Bayesian SPECT Image Reconstruction with Scale Hyperparameter Estimation for Scalable Prior *

Antonio López[1], Rafael Molina[2], and Aggelos K. Katsaggelos[3]

[1] Universidad de Granada. Departamento de Lenguajes y Sistemas Informáticos, 18071 Granada, Spain
alopez@ugr.es

[2] Universidad de Granada. Departamento de Ciencias de la Computación e I.A. , 18071 Granada, Spain
rms@decsai.ugr.es

[3] Northwestern University. Department of Electrical and Computer Engineering, Evaston, Illinois 60208-3118
aggk@ece.nwu.edu

Abstract. In this work we propose a new method to estimate the scale hyperparameter for convex priors with scalable energy functions in Single Photon Emission Computed Tomography (SPECT) image reconstruction problems. Within the Bayesian paradigm, Evidence Analysis and circulant preconditioners are used to obtain the scale hyperparameter. The proposed method is tested on synthetic SPECT images using Generalized Gaussian Markov Random Fields (GGMRF) as scalable prior distributions.

1 Introduction

SPECT is a technique used in Nuclear Medicine to take projection data by a gamma-camera following an orbit around the patient's body, at regularly spaced angles. A reconstructed SPECT image is the discrete representation of a cross section of the isotope distribution within the patient.

When Bayesian methods are applied to the reconstruction of SPECT images, the parameters of the prior model incorporating the expected structure in the image need to be selected. These parameters are known as hyperparameters and their selection has been a serious limitation to the use of statistical methods in medical reconstruction problems. Therefore, reliable automatic methods for the selection of the hyperparameters are essential to obtain correct reconstructions.

Several works have been published on the hyperparameter estimation problem in SPECT image reconstruction, see [7, 11] and references therein. In [8, 9] we used diagonal preconditioning methods to estimate the unknown hyperparameters. Circulant and shift-variant preconditioners were applied to GGMRF priors in [10], where the best reconstructions and execution times were obtained

* This work has been partially supported by CICYT project TIC2000-1275

with a circulant preconditioner. In this paper we propose a new circulant approximation to provide an estimation method of the scale hyperparameter that can be used on any prior with *scalable* energy function (see [3] for a justification on the use of scalable priors).

The rest of the paper is organized as follows. In section 2 we describe the degradation and image models and the Evidence Analysis within the Bayesian paradigm. Section 3 describes the proposed estimation method, as well as, the preconditioning method used. Experimental results are presented in section 4. Section 5 concludes the paper.

2 Hierarchical Bayesian paradigm and Evidence Analysis

Within the Bayesian paradigm, the reconstruction of the original images X, denoted by $\hat{X}(\theta)$, is selected as:

$$\hat{X}(\theta) = \arg\max_X P(Y|X)P(X|\theta) \,, \tag{1}$$

where θ is a hyperparameter vector, $P(X|\theta)$ is the prior distribution and $P(Y|X)$ models the process to obtain the data Y from the real underlying image X.

The Hierarchical Bayesian paradigm first defines the distributions $P(Y|X)$ and $P(X|\theta)$. Next, a distribution $P(\theta)$ for the hyperparameters is defined and the joint distribution $P(\theta, X, Y)$ is formed. The Evidence Analysis performs then the following steps to estimate the hyperparameter vector and reconstruct the image:

1. $P(\theta, X, Y)$ is integrated over the whole image space X to obtain $P(\theta, Y)$ and

$$\hat{\theta} = \arg\max_\theta P(\theta, Y) = \arg\max_\theta P(\theta|Y) \,, \tag{2}$$

   is selected as the hyperparameter vector.
2. The original image X is then obtained as

$$\hat{X}(\hat{\theta}) = \arg\max_X P(Y|X)P(X|\hat{\theta}) \,. \tag{3}$$

The degradation model for emission tomography is specified as a product of independent Poisson distributions:

$$P(Y|X) = \prod_{i=1}^{M} \frac{(\sum_{j=1}^{N} A_{i,j}x_j)^{y_i} \exp\{-\sum_{j=1}^{N} A_{i,j}x_j\}}{y_i!} \,, \tag{4}$$

where M is the number of detectors, N is the number of pixels and A is the system matrix. $A_{i,j}$ is the probability that an emitted photon from pixel j reaches detector i.

The prior models we use are convex Markov Random Fields (MRF) with *scalable* energy function. Their density functions have the following form:

$$P(X) = \frac{1}{Z(\sigma, p)} \exp\{-\sum_{i,j\in\mathcal{N}} V(x_i, x_j, \sigma, p)\} = \frac{1}{Z(\sigma, p)} \exp\{-\frac{1}{\sigma}\sum_{i,j\in\mathcal{N}} U(x_i, x_j, p)\}, \tag{5}$$

where V is the potential function which depends on two hyperparameters σ and p, that is, $\theta = (\sigma, p)$, U is the energy function and Z is the partition function. The elements x_i and x_j are neighbouring pixels and $\mathcal{N}$ is the set of all neighbouring pixel pairs. The scale hyperparameter σ determines the overall smoothness of the reconstruction, and p is called the shape parameter. The energy function is *scalable* if for all $X \in \mathcal{R}^N$ and $\alpha > 0$ we have:

$$U(\alpha X, p) = \alpha^p U(X, p). \tag{6}$$

From above it follows that the partition function can be expressed as:

$$Z(\sigma, p) = \sigma^{N/p} Z(1, p). \tag{7}$$

Equation (7) implies that the partition function, for any scalable prior model, is tractable with respect to the scale hyperparamenter σ. Hence, a general method for the estimation of σ can be obtained.

3 Scale Hyperparameter Estimation

We now proceed to estimate the scale hyperparameter for scalable priors in image reconstruction problems. We assume here that $P(\theta) \propto$ const.

In order to solve Eq. (3), we define the following function $M(X, Y|\theta)$

$$M(X, Y|\theta) = \log P(X|\theta) + \log P(Y|X) =$$
$$-\log Z(\sigma, p) - \frac{1}{\sigma} \sum_{i,j \in \mathcal{N}} U(x_i, x_j, p) + \sum_{i=1}^{M} \left(\sum_{j=1}^{N} (-A_{i,j} x_j) + y_i \log \left[\sum_{j=1}^{N} A_{i,j} x_j \right] \right), \tag{8}$$

and then calculate $P(\theta, Y)$ using

$$P(\theta, Y) \propto P(\theta) \int_X \exp\{M(X, Y|\theta)\} dX \; . \tag{9}$$

This integral cannot be evaluated analytically and therefore we resort to Gaussian quadrature approximation. Using Taylor series expansion, we expand $M(X, Y|\theta)$ around the MAP estimate, $\hat{X}(\theta)$, up to second order terms. Hence, we have the following approximation of $P(\theta, Y)$:

$$P(\theta, Y) \propto \exp\left[M(\hat{X}(\theta), Y|\theta) \right] \left| G(\hat{X}(\theta)) + \frac{1}{\sigma} F(\hat{X}(\theta)) \right|^{-1/2} , \tag{10}$$

where the (i, j)th elements of the matrices $G(\hat{X}(\theta))$ and $F(\hat{X}(\theta))$ are given by:

$$G_{i,j}(\hat{X}(\theta)) = \sum_{k=1}^{M} y_k \frac{A_{k,i} A_{k,j}}{\left(\sum_{l=1}^{N} A_{k,l} \hat{x}_l(\theta) \right)^2} , \tag{11}$$

$$F_{i,j}(\hat{X}(\theta)) = \begin{cases} \sum_{k \in \mathcal{N}_i} \frac{\partial^2 U(\hat{x}_i(\theta), \hat{x}_k(\theta), p)}{\partial x_i^2} & i = j \\ -\frac{\partial^2 U(\hat{x}_i(\theta), \hat{x}_j(\theta), p)}{\partial x_i \partial x_j} & i \neq j,\ j \in \mathcal{N}_i \\ 0 & \text{otherwise .} \end{cases} \tag{12}$$

Using Eq. (10) in Eq. (2) we obtain:

$$\sigma = \frac{p}{N} \sum_{i,j \in \mathcal{N}} U(\hat{x}_i(\theta), \hat{x}_j(\theta), p) + \frac{p}{2N} trace \left[\left(G(\hat{X}(\theta)) + \frac{1}{\sigma} F(\hat{X}(\theta)) \right)^{-1} F(\hat{X}(\theta)) \right] . \tag{13}$$

We can now use the following iterative procedure to estimate the reconstruction and the scale hyperparameter. At each step k, we proceed as follows:

1. Given a previously obtained image $\hat{X}^{k-1}$ and a previous estimate of the scale hyperparameter $\hat{\sigma}^{k-1}$, a new value $\hat{\sigma}^k$ is calculated using Eq. (13).
2. A newly estimated image $\hat{X}^k$, is obtained by iterating once a MAP algorithm for $\hat{\sigma}^k$ and $\hat{X}^{k-1}$.

We note that the estimation of σ involves the calculation of $(G(\hat{X}(\theta)) + \frac{1}{\sigma} F(\hat{X}(\theta)))^{-1}$. This inversion is a computationally intensive problem. We use preconditioning to approximate Hessian matrices in order to simplify their inversion.

In this work circulant preconditioning methods are used to approximate the matrix $(G(\hat{X}(\theta)) + \frac{1}{\sigma} F(\hat{X}(\theta)))$. The circulant preconditioner requires the potential function to be convex, symmetric, twice-differentiable with respect to X and have bounded second derivatives (see [10]). This kind of preconditioner improves the approximation of the matrix $(G(\hat{X}(\theta)) + \frac{1}{\sigma} F(\hat{X}(\theta)))$ with diagonal preconditioners and it is of interest since the Discrete Fourier Transform (DFT) can be applied to solve the matrix inversion.

We start by observing that $G(\hat{X}(\theta))$ can be expressed as

$$G(\hat{X}(\theta)) = A^t W(\hat{X}(\theta)) A \ , \tag{14}$$

where $W(\hat{X}(\theta))$ is a diagonal matrix with diagonal entries

$$W_{i,i}(\hat{X}(\theta)) = \frac{y_i}{\left(\sum_{l=1}^{N} A_{i,l} \hat{x}_l(\theta) \right)^2} . \tag{15}$$

We then apply the approximation introduced in [5] to the Fisher information term, which produces:

$$A^t W(\hat{X}(\theta)) A \approx D_{(g)}(\hat{X}(\theta)) A^t A D_{(g)}(\hat{X}(\theta)), \tag{16}$$

where $D_{(g)}(\hat{X}(\theta))$ is a diagonal matrix with diagonal entries

$$D_{(g)j,j}(\hat{X}(\theta)) = \sqrt{\frac{\sum_{i=1}^{M} A_{i,j}^2 W_{i,i}(\hat{X}(\theta))}{\sum_{i=1}^{M} A_{i,j}^2}} . \tag{17}$$

Following [4], the product A^tA is now approximated by a block-circulant matrix G_c. The kernel of G_c is obtained as follows:

1. First we calculate $A^tA\epsilon_j$, where ϵ_j represents a unit vector centered with respect to the image.
2. Then, the kernel obtained in the previous step, is approximated by a shift invariant symmetrical blurring function.

We also apply the approximation proposed in [5] to the matrix $F(\hat{X}(\theta))$, obtaining:

$$F(\hat{X}(\theta)) \approx D_{(f)}(\hat{X}(\theta))(I - \phi C)D_{(f)}(\hat{X}(\theta)) \ , \tag{18}$$

where the diagonal elements of $D_{(f)}(\hat{X}(\theta))$ have the form:

$$D_{(f)j,j}(\hat{X}(\theta)) = \sqrt{\sum_{k \in \mathcal{N}_j} \frac{\partial^2 U(\hat{x}_j(\theta), \hat{x}_k(\theta), p)}{\partial x_j^2}} \ , \tag{19}$$

and $(I - \phi C)^{-1}$ is a covariance matrix, with $C_{i,j} = 1$ for two neighbouring pixels and for the 8-point neighborhood system we are using, ϕ is just less than $1/8$).

The matrices $D_{(g)}(\hat{X}(\theta))$ and $D_{(f)}(\hat{X}(\theta))$ are approximated by constant diagonal matrices $\sqrt{\alpha(\hat{X}(\theta))}I$ and $\sqrt{\beta(\hat{X}(\theta))}I$ respectively, where:

$$\sqrt{\alpha(\hat{X}(\theta))} = \frac{1}{N}\sum_{j=1}^{N} D_{(g)j,j}(\hat{X}(\theta)) \text{ and } \sqrt{\beta(\hat{X}(\theta))} = \frac{1}{N}\sum_{j=1}^{N} D_{(f)j,j}(\hat{X}(\theta)). \tag{20}$$

Using the above approximations we now have,

$$\begin{aligned} & trace\left[\left(G + \frac{1}{\sigma}F\right)^{-1} F\right] \\ & \approx trace\left[\left(D_{(g)}G_cD_{(g)} + \frac{1}{\sigma}D_{(f)}(I - \phi C)D_{(f)}\right)^{-1} D_{(f)}(I - \phi C)D_{(f)}\right] \\ & \approx trace\left[\left(\alpha G_c + \frac{1}{\sigma}\beta(I - \phi C)\right)^{-1} \beta(I - \phi C)\right] \ . \end{aligned} \tag{21}$$

where we have removed the dependency of the matrices on $\hat{X}(\theta)$ for simplicity.

Finally, using the DFT we obtain the following expression:

$$trace\left[\left(G + \frac{1}{\sigma}F\right)^{-1} F\right] \approx \sum_{i=1}^{N} \frac{\Lambda_i}{\frac{\alpha}{\beta}\Gamma_i + \frac{1}{\sigma}\Lambda_i} \ , \tag{22}$$

where Γ_i and Λ_i are the (i)th elements of the DFT of G_c and $(I - \phi C)$, respectively.

Now, we use the described estimation method with GGMRF priors [2]. They are scalable and have the form:

$$P(X) = \frac{1}{Z(\sigma, p)} \exp\{-\frac{1}{p\sigma^p} \sum_{i,j \in \mathcal{N}} b_{i-j} |x_i - x_j|^p\} \ . \tag{23}$$

The potential function is convex when $p \geq 1$. For $p = 2$ we have the Gaussian distribution. We note that work has been devoted to the estimation of shape parameter p when this class of priors are used on $SPECT$ images ([11, 8]). In these papers it was found that the simultaneous estimation of σ and p was not feasible, resulting in values of $p < 1$ for which the potential function is not convex. Based on the literature on $GGMRF$ prior distributions, we fixed $p = 1.1$ in our experiments to estimate the scale hyperparameter.

When $p < 2$ the potential function has non bounded second derivative with differences $x_i - x_j = 0$. This is a problem for the circulant approach, since, as we have already mentioned, circulant preconditioner requires the potential function to be convex, symmetric, twice-differentiable with respect to X and have bounded second derivatives [10]. In order to overcome the problem, we have used the approximation suggested in [1] when dealing with wavelet image reconstructions using GGMRF prior models:

$$\sum_{i,j \in \mathcal{N}} b_{i-j} |x_i - x_j|^p \approx \sum_{i,j \in \mathcal{N}} b_{i-j} \left(\left(|x_i - x_j|^2 + \delta \right)^{p/2} - \delta^{p/2} \right) \ , \tag{24}$$

where $\delta > 0$ is a stabilization constant.

Using Eqs. (24) and (22) in Eq. (13), we obtain the following scale hyperparameter estimation:

$$\hat{\sigma}^p = \frac{1}{N} \sum_{i,j \in \mathcal{N}} b_{i-j} |\hat{x}_i - \hat{x}_j|^p + \frac{p}{2N} \sum_{i=1}^{N} \frac{\Lambda_i}{\frac{\hat{\alpha}}{\hat{\beta}} \Gamma_i + \frac{1}{\hat{\sigma}^p} \Lambda_i} \ , \tag{25}$$

where α and β are obtained from Eq. (20) with:

$$D_{(f)j,j} = \sqrt{\sum_{k \in \mathcal{N}_i} b_{j-k} \left((p/2 - 1) \left(|\hat{x}_j - \hat{x}_k|^2 + \delta \right)^{\frac{p}{2}-2} |\hat{x}_j - \hat{x}_k|^2 + \left(|\hat{x}_j - \hat{x}_k|^2 + \delta \right)^{\frac{p}{2}-1} \right)}$$

4 Experimental Results

We denote by C2 the described estimation method using circulant approximation and GGMRF, the previously proposed estimation methods using diagonal approximation by D1 [8] and C1 the circulant approximation used in [10]. The image update was obtained using the MAP algorithm proposed in [6].

The synthetic image used in the simulations is shown in Fig. 1 (a), with size 128×128 pixels and 128 angles with 128 bins simulated. We have used 8 independent sets of projections. The average estimations provided by the methods

(a) (b)

Fig. 1. (a) Original image. (b) Reconstruction.

are $\hat{\sigma}^{1.1} = 32.4$ (D1), $\hat{\sigma}^{1.1} = 34.3$ (C1) and $\hat{\sigma}^{1.1} = 34.5$ (C2). The reconstructions obtained with these values are visually indistinguishable (Fig. 1 (b) shows the obtained image with the diagonal approach). The mean squared error of the reconstructed images for different values of $\sigma^{1.1}$ was also calculated (it is not shown here for lack of space). The estimated values of the scale hyperparameter lie within the zone of the minimum of the curve.

Using as stopping criterion $\left|(\hat{\sigma}^{1.1})^k - (\hat{\sigma}^{1.1})^{k-1}\right| / (\hat{\sigma}^{1.1})^k \leq 0.0001$ the average number of iterations needed was: 149 (D1), 122 (C1) and 113 (C2).

We have computed $Bias^2 + Variance$ on Regions of Interest (ROI) to test the quality of the reconstruction. The ROIs are shown in Figure 3 (the first ROI is uniform and the others include edges). These quantities are defined as:

$$Bias^2 \equiv \frac{1}{ML} \sum_{i=1}^{M} \left(\sum_{l=1}^{L} \frac{\hat{x}_i^l - x_i(true)}{x_i(true)} \right)^2 \tag{26}$$

$$Variance \equiv \frac{1}{ML} \sum_{i=1}^{M} \sum_{l=1}^{L} \left(\frac{\hat{x}_i^l - \bar{x}_i}{\bar{x}_i} \right)^2 , \tag{27}$$

where M is the number of pixels in each ROI, L is de number of samples of y, $\hat{x}_i^l$ is the estimated activity in pixel i for the lth sample, $x_i(true)$ is the original pixel value and $\bar{x}_i$ is the estimated mean activity at pixel i from the L samples. The ROIs are shown in Fig. 2 in white. The obtained $Bias^2 + Variance$ values of with our approximations are shows in Table 1. We observe that C2 is slightly better than D1 and C1 since it provides the smaller values of $Bias^2 + Variance$.

5 Conclusions

In this paper we have concentrated on the estimation of the scale hyperparameter σ for prior models with scalable energy function in SPECT image reconstruction problems. The application of circulant preconditioner to estimate this unknown

(a) ROI 1 (b) ROI 2 (c) ROI 3 (d) ROI 4

Fig. 2. Regions of Interest

	ROI 1	ROI 2	ROI 3	ROI 4
D1	0.03951	0.03616	0.08670	0.01987
C1	0.03923	0.02959	0.08392	0.01959
C2	0.03917	0.02957	0.08380	0.01958

Table 1. $Bias^2 + Variance$

hyperparameter has been described. Using prior models such as GGMRF, we have found that the estimation method produces satisfactory reconstruction and the circulant preconditioners exhibit a better convergence rate.

References

1. Belge, M., Kilmer, M.E., Miller, E.L.: Wavelet Domain Image Restoration with Adaptative Edge-Preserving Regularization. IEEE Tr. Im. Proc. **9** (2000) 597–608
2. Bouman, C.A., Sauer, K.: A Generalized Gaussian Image Model for Edge-Preserving Map Estimation. IEEE Tr. Im. Proc. **2** (1993) 296–310
3. Brette, S., Idier, J., Mohammad-Djafari, A.: Scale Invariant Markov Models for Bayesian Inversion of Linear Inverse Problems. Maximum Entropy and Bayesian Methods, Kluwer Academic (Skilling, J., Sibisi S., eds.) (1996) 199-212.
4. Clinthorne, N.H., Pan, T., Chiao, P., Rogers, W.L., Starnos, J.A.: Preconditioning Methods for Improved Convergence Rates in Iterative Reconstructions. IEEE Tr. Me. Im. **12** (1993) 78–83
5. Fessler, J.A., Rogers, W.L.: Spatial Resolution Properties of Penalized-Likelihood Image Reconstruction Methods. IEEE Tr. Im. Proc. **5** (1996) 1346–1358
6. Hebert, T.J., Leahy, R.: Statistic-Based MAP Image Reconstruction for Poisson Data Using Gibbs Priors. IEEE Tr. on Sig. Proc. **40** (1992) 2290–2303
7. Lalush, D.S., Frey, E.C., Tsui, B.H.: Fast Maximum Entropy Approximation in SPECT Using the RBI-MAP Algorithm. IEEE Tr. Med. Im. **19** (2000) 286–294
8. López, A., Molina, R., Katsaggeloss, A.K.: Hyperparamenter Estimation for Emission Computed Tomography Data. Proc. of IEEE Intl. Conf. on Image Processing, Vol. 2 (1999) 667–680
9. López, A., Vega, M., Molina, R., Katsaggeloss, A.K., Rodríguez, A., López, J.M., Llamas, J.M.: Bayesian Reconstructions of SPECT Images, an Aide in Nuclear Medicine Diagnosis. Submitted to Tr. Med. Im. (2001)
10. López, Molina, R., Katsaggeloss, A.K.: Scale Hyperparameter Estimation for GGMRF Prior Models with Application to SPECT Images. Proc. of IEEE Intl. Conf. on Digital Signal Processing, Vol. 2 (2002) 521–524
11. Saquib, S.S., Bouman, C.A., Sauer, K.: Parameter Estimation for Markov Random Fields with Application to Bayesian Tomography. IEEE Tr. Im. Proc. **7** (1998) 445–448