Simultaneous Hyperparameter Estimation and Bayesian Image Reconstruction for PET *

Zhenyu Zhou Richard M. Leahy Erkan U. Mumcuoglu
Signal and Image Processing Institute, Department of Electrical Engineering-Systems,
University of Southern California, Los Angeles, CA 90089-2564

Abstract

We present a new iterative algorithm for Bayesian PET image reconstruction that simultaneously estimates the PET image and the global hyperparameter $\beta$ of a Gibbs prior. True maximum likelihood (ML) estimation of $\beta$ is intractable for the PET reconstruction problem due to the complexity and high dimensionality of the probability densities involved. The new algorithm replaces the true likelihood function for the hyperparameter with an approximation in which the marginalization with respect to the image sample space is reduced to the product of a set of one dimensional integrals; one per image pixel. The approximation is closely related to the mean field theory of statistical mechanics. In essence, this reduction in complexity is achieved by approximating the influence of the neighbors of each pixel over their entire sample space with their estimated posterior modes. A preconditioned conjugate gradient algorithm is used to iteratively compute a MAP estimate of the image. At periodic intervals, the most recent image generated by this iterative procedure is used as an estimate of the posterior mode in the approximate marginalized log likelihood for the data given $\beta$, which in turn is used to update the ML estimate of $\beta$. The procedure is repeated until convergence of both the MAP image estimate and the ML estimate of $\beta$. Results of a validation study using Monte Carlo simulations are presented.

1 Introduction

Bayesian methods have proven to be very powerful for reconstruction of high quality PET images. However, a major problem limiting their utility in quantitative PET is the lack of a practical and robust method for selecting the parameters of the prior, the hyperparameters. Of particular importance for the case of Gibbs priors is the global parameter $\beta$ that multiplies the Gibbs energy function. This parameter controls the degree of influence of the Gibbs prior on the solution. If $\beta$ is too large, the prior will tend to have an over-smoothing effect on the solution. Conversely, if it is too small, the MAP estimate may be unstable, reducing to the ML solution as $\beta$ goes to zero.

The smoothing parameter is often chosen in an ad hoc fashion through visual inspection of the resulting images. There are two basic approaches for choosing $\beta$ in a more principled manner: (i) treating $\beta$ as a regularization parameter and applying techniques such as generalized cross validation, the L-curve, and $\chi^2$-squared goodness of fit tests; (ii) estimation theoretic approaches such as maximum likelihood. While the regularization based approaches can select reasonable values of $\beta$, computation times tend to be unacceptable as the PET image must be reconstructed for a number of different values of $\beta$. In the class of estimation theoretic approaches, computing a ML estimate of $\beta$ is desirable, but intractable. In fact for all but the simplest of priors, ML estimation of $\beta$ presents a difficult challenge even if the PET images are observed directly. Several alternatives to exact ML have been proposed in the literature including various approximations to the likelihood function, stochastic sampling methods, and a method of moments. Performance of the alternative methods that we have examined is generally either poor or remains too computationally demanding. Two such schemes are compared with our new method at the end of this paper.

Here we present an approximate ML method for selecting $\beta$ that is closely related to the mean field theory of statistical mechanics. The method allows simultaneous ML hyperparameter estimation and MAP image reconstruction. It can be easily combined with most existing iterative MAP estimation algorithms – in our case we use the gradient-projection preconditioned conjugate gradient method described in [8]. We present a brief summary of the problem formulation and describe the approximate ML estimation scheme below. The paper concludes with the results of a Monte-Carlo simulation study that examines the bias and variance of this estimator for cases where the true value of $\beta$ is known. We also evaluate the performance of the method for simulations involving a realistic brain phantom.

2 Estimation Methods

2.1 The Bayesian method for PET

We use the following model for the PET emission sinogram:

$$y_i = E(y_i) = \left( \frac{1}{A_i} \sum_j P_{ij} \lambda_j + s_i + \hat{r}_i \right) c_i \quad (1)$$
where \( \lambda_j \) is the emission image value at pixel \( j \), \( y_i \) is the number of detected coincidences at detector pair \( i \), \( \bar{y}_i \) is the expected value of \( y_i \), and \( P_{ij} \) is the geometrical probability of detecting an emission from pixel \( j \) at detector pair \( i \). \( A_i \) is the attenuation correction factor and \( c_i \) the detection efficiency at detector pair \( i \); \( s_i \) and \( r_i \) are the scatter and random component respectively. It is assumed that accurate attenuation and detector efficiency corrections are available, and that accurate independent estimates of the scatter and random component in the sinogram are also available. The \( y_i \) are modeled as a set of independent Poisson random variables with mean \( \bar{y}_i \). Thus, the log-likelihood for the data is:

\[
\ln P(y|\lambda) = \sum_i (-\bar{y}_i + y_i \ln \bar{y}_i).
\]  

We will assume a homogeneous isotropic Markov random field model for the image, \( \lambda \), characterized by the Gibbs distribution:

\[
P(\lambda|\beta) = \frac{1}{Z(\beta)} \exp\{-\beta U(\lambda)\}
\]  

with Gibbs energy \( U(\lambda) \), partition function \( Z(\beta) \) and global hyperparameter \( \beta \). The partition function of the prior is given by

\[
Z(\beta) = \int_{\lambda} \exp\{-\beta U(\lambda)\} d\lambda.
\]  

Here we confine the Gibbs energy function to pairwise interactions:

\[
U(\lambda) = \sum_j \sum_{k \in N_j} \kappa_{jk}V(\lambda_j, \lambda_k),
\]  

where \( N_j \) denotes the set of eight nearest neighbors of pixel \( j \) with \( \kappa_{jk} \) unity for horizontal and vertical neighbors and \( 1/\sqrt{2} \) for diagonal neighbors. The specific potential functions, \( V(\cdot, \cdot) \), used in this work are as follows:

\[
V_1(\lambda_j, \lambda_k) = (\lambda_j - \lambda_k)^2, \quad V_2(\lambda_j, \lambda_k) = \begin{cases} \frac{1}{2}(\lambda_j - \lambda_k)^2, & \text{if } |\lambda_j - \lambda_k| \leq \delta \\ |\lambda_j - \lambda_k| - \frac{\delta}{2}, & \text{otherwise} \end{cases}, \quad V_3(\lambda_j, \lambda_k) = \frac{(\lambda_j - \lambda_k)^2}{\delta^2 + (\lambda_j - \lambda_k)^2}.
\]  

The MAP estimator reconstructs \( \lambda \) by maximizing over the log posterior density:

\[
\hat{\lambda} = \arg \max_{\lambda} \ln P(y|\lambda) - \beta U(\lambda).
\]  

The authors term this the generalized ML method. Direct solution of this problem is still difficult as maximization with respect to \( \beta \) requires evaluation of the partition function of the prior. In the comparison with our method below, we use a maximum pseudo-likelihood approach \[1\] when updating \( \beta \) in which we approximate the prior \( P(\lambda|\beta) \) as the product of conditional densities \( \prod_i P(\lambda_i|\lambda_j, j \in N_j, \beta) \). We refer to this as the generalized maximum pseudo-likelihood (GMPL) method in the following.

\[10, 12\]
2.3 Mode Field Approximated ML

The intractability of true ML estimation of $\beta$ is due to the complexity and dimensionality of the joint density $P(y, \lambda | \beta)$ - it is essentially impossible to compute the marginal density in (8) for each new PET data set. Solution is possible if we can introduce a suitable simplifying approximation to the densities involved. One approach to this problem is to approximate the multidimensional densities as the product of separable functions. The complicated multi-dimensional integrals involved in computing the marginal density in (8), partition functions, or moments, can then be approximated with a product of one dimensional integrals with respect to these separable densities.

Approximating Gibbs distributions using separable functions is the basis for the mean field theory in classical statistical mechanics [2]. The basic idea is to replace the influence of statistical fluctuations of the neighbors of each pixel site by their means, where the means are computed with respect to the approximated (reference) field. Mean field approaches have previously been applied to surface reconstruction [5] and image segmentation [10] and several other image processing problems. Here we use a mean field like approximation to maximize the log of the marginalized likelihood $P(y | \lambda)$. This maximization requires evaluation of the derivatives of the prior and posterior partition functions $Z(\beta)$ and $Z(y, \beta)$. Consequently, we use two reference mean fields, one for the prior and one for the posterior.

In the case of the prior, we use the following mean field approximation of the partition function (4):

$$Z(\beta) \approx \prod_j \int_{\lambda_j} \exp(-\beta \sum_{k \in N_j} V(\lambda_j, \lambda_k^{PR})) d\lambda_j,$$

(15)

where $\lambda_k^{PR}$ is the mean of the prior reference field at pixel $k$. Since $Z(\beta)$ is not a function of the data and the original density is homogeneous and isotropic, $\lambda_k^{PR}$ is a constant for all $k$ and equal to the midpoint of the range of the $\lambda_k$s.

Similarly, the posterior partition function (4) is approximated as

$$Z(y, \beta) \approx \prod_j \int_{\lambda_j} \exp\{\ln P(y | \lambda_j, \lambda_k^{PO}) - \beta \sum_{k \in N_j} V(\lambda_j, \lambda_k^{PO})\} d\lambda_j,$$

(16)

where $\lambda_k^{PO}$ denotes the mean of the posterior reference field. The notation $N_j$ denotes the set of all pixels excluding site $j$.

Substituting these approximations into (10), (11) and then using (12), we find that the mean field approximated MLE of the hyperparameter $\beta$ is a root of the equation

$$\sum_j \int_{\lambda_j} U_j^{PO}(\lambda_j) \exp\{\ln P(y | \lambda_j, \lambda_k^{PO}) - \beta U_j^{PO}(\lambda_j)\} d\lambda_j$$

$$= \sum_j \int_{\lambda_j} U_j^{FR}(\lambda_j) \exp(-\beta U_j^{FR}(\lambda_j)) d\lambda_j,$$

(17)

where $U_j^{PR}(\lambda_j) = \sum_{k \in N_j} V(\lambda_j, \lambda_k^{PR})$ and $U_j^{PO}(\lambda_j) = \sum_{k \in N_j} V(\lambda_j, \lambda_k^{PO})$ are the local energy functions of the reference prior and posterior fields, respectively.

In the PET application, we are interested in computing a MAP estimate of the image, i.e., the mode of the posterior density $P(y | \beta)$. Rather than also computing the mean of the posterior reference field, we use a mode-field reference. This mode is computed using an iterative MAP estimation procedure. In cases where the posterior density is unimodal and symmetric, mean and mode fields are equivalent. Such is the case for Gaussian data with the convex potential functions $V_1$ and $V_2$ defined in (6). For PET data, the Poisson likelihood is asymmetric and the two methods are not equivalent. However, for relatively high count data, the Poisson likelihood is well approximated by a symmetric Gaussian function (i.e., the log-likelihood function $\ln P(y | \beta)$ is approximately symmetric). We therefore anticipate only minor differences between the mode-field and mean-field approximations in this case.

2.4 Numerical Considerations

We use a Newton-Raphson root finding procedure to solve the approximated likelihood equation (17). Numerical integration was performed using an adaptive quadrature method [9]. The evaluation of the integrands on the left hand side of (17) require a large number of forward and backward projections of the current estimate of the image for just a single iteration of the Newton-Raphson root finding procedure. This is computationally demanding and impractical. Instead, a quadratic approximation to the Poisson is used that greatly reduces the computational cost since only one forward and two backward projections of the current image are necessary for each Newton-Raphson root finding iteration. In the case of low count data, there is greater asymmetry and the quadratic approximation becomes poor. In the low count case, we could either use the true Poisson density, or perform estimation only with respect to the likelihood $p(y | \beta)$ computed over those higher count data for which the quadratic approximation is good. The results presented below are for the high count case only.

The Newton-Raphson root finding procedure is combined with a MAP image estimator to perform simultaneous image reconstruction and hyperparameter estimation. In the results presented below, we perform several iterations of the gradient projection preconditioned conjugate gradient method described in [8] over the posterior density with fixed $\beta$. The resulting image is then used as the approximation to the mode field in the Newton-Raphson root finding procedure. After one Newton-Raphson iteration for updating the hyperparameter, we fix $\beta$ and again update the image. This procedure is repeated until effective convergence of both the MAP estimator and $\beta$. While we are not able to prove convergence of the method, the method appears well behaved and we have not yet encountered an example which does not converge. We refer to this method as mode-field approximated maximum likelihood (MFAML).
3 Experiments and Results

We have compared the performance of the method outlined above (MFAML) with a generalized maximum pseudo-likelihood (GMPL) technique and the method of moments (MOM) [4] with the statistic $M(y)$ of the same form as the Gibbs energy function of the prior, computed over the sinogram $y$ with 8-nearest neighbor interactions.

3.1 Application to Image Restoration

We performed Monte Carlo studies for image restoration as follows. For each value of the hyperparameter, fifty sample images were drawn from the quadratic prior ($V_i$ in (6)) using a stochastic (Gibbs) sampler. Each sampled image was then blurred by a $3 \times 3$ kernel:

\[
\begin{pmatrix}
0.001 & 0.028 & 0.001 \\
0.028 & 0.884 & 0.028 \\
0.001 & 0.028 & 0.001
\end{pmatrix}.
\]

(18)

Pseudo-random Gaussian noise was generated to contaminate each of these images. The hyperparameters were then estimated using each of the three methods listed above. Ensemble averaging over the 50 hyperparameter estimates was used to compute bias and variance of the estimators. The results shown in Table 1 are for additive Gaussian noise $N(0, 16)$ on an image with intensities in the range (0, 100). Note that in all cases, the MFAML method outperforms both of the other techniques. The differences are very clear for the cases where $\beta$ is large, corresponding to the case of very smooth images. Although the results deteriorate for all three estimators, the performance of the MFAML method is markedly superior to MOM and GMPL.

3.2 Application to PET Image Reconstruction

We also generated a set of pseudo-random Poisson PET data using the full physical model in (1) for a 2D brain phantom - see [8] for more details of this simulation. We used the three priors defined in (6). Fifty independent data sets were generated for this phantom, and MAP images estimated for each set of data, for each of the priors and for a range of values of $\beta$. Since we do not know the true hyperparameter for the brain phantom, we generated a curve of average total squared error in the reconstructed image versus the range of hyperparameter values, see Figure 1. We then ran simultaneous reconstruction and hyperparameter estimation for each of the data sets and for each of the priors, and marked on Figure 1 the mean value of $\beta$ obtained. Note that the hyperparameter value that the MLAML method produced lies very close to the smallest average squared-error. The total number of counts used in this test was about 1.6 million per data set.

4 Conclusions

We have presented an approximate maximum likelihood estimator for the smoothing hyperparameter for use in image reconstruction and restoration problems. The results presented appear to indicate relatively good performance in terms of bias and variance and the method was observed to converge in all tests performed. A more detailed description of the method, its properties and performance is available in [11].

The mean field approximation described above was chosen heuristically. For a restricted class of Gibbs distributions, first order perturbation theory can be used to find an optimized mean field approximation of the partition function [2]. In [11] we apply this approximation to the special case of quadratic Hamiltonians.

References

Table 1: Monte Carlo Test of mode field approximated maximum likelihood (MFAML), generalized maximum pseudo-likelihood (GMPL), and the method of moments (MOM) for hyperparameter estimation. Mean, percentage bias and variance were computed using 50 independent images drawn from the prior using a Gibbs sampler as described in Section 3.1. The prior had the quadratic potential function defined in (6). (* indicates that the algorithm failed to reach a solution.)

<table>
<thead>
<tr>
<th>True $\beta$</th>
<th>0.0004</th>
<th>0.0010</th>
<th>0.0040</th>
<th>0.0100</th>
<th>0.0400</th>
<th>0.100</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMPL Mean</td>
<td>4.134e-4</td>
<td>1.093e-3</td>
<td>7.742e-3</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>GMPL Bias (%)</td>
<td>3.35%</td>
<td>9.30%</td>
<td>93.6%</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>GMPL STD (%)</td>
<td>1.74%</td>
<td>1.60%</td>
<td>6.49%</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>MOM Mean</td>
<td>4.039e-4</td>
<td>1.012e-3</td>
<td>4.175e-3</td>
<td>1.154e-2</td>
<td>0.0775</td>
<td>0.3565</td>
</tr>
<tr>
<td>MOM Bias (%)</td>
<td>0.97%</td>
<td>1.20%</td>
<td>4.37%</td>
<td>15.4%</td>
<td>93.7%</td>
<td>257%</td>
</tr>
<tr>
<td>MOM STD (%)</td>
<td>2.13%</td>
<td>3.21%</td>
<td>10.9%</td>
<td>31.0%</td>
<td>340%</td>
<td>380%</td>
</tr>
<tr>
<td>MFAML Mean</td>
<td>3.932e-4</td>
<td>9.777e-4</td>
<td>3.794e-3</td>
<td>9.889e-3</td>
<td>0.0306</td>
<td>0.0579</td>
</tr>
<tr>
<td>MFAML Bias (%)</td>
<td>-1.70%</td>
<td>-2.23%</td>
<td>-5.14%</td>
<td>-11.1%</td>
<td>-23.4%</td>
<td>-42.1%</td>
</tr>
<tr>
<td>MFAML STD (%)</td>
<td>1.53%</td>
<td>1.45%</td>
<td>1.63%</td>
<td>2.58%</td>
<td>7.08%</td>
<td>16.3%</td>
</tr>
</tbody>
</table>

Figure 1: PET experiment: The top row shows the three different forms of the prior potential functions as defined in (6). The middle row shows sample PET images reconstructed, with simultaneous MFAML hyperparameter estimation, from pseudo-random data using each of the three priors. The bottom row shows the total squared error in the estimated MAP images as a function of $\beta$. These errors were averaged over 50 independent realizations as described in Section 3.2. The star indicates the mean value of $\beta$ obtained using the MFAML method, note its close proximity to the minimum squared error.