

Gauss-Markov-Potts Priors for Images in Computer Tomography Resulting to Joint Optimal Reconstruction and segmentation

Ali Mohammad-Djafari

Laboratoire des signaux et systèmes (LSS)

(UMR 08506 du CNRS - SUPELEC - Univ Paris Sud)

Supélec, Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France

ABSTRACT

In many applications of Computed Tomography (CT), we know that the object under the test is composed of a finite number of materials meaning that the images to be reconstructed are composed of a finite number of homogeneous area. To account for this prior knowledge, we propose a family of Gauss-Markov fields with hidden Potts label fields. Then, using these models in a Bayesian inference framework, we are able to jointly reconstruct the image and segment it in an optimal way. In this paper, we first present these prior models, then propose appropriate Bayesian computational methods (MCMC or Variational Bayes) to compute the Joint Maximum A Posteriori (JMAP) or the posterior mean estimators. We finally provide a few results showing the efficiency of the proposed methods for CT with very limited angle and number of projections.

Keywords: Computed Tomography; Gauss-Markov-Potts Priors; Bayesian computation; MCMC; Joint Segmentation and Reconstruction

2000 Mathematics Subject Classification: 62F15, 62M40, 68U10, 94A08, 44A12, 65R32

1 Introduction

The simplest forward model in CT is the line integration model:

$$g(\mathbf{s}_i) = \int_{L_i} f(\mathbf{r}) \, dl_i \quad (1.1)$$

where $\mathbf{r} = (x, y)$ is a pixel or $\mathbf{r} = (x, y, z)$ a voxel position, \mathbf{s}_i is a detector position, L_i is a line connecting the X ray source position to the detector position \mathbf{s}_i and dl_i is a unit element on this line (Brooks and Di Chiro, 1975; Budinger, Gullberg and Huesman, 1979). This model becomes equivalent to the Radon Transform (RT) in 2D case. Figure 1 shows three configurations: a 3D and a 2D parallel geometry and a 2D fan beam geometry data acquisition.

No matter the type of the acquisition geometries, when discretized, this equation becomes

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon} \quad (1.2)$$

where the vector $\mathbf{g} = \{g(\mathbf{s}_i), i = 1, \dots, M\}$ contains all the measured data, the vector $\mathbf{f} = \{f_1, \dots, f_N\}$ contains the pixel or voxel values of the discretized images, the matrix \mathbf{H} is a

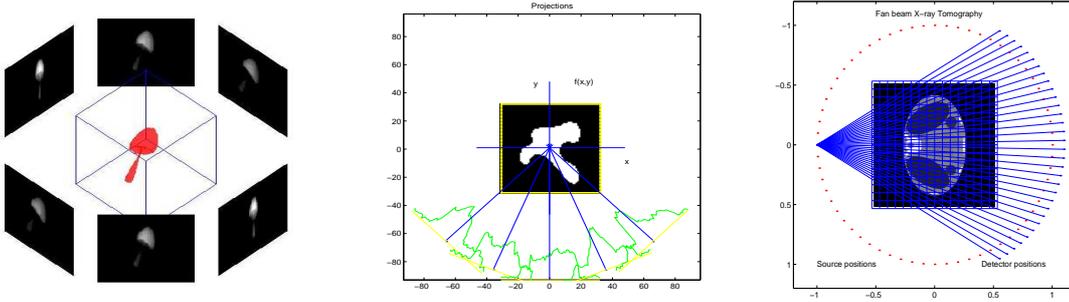


Figure 1: Three examples of CT geometries: 3D parallel, 2D parallel and 2D fan beam.

huge dimensional sparse matrix whose elements $H_{i,j}$ represent the length of the i -th ray in pixel or voxel j , and finally, the vector $\epsilon = \{\epsilon_1, \dots, \epsilon_M\}$ contains the errors.

If we want to distinguish different projections at different angles, then, we can split the vector g in a set of subvectors g_l and thus the projection matrix H into a set of block matrices H_l and the error vector ϵ in a set of subvectors ϵ_l :

$$g = \begin{bmatrix} g_1 \\ \vdots \\ g_L \end{bmatrix}, H = \begin{bmatrix} H_1 \\ \vdots \\ H_L \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_L \end{bmatrix} \quad (1.3)$$

in such a way that we can write

$$g_l = H_l f + \epsilon_l, \quad l = 1, \dots, L \quad (1.4)$$

This discretized presentation of CT, gives the possibility to analyse the most classical methods of image reconstruction (Eggermont and Herman, 1981; Herman and Lent, 1976). For example, it is very easy to see that the solution

$$\hat{f} = H^t g = \sum_l H_l^t g_l \quad (1.5)$$

corresponds to the classical *Backprojection* (BP), the minimum norm solution of $Hf = g$:

$$\hat{f} = H^t (HH^t)^{-1} g = \sum_l H_l^t (H_l H_l^t)^{-1} g_l \quad (1.6)$$

can be identified to the classical *Filtered Backprojection* (FBP), and the least squares (LS) solution

$$\hat{f} = (H^t H)^{-1} H^t g = (H^t H)^{-1} \sum_l H_l^t g_l \quad (1.7)$$

can be identified to the *Backprojection and Filtering* (BPF).

Also, defining the LS criterion

$$J_0(f) = \|g - Hf\|^2 = \sum_l \|g_l - H_l f\|^2 \quad (1.8)$$

and its gradient

$$\nabla J_0(f) = -2H^t(g - Hf) = -2 \sum_l H_l^t (g_l - H_l f) \quad (1.9)$$

it can easily be shown that the following iterative method

$$\mathbf{f}^{(k+1)} = \mathbf{f}^{(k)} + \alpha \sum_l \mathbf{H}_l^t (\mathbf{g}_l - \mathbf{H}_l \mathbf{f}^{(k)}) \quad (1.10)$$

which is a gradient type method to obtain the LS solution, corresponds to the classical Landweber reconstruction method.

In a Bayesian inference framework for this inverse problem (Hanson and Wechsung, 1983), one starts by writing the expression of the posterior law:

$$p(\mathbf{f}|\boldsymbol{\theta}, \mathbf{g}; \mathcal{M}) = \frac{p(\mathbf{g}|\mathbf{f}, \boldsymbol{\theta}_1; \mathcal{M}) p(\mathbf{f}|\boldsymbol{\theta}_2; \mathcal{M})}{p(\mathbf{g}|\boldsymbol{\theta}; \mathcal{M})} \quad (1.11)$$

where \mathcal{M} represents the whole mathematical model, $p(\mathbf{g}|\mathbf{f}, \boldsymbol{\theta}_1; \mathcal{M})$, called the *likelihood*, is obtained using the forward model (1.2) and the assigned probability law $p_\epsilon(\epsilon)$ of the errors ϵ and $p(\mathbf{f}|\boldsymbol{\theta}_2; \mathcal{M})$ is the assigned prior law for the unknown image \mathbf{f} and

$$p(\mathbf{g}|\boldsymbol{\theta}; \mathcal{M}) = \int p(\mathbf{g}|\mathbf{f}, \boldsymbol{\theta}_1; \mathcal{M}) p(\mathbf{f}|\boldsymbol{\theta}_2; \mathcal{M}) d\boldsymbol{\theta}. \quad (1.12)$$

which is called the evidence of the model \mathcal{M} with hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. With the following particular Gaussian prior modeling of the errors probability law $p_\epsilon(\epsilon)$ and $p(\mathbf{f}|\boldsymbol{\theta}_2; \mathcal{M})$:

$$\begin{aligned} p(\mathbf{g}|\mathbf{f}, \theta_\epsilon; \mathcal{M}) &= \mathcal{N}(\mathbf{H}\mathbf{f}, (1/\theta_\epsilon)\mathbf{I}), \\ p(\mathbf{f}|\theta_f; \mathcal{M}) &= \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_f) \text{ with } \boldsymbol{\Sigma}_f = (1/\theta_f)(\mathbf{D}^t \mathbf{D})^{-1} \end{aligned} \quad (1.13)$$

it is easy to see that

$$p(\mathbf{f}|\mathbf{g}, \theta_\epsilon, \theta_f; \mathcal{M}) \propto p(\mathbf{g}|\mathbf{f}, \theta_\epsilon; \mathcal{M}) p(\mathbf{f}|\theta_f; \mathcal{M}) = \mathcal{N}(\hat{\mathbf{f}}, \hat{\boldsymbol{\Sigma}}_f), \quad (1.14)$$

with

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_f &= [\theta_\epsilon \mathbf{H}^t \mathbf{H} + \theta_f \mathbf{D}_f^t \mathbf{D}_f]^{-1} \\ &= \frac{1}{\theta_\epsilon} [\mathbf{H}^t \mathbf{H} + \lambda \mathbf{D}^t \mathbf{D}]^{-1}, \text{ with } \lambda = \frac{\theta_f}{\theta_\epsilon} \end{aligned} \quad (1.15)$$

and

$$\hat{\mathbf{f}} = \theta_\epsilon \hat{\boldsymbol{\Sigma}}_f \mathbf{H}^t \mathbf{g} = [\mathbf{H}^t \mathbf{H} + \lambda \mathbf{D}^t \mathbf{D}]^{-1} \mathbf{H}^t \mathbf{g} \quad (1.16)$$

where \mathbf{D} is a first or second order finite differences matrix.

It is also easy to see that $\hat{\mathbf{f}}$ in (1.16) can also be computed as the solution of the optimization of

$$J_1(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda \|\mathbf{D}\mathbf{f}\|^2 \quad (1.17)$$

where we can see the link with the classical regularization theory.

Keeping the Gaussian prior for the errors, but choosing other more appropriate priors for $p(\mathbf{f}|\boldsymbol{\theta}_2; \mathcal{M})$ and choose the Maximum *a posteriori* (MAP) estimate

$$\hat{\mathbf{f}} = \arg \max_{\mathbf{f}} \{p(\mathbf{f}|\boldsymbol{\theta}, \mathbf{g}; \mathcal{M})\} = \arg \min_{\mathbf{f}} \{J_1(\mathbf{f})\} \quad (1.18)$$

with

$$J_1(\mathbf{f}) = -\ln p(\mathbf{g}|\mathbf{f}, \theta_\epsilon; \mathcal{M}) - \ln p(\mathbf{f}|\boldsymbol{\theta}_2; \mathcal{M}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda \Omega(\mathbf{f}) \quad (1.19)$$

where $\lambda = 1/\theta_\epsilon$ and $\Omega(\mathbf{f}) = -\ln p(\mathbf{f}|\theta_2; \mathcal{M})$, we can see the importance of the prior law as a regularization operator. Many different choices has been used. Here we give a synthetic view of them. At a first glance, we can classify them in two categories:

Separables:

$$p(\mathbf{f}) \propto \exp \left[-\theta_f \sum_j \phi(f_j) \right] \quad (1.20)$$

and

Markovians:

$$p(\mathbf{f}) \propto \exp \left[-\theta_f \sum_j \phi(f_j - f_{j-1}) \right] \quad (1.21)$$

where different classical choices for $\phi(t)$ are:

$$\phi(t) = \left\{ |t|^\beta, (\ln(1 + |t|)), \begin{cases} t^2 & \text{if } |t| < \alpha, \\ 2\alpha t - \alpha^2 & \text{else,} \end{cases}, \begin{cases} |t|^2 & \text{if } |t| < \alpha, \\ \alpha^2 & \text{else,} \end{cases}, \frac{\alpha^2 t^2}{1 + t^2}, \log \cosh(t/\alpha) \right\} \quad (1.22)$$

But all these priors translate global properties of images by assuming a global homogeneity property for them. In many applications, we may have more precise prior, for example piecewise homogeneity. A simple way to introduce this prior is to introduce a hidden variable. As an example, by introducing hidden binary contour variables c_j which take the values 1 when there is a contour at position j and 0 when the location j is inside a homogeneous region, we can propose

Nonhomogeneous Markovian models with hidden binary contours:

$$p(\mathbf{f}) \propto \exp \left[-\theta_f \sum_j [(1 - c_j)\phi(f_j - f_{j-1}) + c_j\phi(f_j)] \right] \quad (1.23)$$

Another more precise prior, in particular in CT, is that the object under the test is composed of a finite number of homogeneous materials. This implies the introduction of a hidden variable z_j which can take discrete values $\{1, \dots, K\}$. All the pixels of the image labeled with the same value $z_j = k$ share a homogeneity property and are grouped in compact homogeneous regions. This is exactly the prior information that we would like to account for in this paper. As we will see in the next section, the proposed Gauss-Markov with hidden separable or Markovian (Potts) hidden labels are appropriate models to account for this prior knowledge.

The rest of this paper is then organized as follows. In Section 2, we give the details of the proposed Gauss-Markov-Potts prior models. In Section 3, we use these priors in a Bayesian framework to obtain a joint posterior law of all the unknowns (image pixel values, image pixel labels and all the hyperparameters including the region statistical parameters and the noise variance). Then, in Section 4 we will see how to perform the Bayesian computations with these priors. Finally, in Section 5, we show a few reconstruction results and in Section 6, we present the main conclusions of this paper.

2 Proposed Gauss-Markov-Potts prior models

As we introduced in previous section, we consider here the case of CT imaging systems with applications where our prior knowledge about the object under the test is that those objects are composed of a finite number of known materials. This is the case of Non Destructive Testing (NDT) in industrial applications where, for example, the known materials are air-metal or air-metal-composite, or the medical imaging where the known materials are air-tissue-muscle-bone in X ray CT or gray-white materials in PET. So, here, we propose a model which accounts for this prior knowledge (Snoussi and Mohammad-Djafari, 2004a; Snoussi and Mohammad-Djafari, 2004b; Humblot and Mohammad-Djafari, 2006; Féron and Mohammad-Djafari, 2005; Mohammad-Djafari, 2007).

The main idea is then to consider the pixels or voxels of the unknown object $\mathbf{f} = \{f(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$ to be classified in K classes, each class identified by a discrete valued variable (label) $z(\mathbf{r}) \in \{1, \dots, K\}$. The K -colored image $z(\mathbf{r}) = \{z(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$ represents then the segmentation of the image $f(\mathbf{r})$. Here, and in the following, \mathcal{R} represents the entire image pixel area.

Indeed, the pixels $\mathbf{f}_k = \{f(\mathbf{r}), \mathbf{r} \in \mathcal{R}_k\}$ in the compact regions $\mathcal{R}_k = \{\mathbf{r} : z(\mathbf{r}) = k\}$ have to share some common properties, for example the same means μ_k and the same variances v_k (probabilistic homogeneity). Those pixels may be localized in a compact region or in a finite number of compact and disjointe regions \mathcal{R}_{kl} such that: $\cup_l \mathcal{R}_{kl} = \mathcal{R}_k$ and $\cup_k \mathcal{R}_k = \mathcal{R}$. Naturally, we assume that \mathbf{f}_k and $\mathbf{f}_l \forall k \neq l$ are *a priori* independents.

To each region is associated a contour. If we represent the contours of all the regions in the image by a binary valued variable $c(\mathbf{r})$ we have $c(\mathbf{r}) = 0$ inside any region and $c(\mathbf{r}) = 1$ on the borders of those regions. We may note that $c(\mathbf{r})$ is obtained from $z(\mathbf{r})$ in a deterministic way

$$c(\mathbf{r}) = \begin{cases} 0 & z(\mathbf{r}) = z(\mathbf{r}') \forall \mathbf{r}' \in \mathcal{V}(\mathbf{r}) \\ 1 & \text{elsewhere} \end{cases} \quad (2.1)$$

where $\mathcal{V}(\mathbf{r})$ represents the neighborhood of \mathbf{r} . See Figure 2 for relation between these quantities.

With these prior informations, we can write:

$$p(f(\mathbf{r})|z(\mathbf{r}) = k, m_k, v_k) = \mathcal{N}(m_k, v_k) \quad (2.2)$$

which suggests a Mixture of Gaussians (MoG) model for $p(f(\mathbf{r}))$:

$$p(f(\mathbf{r})) = \sum_k \alpha_k \mathcal{N}(m_k, v_k) \text{ with } \alpha_k = P(z(\mathbf{r}) = k) \quad (2.3)$$

Now, we also need to model the spatial interactions of the elements of the image pixels $\mathbf{f} = \{f(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$ the labels $z = \{z(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$ and the contours $\mathbf{c} = \{c(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$. This can be considered by considering a markovian structure either for \mathbf{f} or for z or for \mathbf{c} . Here, we only consider the markovian structures for \mathbf{f} and z . We may then consider four cases:

Case 1: Independent Mixture of Independent Gaussians (IMIG) :

In this case, which is the easiest, no *a priori* markovian structure is assumed neither for $\mathbf{f}|z$ nor for z .

$$\begin{cases} p(f(\mathbf{r})|z(\mathbf{r}) = k, m_k, v_k) & = \mathcal{N}(m_k, v_k), \quad \forall \mathbf{r} \in \mathcal{R} \\ p(\mathbf{f}|z, \theta_2) & = \prod_{\mathbf{r} \in \mathcal{R}} \mathcal{N}(m_z(\mathbf{r}), v_z(\mathbf{r})) \end{cases} \quad (2.4)$$

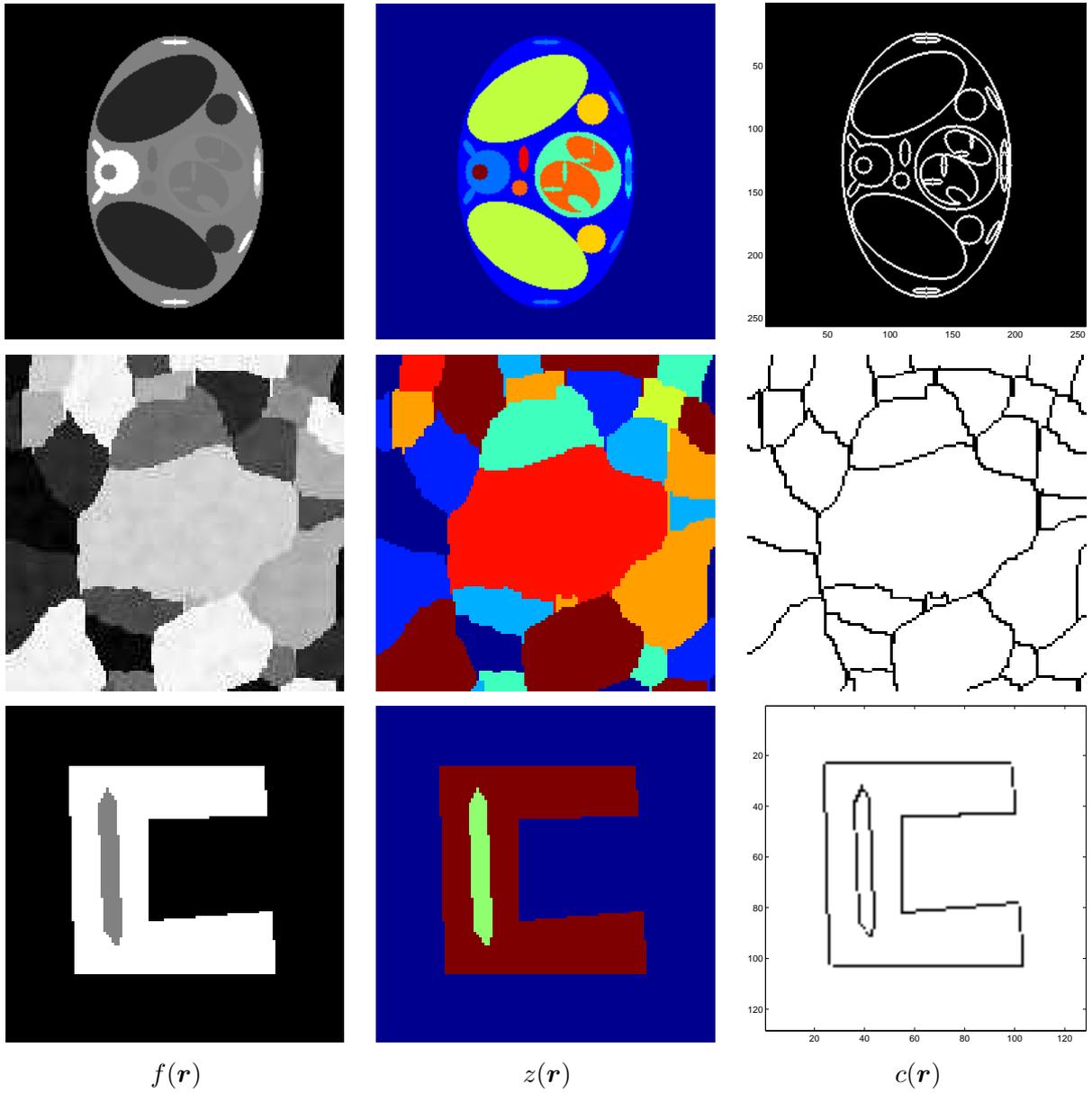


Figure 2: Proposed *a priori* model for the images: the image pixels $f(\mathbf{r})$ are assumed to be classified in K classes, $z(\mathbf{r})$ represents those classes (segmentation) and $c(\mathbf{r})$ the contours of those regions. The three examples of images presented here correspond to three different applications.

with $m_z(\mathbf{r}) = m_k, \forall \mathbf{r} \in \mathcal{R}_k, v_z(\mathbf{r}) = v_k, \forall \mathbf{r} \in \mathcal{R}_k, \boldsymbol{\theta}_2 = \{(m_k, v_k), k = 1, \dots, K\}$ and

$$\begin{cases} p(z(\mathbf{r}) = k) &= \alpha_k, \forall \mathbf{r} \in \mathcal{R} \\ p(\mathbf{z}|\boldsymbol{\theta}_3) &= \prod_{\mathbf{r}} p(z(\mathbf{r}) = k) = \prod_k \alpha_k^{n_k} \end{cases} \quad (2.5)$$

where $n_k = \sum_{\mathbf{r} \in \mathcal{R}} \delta(z(\mathbf{r}) - k)$ is the number of pixels in the class k and $\sum_k n_k = n$ is the total number of pixels and $\boldsymbol{\theta}_3 = \{\alpha_k, k = 1, \dots, K\}$. With this prior model we can write:

$$\begin{aligned} p(\mathbf{f}|\mathbf{z}, \mathbf{m}, \mathbf{v}) &= \prod_{\mathbf{r} \in \mathcal{R}} \mathcal{N}(m_z(\mathbf{r}), v_z(\mathbf{r})) \\ &\propto \exp \left[-\frac{1}{2} \sum_{\mathbf{r} \in \mathcal{R}} \frac{(f(\mathbf{r}) - m_z(\mathbf{r}))^2}{v_z(\mathbf{r})} \right] \\ &\propto \exp \left[-\frac{1}{2} \sum_k \sum_{\mathbf{r} \in \mathcal{R}_k} \frac{(f(\mathbf{r}) - m_k)^2}{v_k} \right] \end{aligned} \quad (2.6)$$

where $\mathbf{m} = \{m_1, \dots, m_K\}$ and $\mathbf{v} = \{v_1, \dots, v_K\}$.

Case 2: Independent Mixture of Gauss-Markovs :

Here, we keep the independance of the labels $z(\mathbf{r})$ as in (2.5), but we account for a local spatial structure of the pixel values \mathbf{f} . This model can be summarized by the following relations:

$$p(f(\mathbf{r})|z(\mathbf{r}), f(\mathbf{r}'), z(\mathbf{r}'), \mathbf{r}' \in \mathcal{V}(\mathbf{r})) = \mathcal{N}(\mu_z(\mathbf{r}), v_z(\mathbf{r})) \quad (2.7)$$

with

$$\begin{cases} \mu_z(\mathbf{r}) &= \frac{1}{|\mathcal{V}(\mathbf{r})|} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \mu_z^*(\mathbf{r}') \\ \mu_z^*(\mathbf{r}') &= \begin{cases} m_z(\mathbf{r}') & \text{if } z(\mathbf{r}') \neq z(\mathbf{r}) \\ f(\mathbf{r}') & \text{if } z(\mathbf{r}') = z(\mathbf{r}) \end{cases} \\ v_z(\mathbf{r}) &= v_k \quad \forall \mathbf{r} \in \mathcal{R}_k \end{cases} \quad (2.8)$$

We may remark that $\mathbf{f}|\mathbf{z}$ is a non homogeneous Gauss-Markov field because the means $m_z(\mathbf{r})$ and the variances $v_z(\mathbf{r})$ are functions of the pixel position \mathbf{r} . We can also write:

$$p(\mathbf{f}|\mathbf{z}, \boldsymbol{\theta}_2) \propto \prod_k \mathcal{N}(m_k \mathbf{1}_k, \boldsymbol{\Sigma}_k) \quad (2.9)$$

where $\mathbf{1}_k = \mathbf{1}_{n_k}, \forall \mathbf{r} \in \mathcal{R}_k$ and $\boldsymbol{\Sigma}_k$ is a covariance matrix of dimensions $n_k \times n_k$. This covariance $\boldsymbol{\Sigma}_k$ is then dependent to the context k .

Noting that $\mu_{z(\mathbf{r}')}^*$ can also be written via the contour variable $c(\mathbf{r}')$:

$$\mu_z^*(\mathbf{r}') = c(\mathbf{r}')m_z(\mathbf{r}') + (1 - c(\mathbf{r}'))f(\mathbf{r}') \quad (2.10)$$

which gives the possibility to write:

$$\begin{aligned} p(\mathbf{f}|\mathbf{z}, \boldsymbol{\theta}_2) &\propto \exp \left[-\frac{1}{2} \sum_{\mathbf{r} \in \mathcal{R}} \frac{(f(\mathbf{r}) - \mu_z(\mathbf{r}))^2}{v_z(\mathbf{r})} \right] \\ &\propto \exp \left[-\frac{1}{2} \sum_{\mathbf{r} \in \mathcal{R}} \frac{\left(f(\mathbf{r}) - \frac{1}{|\mathcal{V}(\mathbf{r})|} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \mu_z^*(\mathbf{r}') \right)^2}{v_z(\mathbf{r})} \right] \end{aligned} \quad (2.11)$$

Case 3: Gauss-Potts :

Here, we keep the first part of the model (2.4 and 2.6) in the Case 1, but we account for the spatial structure of the label image \mathbf{z} with a simple Potts-Markov model:

$$p(z(\mathbf{r})|z(\mathbf{r}'), \mathbf{r}' \in \mathcal{V}(\mathbf{r})) \propto \exp \left[\gamma \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \delta(z(\mathbf{r}) - z(\mathbf{r}')) \right] \quad (2.12)$$

which can equivalently (Hammerslay-Clifford) be written as:

$$p(\mathbf{z}|\gamma) \propto \exp \left[\gamma \sum_{\mathbf{r} \in \mathcal{R}} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \delta(z(\mathbf{r}) - z(\mathbf{r}')) \right] \quad (2.13)$$

The hyperparameters of this prior model are $\theta_2 = \{(m_k, v_k), k = 1, \dots, K\}$ and $\theta_3 = \gamma$.

Case 4: Compound model of Gauss-Markov-Potts :

This is the case where we use the composition of the two last models:

$$\begin{cases} q(f(\mathbf{r})|z(\mathbf{r}), f(\mathbf{r}'), z(\mathbf{r}'), \mathbf{r}' \in \mathcal{V}(\mathbf{r})) = \mathcal{N}(\mu_z(\mathbf{r}), v_z(\mathbf{r})) \\ p(z(\mathbf{r})|z(\mathbf{r}'), \mathbf{r}' \in \mathcal{V}(\mathbf{r})) \propto \exp \left[\gamma \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \delta(z(\mathbf{r}) - z(\mathbf{r}')) \right] \end{cases} \quad (2.14)$$

with $\mu_z(\mathbf{r})$ and $v_z(\mathbf{r})$ as defined in the Case 2.

3 Bayesian jointe reconstruction, segmentation and characterization

In previous section, we proposed four different priors all trying to account for the prior knowledge that the object under the test is composed of finite number K of materials, each material located in disjoint compact regions and charcterized by a discrete value hidden label variable. Each material is thus characterized by a label $z = k$ and a set of statistical properties $\{m_k, v_k, \alpha_k\}$ or $\{m_k, v_k, n_k\}$. We assumed that all the pixels with different labels are independent. The spatial structure is modeled for by introducing a markovian structure either for the labels or for the pixel values or for both. For each case then we obtained the expressions for $p(\mathbf{f}|z, \theta_2; \mathcal{M})$ and $p(z|\theta_3; \mathcal{M})$. Now, if we also know the expression of the likelihood $p(\mathbf{g}|\mathbf{f}, \theta_1; \mathcal{M})$ and assuming that all the hyperparameters $\theta = (\theta_1, \theta_2, \theta_3)$ are known, then we can express the joint posterior of \mathbf{f} and z :

$$p(\mathbf{f}, z|\theta, \mathbf{g}; \mathcal{M}) = \frac{p(\mathbf{g}|\mathbf{f}, \theta_1; \mathcal{M}) p(\mathbf{f}|z, \theta_2; \mathcal{M}) p(z|\theta_3; \mathcal{M})}{p(\mathbf{g}|\theta; \mathcal{M})} \quad (3.1)$$

and then infer on them. However, in a practical application, we also have to estimate the hyperparameters. In a full Bayesian framework, we also have to assign them a prior law $p(\theta|\mathcal{M})$ and then find the expression of the joint posterior law

$$p(\mathbf{f}, z, \theta|\mathbf{g}; \mathcal{M}) = \frac{p(\mathbf{g}|\mathbf{f}, \theta_1; \mathcal{M}) p(\mathbf{f}|z, \theta_2; \mathcal{M}) p(z|\theta_3; \mathcal{M}) p(\theta|\mathcal{M})}{p(\mathbf{g}|\mathcal{M})} \quad (3.2)$$

and use it to infer on all the unknowns \mathbf{f} , z and θ .

Now, to go further in details, let assume a white Gaussian prior law for the noise which results in $p(\mathbf{g}|\mathbf{f}, \theta_\epsilon) = \mathcal{N}(\mathbf{H}\mathbf{f}, \frac{1}{\theta_\epsilon}\mathbf{I})$ which can be expressed equivalently as

$$\begin{aligned} p(\mathbf{g}|\mathbf{f}, \theta_\epsilon) &\propto \exp \left[-\frac{\theta_\epsilon}{2} J_0(\mathbf{f}) \right] \\ \text{with } J_0(\mathbf{f}) &= \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2, \end{aligned} \quad (3.3)$$

The final prior that we need is $p(\theta|\mathcal{M})$. We remember that $\theta = (\theta_1, \theta_2, \theta_3)$ where $\theta_1 = \theta_\epsilon$ is the inverse of the variance of the noise, $\theta_2 = \{(m_k, v_k), k = 1, \dots, K\}$ and $\theta_3 = \{\alpha_k, k = 1, \dots, K\}$

for the models in case 1 and case 2 and $\theta_3 = \{\gamma\}$ for the models in case 3 and case 4. We choose to assign conjugate priors to them.

The associated conjugate priors are: Gamma for θ_ϵ (Inverse Gamma for $1/\theta_\epsilon$), Gaussians for m_k , Inverse Gammas for v_k and Dirichlet for $\alpha = \{\alpha_1, \dots, \alpha_K\}$ due to the constraint $\sum_k \alpha_k = 1$.

$$\begin{cases} p(1/\theta_\epsilon | a_{e0}, b_{e0}) &= \mathcal{IG}(a_{e0}, b_{e0}), \forall k \\ p(m_k | m_0, v_0) &= \mathcal{N}(m_0, v_0), \forall k \\ p(v_k | a_0, b_0) &= \mathcal{IG}(a_0, b_0), \forall k \\ p(\alpha | \alpha_0) &= \mathcal{D}(\alpha_0, \dots, \alpha_0) \end{cases} \quad (3.4)$$

where $a_{e0}, b_{e0}, m_0, v_0, a_0, b_0$ and α_0 are fixed for a given problem. For example, for images normalized between 0 and 1 and $K = 8$, we may fix $m_0 = .5, v_0 = 1, a_0 = 1, b_0 = 1$, and $\alpha_0 = 1/K$ and for a moderate noise variance $a_{e0} = .1, b_{e0} = 1$.

For the Potts model, unfortunately, there is no conjugate prior for γ . So, in this paper we keep fix this parameter.

4 Bayesian computation

Now, we have all the necessary components to find an expression for $p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \mathbf{g}; \mathcal{M})$. However, using directly this expression to compute the Joint Maximum A Posteriori (JMAP) estimates:

$$(\hat{\mathbf{f}}, \hat{\mathbf{z}}, \hat{\boldsymbol{\theta}}) = \arg \max_{(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})} \{p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \mathbf{g}; \mathcal{M})\} \quad (4.1)$$

or the Posterior Means (PM):

$$\begin{cases} \hat{\mathbf{f}} &= \sum_{\mathbf{z}} \int \int \mathbf{f} p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \mathbf{g}; \mathcal{M}) \, d\mathbf{f} \, d\boldsymbol{\theta} \\ \hat{\boldsymbol{\theta}} &= \sum_{\mathbf{z}} \int \int \boldsymbol{\theta} p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \mathbf{g}; \mathcal{M}) \, d\mathbf{f} \, d\boldsymbol{\theta} \\ \hat{\mathbf{z}} &= \sum_{\mathbf{z}} \int \int \mathbf{z} p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \mathbf{g}; \mathcal{M}) \, d\mathbf{f} \, d\boldsymbol{\theta} \end{cases} \quad (4.2)$$

is often too difficult. We then need either the optimisation or numerical integration algorithms to do Bayesian computation. For this, we mainly have two mains approaches:

Numerical exploration and integration via Monte Carlos techniques:

The main idea here is to approximate the computation of the integrations such as (4.2) by the empirical sums of the samples generated according the joint posterior (3.2). The main difficulty then is to generate those samples via Monte Carlos (MC) and more precisely via the Markov Chain Monte Carlo (MCMC) techniques.

To implement this family of method for our problem, we use a Gibbs sampling technic whose basic idea is to generate samples from the posterior law (3.2) using the following general algorithm:

$$\begin{cases} \mathbf{f} &\sim p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}, \mathbf{g}; \mathcal{M}) \propto p(\mathbf{g} | \mathbf{f}, \boldsymbol{\theta}_1; \mathcal{M}) p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}_2; \mathcal{M}) \\ \mathbf{z} &\sim p(\mathbf{z} | \mathbf{f}, \boldsymbol{\theta}, \mathbf{g}; \mathcal{M}) \propto p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}_2; \mathcal{M}) p(\mathbf{z} | \boldsymbol{\theta}_3; \mathcal{M}) \\ \boldsymbol{\theta} &\sim p(\boldsymbol{\theta} | \mathbf{f}, \mathbf{z}, \mathbf{g}; \mathcal{M}) \propto p(\mathbf{g} | \mathbf{f}, \boldsymbol{\theta}_1; \mathcal{M}) p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}_2; \mathcal{M}) p(\boldsymbol{\theta} | \mathcal{M}) \end{cases} \quad (4.3)$$

We have the expressions of all the necessary probability laws in the right hand side of these three conditional laws to be able to sample from them. Indeed, it is easy to show that the first

one $p(\mathbf{f}|z, \boldsymbol{\theta}, \mathbf{g}; \mathcal{M})$ is a Gaussian and then easy to handle. The second $p(z|\mathbf{f}, \boldsymbol{\theta}, \mathbf{g}; \mathcal{M})$ depending on the case is either separable or Potts. There is also many fast methods to generate samples from a Potts field. The last one $p(\boldsymbol{\theta}|\mathbf{f}, z, \mathbf{g}; \mathcal{M})$ is also separable in its components, and thanks to the conjugate property, it is easy to see that the posterior laws are either Inverse Gamma, Inverse Wishart, Gaussians, and Dirichlet for which there are standard sampling schemes.

The main interest of this approach is that by generating those samples we can explore the whole space of the joint posterior law. The main drawback is the computational cost of these techniques which need a great number of iterations to converge and great number of samples to generate after the convergence to obtain stable and low variance estimates. For some more details on the expressions of the conditional probability laws of the right hand sides of (4.3) and the overall computational cost see (Humblot and Mohammad-Djafari, 2006; Féron and Mohammad-Djafari, 2005).

Variational or separable approximation techniques :

The main idea here to propose a simpler joint pdf $q(\mathbf{f}, z, \boldsymbol{\theta})$ for the joint pdf (3.2) where the remaining computations can be done more easily. Between these methods, the separable approximation techniques where $q(\mathbf{f}, z, \boldsymbol{\theta}) = q_1(\mathbf{f}|z) q_2(z) q_3(\boldsymbol{\theta})$ is the one we follow in this paper.

The idea of approximating a joint probability law $p(\mathbf{x})$ by a separable law $q(\mathbf{x}) = \prod_j q_j(x_j)$ is not new (MacKay, 1992; Ghahramani and Jordan, 1997; Penny and Roberts, 1998; Roberts, Husmeier, Penny and Rezek, 1998; Penny and Roberts, 1999; Jaakkola and Jordan, 2000; Miskin and MacKay, 2001). The way to do and the particular choices of parametric families for $q_j(x_j)$ for which the computations can be done easily have been addressed more recently in many data mining and classification problems (Penny and Roberts, 2002; Roberts and Penny, 2002; Cassidy and Penny, 2002; Penny and Friston, 2003; Choudrey and Roberts, 2003; Penny, Kiebel and Friston, 2003; Nasios and Bors, 2004; Nasios and Bors, 2006; Friston, Mattout, Trujillo-Barreto, Ashburner and Penny, 2006; Penny, Kiebel and Friston, 2006; Penny, Everson and Roberts, 2000). However, the use of these techniques for Bayesian computation for the inverse problems in general and in CT in particular is the originality of this paper.

To give a synthetic presentation of the approach, we consider the problem of approximating a joint pdf $p(\mathbf{x}|\mathcal{M})$ by a separable pdf $q(\mathbf{x}) = \prod_j q_j(x_j)$. The first step to do this approximation is to choose a criterion. A natural criterion is the Kullback-Leibler divergence:

$$\begin{aligned} \text{KL}(q : p) &= \int q(\mathbf{x}) \ln \frac{q(\mathbf{x})}{p(\mathbf{x}|\mathcal{M})} d\mathbf{x} \\ &= -H(q) - \langle \ln p(\mathbf{x}|\mathcal{M}) \rangle_{q(\mathbf{x})} \\ &= -\sum_j H(q_j) - \langle \ln p(\mathbf{x}|\mathcal{M}) \rangle_{q(\mathbf{x})} \end{aligned} \quad (4.4)$$

So, the main mathematical problem to study is finding $\hat{q}(\mathbf{x})$ which minimizes $\text{KL}(q : p)$. We may first note two points: a) The optimal solution without any constraint is the trivial solution $\hat{q}(\mathbf{x}) = p(\mathbf{x})$; b) The optimal solution with the constraint $\langle \ln p(\mathbf{x}|\mathcal{M}) \rangle_{q(\mathbf{x})} = c$ where c is a given constant value is the one which maximizes the entropy

$H(q)$ and is given by

$$q_j(x_j) = \frac{1}{C_j} \exp \left[- \langle \ln p(\mathbf{x}|\mathcal{M}) \rangle_{q_{-j}} \right] \quad (4.5)$$

where $q_{-j} = \prod_{i \neq j} q_i(x_i)$ and C_j are the normalizing factors.

However, we may note that, first the expression of $q_j(x_j)$ depends on the expressions of $q_i(x_i), i \neq j$. Thus the computation can only be done in an iterative way. The second point is that to be able to compute these solutions we must be able to compute $\langle \ln p(\mathbf{x}|\mathcal{M}) \rangle_{q_{-j}}$. The only family for which these computations can be done in an easy way is the conjugate exponential family.

Looking at the expression of the joint posterior law $p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}|\mathbf{g})$, one solution is to approximate it by $q(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) = q_1(\mathbf{f}) q_2(\mathbf{z}) q_3(\boldsymbol{\theta})$ breaking all the dependencies. Another solution that we propose and which keeps the strong dependencies of \mathbf{f} and \mathbf{z} , but breaks only the weak dependencies of $\boldsymbol{\theta}$ on \mathbf{f} and \mathbf{z} is to choose

$$q(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) = q_1(\mathbf{f}|\mathbf{z}) q_2(\mathbf{z}) q_3(\boldsymbol{\theta}) \quad (4.6)$$

This is the solution we keep and now we detail its application with the four prior models given in the previous section.

Case 1: IMIG :

In the first case, looking at the structure of the first prior model

$$\begin{cases} p(\mathbf{f}|\mathbf{z}) &= \prod_{\mathbf{r}} p(f(\mathbf{r})|z(\mathbf{r})) &= \prod_{\mathbf{r}} \mathcal{N}(m_z(\mathbf{r}), v_z(\mathbf{r})) \\ p(\mathbf{z}) &= \prod_{\mathbf{r}} p(z(\mathbf{r})) &= \prod_{\mathbf{r}} \alpha_z(\mathbf{r}) = \prod_k \alpha_k^{n_k} \end{cases} \quad (4.7)$$

suggests us the following structure for the approximating probability law:

$$\begin{cases} q(\mathbf{f}|\mathbf{z}) &= \prod_{\mathbf{r}} q(f(\mathbf{r})|z(\mathbf{r})) &= \prod_{\mathbf{r}} \mathcal{N}(\hat{\mu}_z(\mathbf{r}), \hat{v}_z(\mathbf{r})) \\ q(\mathbf{z}) &= \prod_{\mathbf{r}} q(z(\mathbf{r})) &= \prod_{\mathbf{r}} \hat{\alpha}_z(\mathbf{r}) = \prod_k \hat{\alpha}_k^{n_k} \end{cases} \quad (4.8)$$

where we defined $\alpha_z(\mathbf{r}) = \alpha_k, \forall \mathbf{r} \in \mathcal{R}_k$ as well as $m_z(\mathbf{r}) = m_k, \forall \mathbf{r} \in \mathcal{R}_k$ and $v_z(\mathbf{r}) = v_k, \forall \mathbf{r} \in \mathcal{R}_k$.

Now choosing for the hyperparameters $\boldsymbol{\theta} = \{\theta_\epsilon, \{m_k\}, \{v_k\}, \{\alpha_k\}\}$

$$q(\boldsymbol{\theta}) = q(\theta_\epsilon) q(m_k) q(v_k) q(\alpha_k) \quad (4.9)$$

with

$$\begin{cases} q(\theta_\epsilon|\hat{\alpha}_\epsilon, \hat{\beta}_\epsilon) &= \mathcal{G}(\hat{\alpha}_\epsilon, \hat{\beta}_\epsilon), \\ q(m_k|\hat{m}_k, \hat{v}_k) &= \mathcal{N}(\hat{m}_k, \hat{v}_k), \forall k \\ q(v_k|\hat{a}_k, \hat{b}_k) &= \mathcal{IG}(\hat{a}_k, \hat{b}_k), \forall k \\ q(\alpha_k) &\propto \hat{\alpha}_k^{n_k} \\ q(\boldsymbol{\alpha}) &\propto \mathcal{D}(\hat{\alpha}_1, \dots, \hat{\alpha}_K) \end{cases} \quad (4.10)$$

The expressions of $\hat{\mu}_z(\mathbf{r}), \hat{v}_z(\mathbf{r}), \hat{\mu}_k, \hat{v}_k, \hat{m}_k, \hat{v}_k, \hat{a}_k, \hat{b}_k$ and $\hat{\alpha}_k$ are obtained by optimizing the free energy. Here, we omit them, but they can be found in (Ayasso and Mohammad-Djafari, 2007; Ayasso and Mohammad-Djafari, 2008).

Case 2: IMG M : In this case, noting that we had

$$\begin{aligned}
p(\mathbf{f}|\mathbf{z}) &= \prod_{\mathbf{r}} \mathcal{N}(f(\mathbf{r})|\mu_z(\mathbf{r}), v_z(\mathbf{r})) \\
\mu_z(\mathbf{r}) &= \frac{1}{|\mathcal{V}(\mathbf{r})|} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \mu_z^*(\mathbf{r}') \\
\mu_z^*(\mathbf{r}') &= \delta(z(\mathbf{r}') - z(\mathbf{r})) f(\mathbf{r}') + (1 - \delta(z(\mathbf{r}') - z(\mathbf{r}))) m_z(\mathbf{r}') \\
v_z(\mathbf{r}) &= v_k, \quad \forall \mathbf{r} \in \mathcal{R}_k
\end{aligned}$$

we propose the following :

$$\begin{aligned}
q(\mathbf{f}|\mathbf{z}) &= \prod_{\mathbf{r}} \mathcal{N}(f(\mathbf{r})|\hat{\mu}_z(\mathbf{r}), \hat{v}_z(\mathbf{r})) \\
\hat{\mu}_z(\mathbf{r}) &= \frac{1}{|\mathcal{V}(\mathbf{r})|} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \hat{\mu}_z^*(\mathbf{r}') \\
\hat{\mu}_z^*(\mathbf{r}') &= \delta(z(\mathbf{r}') - \hat{z}(\mathbf{r})) \hat{f}(\mathbf{r}') + (1 - \delta(z(\mathbf{r}') - \hat{z}(\mathbf{r}))) \hat{m}_z(\mathbf{r}') \\
\hat{v}_z(\mathbf{r}) &= \hat{v}_k, \quad \forall \mathbf{r} \in \mathcal{R}_k
\end{aligned}$$

where, again, the expressions of $\hat{\mu}_z(\mathbf{r})$, $\hat{v}_z(\mathbf{r})$, \hat{m}_z , \hat{v}_k , \hat{a}_k , \hat{b}_k and $\hat{\alpha}_k$ have to found.

Case 3: MGP : In this cas, we had

$$\begin{cases} p(\mathbf{f}|\mathbf{z}) = \prod_{\mathbf{r}} p(f(\mathbf{r})|z(\mathbf{r})) = \prod_{\mathbf{r}} \mathcal{N}(m_z(\mathbf{r}), v_z(\mathbf{r})) \\ p(\mathbf{z}) = \prod_{\mathbf{r}} p(z(\mathbf{r})|\bar{z}(\mathbf{r}'), \mathbf{r}' \in \mathcal{V}(\mathbf{r})) \propto \prod_{\mathbf{r}} \exp[\gamma \delta(z(\mathbf{r}) - z(\mathbf{r}'))] \end{cases} \quad (4.11)$$

and again, naturally, we propose

$$\begin{cases} q(\mathbf{f}|\mathbf{z}) = \prod_{\mathbf{r}} q(f(\mathbf{r})|z(\mathbf{r})) = \prod_{\mathbf{r}} \mathcal{N}(\hat{\mu}_z(\mathbf{r}), \hat{v}_z(\mathbf{r})) \\ q(\mathbf{z}) = \prod_{\mathbf{r}} q(z(\mathbf{r})|\bar{z}(\mathbf{r}'), \mathbf{r}' \in \mathcal{V}(\mathbf{r})) \propto \prod_{\mathbf{r}} \exp[\gamma \delta(z(\mathbf{r}) - \hat{z}(\mathbf{r}'))] \end{cases} \quad (4.12)$$

This corresponds to the Mean Field Approximation (MFA) of the Potts Markov Field which consists in replacing $z(\mathbf{r}')$ by $\hat{z}(\mathbf{r}')$ computed in previous iteration. This MFA is a classical approximation for the Potts model.

Case 4: MGMP : In this last case, combining the descriptions of Case 2 and Case 3, we propose

$$\begin{cases} p(\mathbf{f}|\mathbf{z}) = \prod_{\mathbf{r}} \mathcal{N}(f(\mathbf{r})|\hat{\mu}_z(\mathbf{r}), \hat{v}_z(\mathbf{r})) \\ \hat{\mu}_z(\mathbf{r}) = \frac{1}{|\mathcal{V}(\mathbf{r})|} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \hat{\mu}_z^*(\mathbf{r}') \\ \hat{\mu}_z^*(\mathbf{r}') = \delta(z(\mathbf{r}') - \hat{z}(\mathbf{r})) \hat{f}(\mathbf{r}') + (1 - \delta(z(\mathbf{r}') - \hat{z}(\mathbf{r}))) \hat{m}_z(\mathbf{r}') \\ \hat{v}_z(\mathbf{r}) = \hat{v}_k, \quad \forall \mathbf{r} \in \mathcal{R}_k \\ q(\mathbf{z}) = \prod_{\mathbf{r}} q(z(\mathbf{r})|\bar{z}(\mathbf{r}'), \mathbf{r}' \in \mathcal{V}(\mathbf{r})) \propto \prod_{\mathbf{r}} \exp[\gamma \delta(z(\mathbf{r}) - \hat{z}(\mathbf{r}'))] \end{cases} \quad (4.13)$$

The following two tables summarize these expressions of prior laws and posterior laws.

Case	IMIG	IMGM	MGP	GMP
$p(\mathbf{f} \mathbf{z})$	$\prod_{\mathbf{r}} \mathcal{N}(m_z(\mathbf{r}), v_z(\mathbf{r}))$	Gauss-Markov	$\prod_{\mathbf{r}} \mathcal{N}(m_z(\mathbf{r}), v_z(\mathbf{r}))$	Gauss-Markov
$p(\mathbf{z})$	$\prod_{\mathbf{r}} \alpha_z(\mathbf{r}) = \prod_k \alpha_k^{n_k}$	$\prod_{\mathbf{r}} \alpha_z(\mathbf{r}) = \prod_k \alpha_k^{n_k}$	Potts	Potts
$p(m_k)$	$\mathcal{N}(m_0, v_0)$	$\mathcal{N}(m_0, v_0)$	$\mathcal{N}(m_0, v_0)$	$\mathcal{N}(m_0, v_0)$
$p(v_k)$	$\mathcal{IG}(a_0, b_0)$	$\mathcal{IG}(a_0, b_0)$	$\mathcal{IG}(a_0, b_0)$	$\mathcal{IG}(a_0, b_0)$
$p(\alpha)$	$\mathcal{D}(\alpha_0, \dots, \alpha_0)$	$\mathcal{D}(\alpha_0, \dots, \alpha_0)$	$\mathcal{D}(\alpha_0, \dots, \alpha_0)$	$\mathcal{D}(\alpha_0, \dots, \alpha_0)$
$p(\theta_\epsilon)$	$\mathcal{G}(a_{\epsilon_0}, b_{\epsilon_0})$	$\mathcal{G}(a_{\epsilon_0}, b_{\epsilon_0})$	$\mathcal{G}(a_{\epsilon_0}, b_{\epsilon_0})$	$\mathcal{G}(a_{\epsilon_0}, b_{\epsilon_0})$

Table 1: Priors for different models.

Case	IMIG	IMGM	MGP	GMP
$q(\mathbf{f} \mathbf{z})$	$\prod_{\mathbf{r}} \mathcal{N}(\hat{\mu}_z(\mathbf{r}), \hat{v}_z(\mathbf{r}))$	$\prod_{\mathbf{r}} \mathcal{N}(\hat{\mu}_z(\mathbf{r}), \hat{v}_z(\mathbf{r}))$	$\prod_{\mathbf{r}} \mathcal{N}(m_z(\mathbf{r}), v_z(\mathbf{r}))$	$\prod_{\mathbf{r}} \mathcal{N}(\hat{\mu}_z(\mathbf{r}), \hat{v}_z(\mathbf{r}))$
$q(\mathbf{z})$	$\prod_{\mathbf{r}} \hat{\alpha}_z(\mathbf{r}) = \prod_k \hat{\alpha}_k^{n_k}$	$\prod_{\mathbf{r}} \hat{\alpha}_z(\mathbf{r}) = \prod_k \hat{\alpha}_k^{n_k}$	Potts	Potts
$q(m_k)$	$\mathcal{N}(\hat{m}_k, \hat{v}_k)$	$\mathcal{N}(\hat{m}_k, \hat{v}_k)$	$\mathcal{N}(\hat{m}_k, \hat{v}_k)$	$\mathcal{N}(\hat{m}_k, \hat{v}_k)$
$q(v_k)$	$\mathcal{IG}(\hat{a}_k, \hat{b}_k)$	$\mathcal{IG}(\hat{a}_k, \hat{b}_k)$	$\mathcal{IG}(\hat{a}_k, \hat{b}_k)$	$\mathcal{IG}(\hat{a}_k, \hat{b}_k)$
$q(\alpha)$	$\mathcal{D}(\hat{\alpha}_1, \dots, \hat{\alpha}_K)$	$\mathcal{D}(\hat{\alpha}_1, \dots, \hat{\alpha}_K)$	$\mathcal{D}(\hat{\alpha}_1, \dots, \hat{\alpha}_K)$	$\mathcal{D}(\hat{\alpha}_1, \dots, \hat{\alpha}_K)$
$q(\theta_\epsilon)$	$\mathcal{G}(\hat{a}_{\epsilon_0}, \hat{b}_{\epsilon_0})$	$\mathcal{G}(\hat{a}_{\epsilon_0}, \hat{b}_{\epsilon_0})$	$\mathcal{G}(\hat{a}_{\epsilon_0}, \hat{b}_{\epsilon_0})$	$\mathcal{G}(\hat{a}_{\epsilon_0}, \hat{b}_{\epsilon_0})$

Table 2: Proposed approximations for different models.

5 Numerical experiment results

Some details on the geometry and applications in Non Destructive Testing (NDT) and other applications can be found in (Mohammad-Djafari, 2007; Humblot and Mohammad-Djafari, 2006; Féron, Duchêne and Mohammad-Djafari, 2005; Snoussi and Mohammad-Djafari, 2004b; Mohammad-Djafari and Robillard, 2006; Mohammad-Djafari, 2002c; Mohammad-Djafari, 2002b; Mohammad-Djafari, 2002a; Ayasso and Mohammad-Djafari, 2008; Mohammad-Djafari, 2008; Bali and Mohammad-Djafari, 2008; Mohammad-Djafari, 2007; Féron, Duchêne and Mohammad-Djafari, 2007).

Here, we report a few results in 2D case, just to show the role of the prior modeling. The first example here is a 2D case reconstruction problem with only two projections. Figure 4 shows a 2D object \mathbf{f} with its two projections \mathbf{g} , its labels \mathbf{z} and its contours \mathbf{c} and Figure 5 shows a typical result which can be obtained by different methods.

6 Conclusion

In this paper we proposed to use different Gauss-Markov-Potts prior models for images to be used in imaging inverse problem of CT. These priors are appropriate tools to translate a prior information that the object under test is composed of a finite number of materials. We used these prior models in a Bayesian estimation framework to propose image reconstruction methods which perform reconstruction and segmentation jointly. However, to be able to implement these methods, we proposed to use either MCMC sampling schemes or the variational

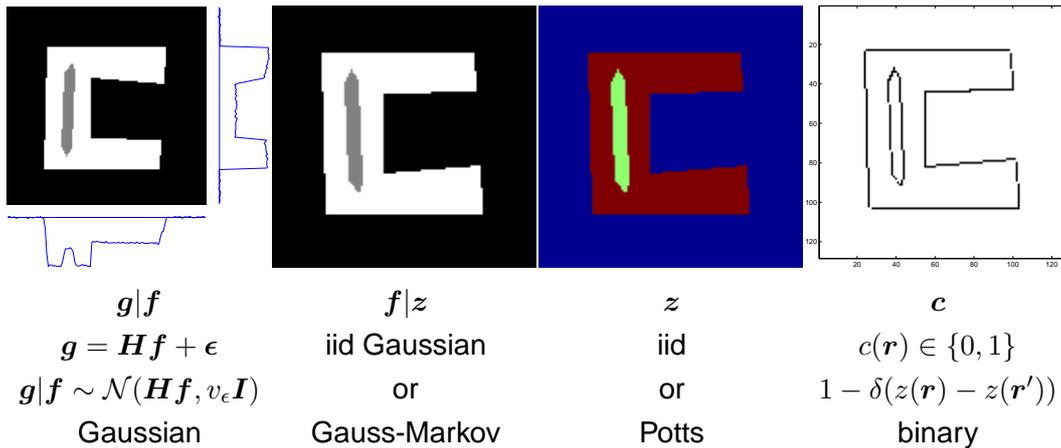


Figure 3: A 2D object f , its two projections g , its material class labels z and its contours c .

Bayesian separable approximations of the joint posterior law of all the unknowns, *i.e.* the image pixel values, the hidden label variables of segmentation and all the hyperparameters of the prior laws. We thus developed iterative algorithms with more reasonable computational cost to compute the posterior means. Finally, we used the proposed models and methods in CT with very small number of projections.

Acknowledgment

I would like to thank Hachem Ayasso for a careful reading of the revised version of this paper.

References

- Ayasso, H. and Mohammad-Djafari, A. 2007. Approche bayésienne variationnelle pour les problèmes inverses. application en tomographie microonde, *Technical report*, Rapport de stage Master ATS, Univ Paris Sud, L2S, SUPELEC.
- Ayasso, H. and Mohammad-Djafari, A. 2008. Variational bayes with gauss-markov-potts prior models for joint image restoration and segmentation, *Visapp Proceedings*, Int. Conf. on Computer Vision and Applications, Funchal, Madaira, Portugal.
- Bali, N. and Mohammad-Djafari, A. 2008. Bayesian approach with hidden markov modeling and mean field approximation for hyperspectral data analysis, *IEEE Trans. on Image Processing* **17**(2): 217–225.
- Brooks, R. A. and Di Chiro, G. 1975. Theory of image reconstruction in computed tomography, *Radiology* **117**: 561–572.
- Budinger, T. F., Gullberg, W. L. and Huesman, R. H. 1979. Emission computed tomography, in G. T. Herman (ed.), *Image Reconstruction from Projections: Implementation and Application*, Springer Verlag, New York, pp. 147–246.

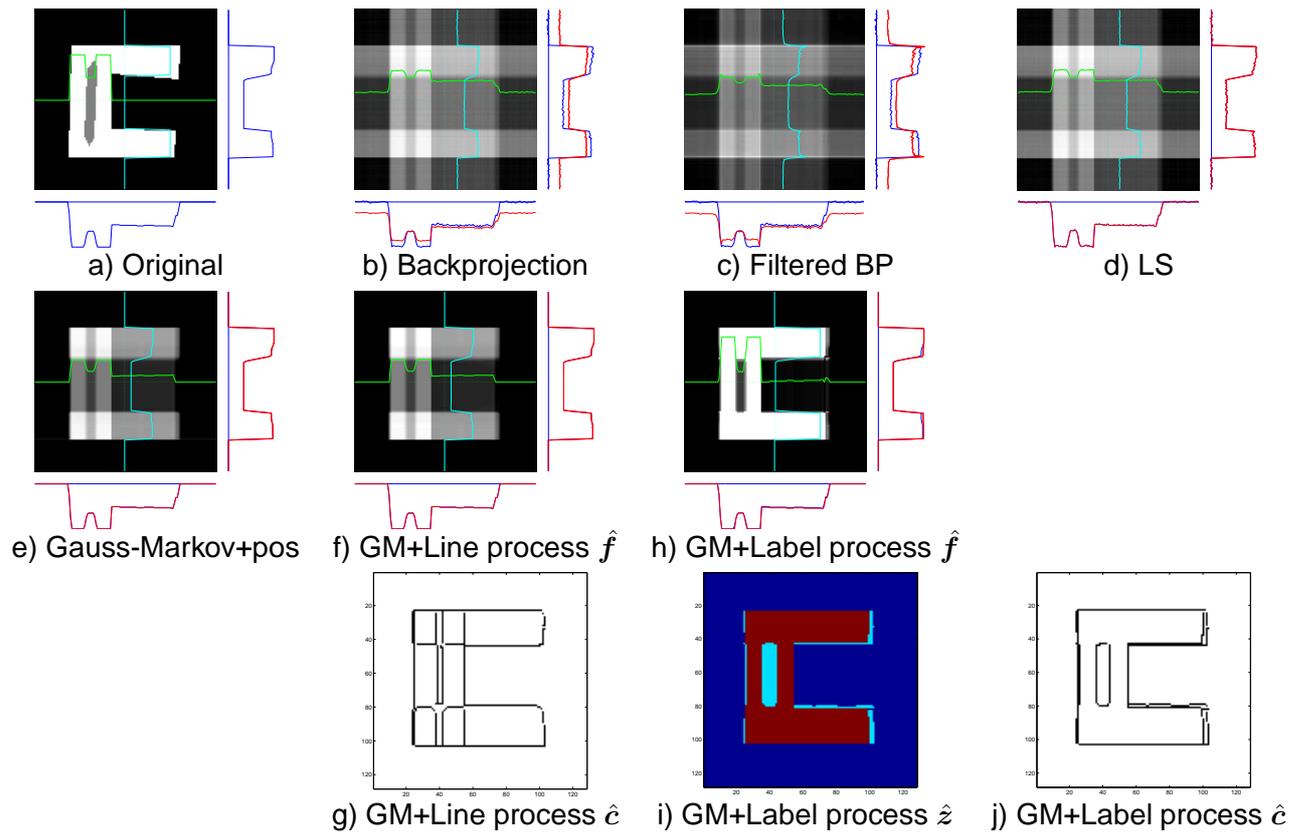


Figure 4: A typical example of reconstruction results using different methods. In this figure, a) original object, b) reconstruction by backprojection, c) reconstruction by filtered backprojection, d) reconstruction by Least Squares, e) reconstruction by a Gauss-Markov (GM) prior and MAP estimate with positivity constraint, f) and g) show \hat{f} and \hat{c} using a Bayesian JMAP with a GM with hidden line process and h), i) and j) show \hat{f} , \hat{z} and \hat{c} using the proposed Gauss-Markov-Potts model and an MCMC based algorithm.

Cassidy, M. and Penny, W. 2002. Bayesian nonstationary autoregressive models for biomedical signal analysis, *IEEE Transactions on Biomedical Engineering* **49**(10): 1142–1152.

Choudrey, R. and Roberts, S. 2003. Variational Mixture of Bayesian Independent Component Analysers, *Neural Computation* **15**(1).

Eggermont, P. and Herman, G. 1981. Iterative algorithms for large partitioned linear systems, with applications to image reconstruction, *Linear Algebra and Its Applications* **40**: 37–67.

Féron, O., Duchêne, B. and Mohammad-Djafari, A. 2005. Microwave imaging of inhomogeneous objects made of a finite number of dielectric and conductive materials from experimental data, *Inverse Problems* **21**(6): 95–115.

Féron, O., Duchêne, B. and Mohammad-Djafari, A. 2007. Microwave imaging of piecewise constant objects in a 2D-TE configuration, *International Journal of Applied Electromagnetics and Mechanics* **26**(6): 167–174.

- Féron, O. and Mohammad-Djafari, A. 2005. Image fusion and joint segmentation using an MCMC algorithm, *Journal of Electronic Imaging* **14**(2): paper no. 023014.
- Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J. and Penny, W. 2006. Variational free energy and the laplace approximation, *Neuroimage* (2006.08.035). Available Online.
- Ghahramani, Z. and Jordan, M. 1997. Factorial Hidden Markov Models, *Machine Learning* (29): 245–273.
- Hanson, K. M. and Wechsung, G. W. 1983. Bayesian approach to limited-angle reconstruction in computed tomography, *Journal of the Optical Society of America* **73**: 1501–1509.
- Herman, G. T. and Lent, A. 1976. Quadratic optimization for image reconstruction I, *Computer Graphics and Image Processing* **5**: 319–332.
- Humblot, F. and Mohammad-Djafari, A. 2006. Super-Resolution using Hidden Markov Model and Bayesian Detection Estimation Framework, *EURASIP Journal on Applied Signal Processing* **Special number on Super-Resolution Imaging: Analysis, Algorithms, and Applications**: ID 36971, 16 pages.
*<http://www.hindawi.com/GetArticle.aspx?doi=10.1155/ASP/2006/36971>
- Jaakkola, T. S. and Jordan, M. I. 2000. Bayesian parameter estimation via variational methods, *Statistics and Computing* **10**(1): 25–37.
- MacKay, D. J. C. 1992. A practical Bayesian framework for backpropagation networks, *Neural Computation* **4**(3): 448–472.
- Miskin, J. W. and MacKay, D. J. C. 2001. Ensemble learning for blind source separation, in S. Roberts and R. Everson (eds), *ICA: Principles and Practice*, Cambridge University Press, Cambridge.
- Mohammad-Djafari, A. 2002a. Bayesian approach with hierarchical markov modeling for data fusion in image reconstruction applications, *Fusion 2002, 7-11 Jul., Annapolis, Maryland, USA*.
- Mohammad-Djafari, A. 2002b. Fusion of x ray and geometrical data in computed tomography for non destructive testing applications, *Fusion 2002, 7-11 Jul., Annapolis, Maryland, USA*.
- Mohammad-Djafari, A. 2002c. Hierarchical markov modeling for fusion of x ray radiographic data and anatomical data in computed tomography, *Int. Symposium on Biomedical Imaging (ISBI 2002), 7-10 Jul., Washington DC, USA*.
- Mohammad-Djafari, A. 2007. Bayesian inference for inverse problems in signal and image processing and applications, *International Journal of Imaging Systems and Technology* **16**: 209–214.
- Mohammad-Djafari, A. 2008. Super-resolution: A short review, a new method based on hidden markov modeling of hr image and future challenges, *The Computer Journal* doi:10,1093/comjnl/bxn005.

- Mohammad-Djafari, A. and Robillard, L. 2006. Hierarchical Markovian models for 3D computed tomography in non destructive testing applications, *EUSIPCO 2006*, EUSIPCO 2006, September 4-8, Florence, Italy.
- Nasios, N. and Bors, A. 2004. A variational approach for Bayesian blind image deconvolution, *IEEE Transactions on Signal Processing* **52**(8): 2222–2233.
- Nasios, N. and Bors, A. 2006. Variational learning for gaussian mixture models, *IEEE Transactions on Systems, Man and Cybernetics, Part B* **36**(4): 849–862.
- Penny, W., Everson, R. and Roberts, S. 2000. Hidden markov independent component analysis, in M. Girolami (ed.), *Advances in Independent Component Analysis*, Springer.
- Penny, W. and Friston, K. 2003. Mixtures of general linear models for functional neuroimaging, *IEEE Transactions on Medical Imaging* **22**(4): 504–514.
- Penny, W., Kiebel, S. and Friston, K. 2003. Variational Bayesian inference for fmri time series, *NeuroImage* **19**(3): 727–741.
- Penny, W., Kiebel, S. and Friston, K. 2006. Variational bayes, in K. Friston, J. Ashburner, S. Kiebel, T. Nichols and W. Penny (eds), *Statistical Parametric Mapping: The analysis of functional brain images*, Elsevier, London.
- Penny, W. and Roberts, S. 1998. Bayesian neural networks for classification: how useful is the evidence framework ?, *Neural Networks* **12**: 877–892.
- Penny, W. and Roberts, S. 1999. Dynamic models for nonstationary signal segmentation, *Computers and Biomedical Research* **32**(6): 483–502.
- Penny, W. and Roberts, S. 2002. Bayesian multivariate autoregressive models with structured priors, *IEE Proceedings on Vision, Image and Signal Processing* **149**(1): 33–41.
- Roberts, S., Husmeier, D., Penny, W. and Rezek, I. 1998. Bayesian approaches to gaussian mixture modelling, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(11): 1133–1142.
- Roberts, S. and Penny, W. 2002. Variational bayes for generalised autoregressive models, *IEEE Transactions on Signal Processing* **50**(9): 2245–2257.
- Snoussi, H. and Mohammad-Djafari, A. 2004a. Bayesian unsupervised learning for source separation with mixture of Gaussians prior, *Journal of VLSI Signal Processing Systems* **37**(2/3): 263–279.
- Snoussi, H. and Mohammad-Djafari, A. 2004b. Fast joint separation and segmentation of mixed images, *Journal of Electronic Imaging* **13**(2): 349–361.