Control and Cybernetics

vol. **25** (1996) No. 5

Positron emission tomography: A review

by

J.M.M. Anderson*, B.A. Mair and Murali Rao****
University of Florida, Gainesville, FL 32611, USA

*Department of Electrical and Computer Engineering,
** Department of Mathematics

Abstract: Positron emission tomography(PET) is a relatively new area of medical imaging, which has been in clinical use for about forty years. Due to its wide applicability in medical and psychological diagnostic procedures, researchers are interested in obtaining accurate quantitative information as to the metabolic activity rate of various parts of the human body from PET scans. Thus, there is much activity in the medical and engineering research communities on the subject. This paper attempts to present a brief review of the important mathematical fundamentals of this research area. It also outlines recent work of the authors on foundational issues. Due to severe space limitations, we cannot include the many new methods which have been formulated for reconstruction of PET images. However, we have tried to include enough material to give a reasonably good idea of the underlying methods on which most of the recent innovations is based.

Keywords: positron emission tomography, maximum likelihood, Poisson, weighted least squares, expectation maximization

1. Introduction

Positron emission tomography (PET) is a medical imaging technique that provides a means for assessing biochemical processes in the human body. By measuring the rates of quantities such as blood flow, oxygen and glucose metabolism, PET becomes a useful pathophysiological and diagnostic tool. It also enables medical researchers to expand their knowledge of brain processes like speech and vision. Modalities such as magnetic resonance imaging and X-ray computed tomography complement PET because they provide mostly anatomical information. Consequently, the advantages of these modalities are being combined to maximize the information available to physicians and researchers.

Regarding pathophysiology, Ter-Pogossian, Raichle and Sobel (1980), state

> "The foundation for the usefulness of PET are the premises that any biological activity stems from, and is accompanied by, regional biochemical changes in the organ or structure in which the biological activity takes place and that pathology is reflected in the alteration of such normal biochemical changes."

Some clinical applications for PET include its use in localizing tumor sites, measuring myocardial perfusion, studying various cancers, grading brain tumors, and determining the effectiveness of radiation therapy, chemotherapy, and surgery. In addition to clinical applications, PET is also being used in psychological studies to understand how the brain works. For example, PET scans are used by researchers to relate regions of the brain to various cognitive and visual skills (Barinaga, 1995, Kosslyn, 1994); and also to psychological disorders. The main steps in PET are:

1. Labeling a selected compound with a positron-emitting radionuclide.
2. Administering the labeled compound to the subject.
3. Estimating the distribution of the labeled compound.
4. Using the estimated distribution to determine parameters of physiological models that provide the information of interest.

The positron-emitting radionuclides that are used most often are ^{11}C, ^{13}N, ^{15}O, and ^{18}F. These radionuclides are used to label compounds such as sugars, amino acids, and neurotransmitter receptor ligands. Because of their short half-lives, positron-emitting radionuclides cannot be stored and must be generated on-site using an accelerator (e.g., cyclotron) which requires a significant amount of capital and technical expertise to maintain.

After it is administered to the subject, the labeled compound is absorbed disproportionately by various regions of the organ of interest. As the radionuclide decays, positrons are emitted in proportion to the distribution of the labeled compound. More positrons are emitted in regions of high absorption, and less positrons are emitted in regions of low absorption. For example, labeled glucose is used more by the regions of the brain that are most active when a certain task is performed. Another example is cancerous tissue's utilization of more glucose than healthy tissue. When a positron is emitted, it annihilates with a nearby electron creating two photons that fly off in nearly opposite, random directions.

Surrounding the organ of interest is one or several rings of detectors that are designed such that a count is incremented whenever a pair of detectors senses two photons in coincidence. Two photons are said to be in coincidence whenever they are registered by a pair of detectors within a small time interval. A pair of detectors defines a volume, which is referred to as a tube. For each tube t, the associated count $n_o^*(t)$, of number of photons detected in coincidence, form the data used in the reconstruction algorithms. These algorithms seek to estimate the unknown annihilation locations or, equivalently, the distribution of the labeled compound.

In reality, positrons migrate a short distance before annihilating with an electron and photons depart at an angle less than 180°. These phenomena impose

a fundamental limit on the accuracy of PET. The estimates of the labeled compound's distribution are also affected by errors due to attenuation (Huang, 1979, Meikle, Dahlbom and Cherry, 1993), accidental coincidences (Hoffman, 1981), scatter (Bergstrom, 1983), and detector inefficiency (Fessler, 1994). Attenuation occurs when photons are absorbed by body tissue and bone causing them to go undetected. The percentage of attenuated photons ranges from $20 - 50\%$ (Politte and Snyder, 1991). Scatter occurs when a photon's line of flight is altered. Accidental coincidences occur when photons arising from separate annihilations are incorrectly registered as a valid coincidence. Thus accidentals usually result in a significant error between the number of observed and actual coincidences. The percentage of accidental coincidences ($5 - 50$ %) depends on the subject and the PET scanner.

Single photon emission computed tomography (SPECT) is a very similar procedure to PET. The only difference is that the decay of the particular radionuclides result in only one photon instead of two. Since the SPECT radionuclides can be stored, an on-site accelerator is unnecessary, thus reducing the cost significantly. However, the radionuclides used in PET have the advantage that they can be incorporated into molecules of biological interest (or their analogs). In SPECT, the same circular configuration of detectors is utilized except that a rotating collimator is used to determine the line-of-flight for the photons. Due to the collimator, many fewer photons are detected in SPECT than in PET so the resulting images are unable to resolve fine details. Also, attenuation correction is more difficult in SPECT. In summarizing the advantage of PET over SPECT, Henry Wagner says that PET provides "better chemistry, better quantitation, and better sensitivity compared to SPECT, and correction for attenuation in the body is more exact" (Tilyou, 1991). Beekman (1995) contains more details and important advances.

X-ray computed tomography (CT) is a much more developed and widespread procedure than PET. In CT, the radiation, the emitters and detectors lie on two parallel planes which rotate around the region of interest. The radioactive source is located outside the human body, and is absorbed as it travels in a straight line through the body, from emitter to the corresponding detector. To obtain sufficient data, these planes are rotated, and data collected at each angle of rotation. In CT, the line of action is completely determined by the emitter and the detector, unlike the random emission directions in PET.

The various aspects of PET generally fall in one of the following categories:

1. Medical research
2. Psychological research
3. Clinical diagnosis
4. Development of labeled compounds
5. Development of detectors and scanners
6. Reconstruction algorithms
7. Accelerators design
8. Physiological modeling

The resolution of open problems in PET require the expertise of chemists, engineers, mathematicians, physicists, physicians, and psychologists. For additional information, some excellent references on PET are Brownell (1982), Reivich and Alavi (1995), Ter-Pogossian, Raichle and Sobel (1980).

In this review, we focus on the mathematical models and reconstruction methods for reconstructing PET images. Section 2 contains a brief derivation of the standard Poisson model introduced by Shepp and Vardi (1982). Existing reconstruction algorithms are discussed in Section 3. We do not attempt to give algorithmic details, but rather focus on the underlying mathematical concepts. Section 4 contains a brief overview of new methods that we have introduced recently, and includes some numerical simulations to compare their performance with existing algorithms.

2. Mathematical model

In this section, we describe the probability model developed by Shepp and Vardi (1982), which relates the means of the emissions to those of the detections.

The seminal work of Shepp and Vardi (1982) obtained the now accepted Poisson model for the PET process. Assume that the data is error free, so that the observed number $n_o^*(t)$ of counts in tube t is the same as the true number $n^*(t)$. To obtain a finite dimensional model suitable for numerical implementation, we consider a fine partition of the region of interest, denoted by Ω (in two or three dimensional space), into B "boxes" or pixels. The emissions can be thought of as the result of randomly falling snowflakes, with areas of high absorption, collecting proportionally more. Thus the emission process is realized as the outcome of a non-homogeneous, spatial Poisson point process with unknown intensity. This intensity is approximately constant on each of the B boxes, so for each $b = 1, 2, \ldots, B$, the random variable $N(b)$ denoting the number of emissions in box b, is Poisson with mean $\lambda(b)$, and since the boxes are disjoint, then the family of $N(b)$'s are independent.

Further, let $N_o^*(t)$ denote the random variable underlying the observation $n_o^*(t)$, $t = 1, 2, \ldots, T$, where $N_o^*(t)$ is Poisson with mean $\lambda_o^*(t) = \mathbb{E}[N_o^*(t)]$ and the $N_o^*(t)$'s are independent. We also introduce the random variables $N(t|b)$ to denote the number of points that are detected in t out of all those that are emitted in b. Thus,

$$N_o^*(t) = \sum_{b=1}^{B} N(t|b). \tag{1}$$

Let the probability that an emission in box b is detected in tube t be denoted by $p(t|b)$. Assuming that all emissions are detected, we have that

$$\sum_{t=1}^{T} p(t|b) = 1, \; b = 1, 2, \ldots, B. \tag{2}$$

These probabilities provide the connection between the emission random variables $N(1), \ldots, N(B)$, and data random variables $N_o^*(1), \ldots, N_o^*(T)$.

To obtain this relationship, observe that by the independence of $N(t|b)$, and (1), it follows that each $N(t|b)$ is Poisson and its mean $\lambda(t|b)$ satisfies

$$\lambda_o^*(t) = \sum_{b=1}^{B} \lambda(t|b). \tag{3}$$

Now, observe that

$$N(b) = \sum_{t=1}^{T} N(t|b) \tag{4}$$

and the probability that $N(t|b) = k$ given that $N(b) = n$ is given by

$$\mathcal{P}\left(N(t|b) = k | N(b) = n\right) = \binom{n}{k} p(t|b)^k (1 - p(t|b)^{n-k}. \tag{5}$$

Interestingly, this binomial property implies that there is some $\omega(b)$ such that

$$\lambda(t|b) = \omega(b)\lambda(b)p(t|b). \tag{6}$$

This is a consequence of the following result which is demonstrated in Mair, Rao and Anderson (1995).

Theorem 1 *Let $X_1, X_2, \ldots, X_n$ be independent non-negative, integer-valued random variables. If there are λ_i's such that*

$$\mathcal{P}(X_i = k | \sum_{j=1}^{n} X_j = m) = \binom{m}{k} \lambda_i^k (\Lambda - \lambda_i)^{m-k}$$

for each i and $0 \leq k \leq m$, where $\Lambda = \sum_{i=1}^{n} \lambda_i$, then there exists $\omega > 0$ such that each X_i is Poisson with mean $\omega\lambda_i$.

The model equation, Shepp and Vardi (1982), follows from (6), (4), (2), and (3).

$$\lambda_o^*(t) = \sum_{b=1}^{B} \lambda(b)p(t|b), \; t = 1, 2, \ldots, T. \tag{7}$$

3. Reconstruction algorithms

Given the data $\boldsymbol{n}_o^* = [n_o^*(1)\ n_o^*(2) \ldots n_o^*(T)]^{\mathrm{T}}$, reconstruction algorithms estimate $\boldsymbol{\lambda} = [\lambda(1)\ \lambda(2)\ \cdots\ \lambda(B)]^{\mathrm{T}}$, where B is the number of boxes and T is the number of tubes. In this way, an estimate for the number of positrons emitted in each box is obtained. Although considerable effort has gone into developing reconstruction algorithms, many questions remain as to which one

should be preferred. Generally, reconstruction algorithms are either based on Fourier or statistical methods. Fourier based algorithms are developed from a deterministic model, while statistically based ones rely on the Poisson model.

Fourier based algorithms are extremely fast, but suffer from artifacts and perform poorly when the number of observed coincidences is low. Low count scenarios are common and occur when radiation doses are low and/or scan times are short. Radiation doses are minimized for patient safety and scan times are kept short for patient comfort. Also, they do not account for the random nature of the PET data and fail to incorporate errors appropriately. In contrast, statistically based algorithms perform well in low count cases and have the ability to account for errors in the data. Unfortunately, they have the disadvantage of requiring computationally expensive iterative methods for estimating the emission intensities. This high computational expense is the main reason that statistically based methods have not yet become popular in "real world" PET.

3.1. Fourier based algorithms

By assuming that the probabilities $p(t|b)$ are independent of the emission pixel b, and allowing the discretizations and detector tubes to become infinitely small, the PET model equation (7) reduces to that of CT, expressed in terms of the Radon transform (Herman, 1980). So, methods of inverting the Radon transform can be applied. Although the mathematical inverse of the Radon transform has been obtained since 1917, it is not immediately applicable due to the lack of sufficient error-free data. Fourier based methods are numerical algorithms developed either from the mathematical formula for inverting the Radon transform or from the projection-slice theorem. The latter type, for example Mersereau (1976), are not as popular as the former ones because they are more difficult to implement and the images are inferior (Lewitt, 1983). An in-depth discussion of these methods can be found in Lewitt (1983).

The convolution backprojection (CBP) algorithm (Shepp and Logan, 1974, Herman, 1980, Natterer, 1986) is based on the inverse Radon transform and was originally developed for X-ray computed tomography. Later modified for PET, it is presently the dominant reconstruction algorithm because its computational efficiency is unmatched and it produces images of reasonable quality when the number of observed coincidences is sufficiently large. Data errors due to attenuation are easily accounted for in the Poisson model (see (7)), and ML-EM iteration (see (8)), by modifying the probabilities $p(t|b)$ to account for experimentally determined survival probabilities (Miller, Snyder and Miller, 1985). However, it is not clear how to account for such errors in the CBP algorithm. The other significant probabilistic errors of scatter and accidental coincidences also present difficulties for the CBP method. We now focus on statistically based algorithms for the remainder of the paper.

3.2. ML-EM algorithm

To address the limitations of the CBP method, Shepp and Vardi (1982) proposed the Poisson model for PET (see (7)) and the use of maximum likelihood estimators to determine the emission intensity vector $\boldsymbol{\lambda}$. The log-likelihood function is given by $L(\boldsymbol{\lambda}) = \log \mathrm{P}[\boldsymbol{N}_o^* = \boldsymbol{n}_o^*]$, where $\boldsymbol{N}_o^* = [N_o^*(1), N_o^*(2), \ldots, N_o^*(T)]^T$. By using the expectation-maximization algorithm (Dempster, Laird and Rubin, 1977), the following iterative procedure was obtained in Shepp and Vardi (1982).

$$\lambda_{n+1}(b) = \lambda_n(b) \sum_{t=1}^{T} \frac{n^*(t)p(t|b)}{\sum_{k=1}^{B} p(t|k)\lambda_n(k)} \tag{8}$$

This algorithm is referred to as the ML-EM algorithm. In Shepp and Vardi (1982), Vardi, Shepp and Kaufman (1985), it was shown that this algorithm converges to a maximum likelihood estimator of $\boldsymbol{\lambda}$ for any initial choice of $\boldsymbol{\lambda}_0$ in which all the components are strictly positive. Furthermore, the likelihood increases with the iterations, and the iterations are all positive. Thus, the ML-EM algorithm has excellent theoretical properties, but converges slowly (especially at locations with zero intensity) and tends to produce images that are speckled. Despite the global convergence property, the limit depends on the initialization. Simulations indicate that a uniformly distributed initial estimate produces the best results. This initialization is almost universally accepted now. Methods to reduce speckle and improve convergence can be found in Coakley (1991), Snyder and Miller (1985), Kaufman (1987), Chen, Lee and Cho (1991). Other ML approaches can be found in Lange and Carson (1984), Rockmore and Macovski (1976).

3.3. MAP methods

Recognizing the drawbacks of the ML-EM algorithm, researchers began considering maximum a posteriori (MAP) methods (De Pierro, 1995, Geman and McClure, 1985, Green, 1990, Hebert and Leahy, 1989, Levitan and Herman, 1987). Instead of just maximizing the log-likelihood function, MAP methods maximize $L(\boldsymbol{\lambda}) + f(\boldsymbol{\lambda})$, where $f(\boldsymbol{\lambda})$ is a known, a priori distribution for the emission intensity vector. These methods are based on the observation that neighboring boxes generally have similar emission intensities. The prior is usually assumed to be a Gibbs distribution which requires the choice of a functional form for the potential, and also the explicit determination of numerical constants. One of these constants (basically a regularization parameter), has to be determined from the data. The methods for doing this inevitably lead to a significant computational burden on any numerical algorithm. As a result of the smoothing effects inherent in MAP methods, they tend to obscure important boundaries. Zhou, Leahy and Mumcuoglu (1993) show an interesting use of magnetic resonance images to provide PET priors.

3.4. Least squares methods

In an effort to improve the computational efficiency of statistically based reconstruction algorithms, least-squares methods were proposed. Least-squares methods use the fact that $n_o^*(t)$ is the ML estimate of $\lambda_o^*(t)$, and minimize the 2-norm of the difference between $\boldsymbol{n}_o^*$ and $\boldsymbol{\lambda}_o^*$, where $\boldsymbol{\lambda}_o^* = [\lambda_o^*(1)\ \lambda_o^*(2) \ldots \lambda_o^*(T)]^T$. In other words, the LS method minimizes $(\boldsymbol{n}_o^* - \boldsymbol{\lambda}_o^*)^{\mathrm{T}}\,(\boldsymbol{n}_o^* - \boldsymbol{\lambda}_o^*)$ with respect to $\boldsymbol{\lambda}$. This approach was suggested by Shepp and Vardi (1982) (see also Vardi, Shepp and Kaufman, 1985), but Daube-Witherspoon and Muehllehner (1986) and Kaufman (1993) actually developed reconstructions algorithms based on it. The image space reconstruction algorithm (ISRA) (Daube-Witherspoon and Muehllehner, 1986), was based in large part on heuristic arguments, but was shown later to converge to an LS estimate of the emission intensity (De Pierro, 1987, Titterington, 1987). The ISRA is computationally efficient and guarantees nonnegative estimates; however, simulation studies indicate that it converges slowlier and produces less accurate estimates than the ML-EM algorithm. Kaufman's methods (Kaufman, 1993) are also computationally efficient, but negative estimates present a problem.

As usual, LS methods can be enhanced by introducing a weight matrix $\boldsymbol{W}$ to emphasize reliable data values while de-emphasizing less reliable data values. Weighted least squares (WLS) approaches have been proposed for both PET (Clinthorne, 1992, Fessler, 1994) and single-photon emission computed tomography (SPECT) (Budinger, Gullberg, 1977, Tsui, Zhao, Frey and Gullberg, 1991). Fessler (1994) contains a WLS approach to account for errors due to accidental coincidences and detector inefficiency in PET. Using a successive overrelaxation method (Marty, 1988) and weights obtained from the data, a reconstruction algorithm was developed that produced nonnegative estimates and converged. The WLS formulation in Clinthorne (1992) accounts for the effect of accidental coincidences, but assumes that their mean is known for each tube.

To minimize the variance of the WLS estimator, we must have $\boldsymbol{W}_{ij} = \mathrm{Cov}(N_o^*(i), N_o^*(j)) = \delta_{i,j}\lambda_o^*(i)$. Due to its dependence on the unknown parameters, the covariance matrix had not been explicitly used in PET. In Tsui, Zhao, Frey and Gullberg (1991), $\boldsymbol{W}$ was estimated by $\mathrm{Diagonal}(n_o^*(1), n_o^*(2), \ldots, n_o^*(T))$. We will refer to this method as the data-based WLS (DWLS) method.

Recently, two novel methods were proposed that allow the entries of $\mathbf{W}$ to be unknown and incorporate them into the estimation procedure (Anderson, Mair, Rao and Wu, 1995a;1995b). The first method minimizes $(\boldsymbol{n}_o^* - \boldsymbol{\lambda}_o^*)^{\mathrm{T}}\,\mathbf{W}^{-1}\,(\boldsymbol{n}_o^* - \boldsymbol{\lambda}_o^*)$, while the second one minimizes this same objective function plus a penalty function. The penalty function is chosen such that the resulting estimate almost preserves the total number of coincidences (a known property of the model that is lost in the LS formulation). This method is referred to as the almost total-coincidence preserving WLS or ATP-WLS method. This is much less restrictive than what was considered in Fessler (1994) and Kaufman (1993). There, the penalties were based on a Markov random field approach (Geman

and McClure, 1985), which correlates the intensities of neighboring pixels. In simulation studies, the penalized and unpenalized algorithms converged nearly twice as fast and had significantly better resolution and contrast than the ML-EM and DWLS methods. They also guarantee nonnegative estimates.

3.5. Other methods

Algebraic reconstruction techniques (ART) (Gordon, Bender and Herman, 1970, Herman, 1980) are a special class of iterative procedures for solving general optimization problems, and so have been implemented for obtaining estimates of various measures of goodness of fit in PET. In Snyder, Schulz and O'Sullivan (1992), it is shown that the ML-EM algorithm can be derived by minimizing the Kullback-Leibler (K-L) distance (Csiszár and Tusnády, 1984) (for a definition, see (18)), from the data to the detector means. Byrne (1993) extended this idea by considering the cross entropy from the data means to the data, and also penalized versions of these measures of divergence. This framework also covers many of the MAP methods. One of the iterative procedures obtained in Byrne (1993) is shown to be identical to an ART, however in our simulations studies it did not perform as well as the ML-EM algorithm. In a manner similar to the CBP algorithm, maximum entropy (MAXENT) methods developed for transmission computed tomography (Gull and Newton, 1986, Minerbo, 1979) have also been modified for PET (see e.g. Kemp, 1980).

4. New reconstruction methods

This section contains an overview of recent methods for PET image reconstruction based on the Poisson model. We feel that these methods introduce new ideas which can serve as building blocks for future enhancements to deal with the important errors in the data. We first introduce a new method of correcting for accidental coincidences. The basic idea here is to let the data determine the accidentals in a natural way. Unlike previous methods, this method requires no significant computational effort over the basic ML-EM algorithm, and, in our simulations, it outperforms the existing methods. We find the LS approach very attractive since it has the capability of dealing with model errors. Unlike all previous methods, our new WLS approach does not require any data-based estimation of the covariance matrix. This represents a fundamental change in the area of WLS estimation theory, and may be applicable to other areas. Our algorithm is iterative, and bears a striking resemblance to the ML-EM algorithm (unlike the ISRA Daube-Witherspoon and Muehllehner, 1986, and the EM-LS obtained in Kaufman, 1993). It also ensures positivity of the estimates and is able to resolve small portions of the image having high intensities.

We also include a brief discussion of a refined mathematical model for PET, which produces an integral equation with an unknown measure, as opposed to a finite linear system. Although this model is more theoretical, it provides an

interesting link between PET and the general theory of inverse problems. This was first noted in Vardi and Lee (1993), where the basic ML-EM algorithm was extended to the continuous case (see (17)). Although this algorithm was demonstrated to be effective in dealing with general inverse problems, there is as yet, no convergence proof. Our contribution here is to suggest a natural topology in which a convergence proof may be possible.

4.1. Accidental correction via EM

Nearly all reconstruction algorithms correct for accidental coincidences by subtracting an estimate of them from the observed coincidences. The accidental coincidences are estimated either by using delayed coincidence circuitry or by calculating them via an estimate of single events (Hoffman, 1981). Unfortunately, this approach quite often results in negative counts which clearly indicates that the accidental coincidences cannot be estimated accurately. Although it is reasonable to model the accidental coincidences as Poisson random variables, the corrected data are not Poisson. Consequently, ML (Lange and Carson, 1984, Shepp and Vardi, 1982) and MAP (Geman and McClure, 1985, Hebert and Leahy, 1989, Levitan and Herman, 1987, Liang and Jaszczak, 1989) methods that are based on the Poisson model may be theoretically inappropriate when the data are corrected in this manner. Lange and Carson (1984) suggested that the accidental rates could be estimated and incorporated in the estimation procedure, however, no algorithm was proposed. Later, Politte and Snyder (1991) incorporated the accidental rates in an ML estimation procedure but unrealistically assumed that they were known. Leahy's algorithm (Leahy and Yan, 1991) estimates both the emission intensity and accidental rate for each tube, however it is very computationally intensive and less effective than the proposed method (Anderson, Mair and Rao, 1995).

We introduce a novel approach which estimates both the emission intensity and the total number of accidental coincidences. Like Politte and Snyder (1991), we assume that the accidental coincidences $n_a^*(t)$ in the t^{th} tube is Poisson with mean $\lambda_a^* = A/T$, that is independent of the detector tubes. Then we jointly determine the ML estimate of the emission intensity $\boldsymbol{\lambda}$ and the total number A, of accidental coincidences.

To develop our algorithm, it is necessary to modify the Poisson model to account for the effects of accidental coincidences. Let $n^*(t)$ and $n_a^*(t)$ denote the number of true and accidental coincidences detected by the t^{th} tube, respectively. Then, the number of photons detected by the t^{th} tube is given by $n_o^*(t) = n^*(t) + n_a^*(t)$, where the true and accidental coincidences are assumed to be independent and Poisson with means $\lambda^*(t) = \sum_{b=1}^{B} p(t|b)\lambda(b)$, and $\lambda_a^*(t) = A/T$, $t = 1, 2, \ldots, T$, respectively. Under these assumptions, we are able to use the EM algorithm (Dempster, Laird and Rubin, 1977) to obtain an iterative algorithm, as in the ideal case in Shepp and Vardi (1982).

To do this, we consider $\boldsymbol{n}_o^* = (n_o^*(1), \ldots, n_o^*(T))$ as the *incomplete* data and

$(n(1),\ldots,n(B),n_a^*,n_o^*(1),\ldots,n_o^*(T))$, where n_a^* is the total number of accidentals, as the *complete* data. The unknowns that need to be estimated are given by $\boldsymbol{\theta}=(\lambda(1),\ldots,\lambda(B),A)$, where A is the mean of n_a^*.

Now, let $\boldsymbol{\theta}_n=(\lambda_n(1),\ldots,\lambda_n(B),A_n)$ be the current estimate of $\boldsymbol{\theta}$ and perform the EM algorithm.

E - step: Use $\boldsymbol{\theta}_n$ and the incomplete data to estimate the *missing* data, by $\hat{n}(b)=\mathbb{E}[n(b)|\boldsymbol{n}_o^*,\theta_n]$ and $\hat{n}_a^*=\mathbb{E}[n_a^*|\boldsymbol{n}_o^*,\theta_n]$. Since $\sum_{b=1}^{B}n(t|b)+n_a^*(t)=n_o^*(t)$, and the random variables are independent and Poisson, we obtain

$$\begin{aligned}\mathbb{E}[n(t|b)|\boldsymbol{n}_o^*,\theta_n] &= \frac{n_o^*(t)p(t|b)\lambda_n(b)}{\sum_{k=1}^{B}p(t|k)\lambda_n(k)+A_n/T}\\ \hat{n}_a^*=\mathbb{E}[n_a^*(t)|\boldsymbol{n}_o^*,\theta_n] &= \frac{n_o^*(t)A_n/T}{\sum_{k=1}^{B}p(t|k)\lambda_n(k)+A_n/T}\end{aligned}$$

M - step: This postulates that an improved estimate $\boldsymbol{\theta}_{n+1}$ of $\boldsymbol{\theta}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$ given the above estimate of the complete data. Since the maximum likelihood estimator of the mean of a Poisson random variable is the random sample, we obtain the following iteration.

$$\lambda_{n+1}(b) = \lambda_n(b)\sum_{t=1}^{T}\frac{n_o^*(t)p(t|b)}{\lambda_n^*(t)+A_n/T},\quad b=1,2,\ldots,B \tag{9}$$

$$A_{n+1} = A_n\sum_{t=1}^{T}\frac{n_o^*(t)/T}{\lambda_n^*(t)+A_n/T} \tag{10}$$

where $\lambda_n^*(t)=\sum_{b=1}^{B}p(t|b)\lambda_n(b)$.

This method guarantees nonnegative estimates, and is stable with increasing iterations. Since it is a true EM method, convergence is guaranteed (Wu, 1983).

To evaluate our algorithm, we simulate a PET scenario with a single ring consisting of 128 equally spaced detectors and a phantom of 128×128 pixels. The probabilities $p(t|b)$s are computed using the angle-of-view method, as described in Shepp and Vardi (1982). The total number of observed coincidences, $n_o^*=2.2\times10^6$, of which 10% are accidentals. The number of true and accidental coincidences, $n^*(t),n_a^*(t)$ were obtained by generating Poisson random variables with means $\lambda^*(t)=\sum_{b=1}^{B}p(t|b)\lambda(b)$, $\lambda_a^*(t)=0.1n_o^*/T$ respectively, where the $\lambda(b)$'s are determined by the phantom.

In the reconstructions, we assumed that the accidentals were only 5% of the total number of observed coincidences. We obtained the initial estimate for the accidentals, by generating random samples $\hat{n}_a(t)$, $t=1,2,\ldots,T$ from Poisson distributions, each having mean $0.05n_o^*/T$. The proposed algorithm was initialized with $A_0=\sum_{t=1}^{T}\hat{n}_a(t)$ and $\lambda_0(b)=(n_o^*-A_0)/T$.

We compare it with the EM-ML with corrected data (conventional), and

Leahy and Yan's (LY) algorithm

$$q_{n+1}(t) = \frac{1}{2}\left[\hat{n}_a(t) + \frac{n_o^*(t)q_n(t)}{\sum_{l=1}^{B} p(t|l)\lambda_n(l) + q_n(t)}\right], \quad t = 1, 2, \ldots, T$$

$$\lambda_{n+1}(b) = \lambda_n(b)\sum_{t=1}^{T} \frac{n_o^*(t)p(t|b)}{\sum_{l=1}^{B} p(t|l)\lambda_n(l) + q_n(t)}, \quad b = 1, 2, \ldots, B$$

where $q_n^*(t)$ is an estimate of the accidental rate (i.e. A/T) for the t^{th} tube.

At the 200^{th} iteration, the results for the proposed, conventional, and LY algorithms indicate more ringing at the "skull" boundary with the conventional and LY methods as compared to the proposed one. These visual differences can be seen from line plots which allow for a more quantitative assessment. Fig. 1 is a line plot of the 90^{th} row of the reconstructed images using the proposed (solid), conventional (dashed), and LY (dotted) reconstruction algorithms. These line plots demonstrate that the proposed method is more accurate than the conventional and LY methods. It is important to observe that the conventional and LY methods exhibit more severe overshoot at the skull boundary, as compared to the proposed method. This overshoot increased with the number of iterations in both the conventional and LY methods.

4.2. A weighted least squares method

Here we propose a novel weighted least-squares (WLS) method, and develop an iterative reconstruction algorithm based on minimizing the objective function. Unlike all other WLS methods, this method does not require any estimation of the covariance matrix from the data. The proposed method guarantees non-negative estimates, and in simulation studies, it converged faster and had much better contrast and resolution than the ML-EM and DWLS algorithms. For this exposition, we assume that the data contains no accidentals. However, the proposed method is easily modified to account for such errors. More importantly, the WLS formulation takes into account errors due to modelling. Using the Poisson model, least-squares estimation is based on the observation that $n^*(t)$ is the ML estimate of $\lambda^*(t)$, where $\lambda^*(t) = \sum_{b=1}^{B} \lambda(b)p(t|b)$, and the $\lambda(b)$'s are estimated by minimizing $\sum_{t=1}^{T}(n^*(t) - \lambda^*(t))^2$. Alternatively, we account for the error in using $n^*(t)$ instead of $\lambda^*(t)$ and model the observed data as $n^*(t) = \lambda^*(t) + e^*(t)$, $t = 1, 2, \ldots, T$, where the $e^*(t)$'s are the model errors. Because the $n^*(t)$'s are assumed to be independent and Poisson with mean $\lambda^*(t)$, the model errors are independent, zero-mean, and have variance $\lambda^*(t)$. In this case, the error covariance matrix is given by $\mathbf{W} = \text{diagonal}(\lambda^*(1), \ldots, \lambda^*(T))$. It is well-known that variance of a least-square estimator is minimized when it is weighted by the inverse of the error covariance matrix; hence, we propose the

following WLS estimator:

$$\hat{\boldsymbol{\lambda}} = arg \min_{\boldsymbol{\lambda} \geq 0} \sum_{t=1}^{T} \frac{(n^*(t) - \lambda^*(t))^2}{\lambda^*(t)}$$

Exploiting the convexity of the above WLS objective function, we obtain that $\hat{\boldsymbol{\lambda}}$ is a minimizer if and only if the following Kuhn-Tucker conditions (Zangwill, 1969), are satisfied.

$$\begin{aligned} \sum_{t=1}^{T} \frac{n^*(t)^2}{(\sum_{k=1}^{B} p(t|k)\hat{\lambda}(k))^2} p(t|b) &= 1 \text{ if } \hat{\lambda}(b) \neq 0 \\ \sum_{t=1}^{T} \frac{n^*(t)^2}{(\sum_{k=1}^{B} p(t|k)\hat{\lambda}(k))^2} p(t|b) &\leq 1 \text{ if } \hat{\lambda}(b) = 0 \end{aligned}$$

The first condition suggests the following fixed point iteration:

$$\lambda_{n+1}(b) = \lambda_n(b) \sum_{t=1}^{T} \frac{(n^*(t))^2 p(t|b)}{\lambda_n^*(t))^2} \tag{11}$$

where $\lambda_n^*(t) = \sum_{b=1}^{B} p(t|b)\lambda_n(b)$.

Simulations suggest that the algorithm is globally convergent, though an analytical proof has not yet been found. Nevertheless, we have proven analytically that the WLS objective function decreases monotonically with increasing iterations. Our simulation results indicate that the proposed algorithm converges nearly twice as fast as the ML-EM algorithm. Since the algorithm is very similar (differs only by the square of a factor) to the ML-EM iteration, the convergence acceleration techniques in Chen, Lee and Cho (1991) and Kaufman (1987) can also be implemented here. On the other hand, this algorithm yields reconstructions which are vastly improved over the ML-EM and the DWLS method.

For our simulation, we use the same basic set-up as in the previous section, except for the following. There were no accidentals, and the total number of coincidences was set at 10^7. The phantom and the reconstructed images after forty iterations using the DWLS, ML-EM and proposed WLS algorithms, are shown in Fig. 2. From the figures, it can be seen that the proposed method is sharper, has greater contrast and does a better job at picking up the details. For example, the thin rectangle on the lower left region of the phantom is much more visible using the WLS algorithm.

Our new WLS method can be further improved by adding various penalty terms. For example, as mentioned above, the iteration does not preserve total counts, so that we could add a penalty term of the form $\alpha(\sum_{b=1}^{B} \lambda(b) - \sum_{t=1}^{T} n^*(t))$. In our simulations (Anderson, Mair, Rao and Wu, 1995a), we found that this penalized method produces images with less bias and more accurate estimates of the emission intensities than the proposed WLS method.

4.3. A refined mathematical model

Despite the reported speckling and numerical divergence of the ML-EM algorithm, with a uniformly distributed initialization, we are able to obtain excellent PET image reconstructions, with no speckling, and exhibiting fast, complete convergence of the algorithm if we generate the detector counts as a random sample drawn from a Poisson distribution with mean $\boldsymbol{\lambda}^*$ determined *exactly* by $\lambda^*(t) = \sum_{b=1}^{B} p(t|b)\lambda(b)$. However, the reconstructed PET images exhibit divergence with severe speckling if the data is simulated according to the following more realistic method. This consists of generating a Poisson point process on Ω with intensity $\boldsymbol{\lambda}$ and then determining the lines of flight from each of these points according to a uniform random distribution. Recall that the derivation of the Shepp–Vardi model was based on the latter principle. We are therefore led to the conclusion that the operator mapping the emission intensity $\boldsymbol{\lambda}$ to the detector means $\boldsymbol{\lambda}^*$ may not be appropriately represented by the matrix of probabilities. This discrepancy may be due to the binning of the probability distribution, and the intensities into finitely many pixels. This is consistent with the fact that, for ill–posed problems, small perturbations in the operator, especially when going from infinite to finite-dimensional representations, dramatically affect attempts to invert the operator. We therefore propose a *semi-infinite model*, which incorporates the following four essential properties of the PET process.

- There are only a finite number of detectors.
- The annihilation of the electrons by the positrons occur at sufficiently microscopic levels to be justifiably represented by points.
- The probabilities governing the emission–detection process are not necessarily constant on "boxes".
- The emission intensity is not necessarily representable by a function (it may be a measure).

The first three of these properties were incorporated in the following first kind Fredholm integral equation model, which was proposed (but not analyzed) in 1982 by Shepp and Vardi (1982):

$$\lambda^*(t) = \int_{\Omega} p(t|x)\lambda(x)dx \tag{12}$$

where $p(t|x)$ is now the probability that an emission at point x in Ω is detected in the t^{th} detector tube, and λ is the assumed intensity density of the Poisson point process governing the emissions.

To account for the fourth property, we assume that emissions occur according to a Poisson point process on Ω with intensity λ, where λ is a finite Borel measure on Ω. Then, for each Borel subset E of Ω, the number of emissions in E is a Poisson random variable with mean $\lambda(E)$. As usual, we assume that $p(t|x)$ is continuous in x for each t and that all emissions are detected.

Fix a detector tube t, and consider the function $p(t|\cdot)$ as a limit of finite sums of step functions determined by a finite partition $B_1, B_2, \ldots, B_m$ of Ω. Then, for each such approximation, the finite-dimensional model (see (7)) holds, and by taking the limiting case as these partitions become finer, we obtain the *semi–infinite* model

$$\lambda^*(t) = \int_\Omega p(t|x)d\lambda(x), \ t = 1, 2, \cdots, T. \tag{13}$$

By the usual arguments, the maximum–likelihood method of estimating λ reduces to the problem of maximizing

$$L(\lambda) = -\lambda(\Omega) + \sum_{t=1}^{T} n^*(t) \log \int_\Omega p(t|x)d\lambda(x) \tag{14}$$

over the set $\mathcal{C}$ of finite, Borel measures on Ω. Since we have only a finite number of detectors, this is equivalent to minimizing the Kullback–Leibler divergence of the data $\boldsymbol{n}^*$ from $\boldsymbol{\lambda}^*$ (Byrne, 1993, Multhei and Schorr, 1989, Vardi, Shepp and Kaufman, 1985). Clearly, the objective function L is concave and the constraint set $\mathcal{C}$ is a convex cone in the infinite-dimensional Banach space consisting of the signed Borel measures on Ω with the total variation norm (Deuschel and Strook, 1989).

From the Kuhn–Tucker first order optimality conditions we obtain that $\hat{\lambda}$ is a maximum–likelihood estimator of the true intensity λ, if and only if the following two conditions are satisfied.

$$\sum_{t=1}^{T} \frac{n^*(t)}{\hat{\lambda}^*(t)} p(t|x) = 1 \ \text{ for } \hat{\lambda} \text{ almost every } x \in \Omega \tag{15}$$

$$\sum_{t=1}^{T} \frac{n^*(t)}{\hat{\lambda}^*(t)} p(t|x) \leq 1 \ \text{ for all } x \in \Omega \tag{16}$$

From (15), a maximum–likelihood estimator $\hat{\lambda}$ satisfies $\hat{\lambda}(\Omega) = \sum_{t=1}^{T} n^*(t)$, hence, from (15), $\hat{\lambda}$ is a fixed point of the map $\lambda \mapsto \sum_{t=1}^{T} \frac{n^*(t)}{\lambda^*(t)} p(t|\cdot)\lambda$ on the convex cone $\mathcal{C}_0 = \{\lambda \in \mathcal{C} : \lambda(\Omega) = \sum_{t=1}^{T} n^*(t)\}$. Hence, it is natural to compute $\hat{\lambda}$ by the fixed point iteration

$$\lambda_{n+1}(x) = \lambda_n(x) \sum_{t=1}^{T} \frac{n^*(t)}{\lambda_n^*(t)} p(t|x) \tag{17}$$

where $\lambda_n^*(t) = \int_\Omega p(t|x)\lambda_n(x)dx$.

A non-probabilistic heuristic argument was used by Kondor (1983) to obtain a similar iteration for solving a general class of first kind integral equations. The same algorithm was rediscovered by Vardi and Lee (1993) by a probabilistic

argument and demonstrated to be very effective in dealing with a wide range of inverse problems. However, the analysis of this method so far has not yielded a general proof of convergence for the iterates in (17) (Kondor, 1983, Multhei and Schorr, 1989, Vardi and Lee, 1993). If the integral equation (13) has a solution which is a finite sum of characteristic functions, then Vardi and Lee (1993) prove that the iteration in (17) converges to a solution of (13). Without this finiteness assumption, Multhei and Schorr prove that, if the iteration converges in mean (i.e. in the L^1 topology) to a continuous function, then the limit function is a maximizer of L. These results were attempting to obtain convergence of the iterates in (17) to a *function*. Examples in Mair, Rao and Anderson (1995) show that such a result cannot be obtained in general, so that the convergence behavior depends on the properties of the kernel $p(t|x)$.

In these examples, $\{\lambda_n\}$ converges in the weak topology (Deuschel and Strook, 1989) to the point mass $\hat{\lambda}$, as opposed to convergence in the first or second mean to a *function*. Thus, we consider the (usually continuous) functions $\{\lambda_n\}$ as measures (absolutely continuous to Lebesgue measure) on Ω, and convergence in the weak topology.

First, for any two measures $\lambda, \nu \in \mathcal{C}_0$, define the K-L distance from λ to ν by

$$\rho(\lambda, \nu) = \int h(x) \log h(x) d\nu(x), \text{where } h = \frac{d\lambda}{d\nu} \tag{18}$$

As before (Multhei and Schorr, 1989, Vardi, Shepp and Kaufman, 1985, Snyder, Schulz and O'Sullivan, 1992), the concavity of the logarithm function and Jensen's inequality can be used to obtain

$$0 \leq \rho(\lambda_{n+1}, \lambda_n) \leq L(\lambda_n) - L(\lambda_{n+1}) \tag{19}$$

Thus, L increases along the sequence of iterates, and hence $\{L(\lambda_n)\}$ converges since it is bounded. The basic problem that prevents the finite–dimensional proof from being extended to the infinite–dimensional case is the fact that the K-L distance from a probability measure μ to another one, ν, is only finite when μ is absolutely continuous with respect to ν (Deuschel and Strook, 1989). However, we conjecture that the following general convergence result holds.

Theorem 2 *The sequence $\{\lambda_n\}$ defined by (17) converges in the weak topology to a Borel measure $\hat{\lambda}$ which is a maximum–likelihood estimator of the true intensity λ.*

Reconstruction algorithms using wavelet bases are presently being developed.

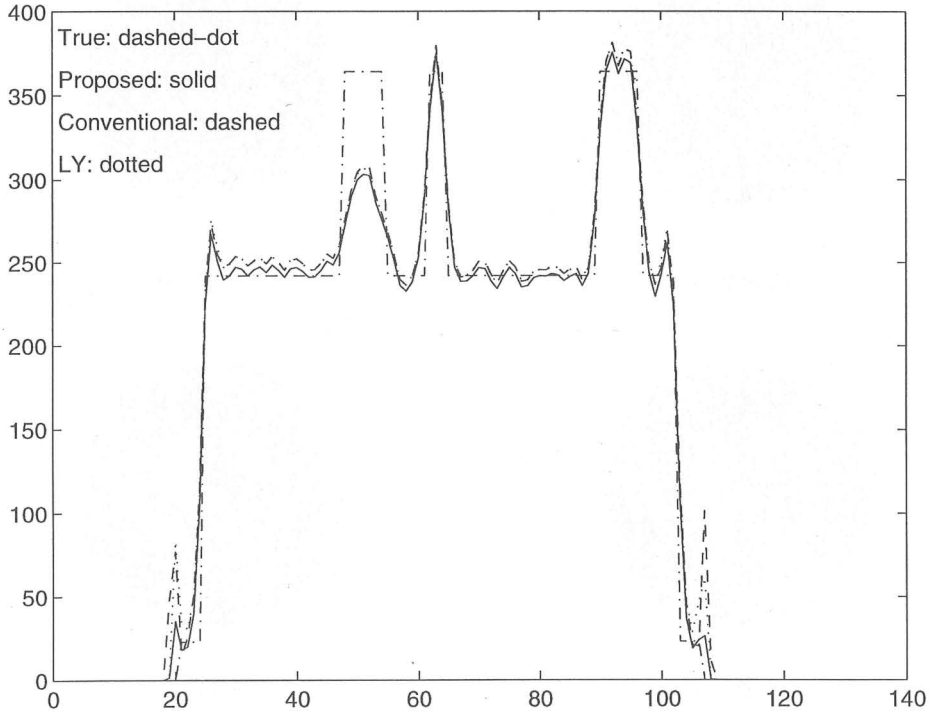
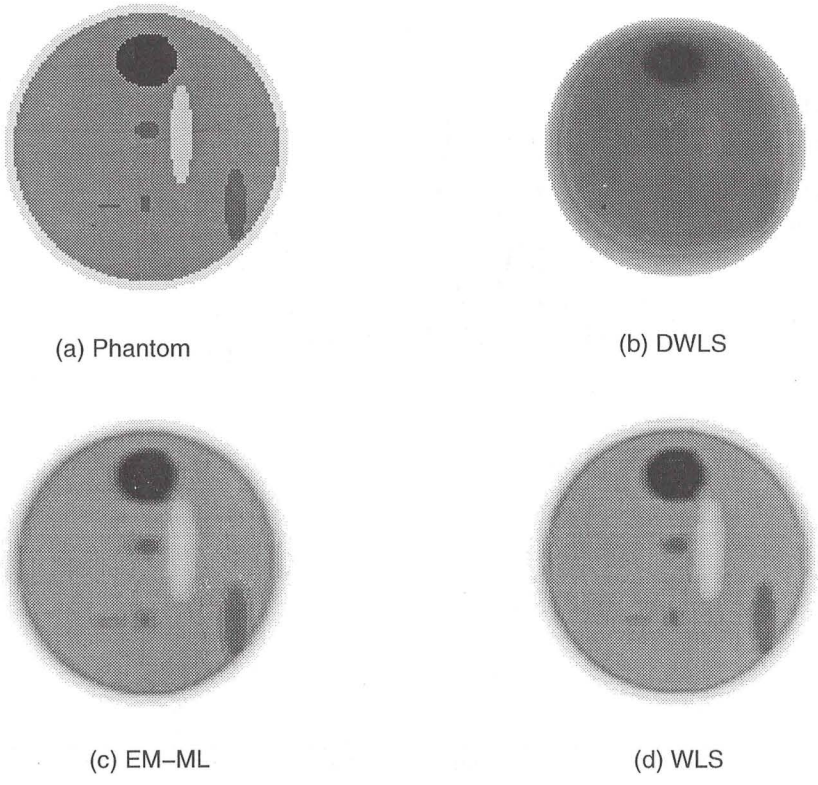


Figure 1. AC-10%, Estimated AC-5%



(a) Phantom (b) DWLS

(c) EM–ML (d) WLS

Figure 2. 40^{th} iteration, $N^* = 10^7$

References

ANDERSON, J.M.M., MAIR, B.A. and RAO, M. (1995) Image reconstruction for PET with iterative correction for accidental coincidences. (Submitted).

ANDERSON, J.M.M., MAIR, B.A., RAO, M. and WU, C.-H. (1995A) Weighted least-squares reconstruction algorithms for positron emission tomography. *Proceedings of the Medical Imaging Conference*, San Francisco.

ANDERSON, J.M.M., MAIR, B.A., RAO, M. and WU, C.-H. (1995B) Weighted least-squares reconstruction algorithms for positron emission tomography. *IEEE Transactions on Medical Imaging.* (Submitted).

BEEKMAN, F. (1995) *Fully 3D SPECT Reconstruction with Object Shape Dependent Scatter Compensation.* Ph.D. Dissertation, Dept. of Medical Physics and Biophysics, University of Nijmegen, The Netherlands.

BARINAGA, M. (1995) Brain researchers speak a common language. *Science*, **270**, 1437-1438.

BERGSTROM, M., et. al. (1983) Correction for scattered radiation in a ring detector positron camera by integral transformation of the projections. *Journal of Computer Assisted Tomography*, **7**, 42-50.

BROWNELL, G.L., et. al. (1982) Positron tomography and nuclear magnetic resonance imaging. *Science*, **215**, 619-626.

BUDINGER, T.F., GULLBERG, G.T. (1977) Transverse section reconstruction of gamma-ray emitting radionuclides in patients. In: M. M. Ter-Pogossian, M. E. Phelps, G. L. Brownell, J. R. Cox, D. O. Davis, R. G. Even, eds., *Reconstruction Tomography in Diagnostic Radiology and Nuclear Medicine*, University Park Press, Baltimore, MD, 315-342.

BYRNE, C.L. (1993) Iterative Image Reconstruction Algorithms Based on Cross-Entropy Minimization. *IEEE Trans. on Image Processing*, **2**, 96-103.

CHEN, C.M., LEE, S.Y. and CHO, Z.H. (1991) Parallelization of the EM algorithm for 3-D PET image reconstructions. *IEEE Trans. on Medical Imaging*, **10**, 513-522.

CLINTHORNE, N.H. (1992) Constrained least-squares vs. maximum likelihood reconstructions for Poisson data. *Proc. of the IEEE Medical Imaging Conference, Orlando, Florida*, **2**, 1237-1239.

COAKLEY, K.J. (1991) A cross-validation procedure for stopping the EM algorithm and deconvolution of neutron depth profiling spectra. *IEEE Trans. on Nuclear Science*, **38**, 9-15.

CSISZÁR, I. and TUSNÁDY, G. (1984) Information geometry and alternating minimization procedures. *Statist. Decisions Suppl.*, **1**, 205-237.

DAUBE-WITHERSPOON, M.E. and MUEHLLEHNER, G. (1986) An iterative image space reconstruction algorithm suitable for volume ECT. *IEEE Trans. on Medical Imaging*, **5**, 61-66.

DE PIERRO, A.R. (1995) A modified expectation maximization for penalized maximum likelihood estimation in emission tomography. *IEEE Trans. on*

Medical Imaging, **14**, 132-137.

De Pierro, A.R. (1987) On the convergence of the iterative image space reconstruction algorithm for volume ECT. *IEEE Trans. on Medical Imaging*, **6**, 174-175.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*, **39**, 1-38.

Deuschel, I.-D., and Strook, D.W. (1989) *Large Deviations.* Academic Press, New York.

Fessler, J.A. (1994) Penalized weighted least-squares image reconstruction for positron emission tomography. *IEEE Trans. on Medical Imaging*, **13**, 290-300.

Geman, S. and McClure, D.E. (1985) Bayesian image analysis: An application to single photon emission tomography. *Proc. Statist. Comput. Sect., Amer. Statist. Assn.*, Washington, D.C., 12-18.

Gordon, R., Bender, R. and Herman, G.T. (1970) Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of Theor. Biol.*, **29**, 471-481.

Green, P. (1990) Bayesian reconstructions from emission tomography using a modified EM algorithm. *IEEE Trans. on Medical Imaging*, **9**, 84-93.

Gull, S.F. and Newton, T.J. (1986) Maximum entropy tomography. *Applied Optics*, **25**, 156-160.

Hebert, T. and Leahy, R. (1989) A generalized EM algorithm for the 3-D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Trans. on Medical Imaging*, **8**, 194-202.

Herman, G.T. (1980) *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography.* Academic Press, San Francisco.

Hoffman, E.J., et. al. (1981) Quantitation in positron-emission computed tomography: 4. Effect of accidental coincidences. *J. of Computer Assisted Tomography*, **5**, 391-400.

Huang, S., et. al. (1979) Quantitation in positron-emission computed tomography: 7. Effects of inaccurate attenuation correction. *J. of Computer Assisted Tomography*, **3**, 804-814.

Kaufman, L. (1987) Implementing and accelerating the EM algorithm for positron emission tomography. *IEEE Trans. on Medical Imaging*, **6**, 37-51.

Kaufman, L. (1993) Maximum likelihood, least squares, and penalized least squares for PET. *IEEE Trans. of Medical Imaging*, **12**, 200-214.

Kemp, M.C. (1980) Maximum entropy reconstruction in emission tomography. *Medical Radionuclide Imaging*, 313-315.

Kondor, A. (1983) Method of convergent weights–An iterative procedure for solving Fredholm's integral equations of the first kind. *Nuclear Instr. Methods*, **216**, 177-181.

KOSSLYN, S.M., et. al. (1994) Identifying objects seen from different viewpoints: A PET investigation. *Brain*, **117**, 1055-1071.

LANGE, K. and CARSON, R. (1984) EM reconstruction algorithms for emission and transmission tomography. *Journal of Computer Assisted Tomography*, **8**, 302-316.

LEAHY, R. and YAN, X. (1991) Incorporation of anatomical MR data for improved functional imaging with PET. *Information Processing in Medical Imaging, International Conf.*, Wye, UK, 105-120.

LEVITAN, E. and HERMAN, G.T. (1987) A maximum a posteriori probability expectation maximization algorithm for reconstruction in emission tomography. *IEEE Trans. on Medical Imaging*, **16**, 185-192.

LEWITT, R.M. (1983) Reconstruction Algorithms: Transform Methods. *Proceedings of the IEEE*, **71**, 390-408.

LIANG, Z. and JASZCZAK, R. (1989) On Bayesian image reconstruction from projections. *IEEE Trans. on Medical Imaging*, **8**, 227-235.

LLACER, J. and VEKLEROV, E. (1989) Feasible images and practical stopping rules for iterative algorithms in emission tomography. *IEEE Trans. on Medical Imaging*, **8**, 186-193.

MAIR, B.A., RAO, M. and ANDERSON, J.M.M. (1995A) Semi-infinite positron emission tomography. *Experimental and Numerical Methods for Solving Ill-Posed Inverse Problems: Medical and Non-Medical Applications*, SPIE Proceedings, **2570**, 292-300.

MAIR, B.A., RAO, M. and ANDERSON, J.M.M. (1995B) A refined mathematical model for positron positron emission tomography. *Conference Record of 1995 IEEE Medical Imaging Conference.*

MAIR, B.A., RAO, M. and ANDERSON, J.M.M. (1995C) Simulation of Poisson point processes. (Submitted).

MARTY, K.G. (1988) *Linear Complementarity, Linear and Nonlinear Programming.* Chapter 9, Helderman Verlag, Berlin.

MEIKLE, S.R., DAHLBOM, M. and CHERRY, S.R. (1993) Attenuation correction using count-limited transmission data in positron emission tomography. *Journal of Nuclear Medicine*, **34**, 143-150.

MERSEREAU, R.M. (1976) Direct Fourier transform techniques in 3-D image reconstruction. *Comput. Biol. Med.*, **6**, 247-258.

MILLER, M.I., SNYDER, D.L. and MILLER, T.R. (1985) Maximum-likelihood reconstruction for single-photon emission computed-tomography. *IEEE Trans. on Nuclear Science*, **32**, 769-778.

MINERBO, G. (1979) MENT: A maximum entropy algorithm for reconstructing a source from projection data. *Computer Graphics and Image Processing*, **10**, 48-68.

MULTHEI, H.N. and SCHORR, B. (1989) On properties of the iterative maximum likelihood reconstruction method. *Math. Meth. in the Appl. Sci.*, **11**, 331-342.

NATTERER, F. (1986) *The Mathematics of Computerized Tomography*, Wiley, New York.

POLITTE, D.G. and SNYDER, D.L. (1991) Corrections for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron emission tomography. *IEEE Trans. on Medical Imaging*, **10**, 82-89.

REIVICH, M. and ALAVI, A. (1995) *Positron Emission Tomography.* eds., Alan R. Liss, NY.

ROCKMORE A.J. and MACOVSKI, A. (1976) A maximum likelihood approach to emission image reconstruction from projections. *IEEE Trans. on Medical Imaging*, **23**, 1428-1432.

SHEPP L.A. and LOGAN, B.F. (1974) The Fourier reconstruction of a head section. *IEEE Trans. on Nuclear Science*, **21**, 21-43.

SHEPP, L.A. and VARDI, Y. (1982) Maximum likelihood reconstruction for emission tomography. *IEEE Trans. on Medical Imaging*, **1**.

SNYDER, D.L. and MILLER, M. (1985) The use of sieves to stabilize images produced with EM algorithm for emission tomography. *IEEE Trans. on Medical Imaging*, **16**.

SNYDER, D.L., SCHULZ, T.J. and O'SULLIVAN, J.A. (1992) Deblurring subject to nonnegativity constraints. *IEEE Trans. on Signal Processing*, **40**, 1143-1150.

SNYDER, D.L., MILLER, M.I., THOMAS, JR., L.J. and POLITTE, D.G. (1987) Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography. *IEEE Trans. on Medical Imaging*, **6**, 228-238.

TER-POGOSSIAN, M.M., RAICHLE, M.E. and SOBEL, B.E. (1980) Positron emission tomography. *Scientific American*, **243**, 170-181.

TILYOU, S.M. (1991) The evolution of positron emission tomography. *Journal of Nuclear Medicine*, Special Issue on Clinical PET, **32**, 15N-26N.

TITTERINGTON, D.M. (1987) On the iterative image space reconstruction algorithm for ECT. *IEEE Trans. on Medical Imaging*, **6**, 52-56.

TSUI, B.M.W., ZHAO, Z., FREY, E.C. and GULLBERG, G.T. (1991) Comparison between ML-EM and WLS-CG algorithms for spect image reconstruction. *IEEE Trans. on Nuclear Science*, **38**, 6, December.

VARDI, Y. and LEE, D. (1993) From image deblurring to optimal investments: maximum likelihood solutions for positive linear inverse problems. *J. R. Statist. Soc. B*, **55**, 569-612.

VARDI, Y., SHEPP, L.A. and KAUFMAN, L. (1993) A statistical model for positron emission tomography. *Journal of the American Statistical Association*, **80**, 8-37.

WU, C.F.J. (1983) On the convergence properties of the EM algorithm. *The Annals of Statist.*, **11**, 95-103.

ZANGWILL, W.I. (1969) *Nonlinear Programming.* Prentice-Hall.

ZHOU, Z., LEAHY, R.M. and MUMCUOGLU, A.U. (1993) A comparative study of the effect of using anatomical priors in PET reconstruction. *IEEE*

Medical Imaging Conference, San Francisco, November.