Joint-MAP Bayesian Tomographic Reconstruction with a Gamma-Mixture Prior

Ing-Tsung Hsiao , Anand Rangarajan , and Gene Gindi*

EDICS: 3-TOMO

Abstract

We address the problem of Bayesian image reconstruction with a prior that captures the notion of a clustered intensity histogram. The problem is formulated in the framework of a joint-MAP (maximum *a posteriori*) estimation with the prior pdf modeled as a mixture-of-gammas density. This prior pdf has appealing properties, including positivity enforcement. The joint MAP optimization is carried out as an iterative alternating descent wherein a regularized likelihood estimate is followed by a mixture decomposition of the histogram of the current tomographic image estimate. The mixture decomposition step estimates the hyperparameters of the prior pdf. The objective functions associated with the joint MAP estimation are complicated and difficult to optimize, but we show how they may be transformed to allow for much easier optimization while preserving the fixed point of the iterations. We demonstrate the method in the context of medical emission and transmission tomography.

Index Terms

Tomographic reconstruction, joint-MAP estimation, gamma mixture, mixture decomposition.

July 24, 2002

I. Introduction

For many applications, image reconstruction from projections may be posed as a statistical estimation problem in which the object is estimated from noisy observed data. For Bayesian approaches, the data is supplemented by object information in the form of a suitable prior. The reconstruction is usually computed as a MAP (maximum *a posteriori*) estimate via minimization of a suitable objective function. A common form of prior information is that of spatial smoothness or piecewise spatial smoothness, perhaps supplemented by a positivity constraint. Historically, much effort has been devoted to the design of objective functions and associated algorithms to carry out such smoothness-constrained MAP estimation. In this work, we explore the use of a different type of prior information for the reconstruction problem, in which knowledge about the object appears in the form of its clustered intensity histogram. This kind of prior pdf would be applicable when the object voxels tend to cluster into few intensity ranges, i.e. the intensity histogram comprises a few peaks. For the application of medical transmission tomography considered here, this is an apt model. For example in the chest region, x-ray attenuation coefficients (the quality to be estimated) tend to cluster around values appropriate for lung, soft tissue and bone. For our other application, medical emission tomography (PET and SPECT), the assumption is good in certain body regions. For example in the brain, the radionuclide density (the quantity to be estimated) tends to cluster around values appropriate for grey matter, white matter and ventricles. This model is not limited to medical applications; see [19] for example.

For this kind of information, a convenient form for the prior pdf is one of mixture models. Unlike smoothing, no voxel neighbor interactions are assumed; instead, the somewhat weaker assumption of intensity clustering is assumed, and the object is assumed to be positive. This leads to certain computational conveniences, including the easy imposition of positivity. In addition, for the joint-MAP estimation strategy discussed below, the problem of estimating hyperparameters becomes more tractable with this form of prior pdf than with a smoothing prior pdf.

In Section II we discuss a joint-MAP framework within which to conveniently couch this problem, and in Section III we discuss details of our mixture-of gammas prior pdf. The

joint-MAP strategy together with the form of the objectives associated with the prior pdf's leads to an objective that is difficult to minimize, and in Section IV , we demonstrate the joint-MAP reconstruction in the context a series of transformations of this objective that solves these difficulties while preserving the fixed point of the objective. In Section V we present our medical imaging application to transmission and emission tomographic reconstruction. A discussion of pitfalls, extensions, and related work is presented in Section VI.

II. Joint-MAP Framework

We couch our development of an estimation method in terms of a joint-MAP framework, in which the unknown object along with hyperparameters of the prior pdf are jointly estimated. (For a survey of alternative full Bayesian approaches, see [1].)

Let the object to be estimated be represented as a vector $\boldsymbol{\theta}$ of N lexicographically ordered voxels $\{\theta_n; n = 1, \ldots, N\}$. The object could be a 2-D or 3-D image, for instance, whose voxels are lexicographically ordered into a vector with N elements. In our Bayesian approach, $\boldsymbol{\theta}$ is a random vector and θ_n a random variable, but notationally, we shall not distinguish between a random quantity and a particular instance of it. The observed, noisy data from which $\boldsymbol{\theta}$ is to be estimated, is given by vector $\mathbf{g}$, comprising lexicographically ordered elements $\{g_m; m = 1, \ldots, M\}$. For realistic problems, $M \approx N$. The details of the imaging system are contained in the likelihood function $p(\mathbf{g}|\boldsymbol{\theta})$, and in Section V we give detailed expressions for this likelihood for our applications to tomography. (For the tomographic case, the measurement data $\mathbf{g}$ will correspond to the sinogram.) For our Bayesian approach, the prior pdf and hyperprior pdf are given by the pdf's $p(\boldsymbol{\theta}|\boldsymbol{\psi})$ and $p(\boldsymbol{\psi})$, respectively, where $\boldsymbol{\psi}$ is a random vector of hyperparameters. In Section III, we supply detailed expressions for these terms for our particular use of a gamma-mixture prior. With the above definitions, the term $p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{g})$ is then the joint posterior distribution on the object and hyperparameters.

From Bayes' theorem:

$$p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{g}) \propto p(\mathbf{g}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\psi})p(\boldsymbol{\psi}) \tag{1}$$

where in (1) we use the fact that the data $\mathbf{g}$ depends on $\boldsymbol{\theta}$ only. The three terms on the right side of (1) correspond to likelihood, prior pdf, and hyperprior pdf. Our MAP

solution is then

$$(\boldsymbol{\theta}^*, \boldsymbol{\psi}^*) \equiv \arg \max_{\boldsymbol{\theta} \geq 0, \boldsymbol{\psi}} p(\boldsymbol{\theta}, \boldsymbol{\psi} | \mathbf{g}). \tag{2}$$

Our tactic is to use a joint-MAP alternating ascent procedure, updating, at iteration k, the object estimate $\boldsymbol{\theta}$ while holding the most recent estimate of $\boldsymbol{\psi}$ fixed, then updating the parameters $\boldsymbol{\psi}$ given the latest estimate of $\boldsymbol{\theta}$. The development in Sections II, III and IV will bear out our choice of a joint-MAP alternating optimization procedure. With this strategy we can write the general framework of our joint-MAP procedure:

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}^k &= \arg \max_{\boldsymbol{\theta} \geq 0} p(\mathbf{g}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\widehat{\boldsymbol{\psi}}^{k-1}) p(\widehat{\boldsymbol{\psi}}^{k-1}) \\
&= \arg \max_{\boldsymbol{\theta} \geq 0} p(\mathbf{g}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\widehat{\boldsymbol{\psi}}^{k-1}), \tag{3} \\
\widehat{\boldsymbol{\psi}}^k &= \arg \max_{\boldsymbol{\psi}} p(\mathbf{g}|\widehat{\boldsymbol{\theta}}^k) p(\widehat{\boldsymbol{\theta}}^k|\boldsymbol{\psi}) p(\boldsymbol{\psi}) \\
&= \arg \max_{\boldsymbol{\psi}} p(\widehat{\boldsymbol{\theta}}^k|\boldsymbol{\psi}) p(\boldsymbol{\psi}). \tag{4}
\end{aligned}$$

To summarize notation, $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$, the object and hyperparameters, are the random vectors that we seek to estimate. The fixed vector $\mathbf{g}$ is the data, and $\widehat{\boldsymbol{\theta}}^k, \widehat{\boldsymbol{\psi}}^k$ are estimates at iteration k of the joint-MAP procedure.[1]

By simply taking the negative logarithm of (1), (3) and (4), we can restate the problem in terms of objective functions. Bayes' Theorem (1) becomes

$$\Phi(\boldsymbol{\theta}|\mathbf{g}) = \Phi_L(\mathbf{g}|\boldsymbol{\theta}) + \Phi_P(\boldsymbol{\theta}|\boldsymbol{\psi}) + \Phi_{HP}(\boldsymbol{\psi}) \tag{5}$$

where Φ_L, Φ_P and Φ_{HP} are objectives corresponding to negative logarithms of the likelihood pdf, prior pdf, and hyperprior pdf, respectively, and the unsubscripted objective Φ corresponds to the posterior.[2] The overall joint-MAP estimation (2) is now expressed as a minimization

$$(\boldsymbol{\theta}^*, \boldsymbol{\psi}^*) = \arg \min_{\boldsymbol{\theta} \geq 0, \boldsymbol{\psi}} [\Phi_L(\mathbf{g}|\boldsymbol{\theta}) + \Phi_P(\boldsymbol{\theta}|\boldsymbol{\psi}) + \Phi_{HP}(\boldsymbol{\psi})] \tag{6}$$

---

[1]The iterations in (3) and (4) will converge to a local maximum of (2) provided certain mild assumptions regarding the number of local maxima obtain [2]. These assumptions are satisfied except in pathological cases, and we have not observed any convergence problems in our many trials using (3) and (4).

[2]We shall henceforth use th term , "likelihood", "prior", and "hyperprior" to refer to Φ_L, Φ_P and Φ_{HP}, respectively. The associated pdf's will be referred to as "prior pdf", "likelihood pdf", etc.

and the joint-MAP alternating ascent (3) and (4) now becomes an alternating descent:

$$\widehat{\boldsymbol{\theta}}^k = \arg\min_{\boldsymbol{\theta} \geq 0} \left[\Phi_L(\mathbf{g}|\boldsymbol{\theta}) + \Phi_P(\boldsymbol{\theta}|\widehat{\boldsymbol{\psi}}^{k-1}) \right], \qquad (7)$$

$$\widehat{\boldsymbol{\psi}}^k = \arg\min_{\boldsymbol{\psi}} \left[\Phi_P(\widehat{\boldsymbol{\theta}}^k|\boldsymbol{\psi}) + \Phi_{HP}(\boldsymbol{\psi}) \right]. \qquad (8)$$

In (5)-(8), the arguments of the various objectives use the same conventions as the pdf notation. Thus in (7), the objective $\Phi_P(\boldsymbol{\theta}|\widehat{\boldsymbol{\psi}}^{k-1}) \equiv -\log p(\boldsymbol{\theta}|\widehat{\boldsymbol{\psi}}^{k-1})$ is a function of $\boldsymbol{\theta}$, with $\widehat{\boldsymbol{\psi}}^{k-1}$ serving as a parameter, but in (8), $\Phi_P(\widehat{\boldsymbol{\theta}}^k|\boldsymbol{\psi})$ is now a function of $\boldsymbol{\psi}$ with $\widehat{\boldsymbol{\theta}}^k$ serving as a parameter.

Note that (7) is in itself a penalized-likelihood reconstruction that uses the latest estimate of the hyperparameters in the penalty, while (8) is a parameter fit to the latest estimate of the object. The diagram in Fig. 1 summarizes this interpretation of the joint-MAP alternation. In the remainder of the paper, we will refer to the left and right side stages of Fig. 1 as the "reconstruction step" and "hyperparameter estimation step" respectively.

Fig. 1. The alternating algorithm (7) and (8) can be summarized in this diagram. The left side is a conventional penalized likelihood reconstruction using current parameter estimates derived from the right side. The right side is a parameter fit to the current object estimate.

In the ensuing development, we shall, without any loss of generality, temporarily model the hyperprior pdf $p(\boldsymbol{\psi})$ as uniform, thus enabling us to conveniently drop the term Φ_{HP} in the derivations in Sections III and IV. In Sec.VI, we discuss the incorporation of hyperpriors.

III. Gamma-Mixture Prior

With the uniform hyperprior, the right side of the alternation (8) becomes a maximum likelihood (ML) parameter fit of $\boldsymbol{\psi}$ to $\widehat{\boldsymbol{\theta}}^k$. Our gamma-mixture prior pdf will capture the notions of positivity and intensity clustering previously mentioned, but as a useful preliminary step, we first introduce an independent gamma prior pdf at each voxel. This prior pdf is independent, but not identically distributed, at each voxel, and preserves positivity. This provisional prior pdf model (*not our final choice!*) is given by

$$p(\widehat{\boldsymbol{\theta}}|\boldsymbol{\psi}) = p(\widehat{\boldsymbol{\theta}}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{n=1}^{N} p(\widehat{\theta}_n|\alpha_n, \beta_n) \tag{9}$$

where the univariate probability is a gamma prior pdf [3] and is given by

$$p(x|\alpha, \beta) = (\frac{\alpha}{\beta})^\alpha \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\alpha/\beta)x}. \tag{10}$$

This prior pdf is introduced provisionally, because its inadequacies (despite its desireable properties) will motivate the introduction of the mixture-of-gammas prior pdf that is a focus of this paper. In Sec. IV, we will show an interesting relation of our mixture-of-gammas prior pdf to (9).

We note (again) in (9) that we do not distinguish between a random variable and its instance. In (9), $\boldsymbol{\psi}$ now comprises two vectors $\boldsymbol{\alpha}$ ($\{\alpha_n; n = 1, \ldots, N\}$) and $\boldsymbol{\beta}$ ($\{\beta_n; n = 1, \ldots, N\}$) which, in turn, parameterize the gamma density in (10) at each location n. The gamma density (10) is defined [3] only for $x > 0$. In (10), $\beta > 0$ is the mean and β^2/α the variance. We assume $\alpha > 1$, in which case the unique mode is given by $\beta(1 - 1/\alpha)$. This form of prior pdf will attract the value of each voxel $\widehat{\theta}_n$ to its local mode $\beta_n(1 - 1/\alpha_n)$ with a relative "strength" given by the inverse standard deviation $(\beta_n/\sqrt{\alpha_n})^{-1}$. Note that, like other positivity preserving asymmetric prior pdfs, the gamma prior pdf will introduce a slight bias due to the difference between mean and mode. The independent gamma prior pdf was proposed for use in emission tomography in [4]. It is convex, and preserves positivity, but impractical for our purposes. The impracticality stems from the fact that we need estimate $2N$ parameters (N θ_n's and N β_n's) from $M \approx N$ measurements, thus leading to an overfitting problem.

If, however, the object $\boldsymbol{\theta}$ comprises intensity levels clustered about a few values, then the intensity histogram comprises a few (L) peaks where $L << N$. This assumption of

a clustered histogram is, in fact, a much weaker one than that of piecewise smoothness, though piecewise smoothness is indeed consistent with a clustered histogram. Nevertheless, the assumption is a good one for our medical applications and would likely apply to many other applications.

To capture the notion of a clustered (and positive valued) intensity histogram, we consider the general form of a *mixture*-of-gammas prior pdf. The prior follows the general form of a mixture density for L classes [5],

$$p(x|\boldsymbol{\pi}, \boldsymbol{\beta}) = \sum_{a=1}^{L} \pi_a p_a(x|\alpha_a, \beta_a). \tag{11}$$

In (11), the prior pdf takes the form of a weighted sum of component distributions p_a, each of which is itself a gamma density (10). Note that each p_a is parametrized by $\{\beta_a\}$, with the L values of α_a being user specified. The mixing proportions (weights) $\{\pi_a\}$ are positive and satisfy the normalization (simplex constraint)

$$\sum_{a=1}^{L} \pi_a = 1. \tag{12}$$

For future convenience, we can rewrite (11) as

$$p(x|\boldsymbol{\psi}) = \sum_{a=1}^{L} \pi_a p_a(x|\alpha_a, \beta_a) \tag{13}$$

where $\boldsymbol{\psi}$ assumes a *new* identity $\boldsymbol{\psi} \equiv (\boldsymbol{\pi}, \boldsymbol{\beta})$ with $\boldsymbol{\pi} = \{\pi_a; a = 1, \ldots, L\}$ and $\boldsymbol{\beta} = \{\beta_a; a = 1, \ldots, L\}$. We summarize the constraints on $\boldsymbol{\pi}$ and $\boldsymbol{\beta}$ with a constraint set Ω given by

$$\Omega = \{\pi_a, \beta_a; \pi_a \geq 0, \beta_a > 0; a = 1, \ldots, L; \sum_a \pi_a = 1\}. \tag{14}$$

If we now apply our mixture pdf (11) *independently* to each pixel, we obtain the form of the prior pdf that we will use for the remainder of the paper:

$$p(\widehat{\boldsymbol{\theta}}|\boldsymbol{\psi}) = \prod_{n=1}^{N} \sum_{a=1}^{L} \pi_a p_a(\widehat{\theta}_n|\alpha_a, \beta_a). \tag{15}$$

whose associated objective is

$$\Phi_P^{\mathrm{mix}}(\widehat{\boldsymbol{\theta}}|\boldsymbol{\psi}) = -\sum_{n=1}^{N} \log \Big\{ \sum_{a=1}^{L} \pi_a p_a(\widehat{\theta}_n|\alpha_a, \beta_a) \Big\} \tag{16}$$

with superscript "mix" denoting the gamma-mixture form of the Φ_P objective. Thus Φ_P^{mix} is our choice for the prior. Note that since the p_a's are all positivity-preserving gamma prior pdfs in (16), the entire objective Φ_P^{mix} preserves positivity, and hence the optimization over $\boldsymbol{\theta}$ needn't be positivity constrained.

Note that with this change to a mixture prior, we now need estimate only $N + 2L$ (N θ_n's, L π_a's and L β_a's) from $M \approx N$ measurements. In our applications involving 2D image reconstruction, for instance, $N \approx 10^4$ and $L < 10$, so that $N + 2L \approx M$.

We do not attempt to estimate α_a, and leave it as a user-specified parameter. The parameter α in practice controls the smoothness of the reconstruction. As α increases, the amount of noise in the reconstruction decreases, but the bias increases. As $\alpha_a \to \infty$, the reconstruction assumes a piecewise flat appearance. As $\alpha \to 1$, its lower limit, the influence of the prior diminishes, and the reconstruction is noisy but unbiased. Thus α controls a bias-variance tradeoff. A user who desires to optimize some task-dependent image quality will usually have to select some particular bias-variance tradeoff by choosing α.

IV. Transformed Objective Function

With our choice of prior (16) and with a uniform hyperprior, we may write an updated version of our alternation (7) and (8):

$$
\widehat{\boldsymbol{\theta}}^k = \arg \min_{\boldsymbol{\theta}} \Big[\Phi_L(\mathbf{g}|\boldsymbol{\theta}) \\
- \sum_{n=1}^{N} \log \Big\{ \sum_{a=1}^{L} \widehat{\pi}_a^{k-1} p_a(\theta_n|\alpha_a, \widehat{\beta}_a^{k-1}) \Big\} \Big],
\tag{17}
$$

$$
\widehat{\boldsymbol{\psi}}^k = (\widehat{\boldsymbol{\pi}}^k, \widehat{\boldsymbol{\beta}}^k) \\
= \operatorname*{argmin}_{\boldsymbol{\psi} \in \Omega} \Big[- \sum_{n=1}^{N} \log \Big\{ \sum_{a=1}^{L} \pi_a p_a(\widehat{\theta}_n^k|\alpha_a, \beta_a) \Big\} \Big].
\tag{18}
$$

The notation $\boldsymbol{\psi} \in \Omega$ implies that $\boldsymbol{\psi} = (\boldsymbol{\pi}, \boldsymbol{\beta})$ obeys the constraints in (14). Note that the left side (17) of the alternation is a nonconvex, complex and difficult to optimize prior. Note also that the hyperparameter estimation (18) is a mixture decomposition problem for which the well-understood and efficient EM (Expectation-Maximization) algorithm is

available [6] [7] [5] for optimization. In Sec. IV-A directly below, we give a brief overview of the EM-mixture algorithm.

In this section, we address these optimization issues. We will do this by transforming the objectives in (17) and (18), and by using a grouped coordinate descent algorithm on (18) and a conventional preconditioned gradient descent on (17). The actual transformation entails substitution of a new objective Φ_P^{hath} for Φ_P^{mix} in both (17) and (18) so as to preserve the fixpoints of (17) and (18).

A. Transforming the Mixture Decomposition Side (18) of the Alternation

We focus first on (18), the mixture decomposition on the right side of the alternation. We will not use the EM approach, but instead adopt an alternative solution due to Hathaway [8]. The latter approach will be seen to have certain advantages when incorporated into the overall joint-MAP scheme. However, we first very briefly summarize the more familiar EM approach to mixture decomposition, as this will help explain the subsequent introduction of the Hathaway alternative.

The EM method for mixtures [7] makes use of a binary complete data space given by random vector $\mathbf{z}$ whose $N \times L$ elements $\{z_{an};\ a = 1, \ldots, L\ n = 1, \ldots, N\}$ are defined as

$$z_{an} = \begin{cases} 1 & : \quad \text{voxel } n \text{ belongs to class } a \\ 0 & : \quad \text{otherwise} \end{cases} \tag{19}$$

subject to the constraint

$$\sum_a z_{an} = 1 \quad \forall n \tag{20}$$

A given instance of $\mathbf{z}$ thus serves as a segmentation of the image into L intensity clusters. Note that if $\mathbf{z}$ were known, the mixture decomposition solution $\{\widehat{\pi}_a^k, \widehat{\beta}_a^k\}$ would be easy to compute. In the EM approach, one forms the following joint probability [5]

$$\log p(\mathbf{z}, \widehat{\boldsymbol{\theta}}^k | \boldsymbol{\psi}) = \sum_n \sum_a z_{an} \log(\pi_a p_a(\widehat{\theta}_n^k | \alpha_a, \beta_a)) \tag{21}$$

and iterates, over an index l, the following alternating expectation (E) step (22) and maximization (M) steps (23) until convergence:

$$Q(\boldsymbol{\psi} \,|\, \widehat{\boldsymbol{\psi}}^{kl}) \quad \equiv \quad E_{\mathbf{z}}\Big[\log p(\mathbf{z}, \widehat{\boldsymbol{\theta}}^k | \, \boldsymbol{\psi}) | \widehat{\boldsymbol{\theta}}^k, \widehat{\boldsymbol{\psi}}^{kl} \Big], \tag{22}$$

$$\widehat{\boldsymbol{\psi}}^{k,l+1} = \operatorname*{argmax}_{\boldsymbol{\psi} \in \Omega} Q(\boldsymbol{\psi} \,|\, \widehat{\boldsymbol{\psi}}^{kl}). \tag{23}$$

The notation $\widehat{\boldsymbol{\psi}}^{kl}$ indicates the lth estimate of $\widehat{\boldsymbol{\psi}}^{k}$, thus $\widehat{\boldsymbol{\psi}}^{k,l=0} \equiv \widehat{\boldsymbol{\psi}}^{k-1}$ and $\widehat{\boldsymbol{\psi}}^{k}$ is the fixed point of the above iteration. The notation $E_{\mathbf{z}}$ denotes an expectation over $\mathbf{z}$. Note that in our context, the EM alternation (index l) is thus an *inner* loop of the general joint-MAP alternating ascent outer loop indexed by k. The EM algorithm reaches a local minimum of (18) by alternating the E and M steps (22) and (23). Remarkably, the grouped coordinate descent algorithm resulting from the Hathaway approach [8] can be shown to be identical to the update equations resulting from the EM approach (22) and (23).

In applying Hathaway's approach [8] to our problem, the E and M steps are replaced by a grouped coordinate descent on a new objective $\Phi_P^{\text{hath}}(\mathbf{z}, \boldsymbol{\theta}|\boldsymbol{\psi})$ where $\boldsymbol{\psi} = (\boldsymbol{\pi}, \boldsymbol{\beta})$ as before, but $\mathbf{z}$ is now reinterpreted as an *analog* variable $0 \leq \mathbf{z} \leq 1$. We can explain this reinterpretation as follows: In the E-step (22), expected values $E_z[\cdot]$ over the binary z are taken. If (22) is carried out, then $Q(\boldsymbol{\psi} \,|\, \widehat{\boldsymbol{\psi}}^{kl})$ will contain *analog* variables $0 \leq w_{an} \leq 1$ where $w_{an} = E_z[z_{an}|\theta_n, \widehat{\boldsymbol{\psi}}^{kl}]$. The optimization (23) is then on the analog $\mathbf{w}$'s. In Hathaway's approach, the EM steps (22) and (23) are replaced by a single objective (24) on $\mathbf{w}$. Optimizing the Hathaway objective on $\mathbf{w}$ leads to the same olution as the EM steps. The symbol $\mathbf{z}$ is commonly used as the binary indication variable in EM-mixture problems. In previous work where the Hathaway approach has been used, it has common to *also* use this same symbol $\mathbf{z}$, but now with the understanding that $\mathbf{z}$ is an analog term $(0 \leq \mathbf{z} \leq 1)$ that replaces the $\mathbf{w}$ mentioned above. We follow this convention of using analog $\mathbf{z}$ in the equations that follow. We shall retain this analog interpretation of $\mathbf{z}$ for the remainder of the paper. Note that (20) still applies to our reinterpreted $\mathbf{z}$. Let $\Lambda = \{z_{an}; a = 1, \ldots, L; n = 1, \ldots, N; 0 \leq z_{an} \leq 1, \sum_a z_{an} = 1\}$.

The Hathaway objective is defined as

$$\Phi_P^{\text{hath}}(\mathbf{z}, \widehat{\boldsymbol{\theta}}^k|\boldsymbol{\psi}) \equiv \sum_a \sum_n \Big\{ z_{an} \log z_{an}$$
$$+ z_{an} \log \frac{1}{\pi_a p_a(\widehat{\theta}_n^k|\alpha, \beta)} \Big\} \ \forall \boldsymbol{\psi} \in \Omega, \, \mathbf{z} \in \Lambda. \tag{24}$$

In [8], Hathaway showed that, for a general mixture decomposition, a constrained minimization of (24) via a grouped coordinate descent on $\mathbf{z}$ and $\boldsymbol{\psi}$, in fact, produces the

same update equations as those derived from the EM approach. Therefore, given identical initial conditions, the Hathaway and EM algorithms yield identical solutions despite the nonconvex nature of the mixture objective.

For our particular case of ML estimation with a gamma mixture, the grouped coordinate descent updates, derived in Appendix -A, become:

$$\widehat{z}_{an}^{kl} = \frac{\widehat{\pi}_{a}^{k,l-1} p_{a}(\widehat{\theta}_{n}^{k}|\alpha_{a}, \widehat{\beta}_{a}^{k,l-1})}{\sum_{b} \widehat{\pi}_{b}^{k,l-1} p_{b}(\widehat{\theta}_{n}^{k}|\alpha_{b}, \widehat{\beta}_{b}^{k,l-1})}, \tag{25}$$

$$\widehat{\pi}_{a}^{kl} = \frac{1}{N} \sum_{n} \widehat{z}_{an}^{kl}, \tag{26}$$

$$\widehat{\beta}_{a}^{kl} = \frac{\sum_{n} \widehat{z}_{an}^{kl} \widehat{\theta}_{n}^{k}}{\sum_{n} \widehat{z}_{an}^{kl}}. \tag{27}$$

In (25), p_a is a gamma pdf as defined in (10). The updates are, for each $\widehat{\boldsymbol{\theta}}^{k}$, iterated over index l till convergence, at which point, the k index is incremented for $(\widehat{\mathbf{z}}, \widehat{\boldsymbol{\pi}}, \widehat{\boldsymbol{\beta}})$.

B. Transforming the Reconstruction Side (17) of the Alternation

Our first transformation is complete: we can replace Φ_{P}^{mix} with Φ_{P}^{hath} in (18). In the second transformation step, the reconstruction step of the alternation (17) will be transformed so that it, too, incorporates the complete data $\mathbf{z}$. This fixpoint-preserving transformation will lead to a great simplification for the reconstruction calculation in (17). The transformation consists of simply replacing $\Phi_{P}^{\text{mix}}(\boldsymbol{\theta}|\boldsymbol{\psi})$ by $\Phi_{P}^{\text{hath}}(\mathbf{z}, \boldsymbol{\theta}|\boldsymbol{\psi})$ in the reconstruction step (17). With this substitution, the reconstruction step then becomes

$$\widehat{\boldsymbol{\theta}}^{k} = \arg \min_{\boldsymbol{\theta}} \left[\Phi_{L}(\mathbf{g}|\boldsymbol{\theta}) + \Phi_{P}^{\text{hath}}(\widehat{\mathbf{z}}^{k-1}, \boldsymbol{\theta}|\widehat{\boldsymbol{\psi}}^{k-1}) \right]. \tag{28}$$

We need to prove that the fixpoints are preserved, i.e. that $\widehat{\boldsymbol{\theta}}^{k}$ from (28) is the same whether we use Φ_{P}^{mix} or Φ_{P}^{hath}.

We first prove the following preliminary result in Appendix -B

$$\min_{\mathbf{z} \in \Lambda} \Phi_{P}^{\text{hath}}(\mathbf{z}, \boldsymbol{\theta}|\boldsymbol{\psi}) = \Phi_{P}^{\text{mix}}(\boldsymbol{\theta}|\boldsymbol{\psi}). \tag{29}$$

Armed with (29), we now demonstrate the fixpoint preservation:

$$\arg \min_{\boldsymbol{\theta}} \left[\Phi_{L}(\mathbf{g}|\boldsymbol{\theta}) + \Phi_{P}^{\text{mix}}(\boldsymbol{\theta}|\boldsymbol{\psi}) \right]$$

$$= \arg \min_{\boldsymbol{\theta}} \left[\Phi_L(\mathbf{g}|\boldsymbol{\theta}) + \min_{\mathbf{z} \in \Lambda} \Phi_P^{\text{hath}}(\mathbf{z}, \boldsymbol{\theta}|\boldsymbol{\psi}) \right]$$

$$= \arg \min_{\boldsymbol{\theta}} \left[\Phi_L(\mathbf{g}|\boldsymbol{\theta}) + \Phi_P^{\text{hath}}(\tilde{\mathbf{z}}(\boldsymbol{\theta}, \boldsymbol{\psi}), \boldsymbol{\theta}|\boldsymbol{\psi}) \right]$$

where $\tilde{\mathbf{z}}(\boldsymbol{\theta}, \boldsymbol{\psi}) \equiv \arg \min_{\mathbf{z} \in \Lambda} \Phi_P^{\text{hath}}(\mathbf{z}, \boldsymbol{\theta}|\boldsymbol{\psi})$ is delivered from the right side of the alternation. This demonstrates the intended result.

Thus the reconstruction step (17) becomes

$$\widehat{\boldsymbol{\theta}}^k = \arg \min_{\boldsymbol{\theta}} \left[\Phi_L(\mathbf{g}|\boldsymbol{\theta}) + \Phi_P^{\text{hath}}(\widehat{\mathbf{z}}^{k-1}, \boldsymbol{\theta}|\widehat{\boldsymbol{\psi}}^{k-1}) \right]$$

$$= \arg \min_{\boldsymbol{\theta}} \left\{ \Phi_L(\mathbf{g}|\boldsymbol{\theta}) - \sum_{an} \left[\widehat{z}_{an}^{k-1}(\alpha_a - 1) \log(\theta_n) \right. \right.$$
$$\left. \left. - \frac{\widehat{z}_{an}^{k-1} \alpha_a}{\widehat{\beta}_a^{k-1}} \theta_n \right] \right\} \tag{30}$$

To obtain the latter step, one simply plugs in (24), eliminating any terms independent of $\boldsymbol{\theta}$, and then uses the definition of the gamma pdf of $p_a(\hat{\theta}_n^k|\alpha, \beta)$ given by (10). The above minimization is easily computed via any suitable gradient method, and in Section V, we illustrate our tomographic reconstruction using a preconditioned conjugate gradient (PCG) method to optimize (30).

Compare (30) to the objective that would result had we used the independent (at each location n) but nonidentically distributed gamma prior pdf of (9) and (10):

$$\widehat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \left\{ \Phi_L(\mathbf{g}|\boldsymbol{\theta}) - \sum_n \left[(\alpha_n - 1) \log(\theta_n) - \frac{\alpha_n}{\beta_n} \theta_n \right] \right\} \tag{31}$$

where now, α_n and β_n comprise the original $2N$ space-varying parameters of the gamma prior. Comparison of (30) and (31) reveals that our reconstruction step update rule takes the form of an independent gamma prior, but crucially, we now have a prescription for generating its parameters:

$$\alpha_n - 1 \equiv \sum_a \widehat{z}_{an}(\alpha_a - 1) \tag{32}$$

$$\frac{\alpha_n}{\beta_n} \equiv \sum_a \widehat{z}_{an} \frac{\alpha_a}{\widehat{\beta}_a} \tag{33}$$

Note that (32) and (33) are valid at any step k of the alternation and consequently we suppress index k in (32) and (33).

With the left side transformation, we have now substituted for the optimization in (17) the optimization in (31)-(33). There are several advantages to using (31)-(33): (i) It is a

---

Joint-MAP Reconstruction

Initial conditions $\{\boldsymbol{\theta}, \pi_a, \beta_a\}$

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

$\quad \boldsymbol{\theta} \leftarrow \arg\min_{\boldsymbol{\theta}} \{\Phi_L(\mathbf{g}|\boldsymbol{\theta}) + \Phi_P^{\text{hath}}(\mathbf{z}, \boldsymbol{\theta}|\boldsymbol{\psi})\}$

Begin B: $l - loop.$ Do B until $\{z_{an}, \pi_a, \beta_a\}$ converge

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

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

$\quad \beta_a \leftarrow \frac{\sum_n z_{an}\theta_n}{\sum_n z_{an}}$
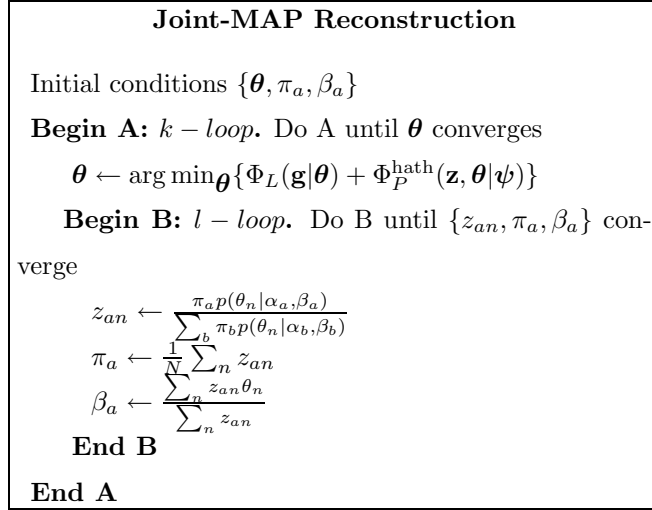
End B

End A

Fig. 2. The pseudocode for the joint-MAP alternating descent reconstruction.

convex objective unlike (17) and is therefore an easier optimization. (ii) As a byproduct, one obtains a segmentation $\{\widehat{z}_{an}\}$ of the object estimate into intensity classes. (iii) It is an independent gamma prior at each voxel, and we now have a prescription for generating its hyperparameters via (32) and (33).

C. Final Objective and Update Algorithm for joint-MAP Reconstruction

With our transformations, the overall optimization (2) has become

$$(\boldsymbol{\theta}^*, \boldsymbol{\psi}^*, \mathbf{z}^*) \equiv \arg \min_{\boldsymbol{\theta}, \boldsymbol{\psi} \in \Omega, \mathbf{z} \in \Lambda} \left[\Phi_L(\boldsymbol{\theta}|\mathbf{g}) + \Phi_P^{\text{hath}}(\mathbf{z}, \boldsymbol{\theta}|\boldsymbol{\psi}) \right]. \tag{34}$$

The final alternation of the joint-MAP reconstruction is

$$\widehat{\boldsymbol{\theta}}^k = \arg\min_{\boldsymbol{\theta}} \left[\Phi_L(\mathbf{g}|\boldsymbol{\theta}) + \Phi_P^{\text{hath}}(\widehat{\mathbf{z}}^{k-1}, \boldsymbol{\theta}|\widehat{\boldsymbol{\psi}}^{k-1}) \right], \tag{35}$$

$$\{\widehat{\mathbf{z}}^k, \widehat{\boldsymbol{\psi}}^k\} = \arg \min_{\mathbf{z} \in \Lambda, \boldsymbol{\psi} \in \Omega} \left[\Phi_P^{\text{hath}}(\mathbf{z}, \widehat{\boldsymbol{\theta}}^k|\boldsymbol{\psi}) \right]. \tag{36}$$

We collect the update equations (25), (26), (27) and (30), and summarize these in the pseudocode of Fig. 2. Note that specific examples of the tomographic likelihood functions $\Phi_L(\mathbf{g}|\boldsymbol{\theta})$ will be given in Section V.

V. APPLICATION TO TOMOGRAPHY

In this section, we apply our joint-MAP formalism to emission and transmission tomography. The imaging model relating the object to projection data is captured in the

definition of our likelihood $\Phi_L(\mathbf{g}|\mathbf{f})$. Some representative 2-D results will be shown.

Fig. 3. The emission phantom has three regions of relative intensity (1:4:8), cold disk (left), background
      ellipse, and hot disk (right).

For any reconstruction, the user must set the parameters L and $\alpha_a, a = 1, \ldots, L$. The
number of classes L is assumed known for our medical examples.[3] As explained earlier,
α_a is left as a free parameter that controls image quality by controlling a bias-variance
tradeoff. For the simulations that follow, we use an empirical scheme to choose α_a in
which the α_a are chosen to minimize the rmse (root-mean-squared error) between the
reconstruction and true object. This yields a reasonable bias-variance tradeoff. Also, we
note that the reconstructions are not especially sensitive to α_a.

A. Emission Tomography

In emission tomography [11] [12], including PET (positron emission tomography) and
SPECT (single-photon emission computed tomography), the goal is to reconstruct the
emission object $\mathbf{f}$ from the projection data $\mathbf{g}$. Here both the vectors $\mathbf{g}$ and $\mathbf{f}$ are lexico-
graphically ordered with elements $\{g_m; m = 1, \ldots, M\}$ and $\{f_n; n = 1, \ldots, N\}$, respec-
tively. The object voxel f_n, the radionuclide density, is defined as the mean number of
photons emitted from voxel n into 4π s.r. per unit time. The sinogram element g_m is the
integer number of counts detected during the imaging time at a detector bin indexed by
m. The observed sinogram $\mathbf{g}$ has an independent (at each detector element m) Poisson
distribution with mean $\bar{g}_m = \sum_n \mathcal{H}_{mn} f_n$ [13]. The system matrix $\mathcal{H}$ comprises elements
$\{\mathcal{H}_{mn}; m = 1, \ldots, M; , n = 1, \ldots, N\}$, indicating a probability per unit time that a photon
emitted from voxel n gets detected in detector bin m. A more complete model of the

[3]This is an excellent assumption for medical transmission tomography but not quite as good an assumption for
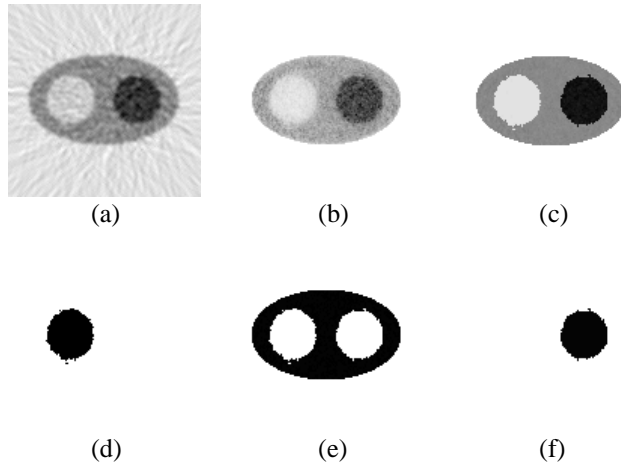medical emission tomography.

Fig. 4. This illustrates the results of (a) FBP emission reconstruction, (b) EM emission reconstruction, (c) joint MAP emission reconstruction, and the estimated z_{an} maps ($0 \leq z_{an} \leq 1$) for three classes (d) a=1 (hot disk) (e) a=2 (background) (f) a=3 (cold disk).

Poisson mean would include an affine term ($\bar{g}_m = \sum_n \mathcal{H}_{mn} f_n + r_m$) to model the effects of randoms and scatter in PET and SPECT, but the exclusion of r_m in our simulations is minor and does not substantially affect our simulations.

With $\mathbf{f}$ now assuming the role of $\boldsymbol{\theta}$, we may use the definition of $\bar{\mathbf{g}}$ and the independent Poisson nature of $\mathbf{g}$ to write the likelihood as:

$$\Phi_L(\mathbf{g}|\mathbf{f}) = -\sum_m \{g_m \log(\bar{g}_m) - \bar{g}_m\} \tag{37}$$

where terms independent of f_n have been dropped. The joint-MAP updates (35) and (36) may now be written using (37), and the reconstruction is carried out according to the prescription in Fig. 2.

We now demonstrate the application of the joint-MAP reconstruction approach with gamma mixture priors to the emission reconstruction problem. (We retain a uniform hyperprior in these simulations.) For the emission object $\mathbf{f}$, we used the 128×128 emission object shown in Fig. 3, which has three regions of relative intensities (1:4:8). (Here, darker regions are more intense.) The regions comprise a hot disk, cold disk and background ellipse. Projections were simulated at 129 angles spaced over 360 degrees with 192 detector bins per angle, and the simulated imaging accumulates 500K total counts into the sinogram.
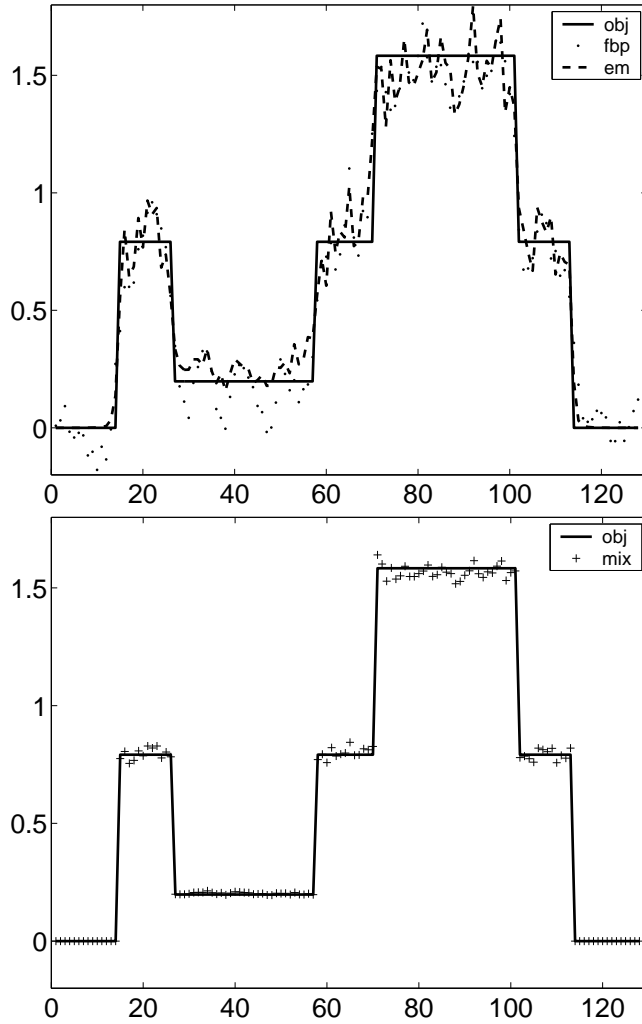
Fig. 5. This illustrates the profile plot of the phantom (solid line), FBP (dotted line), EM (dashed line), and joint-MAP (plus-sign) reconstructions, along central row of each image. The joint-MAP (+ sign) is displayed separately for clarity.

With the projection data **g** available, we can apply the joint-MAP method of Fig. 2 to reconstruct the object. For the reconstruction step of the alternating algorithm (35), we carry out the optimization by an iterative PCG (preconditioned conjugate gradient) algorithm [14], with a diagonal preconditioner. For this trial, the values of α_a are chosen to be 20, 40 and 80 for the three classes, respectively. The number of classes L is assumed known and fixed at L=3 (hot disk, cold disk and background ellipse). We need an initial object estimate for the right side of the alternating algorithm (36), and here a 5-iteration ML-EM reconstruction [15] is used for the initial estimate. The initial values for $\{\pi_a\}$ and

$\{\beta_a\}$ are less critical, and can be obtained easily from the initial estimate.

(a)   (b)   (c)
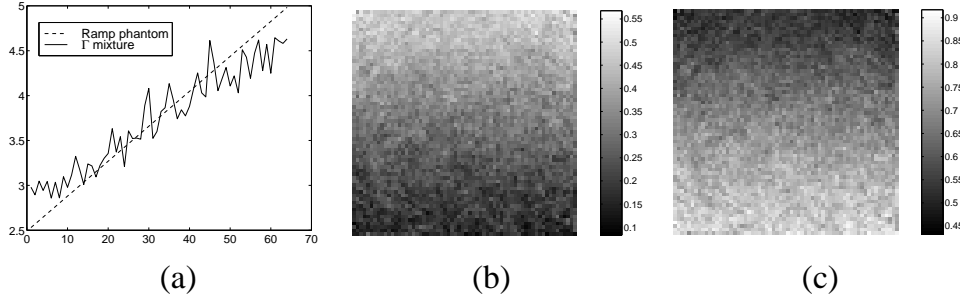
Fig. 6. (a) Profile plots along central column of ramp emission phantom (dotted line) and its joint-MAP
reconstruction (solid line). Axes are intensity vs. pixel number. (b) $\{z_{an}\}$ for class $a = 1$. Intensity
encodes values of $\mathbf{z}$. (c) $\{z_{an}\}$ for class $a = 2$. Intensity encodes values of $\mathbf{z}$. Note the gradual
transition of the $\mathbf{z}$'s.

An anecdotal result of the joint-MAP reconstruction is shown in Fig. 4(c), and its final
z_{an}-maps from the mixture decomposition are displayed in Figs. 4(d)(e)(f) for three classes.
These analog z_{an}-maps constitute a segmentation of the object. For comparison, an FBP
(filtered backprojection) reconstruction [16] and a ML-EM reconstruction [15] are also
presented in Figs. 4(a)(b), respectively. To further illustrate reconstructions, profile plots
along the central row of the reconstructions are illustrated in Fig. 5 along with a profile of
the true phantom. The joint-MAP reconstruction is superior in this anecdotal simulation.

It may appear that the mixture-of-gammas prior works well only with piecewise constant
objects, but the result in Fig. 6 illustrates how the 2-component mixture decomposition
leads to gradual spatial transitions in the z_{an}. The phantom is a ramp intensity object
ranging from a low value of 2.5 at the bottom row to 5.0 at the top row. Noisy projection
data of this phantom are computed and reconstructed using our joint-MAP scheme with
$L = 2$ and α_a chosen as per our rmse scheme. Profile plots of the phantom and its
reconstruction along the center column are shown in Fig. 6 (a). Figures 6(b) and (c) are
images of z_{1n} and z_{2n}. The intensity encodes the value of $\mathbf{z}$. The gradual spatial transitions
in z_{an} show that at a given location, the $\mathbf{z}$'s can take on values far from 1.0 or 0.0, thus
illustrating a noncommittal membership.

B. Transmission Tomography

In transmission tomography, one tries to reconstruct an image of attenuation coefficients from data obtained by projecting rays from many source locations and angles through the attenuating object. X-ray computed tomography (CT) is a familiar application, but in CT as clinically practiced, the data quality is high and analytical algorithms related to FBP can be used. Other applications, such as transmission scanning for attenuation correction in PET and SPECT [17], result in far noisier data, and estimation-theoretic approaches such as ours become useful.

The transmission object is denoted by vector $\boldsymbol{\mu}$ with N lexicographically ordered elements μ_n for each pixel $n \in \{1, \ldots, N\}$. Note that the object $\boldsymbol{\theta}$ now becomes $\boldsymbol{\mu}$. The value μ_n is thus the attenuation coefficient of the material in voxel n. The transmission projection data is denoted again by vector $\mathbf{g}$ of M lexicographically ordered elements g_m, where $m \in \{1, \ldots, M\}$ indexes the detector bins as well as the ray along which the beam is attenuated. The integer value g_m is the number of photons received at bin m. The problem is then to reconstruct $\boldsymbol{\mu}$ from the data $\mathbf{g}$. The transmission projection data $\mathbf{g}$ again follow an independent Poisson distribution at each bin m, but mean $\bar{g}_m$ is now given by

$$\bar{g}_m = u_m e^{-\sum_m l_{mn}\mu_n}, \tag{38}$$

where l_{mn} is the path length of ray m as it traverses voxel n. The quantity $\{u_m; m = 1, \ldots, M\}$, the "blank scan", is the number of counts received when no object is present, and can be obtained by calibration experiments. As with the emission case, terms $\bar{r}_m$ may be added to the right side of (38) to model physical effects that arise in specific applications of attenuation correction in PET and SPECT, but their exclusion does not alter our results substantially.

The likelihood for transmission tomography thus becomes

$$\Phi_L(\mathbf{g}|\boldsymbol{\mu}) = -\sum_m \{g_m \log(\bar{g}_m) - \bar{g}_m\} \tag{39}$$

where constant terms have been dropped, and $\bar{g}_m$ is given by (38).

To simulate PET imaging, we used the attenuation object as shown in Fig. 7(a), which has only two values of narrow-beam attenuation coefficients appropriate at energy 511KeV,
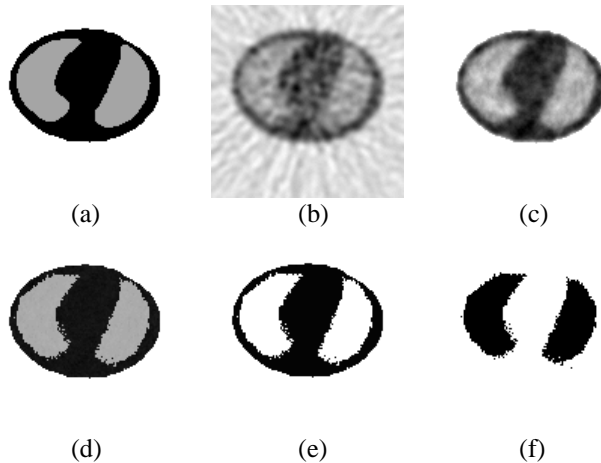
Fig. 7.   This shows (a) the transmission object, and the results of (b) FBP transmission reconstruc-
tion, (c) transmission reconstruction with a quadratic smoothing prior, (d) joint-MAP transmission
reconstruction, and the estimated z_{an} maps for two classes (e) a=1,(f) a=2.

$0.096cm^{-1}$ for soft tissue, $0.035cm^{-1}$ for lung. The object comprises 128x128 pixels. We
used transmission counts totalling 500K. Projections were simulated at 129 angles uni-
formly spaced over 180°, with 192 detector bins per angle.

For the transmission reconstruction in the left hand side of the alternating algorithm in
(35), the optimization is again carried out by an iterative PCG algorithm with a diagonal
preconditioner. For the gamma mixture prior, we used $L = 2$ (lung, soft tissue). The
initial values of π_a and β_a are not especially critical and were easily estimated for the
transmission case from an initial object estimate. The values for α_a in (lung, soft tissue)
were set at $(15, 60)$. We used a 2-iteration ML reconstruction as the initial object estimate.
The reconstruction was then carried out according to the prescription in Fig. 2.

Figure 7 (d) shows an anecdotal reconstruction of the joint-MAP method, while Fig-
ure 7(e)(f) are the estimated z_{an}-maps for two classes. The z_{an}'s thus constitute a seg-
mentation of the object. In addition, we ran an FBP transmission reconstruction shown
in Fig. 7(b). Figure 7(c) shows a penalized likelihood reconstruction obtained with a con-
ventional quadratic smoothing prior: $\Phi_P^{smooth} = \frac{1}{2}\lambda \sum_n \sum_{n' \in \mathcal{N}(n)} (\theta_n - \theta_{n'})^2$. Here $\lambda > 0$
is a parameter that controls the degree of smoothing, and $\mathcal{N}(n)$ a neighborhood system
for the voxel n. The value of the smoothing parameter was set to $\lambda = 2000$ for the best
rmse result and a 4-nearest-neighbor scheme used for $\mathcal{N}(n)$. A profile plot of these recon-
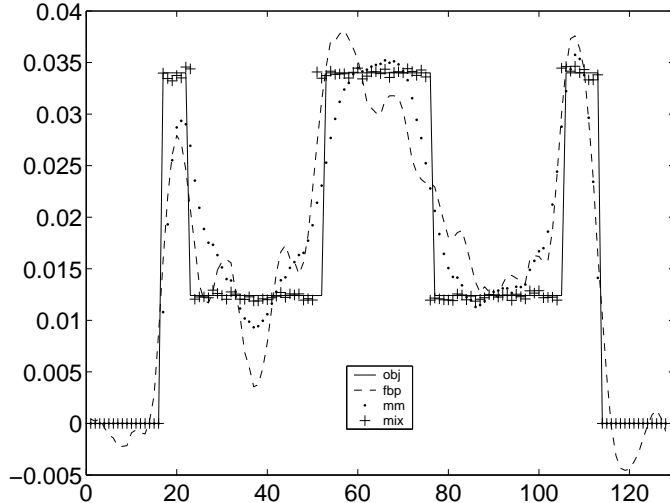
Fig. 8. This illustrates the profile plot of the phantom (solid line - obj), FBP (dashed line - fbp), conven-
tional smoothing prior (dotted line - mm) and joint-MAP (plus-sign - mix) transmission reconstruc-
tions, along the central row of each image. The joint-MAP reconstruction shows better performance
as compared with the FBP and smoothing-prior reconstructions in terms of the qualitative profile
result.

struction along central row, displayed in Fig. 8, illustrates a sense of the quality of the
joint-MAP reconstruction as compared with the FBP and conventional smoothing results.

VI. Discussion

A. Significance of Work

We have presented a joint-MAP reconstruction with a gamma mixture prior and ap-
plied it to statistical tomographic reconstructions. The gamma mixture prior assumes a
clustered intensity histogram in the object, and enforces positivity naturally. Our imple-
mentation of the joint-MAP reconstruction contains an alternation between a penalized
likelihood reconstruction step, and a hyperparameter estimation step (gamma mixture
decomposition). In applying Hathaway's approach [8] to the hyperparameter estimation
step, one can transform the original, complex mixture objective to one that is easier to
optimize. This transformation also allows the utilization of the complete data $\mathbf{z}$ for the
reconstruction step, and leads to a simplification of the penalized likelihood optimiza-
tion. Remarkably, the transformation also leads to an equivalent, independent gamma
prior, where the individual mean and variance parameters of the independent gamma are

computed from the mixture decomposition. We applied the joint-MAP reconstruction to both emission and transmission tomography, and the anecdotal results illustrate the effectiveness of the new prior.

A significant aspect of our work is that it allows one to effectively use the desireable independent-but-not-identically-distributed non-mixture form of the gamma prior pdf in (9). Recall that for (10), if one were able to specify a good value for α_n and β_n at each pixel, then the resulting optimization (31) would be straightforward. These α_n, β_n determine a "external reference value", i.e. a value at each location n, towards which the prior draws the reconstruction. However, assuming α_n is set by the user to control bias variance tradeoff, there is no obvious way to choose β_n. By using an independent-and-identically-distributed mixture pdf (15), we apparently change the problem so that each voxel is drawn by the same multimodel pdf, reflecting our knowledge about the histogram of the medical object of concern. So, in our simulations, each voxel seeks a value consistent with belonging to a given class, e.g. lung or soft tissue. The resulting alternation, (17)(18), is difficult to carry out, so we *transform* (17)(18) to a new alternation (35)(36) that (i) preserves the fixed point of (17)(18) and (ii) introduces a new intermediate variable z_{an} that can be interpreted as a probability of class membership of voxel n in class a. Interestingly, while (36) delivers the values for z_{an} and β_a, the update (35) is exactly of the form (31) needed for the non-mixture gamma prior (9)! In (31), the pdf is the independent, non-mixture, *non*-identically distributed prior pdf of (9), but now, the effective α_n and β_n have been determined, thus "closing the loop". Although α_a is user specified, once it has been specified, the α_n are now computed automatically. Again, the iid mixture form, through transformations of the objective, results in the independent non-identically distributed non-mixture form in (31). The fixpoints remain the same, so optimizing (35)(36) is the same as the optimization (17)(18). This entire scheme does not rely on any particular property of gamma priors; we chose them for their convenient positivity-preserving properties. The entire scheme could easily be repeated with independent Gaussian pdf, for example.

B. Related Previous Work

In [4], an independent gamma prior was first applied in a penalized likelihood reconstruction for both emission and transmission tomography, which is equivalent to our recon-

struction step of the joint-MAP reconstruction. However, there was no hyperparameter estimation step for computing the voxelwise hyperparameters, and these had to be specified empirically. Another pointwise Gaussian prior [18] was combined with a penalized likelihood for emission tomography. As in [4], there was no hyperparameter estimation step for computing the voxel-by-voxel hyperparameters, and the authors resorted to an empirical method of obtaining these from an initial FBP reconstruction. In [19], the gamma mixture model has been applied to estimate the proportions of classes in an SAR (synthetic aperture radar) image, where the object intensity histogram has a few (L) peaks with L known. In their work, only the hyperparameter estimation step of our alternation is performed using a slightly different derivation of the EM update equation to estimate $\boldsymbol{\psi}$ and $\mathbf{z}$. In [20], an independent Gaussian mixture prior was introduced for a penalized likelihood reconstruction for emission tomography, but their approach was not a joint-MAP method with both reconstruction and hyperparameter estimation steps. The Gaussian mixture parameters were computed from a FBP reconstructed image in advance, and kept fixed throughout the reconstruction [20]. Thus there was no hyperparameter estimation step of the alternation as in our joint-MAP reconstruction.

C. Issues for Future Work

The free parameter L, the number of classes in the mixture model, can be confidently fixed in many applications such as thorax transmission tomography or the application in [19]. However, in other applications, such as emission tomography, the estimation of L becomes difficult. One solution is to utilize an information criterion [21], or other estimation-theoretic methods [22] and [23].

Since the mixture model is nonconvex, the result is sensitive to initial conditions. Here, the initial conditions comprise the initial estimate $\widehat{\boldsymbol{\theta}}^0$, and the initial mixture parameters $\widehat{\boldsymbol{\psi}}^0$. For our applications to tomography, the solution is fairly robust to initial conditions. For example, one can use an FBP reconstruction as the initial estimate, and use this to get an initial guess for the mixture parameter $\widehat{\boldsymbol{\psi}}^0$. In particular, for transmission tomography, a reasonable mean value β_a of the attenuation coefficient for each tissue class can be obtained from reference data on tissue properties. Nevertheless, we have investigated the use of a deterministic annealing (DA) method [24] for this problem. DA methods are

useful in finding approximate global minima to nonconvex objectives, and have found use in a wide variety of applications. To use the DA method for our case, one need only make a minor modification [24] to the update of (25). Results of applying DA illustrate a robust solution independent of the initial conditions [24].

It is relatively straightforward to incorporate hyperpriors mathematically. To incorporate hyperpriors, one may simply begin with the general formulation (5) that includes the Φ_{HP} term, and apply all those steps used in deriving (26) and (27). Inclusion of a hyperprior leaves the reconstruction step completely unchanged. In [9], we derived modified versions of (26) and (27) to include a Dirichlet hyperprior [10] on $\boldsymbol{\pi}$ and a gamma hyperprior for $\boldsymbol{\beta}$. However, further development is needed for testing the effects of theses hyperpriors. For our medical applications, hyperpriors on $\boldsymbol{\pi}$ and $\boldsymbol{\beta}$ have natural interpretation in terms of pdf's (over an ensemble of patients) for cluster proportions and means.

The gamma-mixture prior models voxel values as having a clustered histogram, but does not model spatial smoothness. In many applications, one would expect that a spatial correlation exists between neighboring voxels. Since the mixture prior does not assume any spatial interaction, it is possible to see some sparse misclassified (wrong $\mathbf{z}$) voxels in the result. A possible way of including the spatial information is to introduce smoothing interactions on the complete data $\mathbf{z}$. This involves adding smoothing terms in (24), and thus a modification of the update equation (25). Relevant approaches in [25] and [26] discuss the problem of incorporating neighbor correlations in the mixture decomposition. We will investigate this issue in future work.

APPENDIX

A. Derivation of Update equations (25)-(27)

In the following we use the definition of p_a as a gamma prior as in (10). The mixture decomposition problem is stated as a constrained minimization problem on (24):

$$
\arg \min_{\boldsymbol{\psi} \in \Omega, \mathbf{z} \in \Lambda} \Phi_P^{\text{hath}}(\mathbf{z}, \widehat{\boldsymbol{\theta}}^k | \boldsymbol{\psi})
$$

$$
= \arg \min_{\boldsymbol{\psi} \in \Omega, \mathbf{z} \in \Lambda} \sum_n \sum_a \left\{ z_{an} \log z_{an} \right.
$$

$$
\left. + z_{an} \log \frac{1}{\pi_a p_a(\widehat{\theta}_n^k | \alpha_a, \beta_a)} \right\} \tag{40}
$$

which may be reexpressed as an unconstrained minimization via the introduction of Lagrange multipliers. By using Lagrange multipliers and adding two constraints into the objective function in (24), an overall objective function is obtained,

$$
D(\mathbf{z}, \boldsymbol{\psi}) = \sum_n \sum_a \left\{ z_{an} \log z_{an} + z_{an} \log \frac{1}{\pi_a p_a(\widehat{\theta}_n^k | \alpha_a, \beta_a)} \right\} \\
+ \zeta(\sum_a \pi_a - 1) - \sum_n \kappa_n(\sum_a z_{an} - 1) \tag{41}
$$

where Lagrange multipliers ζ and $\{\kappa_n\}$ are used in imposing the constraints $\sum_a \pi_a = 1$ and $\sum_a z_{an} = 1$, respectively. The grouped coordinate descent is then: minimize $D(\mathbf{z}, \widehat{\boldsymbol{\psi}}^{k,l-1})$ w.r.t. $\mathbf{z}$ given the previous estimate $\widehat{\boldsymbol{\psi}}^{k,l-1}$, then minimize $D(\widehat{\mathbf{z}}^{kl}, \boldsymbol{\psi})$ w.r.t. $\boldsymbol{\psi}$ given the current estimate $\widehat{\mathbf{z}}^{kl}$.

We now demonstrate that the update equations (25)(26)(27) can be obtained through this minimization. Setting partial derivatives w.r.t. κ_n and z_{an} to zero as per Karush-Kuhn-Tucker first order conditions [27], one obtains

$$
\frac{\partial D(\mathbf{z}, \widehat{\boldsymbol{\psi}}^{k,l-1})}{\partial \kappa_n} = \sum_a z_{an} - 1 = 0 \tag{42}
$$

$$
\frac{\partial D(\mathbf{z}, \widehat{\boldsymbol{\psi}}^{k,l-1})}{\partial z_{an}} = \log z_{an} + 1 \\
- \log \left[\widehat{\pi}_a^{k,l-1} p_a(\widehat{\theta}_n^k | \alpha_a, \widehat{\beta}_a^{k,l-1}) \right] - \kappa_n = 0 \tag{43}
$$

Solving for z_{an} in (43), we obtain,

$$
z_{an} = e^{\kappa_n - 1} \widehat{\pi}_a^{k,l-1} p_a(\widehat{\theta}_n^k | \alpha_a, \widehat{\beta}_a^{k,l-1}) \tag{44}
$$

and using (42)

$$
\sum_a z_{an} = 1 = e^{\kappa_n - 1} \sum_a \{ \widehat{\pi}_a^{k,l-1} p_a(\widehat{\theta}_n^k | \alpha_a, \widehat{\beta}_a^{k,l-1}) \} \\
e^{\kappa_n - 1} = \frac{1}{\sum_a \{ \widehat{\pi}_a^{k,l-1} p_a(\widehat{\theta}_n^k | \alpha_a, \widehat{\beta}_a^{k,l-1}) \}} \tag{45}
$$

Then by substituting (45) into (44), one can obtain (25) for $\widehat{z}_{an}^{kl}$. With the current estimate $\widehat{z}_{an}^{kl}$, one can estimate the mixture parameters $\boldsymbol{\psi}$ by extremizing $D(\widehat{\mathbf{z}}^{kl}, \boldsymbol{\psi})$ with respect to

ψ and the Lagrange parameters

$$\frac{\partial D(\widehat{\mathbf{z}}^{kl}, \boldsymbol{\psi})}{\partial \boldsymbol{\psi}} = \frac{\partial}{\partial \boldsymbol{\psi}} \bigg[\sum_n \sum_a \Big\{ \widehat{z}_{an}^{kl} \log \frac{1}{\pi_a p_a(\widehat{\theta}_n^k | \alpha_a, \beta_a)} \Big\}$$
$$- \zeta \big(\sum_a \pi_a - 1 \big) \bigg] = 0 \tag{46}$$

where

$$\frac{\partial D(\widehat{\mathbf{z}}^{kl}, \boldsymbol{\psi})}{\partial \boldsymbol{\psi}} = 0 \implies \begin{cases} \frac{\partial D(\widehat{\mathbf{z}}^{kl}, \boldsymbol{\psi})}{\partial \boldsymbol{\pi}} = 0 \\ \frac{\partial D(\widehat{\mathbf{z}}^{kl}, \boldsymbol{\psi})}{\partial \boldsymbol{\beta}} = 0 \end{cases}$$

and the term in [square brackets] in (46) excludes terms independent of $\boldsymbol{\psi}$. First extremizing w.r.t. π_a and ζ,

$$\frac{\partial D(\widehat{\mathbf{z}}^{kl}, \boldsymbol{\psi})}{\partial \pi_a} = \sum_n - \frac{\widehat{z}_{an}^{kl}}{\pi_a} + \zeta = 0, \tag{47}$$

$$\frac{\partial D(\widehat{\mathbf{z}}^{kl}, \boldsymbol{\psi})}{\partial \zeta} = \sum_a \pi_a - 1 = 0. \tag{48}$$

From (47) and (48), one gets $\zeta = N$, and the update equation for π_a in (26). Then zeroing the derivative w.r.t. β_a, one obtains

$$\frac{\partial D(\widehat{\mathbf{z}}^{kl}, \boldsymbol{\psi})}{\partial \beta_a} = \frac{\partial}{\partial \beta_a} \Big\{ \sum_n \sum_a - \widehat{z}_{an}^{kl} \log p_a(\widehat{\theta}_n^k | \alpha_a, \beta_a) \Big\}$$
$$= \frac{\partial}{\partial \beta_a} \Big\{ \sum_n \sum_a - \widehat{z}_{an}^{kl} \Big[- \frac{\alpha_a \widehat{\theta}_n^k}{\beta_a} + (\alpha_a - 1) \log \widehat{\theta}_n^k$$
$$+ \alpha_a \log \frac{\alpha_a}{\beta_a} - \log \Gamma(\alpha_a) \Big] \Big\}$$
$$= \sum_n \widehat{z}_{an}^{kl} \Big[- \frac{\alpha_a \widehat{\theta}_n^k}{\beta_a^2} + \frac{\alpha_a}{\beta_a} \Big] = 0 \tag{49}$$

which leads immediately to the update (27).

B. Derivation for the Transformation (29)

With the definition of the function $\tilde{\mathbf{z}}$

$$\tilde{\mathbf{z}}(\boldsymbol{\theta}, \boldsymbol{\psi}) \equiv \underset{\mathbf{z} \in \Lambda}{\arg\min} \, \Phi_P^{\text{hath}}(\mathbf{z}, \boldsymbol{\theta} | \boldsymbol{\psi}), \tag{50}$$

the left side of (29) may be expressed as

$$\min_{\mathbf{Z} \in \Lambda} \Phi_P^{\text{hath}}(\mathbf{z}, \boldsymbol{\theta} | \boldsymbol{\psi}) = \Phi_P^{\text{hath}}(\tilde{\mathbf{z}}(\boldsymbol{\theta}, \boldsymbol{\psi}), \boldsymbol{\theta} | \boldsymbol{\psi}) \tag{51}$$

Using the result from Appendix-A in obtaining (25), we express $\tilde{\mathbf{z}}(\boldsymbol{\theta}, \boldsymbol{\psi})$ as

$$\tilde{z}_{an}(\boldsymbol{\theta}, \boldsymbol{\psi}) = \frac{\pi_a p_a(\theta_n|\alpha_a, \beta_a)}{\sum_b \pi_b p_b(\theta_n|\alpha_b, \beta_b)} = \frac{A_{an}}{B_n} \tag{52}$$

where for convenience we have summarized the numerator and denominator by $A_{an} \equiv \pi_a p_a(\theta_n|\alpha_a, \beta_a)$ and $B_n \equiv \sum_b \pi_b p_b(\theta_n|\alpha_b, \beta_b)$, respectively. Inserting (52) into (24), we obtain

$$\Phi_P^{\text{hath}}(\tilde{\mathbf{z}}(\boldsymbol{\theta}, \boldsymbol{\psi}), \boldsymbol{\theta}|\boldsymbol{\psi}) = \sum_{an} \left[\frac{A_{an}}{B_n} \log\left(\frac{A_{an}}{B_n}\right) + \frac{A_{an}}{B_n} \log \frac{1}{A_{an}} \right]$$

$$= -\sum_{an} \frac{A_{an}}{B_n} \log(B_n) = -\sum_n \log(B_n) \equiv \Phi_P^{\text{mix}}(\boldsymbol{\theta}|\boldsymbol{\psi}) \tag{53}$$

where we have made use of the normalization $\sum_a A_{an} = B_n$. The relation (29) is thus established.

References

[1] A. Mohammad-Djafari, "A Full Bayesian Approach for Inverse Problems," in *Proc. 15th Inter. Work. Maximum Entropy and Bayesian Methods*, K. Hanson and R. Silver, Eds. 1995, pp. 135–144, Kluwer Academic.

[2] D. Luenberger, *Optimization by Vector Space Methods*, John-Wiley and Sons, New York, NY, 1969.

[3] N. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distribution*, vol. 1, John Wiley and Sons, 2nd edition, 1994.

[4] K. Lange, M. Bahn, and R. Little, "A Theoretical Study of Some Maximum Likelihood Algorithms for Emission and Transmission Tomography," *IEEE Trans. Med. Imaging*, vol. 6, no. 2, pp. 106–114, June 1987.

[5] G. J. McLachlan and K. E. Basford, *Mixture Models*, Marcel Dekker, 1987.

[6] 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*, vol. 39, pp. 1–38, 1977.

[7] R. A. Redner and H. F. Walker, "Mixture Densities, Maximum Likelihood and the EM Algorithm," *SIAM Review*, vol. 26, no. 2, pp. 195–239, April 1984.

[8] R. J. Hathaway, "Another Interpretation of the EM Algorithm for Mixture Distributions," *Stat. Prob. Letters*, vol. 4, pp. 53–56, 1986.

[9] I.-T. Hsiao, *Bayesian Image Reconstruction for Emission and Transmission Tomography*, Ph.D. thesis, Dept of Electrical Engineering, State Univ. of New York at Stony Brook, Stony Brook, New York 11784, USA, May 2000.

[10] D. Titterington, A. Smith, and U. Markov, *Statistical Analysis of Finite Mixture Distributions*, John Wiley and Sons, 1985.

[11] R. J. Jaszczak, "Tomographic Radiopharmaceutical Imaging ," *Proceedings of the IEEE*, vol. 76, pp. 1079–1094, September 1988.

[12] M. Phelps, J. Mazziotta, and H. Schelbert, Eds., *Positron Emission Tomography and Autoradiography: Principle and Applications for the Brain and Heart*, Raven Press, New York, 1986.

[13] H. H. Barrett and W. Swindell, *Radiological Imaging: the Theory of Image Formation, Detection, and Processing*, vol. I and II, Academic Press, Inc., 1981.

[14] 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*, vol. 13, no. 4, pp. 687–701, Dec. 1994.

[15] K. Lange and R. Carson, "EM Reconstruction Algorithms for Emission and Transmission Tomography," *J. Comp. Assist. Tomography*, vol. 8, no. 2, pp. 306–316, April 1984.

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

[17] D. L. Bailey, "Transmission Scanning in Emission Tomography," *Euro. J. Nuc. Med.*, vol. 25, no. 7, pp. 774–787, July 1998.

[18] E. Levitan and G. T. Herman, "A Maximum *A Posteriori* Probability Expectation Maximization Algorithm for Image Reconstruction in Emission Tomography," *IEEE Trans. Med. Imaging*, vol. 6, no. 3, pp. 185–192, Sept. 1987.

[19] R. Samadani, "A Finite Mixture Algorithm for Finding Proportions in SAR Images," *IEEE Trans. Image Processing*, vol. 4, no. 8, pp. 1182–1186, Aug. 1995.

[20] Z. Liang, R. Jaszczak, and R. Coleman, "On Reconstruction and Segmentation of Piecewise Continuous

Images," in *Info. Processing in Med. Imaging*, A. Colchester and D. Hawkes, Eds. 1991, vol. 12, pp. 94–104, Springer Verlag.

[21] Z. Liang, R. Jaszczak, and R. Coleman, "Parameter Estimation of Finite Mixtures Using the EM Algorithm and Information Criteria with Application to Medical Images," *IEEE Trans. Nuclear Science*, vol. 39, no. 4, pp. 1126–1133, 1992.

[22] P. Djurić, "A Reversible Jump Algorithm for Analysis of Gamma Mixtures," in *Proc. SPIE: Bayesian Inference for Inverse Problems*, A. Mohammad-Djafari, Ed., July 1998, vol. 3459, pp. 262–270.

[23] S. Richardson and P. Green, "On Bayesian Analysis of Mixtures with an Unknown Number of Components," *J. Royal Statist. Soc. B*, vol. 59, no. 4, pp. 731–792, 1997.

[24] I.-T. Hsiao, A. Rangarajan, and G. Gindi, "Bayesian Image Reconstruction for Transmission Tomography Using Mixture Model Priors and Deterministic Annealing Algorithms," in *Proc. SPIE: Medical Imaging*, M. Sonka and K. Hanson, Eds., Oct. 2001, vol. 4322.

[25] D. Titterington, "Some Recent Research in the Analysis of Mixture Distributions," *Statistics*, vol. 21, pp. 619–641, 1990.

[26] S. Gopal, *Deterministic and Statistical Solutions for Noninverse Problems in Image Processing*, Ph.D. thesis, Department of Electrical and Computer Engineering, University of Houston, Aug. 1996.

[27] D. Luenberger, *Linear and Nonlinear Programming - 2nd ed.*, Addison-Wesley, Menlo Park, CA, 1984.