Chapter 5

Joint MAP Reconstruction Using Mixture Models as Priors

In previous chapters, we have described some advantages of statistical reconstruction methods. The independent pointwise gamma prior has nice properties, but it is difficult to obtain the individual mean and variance of each component distribution at each location. In chapter 4 we mentioned the use of a mixture model as means of calculating individual means and variances for the gamma prior. In this chapter we show how to accomplish this. A joint MAP (maximum \textit{a posteriori}) procedure, which performs both transmission reconstruction and mixture parameter estimation simultaneously, is proposed. It turns out that the joint MAP procedure with a gamma mixture prior is particularly applicable to transmission imaging. We also consider the application of the joint MAP procedure with gamma mixture prior for emission reconstruction.

5.1 Gamma-Regularized Transmission Reconstruction

One can add a penalty or prior term to the TT likelihood objective function to form a penalized likelihood objective for reconstruction to obtain Eq. (3.64) (repeated
where $\Phi_L(g | \mu)$ is the TT likelihood objective function as given by Eq. (3.18) in chapter 3 (repeated here)

$$
\Phi_L(g | \mu) = \sum_i \{ g_i \log(\tilde{g}_i) - \tilde{g}_i \}
$$

(3.18)

and $\tilde{g}_i = u_i \exp(-\sum_j \mathcal{H}_{ij} \mu_j) + r_i$. For an independent gamma penalty with known parameters, $\alpha$ and $\beta$, the prior objective was shown to be Eq. (3.22) (repeated here with $\theta = \mu$)

$$
\Phi_P^\Gamma(\mu; \alpha, \beta) = \sum_{j=1}^{N} \left[ (\alpha_j - 1) \log \mu_j - \frac{\alpha_j}{\beta_j} \mu_j \right].
$$

(3.22)

Hence, with the MAP objective Eq. (3.64), and any unconstrained algorithm, one can pursue a penalized likelihood reconstruction for TT. (Recall that a MAP reconstruction is entirely equivalent to a penalized likelihood reconstruction for estimating $\mu$ given the posterior objective $\Phi(\mu | g)$.)

As described previously, a big problem for using this independent gamma prior is that there is no clear way to calculate, at each point, the mean and variance of the gamma for each pixel. For an attenuation map of $N$ pixels ($N$ is usually in the order of $10^4$ for a 2D slice), specification of the $2N$ parameters $\alpha_j, \beta_j$ ($j = 1, .., N$) is required. One empirical way is to first get a FBP reconstruction, and set $\beta_j$ equal to it, then empirically set all $\alpha_j$ to some constant. Here, our goal is to retain the nice properties of gamma prior, and use a more principled way to get the $\alpha_j, \beta_j$.

A first step to the solution of this problem is to replace the independent gamma prior with a finite mixture model [31], which comprises $L$ component distributions.
with each one a gamma distribution. As described in chapter 4, a set of independent observations \( \mu = \{ \mu_1, ..., \mu_N \} \), i.e. the pixels of \( \mu \), is said to have a mixture model of \( L \) component densities, if it has the density function given by Eq.(4.1) (repeated here with \( \theta = \mu \))

\[
P(\mu | \alpha, \beta, \pi) = \prod_{j=1}^{N} \sum_{a=1}^{L} \pi_a P(\mu_j | \alpha_a, \beta_a)
\]

(4.1)

where \( \alpha_a, \beta_a \) and \( \pi_a \) are the parameters and mixing proportion of class \( a \), with \( \sum_{a=1}^{L} \pi_a = 1, \pi_a > 0 \), and \( P(\mu_j | \alpha_a, \beta_a) \) is the conditional (gamma) density function given that \( \mu_j \) belongs to class \( a \) [31].

In addition to its use in estimating parameters, another motivation for the mixture model is that it is a good model for the intensity distribution. The intensity histogram of attenuation coefficients in an area like thorax comprises several peaks, each identified with a particular tissue type (index by \( a \)). As illustrated in Fig.5.1, for example, the thorax usually has \( L = 2 \) different tissue classes of soft tissue and lung at the energy level of 511 KeV [141]. Note that this is the intensity histogram of \( \hat{\mu} \), the object estimates, not \( \mu \), the object. Therefore, a finite mixture model can account for this multimodal distribution. As will be shown, we need estimate only the parameters of each class, \( \{ \alpha_a, \beta_a; a = 1, ..., L \} \), instead of each pixel, \( \{ \alpha_j, \beta_j; j = 1, ..., N \} \). Thus, we now need to estimate \( 3L \) (including \( \pi_a \)) additional parameters rather \( 2N \). Since \( L \) is typically less than 5, the complexity of the parameter estimation is greatly reduced. The overall procedure to carry out the reconstruction/mixture-decomposition turns out to be a joint MAP scheme, which we now describe.
Figure 5.1: The intensity histogram of a reconstructed attenuation map for a thorax phantom at 511 KeV has two clustered regions (background pixels not included), soft tissue region with peak at 0.096 cm$^{-1}$, and lung region with peak at 0.036 cm$^{-1}$. Note that this is the histogram that would be decomposed during the reconstruction/segmentation processing as explained later, whereas the underlying histogram of the real object has two “delta function” peaks.

5.2 Joint MAP Strategy

In a conventional MAP reconstruction for transmission tomography one seeks an estimate which maximizes the posterior function given projection data $g$, and one typically assumes the parameters of the prior are also known. Now, we propose a joint-MAP procedure, where both the object $\mu$ and the parameters in the prior are unknown. This is an important difference over the conventional MAP reconstruction as used in medical imaging, and makes the algorithm practically useful in that all the parameters of the prior are determined along with the reconstruction. In conventional
MAP reconstruction, empirical procedures to find the parameters are usually used and more user interaction is needed. Alternatively, many hyperparameter estimation schemes that have been proposed previously are very complex [167]. Our strategy for the joint MAP is to simultaneously solve for $\mu$ as well as parameters, $\alpha, \beta$ and $\pi$. Recall that the MAP reconstruction method is equivalent to a penalized likelihood method for the case of fixed hyperparameters and no hyperprior. Thus the optimizations become the same for both cases. However, if the hyperparameters are estimated at the same time, and a pdf for hyperparameters is assumed, then we have a true Bayesian MAP problem.

The posterior function for the joint MAP reconstruction is $P(\mu, \alpha, \beta, \pi | g)$. From Bayes’ theorem:

\begin{align}
P(\mu, \alpha, \beta, \pi | g) & \propto \Pr(g | \mu) P(\mu, \alpha, \beta, \pi) \\
& \propto \Pr(g | \mu) P(\mu | \alpha, \beta, \pi) P(\alpha, \beta, \pi)
\end{align}

(5.1)

(5.2)

where we used the fact that the data $g$ depends on $\mu$ only. The likelihood $\Pr(g | \mu)$ is again Poisson, and the prior $P(\mu | \alpha, \beta, \pi)$ will be modeled as a gamma mixture. The hyperprior $P(\alpha, \beta, \pi)$ on the prior parameters will be discussed shortly.

Our tactic is to use an alternating iterative ascent procedure, updating, at iteration $k$, the object estimate $\hat{\mu}^k$ while holding $(\alpha, \beta, \pi)$ fixed, then the parameters $(\alpha^k, \beta^k, \pi^k)$ given the latest $\mu$ estimate. With this strategy we can list a joint-MAP procedure, (but this is not the final version.)

\begin{align}
\hat{\mu}^k = \arg \max_{\mu} \Pr(g | \mu) P(\mu | \hat{\alpha}^{k-1}, \hat{\beta}^{k-1}, \hat{\pi}^{k-1}) P(\hat{\alpha}^{k-1}, \hat{\beta}^{k-1}, \hat{\pi}^{k-1}) \\
= \arg \max_{\mu} \Pr(g | \mu) P(\mu | \hat{\alpha}^{k-1}, \hat{\beta}^{k-1}, \hat{\pi}^{k-1})
\end{align}

(5.3)
\[(\hat{\alpha}^k, \hat{\beta}^k, \pi^k) = \arg\max_{\alpha, \beta, \pi} \Pr(g | \mu^k) P(\mu^k | \alpha, \beta, \pi) P(\alpha, \beta, \pi)\]

\[= \arg\max_{\alpha, \beta, \pi} P(\mu^k | \alpha, \beta, \pi) P(\alpha, \beta, \pi) \quad (5.4)\]

Note that Eq. (5.3) is a conventional MAP reconstruction and Eq. (5.4) is a MAP mixture decomposition. However, as discussed in chapter 4 for the ML version, this calculation is intractable for Eq. (5.4) as stated, but may be made tractable by using an EM algorithm and an appropriate complete data space [95]. The complete data is discussed earlier and given by the indicator function, \(Z_{aj}\), and its probability, \(\Pr(Z_{aj} = 1)\), is interpreted as the probability of any pixel \(j\) belong to class \(a\). Note that introducing the complete data \(Z\) into an EM formulation of mixture decomposition will change the final estimate \(\hat{\mu}, \hat{\alpha}, \hat{\beta}\) and \(\hat{\pi}\) because the overall objective function is highly non-convex, and different algorithm will lead to a different fixed point. Also note that complete data satisfies \(\sum_a z_{aj} = 1\) and \(\pi_a = 1/N \sum_j z_{aj}\), as in chapter 4. Thus we can write the mixture prior objective function with complete data \(z\) as

\[\Phi_P^{mix}(\mu, \alpha, \beta, \pi, z) = \sum_{a,j} z_{aj} \text{log} \pi_a + (\alpha_a - 1) \text{log} \mu_j + \alpha_a \text{log}(\frac{\alpha_a}{\beta_a^a}) - \text{log} \Gamma(\alpha_a) - \frac{\alpha_a \mu_j}{\beta_a} \quad (5.5)\]

where \(z_{aj}\) is calculated through the E-step of the EM algorithm for mixture decomposition. Here we have so far considered only the likelihood function of the mixture model without the hyperprior terms. Therefore, the overall objective function with the gamma mixture prior for transmission tomography becomes

\[\Phi(\mu | g) = \Phi_L(g | \mu) + \Phi_P^{mix}(\mu; \alpha, \beta, \pi, z) \quad (5.6)\]

where the gamma mixture objective \(\Phi_P^{mix}\) has replaced the independent gamma objective.
\( \Phi_P^+ \) in Eq (3.64). The new alternating update, replacing Eqs.(5.3) and (5.4) is

\[
\hat{\mu}^k = \arg\max_{\mu} \left[ \Phi_L(g|\mu) + \Phi_P^\text{mix}(\mu; \hat{\alpha}^{k-1}, \hat{\beta}^{k-1}, \hat{\pi}^{k-1}, \hat{z}^{k-1}) \right] \tag{5.7}
\]

\[
\hat{z}^k, \hat{\pi}^k, \hat{\alpha}^k, \hat{\beta}^k = \arg\max_{z, \pi, \alpha, \beta} \Phi_P^\text{mix}(\hat{\mu}^k; \alpha, \beta, \pi, z) \tag{5.8}
\]

This is our final form of the update, though we need to supply additional detail for the various terms in Eqs. (5.7) and (5.8). Note that Eq (5.7) is a conventional MAP transmission reconstruction with known gamma mixture penalties, and the known hyperparameters of the penalties are provided from Eq (5.8), which is simply a mixture decomposition that fits parameters of each class to the current histogram \( \hat{\mu}^k \) using an EM algorithm [95].

### 5.3 Mixture Decomposition

In this section, we supply explicit forms for calculating Eq.(5.8).

#### 5.3.1 ML for Gamma Mixture Decomposition

The ML derivation for the gamma mixture parameters estimation using the EM algorithm can be found in chapter 4. Without a hyperprior on gamma mixture priors, the update derivation for gamma mixture parameters described in chapter 4 is indeed the second step, Eq. (5.8), of the joint-MAP procedure. They are repeated here,

\[
z_{aj}^l = \frac{\hat{\pi}_{a}^l P_a(\mu_j|\hat{\alpha}_{a}^l, \hat{\beta}_{a}^l)}{\sum_b \hat{\pi}_{a}^l P_b(\mu_j|\hat{\alpha}_{b}^l, \hat{\beta}_{b}^l)} \tag{5.9}
\]

\[
\hat{\pi}_{a}^{l+1} = \frac{1}{N} \sum_j z_{aj}^l \tag{5.10}
\]
\[
\hat{\beta}_a^{l+1} = \frac{\sum_j z_{aj}^{l} \mu_j}{\sum_j z_{aj}^{l}}
\]

\[
\hat{\alpha}_a^{l+1} = \arg \max_{\alpha_a} \sum_j z_{aj}^{l} \left[ (\alpha_a - 1) \log \mu_j + \alpha_a \log \left( \frac{\alpha_a}{\hat{\beta}_a^{l+1}} \right) - \log \Gamma(\alpha_a) - \frac{\alpha_a \mu_j}{\hat{\beta}_a^{l+1}} \right]
\]

where \( P_a(\mu_j|\alpha_a, \beta_a) \) is the component gamma pdf.

Note that the joint-MAP procedure, (5.7) and (5.8), includes two iterative loops, the outer loop and the inner loop. The outer loop is the MAP reconstruction indexed by \( k \) with the \( k \)th parameters supplied by the inner loop which is the mixture decomposition. Because the mixture decomposition, Eqs.(5.9)-(5.12), is itself an iterative procedure, its iteration is indexed by \( l \).

### 5.3.2 MAP for Gamma Mixture Decomposition

To make the mixture estimation strictly Bayesian, a hyperprior term is necessary [146]. So far we have presented the reconstruction by implicitly assuming a uniform hyperprior. Given the interpretation associated with a mixture model prior, we naturally interpret \( P(\alpha, \beta, \pi) \) as a hyperprior on the means, variances and mixing proportions of the various transmission histogram peaks. In principle, it is possible to obtain an ensemble of such histograms for a large collection of patients; histograms of CT chest scans are approximations at low, broad-spectrum energies [10]. A model that we adopt (mostly for mathematical convenience) is that of independent distributions for \( \alpha \), and \( \beta \), both modeled as Pareto and gamma, respectively. As for \( \pi \), one simple hyperprior often used is the Dirichlet distribution [14, 146]. We now show a modification of the alternating procedure, Eqs.(5.8), for the case of joint MAP including the hyperprior. In the following we will assume the attenuation map \( \mu \) is given from the latest estimate, and thus \( \mu \) remains fixed.
For the mixture EM algorithm in chapter 4, the estimate of parameters is obtained by optimizing the mixture ML-EM objective Eq.(5.5), where only the likelihood function is used. Now we add a hyperprior term \( P(\psi) \) to the likelihood, and that leads to a new MAP mixture objective. The EM algorithm is easily modified to produce the MAP estimate as described in chapter 3 by simply including the log hyperprior term in the ML-EM objective. That is because the hyperprior is independent of the complete data \( z \) in the EM algorithm. Thus, the MAP-EM objective for mixture model becomes,

\[
Q^{MAP-EM}[\psi|\hat{\psi}^{t}] = Q^{ML-EM}[\psi|\hat{\psi}^{t}] + \log P(\psi)
\]

(5.13)

where \( Q^{ML-EM}[\psi|\hat{\psi}^{t}] \) is the ML-EM mixture objective Eq.(5.5), and \( \log P(\psi) \) the log hyperprior for the mixture parameter. Therefore, the M-step of the EM algorithm to estimate the parameters becomes an optimization of the new MAP-EM mixture objective (5.13). One strategy is to consider that \( \pi, \beta, \) and \( \alpha \) are all independent. hence \( P(\pi, \beta, \alpha) = P(\pi)P(\beta)P(\alpha) \). Below, we discuss forms for \( P(\pi), P(\beta), \) and \( P(\alpha) \).

Hyperprior for \( \pi \)  Assume \( \pi = \{\pi_a; a = 1, .., L\} \) has a Dirichlet distribution [14, 146] of dimension \( L - 1 \), with parameters \( n = \{n_a; n_a > 0, a = 1, .., L\} \) if it has the probability density \( P(\pi) \),

\[
P(\pi) = K \prod_{a}^{L} \pi_a^{n_a - 1}
\]

(5.14)

where \( \sum_{a} \pi_a = 1, 0 \leq \pi_a \leq 1.0 \) and \( K = \frac{\Gamma(\sum_{a} n_a)}{\prod_{a} \Gamma(n_a)} \). Note that if \( L = 2 \), it reduces to a beta density. The mean \( \mu_{\pi_a} \) and the variance \( v_{\pi_a} \) of the Dirichlet distribution are

103
given by

\[
\begin{align*}
    u_{\pi_a} &= \frac{n_a}{\sum_a n_a} \\
    v_{\pi_a} &= \frac{u_{\pi_a}(1 - u_{\pi_a})}{1 + \sum_a n_a}
\end{align*}
\]

Thus, the MAP-EM objective with hyperprior in Eq.(5.13) becomes

\[
Q^{MAP-EM}[\psi | \hat{\psi}^l] = Q^{ML-EM}[\psi | \hat{\psi}^l] + \log P(\pi) + \log P(\beta) + \log P(\alpha)
\]

\[
= \sum_{a} \hat{z}_{aj}^l \left\{ (\alpha_a - 1) \log \mu_j + \alpha_a \log \frac{\alpha_a}{\beta_a} - \mu_j \frac{\alpha_a}{\beta_a} - \log \Gamma(\alpha_a) \right\} \\
+ \sum_{a} \hat{z}_{aj}^l \log \pi_a + \sum_{a} \left\{ (n_{\pi_a} - 1) \log \pi_a \right\} + \log P(\beta) + \log P(\alpha)
\]

(5.15)

By adding the constraint \(\sum_a \pi_a = 1\) through a Lagrange multiplier \(\zeta\), and zeroing the first derivative w.r.t \(\pi_a\) and \(\zeta\) similarly to that in chapter 4, one can get

\[
\frac{\partial \left\{ Q^{MAP-EM} - \zeta (\sum_a \pi_a - 1) \right\}}{\partial \zeta} = \sum_a \pi_a - 1 = 0
\]

(5.16)

\[
\frac{\partial \left\{ Q^{MAP-EM} - \zeta (\sum_a \pi_a - 1) \right\}}{\partial \pi_a} = \sum_j \hat{z}_{aj}^l \frac{1}{\pi_a} + \frac{n_a - 1}{\pi_a} - \zeta = 0
\]

(5.17)

Rearranging Eq. (5.17) and using Eq. (5.16), one can get

\[
\zeta \sum_a \pi_a = \sum_a \left\{ \sum_j \hat{z}_{aj}^l + n_a - 1 \right\} \\
\zeta = N + \sum_a n_a - L
\]

(5.18)

which follows the constraints \(\sum_a \pi_a = 1\), and \(\sum_a \hat{z}_{aj} = 1\). Therefore, by putting

\[\zeta = N + \sum_a n_a - L\]

into Eq. (5.17), and rearranging it, one gets the final update
equation for \( \pi_a \),

\[
\hat{\pi}_{a}^{t+1} = \frac{\sum_j z_{aj}^t + n_a - 1}{N + \sum_a n_a - L}
\]  

(5.19)

**Hyperprior for \( \beta \)** Assume each \( \beta_a \) is distributed as a gamma with parameters \( \alpha \leftrightarrow u_{\beta_a} \) and \( \beta \leftrightarrow v_{\beta_a} \) for \( a = 1, \ldots, L \). That is,

\[
\log P(\beta) = \sum_a \left[ (u_{\beta_a} - 1) \log \beta_a - u_{\beta_a} \frac{\beta_a}{\beta_a} \right]
\]  

(5.20)

Thus, the MAP-EM objective with hyperprior \( P(\beta) \) becomes

\[
Q_{MAP-EM}^{\hat{\psi}}[\psi] = Q_{ML-EM}^{\hat{\psi}}[\psi] + \log P(\beta) + \log P(\pi) + \log P(\alpha)
\]

\[
= \sum_{a} z_{aj}^t \left\{ \log \pi_a + (\alpha_a - 1) \log \mu_j + \alpha_a \log \frac{\alpha_a}{\beta_a} - \mu_j \frac{\alpha_a}{\beta_a} - \log \Gamma(\alpha_a) \right\}
\]

\[
+ \sum_{a} \left\{ (u_{\beta_a} - 1) \log \beta_a + \log P(\pi) + \log P(\alpha) \right\} (5.21)
\]

The first derivative w.r.t. \( \beta_a \) is given by

\[
\frac{\partial Q_{MAP-EM}}{\partial \beta_a} \bigg|_{\pi_a = \hat{\pi}_{a}^{t+1}, \alpha_a = \hat{\alpha}_a^t} = \left\{ \sum_j z_{aj}^t \mu_j \frac{\alpha_a}{\beta_a} - \mu_j \frac{\alpha_a}{\beta_a} - \frac{u_{\beta_a} - 1}{\beta_a} \right\} = 0
\]

Thus, one can derive a closed form update equation for \( \beta_a \),

\[
\hat{\beta}_{a}^{t+1} = \frac{-b_3 \pm \sqrt{b_3^2 + 4a_3c_3}}{2a_3}
\]  

(5.22)

where \( A_3 = \frac{u_{\beta_a}}{v_{\beta_a}}, B_3 = (\sum_j z_{aj}^t \hat{\alpha}_a^t - u_{\beta_a} + 1), \) and \( C_3 = \sum_j z_{aj}^t \mu_j \hat{\alpha}_a^t \). Also, \( \hat{\beta}_{a}^{t+1} \) must be positive.
Hyperprior for $\alpha$  An independent Pareto hyperprior is used for $\alpha$. A random variable $X = x$ is said to have a Pareto distribution $[21, 65]$ if its pdf has the form

$$P(x|u, v) = \frac{v^u}{x^{v+1}} \quad u < x < \infty, \; u > 0, \; \text{and} \; v > 0 \quad (5.23)$$

where its mean is $\frac{uv}{v-1}$ for $v > 1$, and its variance $\frac{u^2v}{(v-1)^2(v-2)}$ for $v > 2$. In order to satisfy the constraint for $\alpha_a$, $\alpha_a > 1$, we have to make sure $1 \leq u$. That is, assume each $\alpha_a$ is independently distributed as a Pareto with parameters $u \leftrightarrow u_{\alpha_a} > 1$, $v \leftrightarrow v_{\alpha_a}$, and $v_{\alpha_a} > 0$. Thus, the hyperprior pdf for $\alpha_a$ becomes

$$P(\alpha) = \prod_a \frac{v_{\alpha_a}u_{\alpha_a}v_{\alpha_a}}{\alpha_{\alpha_a}+1} \quad (5.24)$$

By adding the log hyperprior $\log P(\alpha)$, the MAP objective turns out to be

$$Q_{MAP-EM}^{\psi|\hat{\psi}} = Q_{ML-EM}^{\psi|\hat{\psi}} + \log P(\alpha) + \log P(\pi) + \log P(\beta)$$

$$= \sum_{a_j} z_{aj} \left\{ \log \pi_a + (\alpha_a - 1) \log \mu_j + \alpha_a \log \frac{\alpha_a}{\beta_a} - \frac{\alpha_a}{\beta_a} - \log \Gamma(\alpha_a) \right\}$$

$$+ \sum_a \left\{ \log (v_{\alpha_a}) - (v_{\alpha_a} + 1) \log \alpha_a \right\} + \log P(\pi) + \log P(\beta) \quad (5.25)$$

However, as in the ML-EM for estimating $\alpha$, Eq. (5.25) cannot be solved explicitly and an iterative 1-D optimization method, such as Newton’s method, can be applied to obtain

$$\hat{\alpha}_a^{l+1} = \arg \max_{\alpha_a} \left\{ Q_{MAP-EM}^{\psi|\hat{\psi}} \right\}_{\pi_a=\hat{\pi}_a^{l+1}, \beta_a=\hat{\beta}_a^{l+1}} \quad (5.26)$$

Note that the update for $\hat{\psi}_{aj}$ above remains the same as in Eqs. (5.9), and the new MAP updates Eq. (5.22) for $\beta_a$, Eq. (5.26) for $\alpha_a$ and Eq. (5.19) for $\pi_a$ have replaced Eqs. (4.22), (4.24) and (4.20).
5.4 MAP Reconstruction for Transmission Tomography

In this section, we supply explicit means for calculating Eq. (5.7). Now, let’s go back to the first step, Eq. (5.7), of the joint MAP procedure. This is an ordinary MAP reconstruction method (or penalized-likelihood reconstruction) for transmission tomography as in section 3.4.5. This part can be accomplished by any suitable unconstrained method. One might, for instance, use the EM reconstruction algorithm [59] listed as Eq.(3.59). However, EM for transmission tomography is too slow due to too many calculations of exponentials at each iteration, and is also inexact with an approximation in the M-step as discussed in chapter 3. To avoid the inexact approximation, here we simply use the method of preconditioned conjugate gradients [156].

With the mixture decomposition in Eq (5.8) at iteration \( k \) resolved, then the MAP reconstruction Eq.(5.7) has the objective given by Eq.(5.6). The gamma mixture objective for the reconstruction now takes the form

\[
\Phi^{mix}_P(\boldsymbol{\mu}; \alpha^k, \beta^k, \pi^k, \hat{z}^k) = \sum_{a,j} \hat{z}^k_{a,j} \left[ \log \mu_j - \frac{\hat{\alpha}^k_a}{\hat{\beta}^k_a} \mu_j \right]
\]  

(5.27)

Note that this equation is similar to Eq. (5.5) but its \( k \)th parameters \((\alpha^k, \beta^k, \hat{z}^k)\) are known and the terms independent of object \( \boldsymbol{\mu} \) have been dropped. Comparison of the objective for the prior in Eq (5.1) to that of Eq (5.27) reveals the final solution for the hyperparameters at each location as an appropriate \( z \)-weighted combination of \( \alpha_a \) and \( \beta_a \):

\[
\alpha_j - 1 = \sum_a \hat{z}^k_{a,j} (\hat{\alpha}^k_a - 1) \tag{5.28}
\]

\[
\frac{\alpha_j}{\beta_j} = \sum_a \hat{z}^k_{a,j} \frac{\hat{\alpha}^k_a}{\hat{\beta}^k_a} \tag{5.29}
\]
Thus, the difficulty of calculating a great number of parameters is overcome by an easy mixture decomposition of a few mixture component parameters. Equations (5.28) and (5.29) explicitly show the reduction of $2N$ parameters to $3L$ parameters, a reduction of about 20000 to 12 for a realistic TT problem.

With hyperparameters known, we can then optimize the MAP objective in Eq (5.6) with a suitable algorithm, and here we use PCG algorithm described in chapter 3. The PCG update is exactly the same as in chapter 3 for MAP transmission reconstruction using a gamma prior, but here the hyperparameters are given by Eqs. (5.28) and (5.29).

In sum, a Bayesian means for using a gamma prior in transmission reconstruction is formulated by computing the parameters of the independent gamma at each location through a mixture decomposition procedure.

## 5.5 Final Joint MAP Procedure

The final update procedure follows the form of the joint-MAP Eqs. (5.7) and (5.8). Here, $k$ indexes an outer loop, and $l$ an inner loop. Start with initial estimate $\hat{\mu}^0$ and perform Eq.(5.8). That is, at iteration $k$ compute mixture parameters $\hat{\kappa}^k_a, \hat{\tau}^k_a, \hat{\beta}^k_a, \hat{\gamma}^k_a$ using the iterations (until stable) over $l$ given by Eqs. (5.9) (5.11) (5.12). Using these parameters, perform the MAP estimate for $\hat{\mu}^k_j$ given by a PCG optimization procedure. Then find the next mixture parameters using known $\hat{\mu}^k_j$. The cycle of mixture resolution and MAP estimation, indexed by $k$, repeats until stability is reached. The following pseudocode and the diagram Fig. 5.2 summarizes the procedure.
Figure 5.2: The alternating algorithm can be summarized in this diagram. The left side is a conventional MAP reconstruction (same as in Eq.(5.7)) with gamma mixture priors and the parameters \((\hat{\alpha}, \hat{\beta}, \hat{\pi}, \hat{z})\) provided by the right side. Given the reconstruction \(\hat{\mu}\) from the left side, then the right side becomes a mixture decomposition (same as in Eq.(5.8)).

Final Updating Procedure

Initialize \((z, \pi, \alpha, \beta, \mu)\)

**Begin: Outer loop**

Do until \(\mu\) converges

\[
\mu \leftarrow \arg \max \mu \{ \Phi_L(g|\mu) + \Phi^\text{mix}_P(\mu; \alpha, \beta, z) \} ; \text{ using PCG algorithm}
\]

**Begin: Inner loop**

Do until \((z, \pi, \alpha, \beta)\) converge

\[
z_{aj} \leftarrow \frac{\pi_a P(\mu_j | \alpha_a, \beta_a)}{\Sigma_b \pi_b P(\mu_j | \alpha_b, \beta_b)}
\]

\[
\pi_a \leftarrow \frac{1}{\sum_j z_{aj}}
\]

\[
\beta_a \leftarrow \frac{\Sigma_j z_{aj} \mu_j}{\Sigma_j z_{aj}}
\]

\[
\alpha_a \leftarrow \arg \max \alpha_a \sum_j z_{aj} \left[ -\frac{\alpha_a \mu_j}{\beta_a} + (\alpha_a - 1) \log(\mu_j) + \alpha_a \log \frac{\beta_a}{\beta_a} - \log \Gamma(\alpha_a) \right]
\]

**End: Inner loop**

**End: Outer loop**
5.6 Discussion

5.6.1 Spatial Correlation

The gamma-mixture approach depends on pixel values having a clustered histogram, with individual peak in each class not overlapping too much. Thus it is likely to work better for transmission than emission scans.

In most applications, measurements at neighboring pixels are spatially correlated. Thus it would help if one can apply the spatial interaction into the model. Because the mixture model does not assume any spatial information, it is possible to see some sparse mis-classified pixels in the result of the mixture decomposition on an image. One improvement is to reintroduce smoothing through interaction terms on the complete data $z$ [145]. Smoothing on $z$ helps eliminate sparse pixels, but may have undesirable side effects, for example, oversmoothing the result and causing more mis-classification. Thus careful study is needed to choose a good smoothing form. It is also considerably complicated [145]. (The problems with smoothing on $z$ are very different than those associated with conventional smoothing priors on $\mu$ itself.) Other choice for imposing the spatial interaction is to use a spatially variant mixture model proposed in [46], where they claim that the application of smoothing to their model is more appropriate and less complicated than that to the finite mixture model.

5.6.2 Initial Conditions

A problem with the mixture model is that its solution depends on the initial conditions for the parameters. However, the initial condition for the mixture model prior in TT is not difficult to obtain. The initial condition of $\beta$ in the mixture model can be set easily since we know approximately the typical attenuation coefficients for
different anatomical regions. As for the initial values for $\pi$, the mixing proportion, one can get an initial estimate from an FBP or EM reconstruction, and apply a threshold-based segmentation method [160] to extract the initial region information, and then obtain initial values for $\pi$.

5.6.3 Number of Classes

As for the free parameter $L$, the number of classes in the mixture model, it can be confidently set for thorax imaging in the transmission case. However, for emission tomography, the estimation for $L$ becomes complicated. One solution is to utilize an information criterion [90, 95]. Others use the reversible Markov Chain Monte Carlo (MCMC) methods [27, 126]. One empirical way is to choose $L$ by simply inspecting an FBP reconstruction [164].

5.6.4 Hyperparameter Estimation

Compared to the hyperparameter ($\lambda$) estimation in a MAP reconstruction problem using a smoothing prior, the hyperparameter ($\alpha, \beta$) estimation in the gamma mixture model is much easier. The problem in estimating the hyperparameter $\lambda$ in a smoothing prior is due to the difficulty in evaluating the partition function $Z$ of Eq.(3.25),

$$Z(\lambda) = \int_{\theta} \exp(-\lambda U(\theta))$$

(3.25)

where $Z$ is dependent on the hyperparameter $\lambda$ [56]. The hyperparameter estimation for a smoothing prior, i.e. Eq.(5.4) for a smoothing prior, becomes

$$\hat{\lambda} = \arg \max_{\lambda} \log P(\theta | \lambda) = \arg \max_{\lambda} \log Z - \lambda U(\theta) = \arg \max_{\lambda} \log \left\{ \int_{\theta} \exp(-\lambda U(\theta)) \right\} - \lambda U(\theta)$$

(5.30)
where $P(\theta|\lambda)$ is the density function of a smoothing prior given in Eq.(3.24). To estimate $\lambda$, one has to evaluate the multidimensional integrals of $Z$ over the sample space for $\theta$. Since the energy function $U(\theta)$ usually contains a complex coupling structure amongst the elements of $\theta$, the integration for $Z$ becomes very complicated, and thus the estimation for $\lambda$ becomes very difficult. To overcome the difficulty, various methods have been proposed, and a good review on this issue can be seen in [167]. However, some of the methods are intractable due to the high dimensionality of the problem, or impractical due to the requirement of a large computational cost, and some empirical approximations have been proposed.

For our gamma mixture model, the tractable ML estimation of the hyperparameter using an EM algorithm has less computational cost when compared to that for a smoothing prior. Such nice properties simply come naturally with the use of the gamma mixture prior.

### 5.6.5 Computational Problems with $\alpha$

Although the hyperparameter estimation of the mixture model is less complicated than that of a smoothing prior, however, there is a computational problem with the estimate of the hyperparameter $\alpha$. We found that the values of $\hat{\alpha}_a^k$ were increasing toward infinity as the iterations proceeded. Such a divergence problem of $\hat{\alpha}_a^k$ still happens even when imposing a hyperprior on $\alpha_a$. It is not an artifact, but is implicit in the objective function itself. We do not have any answer for this problem yet. Our strategy for now is to fix $\alpha_a$. One possible empirical way is to limit $\alpha_a$ to a range, for example, $\alpha_a$ is only allowed to update between $[5,500]$. The other way is to let $\alpha_a$ be a free parameter, and chosen by a user, like the choice of filter and cutoff frequency in FBP reconstruction. This analogy is appropriate, since $\alpha_a$ controls smoothness in
effect. The divergence problem can be illustrated in a simple 1-D Gaussian model with a Gaussian prior, where the variance of the prior has the same divergence problem as the $\alpha_n$. Future studies on this simple model may lead to some insights on this problem. It appears that the joint-MAP strategy will have to be modified.

5.7 Applying the Gamma Mixture Model to Emission Reconstruction

Since the joint MAP scheme with gamma mixture prior enjoyed some success in the application to transmission tomographic reconstruction, we wanted to explore its use for emission reconstruction.

The joint MAP procedure described in previous sections can be applied to emission case simply by changing the object $\mu$ to $f$ and using the likelihood expressions $\Phi_L(g|f)$ suitable for emission rather than transmission imaging,

$$
\hat{f}^k = \arg\max_f \Phi_L(g|f) + \Phi_P^{\text{mix}}(f|\hat{\alpha}^{k-1}, \hat{\beta}^{k-1}, \hat{\pi}^{k-1}, \hat{z}^{k-1}) \tag{5.31}
$$

$$
\hat{z}^k, \hat{\pi}^k, \hat{\alpha}^k, \hat{\beta}^k = \arg\max_{z,\pi,\alpha,\beta} \Phi_P^{\text{mix}}(f|\alpha, \beta, \pi, z) \tag{5.32}
$$

The second step (5.32) of the joint MAP is a mixture decomposition and follows same derivations as for the TT case, while the first step is a MAP emission reconstruction as in chapter 3, but with a gamma-mixture object model. The first step, MAP reconstruction (5.31), can be accomplished by any suitable algorithm. Here, we use the EM algorithm for the MAP emission reconstruction as in chapter 3. The reason is that the EM algorithm for emission reconstruction is easy to implement and involves no approximation form as is the case for TT.

The MAP-EM objective function with a gamma-mixture-model objective function
for the emission case becomes

\[ Q^{\text{MAP-EM}}(\mathbf{f} | \hat{\mathbf{f}}^k) = Q^{\text{ML-EM}}(\mathbf{f} | \hat{\mathbf{f}}^k) + \Phi^{\text{mix}}_P (\hat{\mathbf{f}}^k | \alpha, \beta, \pi, \mathbf{z}) \]  

(5.33)

where the mixture prior has fixed mixture parameters \((\hat{z}^k, \hat{\pi}^k, \hat{\alpha}^k, \hat{\beta}^k)\) given from mixture decomposition procedure in Eq. (5.32), and \(Q^{\text{ML-EM}}(\mathbf{f} | \hat{\mathbf{f}}^k)\) is the ML-EM objective function given in Eq. (3.39). By optimizing the MAP-EM objective function with respect to \(\mathbf{f}\), one can get the MAP-EM reconstruction equation:

\[ \hat{f}^k_j = \frac{\sum_i g^k_{ij} \hat{f}_i^{k-1} + \sum_a \hat{z}_a^k (\hat{\alpha}_a^k - 1)}{\sum_i \mathcal{H}_{ij} + \sum_a \hat{\alpha}_a^k \hat{f}_a^k}. \]  

(5.34)

Note that, the independent gamma prior appears as terms \(\alpha_j - 1\) and \(\frac{\alpha_j}{\beta_j}\) in Eq. (3.42), but in its original form, there is no prescription for specifying the values of these parameters at each location \(j\). Here, as in the case for transmission tomography, we have shown that these two terms are substituted by \(\sum_a \hat{z}_a^k (\hat{\alpha}_a^k - 1)\) and \(\sum_a \hat{\alpha}_a^k \hat{f}_a^k\), with all quantities available from the mixture decomposition in Eq. (5.32). Thus a hyperparameter at location \(j\) becomes a weighted (by \(z_{aj}\)) sum of contributions from each class \(a\).

5.8 Summary

In summary, we have developed a joint MAP scheme for the reconstruction problem both in transmission and emission tomography. To retain the nice properties of independent gamma priors, the gamma mixture model is introduced as the prior for the MAP reconstruction. Thus the hyperparameters are calculated through a mixture decomposition. Note that the use of the gamma mixture in image processing
is not new, but the idea of applying gamma mixtures to transmission tomography in order to calculate the individual parameters for pointwise priors is new.

Since our aim is to use this TT method for attenuation correction in the emission reconstruction, the joint MAP transmission reconstruction using the gamma mixture model will be compared to other attenuation correction methods. The results of performance comparisons are presented in the next chapter.