A DISCRETE-CONTINUOUS BAYESIAN MODEL FOR EMISSION TOMOGRAPHY

Mame Diarra Fall †⋆, *Éric Barat* ⋆, *Claude Comtat* ‡, *Thomas Dautremer* ⋆, *Thierry Montagu* ⋆ *and Ali Mohammad-Djafari* †

† Laboratoire des Signaux et Systèmes, (UMR8506: CNRS–SUPELEC–Univ Paris-Sud), 91192 Gif-sur-Yvette, FRANCE.
⋆ Laboratoire de Modélisation, Simulation et Systèmes (LM2S), CEA-Saclay, 91191 Gif-sur-Yvette, FRANCE
‡ Service Hospitalier Frédéric Joliot, CEA/DSV/I2BM, 91401 Orsay, FRANCE

ABSTRACT

In this contribution, we propose a discrete-continuous reconstruction method for Positron Emission Tomography (PET). The goal is to reconstruct a continuous radiotracer activity distribution from a finite set of measurements (namely, the discrete projections of detected random emissions). Our approach can be viewed as an indirect density estimation problem, *i.e*, the problem of recovering a probability density function based on indirect observations. We cast the reconstruction problem in a *Bayesian nonparametric estimation* framework where regularization of the ill-posed inverse problem is achieved by putting a prior on the investigated radiotracer activity distribution.

We propose a hierarchical model and use it for the MCMC schemes to generate samples from the posterior activity distribution and compute its functionals (mean, standard deviation etc.). Results will illustrate the performances of the proposed method and we compare our approach to another Bayesian method, the maximum a posteriori estimation (MAP), which is based on a fully discrete-discrete problem formulation.

***Index Terms*—** Bayesian nonparametrics, indirect density estimation, MCMC sampling, Positron Emission Tomography

1. INTRODUCTION

An interesting feature of Emission computed Tomography is in producing 3D images that give information about the metabolic activity of an organ (the brain, for example). In Positron Emission Tomography (PET), the subject is injected with a molecule labelled with a positron emitter radionuclide. The emitted positrons travel a small distance before annihilating with an electron. This annihilation generates two gamma photons that fly-off at the speed of light in opposite directions. These photons are detected by rings of detectors around the subject. If two photons are detected within a short time interval, a coincidence event is recorded revealing that a positron emission occurred in the virtual line joining both detectors (called line-of-response (LOR)). Tomographic algorithms are used to reconstruct 3D images of the radioactivity distribution from the collected data sets.

Conventional image reconstruction algorithms in Emission Tomography are roughly divided into two categories: analytical and statistical methods. Analytical methods like filtered back-projection algorithms are based on the direct inversion of the X-ray transform [1] and address the reconstruction problem (at least theoretically) in continuous data and image spaces. They provide fast reconstructions, however they are based on over-simplified models of the physical processes and do not account for the stochastic nature of positron emission and photon detection. As a consequence, reconstructed images suffer from a lot of artifacts and significant noise. To overcome these deficiencies, statistical methods have been proposed to model more realistically the process of acquiring data by taking into account the random nature of the phenomenon. They are based on optimization methods, typically maximum likelihood approaches (ML) ([2], [3]). They usually provide reconstructed images with a higher signal-to-noise ratio than analytical methods but nevertheless have high noise due to ill-conditioned nature of the problem. This can be solved by adding a regularization term in the form of a priori density for the image using Bayesian approaches.When Bayesian methods are used in Emission Tomography, they are based on the maximum a posteriori approach (MAP) ([4], [5]). However, all these statistical methods require a discrete formulation of the radiotracer distribution in the space over which the reconstruction is needed, which is not natural because emission distribution is continuous.

The approach presented in this paper is a statistical one but is fundamentally different from those mentioned above. The continuous radioactivity distribution is directly reconstructed from the discrete data. Thus, this can be viewed as a third category in image reconstruction algorithms for Emission Tomography. We consider the radiotracer activity distribution as being a probability density on $\mathbb{R}^3$ and infer on it. This approach is called nonparametric because the parameter of interest, which is a probability distribution, is infinite dimensional and it is Bayesian since we put a prior on this distribution and infer on its posterior. So, the advantage of this approach is in offering a Bayesian regularization context for nonparametric estimation. In addition, we have access to the entire distribution of the posterior activity distribution, not only a point estimate like in ML and MAP approaches. Therefore, any posterior uncertainty (like variance, highest posterior density regions etc.) can be estimated, which is well suited for quantitative imaging in "small samples" (referred to as "low doses" in PET).

The Bayesian nonparametric approach was first used for the two-dimensional (2D) PET reconstruction in [6]. In this paper, we adopt the same approach for the three dimensional case. The 3D case is, however, more complicated since it requires to deal with truncated data due to the limited field-of-view of the system (FOV). The sampling scheme used in this work is different from that in [6] and is not based on any truncation of the infinite dimensional distribution.

The rest of this paper is structured as follows. In section 2, we describe the general problem formulation in the Bayesian nonparametric context. In section 3, the nonparametric prior model for the unknown parameter of interest is given. Our hierarchical model for emission tomographic data is presented in section 4. Afterwards, the derived MCMC sampling is outlined in section 5. Results illustrating the quality of the proposed algorithm compared to MAP-EM are shown in section 6 and we conclude the paper in section 7.

2. PROBLEM STATEMENT IN THE BAYESIAN NONPARAMETRIC FRAMEWORK

In this section, we develop the statiscal framework for the Bayesian nonparametric reconstruction applied to PET data.

We are interested in reconstructing the 3D image of the radiotracer activity distribution from measured data. We assume that data are stored in the so-called *list-mode* format such that the coordinates of the detectors receiving two coincident photons are observed (the approach is also applicable to *sinogram mode*, see [6]). We denote by the term "LOR" the virtual line that connects two detectors in coincidence. A LOR l is parametrized by $(\mathbf{s}_l, \varpi_l, \mathbf{r}_l, \varphi_l)$ where $\mathbf{s}_l$ is the distance between the z-axis (ring center) and the LOR in a transaxial plane (xy); the LOR and the x-axis define the azimuthal angle ϖ_l; $\mathbf{r}_l$ is the axial coordinate of the LOR (if the coordinates of the two detector crystals in coincidence are (z_1, z_2), $r_l = (z_1 + z_2)/2$). Finally, the LOR forms the angle φ_l with the transaxial plane. We map LOR parameters to the index l such that the random variable y_i is the index of the LOR corresponding to the i^{th} observed coincidence. The inverse problem can be formulated as follows in the Bayesian nonparametric context [7]

$$\begin{aligned} &G \sim \mathcal{G} \\ &F(\cdot) = \int_{\mathcal{X}} \mathcal{P}(\cdot|\mathbf{x})\, G(d\mathbf{x}) \\ &y_i \overset{\text{iid}}{\sim} F, \text{ for } i = 1, \dots, n \end{aligned} \tag{1}$$

where $\mathcal{X} \subset \mathbb{R}^3$ is the object space (*i.e*, the space over which reconstruction is carried out); G stands for the spatial distribution of recorded events we want to estimate from the observed F-distributed data sets $\mathbf{Y} = (y_1, \dots, y_n)$. In Bayesian nonparametrics, we must put a prior over the unknown distribution G. This is given by the prior law $\mathcal{G}$ of G. The relation between the data y_i and the emissions locations $\mathbf{x}_i$ is embodied in a probability distribution $\mathcal{P}(\cdot|\mathbf{x})$ called projection distribution, *i.e*, the probability for detecting photons pair in a LOR l given that the corresponding positron emission occurred in $\mathbf{x}$.

In the formulation (1) of the inverse problem, $G(\cdot)$ is the spatial distribution for *recorded events*. Nevertheless, data are truncated in the 3D case due to the finite axial extent of the PET scanner and detector efficiencies. Let introduce the random variable $y^* := y > 0$ to denote a recorded data. So, $G(d\mathbf{x}) = G^*(d\mathbf{x}|y^*)$, where $G^*(d\mathbf{x}|y^*)$ is the distribution of $\mathbf{x}$ conditioned on the fact that the event has been detected. The distribution $G^\dagger(\cdot)$ for whole events (not only the recorded) can be expressed using a derivation of Bayes' rule,

$$G^\dagger(d\mathbf{x}) = \frac{G^*(d\mathbf{x}|y^*)\,\Pr(y^*)}{\Pr(y^*|\mathbf{x})} \tag{2}$$

where for all $\mathbf{x} \in \mathcal{X}$, $\Pr(y^*|\mathbf{x}) = \sum_{l=1}^{L} \Pr(y = l|\mathbf{x})$ is the probability for recording an emission located in $\mathbf{x}$ by the system accounting for system's geometry and physical properties. Since $G^\dagger(d\mathbf{x})$ is a normalized version of $G^*(d\mathbf{x}|y^*)/\Pr(y^*|\mathbf{x})$, posterior inference on $G^*(d\mathbf{x}|y^*)$ is sufficient to estimate the posterior of $G^\dagger(d\mathbf{x})$ as soon as $\Pr(y^*|\mathbf{x}) > 0$, for all $x \in \mathcal{X}$. For sake of simplicity in the notations, in the rest of this paper we omit the conditioning since we deal with the recorded events in the inference. We use $G(d\mathbf{x})$ to stand for $G^*(d\mathbf{x}|y^*)$, the distribution of a recorded event.

We model the detection probability using Bayes' theorem:

$$\Pr(y = l|\mathbf{x}) = \frac{f_{\text{geom}}(\mathbf{x}|y = l)\Pr(y = l)}{\sum_{l=0}^{L} f_{\text{geom}}(\mathbf{x}|y = l)\Pr(y = l)} \tag{3}$$

where $y = 0$ stands for an unobserved event, $f_{\text{geom}}(\mathbf{x}|y = l)$ is the geometric probability density that a photons pair detected in a LOR l originated from a spatial location $\mathbf{x}$. This distribution accounts for the fact that the probability to detect in coincidence emissions from inside the system's FOV is position-dependent. Indeed, if we consider perfect detectors, the probability to record true events is merely geometric. It is determined by computing for each location $\mathbf{x}$, the solid angle subtended from this point to the faces of the detectors. Other effects such as non-collinearity of photons and positron range due to the small distance the emitted positron travels before annihilating with an electron can also be included in f_{geom}.

The probability for observing an event in a LOR l is modeled as follows, for any $l > 0$

$$\Pr(y = l) \propto n(l)\, a(l). \tag{4}$$

For each detector pair l, $n(l)$ is a normalizing factor standing for its detection efficiency and is obtained from calibration procedures; $a(l)$ is the attenuation factor and can be estimated by a preliminary transmission scan of the object. In absence of attenuation and imperfections in detectors, $\Pr(y = l) = 1/L$, where L is the total number of possible LORs.

3. NONPARAMETRIC PRIOR FOR THE RANDOM ACTIVITY DISTRIBUTION

The formulation (1) requires to put a prior distribution $\mathcal{G}$ over the random activity distribution G of recorded events. We choose the prior on the density of G (denoted f_G), as being a Dirichlet process mixture (DPM). We refer to [7] for more details and [6] for a short introduction to this process and its application to 2D PET reconstruction. We briefly recall that this is obtained by first generating a random discrete distribution from a Dirichlet process (DP) with parameters $\alpha > 0$ which tunes the prior number of components and G_0 a probability distribution of components locations. Second, the realization H is smoothed out with a parametric kernel. We choose this kernel as being a multivariate normal distribution (with density $f_{\mathcal{N}}$, whose parameters are the mean $\mathbf{m}$ and the covariance matrix $\mathbf{\Sigma}$), and G_0 the Normal-Inverse Wishart model ($\mathcal{NIW}_{\rho, n_0, \boldsymbol{\mu}_0, \mathbf{\Sigma}_0}$).

$$\begin{aligned} &H \sim DP(\alpha, G_0) \Rightarrow H(\cdot) = \sum_{k=1}^{\infty} w_k \delta_{\boldsymbol{\theta}_k}(\cdot) \\ &f_G(\mathbf{x}) = \int f_{\mathcal{N}}(\mathbf{x}|\boldsymbol{\theta}) H(d\boldsymbol{\theta}) = \sum_{k=1}^{\infty} w_k\, f_{\mathcal{N}}(\mathbf{x}|\mathbf{m}_k, \mathbf{\Sigma}_k) \end{aligned} \tag{5}$$

where

- the sequence of weights $\mathbf{w} = (w_1, w_2, \ldots)$ is constructed as follows: for all i, $V_i \sim \text{Beta}(1, \alpha)$ and $w_1 = V_1$ and for all $k \geq 2$, $w_k = V_k \prod_{j=1}^{k-1}(1 - V_j)$. This is called GEM distribution, written $\mathbf{w} \sim \text{GEM}(\alpha)$,
- $\boldsymbol{\theta}_k = (\mathbf{m}_k, \boldsymbol{\Sigma}_k) \sim G_0$

In the following section, we introduce our hierarchical model for PET data.

4. THE BAYESIAN HIERARCHICAL MODEL

Let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ be n emissions locations in the domain of interest. In emission imaging systems, these locations are never observed directly, only their projections $\mathbf{y} = (y_1, \ldots, y_n)$ are seen. Then our problem is an indirect density estimation since observations are not distributed according to the distribution of interest G but instead according to F, which is related to G via the integral in (1). We introduce emissions locations as hidden variables to complete the data. We also introduce allocation variables $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ such that $c_i = k$ if $\boldsymbol{\theta}_i = (\mathbf{m}, \boldsymbol{\Sigma})_i = \boldsymbol{\theta}_k$. Then the joint distribution of all variables in the model can be written as

$$\begin{aligned} f(\mathbf{y}, \mathbf{X}, \mathbf{c}, \boldsymbol{\Theta}, \mathbf{w}) = & P(\mathbf{y}|\mathbf{X}) \times f(\mathbf{X}|\mathbf{c}, \boldsymbol{\Theta}) \\ & \times P(\mathbf{c}|\mathbf{w}) \times f(\mathbf{w}) \times f(\boldsymbol{\Theta}) \end{aligned} \tag{6}$$

with distributions given by the following generative hierarchical model

$$\begin{aligned} y_i|\mathbf{x}_i &\overset{\text{ind}}{\sim} \mathcal{P}(y_i|\mathbf{x}_i) \\ \mathbf{x}_i|c_i, \boldsymbol{\Theta} &\overset{\text{ind}}{\sim} \mathcal{N}(\mathbf{x}_i|\boldsymbol{\theta}_{c_i}) \\ c_i|\mathbf{w} &\overset{\text{iid}}{\sim} \sum_{k=1}^{\infty} w_k \delta_k(\cdot) \\ \mathbf{w} &\sim \text{GEM}(\alpha) \\ \boldsymbol{\theta}_k &\overset{\text{iid}}{\sim} \mathcal{NIW}_{\rho, n_0, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0} \end{aligned} \tag{7}$$

5. MCMC SAMPLING

Having our hierarchical model (7), we can obtain samples from the posterior distribution of the random activity $G(\cdot)|\mathbf{y}$. For this purpose, we need to get draws from the posterior laws of model variables. Since these posteriors have intractable forms, we opted for a MCMC sampling to generate these draws. The algorithm will update each block of variables in turn according to its conditional posterior as follows

$$\begin{array}{ll} \text{Proposal of annihilations locations:} & \mathbf{X}|\mathbf{w}, \boldsymbol{\Theta}, \mathbf{y} \\ \text{Allocation of locations to components:} & \mathbf{c}|\mathbf{w}, \boldsymbol{\Theta}, \mathbf{X} \\ \text{Updating of the components parameters:} & \boldsymbol{\Theta}|\mathbf{c}, \mathbf{X} \\ \text{Updating of the components weights:} & \mathbf{w}|\mathbf{c}. \end{array} \tag{8}$$

By choosing conjugate priors, these conditional posteriors are directly generated by a Gibbs sampler. However, the sampling of $\mathbf{X}|\mathbf{w}, \boldsymbol{\Theta}, \mathbf{y}$ is more complex since the distribution has no known form. We use the Metropolis-Hastings algorithm with a proposal distribution being the product of the Dirichlet process mixture and a normal distribution extended in each observed LOR direction. The result is a re-weighted mixture of Gaussian distributions in the LOR.

The major problem in the sampling is to tackle the infinite number of dimensions that appears in (7), without any hard truncation. We developed an algorithm belonging to the class of slice sampling methods [8]. The general idea of the slice sampling strategy is to introduce latent auxiliary variables that allow to work with a conditionally finite number of components at each iteration of the sampler.

6. SIMULATION RESULTS

We applied our algorithm to PET simulated data and compared its performances to those obtained with a Bayesian voxels-based algorithm using the maximum a posteriori approach (MAP) with a Gibbs distribution as prior.

We randomly generated decays from a 3D voxelized brain phantom by a Monte Carlo technique such that $n = 10^7$ events were recorded. We have not included scattered and random coincidences in the model and detection probabilities were assumed purely geometrical.

We give now the implementation details for the tested methods.

Proposed method

We choose the parameters of the DPM as follows: $\alpha = 500$, $\boldsymbol{\Sigma}_0 = 6.25 \times \mathbb{I}_3$, $n_0 = 4$. We first computed 5000 iterations for the burn-in period followed by 10000 for computing functionals of the posterior spatial emission distribution. After the burn-in, each draw $(\mathbf{X}^{(t)}, \mathbf{c}^{(t)}, \mathbf{w}^{(t)}, \boldsymbol{\Theta}^{(t)})$ obtained at iteration t is used to construct a sample of the random probability density of recorded events,

$$f_{G^{(t)}}(\mathbf{x}) \approx \sum_{k=1}^{K^*} w_k^{(t)} f_{\mathcal{N}}\left(\mathbf{x}|\theta_k^{(t)}\right) \tag{9}$$

where K^* is the number of components required by the slice sampler. Using equations (2), (3) and (4), the probability density of all events is given by

$$f_{G^{\dagger(t)}}(\mathbf{x}) \propto \frac{f_{G^{(t)}}(\mathbf{x})}{\sum_{l=1}^{L} n(l) a(l) f_{geom}(\mathbf{x}|y = l)}. \tag{10}$$

We choose as estimator for the density the conditional expectation of $G^{\dagger}$,

$$\mathbb{E}(f_{G^{\dagger}}|\mathbf{y}) \approx \frac{1}{N} \sum_{t=1}^{N} f_{G^{\dagger(t)}}$$

with N being the number of iterations after the burn in.

MAP method

The Gibbs prior is used to express spatial constraints such that neighboring voxels must have the same intensity value while allowing curt changes at tissues boundaries. The potential function used was the *log cosh* function suggested by [4]. The parameters of the Gibbs prior were chosen such that to minimize the mean squared error (MSE) w.r.t to the brain phantom. We used a 5-voxels neighboorhood, the weights are

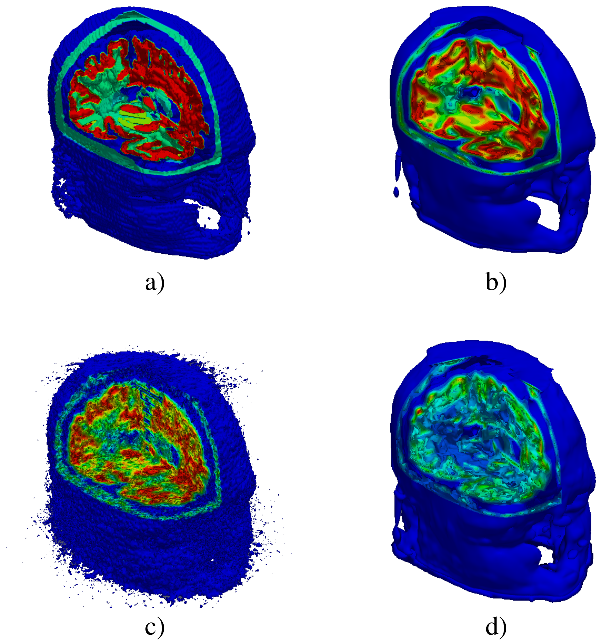

Fig. 1. Reconstruction results: a) 3D brain fantom; b) proposed estimator; c) MAP-EM estimator; d) proposed estimator standard deviation.

the inverse of the L2-distance between the two voxels considered. The algorithm used to maximize the posterior distribution is a modified version of the Expectation Maximization algorithm.

Results

Results obtained by the two algorithms are depicted in [Fig.1]. Image a) is the 3D brain phantom used to generate the data, image b) the conditional mean of our method and image c) the MAP-EM reconstruction. Since any posterior functional is available, we display in image d) the conditional standard deviation w.r.t. the phantom. Credible intervals are also available for each region of interest (figures not shown). It is substantially notable on 3D curves [Fig.1] that our method provides smooth isosurfaces while not removing boundaries and edges in the image. In contrast, the MAP-EM approach exhibits very noisy images. This underlines that the proposed method is suitable to improve signal-to-noise ratio and image resolution. It is worth noting that in our results, the discretization of the spatial distribution is only for visualization and was chosen *a posteriori* to $256 \times 256 \times 128$ corresponding to the phantom size. The same discretization was used for MAP-EM.

7. CONCLUSION

We have presented an indirect-density estimation model for image reconstruction in emission tomography. The inference scheme was done through a specific MCMC method to deal with the infiniteness of the distribution. The structure of the sampler allows its easy parallelization. Simulation results on PET data have provided good quality images in the context of a relatively low number of events (10 millions) and our reconstructions are by far better than those obtained by the MAP-EM method. However, we did not account for background effects (randoms and scattered for example) in our simulations. Future works will be devoted to include these effects.

The major drawback of our approach is its computational demands, although it was implemented on a parallel computer ($\approx$ one day on a high performance computer). To circumvent the time due to the MCMC sampling, an alternative is the variational Bayesian method which approximates the posterior distributions analytically [9]. Another interesting way to speed up our algorithm can be achieved by implementing it on Graphical processor units (GPU) hardware. Since we work directly with functionals belonging to $\mathbb{R}^3$, this is appealing for these architectures.

8. REFERENCES

[1] F. Natterer, *The Mathematics of Computerized Tomography*, Society for Industrial and Applied Mathematics, 1986.

[2] L.A. Shepp and Y. Vardi, “Maximum likelihood reconstruction in positron emission tomography,” *IEEE Trans. Med. Imag.*, vol. 1, pp. 113–122, 1982.

[3] K. Lange and R. Carson, “EM reconstruction algorithm for emission and transmission tomography,” *J. Comp. Assist. Tomo.*, vol. 8, pp. 306–316, 1984.

[4] P. J. Green, “Bayesian reconstructions from emission tomography data using a modified EM algorithm,” *IEEE Trans. Med. Imaging*, vol. 9, no. 1, pp. 84–93, 1990.

[5] S. Geman and D. E. McClure, “Statistical methods for tomographic image reconstruction,” in *Proceedings of the 46th Session of the International Statiscal Institute*, Bulletin of the ISI, Ed., 1987, vol. 52, pp. 5–21.

[6] É. Barat, C. Comtat, T. Dautremer, T. Montagu, and R. Trébossen, “A Bayesian nonparametric approach for PET reconstruction,” in *IEEE Med. Imaging Conf. Rec., NSS/MIC ’07.*, 2007, vol. 6, pp. 4155–4162.

[7] N.L. Hjort, C. Holmes, P. Müller, and S.G. Walker, *Bayesian Nonparametrics*, Cambridge University Press, April 2010.

[8] S.G. Walker, “Sampling the Dirichlet mixture model with slices,” *Comm. Statist.*, vol. 36, pp. 45–54, 2007.

[9] D.M. Blei and M.I. Jordan, “Variational inference for dirichlet process mixtures,” *Bayesian Analysis*, vol. 1, no. 1, pp. 121–144, 2006.