Bayesian Image Reconstruction for Transmission Tomography Using Deterministic Annealing

Ing-Tsung Hsiao[♯], Anand Rangarajan[♭], and Gene Gindi[†]

[♯]School of Medical Technology, Chang Gung University,
Kwei-Shan, Tao-Yuan, 333, Taiwan, R.O.C.

[†]Departments of Radiology and Electrical & Computer Engineering,
SUNY Stony Brook, Stony Brook, NY 11794, USA

[♭]Department of Computer & Information Science and Engineering,
University of Florida, Gainesville, FL 32611, USA

Corresponding Author:

Gene Gindi, Ph.D.
Department of Radiology
SUNY at Stony Brook
Stony Brook, NY 11794
Tel: (631)444-2539
Fax: (631)444-6450
Email: gindi@clio.rad.sunysb.edu

ABSTRACT

We previously introduced a new, effective Bayesian reconstruction method[1] for transmission tomographic reconstruction that is useful in attenuation correction in SPECT and PET. The Bayesian reconstruction method used a novel object model (prior) in the form of a mixture of gamma distributions. The prior models the object as comprising voxels whose values (attenuation coefficients) cluster into a few classes. This model is particularly applicable to transmission tomography since the attenuation map is usually well-clustered and the approximate values of attenuation coefficients in each anatomical region are known. The reconstruction is implemented as a maximum *a posteriori* (MAP) estimate obtained by iterative maximization of an associated objective function. As with many complex model-based estimations, the objective is nonconcave, and different initial conditions lead to different reconstructions corresponding to different local maxima. To make it more practical, it is important to avoid such dependence on initial conditions. Here, we propose and test a deterministic annealing (DA) procedure for the optimization. Deterministic annealing is designed to seek approximate global maxima to the objective, and thus robustify the problem to initial conditions. We present the Bayesian reconstructions with and without DA and demonstrate the independence of initial conditions when using DA. In addition, we empirically show that DA reconstructions are stable with respect to small measurement changes.

Keywords: Bayesian image reconstruction, attenuation correction, gamma mixture model, deterministic annealing, transmission tomography

1 INTRODUCTION

Transmission tomography is perhaps most familiar in its medical guise as CT (computed tomography). A CT image is a map of attenuation coefficients and is computed by a tomographic reconstruction procedure. Because of the relatively high dose in conventional CT, the measured projection data contain little noise, and the subsequent reconstruction from projections can often be obtained with deterministic methods, such as modifications of the familiar filtered backprojection algorithm.[2] In other medical and non-medical applications, however, the measured projection data are photon limited and the imaging geometry is complex. Here, model-based statistical reconstruction methods find application.

Our own interest in transmission tomography[3,4] has been in its application to the nuclear-medical modalities of PET and SPECT. In these two forms of emission tomography, knowledge of the patient's *transmission* image is needed to perform an attenuation correction on the *emission* reconstruction.[5] In many imaging scenarios, particularly in imaging of chest regions, the lack of an effective attenuation correction can lead to rather serious image artifacts and inaccuracies in the emission reconstruction. Therefore, knowledge of an accurate attenuation map is needed.

To measure transmission data in this setting, radioactive sources are mounted within the PET or SPECT camera, and the emitted gamma rays passed through the patient to thus measure projection data suitable for subsequent 2D transmission reconstructions. Because of dose and instrumental limitations, this data is photon limited and also corrupted by other sources of uncertainty. In PET, an actual reconstruction of the transmission data need not be pursued; attenuation correction factors can be derived from the transmission projection data itself.[6] However, this form of attenuation correction is suboptimal,[7] and for both PET and SPECT, it is more effective to first perform a transmission reconstruction, and use this for attenuation correction.

PET and SPECT attenuation correction is useful in chest and whole-body slices, where the average attenuation coefficients in lung, bone, and soft-tissue regions are quite different. These chest-region attenuation maps are stereotypical, and the topology of the anatomical regions comprising the 2D slice is the same from patient to patient. In addition, much is known *a priori* about the

actual ranges of values of attenuation coefficient for each anatomical region. That is, the histogram of attenuation coefficients is stereotypical. Finally, it is known that the attenuation values vary smoothly, with occasional discontinuities at anatomical borders. With all this prior knowledge, one can pose a variety of geometric or statistical models for use in a model-based transmission reconstruction. Examples have included schemes using deformable contours,[8] and level sets,[9] as well as a variety of edge-preserving smoothness constraints.[10]

In this paper, we focus on developing a model that takes advantage of knowledge of the histogram of attenuation coefficients. Our model is that the histogram is lumped into a few peaks of different location and width, with each peak corresponding to an anatomical region of the image. In statistical terms, we use a mixture model as a prior density on the histogram of attenuation coefficients, and use this prior in a Bayesian context to reconstruct the noisy transmission data. This type of intensity prior has garnered interest well beyond our current application to PET and SPECT attenuation correction. For instance, similar models[11,12] have been used for Bayesian segmentation.

We take a Bayesian MAP (maximum *a posteriori*) approach, in which the reconstruction is computed by maximizing an associated objective function via an iterative algorithm. Since our prior is a mixture model, there are additional prior parameters that have to be estimated as well. Hence, MAP estimation in our case is technically a joint-MAP strategy since the reconstruction *and* prior parameters are to be estimated. While joint-MAP strategies can and indeed often do lead to biased estimators, we have not observed this in our case. A careful treatment of this issue could perhaps be experimentally addressed by analyzing samples from the posterior. This is beyond the scope of the present work. A more important issue is that as with many other Bayesian MAP estimation procedures using complex prior models, the posterior objective is nonconcave, i.e. has many local maxima. Therefore, the reconstruction can depend on initial parameter settings and the initial object estimate. (Here, we term the true attenuation map the "object" and its reconstruction the "object estimate".) That the reconstruction depends on initial settings is a critical failure of this and many similar model-based estimation techniques. In this paper, we adopt a methodology, deterministic annealing, for avoiding excessive dependence on initial conditions.[13–15] We show that

with proper use of such an annealing scheme, we can indeed robustify the problem to variation in initial conditions and compute an effective object estimate.

When standard gradient descent-based methods are used on these nonconcave objective functions, the results usually depend on the choice of initial conditions. Typically, there are no reasonable alternatives to gradient-based procedures. For example, while simulated annealing and algorithms based on evolutionary programming avoid poor local maxima, they are still too slow for most problems (when implemented on conventional workstations). In these situations, deterministic annealing is an attractive choice since, as we will show, we can continue to use gradient-based procedures while avoiding poor local maxima.

Deterministic annealing methods index an objective function by a computational temperature. When deterministic annealing objective functions are carefully designed, at high temperatures the objective function is concave or almost concave. Consequently, there is just a single maximum which is easily attained. Then, as the temperature is reduced, the maximum from higher temperatures is tracked. This is achieved by using the solution at one temperature as the initial condition at the next, lower temperature. In this way, the DA algorithm becomes independent of initial conditions. Despite this, the original nonconcave MAP objective function is approached at lower temperatures.[13]

In Sec.2, we describe the transmission tomography problem. In Sec.3 and Sec.4, we develop a model-based MAP reconstruction method based on mixture model priors, and in Sec.5 discuss how to robustify the MAP optimization using deterministic annealing. In Sec.6, we demonstrate that deterministic annealing indeed leads to a reconstruction that is stable and independent of initial estimates.

Our emphasis in this paper is a demonstration of the effectiveness of DA on a complicated model-based optimization. A few necessary mathematical derivations leading up to our DA objective functions are complicated, and we will, on three particular instances, simply cite crucial results without proof rather than include an intractably lengthy mathematical development.

2 TRANSMISSION TOMOGRAPHY

In developing our models, we will denote the object (attenuation map) by vector $\boldsymbol{\mu}$ ($\boldsymbol{\mu} = \{\mu_n; n = 1, \ldots, N\}$) where μ_n is the attenuation coefficient at voxel n. (Note that attenuation coefficients depend on the energy of the irradiating beam, and here we implicitly assume a monoenergetic source.) The projection data, or sinogram, are denoted by vector $\mathbf{g}$ ($\mathbf{g} = \{g_m; m = 1, \ldots, M\}$) where g_m is the number of photons sensed at detector element m. Note that $M \approx N$ for real systems. We shall also denote the reconstructed estimate at iteration k as vector $\hat{\boldsymbol{\mu}}^k$ with components $\hat{\mu}_n^k$.
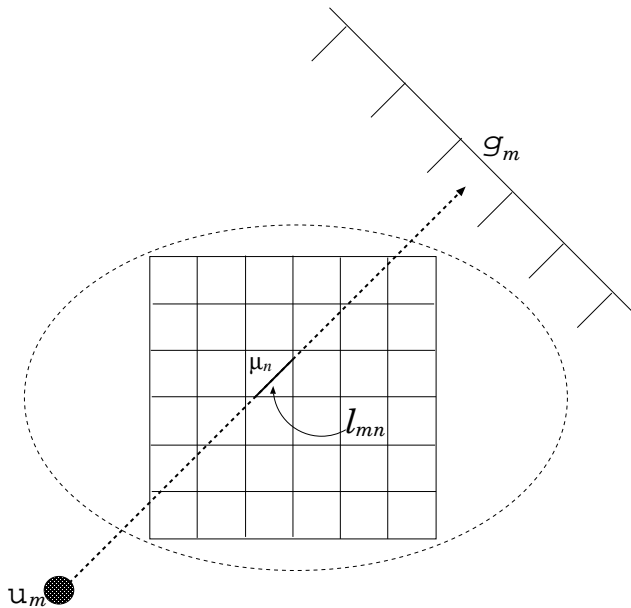
Figure 1. Geometry for transmission tomography with u_m, the external source strength along ray m, μ_n the linear attenuation coefficient at pixel n, l_{mn} the chord length through pixel n, and g_m the counts detected at detector m.

Figure 1 summarizes the data acquisition. Radiation from an external source (e.g. a radioactive line source in transmission SPECT or PET) passes through the object and is attenuated, with the diminished beam arriving at detector element m. We may index this ray by m, and designate the source strength at this position by u_m, an element of the vector $\mathbf{u}$. Figure 1 shows one ray, but the source can be translated and gantry rotated to generate many (M) such rays. With no attenuating object in the scanner, the detector readings record a "blank scan" u_m. The u_m obey Poisson statistics, but the number of counts is so high that the noisy measure of u_m is considered equal

to its underlying mean. Thus u_m is considered a known constant in our mathematical expressions. With an object $\boldsymbol{\mu}$, photons are scattered or absorbed,[16] and the number of photons reaching g_m follows a Poisson distribution[16] with mean given by

$$\bar{g}_m = u_m e^{-\sum_n l_{mn}\mu_n}. \tag{1}$$

Here l_{mn} is the path length of ray m as it traverses voxel n, and the exponential factor reflects a Beer's law attenuation. Note also that the form of Eq.(1) flexibly models a variety of imaging geometries, and makes no restriction to parallel beam geometries. In our simulations, however, we shall use a series of parallel beams at equispaced angles.

Since $\mathbf{g}$ is independent Poisson at each m, it follows a probability law

$$p(\mathbf{g}|\boldsymbol{\mu}) = \prod_{m=1}^{M} \frac{\bar{g}_m e^{-\bar{g}_m}}{g_m!} \tag{2}$$

The reconstruction problem is then to obtain an estimate $\hat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$ given $\mathbf{g}$. Note that our imaging model, Eq.(1), models only photon noise. For SPECT and PET attenuation correction, other terms reflecting randoms, scatter, and crosstalk effects must be appended to the model.[17] However, as a generic problem in photon-limited transmission tomography, Eq.(1) captures the essence of the imaging model.

3 THEORY

We take a Bayesian approach to reconstruction, and in our context Bayes' theorem becomes

$$p(\boldsymbol{\mu}, \boldsymbol{\psi}|\mathbf{g}) = \frac{p(\mathbf{g}|\boldsymbol{\mu})p(\boldsymbol{\mu}|\boldsymbol{\psi})p(\boldsymbol{\psi})}{p(\mathbf{g})} \tag{3}$$

where $p(\boldsymbol{\mu}, \boldsymbol{\psi}|\mathbf{g})$ is the joint posterior, and $p(\mathbf{g}|\boldsymbol{\mu})$, the likelihood contains a forward model of the imaging system. The term $p(\boldsymbol{\mu}|\boldsymbol{\psi})$ is the prior, and depends on as yet unspecified parameters $\boldsymbol{\psi}$. The term $p(\boldsymbol{\psi})$ is a hyperprior on $\boldsymbol{\psi}$. Note [from Eq.(3)] that $\mathbf{g}$ is conditionally independent of any model parameters $\boldsymbol{\psi}$. The parameter vector $\boldsymbol{\psi}$ shall assume different identities as we develop

towards our final model, and we shall be clear to track its identities. By taking the log of Eq.(3), and invoking the MAP principle, we may state the reconstruction in terms of a maximization

$$
\begin{aligned}
\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\psi}} &= \underset{\boldsymbol{\mu}, \boldsymbol{\psi}}{\arg\max} \, \Phi(\boldsymbol{\mu}, \boldsymbol{\psi}|\mathbf{g}) \\
&= \underset{\boldsymbol{\mu}, \boldsymbol{\psi}}{\arg\max} \{ \Phi_L(\mathbf{g}|\boldsymbol{\mu}) + \Phi_P(\boldsymbol{\mu}|\boldsymbol{\psi}) + \Phi_{HP}(\boldsymbol{\psi}) \}
\end{aligned} \tag{4}
$$

where Φ is the overall posterior objective and Φ_L, Φ_P, Φ_{HP} are objective functions corresponding to the logarithm of the likelihood, prior, and hyperprior terms, respectively. (The arguments in each objective follow the notational conventions of the corresponding probability expression. Hence $p(\mathbf{g}|\boldsymbol{\mu}) \rightarrow \Phi_L(\mathbf{g}|\boldsymbol{\mu})$ for example.) From Eq.(2), we immediately have our expression for $\Phi_L(\mathbf{g}|\boldsymbol{\mu})$

$$
\Phi_L(\mathbf{g}|\boldsymbol{\mu}) = \sum_m \{ g_m \log(\bar{g}_m) - \bar{g}_m \} \tag{5}
$$

where we have dropped terms independent of $\boldsymbol{\mu}$. At this point, we assume a uniform hyperprior and take $\Phi_{HP} = $ constant, thus eliminating Φ_{HP} from the optimization. In the ensuing development, we first consider a provisional model-based prior before arriving at our final, proposed model. We then focus on the difficult task of optimizing this final objective.

3.1 A Provisional Model: Regularized Likelihood with Independent Gamma Prior

Our provisional model is the independent gamma prior[18] which takes the form

$$
p(\boldsymbol{\mu}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{n=1}^{N} \frac{\alpha_n^{\alpha_n}}{\beta_n^{\alpha_n} \Gamma(\alpha_n)} \mu_n^{\alpha_n - 1} e^{-\frac{\alpha_n}{\beta_n} \mu_n} \tag{6}
$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are N-dimensional vectors with components α_n, β_n at each pixel n. Here, $\Gamma(x)$ is a gamma function. Also, $\alpha_n > 1$. At each pixel n, the gamma pdf (probability density function) controls the value (mode of pdf) towards which μ_n is attracted, as well as the "strength" of that attraction. It does this through its mode, $\beta_n(1 - 1/\alpha_n)$, and variance, β_n^2/α_n. The gamma density $\rightarrow 0$ as $\boldsymbol{\mu} \rightarrow 0$, so positivity is preserved. Thus, if one somehow knew the true value of μ_n (governed by β_n) to within an uncertainty (governed by β_n^2/α_n), one could apply such a prior at each point

n. But possession of this type of knowledge is unreasonable to expect. Nevertheless, this gamma prior serves as a crutch to explain our subsequent final prior.

In terms of Bayes theorem Eq(3), the independent gamma prior has $\boldsymbol{\psi} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$. Using Eq.(5) and taking the log of Eq.(6), the objective using the gamma prior becomes

$$\Phi(\boldsymbol{\mu}) = \Phi_L(\mathbf{g}|\boldsymbol{\mu}) + \Phi_P(\boldsymbol{\mu}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_m \{g_m \log(\bar{g}_m) - \bar{g}_m\} + \sum_{n=1}^N \{(\alpha_n - 1) \log \mu_n - \frac{\alpha_n}{\beta_n} \mu_n\} \qquad (7)$$

One can then maximize $\Phi(\boldsymbol{\mu})$ w.r.t. $\boldsymbol{\mu}$ to obtain a MAP reconstruction $\hat{\boldsymbol{\mu}}$ by applying any suitable optimization algorithm to this very tractable objective,[18] which is concave and has positivity imposed directly by the gamma prior.

We will later show that optimization using our final, model based, prior involves optimizing an objective of the *form* in Eq.(7). However, as is, we cannot use Eq.(7). It requires specification of the $2N$ parameters, α_n, β_n, $n = 1, .., N$. With this strategy, one would have to thus specify $3N \gg M$ parameters (N for $\boldsymbol{\mu}$, N for $\boldsymbol{\alpha}$, N for $\boldsymbol{\beta}$) given the M measurements $\mathbf{g}$. Instead, to retain the nice properties of the gamma prior, we propose a new scheme to greatly reduce the number of parameters and simultaneously estimate them.

3.2 A New Prior: Gamma Mixture Model

Our extension of the independent gamma prior is to a gamma mixture model. Assume we are given a set of independent observations $\boldsymbol{\mu} = \{\mu_1, ..., \mu_N\}$ and it obeys a mixture model[19] of L component densities. Then it has the mixture density function

$$p(\boldsymbol{\mu}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi}) = \prod_{j=1}^N \sum_{a=1}^L \pi_a p(\mu_n|\alpha_a, \beta_a) \qquad (8)$$

where the parameter vector $\boldsymbol{\psi}$ now comprises 3 vectors, $\boldsymbol{\psi} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi}\}$. Equation (8) can be thought of as modeling an intensity histogram of $\boldsymbol{\mu}$ comprising $a = 1, \ldots, L$ (typically $L = 3$ or 4) peaks. Each peak is itself modeled as a gamma density $p(\mu_n|\alpha_a, \beta_a)$ with gamma density parameters α_a, β_a. The "mixing proportions" π_a gauge the area under each peak, and thus $\sum_{a=1}^L \pi_a = 1$ with

$\pi_a > 0$. The vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\pi}$ are now L-element vectors $(\alpha_a, \beta_a, \pi_a; a = 1, \ldots, L)$ and not N-element vectors.

The motivation for using the mixture model is that the histogram of attenuation coefficients in an area like thorax comprises several peaks, each identified with a particular tissue type (indexed by a), e.g. soft tissue, lung and bone in a SPECT attenuation correction problem. A finite mixture model can account for this multimodal distribution. For the mixture model, we now need estimate at most $3L$ additional parameters rather than $2N$. Since L is typically less than 5, our predicament is improved. With so few parameters, we may consider estimating them directly from the data via a joint MAP scheme.

By taking the log of Eq.(8), we obtain an objective function Φ_P for our new mixture-model prior:

$$
\begin{aligned}
\Phi_P(\boldsymbol{\mu}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi}) &= \log p(\boldsymbol{\mu}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi}) \\
&= \sum_n \log \left(\sum_{a=1}^{L} \pi_a p(\mu_n | \alpha_a, \beta_a) \right)
\end{aligned}
\tag{9}
$$

This function is nonconcave with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\psi}$, and using Eq.(9) in the optimization Eq.(4) would lead to a difficult optimization with many local maxima. While the single gamma prior in Eq.(7) leads to a concave objective, we have a more expressive mixture-model prior in Eq.(9) which in turn leads to a difficult to optimize nonconcave objective. The main source of nonconcavity is the summation over the different mixture components [denoted by the index a in Eq.(9)].

At this point, we shall consider α_a to be a user specified parameter and not consider its estimation. This parameter will have the property of controlling smoothness in the reconstruction. Thus $\boldsymbol{\psi}$ *now* assumes the identity $(\boldsymbol{\pi}, \boldsymbol{\beta} = \{\pi_a, \beta_a; a = 1, \ldots, L\})$.

3.3 Joint MAP Reconstruction via An Alternating Algorithm

We now consider the optimization Eq.(4) in more detail. Our strategy is to apply an iterative (with iteration index k) alternating ascent to Eq.(4), first updating $\hat{\boldsymbol{\mu}}$ while holding $\hat{\boldsymbol{\psi}}^{k-1}$ fixed, then

updating $\hat{\boldsymbol{\psi}}$ while holding $\hat{\boldsymbol{\mu}}^k$ fixed. The resulting alternation, a form of grouped coordinate ascent, becomes

$$\hat{\boldsymbol{\mu}}^k = \arg\max_{\boldsymbol{\mu}}\{\Phi_L(\mathbf{g}|\boldsymbol{\mu}) + \Phi_P(\boldsymbol{\mu}|\hat{\boldsymbol{\psi}}^{k-1})\} \tag{10}$$

$$\hat{\boldsymbol{\psi}}^k = \arg\max_{\boldsymbol{\psi}}\{\Phi_P(\hat{\boldsymbol{\mu}}^k|\boldsymbol{\psi})\} \tag{11}$$

with initial estimate $\hat{\boldsymbol{\mu}}^0$

Examination of Eq.(10) shows that it is a regularized reconstruction, albeit it with the nonconcave prior given in Eq.(9). Equation (11) fits parameters to a mixture model, and so is a mixture decomposition.[19] The alternation can be summarized by the diagram in Fig.2, illustrating two separate optimization problems.
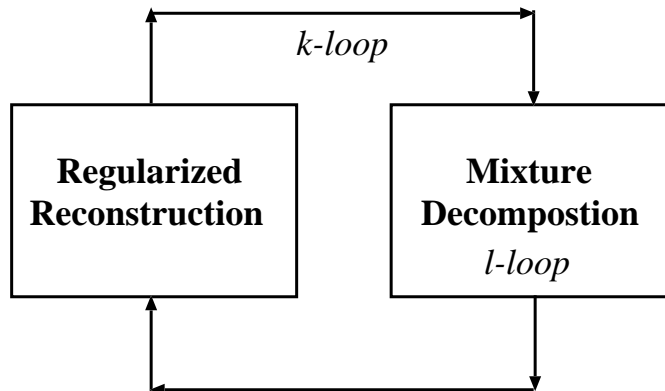
Figure 2. The alternating algorithm can be summarized in this diagram. The left side is a conventional regularized reconstruction with gamma mixture priors, and the parameters are provided by the right side. Given the reconstruction from the left side, then the right side becomes a mixture decomposition.

If Eqs.(10)(11) were easy to maximize and insensitive to initial conditions, then we would be done. However, the mixture decomposition Eq.(11) is highly nonconcave, and unfavorable maxima using Eq.(11) lead to bad reconstructions $\hat{\boldsymbol{\mu}}$. We now recast Eqs.(10)(11) in form amenable to deterministic annealing.

3.4 A Modified Objective for Mixture Decomposition

We now focus specifically on the mixture decomposition, Eq.(11), i.e. the optimization on the right side of Fig.2. There are many methods available for estimating the mixture parameters, and ML estimation via the EM (expectation-maximization) algorithm is a popular one.[20] The EM algorithm for mixture problems can be shown to be equivalent[21] to a method of coordinate ascent on a particular objective that replaces Eq.(9) and leads to mathematical conveniences. We state this new mixture decomposition objective, which we term Φ_P^{mix}, without development or proof, relying on our reference[21] for details.

First, let's define an $N \times L$ complete data $\mathbf{z}$ with element z_{an}, where $\sum_a z_{an} = 1$ and $0 \leq z_{an} \leq 1$, Here, z_{an} is analog and indicates a degree of membership of pixel n in class a. The objective function for mixture decomposition is then,[21]

$$\Phi_P^{mix}(\boldsymbol{\mu}|\mathbf{z}, \boldsymbol{\pi}, \boldsymbol{\beta}) = -\left[\sum_n \sum_a \{z_{an} \log z_{an} + z_{an} \log \frac{1}{\pi_a p(\mu_n|\alpha_a, \beta_a)}\}\right] - \eta(\sum_a \pi_a - 1) + \sum_n \kappa_n(\sum_a z_{an} - 1), \quad (12)$$

where Lagrange multipliers η and κ_n are used in imposing the constraints $\sum_a \pi_a = 1$ and $\sum_a z_{an} = 1$. In Eq.(12), the term $-\sum_n \sum_a z_{an} \log z_{an}$ is an entropy term. Consequently, Eq.(12) can be interpreted as *maximizing* the entropy of $\mathbf{z}$ to the extent possible while simultaneously trying to maximize the second term which depends on the data. The mixture parameters are estimated by optimizing Eq.(12), so that Eq.(11) in our joint MAP scheme is transformed to

$$\hat{\mathbf{z}}, \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\beta}} = \arg \max_{\mathbf{z}, \boldsymbol{\pi}, \boldsymbol{\beta}} \{\Phi_P^{mix}(\boldsymbol{\mu}|\mathbf{z}, \boldsymbol{\pi}, \boldsymbol{\beta})\}. \quad (13)$$

Note that, so far, $\mathbf{z}$ in Eq.(12) is simply a mathematical convenience[21] leading to the easily implemented Eqs.(14)-(16) below. The optimization in Eq.(13) will not in itself solve the local maximum problem associated with mixture decomposition.

It is possible to solve Eq.(13) through a series of iterative updates. If we index this iteration by l, then iteration over l is a sub-iteration of the main loop indexed by k. (This is indicated in Fig.2.) First, by optimizing the unconstrained objective function Eq.(12) w.r.t z_{an} and the Lagrange

multipliers κ_n, one can obtain a closed-form update for $\hat{z}_{an}$. Then by setting the first derivatives w.r.t π_a and the Lagrange multiplier η to zero in Eq.(12), one can also get a closed-form solution for $\hat{\pi}_a$. The update equation for $\hat{\beta}_a$ is derived by setting the first derivatives of the objective function w.r.t. β_a to zero. The update equations for mixture decomposition then become

$$\hat{z}_{an}^{k,l} = \frac{\hat{\pi}_a^{k,l-1} p(\mu_n|\alpha_a, \hat{\beta}_a^{k,l-1})}{\sum_b \hat{\pi}_b^l p(\mu_n|\alpha_b, \hat{\beta}_b^{k,l-1})} \tag{14}$$

$$\hat{\pi}_a^{k,l} = \frac{1}{N} \sum_n \hat{z}_{an}^{k,l} \tag{15}$$

$$\hat{\beta}_a^{k,l} = \frac{\sum_n \hat{z}_{an}^{k,l} \mu_n}{\sum_n \hat{z}_{an}^{k,l}}. \tag{16}$$

In Eqs.(14)-(16), we use two superscripts k, l to index the 2 levels of iteration. Note that these update equations are also derivable using an EM approach for mixture decomposition.[21] Note also that in the first (k=0) pass through the mixture decomposition, we need initial conditions for parameters, $\hat{\pi}_a^{0,0}$, $\hat{\beta}_a^{0,0}$. Once the iterations over l converge, we then increment $k \to k+1$.

3.5 Reconstruction Algorithm without Annealing

In Eq.(11), we have replaced the Φ_P of Eq.(9) with the modified objective Φ_P^{mix} of Eq.(12). This change preserves local maxima, and leads to the rapidly computed updates Eqs.(14)(15)(16). Remarkably, it is also possible to replace Φ_P in Eq.(10) by Φ_P^{mix}, again preserving all fixed points. We state this fact without proof, but arguments in[22] motivate this replacement. With this replacement, the alternation Eq.(10) $\leftrightarrow$ Eq.(11) is transformed to

$$\hat{\boldsymbol{\mu}}^k = \arg\max_{\boldsymbol{\mu}} \{\Phi_L(\mathbf{g}|\boldsymbol{\mu}) + \Phi_P^{mix}(\boldsymbol{\mu}|\hat{\boldsymbol{\pi}}^{k-1}, \hat{\boldsymbol{\beta}}^{k-1}, \hat{\mathbf{z}}^{k-1})\} \tag{17}$$

$$(\hat{\boldsymbol{\pi}}^k, \hat{\mathbf{z}}^k, \hat{\boldsymbol{\beta}}^k) = \arg\max_{\boldsymbol{\pi}\mathbf{z}\boldsymbol{\beta}} \Phi_P^{mix}(\hat{\boldsymbol{\mu}}^k|\boldsymbol{\pi}, \boldsymbol{\beta}, \mathbf{z}) \tag{18}$$

Eliminating terms independent of $\boldsymbol{\mu}$, the second term in Eq.(17) becomes

$$\Phi_P^{mix}(\boldsymbol{\mu}|\hat{\boldsymbol{\pi}}^k, \hat{\boldsymbol{\beta}}^k, \hat{\mathbf{z}}^k) = \sum_n \sum_a [\hat{z}_{an}^k(\alpha_a - 1)\log\mu_n - \hat{z}_{an}^k \frac{\alpha_a}{\hat{\beta}_a^k}\mu_n]. \tag{19}$$

Remarkably, with this form, the reconstruction side (left-side of Fig.2) of the alternation again assumes the form of a concave and easily optimized objective with the independent gamma prior seen in Eq.(7). From comparison of the prior objective in Eq.(19) to that in Eq.(7), we can identify the parameters of the effective pointwise gamma priors α_n β_n as a z-weighted combination of mixture class parameters, α_a, β_a:

$$\alpha_n - 1 \;=\; \sum_a \hat{z}_{an}(\alpha_a - 1) \tag{20}$$

$$\frac{\alpha_n}{\beta_n} \;=\; \sum_a \hat{z}_{an}\frac{\alpha_a}{\hat{\beta}_a}. \tag{21}$$

In essence, we now have a procedure for specifying the original pointwise parameters α_n, β_n in terms of class parameters α_a, β_a, and we have an easy concave maximization for Eq.(17). But we still have the problem of local maxima.

We may now collect our results and state our non-annealing version of the reconstruction algorithm. We refer to this version by the ungainly but useful title of "non-DA algorithm".

(a) State initial conditions $\hat{\boldsymbol{\mu}}^0$, $\hat{\boldsymbol{\beta}}^{0,0}$, $\hat{\boldsymbol{\pi}}^{0,0}$.

(b) Optimize Eq.(18) using Eqs.(14)(15)(16) till convergence.

(c) Optimize Eq.(17) till convergence.

(d) Alternate (b)-(c) till convergence.

The non-DA algorithm is also summarized in the following pseudocode where for clarity, we have suppressed k, l superscripts.

Pseudocode: Non-DA Algorithm

Initial conditions $(\boldsymbol{\mu}, \pi_a, \beta_a)$

Begin A: $k - loop$**.** Do A until $\boldsymbol{\mu}$ converges

$$\boldsymbol{\mu} \leftarrow \arg\max_{\boldsymbol{\mu}} \{\Phi_L(\mathbf{g}|\boldsymbol{\mu}) + \Phi_P^{mix}(\boldsymbol{\mu}|\boldsymbol{\pi}, \boldsymbol{\beta}, \mathbf{z})\}$$

Begin B: $l - loop$**.** Do B until (z_{an}, π_a, β_a) converge

$$z_{an} \leftarrow \frac{\pi_a p(\mu_n|\alpha_a, \beta_a)}{\sum_b \pi_b p(\mu_n|\alpha_b, \beta_b)}$$

$$\pi_a \leftarrow \frac{1}{N} \sum_n z_{an}$$

$$\beta_a \leftarrow \frac{\sum_n z_{an}\mu_n}{\sum_n z_{an}}$$

End B

End A

---

4 RECONSTRUCTION ALGORITHM WITH ANNEALING

Since the overall objective with the gamma mixture model remains nonconcave, it is sensitive to the initial conditions, $\hat{\boldsymbol{\mu}}^0$, $\hat{\boldsymbol{\beta}}^{0,0}$, $\hat{\boldsymbol{\pi}}^{0,0}$. In order to make the method more practical, it is necessary to avoid such dependence on the initial conditions.

Since the mixture decomposition stage (right side of Fig.2) of the alternation contains the nonconcavity, one might initially try using DA only on this part of the optimization. It turns out that the problem of using DA on mixtures has indeed been addressed,[13,14] and we state the result without proof. In terms of our non-DA algorithm, incorporation of DA into the mixture problem is surprisingly simple. The principal change is the introduction of a temperature parameter into the entropy term of the mixture objective. When this is done, Eq.(12) gets replaced by its DA version

$$\Phi_P^{mix} = -\sum_{an} \left[T z_{an} \log z_{an} + z_{an} \log \frac{1}{\pi_a p(\hat{\mu}_n^k|\alpha_a, \beta_a)} \right] - \eta(\sum_a \pi_a - 1) + \sum_n \kappa_n (\sum_a z_{an} - 1). \quad (22)$$

Eq.(22) is very similar to Eq.(12). Just as in Eq.(12), the term $-\sum_n \sum_a z_{an} \log z_{an}$ is an entropy term. However, in Eq.(22), the entropy is modulated by the temperature parameter. At high temperatures, the entropy term dominates forcing close to equal occupancy in all the clusters. As the temperature is lowered, the entropy term is not emphasized as much and the second term which depends on the data now has a greater influence in the overall objective. The optimization is

conducted by gradually lowering T. The DA temperature T starts at a maximum value T_{max} and decreases at a rate of ϵ at each temperature iteration index t. That is, for each iteration t, the DA temperature becomes $T = T_{max} \times \epsilon^t$.

Exactly as in Eqs.(14)(15)(16), we can derive closed-form expressions for $\mathbf{z}$, $\boldsymbol{\pi}$ and $\boldsymbol{\beta}$ for the DA algorithm. When this is done, we get for $\mathbf{z}$

$$\hat{z}_{an}^{k,l} = \frac{[\hat{\pi}_a^{k,l-1} p(\mu_n|\alpha_a, \hat{\beta}_a^{k,l-1})]^{\frac{1}{T}}}{\sum_{b=1}^{L} [\hat{\pi}_b^{k,l-1} p(\mu_n|\alpha_b, \hat{\beta}_b^{k,l-1})]^{\frac{1}{T}}}. \tag{23}$$

The remaining update equations Eqs.(15) and (16) are identical to the previous non-DA development (in Sec.3.4) and will not be repeated.

Examine the update equation in (23). It differs from its counterpart in Eq.(14) via the inverse temperature exponentiation factor appearing in both the numerator and the denominator. When the temperature is very high, the exponentiation factor is small and the z_{an} at voxel n corresponding to the different classes all approach $\frac{1}{L}$. As the temperature is reduced, the exponentiation factor increases causing the z_{an} to approach binary values.[13] Thus as temperature is lowered, the commitment to class assignment increases.
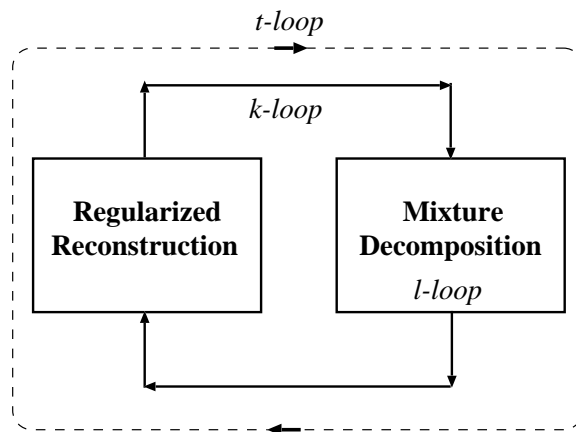
Figure 3. This illustrates the $t - k - l$ loops of the joint MAP reconstruction with DA where t indexes the temperature loop, k indexes the alternating MAP loop, and l indexes the mixture decomposition loop.

If the DA were strictly confined to the mixture decomposition side of the alternation, we would have to repeat the entire annealing sequence each time we entered the mixture phase. To avoid this computational burden, the annealing sequence is transferred to the outermost loop as shown in Fig.3 (compare this to the non-DA version in Fig.2). The outermost $t - loop$ indexes a temperature-lowering (annealing) schedule. The two phases of the alternation - attenuation coefficient estimation and mixture decomposition - are performed at fixed temperature. We have empirically demonstrated that this approach gives us the independence of initial conditions and is not prohibitively computationally expensive. This is shown in the next section.

Our annealing version of the reconstruction is the same as the non-annealing version but for two changes: (1) The $k - l$ loop is run to stability, the temperature ($t - loop$) is updated, and the $k - l$ loop repeated. (2) The update for $\hat{z}_{an}$ Eq.(14) gets replaced by Eq.(23). The DA algorithm is summarized in the following pseudocode 2:

Pseudocode: DA Algorithm

Initial conditions $(\boldsymbol{\mu}, \pi_a, \beta_a)$

$T = T_{max}$

Begin C: $t - loop$. Do C until $\boldsymbol{\mu}$ converges

$T = T_{max} \times \epsilon^t$

    Begin A: $k - loop$. Do A until $\boldsymbol{\mu}$ converges

    $\boldsymbol{\mu} \leftarrow \arg\max_{\boldsymbol{\mu}} \{\Phi_L(\mathbf{g}|\boldsymbol{\mu}) + \Phi_P^{mix}(\boldsymbol{\mu}|\boldsymbol{\pi}, \boldsymbol{\beta}, \mathbf{z})\}$

    Begin B: $l - loop$. Do B until (z_{an}, π_a, β_a) converge

$$z_{an} \leftarrow \frac{[\pi_a p(\mu_n|\alpha_a, \beta_a)]^{\frac{1}{T}}}{\sum_b [\pi_b p(\mu_n|\alpha_b, \beta_b)]^{\frac{1}{T}}}$$

$$\pi_a \leftarrow \frac{1}{N} \sum_n z_{an}$$

$$\beta_a \leftarrow \frac{\sum_n z_{an} \mu_n}{\sum_n z_{an}}$$

    End B

    End A

End C

---

5 SIMULATION RESULTS

The non-DA version of our reconstruction has shown good results for attenuation correction in PET.[3,4] However, these results required careful selection of initial conditions. Here, to make it more practical, we apply the DA procedure to the alternating algorithms to make it robust to initial conditions. In this section, we compare results using the non-DA version with those of the DA version.

5.1 Reconstruction Details

We used the attenuation object as shown in Fig.4, which has two values of narrow-beam attenuation coefficients appropriate for the energies used in PET (511KeV), $\mu = 0.095 cm^{-1}$ for soft tissue and $\mu = 0.035 cm^{-1}$ for lung. The object comprises $N = 128 \times 128$ pixels. We used transmission counts of 500K. The sinogram had dimensions of 129 angles by 192 detector pairs (rays) per angle so that $M = 129 \times 192$.

Figure 4. The attenuation object used in the simulations.

To test different initial conditions (IC's) for joint MAP reconstructions with and without DA, we generated four different initial estimates $\hat{\boldsymbol{\mu}}^0$ from the same noisy sinogram data $\mathbf{g}$. The first initial estimate is an image with a constant value everywhere (row 1, col 1 of Fig.5). The second initial estimate was generated by an EM-ML transmission reconstruction[23] stopped at iteration 10 (row 2, col 1 of Fig.5). A 2-iteration ML transmission reconstruction using a preconditioned conjugate gradient algorithm was used for the third initial estimate (row 3, col 1 of Fig.5), while a filtered backprojection transmission reconstruction[2] with a Hamming filter was produced for the fourth initial estimate (row 4, col 1 of Fig.5).

We then applied the four IC's to the joint MAP reconstructions with and without DA. For the mixture prior, we used $L = 2$ (lung, soft tissue). The values for α_a in (lung, soft tissue) for the case without DA were set at (15, 60), while for the case with DA were (50, 50). The initial values for class means β_a were (0.028, 0.084) for (lung, soft tissue) for both DA and non-DA. Note that here the true class means for (lung, soft tissue) are (0.035, 0.095) at 511KeV. The initial settings of π_a were not especially critical, and were set to a uniform value of $\frac{1}{L}$ for both DA and non-DA.

The maximum temperature of the DA was set at $T_{max} = 500$, and the rate $\epsilon = 0.95$. Note that T_{max} should be set as large as possible and ϵ should be as close to unity as possible. And, please note that annealing takes place in the outermost loop. Within the annealing loop, the main cost of the DA procedure is the left side of the alternation which is mostly the cost of a projection and backprojection. This cost is about the same as one iteration of our EM-ML and non-DA algorithm, while FBP costs 0.5 such iterations. The left side of the alternation is not carried through to completion. In our implementation, the left side is executed just once. In practice, our DA results typically utilized about 8 temperatures (t loop), and a few iterations per temperature (k loop) leading to a typical total number of iterations for DA of about 50–60 iterations. While slow, we observe that much of the reconstruction happens during 2 or 3 temperatures and hence speedups are possible. We are currently exploring this issue. Convergence is established whenever the relative change in the reconstruction estimate $\hat{\boldsymbol{\mu}}^k$ is less than 10^{-8}.
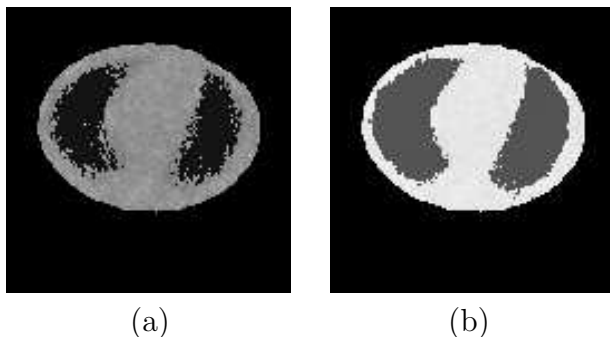
5.2 Results

The anecdotal results for the joint MAP reconstructions for 4 IC's are shown in Fig.5, column 2, without DA, and in Fig.5, column 3, with DA. The reconstructions without DA display 4 different results, and thus indicate high dependence on the IC's. However, the reconstructions with DA illustrate nearly identical results for 4 IC's. Thus DA leads to robust invariance to initial object estimate $\hat{\boldsymbol{\mu}}^0$.

What about robust invariance to initial values of class mean β_a? Here, we used the initial object estimate $\hat{\boldsymbol{\mu}}^0$ obtained from the 2-iteration ML transmission reconstruction, but with a different
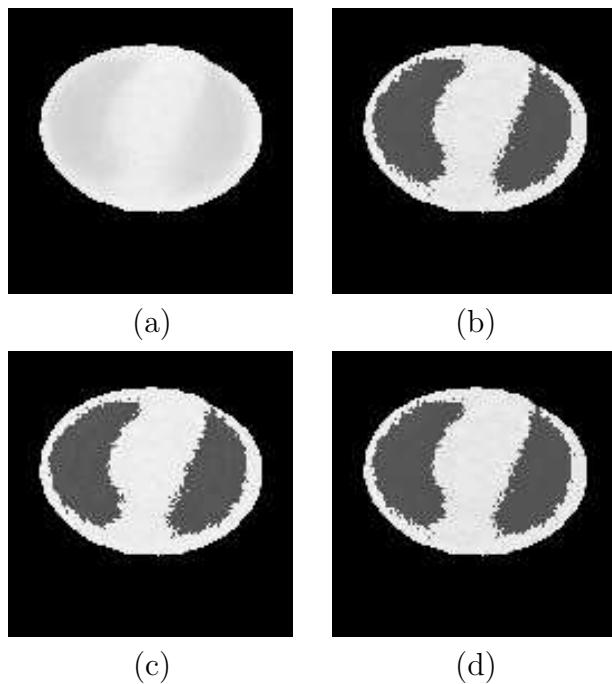
| IC | without DA | with DA |
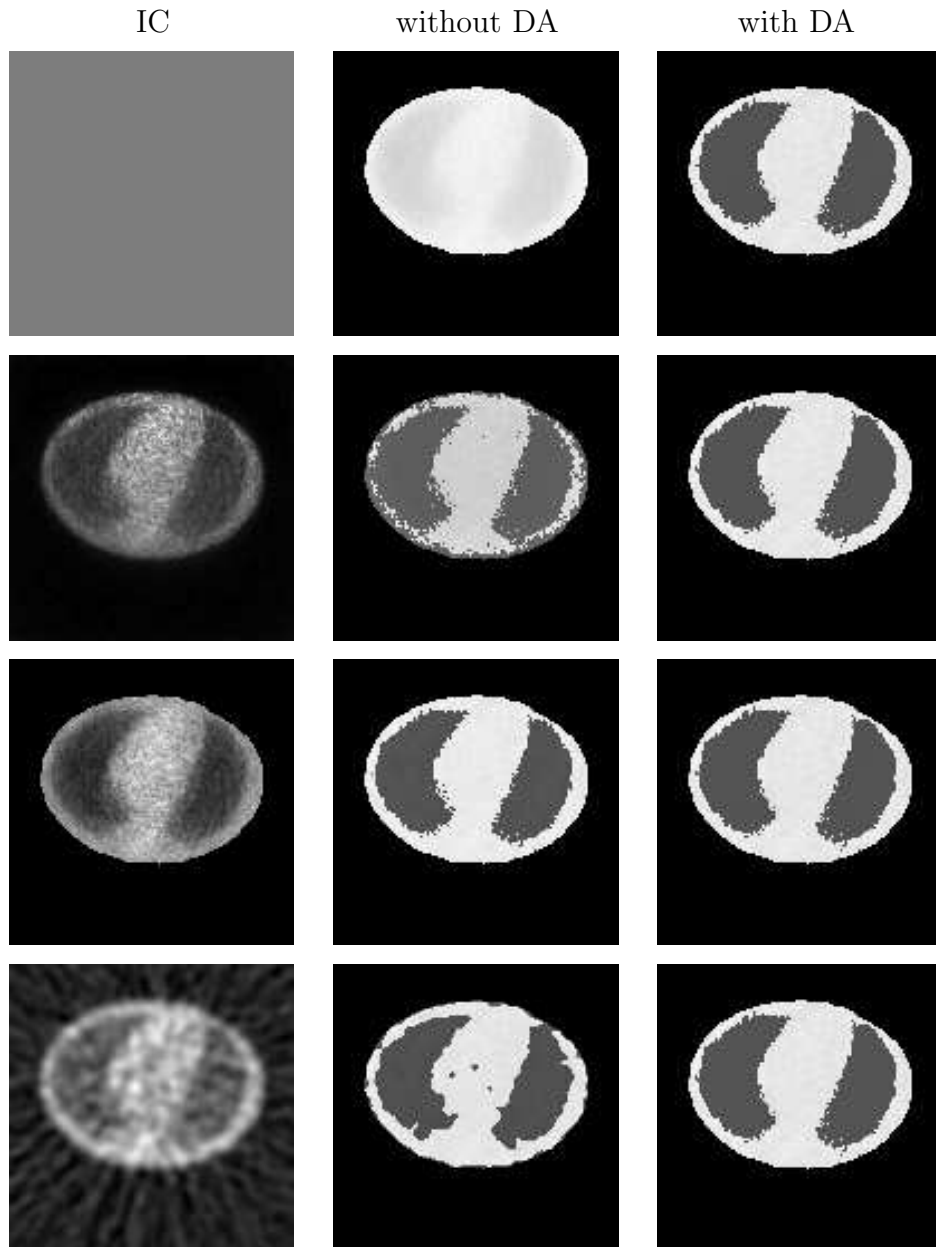
Figure 5. This illustrates the results of joint MAP reconstructions with and without DA for different initial conditions. Column 1 illustrates four different IC's. Columns 2 and 3 indicate joint MAP reconstructions with and without DA. Rows indicate results for different IC's of uniform, 10 iteration of EM-ML, 2 iteration of ML, and FBP transmission reconstructions. The results show that DA is nearly independent of IC's. The third column displays nearly identical reconstructions from DA but still shows small differences especially in the boundary pixels between lung and soft tissue. Without DA, the reconstructions differ considerably.

initial value of (0.056, 0.056) for β_a of (lung, soft tissue). The noisy sinogram **g** remained the same as in fig.5. The results with and without DA are shown in Fig.6. Again, the result with DA in Fig.6(b) shows the robust invariance to initial values of β_a. The reconstruction in Fig.6(b) is nearly identical to those in Fig.5 col 3. However, Fig.6(a) shows that $\hat{\boldsymbol{\mu}}$ has changed considerably with the change in initial β_a, as can be seen by comparing Fig.6(a) with col2, row 3 of Fig.5.

(a)   (b)

Figure 6. This illustrates the results of joint MAP reconstructions (a) without and (b) with DA at different initial value of β_a. (Note: images are displayed at different grey scales due to higher dynamic range of image in (a).

While the above results show an invariance to initial condition for a given set of noisy data, it is also important to demonstrate that DA is stable, i.e. that the final estimate $\hat{\boldsymbol{\mu}}$ changes slowly as the data **g** changes. To do this, we generated another noisy sinogram with the same count level. We then reconstructed the new noisy sinogram without DA and with DA by using two different initial conditions of uniform and 2-iteration ML reconstruction. The values of fixed α_a and initial $\hat{\beta}_a$ are the same as the ones used in Fig.5. The reconstructions on the new noisy sinogram without DA are displayed in Fig.7(a) and (c) for initial estimates of uniform and 2-iteration ML reconstruction, while the results with DA are shown in Fig.7(b) and (d) for the two different initial estimates, respectively. The resemblance of Figs.7(b), (d) to the other DA results demonstrates the stability. This stability is retained when other noisy sinograms with the same count level are reconstructed with and without DA (these results are not shown here).

Figure 7. This figure shows the reconstructions of different noisy projection data without DA using (a) uniform initial condition and (c) 2 iteration ML reconstruction, while the reconstructions with DA are shown in (b) and (d) using initial estimates of uniform and 2-iteration ML reconstruction, respectively.

6 DISCUSSION AND CONCLUSION

In this paper we have proposed a particular type of statistical model - one that captures aspects of the clustered histogram of the object - and used this in the context of transmission reconstruction. Imposing this type of model information leads inevitably to an optimization of a nonconcave objective, so that our MAP solutions are sensitive to initial conditions. The use of a deterministic annealing procedure has been effective in removing this sensitivity, and we consider this fact to be a main result of the paper. We have observed that in a separate emission reconstruction problem involving "mechanical" models of smoothness,[24] a similar sensitivity to initial conditions arises due to a nonconcave objective. The application of a different DA procedure[25] again results in a robustness to initial conditions. It is interesting that DA has been effective in two separate tomographic applications.

In our results, the DA algorithm appears to yield virtually the same solution $\hat{\boldsymbol{\mu}}$ for *any* initial condition. It is possible, however, to get a very different $\hat{\boldsymbol{\mu}}$ with DA for the case where the α_a are set to low (nearly unity) values. However, as $\alpha_a \to 1$, the overall influence of the prior diminishes so that $\hat{\boldsymbol{\mu}}$ approaches a very noisy ML (maximum likelihood) solution. Therefore, it is not surprising that $\hat{\boldsymbol{\mu}}$ varies. Interestingly, eliminating the π_a-update in the DA algorithm helps even in this case. The π_a are adjunct variables in this setup and do not need to be estimated.

While DA allows arbitrary specification of initial object and initial mixture parameters, the reconstruction can be slow if the annealing schedule is not chosen well. Choosing T_{max} too low will result in initial condition dependence, and choosing a low annealing rate ϵ will also result in problems. One can always set T_{max} high and ϵ near to unity, but then the reconstruction is slow to converge. More analysis is required for setting initial temperatures and annealing schedules.

Acknowledgements

References

1. I.-T. Hsiao, A. Rangarajan, and G. Gindi. Joint-MAP Reconstruction/Segmentation for Transmission Tomography Using Mixture-Models as Priors. *Proc. IEEE Nuc. Sci. Symp. and Med. Imag. Conf.*, II:1689–1693, Nov. 1998.

2. A. Rosenfeld and A. C. Kak. *Digital Picture Processing*, volume 1. Academic Press, New York, 2nd edition, 1982.

3. I.-T. Hsiao and G. Gindi. Comparison of Gamma-Regularized Bayesian Reconstruction to Segmentation-Based Reconstruction for Transmission Tomography. *J. Nuclear Medicine*, 40:74P, May 1999.

4. I.-T. Hsiao, W. Wang, and G. Gindi. Performance Comparison of Smoothing and Gamma Priors for Transmission Tomography. In *Proc. IEEE Nuc. Sci. Sym. Med. Imaging Conf.*, volume II, pages 860–864, Oct. 1999.

5. D. L. Bailey. Transmission Scanning in Emission Tomography. *Euro. J. Nuc. Med.*, 25(7):774–787, July 1998.

6. J. M. Ollinger and J. A. Fessler. Positron Emission Tomography. *IEEE Signal Proc. Mag.*, pages 43–55, Jan. 1997.

7. E. U. Mumcuoglu, R. Leahy, S. R. Cherry, and Z. Zhou. Fast Gradient-Based Methods for Bayesian Reconstruction of Transmission and Emission PET Images. *IEEE Trans. Med. Imaging*, 13(4):687–701, Dec. 1994.

8. X. Battle, C. Le Rest, A. Turzo, and Y. Bizais. Three-Dimensional Attenuation Map Reconstruction Using Geometrical Models and Free-Form Deformations [SPECT Application]. *IEEE Trans. Med. Imaging*, 19:404–411, May 2000.

9. D. Yu and J. Fessler. Three-Dimensional Non-Local Edge-Preserving Regularization for PET Transmission Reconstruction. In *Proc. IEEE Nuc. Sci. Sym. Med. Imaging Conf.*, Oct. 2000.

10. C. Bouman and K. Sauer. A Generalized Gaussian Image Model for Edge-Preserving MAP Estimation. *IEEE Trans. Image Proc.*, 2(3):296–310, July 1993.

11. Z. Liang, J. R. MacFall, and D. P. Harrington. Parameter Estimation and Tissue Segmentation from Multispectral MR Images. *IEEE Trans. Med. Imaging*, 13(3):441–449, Aug. 1994.

12. R. Samadani. A Finite Mixture Algorithm for Finding Proportions in SAR Images. *IEEE Trans. Image Processing*, 4(8):1182–1186, Aug. 1995.

13. A. Yuille and J. Kosowsky. Statistical Physics Algorithms That Converge. *Neural Comput.*, 6(3):341–356, Apr. 1994.

14. A. Yuille, P. Stolorz, and J. Utans. Statistical Physics, Mixtures of Distributions and the EM Algorithm. *Neural Comput.*, 6(2):334–340, Mar. 1994.

15. A. Rangarajan, S. Gold, and E. Mjolsness. A Novel Optimizing Network Architecture with Applications. *Neural Comput.*, 8(5):1041–1060, Dec. 1996.

16. H. H. Barrett and W. Swindell. *Radiological Imaging: the Theory of Image Formation, Detection, and Processing*, volume I and II. Academic Press, Inc., Sept. 1981.

17. E. U. Mumcuoglu, R. M Leahy, and S. R. Cherry. Bayesian Reconstruction of PET Images: Methodology and Performance Analysis. *Phys. Med. Bio.*, 41:1777–1807, Sept. 1996.

18. K. Lange, M. Bahn, and R. Little. A Theoretical Study of Some Maximum Likelihood Algorithms for Emission and Transmission Tomography. *IEEE Trans. Med. Imaging*, 6(2):106–114, June 1987.

19. B.S. Everitt and D.J. Hand. *Finite Mixture Distributions*. Chapman and Hall, 1981.

20. A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm. *J. Royal Statist. Soc. B*, 39:1–38, Jan. 1977.

21. R. J. Hathaway. Another Interpretation of the EM Algorithm for Mixture Distributions. *Stat. Prob. Letters*, 4:53–56, Jan. 1986.

22. E. Mjolsness and C. Garrett. Algebraic Transformations of Objective Functions. *Neural Networks*, 3:651–669, Aug. 1990.

23. K. Lange and R. Carson. EM Reconstruction Algorithms for Emission and Transmission Tomography. *J. Comp. Assist. Tomography*, 8(2):306–316, Apr. 1984.

24. S. J. Lee, A. Rangarajan, and G. R. Gindi. Bayesian Image Reconstruction in SPECT Using Higher Order Mechanical Models as Priors. *IEEE Trans. Med. Imaging*, 14:669–680, June 1995.

25. S.-J. Lee. *Bayesian Image Reconstruction in Emission Computed Tomography Using Mechanical Models as Priors*. PhD thesis, Dept of Electrical Engineering, State University of New York at Stony Brook, Stony Brook, New York 11784, USA, Aug. 1995.