

USC-SIPI REPORT #241

Fast Gradient-Based Methods for Bayesian Reconstruction of Transmission and Emission

by

**Erkan U. Mumcuoglu, Richard Leahy, Simon R. Cherry
and Zhenyu Zhou**

September 1993

**Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 404
Los Angeles, CA 90089-2564 U.S.A.**

Fast Gradient-Based Methods for Bayesian Reconstruction of Transmission and Emission PET Images^{*†}

Erkan U. Mumcuoglu[†] Richard Leahy[†] Simon R. Cherry[†]
Zhenyu Zhou[†]

[†]Signal and Image Processing Institute,
Department of Electrical Engineering-Systems,
University of Southern California, Los Angeles, CA 90089-2564

[‡]Imaging Sciences Division, Crump Institute for Biological Imaging,
B2-086 CHS, UCLA School of Medicine, 10833 LeConte Ave,
Los Angeles, CA 90024-6948

September 3, 1993

Abstract

We describe conjugate gradient algorithms for reconstruction of transmission and emission PET images. The reconstructions are based on a Bayesian formulation where the data are modeled as a collection of independent Poisson random variables and the image is modeled using a Markov random field. A conjugate gradient algorithm is used to compute a maximum a posteriori (MAP) estimate of the image by maximizing over the posterior density. To ensure non-negativity of the solution a penalty function is used to convert the problem to one of unconstrained optimization. Preconditioners are used to enhance convergence rates. These methods generally achieve effective convergence in 15-25 iterations. Reconstructions are presented of an ¹⁸FDG whole body scan from data collected using a Siemens/CTI ECAT931 whole body system. These results indicate significant improvements in emission image quality using the Bayesian approach, in comparison to filtered backprojection, particularly when reprojections of the MAP transmission image are used in place of the standard attenuation correction factors.

^{*}This work was supported by a grant from the Whitaker Foundation

[†]Please send all correspondence to Richard Leahy at the above address

1 Introduction

One of the major factors limiting the clinical utility of iterative positron emission tomography (PET) image reconstruction schemes based on statistical formulations is the computational cost involved in estimating the image. Although efficient implementation on a fast workstation or specialized processor can result in low per-iteration costs, reconstruction times can still be prohibitive if a large number of iterations are required. Here we describe penalized preconditioned conjugate gradient methods for both emission and transmission reconstruction in PET where effective convergence can often be attained in 15-25 iterations. The method forms a *maximum a posteriori* (MAP) estimate of the image based on a Poisson data model, which explicitly includes contributions from scatter and random coincidences, and a Markov random field prior which favors locally smooth images.

Our motivation for using Bayesian procedures for the reconstruction of PET images is that this framework allows for the development of an accurate statistical model for the observed data and also allows the inclusion of additional information concerning the unknown image in the form of a prior distribution. In the following we describe Bayesian procedures not only for reconstruction of the emission image, but also for reconstruction of a transmission or attenuation image and for nonlinear smoothing of measured randoms and blank sinograms. The smoothed randoms sinograms are used in place of direct randoms subtraction in the coincidence data as described in Sections (2.1) and (2.2). The smoothed blank sinogram is used as an estimate of the effective transmission source activity as described in Section (2.1).

The methods described here are applicable to the general PET image reconstruction problem. However, we emphasize the application to whole body PET. Whole body protocols call for multiple data sets (frames) corresponding to multiple positions of the patient within the scanner. Consequently, the acquisition time per frame is necessarily short. This results in poor signal to noise ratios in the data and correspondingly poor images using standard correction and reconstruction methods [5]. Here we demonstrate the potential for substantial improvements in image quality using the Bayesian estimation procedures described.

The primary problem in PET is to reconstruct spatial distributions of radio-labeled compounds from measurements of coincidence emission data, i.e. the emission sinogram. The main purpose for collecting the transmission sinogram is for computing attenuation correction factors for the emission data as the ratio of the blank to transmission sinograms. However, there are several reasons why reconstruction of a transmission image from this data is appropriate. First, correction factors based on the ratio of the blank and transmission sinograms provide noisy and biased estimates of the true attenuation factors. Linear smoothing of the transmission sinogram reduces noise but does not provide sufficient accuracy [27]. Reconstruction of a transmission image, followed by reprojection, can improve the accuracy of the estimated correction factors provided additional information, in the form of a statistical data model and a prior density on the image, is introduced. It is important to note that simply using fil-

tered backprojection followed by reprojection of the reconstructed transmission image is essentially equivalent to direct linear filtering of the attenuation correction factors and should not be expected to yield further improvements. There are other advantages to reconstruction of the transmission image. For example, if the patient moves between the transmission and emission scans, they can be re-registered before reprojection for computation of the attenuation correction factors. The attenuation images can also be used in estimating scatter components in emission sinograms [30]. Finally, the transmission images provide anatomical landmarks that are often not visible in the emission images. These can be used when interpreting the emission images or for registration of the PET images with other anatomical modalities [23].

Since its development in [36], the expectation-maximization (EM) algorithm for maximum likelihood (ML) reconstruction of emission PET images has been the most widely used iterative scheme for statistical reconstruction. There are however two major problems with the application of this scheme to PET data. First, convergence of the EM algorithm can be slow [20], [19] and second, the ML estimation problem in PET is typically ill-conditioned. Here we avoid this ill-conditioning by using a Bayesian formulation and consider alternatives to EM which exhibit faster convergence.

Lange *et al* [20] and others have noted that the EM algorithm for emission PET can be recast as a gradient ascent algorithm in which the update vector is the product of the gradient vector with a diagonal matrix with elements equal to a scaled version of the current estimates of the pixel intensities (henceforth referred to as a preconditioner). Over-relaxation schemes based on this observation have been proposed [25]. An alternative approach which uses a vector extrapolation technique to enhance the convergence of the EM algorithm was proposed in [34]. This technique leads to impressive gains in the convergence rate but no method for imposing a non-negativity constraint is described. In [19] it was noted that the effect of the preconditioner associated with the EM algorithm was to modify the update vector from the direction of steepest ascent to one which reduces the gradient towards zero for pixels as their intensities approach zero. In this manner the image is always constrained to be non-negative. It was also argued that the effect of the preconditioner is to improve the convergence rate by effectively improving the conditioning of the preconditioned Hessian relative to the original Hessian of the log-likelihood function. Kaufman [19] also proposed two important modifications to the basic EM approach: (i) to use the EM-direction vector but to perform a line search for the optimal step length and (ii) to utilize a conjugate gradient method with the EM preconditioner. Both techniques lead to improved convergence rates for ML estimation but neither method is implicitly constrained to the non-negative orthant, as is the case for the standard EM algorithm. To overcome this problem, possible alternatives include using a suboptimal step size to avoid negative values at each iteration or to truncate any negative values to zero once the pixels have been updated using the optimal step size [19]. In either case, the effect of the positivity constraint is to largely negate the improvements which may have otherwise resulted from using a conjugate gradient method. In this paper we deal with the non-negativity constraint using a penalty function. In this way we can replace the constrained problem with an unconstrained one and apply conjugate gradient techniques directly.

In the case of transmission PET, attempts at developing an EM algorithm have not resulted in attractive closed form update procedures [21]. An approximate EM algorithm was proposed in [3] but the computational cost remains high. Lange [20] proposed a preconditioned gradient descent method by drawing an analogy to the EM algorithm for the emission case. In this algorithm the update vector is again formed by multiplying the gradient vector of the log likelihood function by a diagonal preconditioner which is a function of the current estimate of the transmission image. When combined with a line search and a positivity constraint, the algorithm should exhibit monotonic increases in the log-likelihood. Similarly, the conjugate gradient scheme proposed in [19] could be adapted to the transmission problem with an appropriate preconditioner. However the benefits of the conjugate gradient approach will be limited by the presence of the positivity constraint. Below, we again apply the penalty function approach to overcome this problem.

In both emission and transmission CT the standard ML estimation problem is often ill-conditioned. This property is manifested in the high variance that is often observed in ML reconstructions at high iteration numbers. Here we avoid this problem by maximizing over the log posterior density rather than the log likelihood function. Alternative ways of dealing with this problem include the use of sieves [28] or early termination of the ML algorithm using a stopping rule [40]. To specify the posterior density, we need to introduce a prior density which models the spatial configuration of the image. A wide range of priors have been investigated for both transmission and emission tomography as discussed in Section (2.3). In this work we use Markov random field (MRF) priors. Using the Poisson data model and the MRF prior we form a MAP estimate by maximizing over the posterior density. This can be achieved by extending the EM ML algorithm to the MAP problem [24], [12], [14]. With the exception of a restricted class of priors, there is no closed form solution for the M-step for MAP estimation and the EM procedure is often relaxed to a generalized EM (GEM) algorithm in which the maximization step in the EM algorithm is replaced with an updating procedure which guarantees monotonic increase in the posterior probability at each iteration. An alternative EM-type algorithm is described in [32] [15] in which a closed form modified M-step is used. This algorithm has been shown to converge to a MAP estimate when the log prior is concave. In the case where a line process is also included in the MRF model optimization during the M-step becomes even more difficult. Deterministic annealing schemes which are embedded in the GEM algorithm are described in [22] and [11]. Here we avoid the use of the EM formalism entirely and work directly with gradient based techniques similar to those discussed above. We examine two MRF priors. The first of these, the Geman and McClure prior [10], is differentiable and fits directly into the gradient based approach. The second is the so-called weak membrane model [11] which includes a line process to represent discontinuities in image intensity. In this case we use a continuation method similar to that described in [11] except that again we use a gradient based procedure in place of the GEM algorithm.

2 The Statistical Model

Both transmission and emission PET data can be well modeled as a set of independent Poisson random variables. Here we briefly describe the models for the data and the manner in which scattered and random coincidences are included in the model. The form of the prior is also discussed.

2.1 Transmission Data

Transmission PET data are collected using an external positron source. The two common configurations available are a fixed ring source [16] or a rotating rod source [38]. The major advantages of the rod source are that the scatter component in the transmission data is greatly reduced and also that there is the potential for simultaneous acquisition of transmission and emission data. The model below is appropriate for rod and ring sources.

In order to reconstruct a transmission PET image, or to compute attenuation correction factors, it is necessary to collect both blank and transmission sinograms. The blank sinogram is collected in the absence of a patient or other attenuating medium and reflects the activity of the source. The transmission sinogram is then collected with the patient in position. Both the blank and transmission sinograms are observations of a random process and consequently should both be modeled as such. The likelihood function for the transmission sinogram is a function of the expected value of the blank sinogram, rather than the blank sinogram itself. In the Appendix we describe a non-linear Bayesian smoothing technique for estimating the mean of the blank sinogram using the observed blank sinogram.

Let y_i denote the number of coincidences between the i^{th} detector pair in the transmission sinogram and \hat{z}_i the expected value of the corresponding element in the blank sinogram. The transmission data, y_i , can be modeled as independent Poisson observations which consist of true, scattered and random coincidences, with respective means $\hat{t}_i(\mu)$, \hat{s}_i and \hat{r}_i :

$$\hat{y}_i = E(y_i) = \hat{t}_i(\mu) + \hat{s}_i + \hat{r}_i \quad (1)$$

We assume that the image can be adequately represented as a set of pixels, each with constant linear attenuation coefficient μ_j . The total attenuation between the i^{th} detector pair is then determined by the sum of the attenuation coefficients μ_j multiplied by their respective volume of intersection l_{ij} with the i^{th} projection strip. Let τ_y and τ_z denote the data acquisition times for the transmission and blank scans respectively. Similarly let δ_y and δ_z denote the system dead time (as a fraction of the total data collection time) for the transmission and blank scans respectively. The mean of the true coincidences in the transmission sinogram is then:

$$\hat{t}_i(\mu) = u_i e^{(-\sum_j l_{ij} \mu_j)}, \quad u_i = \frac{\tau_y \delta_y}{\tau_z \delta_z} \hat{z}_i. \quad (2)$$

The normalization factor $\frac{\tau_y \delta_y}{\tau_z \delta_z}$ compensates for the difference in observation times

for blank and transmission scans, and also for variations in system dead time, which will generally be higher for the blank scan due to the higher singles rate. This model assumes that for a given data set, deadtime is constant across all detector pairs in a given plane and also that the deadtime of the system does not violate the assumption, implicit in the Poisson model, that the probabilities of detection of each photon pair at each detector pair are statistically independent. Currently, we simply use the deadtime correction factors supplied in the header-file of the data which are calculated using an empirical formula [37].

The scatter and randoms components are also functions of the unknown attenuation coefficients. Typically the data are precorrected for randoms by subtracting a randoms sinogram from the transmission coincidence sinogram. The randoms sinogram is collected using a delayed coincidence circuit with the same resolving time as that of the coincidence sinogram. The delay that is introduced between the detection of the first photon and the onset of the coincidence resolving window is sufficiently long that a 'coincidence' event detected during this time must be a random coincidence. The subtraction of these two sinograms is performed automatically during data acquisition. The effect of this correction is to compensate the data, in its mean, for the effect of randoms. This follows since we are subtracting a random variable whose mean is equal to the mean of the randoms component in the transmission sinogram. Unfortunately, since corresponding elements of the transmission sinogram and randoms sinogram are independent Poisson random variables, the randoms corrected sinogram is the difference of two Poissons whose variance is equal to the sum of their means. In other words, randoms subtraction compensates in mean for randoms, but it also increases the variance of the measurement by an amount equal to the mean of the randoms. The measurement is no longer Poisson. To avoid this problem we must maintain the transmission and randoms measurement as two separate sinograms.

There are many ways in which randoms may be treated within the model in equation(1). We could jointly estimate the attenuation image and the randoms sinogram as proposed in [22]. Alternatively, we could assume a constant randoms contribution as described in [33]. The method we have used here is to estimate the mean of the randoms sinogram using the nonlinear Bayesian smoothing technique which is described in the Appendix. This technique exploits the property that the randoms sinogram at the surface of the detectors (i.e. before it is subject to spatial variations in detector pair sensitivities) is usually very smooth. We use a Gaussian MRF model for this smooth sinogram and compute a MAP estimate using a Poisson model for the observed randoms. The resulting estimated mean randoms sinogram is then treated as a constant in (1) when we optimize over the attenuation image.

The scatter component in (1) is highly dependent on the transmission image since Compton scatter is the dominant attenuating mechanism at 511keV in human tissue. In principal, the scatter component could be estimated at each iteration from the current estimate of the transmission image [6]. Such a procedure would be exorbitantly costly. However, it may be sufficient to use a fixed estimate of the scatter sinogram computed from a preliminary estimate of the transmission image. Our basis for this claim is that the scatter sinogram is typically very smooth and is robust to small variations in

the image. Consequently, computation from an approximate attenuation image should not differ considerably from computation using a more accurate image. Ollinger [30] recently presented a technique for computation of the 3D scatter sinogram for emission PET data. In [29] we describe a fast in-plane scatter computation for the transmission case. In both cases, the scatter model is based on the Klein-Nishina distribution.

Given independent estimates of the blank sinogram, randoms sinogram and scatter sinogram, we can write the likelihood function for the transmission image as:

$$p(y|\mu) = \prod_i e^{-\hat{y}_i} \frac{\hat{y}_i^{y_i}}{y_i!}, \quad \hat{y}_i = \hat{t}_i(\mu) + \hat{s}_i + \hat{r}_i \quad (3)$$

Thus, the log-likelihood function is

$$\ln p(y|\mu) = \sum_i \{-\hat{y}_i + y_i \ln(\hat{y}_i)\} \quad (4)$$

2.2 Emission Data

The development of the emission model is very similar to that of the transmission model. Again the data can be represented as a set of independent Poisson random variables x_i with mean \hat{x}_i equal to the sum of three components: true coincidences \hat{e}_i , randoms \hat{r}_i , and scatter \hat{s}_i . The major difference is that in this case the unknown parameters, the emission intensities in each pixel, are related to the mean of the data through an affine transform rather than an exponential.

$$\hat{x}_i = E(x_i) = \hat{e}_i(\lambda) + \hat{s}_i + \hat{r}_i \quad (5)$$

where

$$\hat{e}_i = \sum_j p_{ij} \lambda_j \quad (6)$$

In this case the elements p_{ij} of the matrix \mathbf{P} denote the probability of detecting an emission from pixel site j at detector pair i . Assume that there are M detector pairs and N^2 pixels in the image. Then \mathbf{P} is an $M \times N^2$ matrix which may be factored as follows:

$$\mathbf{P} = \mathbf{P}(\text{detectors})\mathbf{P}(\text{attenuation})\mathbf{P}(\text{geometric})\mathbf{P}(\text{positron}) \quad (7)$$

where $\mathbf{P}(\text{detectors})$ is a banded $M \times M$ matrix that can include factors related to the detectors including intrinsic and geometrical sensitivity, crystal penetration, inter-crystal scatter and system deadtime; $\mathbf{P}(\text{attenuation})$ is a diagonal $M \times M$ matrix that contains the attenuation correction factors; $\mathbf{P}(\text{geometric})$ is an $M \times N^2$ matrix that contains the probabilities that a photon pair produced in each pixel site, in the absence of an attenuating medium, reaches each of the detector pairs - this matrix contains terms for both the geometrical relationship between the source volume and the detectors and for the effect of uncertainties in the angular separation of the photon pair produced by annihilation; $\mathbf{P}(\text{positron})$ is an $N^2 \times N^2$ matrix that can model the effects of positron range. A

similar factorization scheme for PET is described in [1]. The reconstruction schemes described below utilize forward and backprojection operations. Forward projection using equation (7) involves multiplying the image vector by each of these (sparse) matrices in turn. Similarly, backprojection requires multiplying a vector of length M by the transpose of each of these matrices in reverse order. Since these matrices are all very sparse and can, for the most part, be pre-calculated, these computations are very fast. A full model would include all of the above factors in \mathbf{P} . Currently, we have not included factors to model positron range and angular separation or inter-crystal penetration and scatter. These factors while important, particularly in the latest generation of high resolution scanners, are probably second order in comparison to the other factors listed above.

As with the transmission data model, we assume that the randoms and scatter components are additive processes which can be estimated independently of the iterative reconstruction process. Randoms can be estimated from a separate randoms sinogram using the nonlinear Bayesian smoothing technique described in the Appendix. The scatter component can be estimated using a forward scatter computation based on a preliminary reconstruction of the emission image as described above using the methods in [6] or [30].

Using this model we can then write the log likelihood as

$$p(x|\lambda) = \prod_i e^{-\hat{x}_i} \frac{\hat{x}_i^{x_i}}{x_i!}, \quad \hat{x}_i = \hat{e}_i(\lambda) + \hat{s}_i + \hat{r}_i \quad (8)$$

Thus, the log-likelihood function is

$$\ln p(x|\lambda) = \sum_i \{-\hat{x}_i + x_i \ln(\hat{x}_i)\} \quad (9)$$

Note that the only difference between the emission model (8) and the transmission model (3) is the form of the relationship between the mean of the true coincidences and the unknown parameters. Consequently, the algorithms for emission and transmission reconstruction can be developed simultaneously.

2.3 The Prior and Posterior Densities

The reconstruction methods developed below are based on a Bayesian formulation with the prior drawn from the class of Markov random fields. A wide range of MRF priors have been studied for emission and transmission tomography including intensity process models [10], [35], [12], [14] and intensity and line process models [18], [22], [11]. In this work we consider priors of both types. Since we will be using gradient based optimization it is advantageous to work with a prior which has continuous first derivatives such as the Geman and McClure [10] prior:

$$p(\theta) = \frac{1}{Z} e^{-\frac{1}{\beta} U(\theta)} = \frac{1}{Z} e^{-\frac{1}{\beta} \sum_i \sum_{j>i} v_{ij}(\theta_j, \theta_i)} \quad (10)$$

where θ denotes either the attenuation image μ or the emission image λ . $U(\theta)$ is a Gibbs energy function defined as the sum of potential functions $V_{ij}(\theta_j, \theta_i)$ on pairwise cliques of neighboring pixels:

$$V_{ij}(\theta_j, \theta_i) = \frac{k_{ij}(\theta_j - \theta_i)^2}{\delta^2 + (\theta_j - \theta_i)^2} \quad (11)$$

The constants k_{ij} are generally chosen as the inverse of the Euclidean distance between the pixel centers for each pixel pair. This prior is applicable to both two and three dimensional images with an appropriately defined neighborhood.

The potential function in (11) acts like a quadratic penalty for $(\theta_j - \theta_i)^2 \ll \delta^2$ but is bounded above by unity. Consequently, the prior progressively penalizes intensity differences and thus favors locally smooth images, but by controlling the behavior for large intensity differences through δ , true discontinuities in the image are not over penalized. We have found that with an appropriate choice of δ this prior yields reconstructions which are locally smooth but do not appear to over-smooth across intensity boundaries. In the case of transmission images, we expect the discontinuities in the images to correspond to bone/soft-tissue and soft-tissue/air boundaries. At 511keV, the attenuation coefficient for soft tissue is almost constant and the image histogram should contain three distinct peaks centered around 0.0cm^{-1} (air), 0.095cm^{-1} (soft-tissue) and 0.151cm^{-1} (bone) [17]. In the case of chest scans, lung tissue can have values distributed from zero to 0.095cm^{-1} . We have chosen the value of $\delta = 0.025$ as approximately half the intensity difference between bone and soft-tissue as a compromise between the desire for clean anatomical boundaries and smooth variations inside anatomical structures. More restrictive priors have been proposed in [7] and [31]. These priors specifically include the mean values for each tissue type. For emission imaging the choice of a value for δ in (11) is more difficult. Clearly the value should reflect the absolute concentrations in each pixel, so that δ should change linearly with dose. Currently we use a filtered backprojection reconstruction and choose δ as approximately half the intensity of the significant background activity (e.g. the white matter intensity in an ^{18}F FDG brain image).

An alternative to (11) which explicitly models the presence of discontinuities in the image are the class of MRFs which include a binary line process to represent these discontinuities. In general the Gibbs energy function for these models may include potentials on cliques which include neighboring line sites - these potentials are typically defined to encourage the formation of continuous boundaries in the image. A more restricted version of this model, the so-called weak membrane model, does not include cliques containing more than one line site and has the following form:

$$p(\theta, l) = \frac{1}{Z} e^{-\frac{1}{\beta} U(\theta, l)} \quad (12)$$

where

$$U(\theta, l) = \sum_i \sum_{j>i} k_{ij} \{ (\theta_i - \theta_j)^2 (1 - l_{ij}) + \alpha_{ij} l_{ij} \} \quad (13)$$

Here the line process l_{ij} represents the presence ($=1$) or absence ($=0$) of a discontinuity between neighboring pixels i and j . The model is general for two or three dimensional

images and the line process can be defined for both directly and diagonally adjacent pixel pairs. The constants k_{ij} are again generally chosen as the inverse of the Euclidean distance between pixels i and j and the constants α_{ij} represent the penalty associated with introducing a discontinuity between pixels i and j (note that when the line process is 'on' the first term in the energy function (13) becomes zero). The simplest choice of α_{ij} would be a constant for all pairs $\{ij\}$. However, if additional information were available, in the form of anatomical boundary information for example, then this could be used to influence the formation of the line process through spatially variant α_{ij} 's [11]. In Section (3.5) we show how the gradient-based optimization schemes presented below can be adapted to MAP estimation using this weak membrane model via a continuation method similar to those in [2] and [11].

In the following section we describe a procedure for computing a MAP estimate by maximizing over the posterior densities $p(\mu|y)$ and $p(\lambda|x)$ for transmission and emission reconstruction respectively. Applying Bayes theorem for the Geman and McClure prior (11) we can write the log posterior probabilities as:

$$\begin{aligned} \ln p(\mu|y) &= \ln p(y|\mu) - \frac{1}{\beta} U(\mu) + K_t \\ &= \sum_i \{-\hat{y}_i + y_i \ln(\hat{y}_i)\} - \frac{1}{\beta} \sum_i \sum_{j>i} V_{ij}(\mu_i, \mu_j) + K_t \end{aligned} \quad (14)$$

and

$$\begin{aligned} \ln p(\lambda|x) &= \ln p(x|\lambda) - \frac{1}{\beta} U(\lambda) + K_e \\ &= \sum_i \{-\hat{x}_i + x_i \ln(\hat{x}_i)\} - \frac{1}{\beta} \sum_i \sum_{j>i} V_{ij}(\lambda_i, \lambda_j) + K_e \end{aligned} \quad (15)$$

where K_t and K_e are constants. The extensions to the weak membrane prior are described in Section (3.5).

3 Computing the MAP Estimate

3.1 The Preconditioned Conjugate Gradient Method

The classical approach to the maximization of an unconstrained functional $\Phi(\theta)$ with continuous derivatives is to use some form of gradient ascent:

$$\theta^{(n+1)} = \theta^{(n)} + \alpha^{(n)} p^{(n)} \quad (16)$$

where the direction vector $p^{(n)}$ is a function of the gradient vector $g^{(n)} = \nabla \Phi(\theta)_{\theta=\theta^{(n)}}$. The scalar parameter $\alpha^{(n)}$ is usually chosen using a line search to maximize $\Phi(\theta^{(n+1)})$. In the steepest ascent algorithm, the direction vector is simply the gradient vector $g^{(n)}$. In this case the convergence rate depends on the condition number of the Hessian $H(\theta^*)$ of $\Phi(\theta^*)$. The rate of convergence can often be dramatically improved by modifying the

direction vector. For example, Newton's method uses the direction vector formed by the product of the inverse of the Hessian with the gradient vector. Newton's method exhibits quadratic convergence properties, but exact computation of the inverse of the Hessian is impractical and is often replaced with some approximate inverse (C). In this case the convergence rate is determined by the condition number of $CH(\theta^*)$ [26].

An alternative to the steepest ascent and Newton methods is the method of conjugate gradients. For quadratic cost functions convergence is attained in a finite number of steps. When modified to non-quadratic problems, while the method no longer converges in a finite number of steps, the conjugate gradient method often exhibits favorable convergence rates. Convergence properties can be further enhanced by the introduction of a positive definite preconditioning matrix C . The improvement gained by the use of the preconditioner again depends on the condition number of the preconditioned Hessian CH . The preconditioned Polak-Ribiere form of conjugate gradient (PCG) is as follows:

$$\begin{aligned}\theta^{(n+1)} &= \theta^{(n)} + \alpha^{(n)} p^{(n)} && : \alpha^{(n)} \text{ by line search} \\ p^{(n)} &= d^{(n)} + \beta^{(n-1)} p^{(n-1)} \\ d^{(n)} &= C g^{(n)} \\ \beta^{(n-1)} &= \frac{(g^{(n)} - g^{(n-1)})^T d^{(n)}}{g^{(n-1)T} d^{(n-1)}}\end{aligned}\tag{17}$$

The PCG algorithm is initialized with $p^{(0)} = d^{(0)}$, and iteratively computes the conjugate directions, $p^{(n)}$. Care must be exercised to ensure that $p^{(n)}$ is an ascent direction. If the $p^{(n)}$ update equation is premultiplied by $g^{(n)T}$, then

$$g^{(n)T} p^{(n)} = g^{(n)T} C g^{(n)} + \beta^{(n-1)} g^{(n)T} p^{(n-1)}\tag{18}$$

The first term on the right hand side is positive by the positive definiteness of C . The second term, though, can become negative which can potentially result in

$$g^{(n)T} p^{(n)} < 0\tag{19}$$

which makes $p^{(n)}$ a descent direction. In these cases, we reinitialize the PCG algorithm with $p^{(n)} = d^{(n)}$.

3.2 Non-negativity and the Penalty Function

A major problem that arises with any of the gradient based methods discussed above is that they are designed for unconstrained optimization, yet for both transmission and emission PET, it is necessary to confine the image to the non-negative orthant. This can be easily achieved by limiting the step size $\alpha^{(n)}$ so that the image remains non-negative at each iteration. In images with many zero valued pixels, convergence is very slow, since the constraint boundary is hit at many iterations [19]. A more successful approach is to use a clipping operator [39], i.e. if the n^{th} iteration produces negative pixel values, they are projected back onto the constraint surface using a clipping operator.

A major disadvantage with this however, is that any time the clipping operator is applied, a second forward projection of the new image will be required to proceed to the next iteration because the new image is now a nonlinear function of the previous image and the unconstrained update vector. This can almost double the per-iteration computational cost.

An alternative to the two schemes outlined above is to convert the constrained optimization problem into an unconstrained problem using either a barrier or penalty function [26]. The barrier function method confines the search to the interior of the constraint region (in our case the non-negative orthant) by adding to the cost function a term which increases unboundedly as the boundary of the constraint region is approached. The penalty function method adds to the cost function a penalty term which penalizes values which lie outside the allowed interval. Associated with these methods is a parameter γ that determines the relative weight of the penalty or barrier function and the original cost function. In the limit as this parameter goes to zero, the solution of the new unconstrained problem approaches the solution of the original constrained problem.

In the following we use a penalty function method. Consider the problem

$$\begin{aligned} & \text{maximize } \Phi(\theta) \\ & \text{subject to } \theta \geq 0 \end{aligned} \quad (20)$$

where Φ is a continuous function on R^n . The idea of the penalty function method is to replace the problem (20) by an unconstrained problem of the form

$$\text{maximize } q(\gamma, \theta) = \Phi(\theta) - \frac{1}{\gamma} P(\theta) \quad (21)$$

where γ is a positive constant and P is a continuous function on R^n satisfying (i) $P(\theta) \geq 0$ for all $\theta \in R^n$, and (ii) $P(\theta) = 0$ iff $\theta \geq 0$. When a gradient based technique is used to solve the new problem, it is also preferable that the function has continuous first derivatives. The procedure for solving (20) is the following : Let $\{\gamma_k\}$, $k = 1, 2, \dots$ be a sequence tending to zero such that $\gamma_k > 0$, $\gamma_k > \gamma_{k+1}$. For each k , find the θ_k that maximizes $q(\gamma_k, \theta)$. Then any limit point of the sequence θ_k is a solution to (20) [26].

For the transmission problem we use the following quadratic penalty function:

$$\frac{1}{\gamma} P(\mu) = \sum_j \left(\frac{\mu_j}{\gamma} \right)^2 u(-\mu_j) \quad (22)$$

where $u(\cdot)$ is the unit step function. Although in theory a sequence of solutions should be generated corresponding to a decreasing sequence in the parameter γ , in practice we find that provided an appropriate value of γ is used it can be held constant throughout the iteration process without either significantly reducing the convergence rate of the algorithm or resulting in significant negative pixel values. A value of $\gamma = 0.01\theta_{max}$, where θ_{max} is the maximum pixel intensity in the image was experimentally observed to be a good choice for many computed phantom and clinical studies in both transmission and emission reconstructions. In the transmission case $\theta_{max} \approx 0.151 \text{ cm}^{-1}$, but in the emission case a preliminary FBP reconstruction is necessary to determine θ_{max} .

In the emission case we use a modified form of (22):

$$\frac{1}{\gamma} P(\lambda) = \sum_j \left(\frac{\lambda_j - \gamma_s}{\gamma} \right)^2 u(-\lambda_j + \gamma_s), \quad \gamma_s \geq 0 \quad (23)$$

We found that this produced faster convergence than (22). We believe that the slower convergence when using (22) is due to the fact that the log likelihood component of the posterior in (15) diverges as the mean of any element \hat{x}_i of the sinogram, for which the data x_i is non-zero, approaches zero. Therefore a negative gradient sum over pixels associated with a projection which is already close to zero (for the current image estimate) can produce a very small step size which in turn reduces the convergence rate of the algorithm. The modified penalty (23) adds a small positive gradient to small positive pixels and results in larger step sizes and faster convergence. As $\gamma_s \rightarrow 0$ this effect is gradually removed. We have observed empirically that reducing γ_s according to the sequence $\gamma_s(n) = \frac{(0.8)^n \theta_{max}}{100}$ at iteration n , results in fast convergence. We note that the slow convergence when using (22) was encountered primarily with real data for which separate scatter and random components were not available. Since these components are usually non-zero in all elements of the sinogram, we anticipate that this problem will not arise when these components are included in the procedure. Finally we note that in the transmission case there is no such problem, since even when projections become negative, the log-likelihood function is still bounded.

3.3 Preconditioners

The trick to achieving fast convergence of the conjugate gradient method is to use an appropriate preconditioner. Recall that the convergence characteristics of the conjugate gradient method are determined by the eigenstructure of the Hessian. When a preconditioner is used, the convergence behavior is then dependent on the product of the preconditioner and the Hessian. There are two problems that the preconditioner is capable of dealing with. Firstly, the poor conditioning of the Hessian of the original posterior probabilities can be circumvented through the use of a preconditioner. Second, the conditioning of the Hessian can be greatly affected by the penalty function if any of the constraints are active, i.e. if any pixels are negative. An appropriate preconditioner can compensate for this effect.

Lange *et al* [20] noted that the EM-ML algorithm for emission reconstruction can be viewed as a gradient ascent method where the gradient is modified by a diagonal preconditioner:

$$C_E^{(n)} = \text{diag} \left\{ \frac{\lambda_k^{(n)}}{\sum_i p_{ik}} \right\} \quad (24)$$

Kaufman [19] used this idea as the basis for a preconditioned conjugate gradient technique for maximum likelihood estimation. In the case of transmission imaging, Lange

et al [20] propose the following preconditioner:

$$C_T^{(n)} = \text{diag}\left\{\frac{\mu_k^{(n)}}{\sum_i Y_i l_{ik}}\right\} \quad (25)$$

In the following we use the preconditioner C_E for emission reconstruction and the following modified form of Lange's preconditioner for transmission reconstruction:

$$C_T^{(n)} = \text{diag}\left\{\frac{\mu_k^{(n)}}{\sum_i l_{ik}}\right\} \quad (26)$$

We find in practice that there is no noticeable difference in convergence behavior between (25) and (26). Note that the preconditioners above are positive definite only when the current estimates of the image are strictly positive. They can be made positive definite by simply thresholding the λ 's in (24) and the μ 's in (26) at a small positive constant. We use $0.01\lambda_{\max}$ for (24) and $0.01\mu_{\max}$ for (26) as threshold values. We have also studied other more complex, non-diagonal forms of preconditioners similar to those investigated in [4], but again found no further improvement in convergence rate.

When a penalty function is used to impose non-negativity, the preconditioners above need to be modified. Since it is not practical to compute optimal solutions to the unconstrained problem (21) for a large number of values of the parameter γ , we can instead try to maximize $q(\gamma_k, \theta)$ for a single small value of γ_k . Unfortunately, as shown in [26], the eigenvalue structure of the Hessian of the modified objective function becomes increasingly unfavorable as $\gamma \rightarrow 0$. Following [26], we can partition the Hessian (\hat{H}) of the modified cost function $q(\gamma_k, \theta)$ as

$$\hat{H} = H + \Gamma = \begin{bmatrix} H_{11} & H_{12} \\ H_{12}^T & H_{22} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \text{diag}\{-\frac{2}{\gamma^2}\} \end{bmatrix} \quad (27)$$

where H denotes the Hessian of the unpenalized functional $\Phi(\theta)$ and Γ denotes the Hessian of the penalty function (22). Here, the elements of the vector θ have been rearranged as $[\theta_1^T \theta_2^T]$ where θ_1 are the elements of θ which currently do not violate the positivity constraint, and θ_2 are the remaining elements which do violate the constraint. The submatrices H_{kl} $k, l = 1, 2$ represent the corresponding partitions of the unpenalized Hessian. Clearly \hat{H} becomes increasingly ill-conditioned as γ is reduced, if any of the pixels violate the positivity constraint.

To overcome this problem we define a preconditioner which compensates for the ill-conditioning:

$$C_{\text{penalty}}^{(n)} = \begin{bmatrix} I & 0 \\ 0 & \psi \text{diag}\{-\frac{\gamma^2}{2}\} \end{bmatrix}. \quad (28)$$

where ψ is a scalar constant. For γ sufficiently small, the lower-right block of \hat{H} in (27) is dominated by the penalty term and the rate of convergence of this method is

determined by the eigenvalues of

$$C_{\text{penalty}}^{(n)} \hat{H} = \begin{bmatrix} H_{11} & H_{12} \\ 0 & \psi I \end{bmatrix} \quad (29)$$

of which k should equal ψ , where k is the number of pixels violating the constraint. The remaining eigenvalues are the eigenvalues of H_{11} . Thus, if the smallest and largest eigenvalues of H_{11} $\{a, A\}$ are located such that the interval $[a, A]$ contains ψ , the convergence rate of the conjugate gradient will be equal to the canonical rate $[(A - a)/(A + a)]^2$. To prevent the penalty function slowing the convergence rate it is necessary to choose ψ to lie in the interval $[a, A]$. Currently, we choose ψ by trial and error. However, we have found that for fixed γ , (our choice of γ is described in Section (3.2)), a single value of ψ results in fast convergence over a wide range of clinical data sets.

In conclusion, we can introduce preconditioners to deal with the illconditioning associated with both the original unconstrained function (using (24) or (26)) and the penalty function (using (28)). The overall preconditioner is then the product $C_{\text{penalty}} C_T$ for transmission reconstruction and $C_{\text{penalty}} C_E$ for emission reconstruction. In the standard applications of preconditioners, the preconditioner is not changed at each iteration. Changing the preconditioner can certainly have adverse effects on the convergence behavior, particularly for a conjugate gradient method. However, one could argue that the preconditioner will not change rapidly from one iteration to the next. It is certainly the case that practical benefits, in the form of fast convergence, are realized.

3.4 Optimal Line Searches

The preconditioned conjugate gradient method in (17) requires a line search at each step to compute the optimal step size $\alpha^{(n)}$. To perform this line search we use a Newton-Raphson procedure:

$$\alpha^{(k+1)} = \alpha^{(k)} - \frac{(\partial q(\gamma, \theta^{(n+1)})/\partial \alpha)}{(\partial^2 q(\gamma, \theta^{(n+1)})/\partial \alpha^2)} \quad (30)$$

where $\theta^{(n+1)} = \theta^{(n)} + \alpha^{(n)} p^{(n)}$ in (17) and $q(\gamma, \theta^{(n+1)})$ is the penalized posterior density evaluated at $\theta^{(n+1)}$. Here we consider the computation involved in (30) for both the transmission and emission reconstruction.

In the transmission case, the penalized posterior density has the form:

$$\ln p(\mu^{(n+1)} = \mu^{(n)} + \alpha p^{(n)} | y) = \sum_i \{ -(v_i^{(n)} + \hat{s}_i + \hat{r}_i) + y_i \ln(v_i^{(n)} + \hat{s}_i + \hat{r}_i) \} - \frac{1}{\beta} U(\mu^{(n+1)}) - \frac{1}{\gamma} P(\mu^{(n+1)}) \quad (31)$$

where

$$v_i^{(n)} = u_i e^{(-\sum_m l_{im} \mu_m^{(n)})} e^{(-\alpha \sum_m l_{im} p_m^{(n)})} \quad (32)$$

Taking first and second derivatives of this function with respect to the step size α gives the following expressions:

$$\begin{aligned}
\frac{\partial g}{\partial \alpha} &= \sum_i \{v_i^{(n)} (\sum_m l_{im} p_m^{(n)}) (1 - \frac{y_i}{v_i^{(n)} + \hat{s}_i + \hat{r}_i})\} - \frac{1}{\beta} \nabla U(\mu^{(n+1)}) - \frac{1}{\gamma} \nabla P(\mu^{(n+1)}) \\
\frac{\partial^2 g}{\partial \alpha^2} &= - \sum_i \{v_i^{(n)} (\sum_m l_{im} p_m^{(n)})^2 (1 - \frac{y_i (\hat{s}_i + \hat{r}_i)}{\{v_i^{(n)} + \hat{s}_i + \hat{r}_i\}^2})\} - \frac{1}{\beta} \nabla^2 U(\mu^{(n+1)}) - \frac{1}{\gamma} \nabla^2 P(\mu^{(n+1)})
\end{aligned} \tag{33}$$

Examination of equations (31) and (33) reveals that once the update direction vector $p^{(n)}$ has been computed and forward projected, evaluation of the first and second derivatives in (33) is fast. Therefore, at each iteration the heaviest computational work is a backward projection to find the gradient vector, $g^{(n)}$, and a forward projection of the direction vector, $p^{(n)}$. The forward projection of $\mu^{(n+1)}$ can be computed as a linear combination of the projection of the previous estimate $\mu^{(n)}$ and the update direction vector $p^{(n)}$. Using an HP 9000(series 700) workstation we found that for a 128×128 pixel image forward and back projection require approximately five seconds each and two to three iterations of the Newton-Raphson search requires less than one second. The per-iteration computation cost of GEM and PCG are almost equal. The PCG line search procedure is replaced in GEM by a coordinate ascent search procedure.

Although, the log-likelihood function for the transmission model without additional scatter and randoms terms is concave, when these terms are added this is no longer true. Addition of the prior introduces additional non-concavity. Consequently, at best we would expect to converge on a local maximum using our conjugate gradient method. In the line search, the Newton-Raphson will fail for values of the step-size parameter for which the corresponding behaviour of the cost function is non-concave. This is simply checked by looking at the sign of the second derivative in (30). In the event that the Newton-Raphson procedure fails, an approximate line search can be performed using Armijo's rule [26].

The development for the emission case is very similar to that described above. The penalized posterior density has the form:

$$\begin{aligned}
\ln p(\lambda^{(n+1)} = \lambda^{(n)} + \alpha p^{(n)} | x) &= \sum_i \{-(v_i^{(n)} + \hat{s}_i + \hat{r}_i) + x_i \ln(v_i^{(n)} + \hat{s}_i + \hat{r}_i)\} \\
&\quad - \frac{1}{\beta} U(\lambda^{(n+1)}) - \frac{1}{\gamma} P(\lambda^{(n+1)})
\end{aligned} \tag{34}$$

where

$$v_i^{(n)} = \sum_j p_{ij} \lambda_j^{(n)} + \alpha \sum_j p_{ij} p_j^{(n)} \tag{35}$$

Taking first and second derivatives of this function with respect to the step size α gives the following expressions:

$$\begin{aligned}
\frac{\partial g}{\partial \alpha} &= - \sum_i \{(\sum_j p_{ij} p_j^{(n)}) (1 - \frac{x_i}{v_i^{(n)} + \hat{s}_i + \hat{r}_i})\} - \frac{1}{\beta} \nabla U(\lambda^{(n+1)}) - \frac{1}{\gamma} \nabla P(\lambda^{(n+1)}) \\
\frac{\partial^2 g}{\partial \alpha^2} &= - \sum_i \{ \frac{(\sum_j p_{ij} p_j^{(n)})^2 x_i}{\{v_i^{(n)} + \hat{s}_i + \hat{r}_i\}^2} \} - \frac{1}{\beta} \nabla^2 U(\lambda^{(n+1)}) - \frac{1}{\gamma} \nabla^2 P(\lambda^{(n+1)})
\end{aligned} \tag{36}$$

As with the transmission case, once the new search direction vector $p^{(n)}$ has been forward projected, the additional cost of performing the line search is small.

In the transmission case, range for the Newton-Raphson line search over which $P(\mu^{(n+1)}|y)$ is bounded is $\alpha^n \epsilon [0, \infty)$. In the emission case, $P(\lambda^{(n+1)}|x)$ is bounded for $\alpha^n \epsilon [0, \alpha_{max}^n)$ where

$$\alpha_{max}^n = \min_{i : \sum_j p_{ij} p_j^{(n)} < 0} \left\{ \frac{-\hat{s}_i - \hat{r}_i - \sum_j p_{ij} \lambda_j^{(n)}}{\sum_j p_{ij} p_j^{(n)}} \right\} \quad (37)$$

3.5 Deterministic Annealing Techniques

The preconditioned conjugate gradient method can be modified for use in MAP estimation in conjunction with the weak membrane prior defined in (12) using a continuation method. The basic idea behind the continuation methods is to approximate a non-concave function which has discontinuous derivatives with a sequence of smooth functions that converges uniformly to the original function. Through this procedure it is possible to avoid local maxima of the original function, and thus converge on a more desirable maximum. Although in general no global convergence properties are known, favorable behavior has been observed by several authors [2], [11] and [9]. Here we briefly describe how this approach can be combined with the gradient methods described here. Readers are referred to [2], [11] and [9] for a treatment of the underlying theory.

It is straightforward to show that the MAP solution for the weak membrane model (12) (with intensity and line processes) is identical to that obtained using a "broken parabola" (with an intensity process only) with the following Gibbs energy function:

$$U(\theta) = \sum_i \sum_{j>i} V_{ij}(\theta_i, \theta_j) \quad (38)$$

where

$$V_{ij}(\theta_i, \theta_j) = \begin{cases} k_{ij}(\theta_i - \theta_j)^2 & \text{if } (\theta_i - \theta_j)^2 \leq \alpha_{ij} \\ k_{ij}\alpha_{ij} & \text{if } (\theta_i - \theta_j)^2 > \alpha_{ij} \end{cases} \quad (39)$$

Direct optimization using (38) is difficult as the function does not have continuous derivatives and its gradient is zero once the intensity difference exceeds a given threshold. The continuation methods approximate this function with a sequence of smooth functions which converge to this broken parabola. In [11] it is shown that an appropriate sequence can be found using a marginal posterior parameterized by a temperature parameter T . To briefly summarize this approach we first note that maximizing over the posterior $p(\theta, l/y)$ (where y is the data) is equivalent to maximizing over the modified posterior:

$$p_T(\theta, l/y) = \frac{1}{K_T} p(\theta, l/y)^{\frac{1}{T}} \quad (40)$$

where K_T is the appropriate normalizing constant. Now let $p_T(\theta/y)$ denote the marginal of the modified posterior:

$$p_T(\theta/y) = \sum_{\{l\}} p_T(\theta, l/y) = \frac{1}{K_T} \frac{p(y/\theta)^{\frac{1}{T}}}{p(y)^{\frac{1}{T}}} \sum_{\{l\}} p(\theta, l)^{\frac{1}{T}} \quad (41)$$

where the sum over $\{l\}$ denotes the sum over all possible line configurations. It can be shown that in the limit as $T \rightarrow 0$ any θ which maximizes $p_T(\theta/y)$ is also a maximizer of the posterior with the broken parabolic prior (38). In turn, any θ which is a MAP estimate over the broken parabolic prior is also an image which maximizes the original posterior density with the weak membrane prior. Thus T parameterizes a sequence of functions such that the θ_T^* which maximizes these functions converges to the θ^* which solves the original MAP optimization problem as $T \rightarrow 0$.

It can be shown [2][11] that integrating over the line process yields the following Gibbs potential function:

$$V_{ij}^T(\theta_i, \theta_j) = -T \ln [e^{-\frac{k_{ij}}{T\beta}(\theta_i - \theta_j)^2} + e^{-\frac{k_{ij}\alpha_{ij}}{T\beta}}] \quad (42)$$

It is easy to prove that

$$\lim_{T \rightarrow 0} V_{ij}^T(\theta_i, \theta_j) = \begin{cases} k_{ij}(\theta_i - \theta_j)^2 & \text{if } (\theta_i - \theta_j)^2 \leq \alpha_{ij} \\ k_{ij}\alpha_{ij} & \text{if } (\theta_i - \theta_j)^2 > \alpha_{ij} \end{cases} \quad (43)$$

i.e. the broken parabola.

For each T we compute θ_T^* as the maximizer over the log of the modified marginal posterior

$$\ln p_T(\theta/y) = \frac{1}{T} \ln p(y/\theta) - \frac{1}{T} \sum_i \sum_{j>i} V_{ij}^T(\theta_j, \theta_i) \quad (44)$$

By slowly reducing T as each new θ_T^* is reached, we converge to a maximum of the original posterior density. Since for large T the function $\ln p_T(\theta/y)$ is smoother than $\ln p(\theta/y)$, this annealing process can avoid local minima which may otherwise be encountered by optimizing directly over the original posterior density. However, this is a deterministic annealing scheme, so we would expect convergence to a local maximum only.

A GEM-based continuation method for emission tomography which uses this approach is described in [11]. In order to obtain a closed form M-step in this algorithm the line process (or more precisely, its conditional mean) is reintroduced at each iteration. As an alternative, we can apply the penalized conjugate gradient procedure to optimize over (44) with only minor modifications. Ideally, the algorithm should track the solution θ_T^* continuously with respect to T . In practice, optimization can only be performed at a finite number of temperatures. In the results presented below using this method we used only six different values of T starting with an initial value $T = 2.0$. T was then halved every 10 iterations. The PCG algorithm was also reinitialized every 10 iterations, starting at the 20th iteration. The algorithm was terminated at the 60th iteration as compared to the 30th iteration for the other PCG methods.

4 Results and Discussion

Below we examine the performance of the PCG methods using simulated and phantom data. In all these cases the Geman and McClure prior was used and two dimensional images were reconstructed using an 8-nearest neighbor model, i.e. each plane was reconstructed independently. Clinical results are then presented for a whole body study. In this case the transmission images were reconstructed using a two dimensional 8-nearest neighbor system and the emission images were reconstructed using both a two dimensional 8-nearest neighbor system and a three dimensional MRF model with 18-neighbors per interior pixel - 8 in the transaxial plane and 5 in each adjacent transaxial plane. Three dimensional reconstructions are presented for both the Geman and McClure prior and the weak membrane prior.

4.1 Simulations and Phantom Data

In Figure (1) we show convergence behavior using computer generated transmission data for a chest phantom and the Geman and McClure prior for the following algorithms: penalized preconditioned conjugate gradient (PCG) and preconditioned steepest ascent (PSA) as well as conjugate gradient (CG) and steepest ascent (SA) without preconditioning. Figures (1a) and (1b) show the posterior probability as a function of iteration number. These figures indicate significantly faster convergence using preconditioning and further improvements using conjugate gradient as opposed to steepest ascent. Figure (1c) shows the total squared error between the true image and the reconstructed image as a function of iteration. Also shown are the values of the total squared error for a PCG based maximum likelihood estimator and the fixed error associated with filtered backprojection. While squared error is a crude measure of image quality we can see from this plot the regularizing influence of the prior (relative to the maximum likelihood estimator) although at higher iterations the PCG method appears to exhibit a slow deterioration. This could probably be reduced, at the risk of over smoothing, by increasing the influence of the prior.

The convergence behavior for emission data for a computer based brain phantom is shown in Figure (2). The emission data was simulated with a 10% randoms component. Figure (2a) compares the posterior density as a function of iteration number for PCG and the GEM [14] methods. Figure (2b) compares the l_2 error (scaled by the l_2 norm of the true image) in the reconstruction using the PCG-MAP algorithm, in comparison to the standard EM-maximum likelihood algorithm, the GEM algorithm and FBP. Again we see the fast convergence behavior of PCG compared to GEM as well as the superior performance of the MAP estimate compared to ML and FBP. In these simulations, the reconstruction region was chosen as an elliptical region around the head. Our simulations reveal that in the case when the reconstruction region contains many zero-valued pixels outside the object, the convergence rate of PCG drops while that of GEM does not. For practical purposes, smaller reconstruction regions can be automatically estimated from the emission or transmission sinograms directly or from a transmission image if one is available.

In Figure (3) we show logarithmic plots of the l_1 and l_∞ norm of the update vector (times the step size) as a function of iteration number. These norms are scaled by the corresponding norm of the true image. It is interesting to note that the PCG method produces larger changes in the image in the earlier iterations than the GEM method. At later iterations, the changes in the image using PCG decrease rapidly compared to those using GEM. The l_∞ norm may prove particularly useful as a practical measure of convergence. Since this function is approximately monotonic in the earlier iterations, it is possible to (approximately) bound the maximum change that can occur in the image if further iterations are performed. To produce these plots, the software was run with all variables declared as double-precision. If single precision is used, similar behaviour is observed for approximately the first one hundred iterations. Beyond this numerical errors begin to dominate.

Figure (4) illustrates the effect of applying FBP and PCG-MAP techniques to transmission and emission data collected using a chest phantom in a Siemens/CTI ECAT931 scanner. Figures (4a) and (4b) compare reconstructions of the transmission images from data collected for 2 minutes. The remaining images are emission reconstructions comparing FBP and PCG-MAP with attenuation factors computed using various techniques. To distinguish the effects of different attenuation correction methods from the choice of emission reconstruction algorithm, we show in Figures (4c) and (4d) MAP and FBP emission reconstructions respectively from attenuation factors calculated using the standard procedure but using a transmission data set which was collected for three hours and hence exhibits a very high signal to noise ratio. Figure (4e) shows a MAP reconstruction from the same emission data set but using a reprojection of the two minute scan MAP-transmission image for attenuation correction. For comparison Figures (4f) and (4g) show FBP and MAP reconstructions respectively, both using standard attenuation correction factors calculated from the two minute transmission scan. These images show that it is possible to reconstruct reasonable attenuation images, and to achieve reasonably accurate attenuation correction, using relatively short transmission scans. Of these results, using reprojections of the MAP transmission image in conjunction with MAP reconstruction of the emission image appears to give superior results.

4.2 Clinical Application to Whole Body Data

We have also applied the PCG techniques described here to the reconstruction of transmission and ^{18}F FDG emission whole body scans collected using the 15 slice, Siemens/CTI ECAT931 scanner. A total of 14 frames (4 minutes each) of emission data were collected. To increase axial sampling, pairs of frames were interleaved by 1/2 of the slice thickness resulting in an axial sample interval of 0.3375cm. Transmission sinograms were also collected for 4 minutes each but no interleaving was performed resulting in an axial sample interval of 0.675cm. In all reconstructions the transaxial pixel size is 0.44cm by 0.44cm. The method used to reconstruct the whole body images differed slightly from that described above in one important respect. For a 3D reconstruction the conjugate gradient method should update all pixels simultaneously. The dimen-

sions of the image and the data are such that this becomes cumbersome. Rather, we update each plane sequentially. We use a true 3D prior with an 18-nearest neighbor model. Thus when we compute the gradient of the posterior, the contributions from the prior require computation of the derivatives of the Gibbs potential functions across planes. The resulting procedure can be viewed as a compromise between coordinate ascent, such as that found in GEM-type algorithms, and true gradient based procedure in which all pixels are updated simultaneously. When compared with 2D, the 3D computation speed is reduced by about 20% due to the larger number of computations associated with the larger neighborhood for the 3D prior and the delay incurred in reading and writing intermediate vectors to the hard disk.

The transmission data were collected simultaneously with a randoms sinogram. The emission data were collected using standard randoms subtraction techniques. Transaxial images were reconstructed using the PCG method with 30 iterations per image. Using an HP 9000(series 700) workstation we are able to reconstruct a full frame of 15 transaxial slices in approximately 75 minutes. An additional 75 minutes/frame was required to reconstruct the corresponding emission images. This time could probably be reduced by 33-50% by reducing the number of iterations with little noticeable loss in image quality. Further savings could be achieved by dropping zero valued pixels outside the patient from our updating procedure. Using the precomputed projection matrices, the corresponding savings would be approximately linearly related to the fraction of pixels removed.

Figures (5), (6) and (7) show transaxial, coronal and sagittal sections through reconstructions of the transmission and ^{18}F FDG emission images. The lettering (a) - (h) below is consistent for each of Figures (5), (6) and (7). In (a) and (b) we compare transmission reconstructions using filtered backprojection and MAP. The MAP images exhibit significantly lower variance and sharper boundary definition although it is difficult to clearly see the bone except for the skull and arms due to low count rates. FBP reconstructions of the emission images using standard attenuation correction and reprojections of the MAP transmission image are shown in (c) and (d) respectively. The remainder of the images compare statistical emission reconstructions (all using reprojections of the MAP transmission image) as follows: (e) ML; (f) MAP 2D Geman McClure; (g) MAP 3D Geman McClure; and (h) MAP 3D weak membrane prior. The overall quality of the MAP emission reconstructions, particularly in areas of low background activity, is superior to FBP, even when the FBP method uses the improved attenuation correction. The results also indicate that it is beneficial to extend the 2D spatial interaction terms in the prior to 3D interactions, since reconstructions with 2D interactions only, show a higher degree of local smoothness in the transaxial direction than in the axial direction.

5 Conclusions

We have presented a practical scheme for iterative Bayesian reconstruction of transmission and emission PET images. Simulation studies indicate that the penalized,

preconditioned conjugate gradient methods exhibit very favorable convergence rates and the use of MRF priors produce relatively stable behavior. Phantom studies reveal the potential for substantial improvement in image quality using MAP estimation in the computation of both the attenuation factors (via the transmission image) and the emission image, as compared to FBP. Results using clinical whole body data also appear to show a significant improvement in image quality compared to images formed using current clinical methods. We hope to report shortly on whether this is borne out by measurable improvements in lesion detectability.

The goal of this work was to produce a fast iterative procedure for PET reconstruction. In order to achieve this goal we have departed somewhat from the standard procedures for the use of conjugate gradient techniques with preconditioning and penalty functions. For example, the current preconditioner is changed at each iteration - this has the potential to negate the advantages of using a conjugate gradient procedure. However, our convergence plots appear to indicate that in practice the conjugate gradient procedure does converge faster than a steepest ascent method. Furthermore, we do check at each iteration to ensure that the update vector does increase the cost function at each iteration - if not, the procedure is reinitialized. A second potential source of problems is that the parameter of the non-negativity penalty function is held constant throughout the reconstruction procedure. Although the theory calls for this parameter to be decreased as each new optimum is reached, we find that in practice it is fairly easy to choose a value of the parameter that neither results in significant negative pixel values or noticeably reduces the convergence rate. As presented, these approximate procedures appear to work well in the sense that we observe significant qualitative improvement in computer generated, phantom and clinical images using these, as opposed to standard procedures. We are now in the process of performing objective evaluations of these techniques.

Acknowledgements

This work was supported by a grant from the Whitaker Foundation. The authors would like to thank Carl Hoh, M.D., and Arion Chatziioannou for providing us with the whole body PET data, and Steven Meikle for providing the chest phantom PET data presented in this paper.

References

- [1] J. Baker. *Spatially variant tomographic imaging: estimation, identification and optimization*. PhD thesis, Lawrence Berkeley Laboratory, University of California, Nov. 1991.
- [2] G. Bilbro, W. Snyder, S. Garnier, and J. Gault. Mean field annealing: A formalism for constructing gnc-like algorithms. *IEEE Transactions on Neural Networks*, 3, 1992.

- [3] J. Browne and T. Holmes. Developments with maximum likelihood x-ray computed tomography. *IEEE Transactions on Medical Imaging*, 11:40–52, 1992.
- [4] N. Clinthorne, T.S. Pan, et al. Preconditioning methods for improved convergence rates in iterative reconstruction. *IEEE Transactions on Medical Imaging*, 12(1):78–83, March 1993.
- [5] M. Dahlbom, E.J. Hoffman, et al. Whole-body pet: Part i. methods and performance characteristics. *Journal of Nuclear Medicine*, 33(6):1191–1199, June 1992.
- [6] M. Daube-Witherspoon, R. Carson, Y. Yan, and T. Yap. Scatter correction in ml reconstruction of pet data. In *Proc. IEEE Nuclear Science and Medical Imaging Conference*, 1992.
- [7] J.A. Fessler. Segmented attenuation correction for pet. In *Proc. IEEE Nuclear Science and Medical Imaging Conference*, 1992.
- [8] J.A. Fessler. Tomographic reconstruction using information-weighted spline smoothing. In *Proc. Information Processing in Medical Imaging Conference*, 1993.
- [9] D. Geiger and F. Girosi. Parallel and deterministic algorithms from mrf's: Surface reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (5), 1991.
- [10] S. Geman and D.E. McClure. Bayesian image analysis: An application to single photon emission tomography. *Proceedings of the American Statistical Association*, pages 12–18, 1985.
- [11] G. Gindi, A. Rangarajan, M. Lee, and I.G. Zubal. Bayesian reconstruction for emission tomography via deterministic annealing. In *Proc. Information Processing in Medical Imaging Conference*, 1993.
- [12] P.J. Green. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Transactions on Medical Imaging*, 9(1):84–93, March 1990.
- [13] T. Hebert. A union of deterministic and stochastic methods for image reconstruction. In *Proc. IEEE Nuclear Science and Medical Imaging Conference*, 1992.
- [14] T. Hebert and R. Leahy. A generalized EM algorithm for 3D bayesian reconstruction from poisson data using Gibbs priors. *IEEE Transactions on Medical Imaging*, 8(2):194–202, June 1989.
- [15] G.T. Herman and A.R. De Pierro. On methods for maximum a posteriori image reconstruction with a normal prior. *Journal of Visual Communication and Image Representation*, 3(4):316–324, December 1992.
- [16] E. Hoffman, T. Guerrero, G. Germano, W. Digby, and M. Dahlbom. PET system calibrations and corrections for quantitative and spatially accurate images. *IEEE Transactions on Nuclear Science*, NS-36:1108–1112, 1989.

- [17] J.H. Hubbell. Photon mass attenuation and energy-absorption coefficients from 1keV to 20meV. *Int.J.Appl.Radiat.Isot.*, 33:1269–1290, 1981.
- [18] V.E. Johnson, W.H. Wong, X. Hu, and C.T. Chen. Bayesian restoration of PET images using Gibbs priors. *Proc. 1989 International Conference on Information Processing in Medical Imaging*, Wiley-Liss, pages 15–28, 1991.
- [19] L. Kaufman. Implementing and accelerating the em algorithm for positron emission tomography. *IEEE Transactions on Medical Imaging*, 6(1):37–51, 1987.
- [20] K. Lange, M. Bahn, and R. Little. A theoretical study of some maximum likelihood algorithm for emission and transmission tomography. *IEEE Transactions on Medical Imaging*, MI-6(2):106–125, June 1987.
- [21] K. Lange and R. Carson. EM reconstruction algorithms for emission and transmission tomography. *Journal of Computer Assisted Tomography*, 8(2):306–316, April 1984.
- [22] R. Leahy and X.H. Yan. Incorporation of anatomical MR data for improved functional imaging with PET. *Proc. XIIIth International Conference on Information Processing in Medical Imaging*, Springer-Verlag, pages 105–120, 1991.
- [23] D.N. Levin, C.A. Pelizzari, G.T.Y. Chen, C.T. Chen, and M.D. Cooper. Retrospective geometric correlation of MR, CT, and PET images. *Radiology*, pages 817–823, 1988.
- [24] E. Levitan and G.T. Herman. A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography. *IEEE Transactions on Medical Imaging*, MI-6(3):185–192, September 1987.
- [25] R. Lewitt and G. Muehllehner. Accelerated iterative reconstruction for positron emission tomography based on the em algorithm for maximum likelihood estimation. *IEEE Transactions on Medical Imaging*, 5(1):16–22, 1986.
- [26] D.G. Luenberger. *Linear and Nonlinear Programming*. Addison Wesley, second edition, 1984.
- [27] S. Meikle, M. Dahlbom, and S.R. Cherry. Attenuation correction using count-limited transmission data in PET. *Journal of Nuclear Medicine*, 1993.
- [28] M. Miller, D. Snyder, and S. Moore. An evaluation of the use of sieves for producing estimates of radioactivity distribution with the EM algorithm for PET. *IEEE Transactions on Nuclear Science*, 33, 1986.
- [29] E. Mumcuoglu, R. Leahy, and S. Cherry. A statistical approach to transmission image reconstruction from ring source calibration measurements in PET. *Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference, Florida*, November 1992.

- [30] J. Ollinger and G. Johns. Model-based scatter correction in three dimensions. In *Proc. IEEE Nuclear Science and Medical Imaging Conference*, 1992.
- [31] J.M. Ollinger. The use of maximum a posteriori and maximum likelihood transmission images for attenuation correction in pet. In *Proc. IEEE Nuclear Science and Medical Imaging Conference*, 1992.
- [32] A.R. De Pierro. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. *Submitted for publication*, 1992.
- [33] D.G. Politte and D.L. Snyder. Corrections for accidental coincidences and attenuation in maximum likelihood image reconstruction for positron emission tomography. *IEEE Transactions on Medical Imaging*, MI-10(1):82–89, March 1991.
- [34] N. Rajeevan, K. Rajgopal, and G. Krishna. Vector-extrapolated fast maximum likelihood estimation algorithms for emission tomography. *IEEE Transactions on Medical Imaging*, 11(1):9–20, March 1992.
- [35] K. Sauer and C. Bauman. Bayesian estimation of transmission tomograms using segmentation based optimization. *IEEE Nucl. Med. Conf.*, II:165–170, 1990.
- [36] L.A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, MI-1(2):113–122, October 1982.
- [37] Siemens. *Technical Introduction to ECAT Scanner*, 1990.
- [38] C.J. Thompson, A. Dagher, D.N. Lunney, S.C. Strother, and A.C. Evens. A technique to reject scatter radiation in pet transmission scans. *SPIE Proceedings*, 671:244–253, 1986.
- [39] H.J. Trussell. Convergence criteria for iterative restoration methods. *IEEE Transactions on Acoustic, Speech and Signal processing*, 31:129–136, 1983.
- [40] E. Veklerov and J. Llacer. Stopping rule for the MLE algorithm based on statistical hypothesis testing. *IEEE Transactions on Medical Imaging*, MI-6(4):313–319, December 1987.

Appendix: Mean Sinogram Estimation

The elements of the blank sinogram (z_i) are Poisson random variables. Here we describe a Bayesian technique for estimating the mean of the sinogram from a single observation. We then describe the extension of this technique to estimating the mean of a randoms only sinogram.

Consider the following model parameters: (i) the intrinsic detector efficiencies, $\epsilon(k)$, and (ii) the source sinogram f_i which is a smoothly changing function that models

the source ring projections and the geometric detector efficiencies. Then we can model the mean of the true coincidences in the observed blank sinogram as:

$$\lambda(i) = f_i \epsilon(i_1) \epsilon(i_2) \quad (45)$$

where i_1 and i_2 are the indices of the detector pair corresponding to the i^{th} element of the sinogram. To model the blank sinogram accurately, randoms should also be included in the model. The randoms component of the sinogram can be modeled as a second Poisson process with mean

$$\hat{r}_i = 2\tau \text{ singles}(i_1) \text{ singles}(i_2) \quad (46)$$

where τ is the coincidence timing window, and $\text{singles}(\cdot)$ represents the singles rate at the individual detectors. Assuming that singles rates should vary smoothly (when intrinsic detector efficiencies are factored out), an equivalent model for the randoms is

$$\hat{r}_i = \bar{r}_i \epsilon(i_1) \epsilon(i_2) \quad (47)$$

where \bar{r}_i represents the smoothly changing mean of the randoms sinogram compensated for intrinsic detector efficiencies. Since (45) and (47) are in the same form, the blank sinogram can be modeled as

$$p(z|f) = \prod_i e^{-\lambda(i)} \frac{\lambda(i)^{z_i}}{z_i!}, \quad \lambda(i) = \hat{f}_i \epsilon(i_1) \epsilon(i_2), \quad \hat{f}_i = f_i + \bar{r}_i \quad (48)$$

where \hat{f}_i now includes both random and true coincidences. The log-likelihood function now becomes,

$$\ln p(z|f) = \sum_i \{-\hat{f}_i \epsilon(i_1) \epsilon(i_2) + z_i \ln[\hat{f}_i \epsilon(i_1) \epsilon(i_2)]\} \quad (49)$$

where the detector pair efficiencies are $\epsilon(i_1) \epsilon(i_2)$. The detector normalization file, which is usually acquired once a week using a planar source, provides a high count measurement which can be used as a good estimate of $\epsilon(i_1) \epsilon(i_2)$ after appropriate normalization for geometric efficiency.

We use a Bayesian technique to compute an estimate of the mean of the blank sinogram. A Gaussian Gibbs prior is used to reflect the expected smoothness of the blank sinogram.

$$p(f) = \frac{1}{\lambda} e^{-\frac{1}{\lambda} \sum_i \sum_{j>i} V_{ij}(f)} \quad (50)$$

where

$$V_{ij}(f) = k_{ij} (f_i - f_j)^2. \quad (51)$$

for all two pixel cliques on an eight nearest neighbor system with $k_{ij} = 1$ for vertical and horizontal neighbors and $k_{ij} = \frac{1}{\sqrt{2}}$ for diagonal neighbors. Maximizing the log of the posterior density

$$\ln(f|z) = \ln(z|f) - \frac{1}{\beta} \sum_i \sum_{j>i} V_{ij}(f). \quad (52)$$

can be performed using any gradient type algorithm since the posterior density in this case is concave. Once the mean blank sinogram has been estimated, it is possible to subtract out the mean random component provided an independent randoms sinogram is collected simultaneously. We use a simple modification of the above method to smooth the randoms sinogram before subtraction.

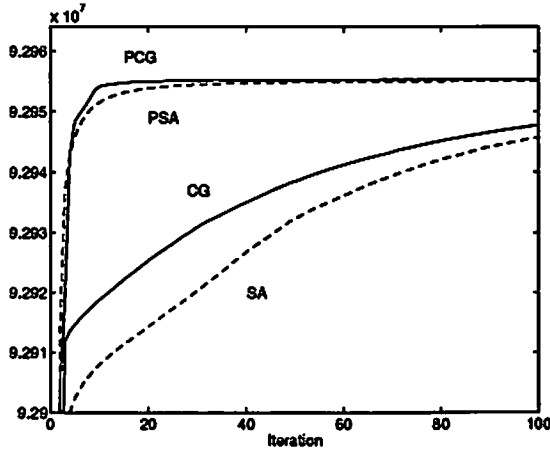
Daily changes in intrinsic detector efficiencies are possible [16]. An alternative to the use of the detector normalization file is to estimate the ϵ vector directly from the blank data set. Equating the derivative of (49) with respect to $\epsilon(j)$ to zero yields a simple update equation:

$$\epsilon^{(n+1)}(j) = \frac{\sum_i z_i}{\sum_i f_i \epsilon^{(n)}(i_2)} \quad (53)$$

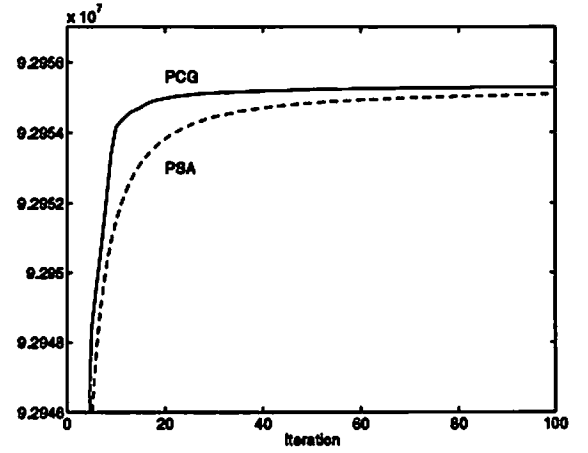
where the summation index i is for all sinogram components in coincidence with detector j . A logical extension is to estimate \hat{f} and ϵ consecutively. The computation cost is small in comparison to that of the MAP reconstruction of the transmission image.

The same idea can also be applied to perform a nonlinear smoothing of a standard transmission or emission sinogram using the Bayesian method and the estimated ϵ 's. Although transmission or emission sinograms exhibit more variation than the blank sinogram, a careful choice of prior will be very effective in reducing Poisson noise in the measurement. Thus, an alternative to full MAP reconstruction of transmission or emission PET images is to perform nonlinear Bayes smoothing of the sinograms via MAP estimation and then apply filtered backprojection to the result. This is similar to the methods presented in [13] and [8].

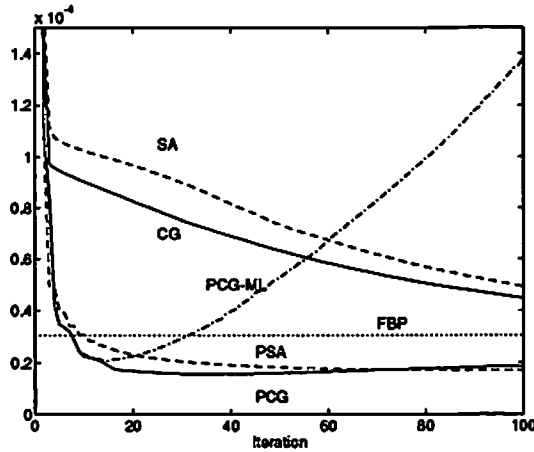
Finally, estimation of intrinsic detector efficiencies also provides an opportunity to estimate a smoothed randoms sinogram, \hat{r}_i , using a separately acquired randoms sinogram. Using the model for randoms in (47) (with estimated detector efficiencies) the Bayesian estimation scheme described in (49)-(52) can be applied directly to estimate \hat{r}_i .



(a)



(b)



(c)

Figure 1: (a) Values of the posterior density versus iteration number for reconstructions of a computer generated transmission phantom using steepest ascent (SA) and conjugate gradient (CG) methods, and preconditioned version of these methods (PCG,PSA). (b) Zoomed in version of (a) showing (PSA,PCG) only. (c) Total squared error between the true and reconstructed image versus iteration number using steepest ascent (SA) and conjugate gradient (CG) methods, preconditioned versions of these methods (PSA,PCG), filtered backprojection (FBP) and maximum-likelihood (PCG-ML).

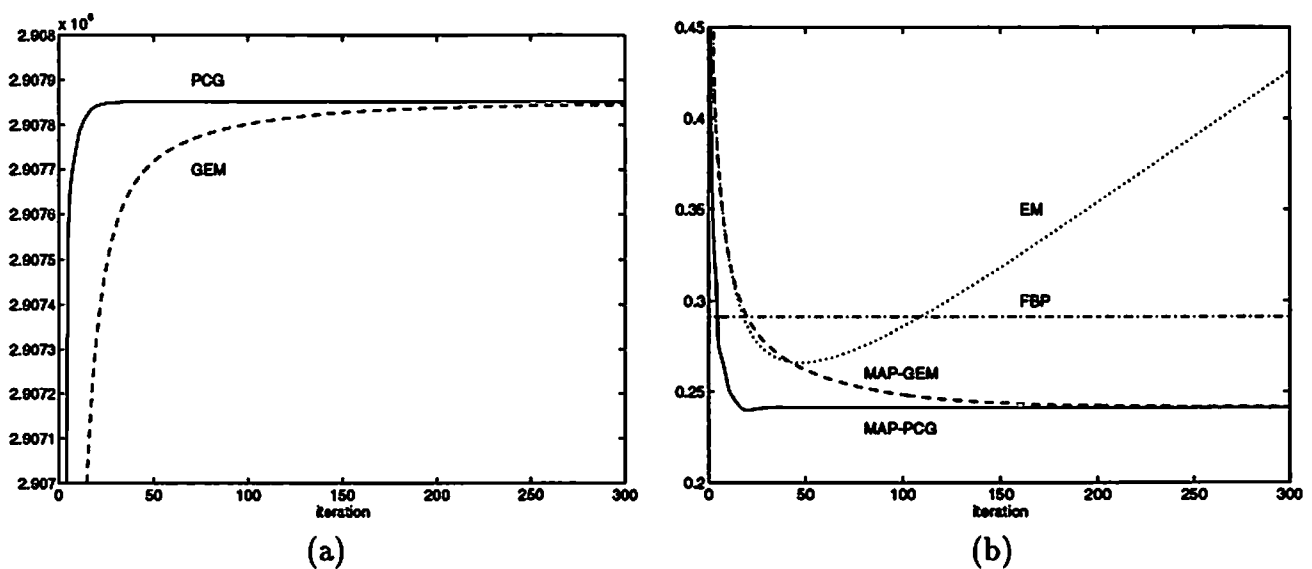


Figure 2: (a) Values of the posterior density versus iteration number for GEM and PCG for a computer generated emission phantom. (b) A comparison of the l_2 norm of the error between the reconstructed image and the true image (normalized by the l_2 norm of the true image) as a function of iteration number for maximum likelihood estimation via EM, and maximum a posteriori estimation using GEM and PCG. Also shown is the total squared error of the filtered backprojection reconstruction.

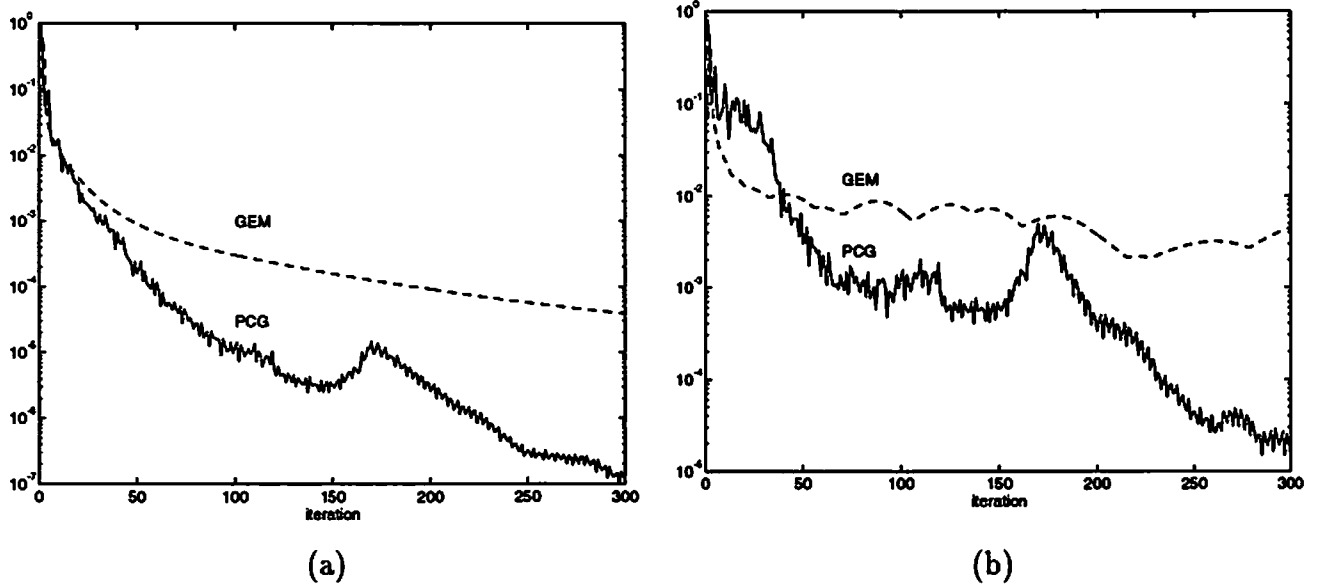


Figure 3: (a) l_1 norm of the update vector (times the step size) normalized to the l_1 norm of the true image for PCG and GEM for a simulated emission brain phantom. (b) l_∞ norm of the update vector (times the step size) normalized to the l_∞ norm of the true image for PCG and GEM for the same emission brain phantom.

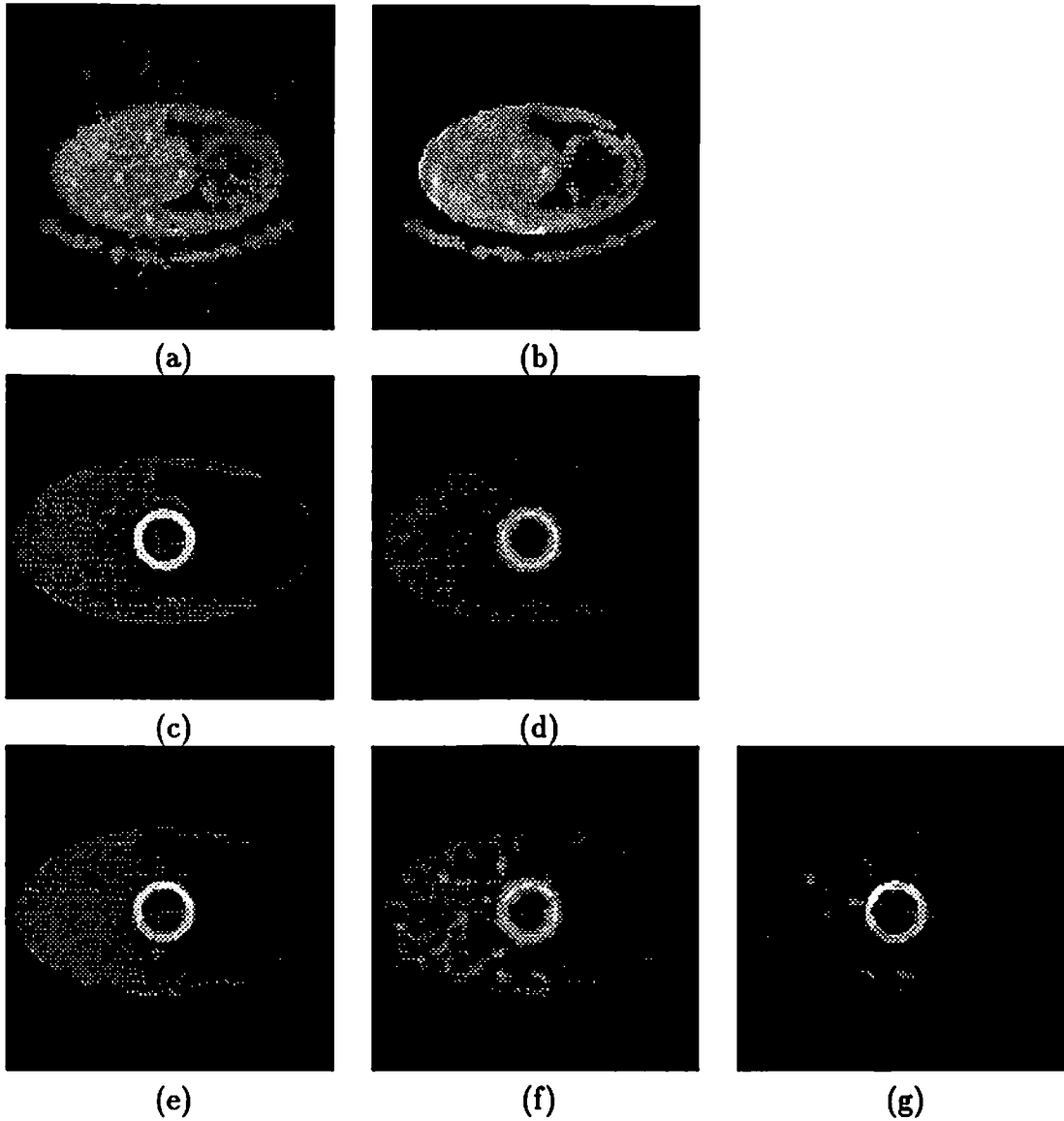


Figure 4: Transmission reconstructions of a chest phantom. (a) FBP reconstruction using two minute data (2MD); (b) MAP reconstruction using 2MD. Emission reconstructions using correction factors computed in each case from either 2MD or three hour (3HD) transmission scans. (c) MAP reconstruction using 3HD, standard correction technique; (d) FBP reconstruction using 3HD, standard technique; (e) MAP reconstruction using 2MD, MAP-based attenuation correction; (f) FBP reconstruction using 2MD, standard technique; (g) MAP reconstruction using 2MD, standard technique.

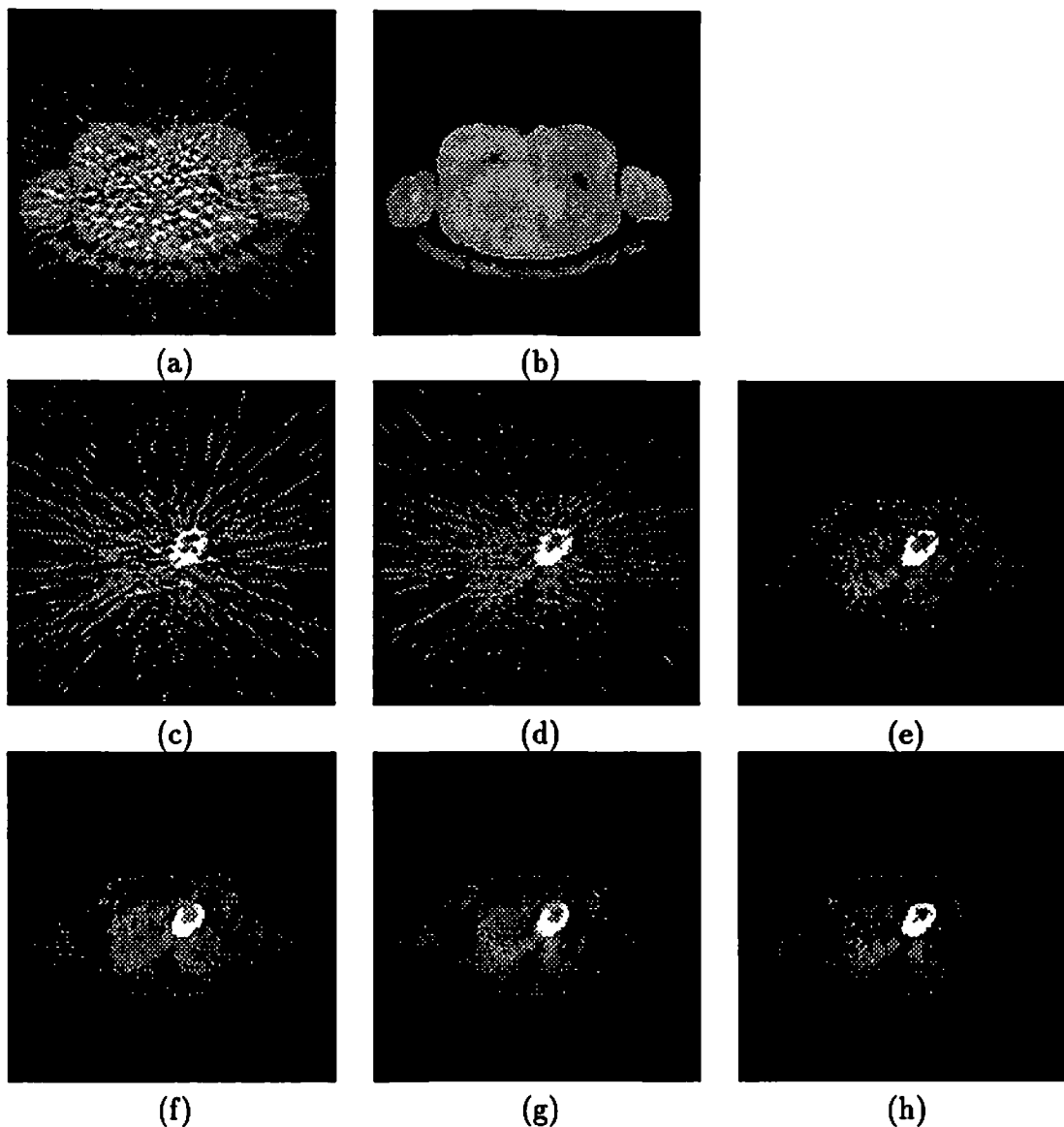


Figure 5: Transaxial slices from the whole body study: (a) FBP reconstruction of transmission data using the smoothed blank/transmission sinogram; (b) MAP reconstruction from the transmission sinogram; (c) FBP reconstruction of the emission image using smoothed blank/transmission attenuation correction factors; (d) FBP reconstruction of the emission image using reprojections of the MAP transmission image; (e) ML reconstruction of the emission image using reprojection of the MAP transmission reconstruction; MAP reconstructions of the emission image using reprojection of the MAP transmission reconstruction using different priors: (f) 2D Geman McClure prior; (g) 3D Geman McClure prior; (h) 3D Continuation method.

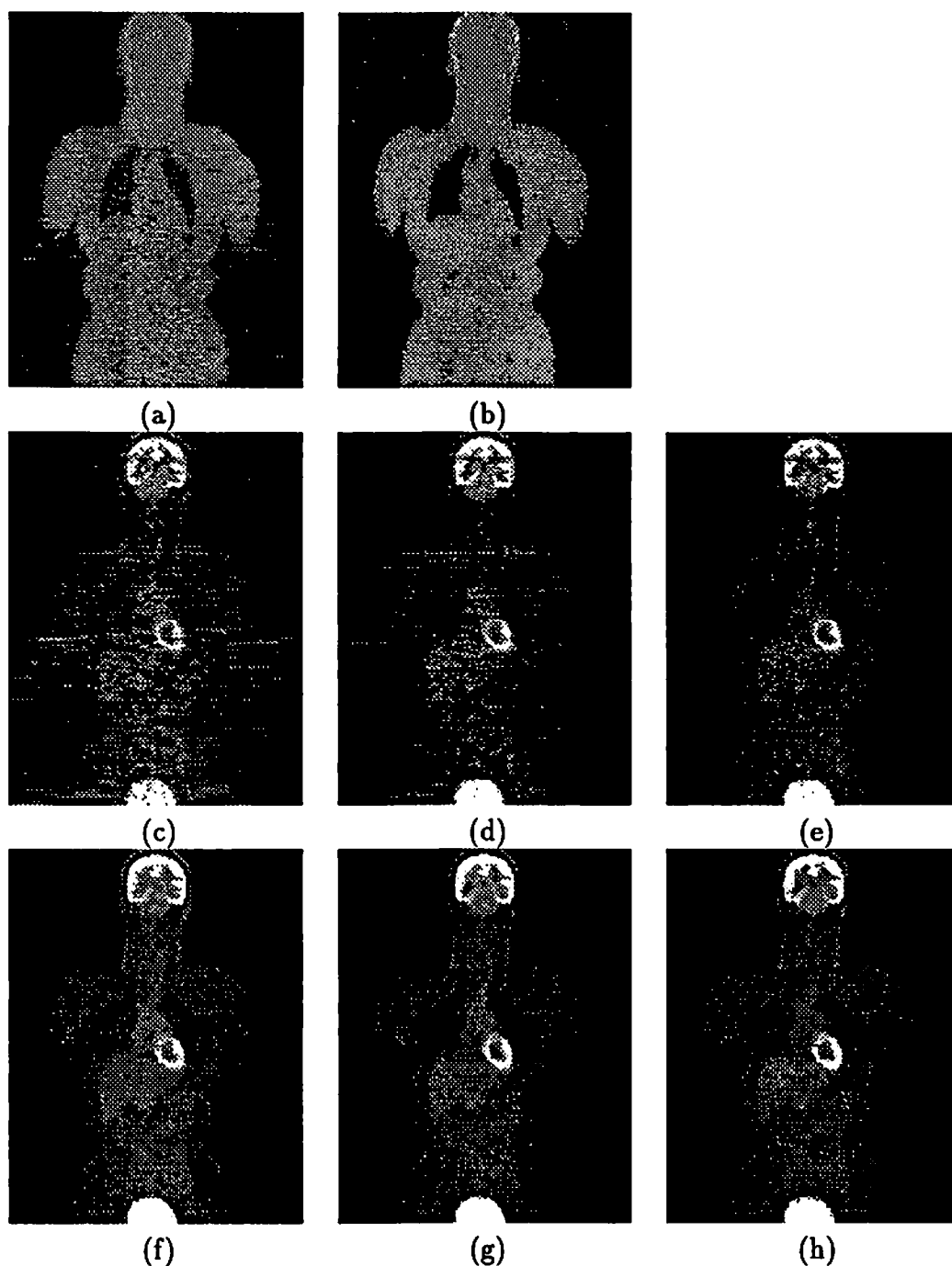


Figure 6: Coronal images from the whole body study: (a) FBP reconstruction of transmission data using the smoothed blank/transmission sinogram; (b) MAP reconstruction from the transmission sinogram; (c) FBP reconstruction of the emission image using smoothed blank/transmission attenuation correction factors; (d) FBP reconstruction of the emission image using reprojections of the MAP transmission image; (e) ML reconstruction of the emission image using reprojection of the MAP transmission reconstruction; MAP reconstructions of the emission image using reprojection of the MAP transmission image using different priors: (f) 2D Geman McClure prior; (g) 3D Geman McClure prior; (h) 3D Continuation method.

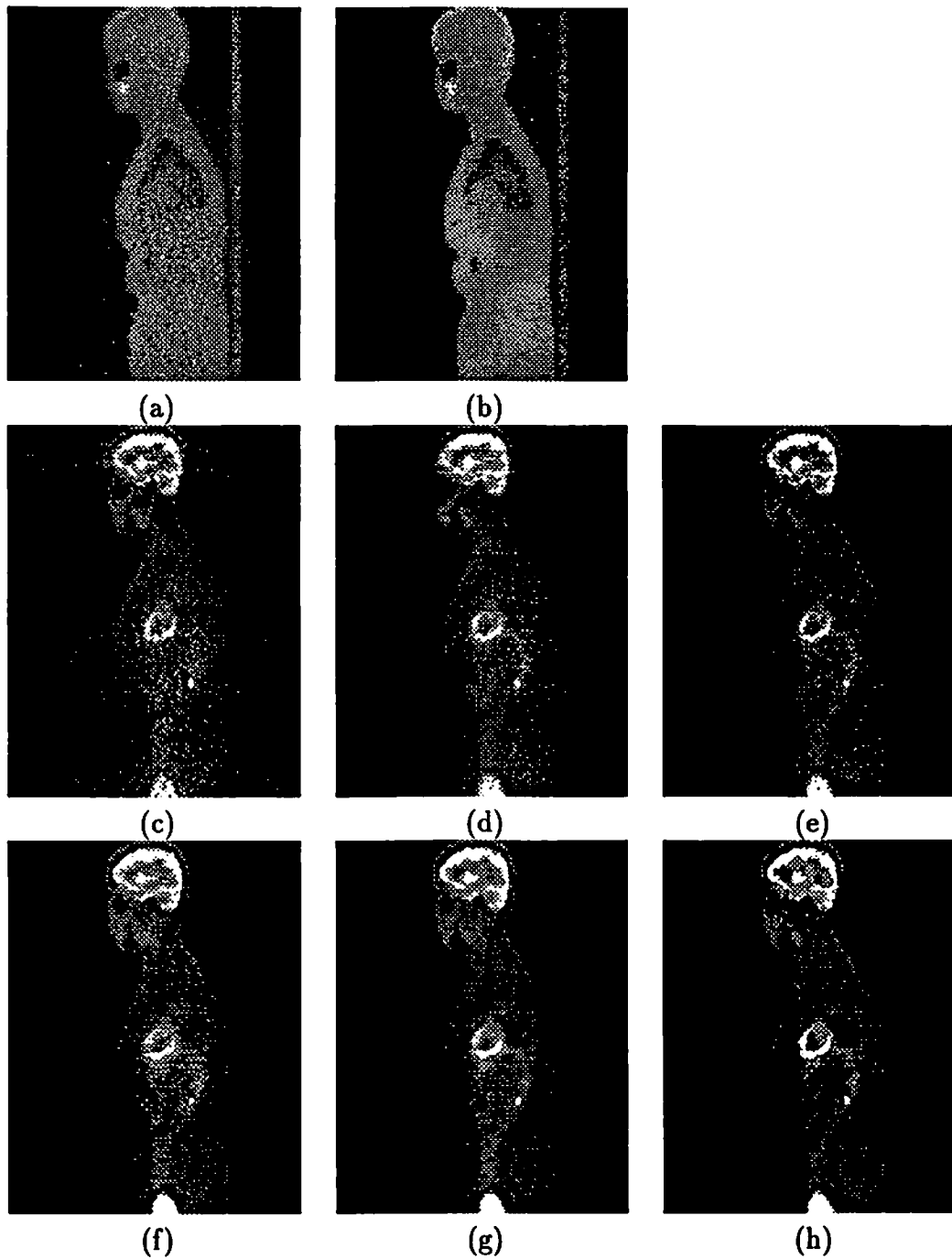


Figure 7: Saggital sections of the whole body study: (a) FBP reconstruction of transmission data using the smoothed blank/transmission sinogram; (b) MAP reconstruction from the transmission sinogram; (c) FBP reconstruction of the emission image using smoothed blank/transmission attenuation correction factors; (d) FBP reconstruction of the emission image using reprojections of the MAP transmission image; (e) ML reconstruction of the emission image using reprojection of the MAP transmission reconstruction; MAP reconstructions of the emission image using reprojection of the MAP transmission image using different priors: (f) 2D Geman McClure prior; (g) 3D Geman McClure prior; (h) 3D Continuation method.