

# Adaptively altering hyper-parameter for Improved Reconstruction in PET

P. P. Mondal, K. Rajan, Senior Member, IEEE, and L. M. Patnaik, Fellow, IEEE

**Abstract**—It is well known that as the iterations of the maximum likelihood algorithm are continued, density estimates become more and more noisy. In situations where some prior knowledge about the object distribution is available, one may utilize such information in the reconstruction procedure for improving the reconstruction. Fixed prior based image reconstruction process produces over-smooth reconstruction due to the penalizing nature of the potential. As the reconstruction process builds up, image properties like smoothness, frequency content etc. change and hence fixed prior based image reconstruction process may not serve the purpose. A new prior is proposed which is capable of producing improved reconstruction over those obtained by existing fixed prior based Bayesian algorithms. These are termed as dynamic priors, which unlike other priors modify itself recursively according to the noise level in the reconstruction. It is found that inclusion of prior knowledge in the reconstruction algorithm results in local minimums. In the present approach, appropriate prior energy is supplied to the estimate to overcome local minimums by shedding Gibbs hyperparameter and subsequently the effect of prior is removed recursively as the estimate approaches global minimum.

**Index Terms**—Hyperparameter Scheduling, Maximum A-Posteriori, Maximum Likelihood, Positron Emission Tomography, Dynamic Priors, Local Minimum.

## I. Introduction

**P**OSITRON emission tomography (PET) is an imaging technique for imaging brain activities like, motor action, pain etc. when different stimulus is induced artificially in the brain. These studies require injection or inhalation of radio-compounds in the body. Proportional to the physiological activity of the region, cells concentrate these radio-compounds in different proportions. These radio-compounds emit sub-atomic particles which gets detected in a ring detector system surrounding the object under study. The aim is the reconstruction of an accurate map of spatial and temporal distribution of the radio-pharmaceuticals from the detected counts.

Image reconstruction techniques plays a vital role in positron emission tomography. Frequently used image

This work was supported by Council of Scientific and Industrial Research, New Delhi, INDIA.

P.P.Mondal is with Department of Physics, Indian Institute of Science, Bangalore 560012, India (corresponding author); e-mail: partha@physics.iisc.ernet.in

K. Rajan, is with Department of Physics, Indian Institute of Science, Bangalore 560012, India. e-mail: rajan@physics.iisc.ernet.in

L. M. Patnaik, is with Department of computer Science and Automation, Indian Institute of Science, Bangalore 560012, India. e-mail: lalit@micro.iisc.ernet.in

reconstruction techniques are convolution backprojection (CBP) [1], maximum likelihood (ML) [2] and maximum a-posteriori (MAP) [3][4]. MAP-reconstruction technique has advantage over all other techniques because of its ability to incorporate prior knowledge in the reconstruction process. Prior knowledge helps in producing smooth looking reconstructed images which are more appealing. This is where MAP-estimation out-performs CBP and ML estimation. It is shown by Girodias et. al. [5] that inclusion of prior knowledge in the reconstruction technique results in the creation of local minimum in addition to global minimum in the posterior energy landscape. Since MAP-algorithm is iterative in nature, the task is to reach global minimum iteratively with minimal number of iterations. Girodias et. al. have used simulated annealing technique to overcome local minimums and speeded up the annealing process by executing it parallelly to gain the benefit of simulated annealing for real time image reconstruction [5]. In this paper an algorithmic approach is proposed for reaching global minimum.

## II. Image Reconstruction in Bayesian Domain

Image reconstruction in Bayesian domain has several advantages over ML and CBP. The biggest benefit is the inclusion of desired image properties in the reconstruction algorithm via prior knowledge. Proper specification of prior knowledge is crucial since this may lead to over-smoothing of the estimates. Inclusion of prior results in the creation of local minimums in the posterior energy landscape. Once the estimate gets trapped in the local minima well, further iteration doesnot help in producing better estimates. Hence at each iterative stage, appropriate Gibbs energy should be provided to the estimate to ensure its smooth travel towards global minimum. So, the sole objective is the determination of that estimate  $\hat{\lambda}$  which minimizes the posterior energy.

For PET, the log-likelihood function  $L(\lambda)$  is given by [2],

$$L(\lambda) = \prod_{j=1}^M \left[ \frac{\exp(-\sum_{i=1}^N \lambda_i p_{ij}) (\sum_{i=1}^N \lambda_i p_{ij})^{y_j}}{y_j!} \right] \quad (1)$$

Image field is assumed as Markov random field (MRF) and by Hammerseley-Clifford theorem [6], image  $\lambda$  is characterized by Gibbs distribution [6],

$$P(\lambda) = \frac{1}{Z} \exp\left[-\frac{1}{\beta} \sum_{i,(j>i)} \sum_{j \in N_i} V(\lambda_i, \lambda_j)\right] \quad (2)$$

where,  $Z$  is the normalizing constant for the distribution,  $\beta$  is the Gibbs hyper-parameter,  $U(\lambda)$  is Gibbs energy,  $N_i$  are the nearest neighbors of pixel  $i$  and  $V(\lambda_i, \lambda_j)$  is the potential at the site containing the subset  $N_i$ .

Posterior energy  $E(\lambda/y, \beta)$  is given by,

$$E(\lambda/y, \beta) = E(y/\lambda) + E(\lambda/\beta) \quad (3)$$

where,  $E(y/\lambda)$  and  $E(\lambda/\beta)$  are respectively the likelihood and prior energy.

Writing explicitly, the posterior energy becomes,

$$E(\lambda/y, \beta) = \sum_{j=1}^M \left[ \sum_{i=1}^N \lambda_i p_{ij} - y_j \log \left( \sum_{i=1}^N \lambda_i p_{ij} \right) + \log(y_j!) \right] + \frac{1}{\beta} \sum_{i,(j>i)} \sum_{j \in N_i} V(\lambda_i, \lambda_j) \quad (4)$$

Seeking the minimization of posterior energy function is equivalent to the maximization of MAP-function,

$$\max_{\lambda \geq 0} \log P(\lambda/y) = l(\lambda) + \log P(\lambda) \quad (5)$$

where,  $l(\lambda)$  is the log-likelihood function.

Solution for eqn.(5) is very difficult due to the complicated nature of the prior. Green [7] has proposed one step late (OSL) approximation for an iterative update to the imposed MAP-problem,

$$\hat{\lambda}_i^{k+1} = \frac{\hat{\lambda}_i^k}{\left[ \sum_{j=1}^M p_{ij} + \frac{1}{\beta} \sum_{j \in N_i} \left( \frac{\partial V(\lambda_i, \lambda_j)}{\partial \lambda_i} \right)_{\lambda_i = \hat{\lambda}_i^k} \right]} \times \sum_{j=1}^M \frac{y_j p_{ij}}{\sum_{i=1}^N \hat{\lambda}_i^k p_{ij}} \quad (6)$$

Proper choice of hyperparameter  $\beta$  plays a crucial role for quality MAP image reconstruction [3] [4]. A new methodology is presented in which the estimate is left unchanged as long as the measurement data satisfies  $\chi^2$ -test [8] and thereafter  $\beta$  is changed according to the pre-defined hyperparameter schedule.

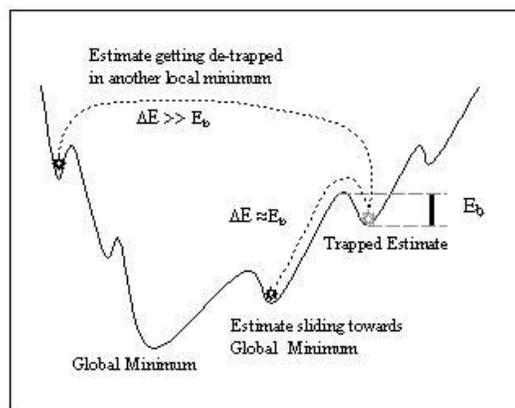


Fig. 1. Schematic diagram of posterior energy landscape.

### III. Gibbs Hyperparameter Scheduling

Gibbs hyperparameter controls the amount of smoothness in the reconstructed image. Imbalance in the hyperparameter value results either in over-smoothing of the estimates or malfunctioning of MAP-algorithm. Due to noisy and incomplete measurement data in emission tomography, prior knowledge is incorporated in the reconstruction process. The energy function for MAP contains likelihood energy and an additional term called Gibbs energy.

When the prior is included in the reconstruction process, local minimums appear in the posterior energy landscape [5]. It becomes extremely difficult for the estimate to overcome once they are trapped in these local energy wells. Then the key is to alter the prior energy by updating hyperparameter after each MAP-iteration, such that appropriate prior energy is supplied to the estimate to overcome the local energy barrier and there-after iteratively removing the effect of prior as global minimum is reached. Similar technique is used in [9].

Non-smoothness in the image is the indicator of noise. With increasing iteration, the noise artifact becomes dominant and therefore requires more contribution from prior for noise suppression. Thus, noise can be viewed as an indicator of the presence of local minimum in the posterior energy landscape. As we approach towards the ML-estimate, more and more noise start appearing, confirming the presence of numerous local minimums. This forces proper selection of the hyperparameter  $\beta$ . Once the estimate gets trapped in the local minimum, the chance of its coming out entirely depends on the prior energy (see Fig.1). This is the reason why posterior distribution works better than likelihood distribution as a cost function. In the posterior energy landscape, prior energy provides the additional energy required for overcoming the local minima wells. Since hyper-parameter  $\beta$  tunes the Gibbs energy, a proper selection of  $\beta$  becomes crucial.  $\beta$  determines the amount of additional energy required to overcome the local energy barrier  $E_b$ . Smaller values of  $\beta$  means effectively more prior energy to overcome the local minima trap and possibly

getting de-trapped in another local minima (Fig.1) while larger  $\beta$  means inefficient energy to overcome local energy barriers.

In view of the above discussion the following three rules seem essential for MAP image reconstruction :

(1)  $\Delta E \ll E_b$  i.e, the prior energy is much lower than the barrier energy; the estimate remains trapped in the local minimum (see Fig. 1).

(2)  $\Delta E = E_b + \epsilon$  with  $\epsilon$  an infinitesimally small positive quantity; the prior energy is just greater than the barrier energy, then the estimate is ensured to move towards global minimum (see Fig. 1).

(3)  $\Delta E \gg E_b$  i.e, the prior energy is much greater than the barrier energy, so the estimate hops in local minimums (see Fig.1).

In the posterior energy landscape, we are looking for posterior energy minimization rather than minimization of likelihood energy. Inclusion of Gibbs prior in the image formation procedure facilitates the estimate to overcome local energy barriers and makes it easier for the estimate to slide towards global minimum. The success of this approach entirely depends upon the proper selection of  $\beta$  after each MAP-iteration. Hence Gibbs energy should be tuned accordingly so that the estimate strictly obeys rule(2). Higher Gibbs energy results in the violation of rule(2),and favors rule(3). In other words, choosing hyperparameter obeying rule (2) will lead the estimate towards global minimum.

The following smoothening schedules are proposed for recursively updating  $\beta$ ,

$$\beta = \beta_0 k \quad (7)$$

$$\beta = \beta_0 e^k \quad (8)$$

where,  $\beta_0$  is a constant and  $k$  represents the iteration number.

Due to ill-conditioned nature of the posed non-linear problem, large number of minimums appear in the posterior energy landscape. Prior energy helps in overcoming these local minimums and its effect diminishes as the estimate approaches global minimum. The above two smoothening schedule given by eqn.(10) and eqn.(11) are meant for this purpose only.

#### IV. Simulated Experimental Results on PET System

Implementation of the proposed algorithm is performed on a simulated PET system. The PET system consists of a ring detector system with 64 detectors and the object space is decomposed into  $64 \times 64$  square pixels. The object space is a square region inscribed within the circle of detectors. To exploit the eight-fold symmetry we have assumed the object to be bounded by a circle inscribed in this square

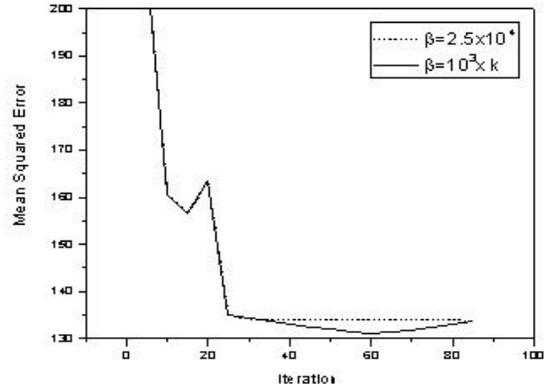


Fig. 2. Mean squared error versus iterations for (a)  $\beta = 2.5 \times 10^4$  (dotted line) and  $\beta = \beta_0 k, \beta_0 = 10^3$  (solid line)

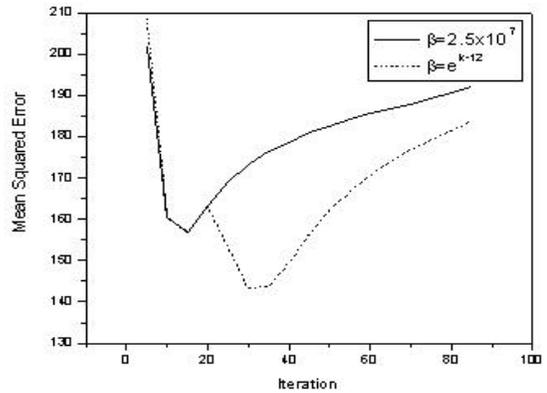


Fig. 3. Mean squared error versus iterations for (a)  $\beta = 2.5 \times 10^7$  (solid line) and  $\beta = \beta_0 e^k, \beta_0 = e^7$  (dotted line)

object space [10][11]. An electron-positron annihilation event occurring inside a pixel is governed by a number of physical phenomena such as attenuation, scattering, absorption and detector characteristics. All these physical processes have a bearing on the probability matrix. Each element of the probability matrix defines the probability of a photon getting detected in the detector  $j$  after emanating from the object pixel  $i$ . For simplicity, we assume that the probability of emission in pixel  $i$  which gets detected in a detector tube  $j$  is proportional to the angle  $\theta$  seen by the center of the pixel  $i$  into the detector tube  $j$  [1], given by  $p_{ij} = \frac{\theta_{ij}}{\pi}$ .

Before the reconstruction begins, the probability matrix  $\mathbf{P} = [p_{ij}]$ ,  $i = 1, \dots, N$  and  $j = 1, \dots, M$  should be computed and stored. It is found that using symmetry available in the system the data size of  $\mathbf{P}$  matrix can be reduced considerably [10]. For simulating measurement data, a Monte Carlo procedure is used [1][11]. Shepp and Logan mathematical phantom with 100,000 emissions has

been used for simulation studies.

Large number of test runs have been performed for correct selection of  $\beta$ -value that gives the lowest mean squared error (MSE), when compared with the original test phantom. The value is found to be  $2.5 \times 10^4$ . It should be noted from expression(9) that as  $\beta$  tends to infinity, the OSL algorithm becomes ML-algorithm. Better estimate can only be obtained if we choose  $\beta$  wisely. But correct selection of  $\beta$ -value is seldom possible in a single iterative process. To overcome this problem, two recursive schedules are proposed for altering  $\beta$  (see eqn.(10) and eqn.(11)). These recursive priors outperforms static prior based reconstruction techniques (see Fig.2 and Fig.3). These MSE plots further exemplifies the superiority of the proposed technique by showing that lower MSE values are possible when  $\beta$  is varied recursively using the proposed linear and exponential schedule. It is found that for  $\beta < 2.5 \times 10^4$  the MAP-estimate becomes unstable. Hence  $\beta$  was varied from  $2.5 \times 10^4$  to  $2.5 \times 10^7$  to ensure stable reconstruction. We have used  $\beta_0 = 10^3$  for linear scheduling and  $\beta_0 = e^7$  for exponential scheduling for stable reconstruction.

## V. Conclusion

In this work, we have presented a new approach for improved reconstruction by iteratively changing Gibbs hyper-parameter after each MAP-iteration, according to the noise intensity. We have argued that noise is the indicator of the presence of local minimums in the posterior energy landscape. Only when appropriate Gibbs energy is provided such that the estimate gains enough energy to overcome the local minimum, the estimate can move towards global minimum and subsequently results in noise free reconstruction. Results have shown improved reconstruction by scheduling hyper-parameter over fixed hyperparameter based MAP-reconstruction.

## Acknowledgment

The first author would like to thank Mr. Hemanth, research scholar at Indian Institute of Science for his help during the simulation experiment.

Also, he wishes thanks to Council of Scientific and Industrial research (CSIR), Government of India, for providing Junior Research fellowship. He declares that this work is partially supported by CSIR, New Delhi, India.

## References

- [1] L. A. Shepp and B. F. Logan, "The fourier reconstruction of a head section", IEEE Trans. Nuclear Science, NS-21, pp.21-43, 1974.
- [2] L.A. Shepp and Y. Vardi, " Maximum likelihood estimation for emission tomography ", IEEE Trans. on Medical Imaging ., MI-1: pp. 113-121, 1982
- [3] Z. Zhou, R. M. Leahy and J. Qi, " Approximate maximum likelihood hyperparameter estimation for Gibbs prior ", IEEE Trans. on Img. proc., Vol.6, No.6, pp.844-861, June, 1997.
- [4] T. Hebert and R. Leahy, " Statistic based MAP image reconstruction from Poisson data using Gibbs Priors ", IEEE Trans. on Sig. Proc., vol. 40, No.2, pp.2290-2303, Sept. 1992.

- [5] K. A. Girodias, H. H. Barrett and R. L. Shoemaker " Parallel simulated annealing for emission tomography ", Phys. Med. Biol., Vol. 36, No.7, pp. 921-938, 1991.
- [6] J. Besag, " On the statistical analysis of dirty pictures", JI. Of Royal Statistical Soc., B, vol. 48, No. 3, pp. 259-302, 1986
- [7] P.J. Green, "Bayesian reconstruction from emission tomography data using a modified EM algorithm", IEEE Trans. on Med. Img., vol.9, No.1, March, 1990.
- [8] E. Veklerov and J.Llacer, " Stopping rule for MLE algorithm based on statistical hypothesis testing ", IEEE Trans. on Med. Img., MI-6 : pp. 313-319, 1987.
- [9] M. Svensen, F. Kruggel and D Yves von Cramon, "Praobabilistic Modelling of Single Trial fMRI Data", IEEE Trans. Med. Img.,Vol.19, pp.25-35, Jan. 2000
- [10] L Kaufmann, " Implementing and accelerating the EM-algorithm for positron emission tomography ",IEEE Trans. Med. Imaging, vol. MI-6, pp.37-51, Mar. 1987.
- [11] N. Rajeevan, K. Rajgopal and G. Krishna. " Vector-Extrapolated fast maximum likelihood estimation algorithms for emission tomography ", IEEE Trans. on Medical Imaging ., Vol. 11, No.1, March 1992