\lambda_j^n}{\sum_{i\in J_j} c_{ij}} \sum_{i\in J_j} \frac{c_{ij}\,Y_i}{\sum_{k\in I_i} c_{ik}\,\lambda_k^n} \tag{6}$$

for the jth component of the new parameter vector λ^{n+1}. Since $E(\ln f(X,\lambda)|Y,\lambda^n)$ is concave in λ, Eq. 6 must represent the maximum.

Equation 6 has a simple intuitive interpretation. N_{ij} is the best current estimate of the number of detected photons (pairs of photons) that originate from pixel j during the ith projection. The sum $\sum_{i\in J_j} N_{ij}$ is the estimated number that originate from pixel j summed over all projections that pass through pixel j. This estimated number is set equal to the expected number $\sum_{i\in J_j} c_{ij}\,\lambda_j$ and the resulting equation solved for λ_j.

Equation 6 is also remarkable for two other reasons. First, the non-negativity constraints $\lambda_j^n \geq 0$ are automatically satisfied. However, note that the initial λ_j^0 must be positive; otherwise, all subsequent $\lambda_j^n = 0$. Second, at each iteration it is easy to verify that

$$\sum_i \sum_{j\in I_i} c_{ij}\,\lambda_j^n = \sum_i Y_i.$$

In other words, the current expected number of photons (pairs of photons) stays fixed at the total observed number.

Certain equality constraints also become trivial to implement in the above framework. For instance, any intensity λ_j can be held at a constant value. The update Eq. 6 is simply ignored at each iteration for pixel j. Neighboring pixels can be forced to have the same intensities. If S is a set of pixels with common intensity λ_S, the M-step of the algorithm is modified so that λ_S is updated by

$$\lambda_S = \frac{\sum_{j\in S}\sum_{i\in J_j} N_{ij}}{\sum_{j\in S}\sum_{i\in J_j} c_{ij}}.$$

Transmission Computed Tomography

The goal in transmission CT is to compute local attenuation coefficients in the object of interest. In contrast with emission CT the source of photons is external to the object. The parameters to be estimated are the attenuation coefficients μ_j for each pixel j. Corresponding to each projection i, there is also an associated mean source intensity α_i. (See Table 1 for notation.) The α_i depends on the geometries of the source and the detector and can be estimated empirically simply by removing the object to be reconstructed. Hence, α_i is taken as known.

For each projection i, let W_i be the total number of photons leaving the source and headed toward the detector. In the absence of attenuation, all W_i photons would be detected, and by our earlier remarks the W_i are independent. Let Y_i be the actual number of photons detected. The Y_i constitute the observed data. Assuming an infinitely thin beam, each of the W_i photons leaving the source has the same probability

$$e^{-\sum_{j\in I_i} l_{ij}\mu_j} \tag{7}$$

of reaching the detector. It follows that Y_i has a Poisson distribution with mean the product of the mean of W_i, $d_i = \Delta t_i\alpha_i$, times the probability (Eq. 7). As a consequence of the independence of the W_i, the Y_i are also independent.

In summary, the log likelihood over all projections reduces to

$$\ln g(Y,\mu) = \sum_i \left\{ -d_i e^{-\sum_{j\in I_i} l_{ij}\mu_j} - Y_i\sum_{j\in I_i} l_{ij}\mu_j + Y_i\ln d_i - \ln Y_i! \right\}, \tag{8}$$

where Y and μ are vectors whose components are the Y_i and μ_j, respectively.

To check the strict concavity of $\ln g(Y,\mu)$ note that

$$\frac{\partial^2}{\partial\mu_k\partial\mu_l}\ln g(Y,\mu) = -\sum_i a_{ik}\, d_i\, e^{-\sum_{j\in I_i} l_{ij}\mu_j} a_{il},$$

where now

$$a_{ik} = \begin{cases} l_{ik} & k\in I_i \\ 0 & k\notin I_i. \end{cases}$$

The conditions on the matrix $A = (a_{ik})$ for the associated quadratic form

$$\sum_{k,l} v_k \frac{\partial^2}{\partial\mu_k\partial\mu_l}\ln g(Y,\mu) v_l$$

to be strictly negative definite are the same as in the emission case.

To derive the E-step of the EM algorithm we focus on a single projection. It simplifies matters to abandon temporarily our previous notation and drop the projection subscript i. We also relabel pixels, starting with pixel 1 adjacent to the source and ending with pixel $m-1$ adjacent to the detector. Note the irrelevance of pixels disjoint from the projection line.

Natural candidates for the complete data are the numbers of photons entering each pixel along the projection. Let X_j, $1 \leq j \leq m-1$, denote the number of photons entering pixel j. $X_1 = W$ is the number of photons emitted by the source and has mean $d = \Delta t\alpha$. X_m we identify with Y, the number of photons detected.

The key to writing the correct complete data likelihood is to note that the number of photons X_{j+1} leaving pixel j depends only on the number X_j entering and on the product $l_j\mu_j$ of the projection path length and attenuation coefficient. The number leaving should therefore follow a binomial distribution with probability of success $e^{-l_j\mu_j}$ and number of trials X_j. By an appropriate conditioning argument which inductively adds pixels from source to detector, it is possible to show that the complete data log likelihood is

$$-d + X_1\ln d - \ln X_1! + \sum_{j=1}^{m-1}\left\{\ln\binom{X_j}{X_{j+1}} + X_{j+1}\ln(e^{-l_j\mu_j}) + (X_j - X_{j+1})\ln(1 - e^{-l_j\mu_j})\right\} \tag{9}$$

To perform the E-step we must compute the expectations

$$E(X_j|X_m,\ \mu^n),$$

of the complete data X_j conditional on the incomplete data X_m and the current parameter vector μ^n. The conditional expectations of $\ln X_1!$ and $\ln\binom{X_j}{X_{j+1}}$ are naturally irrelevant in the M-step.

Just as in the case $X_m = Y$, the random variables X_j, $1 \leq j \leq m-1$, follow Poisson distributions with means

$$E(X_j) = \gamma_j = de^{-\sum_{k=1}^{j-1} l_k\mu_k}.$$

Furthermore, the distribution of X_m conditional on X_j is binomial with success probability

$$e^{-\sum_{k=j}^{m-1} l_k\mu_k} = \gamma_m/\gamma_j.$$

Hence, the joint probability

$$P(X_j,X_m) = e^{-\gamma_j}\frac{\gamma_j^{X_j}}{X_j!}\binom{X_j}{X_m}\left(\frac{\gamma_m}{\gamma_j}\right)^{X_m}\left(1 - \frac{\gamma_m}{\gamma_j}\right)^{X_j - X_m},$$

and the conditional probability

$$P(X_j|X_m) = \frac{e^{-\gamma_j}\frac{\gamma_j^{X_j}}{X_j!}\binom{X_j}{X_m}\left(\frac{\gamma_m}{\gamma_j}\right)^{X_m}\left(1 - \frac{\gamma_m}{\gamma_j}\right)^{X_j - X_m}}{e^{-\gamma_m}\frac{\gamma_m^{X_m}}{X_m!}}$$

$$= e^{-(\gamma_j - \gamma_m)}\frac{(\gamma_j - \gamma_m)^{X_j - X_m}}{(X_j - X_m)!}$$

In other words, conditional on X_m, $X_j - X_m$ follows a Poisson distribution with mean $\gamma_j - \gamma_m = E(X_j) - E(X_m)$. Thus,

$$E(X_j - X_m|X_m) = E(X_j) - E(X_m),$$

or

$$E(X_j|X_m) = X_m + E(X_j) - E(X_m). \tag{10}$$

This is precisely the conditional expectation needed in the E-step.

To perform the M-step of the EM algorithm we revert to our previous notation. Let M_{ij} and N_{ij} be the expected number of photons entering and leaving pixel j, respectively, for projection i. These expectations are conditional on the number of photons Y_i observed and the current parameter vector μ^n. They are given by Eq. 10 above. Summing the values (Eq. 9) derived from the E-step over all projections now yields

$$\sum_i\sum_{j\in I_i}\{N_{ij}\ln(e^{-l_{ij}\mu_j}) + (M_{ij} - N_{ij})\ln(1 - e^{-l_{ij}\mu_j})\} + R, \tag{11}$$

where the term R includes all those quantities that do not depend on the new parameters μ_j.

The M-step is carried out by computing the partial derivative of Eq. 11 with respect to μ_k and equating it to zero for each k. This yields

$$0 = \sum_{i\in J_k} -N_{ik}l_{ik} + \sum_{i\in J_k}(M_{ik} - N_{ik})\frac{l_{ik}\, e^{-l_{ik}\mu_k}}{1 - e^{-l_{ik}\mu_k}}$$

$$= \sum_{i\in J_k} -N_{ik}l_{ik} + \sum_{i\in J_k}(M_{ik} - N_{ik})\frac{l_{ik}}{e^{l_{ik}\mu_k} - 1}. \tag{12}$$

The transcendental Eq. 12 cannot be solved exactly. However, the right hand side of Eq. 12 is monotone in μ_k and has one positive solution. Furthermore, an easy convexity argument shows that the solutions correspond to a maximum of Eq. 11.

All practical applications will have sufficiently small pixels so that terms such as $l_{ik}\mu_k$ in Eq. 12 will be small. This suggests using the approximation

$$\frac{1}{e^s - 1} = \frac{1}{s} - \frac{1}{2} + \frac{s}{12} + O(s^3) \tag{13}$$

with $s = l_{ik}\mu_k$. (To derive Eq. 13 expand $s/(e^s - 1)$ in a Taylor series.) Retaining only the $1/s$ term in Eq. 13 and substituting into Eq. 12 gives the approximate solution

$$\mu_k^{n+1} = \frac{\sum_{i\epsilon J_k}(M_{ik} - N_{ik})}{\sum_{i\epsilon J_k} N_{ik} l_{ik}}. \quad (14)$$

A better approximate solution can be obtained by retaining the first two terms of Eq. 13. Substitution of

$$\frac{1}{e^s - 1} \simeq \frac{1}{s} - \frac{1}{2}, \; s = l_{ik}\mu_k,$$

into Eq. 12 and rearrangement yields

$$\mu_k^{n+1} = \frac{\sum_{i\epsilon J_k}(M_{ik} - N_{ik})}{1/2\sum_{i\epsilon J_k}(M_{ik} + N_{ik})l_{ik}}. \quad (15)$$

This method of moments estimate can be derived heuristically by equating the expected number of photons captured by pixel k,

$$\sum_{i\epsilon J_k} 1/2(M_{ik} + N_{ik})l_{ik}\mu_k,$$

to the current estimated number of captured photons,

$$\sum_{i\epsilon J_k}(M_{ik} - N_{ik}),$$

and solving for μ_k.

The two Eqs. 14 and 15 actually furnish upper and lower bounds, respectively, to the true solution μ_k^{n+1}. This follows from the two elementary inequalities

$$\frac{1}{s} - \frac{1}{2} \leq \frac{1}{e^s - 1} \leq \frac{1}{s}, \; s > 0.$$

Our third and most accurate solution is based on using all three terms of Eq. 13. Substitution into Eq. 12 and rearrangement gives the quadratic equation

$$0 = \mu_k^2 \sum_{i\epsilon J_k}(M_{ik} - N_{ik})\frac{l_{ik}^2}{12} - \mu_k \sum_{i\epsilon J_k}(M_{ik} + N_{ik})\frac{l_{ik}}{2} + \sum_{i\epsilon J_k}(M_{ik} - N_{ik}). \quad (16)$$

Equation 16 can be rewritten in obvious notation as

$$0 = A\mu_k^2 - B\mu_k + C. \quad (17)$$

Provided the discriminant $B^2 - 4AC > 0$, Eq. 15,

$$\mu_k^{n+1} = C/B,$$

can be viewed as the leading term of the expansion

$$\mu_k^{n+1} = B/2A - \sqrt{B^2 - 4AC}/2A = \frac{B}{2A}(1 - \sqrt{1 - 4AC/B^2}) = C/B + AC^2/B^3 + \ldots,$$

for the smaller root of Eq. 17.

Two criteria are important to remember in assessing which of these solutions to use for the M-step. One is obviously accuracy, and the other, computational simplicity. On the basis of both these criteria, we prefer the second approximate solution (Eq. 15). But it must be recognized that close to the maximum, an inaccurate approximation may not lead to an increase in $\ln g(Y,\mu)$ from one iteration to the next. In theory, this difficulty can be circumvented by switching approximations in the course of iterating.

It should be reemphasized that all three estimates satisfy $\mu_k^{n+1} > 0$ whenever $\mu_k^n > 0$. If $\mu_k^n = 0$, $M_{ik} = N_{ik}$ in Eq. 11, and the only way to maximize Eq. 11 is to take $\mu_k^{n+1} = 0$ also. Finally, the presence of simple equality constraints on the μ_k can be handled just as in the emission case.

DISCUSSION

General Characteristics of the EM Reconstruction Algorithms

The EM reconstruction algorithms are based on an exact stochastic model of the projection measurements. Fourier/convolution techniques (15) are deterministic in nature and completely ignore the stochastic element of the data. Iterative approaches (2) can accommodate the presence of noise in the observations by means of the definition of criteria for an acceptable reconstruction. For instance, some algebraic techniques implement a weighted least squares approach with weights proportional to the estimated variance of each measurement (16). In scanning protocols in which counting statistics are high, ignoring the true statistical nature of the observations is not a serious limitation because Poisson counting noise is only a component in total system noise. This is the case currently with most transmission tomography and with some applications of emission tomography. For example, a high statistics fluorodeoxyglucose scan on the positron tomograph NeuroECAT (17) might collect from 2 to 5 million counts, divided among 8,448 projections, for an average value of 300 to 600 counts/projection. In emission tomography, however, some protocols are limited by statistical noise and the use of an inappropriate model is detrimental. This will be particularly true for future brain ligand tracers (18). It is precisely in these low count cases that the EM algorithms are expected to provide the greatest improvement in reconstruction quality. Furthermore, better reconstructions at lower count rates should allow reduction in patient radiation dose.

Unlike the Fourier-based techniques, algebraic reconstruction methods allow the specification of a detailed physical model relating the measured projections to the pixel values to be estimated. For the emission case, it is not necessary to assume that each projection is infinitely thin, uniform in depth, and unattenuated. Such deviations from the ideal cannot be directly handled by Fourier techniques, although the cutoff frequency and shape of the reconstruction filter function can be modified to adjust the noise level and resolution of the reconstruc-

tion (19,20). Furthermore, the physical model and the computational details of the EM algorithms generalize directly to three-dimensional reconstruction.

Built into the iteration structure of the EM algorithm is the non-negativity of the final reconstruction. In most algebraic methods non-negativity constraints are extraneous complications that slow the iterations and complicate the computer code (21). Other constraints, such as assigning a zero value to pixels outside the patient, can be directly incorporated in the EM algorithms. Fourier reconstructions can produce negative values and provide no direct approaches for use of constraints.

Unlike other iterative techniques, the EM algorithms provide a natural objective measure of the quality of a reconstruction, namely the log likelihood of the original observations given the model and the current pixel estimates (Eqs. 3 and 8). Since this function incorporates all the stochastic and physical information of the model, it should offer a better and more natural measure than maximum entropy (22,23), least squares (16), minimum norm (21), and their generalizations (24). The likelihood is guaranteed to increase at each step, and experience should provide useful criteria for termination of the iterations.

Emission Algorithm

There are a large number of physical factors affecting the projection measurements in single photon emission CT (SPECT) (25) and positron emission tomography (PET) (26), which, if not accounted for, can introduce errors in the final reconstruction. To reduce the effect of these factors in practical applications, scanners are designed to minimize them (e.g., scatter) (27), the projection data are "corrected" in some manner before reconstruction (e.g., attenuation correction), or the effects are ignored (e.g., positron range). In emission tomography, many of these phenomena have not been significant due to the relatively poor spatial resolution of current systems. But with the expectation of smaller spatial resolution in the near future (28), a physically accurate reconstruction model becomes vitally important.

The EM emission algorithm allows introduction of these physical factors directly into the reconstruction model through specification of the c_{ij} in Eq. 2. The following factors can be included in the c_{ij}:

1. Counting time per projection and known radioactive decay.

2. Normalization for redundant projection sampling and detector pair efficiency.

3. Variation in spatial resolution. This is particularly useful when there is a significant variation of resolution with depth from the detectors, an effect which is a more significant problem in SPECT (25), but should also be considered in PET (29). Accurate specification of these variations into the physical model should noticeably improve the quality of the reconstructed image.

4. Attenuation. There are a variety of schemes for performing attenuation correction in emission tomography (16,30,31). For the EM algorithm, when the attenuation medium is known or can be estimated, this information can be incorporated in the c_{ij} instead of "correcting" the projection data before reconstruction. For PET, the probability that an event on any projection line i will not be attenuated is nearly independent of the pixel of origin. Using the notation of Table 1, this probability is given in Eq. 7. For SPECT, events from each pixel j that contributes to projection i have differing probabilities of passing to the detector without attenuation. The analog of Eq. 7 is

$$e^{-\sum_{k \in K_{ij}} l_{ij}\mu_k},$$

where K_{ij} denotes the set of pixels between pixel j and detector i. Note that the above formulas are approximations due to the finite size of the detectors (32). The exact specification of attenuation effects through the c_{ij}, although computationally burdensome, should provide a much more quantitatively accurate reconstruction. An average value for attenuation coefficient for the entire field could also be used to reduce the computations at the expense of accuracy. A compromise is to use boundary information to subdivide the field into regions of uniform attenuation (33).

5. Scatter. A photon emitted from any pixel can be deflected and ultimately detected along an inappropriate projection line. This effect causes all projection measurements to be affected by the activity in all pixels to a small degree. Theoretically a quite exact scatter correction can be incorporated into the c_{ij}, given complete knowledge of the attenuation medium. However, such an addition to the algorithm requires many assumptions and would add substantially to computation time, which is directly related to the number of c_{ij} that are nonzero.

6. Accidental coincidences. Random coincidences occur in PET due to the accidental arrival of two independent photons at opposing detectors within the coincidence timing window of the scanner. One corrective (17) is to independently "measure" the randoms and subtract them from the coincidence measurements to provide "true coincidence" data. In this case the actual projection measurement is a difference of two independent Poisson measurements, which is not itself Poisson. This deviation will cause significant errors in the use of the EM algorithm if the random fraction is high. The EM emission algorithm can be extended to include the effect of randoms by recognizing that

$$Y_i = \sum_{j \in I_i} X_{ij} + A_i,$$

where A_i is the random number of accidental coincidences detected along projection i. The A_i have Poisson distribution with expected value a_i and are independent of the X_{ij}. By extending the vector of parameters to include the a_i, the E-step becomes

$$N_{ij} = \frac{c_{ij}\,\lambda_j^n\,Y_i}{\sum_{k \in I_i} c_{ik}\,\lambda_k^n + a_i^n},$$

where a_i^n is the estimate of a_i at the nth iteration. A good initial estimate for the a_i can be supplied by the independent random measurement. Unfortunately, this approach will double the size of the parameter vector, causing the estimation to be underdetermined. A good compromise is to treat the randoms as a known constant, i.e., replace a_i^n in the above equation with the measured a_i. This approximation will introduce a percentage error in the calculation of N_{ij} on the order of $a_i^{1/2}/Y_i$ near convergence. The error of the original method (subtracting measured randoms) is larger by a factor of at least $1 + (1.5\ a_i/Y_i)$. Note, however, that the main application of the EM algorithm is expected to be in low count conditions when the random fraction and all these errors are small.

7. Positron range and angulation effects. As spatial resolution improves in PET systems, errors due to the range of the positron before annihilation and due to deviations in the angle of flight of the two photons from 180° will become significant (34). One approach for an analytic incorporation of these effects into the c_{ij} has already been suggested (3).

8. Time-of-flight information. For positron tomographs designed to collect photon time-of-flight data (35), the EM algorithm can be extended by subdividing each observation Y_i into a finite number of locations along the projection line. The effects of timing resolution can then be incorporated into the c_{ij}.

The practical application of the EM algorithm to an emission reconstruction should involve a preliminary direct measurement of as many of the aforementioned physical effects as possible. For example, a high count, pixel-sized source could be sequentially positioned at each pixel position and projection measurements could be collected. This procedure will provide information about detector efficiencies, variation of resolution with position, positron range, and angulation effects. An individual reconstruction would employ this base data set and incorporate scanning time, decay rates, and individual attenuation data, if known. Note that the absolute scaling of the c_{ij} is not important.

Transmission Algorithm

The EM transmission reconstruction algorithm does not provide the degree of generality in specification of the physical model that is possible in the emission case. It is a fundamental assumption that each projection is infinitely thin. The physical model is expressed through the l_{ij}, which are directly derivable from the scanner configuration. It may, however, be possible to incorporate a correction for the beam hardening effect (36,37) by using measurements of the relative effects of a given attenuation coefficient on a range of photon energies (38). First a rule must be defined that summarizes these measurements into a number of energy classes. The algorithm is modified so that a typical complete data random variable X_{ijk} becomes the number of photons from projection i entering pixel j of photon energy class k. With knowledge of the energy distribution of the source for each projection, the E-step can be performed using Eq. 9, modified by the supplied rule. The M-step, Eq. 15, is extended so that the numerator and denominator sum over all energy levels as well as projections.

Comparison with Other Reconstruction Algorithms

The distinct advantages of the EM reconstruction algorithms over other techniques lie in their use of a proper statistical model and the log likelihood criterion as a measure of the quality of a reconstruction. Other iterative, algebraic schemes can incorporate the same physically reasonable constants, c_{ij} and l_{ij}, but they are not based on the Poisson assumptions. They can also accommodate some constraints, but they cannot deal with the crucial nonnegativity constraints as easily as the EM algorithms. Fourier and convolution techniques do have many practical advantages, the most obvious of which is computation speed. Without very specialized computer hardware, it is unlikely that any implementation of the EM algorithms will produce computation times much better than an order of magnitude greater than that of comparable Fourier algorithms. EM performance times relative to other iterative techniques need to be checked. Fourier techniques have two other useful features. One is the ability to give reasonable reconstructions even when there are fewer projections than pixels. In such underdetermined systems, the final pixel values for algebraic techniques depend on the initial pixel extimates. Another feature of Fourier methods is their ability to reconstruct only a portion of the field of view, whereas all algebraic techniques must reconstruct the entire field.

CONCLUSIONS

The EM algorithms, presented above, provide a highly attractive approach to the reconstruction problem. They are based on a stochastic model that allows direct specification of most physical features

affecting the projection measurements. The formulation is based on well accepted maximum likelihood techniques, and appropriate use of the EM algorithm provides a practical approach toward finding the maximum likelihood solution. If the model is specified correctly, this algorithm should provide the best possible image reconstruction from the data. Therefore, even if implementation times are burdensome, the EM algorithms will provide a benchmark to define the limits of a particular scanning protocol in terms of noise and spatial resolution. However, as stated previously, the EM reconstruction should offer a particularly substantial improvement in two areas. First, for SPECT, the problems of variation of resolution with depth and attenuation correction can be handled appropriately, providing a more quantitative reconstruction. Second, for both emission and transmission tomography, but particularly in the emission case, the use of an accurate statistical model should dramatically improve the reconstruction of images with low counts.

APPENDIX

Convergence of the EM Algorithms

Our aim here is to prove that the EM iterates converge to the appropriate maximum likelihood estimate regardless of their initial value. In the emission case Shepp and Vardi (3) demonstrate that, if the iterates converge, they necessarily converge to the maximum likelihood estimate. This is a weaker property than global convergence. Our proof closely follows the developments in (12) and (13) except that we must contend with boundaries. The concavity of the log likelihood more than compensates for this added complication.

To proceed with the theory, we revert to the neutral parameter vector θ representing either intensities or attenuation coefficients. Each component θ_j satisfies the constraint $\theta_j \geq 0$. We have already remarked that $\theta_j^{n+1} > 0$ if and only if $\theta_j^n > 0$. To avoid getting stuck permanently on a boundary, our initial estimates θ^0 will have all components positive. This does not preclude the possibility

$$\lim_{n \to \infty} \theta_j^n = 0$$

for some j. In the emission case the iteration map $\lambda^n \to \lambda^{n+1}$ is undefined if both

$$\sum_{k \in I_i} c_{ik} \lambda_k^n = 0$$

and $Y_i > 0$ for the same i. However, such a singularity can never be approached because $\ln g(Y,\lambda)$ increases at each iteration. For notation convenience, we let

$$L(\theta) = \ln g(Y,\theta)$$
$$Q(\theta|\theta^n) = E(\ln f(X,\theta)|Y,\theta^n).$$

Implicit in our previous discussion is the fact that the maximum of $L(\theta)$ exists. This fact is a straightforward consequence of the continuity of $L(\theta)$ and the fact that $L(\theta)$ tends to $-\infty$ as any θ_j tends to $+\infty$. To ensure the falloff of $L(\theta)$ at ∞ in the transmission case it is necessary to assume that for each pixel j, there is some projection $i \in J_j$ with $Y_i > 0$. We will make this reasonable assumption. We will also assume that $L(\theta)$ is strictly concave so as to eliminate the possibility of more than one maximum point. With these preliminaries attended to we can state the first of three lemmas.

Lemma 1: The iterates θ^n all belong to the same compact convex set.

Proof: Due to the monotone nature of the EM algorithm, the iterates all belong to the set

$$\{\theta: L(\theta) \geq L(\theta^0)\}.$$

This set is compact because of the continuity of L and its behavior at ∞. It is convex because L is concave.

Lemma 2: The Euclidean distance $\|\theta^{n+1} - \theta^n\|$ between successive iterates tends to 0 as n tends to ∞.

Proof: The inequality

$$L(\theta^{n+1}) - L(\theta^n) \geq Q(\theta^{n+1}|\theta^n) - Q(\theta^n|\theta^n) \qquad (18)$$

holds because $\theta \to H(\theta|\theta^n) = Q(\theta|\theta^n) - L(\theta)$ achieves its maximum at $\theta = \theta^n$. Now expand $\theta \to Q(\theta|\theta^n)$ to second order about the point θ^{n+1} using the fact that the first partials vanish:

$$Q(\theta^{n+1}|\theta^n) - Q(\theta^n|\theta^n)$$
$$= -\tfrac{1}{2}(\theta^{n+1} - \theta^n)^t \frac{\partial^2}{\partial\theta^2} Q(\theta|\theta^n)\Big|_{\theta=\bar{\theta}} (\theta^{n+1} - \theta^n), \qquad (19)$$

where $\bar{\theta}$ is on the line segment between θ^n and θ^{n+1} and where $\frac{\partial^2}{\partial\theta^2} Q(\theta|\theta^n)$ is the matrix of second partials of Q. The Eqs. 5 and 12 for the first partials show that only the diagonal terms in this matrix are nonzero. Hence, combining Eqs. 18 and 19 yields

$$L(\theta^{n+1}) - L(\theta^n)$$
$$\geq \sum_j (\theta_j^{n+1} - \theta_j^n)^2 \left(-\tfrac{1}{2} \frac{\partial^2}{\partial\theta_j^2} Q(\theta|\theta^n)\Big|_{\theta=\bar{\theta}} \right).$$

Suppose we can show that

$$\min_j -\tfrac{1}{2} \frac{\partial^2}{\partial\theta_j^2} Q(\theta|\theta^n)\Big|_{\theta=\bar{\theta}} \qquad (20)$$

is bounded below by a positive constant c, which does not depend on the particular choice of n. Then

$$\sum_j (\theta_j^{n+1} - \theta_j^n)^2 \leq c^{-1} \{L(\theta^{n+1}) - L(\theta^n)\}. \qquad (21)$$

Since $L(\theta^n)$ is bounded and monotone increasing in n, the right-hand side of Eq. 21 tends to 0.

Hence, we need to bound the Eq. 20. Consider the emission case first. Direct differentiation of Eq. 5 shows

$$-\frac{\partial^2}{\partial\lambda_j^2} Q(\lambda|\lambda^n) = \sum_{i \in J_j} N_{ij} \lambda_j^{-2}$$
$$= \lambda_j^{-1} \left\{ \frac{\partial}{\partial\lambda_j} Q(\lambda|\lambda^n) + \sum_{i \in J_j} c_{ij} \right\}.$$

Obviously, $-\frac{\partial^2}{\partial\lambda_j^2} Q(\lambda|\lambda^n)$ depends on λ only through λ_j

and is monotone decreasing in λ_j. Furthermore,

$$-\frac{\partial^2}{\partial\lambda_j^2}Q(\lambda|\lambda^n)|_{\lambda=\lambda^{n+1}} = \frac{1}{\lambda_j^{n+1}}\sum_{i\epsilon J_j} c_{ij} \tag{22}$$

$$-\frac{\partial^2}{\partial\lambda_j^2}Q(\lambda|\lambda^n)|_{\lambda=\lambda^n} = \sum_{i\epsilon J_j} N_{ij}\frac{1}{(\lambda_j^n)^2} = \frac{1}{\lambda_j^n}\sum_{i\epsilon J_j} c_{ij}\, Y_i \frac{1}{\sum_{k\epsilon I_i} c_{ik}\,\lambda_k^n}. \tag{23}$$

Since $\bar{\lambda}_j$ lies on the line segment between λ_j^n and λ_j^{n+1}, it is clear that

$$-\frac{\partial^2}{\partial\lambda_j^2}Q(\lambda|\lambda^n)|_{\lambda=\bar{\lambda}}$$

is greater than the smaller of the two quantities (Eqs. 22 and 23), both of which are bounded below by a strictly positive constant on $\{\lambda:L(\lambda) \geq L(\lambda^0)\}$.

The transmission case is much the same. Differentiating Eq. 12 yields

$$-\frac{\partial^2}{\partial\mu_j^2}Q(\mu|\mu^n) = \sum_{i\epsilon J_j}(M_{ij} - N_{ij})\frac{l_{ij}^2\, e^{l_{ij}\mu_j}}{[e^{l_{ij}\mu_j} - 1]^2}$$

$$\geq [(\min_{i\epsilon J_j} l_{ij}\, e^{l_{ij}\mu_j}/(e^{l_{ij}\mu_j} - 1)]\left\{\sum_{i\epsilon J_j}(M_{ij} - N_{ij})l_{ij}/(e^{l_{ij}\mu_j} - 1)\right\}.$$

The first term

$$\min_{i\epsilon J_j} l_{ij}e^{l_{ij}\mu_j}/(e^{l_{ij}\mu_j} - 1)$$

is bounded below by a positive constant on the compact set $\{\mu:L(\mu) \geq L(\mu^0)\}$. The second term

$$\sum_{i\epsilon J_j}(M_{ij} - N_{ij})\frac{l_{ij}}{e^{l_{ij}\mu_j} - 1} = \frac{\partial}{\partial\mu_j}Q(\mu|\mu^n) + \sum_{i\epsilon J_j} N_{ij}l_{ij} \tag{24}$$

is monotone decreasing in μ_j. At the endpoint μ^{n+1}, Eq. 24 equals $\sum_{i\epsilon J_j} N_{ij}l_{ij}$. At the other endpoint μ^n, it amounts to

$$\sum_{i\epsilon J_j}(M_{ij} - N_{ij})\, l_{ij}/(e^{l_{ij}\mu_j^n} - 1)$$

$$= \sum_{i\epsilon J_j}\{E(M_{ij}) - E(N_{ij})\}\, l_{ij}/(e^{l_{ij}\mu_j^n} - 1) = \sum_{i\epsilon J_j} E(N_{ij})\, l_{ij}, \tag{25}$$

using Eq. 10 with $E(M_{ij})$ and $E(N_{ij})$ as the expected number of photons entering and leaving pixel j, respectively. The sum (Eq. 25) is also bounded below by a positive constant on $\{\mu:L(\mu) \geq L(\mu^0)\}$.

Lemma 3: Let θ^* be a limit point of the sequence θ^n. Then

$$\frac{\partial}{\partial\theta_j}L(\theta^*) = 0$$

whenever $\theta_j^* > 0$. Furthermore, there are only a finite number of limit points.

Proof: Let θ^{n_k} be a subsequence that approaches θ^*. Also, let j be an index with $\theta_j^* > 0$. Since $\|\theta^{n_k+1} - \theta^{n_k}\| \to 0$,

$$\frac{\partial}{\partial\theta_j}L(\theta^*) = \lim_{k\to\infty}\frac{\partial}{\partial\theta_j}L(\theta^{n_k}) = \lim_{k\to\infty}\frac{\partial}{\partial\theta_j}Q(\theta|\theta^{n_k})|_{\theta=\theta^{n_k}}$$

$$= \lim_{k\to\infty}\frac{\partial}{\partial\theta_j}Q(\theta|\theta^{n_k})|_{\theta=\theta^{n_k+1}} = 0.$$

Here we have employed the joint continuity of $\frac{\partial}{\partial\theta_j}Q(\theta|\theta')$ in (θ,θ') for $\theta,\theta' > 0$.

To show that there are only a finite number of limit points, consider two such points θ^* and θ^{**}. Suppose the two sets

$$S^* = \{j:\theta_j^* = 0\}$$
$$S^{**} = \{j:\theta_j^{**} = 0\}$$

coincide. Let $L_{S^*}(\theta)$ be the restriction of $L(\theta)$ to the set

$$\{\theta:\theta_j = 0 \text{ for } j\epsilon S^*\}.$$

$L_{S^*}(\theta)$ is a strictly concave function of the reduced number of parameters. As such, it can have at most one stationary point, which is necessarily the maximum point. It follows that θ^* and θ^{**} coincide. The number of limit points is obviously bounded above by the number of subsets of the set of parameter indices.

Theorem: $\lim_{n\to\infty}\theta^n = \theta^\infty$, where θ^∞ is the maximum point of $L(\theta)$.

Proof: It suffices to prove that there is at most one limit point θ^∞ and that θ^∞ is the maximum point. To prove the first contention we invoke a result highlighted in (13); namely, the set of limit points of a bounded sequence θ^n is connected and compact provided $\|\theta^{n+1} - \theta^n\| \to 0$ (39). According to Lemma 3, the set of possible limit points is finite in the present situation. The only way a finite set can be connected is for it to consist of a single point. Thus, $\lim_{n\to\infty}\theta^n = \theta^\infty$ exists.

To complete the proof we must show that θ^∞ is a Kuhn-Tucker point of $L(\theta)$ (40,41). In the presence of boundaries, such points are the analogs of stationary points. Since $L(\theta)$ is strictly concave, there is at most one Kuhn-Tucker point, and this necessarily gives the maximum.

With the present simple boundaries, θ^∞ is a Kuhn-Tucker point if and only if

$$\frac{\partial}{\partial\theta_j}L(\theta^\infty)\begin{cases} = 0 & \theta_j^\infty > 0 \\ \leq 0 & \theta_j^\infty = 0. \end{cases}$$

Since the case $\theta_j^\infty > 0$ is already taken care of by Lemma 3, suppose $\theta_j^\infty = 0$ and

$$\frac{\partial}{\partial\theta_j}L(\theta^\infty) > 0.$$

By continuity,

$$\frac{\partial}{\partial\theta_j}L(\theta^n) = \frac{\partial}{\partial\theta_j}Q(\theta|\theta^n)|_{\theta=\theta^n} > 0$$

holds for all n large enough. But

$$\frac{\partial}{\partial\theta_j}Q(\theta|\theta^n)$$

is strictly decreasing in θ_j. This means $\theta_j^n < \theta_j^{n+1}$ for all large n, a contradiction to the assumption $\theta_j^n \to 0$. The proof is complete.

Acknowledgment: The authors wish to thank Dr. Sungcheng Huang for offering helpful suggestions and Avis Williams for preparation of the manuscript. This work was supported in part by the University of California at Los Angeles, NIH Research Career Development Award K04 HD00307, and NRSA Training Grant GM7191.

REFERENCES

1. Herman GT, ed. *Topics in applied physics, vol. 32. Image reconstruction from projections: implementation and applications*. Berlin: Springer-Verlag, 1979.
2. Herman GT. *Image reconstruction from projections: the fundamentals of computerized tomography*. New York: Academic Press, 1980.
3. Shepp LA, Vardi Y. Maximum likelihood reconstruction for emission tomography. *IEEE Trans Med Imaging* 1982;MI-1:113–22.
4. Rockmore AJ, Macovski A. A maximum likelihood approach to image reconstruction. *Proc Joint Automatic Control Conf* 1977;782–6.
5. Rao CR. *Linear statistical inference and its applications*. 2nd ed. New York: Wiley, 1973:59.
6. Snyder DL, Politte DG. Image reconstruction from list-mode data in an emission-tomography system having time-of-flight measurements. *IEEE Trans Nucl Sci* 1983;NS-30:1843–9.
7. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Series B* 1977:39:1–38.
8. Baum LE, Petrie T, Soules G, Weiss N. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann Math Stat* 1970;41:164–71.
9. Hartley HO, Hocking RR. The analysis of incomplete data. *Biometrics* 1971;27:783–808.
10. Orchard T, Woodbury MA. A missing information principle: theory and applications. In: *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability*. Univ. Calif. Press: 1972;1:697–715.
11. Sundberg R. Maximum likelihood theory for incomplete data from an exponential family. *Scand J Stat* 1974;1:49–58.
12. Boyles RA. On the convergence of the EM algorithm. *J R Stat Soc Series B* 1983;45:47–50.
13. Wu CF. On the convergence properties of the EM algorithm. *Ann Stat* 1983;11:95–103.
14. Feller W. *An introduction to probability theory and its applications. Vol. 1*. 3rd ed. New York: Wiley, 1968:216–37.
15. Shepp LA, Logan BF. The Fourier reconstruction of a head section. *IEEE Trans Nucl Sci* 1974;NS-21:21–43.
16. Budinger TF, Gullberg GT. Transverse section reconstruction of gamma-ray emitting radionuclides in patients. In: Ter-Pogossian MM, Phelps ME, Brownell GL, Cox JR, Davis DO, Even RG, eds. *Reconstruction tomography in diagnostic radiology and nuclear medicine*. Baltimore: University Park Press, 1977:315–42.
17. Hoffman EJ, Phelps ME, Huang SC, et al. A new tomograph for quantitative positron emission computed tomography of the brain. *IEEE Trans Nucl Sci* 1981;NS-28:99–103.
18. Phelps ME, Mazziotta JC, Huang SC. Study of cerebral function with positron computed tomography. *J Cereb Blood Flow Metabol* 1982;2:113–62.
19. Chesler DA, Riederer SJ, Pelc NJ. Noise due to photon counting statistics in computed X-ray tomography. *J Comput Assist Tomogr* 1977;1:64–74.
20. Huang SC, Phelps ME, Hoffman EJ, Kuhl DE. Cubic splines for filter design in CT. *IEEE Trans Nucl Sci* 1980;NS-27:1368–74.
21. Herman GT, Lent A, Lutz PH. Relaxation methods for image reconstruction. *Comm Assoc Comput Mach* 1978;21:153–8.
22. Gordon R, Bender R, Herman GT. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *J Theor Biol* 1970;29:471–81.
23. Gull SF, Daniell GJ. Image reconstruction from incomplete and noisy data. *Nature* 1978;272:686–90.
24. Herman GT, Lent A. Iterative reconstruction algorithms. *Comput Biol Med* 1976;6:273–94.
25. Budinger TF. Physical attributes of single-photon tomography. *J Nucl Med* 1980;21:579–92.
26. Hoffman EJ, Phelps ME. An analysis of some of the physical aspects of positron transaxial tomography. *Comput Biol Med* 1976;6:345–60.
27. Phelps ME, Hoffman EJ, Huang SC, Kuhl DE. Design considerations in positron computed tomography (PCT). *IEEE Trans Nucl Sci* 1979;NS-26:2746–51.
28. Phelps ME, Huang SC, Hoffman EJ, Plummer D, Carson R. An analysis of signal amplification using small detectors in positron emission tomography. *J Comput Assist Tomogr* 1982;6:551–65.
29. Hoffman EJ, Huang SC, Plummer D, Phelps ME. Quantitation in positron emission computed tomography: 6. Effect of nonuniform resolution. *J Comput Assist Tomogr* 1982;6:987–99.
30. Phelps ME. Emission computed tomography. *Semin Nucl Med* 1977;7:337–65.
31. Lewis ML, Willerson JT, Lewis SE, Bonte FJ, Parkey RW, Stokely EM. Attenuation compensation in single-photon emission tomography: a comparative evaluation. *J Nucl Med* 1982;23:1121–7.
32. Huang SC, Hoffman EJ, Phelps ME, Kuhl DE. Quantitation in positron emission computed tomography: 2. Effects of inaccurate attenuation correction. *J Comput Assist Tomogr* 1979;3:804–14.
33. Huang SC, Carson RE, Phelps ME, Hoffman EJ, Schelbert HR, Kuhl DE. A boundary method for attenuation correction in positron computed tomography. *J Nucl Med* 1981;22:627–37.
34. Phelps ME, Hoffman EJ, Huang SC, Ter-Pogossian MM. Effect of positron range on spatial resolution. *J Nucl Med* 1975;16:649–52.
35. Ter-Pogossian MM, Mullani NA, Ficke DC, Markham J, Snyder DL. Photon time-of-flight assisted positron emission tomography. *J Comput Assist Tomogr* 1981;5:227–39.
36. Brooks RA, DiChiro G. Beam hardening in X-ray reconstructive tomography. *Phys Med Biol* 1976;21:390–8.
37. Nalcioglu O, Lou RY. Post-reconstruction method for beam hardening in computerised tomography. *Phys Med Biol* 1979;24:330–40.
38. Phelps ME, Hoffman EJ, Ter-Pogossian MM. Attenuation coefficients of various body tissues, fluids, and lesions at photon energies of 18 to 136 keV. *Radiology* 1975;117:573–83.
39. Ostrowski AM. *Solution of equations in Euclidean and Banach spaces*. New York: Academic Press, 1973:173.
40. Beltrami EJ. *An algorithmic approach to nonlinear analysis and optimization*. New York: Academic Press, 1970.
41. Hestenes MR. *Optimization theory: the finite dimensional case*. Huntington, NY: Krieger Publishing, 1981.