

# Joint Maximum Likelihood Estimation of Emission and Attenuation Densities in PET

Neal H. Clinthorne, Jeffrey A. Fessler, Gary D. Hutchins, and W. Leslie Rogers  
The University of Michigan, Division of Nuclear Medicine  
Ann Arbor, MI 48109-0552.

## Abstract

Accurate attenuation correction can be performed in PET using transmission scanning to estimate the survival probabilities along each coincidence line. However, since these measurements are typically corrupted by Poisson counting noise, they propagate additional uncertainty into reconstructed images and kinetic parameter estimates. This can be especially true in the thorax where the attenuating medium is heterogeneous and the statistical precision of the transmission scan may be approximately the same as that of the emission data. To account for the Poisson noise in the transmission measurement, we have developed a sieve-constrained maximum likelihood algorithm that jointly estimates both the survival probability and emission intensity.

## I. INTRODUCTION

In positron emission tomography, accurate quantification of emission intensity will result only if  $\gamma$ -ray absorption is accounted for in the reconstruction process [1]. This usually entails pre-correcting the projection data by dividing each projection element by the *survival probability* of  $\gamma$ -rays emitted along each coincidence chord. In [2] Politte and Snyder noted that this pre-correction is inconsistent with the Poisson projection data model and developed constrained maximum likelihood reconstruction algorithms for the case in which the survival probabilities are known exactly. It is usually the case, however, that these parameters are not known and must be estimated using a transmission scan with a positron-emitting transmission source. The counting noise inherent in this scan propagates additional errors into the reconstruction. In regions where the distribution of attenuator is relatively constant, it may be reasonable to simply smooth the resulting survival probability estimates. But in regions such as the thorax, where the attenuating medium is extremely heterogeneous, this may result in additional reconstruction errors due to the bias introduced by smoothing.

In this paper we offer an interesting alternative—joint maximum likelihood estimation of both the survival probabilities and the emission intensity. In the next section, we

describe the assumptions and extensions necessary to formulate the joint estimation problem. In Section III we describe alternative methods for solving the likelihood maximization problem and develop a constrained joint alternate and maximize (JAM) version of the generalized EM algorithm. Section IV presents some of our initial experiences in using the JAM algorithm with simulated PET data.

## II. THEORY

### A. Emission and Transmission Response Model

To form the likelihood it is necessary to have a statistical model relating our measured projections to the underlying parameters of interest, the emission intensity and the survival probabilities. We assume that effects of the system geometry on the measurements are known and can be represented as a  $D \times B$  matrix  $A$  where  $D$  is the total number of coincidence lines and  $B$  is the number of object pixels. Moreover, we require that the random projection vector  $\mathbf{Y} = [Y_1, \dots, Y_D]^T$  has the following distribution

$$\mathbf{Y} \sim \text{Pois}(\underline{\mu} \odot A \underline{\lambda} + \underline{\eta}), \quad (1)$$

where  $\odot$  signifies element-by-element vector multiplication,  $\underline{\lambda} = [\lambda_1, \dots, \lambda_B]^T$  is the vector of emission intensities,  $\underline{\mu} = [\mu_1, \dots, \mu_D]^T$  is the vector of survival probabilities, and in accordance with the random coincidence model given in [3],  $\underline{\eta} = [\eta_1, \dots, \eta_D]$  is a vector representing the random coincidence rate. In (1), the survival probabilities  $\underline{\mu}$  have been separated from the response matrix  $A$  by assuming that they are invariant with respect to the emission location along each coincidence line (a reasonable model for attenuation in PET tomographs having good resolution).

For the transmission scan we assume that this response-invariance extends to the location of the transmission source and that the random vector of transmission measurements  $\mathbf{M} = [M_1, \dots, M_D]^T$  has distribution

$$\mathbf{M} \sim \text{Pois}(\underline{\mu} \odot \underline{\Delta}). \quad (2)$$

In (2) the vector  $\underline{\Delta}$  summarizes the effects of transmission source strength and transmission scanning time relative to the total emission scanning time.

<sup>0</sup>This work was supported by the National Cancer Institute, DHHS, under Grant RO1-CA54362.

### B. Joint Likelihood

Armed with the above model for the projection and transmission measurements, and with the knowledge that these measurements are independent, we can immediately form the joint log-likelihood by simply adding the individual log-likelihoods,

$$L(\underline{\lambda}, \underline{\mu}) = \mathbf{y}^T \log(\underline{\mu} \odot A\underline{\lambda} + \underline{\eta}) - \mathbf{1}^T(\underline{\mu} \odot A\underline{\lambda}) + \mathbf{m}^T \log \underline{\mu} \odot \underline{\Lambda} - \mathbf{1}^T(\underline{\mu} \odot \underline{\Lambda}) + C, \quad (3)$$

where  $\mathbf{y} = [y_1, \dots, y_D]^T$  and  $\mathbf{m} = [m_1, \dots, m_D]^T$  represent realizations of the random vectors  $\mathbf{Y}$  and  $\mathbf{M}$  respectively,  $\mathbf{1}$  represents a vector of ones of the appropriate length, and  $C$  is a term independent of  $\underline{\lambda}$  or  $\underline{\mu}$ .

We maximize (3), by simultaneously choosing vectors  $\underline{\lambda}$  and  $\underline{\mu}$  from the admissible domain

$$\mathcal{D} = \{ \underline{\lambda}, \underline{\mu} : \lambda_b \geq 0, b = 1, \dots, B \\ 0 < \mu_d \leq 1, d = 1, \dots, D \}.$$

If these maximizing vectors lie in the interior of  $\mathcal{D}$ , they will satisfy  $\nabla_{\underline{\lambda}} L(\underline{\lambda}, \underline{\mu}) = \nabla_{\underline{\mu}} L(\underline{\lambda}, \underline{\mu}) = 0$  where

$$\nabla_{\underline{\lambda}} L(\underline{\lambda}, \underline{\mu}) = A^T((\mathbf{y} \odot \underline{\mu}) \odot (\underline{\mu} \odot A\underline{\lambda} + \underline{\eta})) - A^T \underline{\mu} \quad (4)$$

$$\nabla_{\underline{\mu}} L(\underline{\lambda}, \underline{\mu}) = (\mathbf{y} \odot A\underline{\lambda}) \odot (\underline{\mu} \odot A\underline{\lambda} + \underline{\eta}) - A\underline{\lambda} + \mathbf{m} \odot \underline{\mu} - \underline{\Lambda}, \quad (5)$$

and the symbol  $\odot$  denotes element-by-element vector division. Furthermore, if the likelihood is strictly concave, these vectors will be the unique, global maximizer of  $L(\underline{\lambda}, \underline{\mu})$ .

### C. Properties of Joint Likelihood

In addition to being a sufficient condition for the existence of a unique global maximum of (3), strict concavity is a necessary ingredient in many of the global convergence proofs for iterative maximization techniques [4,5]. For this condition to hold for (3), the following Hessian matrix must be positive-definite:

$$-\nabla^2 L(\underline{\lambda}, \underline{\mu}) = \begin{bmatrix} A^T & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}, \quad (6)$$

where

$$B_{11} = \text{diag}(\mathbf{y} \odot \underline{\mu}^2 \odot \hat{\mathbf{y}}^2(\underline{\lambda}, \underline{\mu})), \\ B_{12} = B_{21} = \text{diag}(\mathbf{1} - \mathbf{y} \odot \underline{\eta} \odot \hat{\mathbf{y}}^2(\underline{\lambda}, \underline{\mu})), \\ B_{22} = \text{diag}(\mathbf{y} \odot (A\underline{\lambda})^2 \odot \hat{\mathbf{y}}^2(\underline{\lambda}, \underline{\mu}) + \mathbf{m} \odot \underline{\mu}^2),$$

and

$$\hat{\mathbf{y}}(\underline{\lambda}, \underline{\mu}) = \underline{\mu} \odot A\underline{\lambda} + \underline{\eta}. \quad (7)$$

Unfortunately, we see that  $L(\underline{\lambda}, \underline{\mu})$  is not necessarily strictly concave and the global convergence properties of iterative methods may be difficult to establish. Nevertheless, the following analysis shows that under reasonable

conditions, the likelihood has a large area over which strict concavity exists. If we restrict our estimates to this region, any algorithm which increases the likelihood at each iteration will converge to the regional maximum. (However, we have yet to establish that the global maximum on  $\mathcal{D}$  always lies within this region.)

For (6) to be positive-definite, it is sufficient that the  $B$ -matrix be positive definite. This implies that  $b_{ii}b_{jj} > b_{ij}b_{ji}, \forall i, j$  and, in the absence of random coincidences, leads to the following conditions

$$y_d(y_d + m_d) > \hat{\mu}_d^2 \left( \sum_{b=1}^B a_{db} \hat{\lambda}_b \right)^2, \quad d = 1, \dots, D. \quad (8)$$

If there are a sufficient number of events in the emission and transmission data, it is extremely unlikely that we would consider estimates  $\hat{\underline{\lambda}}$  and  $\hat{\underline{\mu}}$  that violate these conditions. This also applies when random coincidences comprise a small fraction of the total events; however, we have not completed our analysis in the high random coincidence regime.

## III. ALGORITHMS FOR MAXIMIZATION

### A. The EM Algorithm—Choice of Complete Data

To develop the joint maximum likelihood reconstruction algorithm, we start with the “natural” complete data space for emission tomography introduced by Shepp and Vardi [6] and extended to include random coincidences by Politte and Snyder in [2] where it is termed “**ML-IB**.” In this complete-data space, each observed projection element is composed as a sum of the unobserved complete data

$$Y_d = \sum_b N_{db} + R_d, \quad \begin{matrix} d = 1, \dots, D \\ b = 1, \dots, B \end{matrix}, \quad (9)$$

where  $Y_d$  represents the  $d$ th component of the random projection vector  $\mathbf{Y}$ ,  $N_{db}$  is the number of events emitted from the  $b$ th pixel in the object detected on coincidence line  $d$ , and  $R_d$  represents an additive, random coincidence component. To perform the joint estimation, we augment this complete data in the obvious way—by simply including the measured transmission data  $\mathbf{M}$ . The resulting complete data space, which we term **ML-JB**, can be represented as  $\mathbf{X}_{JB} = \{ \{N_{db}\}, \{R_d\}, \{M_d\} \}$ .

To develop the E-step of the EM algorithm we form the expected value of the log-likelihood of the complete data, conditioned on the incomplete, observed data and the current parameter estimates:

$$E[\log f(\mathbf{X} | \hat{\underline{\lambda}}, \hat{\underline{\mu}}) | \mathbf{y}, \mathbf{m}, \hat{\underline{\lambda}}^{(k)}, \hat{\underline{\mu}}^{(k)}] = \sum_{d,b} [E[N_{db} | \mathbf{y}, \hat{\underline{\lambda}}^{(k)}, \hat{\underline{\mu}}^{(k)}] \log \hat{\mu}_d a_{db} \hat{\lambda}_b - \hat{\mu}_d a_{db} \hat{\lambda}_b] + \sum_d [E[R_d | \mathbf{y}, \hat{\underline{\lambda}}^{(k)}, \hat{\underline{\mu}}^{(k)}] \log \eta_d - \eta_d] + \sum_d [E[M_d | \mathbf{m}, \hat{\underline{\mu}}^{(k)}] \log \mu_d \Lambda_d - \mu_d \Lambda_d]. \quad (10)$$

The expectations on the right side of (10) are given by

$$E[N_{ab}|\mathbf{y}, \hat{\lambda}, \hat{\mu}] = y_d \frac{\hat{\mu}_d a_{ab} \lambda_b}{\hat{y}_d(\hat{\lambda}, \hat{\mu})}, \quad (11)$$

$$E[R_d|\mathbf{y}, \hat{\lambda}, \hat{\mu}] = y_d \frac{\eta_d}{\hat{y}_d(\hat{\lambda}, \hat{\mu})}, \quad (12)$$

$$E[M_d|\mathbf{m}, \hat{\mu}] = m_d, \quad (13)$$

where

$$\hat{y}_d(\hat{\lambda}, \hat{\mu}) = \hat{\mu}_d \sum_{b=1}^B a_{db} \hat{\lambda}_b + \eta_d. \quad (14)$$

The maximization, or M-step of the algorithm requires that we choose new parameter estimates  $\hat{\lambda}$  and  $\hat{\mu}$  that maximize (10). Substituting the above expectations into (10), and differentiating the resulting expression with respect to  $\hat{\lambda}$  and  $\hat{\mu}$  yields the following equations which must be solved for the M-step:

$$\frac{1}{\hat{\lambda}_b} \sum_{d=1}^D \left[ \frac{y_d a_{db} \hat{\mu}_d^{(k)}}{\hat{\mu}_d^{(k)} \sum_b a_{db} \hat{\lambda}_b^{(k)} + \eta_b} \right] - \sum_{d=1}^D a_{db} \hat{\mu}_d = 0, \quad (15)$$

$$\frac{1}{\hat{\mu}_d} \left[ \frac{y_d \sum_b a_{db} \hat{\lambda}_b^{(k)}}{\hat{\mu}_d^{(k)} \sum_b a_{db} \hat{\lambda}_b^{(k)} + \eta_d} + m_d \right] - \left( \sum_{b=1}^B a_{db} \hat{\lambda}_b + \Lambda_d \right) = 0. \quad (16)$$

Unfortunately, these equations are coupled in  $\hat{\lambda}$  and  $\hat{\mu}$  making the M-step with the **ML-JB** complete data computationally unattractive. But, they can form the basis for the one-step late (OSL) iteration introduced by Green [7] or a generalized EM (GEM) iteration [8]. In the next section we develop a version of the GEM method that we call the joint alternate and maximize (JAM) method.

First however, we note in passing, that by augmenting the **ML-JB** complete data by the events which originate from object pixel  $b$ , propagate along coincidence line  $d$ , but due to absorption are *not* detected, we can uncouple the equations in the M-step of the update to derive a computationally feasible EM algorithm. This is equivalent to adding the observed transmission measurements to the complete data space **ML-IA** in [2]. The emission update for the resulting EM algorithm is linear, while the survival update requires solving a quadratic equation for each component of  $\hat{\mu}$  [4]. However, based on the analysis in [4] we expect that algorithms based on the **ML-JA** complete data to converge more slowly than those based on **ML-JB**.

### B. The Joint Alternate-Maximize (JAM) Algorithm

Although the coupling between  $\hat{\lambda}$  and  $\hat{\mu}$  in (15) and (16) makes the necessary maximization for the EM algorithm difficult, it is quite straightforward to derive a GEM algorithm, we simply alternate between updating  $\hat{\lambda}$  and  $\hat{\mu}$  using their most recent estimates at each step of the iteration.

Such iterations have been called alternate and maximize methods and often demonstrate improved asymptotic convergence properties over the EM algorithm [9]. The JAM iteration takes the form:

$$\begin{aligned} \hat{\lambda}^{(k+1)} &= \hat{\lambda}^{(k)} \odot A^T((\hat{\mu}^{(k)} \odot \mathbf{y}) \\ &\odot (\hat{\mu}^{(k)} \odot A\hat{\lambda}^{(k)} + \boldsymbol{\eta})) \odot A^T \hat{\mu}^{(k)}, \end{aligned} \quad (17)$$

$$\begin{aligned} \hat{\mu}^{(k+1)} &= [\hat{\mu}^{(k)} \odot ((\mathbf{y} \odot A\hat{\lambda}^{(k+1)}) \\ &\odot (\hat{\mu}^{(k)} \odot A\hat{\lambda}^{(k+1)} + \boldsymbol{\eta})) + \mathbf{m}] \\ &\odot (A\hat{\lambda}^{(k+1)} + \underline{\Lambda}). \end{aligned} \quad (18)$$

There are two features of particular interest in the above iteration: (1) the likelihood increases at each step and (2) it is not necessary to perform both emission and survival updates at each iteration. We typically update the survival probabilities once for every ten emission updates.

### C. Solution Stabilization

Because the tomographic reconstruction problem is ill-posed, it is usually necessary to either constrain the estimates to lie in a subspace consisting of smooth solutions or penalize estimates that are inconsistent with our prior notions of the smoothness of emitter or attenuator in the object.

Here we take the former approach and apply kernel sieves [10] to constrain both the survival probability estimate and the emission estimate. If the true survival probability and emission intensity are members of their respective sieves, then this method will result in a joint maximum-likelihood estimate that is asymptotically unbiased. If, on the other hand, they are not sieve members, the method will inevitably introduce bias—although hopefully, with a correspondingly larger reduction in estimator variance.

To implement the sieve-constrained reconstruction we assume the existence of both “pre-emission” and “pre-survival” spaces and signify individual elements of these spaces by  $\xi$  and  $\rho$  respectively. We define the transformations that take elements of these spaces to corresponding elements  $\lambda$  and  $\mu$  in the desired, constrained spaces by  $G$  and  $H$

$$\begin{aligned} \lambda &= G\xi, \\ \mu &= H\rho. \end{aligned} \quad (19)$$

By defining the complete pre-emission and pre-survival data as

$$\mathbf{X}_{JA}^P = \{\{N_{dq}^P\}, \{R_d\}, \{L_{dr}^P\}\}, \quad (20)$$

where

$$Y_d = \sum_{q=1}^Q N_{dq}^P + R_d, \quad (21)$$

$$M_d = \sum_{r=1}^R L_{dr}^P, \quad (22)$$

and taking  $G$  and  $H$  to be  $B \times Q$  and  $D \times R$  matrices, we can develop the following sieve-constrained JAM algorithm:

$$\begin{aligned} \hat{\underline{\xi}}^{(k+1)} &= \hat{\underline{\xi}}^{(k)} \odot [G^T A^T ((\hat{\underline{\mu}}^{(k)} \odot \mathbf{y}) \\ &\odot (\hat{\underline{\mu}}^{(k)} \odot A G \hat{\underline{\xi}}^{(k)} + \underline{\eta}))] \odot (A^T \hat{\underline{\mu}}^{(k)}), \end{aligned} \quad (23)$$

$$\hat{\underline{\lambda}}^{(k+1)} = G \hat{\underline{\xi}}^{(k)}, \quad (24)$$

$$\begin{aligned} \hat{\underline{\rho}}^{(k+1)} &= \hat{\underline{\rho}}^{(k)} \odot [H^T ((\mathbf{y} \odot A \hat{\underline{\lambda}}^{(k+1)}) \\ &\odot (\hat{\underline{\mu}}^{(k)} \odot A \hat{\underline{\lambda}}^{(k+1)} + \underline{\eta}) + \mathbf{m} \odot \hat{\underline{\mu}}^{(k)})] \\ &\odot H^T (\underline{\Delta} + A \hat{\underline{\lambda}}^{(k+1)}), \end{aligned} \quad (25)$$

$$\hat{\underline{\mu}}^{(k+1)} = H \hat{\underline{\rho}}^{(k+1)}. \quad (26)$$

#### IV. APPLICATION

We have implemented the JAM algorithm on a Stardent ST3000 computer system using C in conjunction with the Application Visualization System (AVS). We have performed preliminary evaluations of the joint estimation method using simulated PET data.

A PET imaging system having 100 projections of 64 parallel rays with an invariant Gaussian projection resolution of 9 mm FWHM was simulated. The object domain was an elliptical region circumscribed within a  $50 \times 64$  6 mm pixel rectangle. Projection images were generated from the emission phantoms shown in Fig. 1. The emission image shown in the center simulates a tumor imaging situation while the image on the right corresponds to cardiac study. The survival probabilities for all reconstructions were generated using the phantom on the left in Fig. 1. We used linear attenuation coefficients of  $0.156 \text{ cm}^{-1}$  for bone,  $0.095 \text{ cm}^{-1}$  for soft tissue, and  $0.022 \text{ cm}^{-1}$  for lung tissue.

Figure 2 shows reconstructions from an extreme example; there are 2 million total counts in the transmission scan and 5 million events in the emission data. The randoms fraction of total events in the emission scan is about 2%. The data were reconstructed using 1000 iterations of the EM algorithm, for the cases in which either the known survival probabilities or survivals estimated solely from transmission data were used, or using 1000 iterations of the JAM algorithm for joint survival-emission estimation. The emission estimates were subjected to a 1.5 pixel fwhm Gaussian sieve constraint and no survival sieve was employed. The images on the left in Fig. 2 result from using the noiseless survival probabilities. Images in the center were reconstructed using noisy transmission measurements and the JAM algorithm. The images on the right were reconstructed using the noisy transmission data

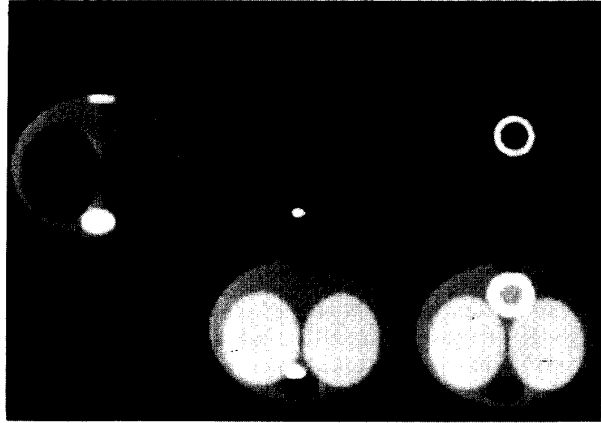


Figure 1: Top Row: Left. Map of the attenuation density in the simulated object. Center. Emission intensity for tumor simulation. Right. Emission intensity for cardiac imaging simulation. Bottom Row: Emission intensities displayed on log scale to enhance low-level detail

and the conventional EM method. Clearly, the joint estimation images are visually better than when the survival probabilities are not jointly estimated. The survival estimates for the three reconstructions are shown in Fig. 3. Note the improvement in noise here as well.

To investigate the quantitative performance of joint estimation, we used the cardiac phantom (Fig. 1, right) and the same attenuation distribution as above to generate 30 Poisson realizations each having 3 million emission events and 3 million transmission events. We reconstructed each realization using a 1.5 pixel Gaussian kernel sieve and 500 iterations of the appropriate algorithm. Reconstructions of one realization are shown in Fig. 4 (same order as Fig. 2). From the 30 reconstructions, we estimated the pixel-wise mean-squared error in the emission estimate. The square-root of this error distribution is shown in Fig. 5. The total error in each image is: (1) Survivals known exactly: 18700. (2) Noisy survivals, joint estimation: 27000. (3) Noisy survivals, emission estimation only: 41000.

#### V. DISCUSSION AND CONCLUSION

While the simulation results in the previous section are encouraging, there remain many areas requiring investigation. In particular the JAM algorithm may have considerably poorer convergence properties than algorithms that estimate the emission intensity only. This is no surprise, since there is inherently a larger number of degrees-of-freedom in the joint estimation problem. While conceivably the EM algorithm we have derived for joint estimation using complete data ML-JA could be faster, we are not

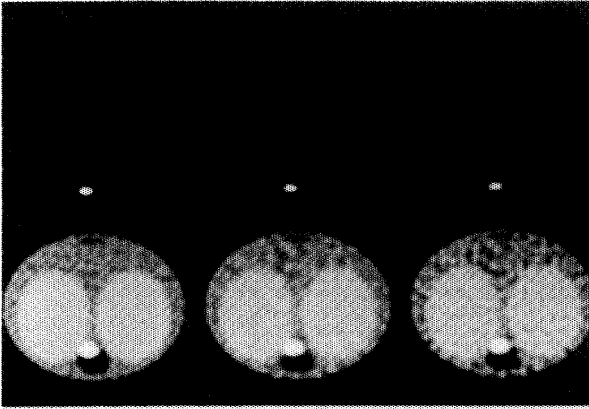


Figure 2: Top Row: Left. Reconstruction of tumor simulation with the survival probabilities known exactly. Center. Reconstruction jointly estimating survivals. Right. Reconstruction using survival probabilities directly estimated from transmission measurements. Bottom Row: Logarithmic display of top row.

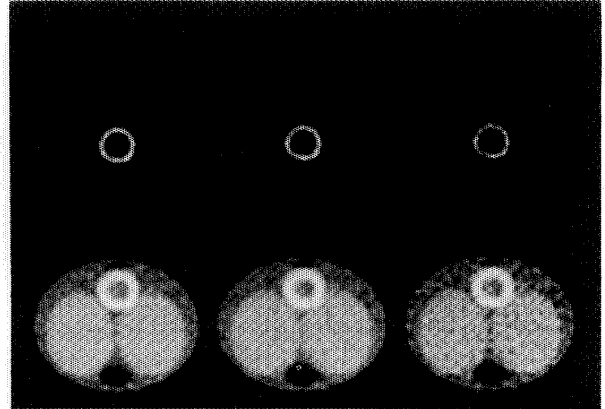


Figure 4: Reconstructions of simulated cardiac image. Order is the same as in Fig. 2.

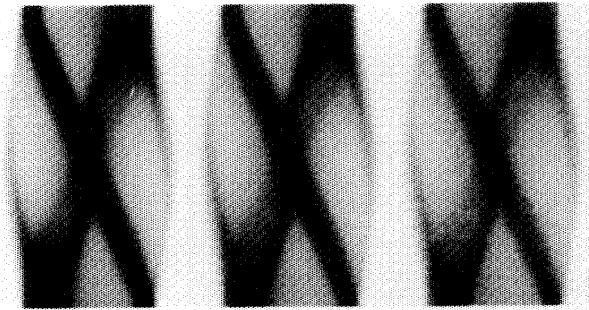


Figure 3: Left. True survival probabilities. Center. Jointly estimated survival probabilities for reconstructions in Fig. 2. Right. Survival estimated solely from transmission data.

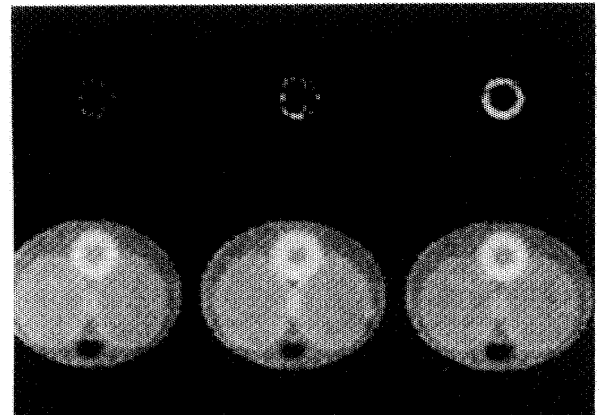


Figure 5: Estimates of the mean-squared error from the true object for the simulated cardiac image using 30 realizations. Top Row: Left. Known survival probabilities. Center. Joint estimation. Right. Survival probabilities estimated from transmission measurements only. Bottom Row: Logarithmic display.

encouraged by the fact that in practice the EM method using the **ML-IA** complete data converges much more slowly than its **ML-IB** counterpart and, moreover, that AM algorithms often exhibit better convergence properties than do EM methods. In any case, the use of appropriate sieve constraints should help the convergence properties of all algorithms.

The accuracy of the model relating the parameters to the observed data can also be brought into question. Both the emission scan and transmission scan are typically corrupted by Compton-scatter, which we do not model here, and additionally, the transmission scan is degraded by random coincidences. Although, this can be included in the joint likelihood model in a straightforward manner, the questions always arise: Which degradations are most significant? How much modeling is too much?

Ultimately, however, we feel that the joint estimation method and the constrained JAM algorithm we have outlined here will prove useful in PET reconstruction. The method may have the most to offer in thoracic imaging, where the distribution of attenuator in the object is extremely heterogeneous and the transmission scan often noisy due to long attenuation paths.

### References

- [1] S Huang, EJ Hoffman, ME Phelps, D. Kuhl, "Quantitation in positron emission tomography: 2. Effects of inaccurate attenuation correction." *JCAT*, 3(6):804-814, 1979.
- [2] DG Politte, DL Snyder, "Corrections for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron emission tomography." *IEEE Trans. Med. Imaging*, 10(1):82-89, 1991.
- [3] TJ Holmes, DL Snyder, DC Ficke, "The effect of accidental coincidences in time-of-flight positron emission tomography." *IEEE Trans. Med. Imaging*, MI-3:68-79, 1984.
- [4] JA Fessler, NH Clinthorne, WL Rogers, "On complete-data spaces for PET reconstruction algorithms." Submitted to *IEEE Trans. Med. Imaging*, September, 1991.
- [5] K Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing." *IEEE Trans. Med. Imaging*, 9(4):439-446, 1990.
- [6] LA Shepp, Y Vardi, "Maximum likelihood reconstruction for emission tomography." *IEEE Trans. Med. Imaging*, MI-1:112-121, 1982.
- [7] PJ Green, "On use of the EM algorithm for penalized likelihood estimation." *J. Royal Stat. Soc. B*, 52(3):443-452, 1990.
- [8] AP Dempster, NM Laird, DB Rubin, "Maximum likelihood from incomplete data via the EM algorithm (with discussion)." *J. Royal Stat. Soc. B*, 39:1-38, 1977.
- [9] JA Fessler, "On the convergence rates of EM and AM algorithms for maximum-likelihood estimation of superimposed signals." Submitted to *ICASSP 1992*, April 8-11, 1992, San Francisco, CA.
- [10] DL Snyder, MI Miller, LJ Thomas Jr., DG Politte, "Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography." *IEEE Trans. Med. Imaging*, MI-6:228-238, 1987.