HAL
open science

Bayesian 3D X-ray computed tomography image reconstruction with a scaled Gaussian mixture prior model

Li Wang, Nicolas Gac, Ali Mohammad-Djafari

To cite this version:

Li Wang, Nicolas Gac, Ali Mohammad-Djafari. Bayesian 3D X-ray computed tomography image reconstruction with a scaled Gaussian mixture prior model. 34th International Workshop on Bayesian Inference and Maximun Entropy Methods in Science and Engineering (MaxEnt'14), Sep 2014, Amboise, France. pp.556-563, 10.1063/1.4906022 . hal-01338706

HAL Id: hal-01338706

https://hal.science/hal-01338706v1

Submitted on 29 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bayesian 3D X-ray Computed Tomography image reconstruction with a Scaled Gaussian Mixture prior model

Li WANG[1], Nicolas GAC and Ali MOHAMMAD-DJAFARI

Laboratoire des Signaux et Systèmes 3, Rue Joliot-Curie 91192 Gif sur Yvette

Abstract. In order to improve quality of 3D X-ray tomography reconstruction for Non Destructive Testing (NDT), we investigate in this paper hierarchical Bayesian methods. In NDT, useful prior information on the volume like the limited number of materials or the presence of homogeneous area can be included in the iterative reconstruction algorithms. In hierarchical Bayesian methods, not only the volume is estimated thanks to the prior model of the volume but also the hyper parameters of this prior. This additional complexity in the reconstruction methods when applied to large volumes (from 512^3 to 8192^3 voxels) results in an increasing computational cost. To reduce it, the hierarchical Bayesian methods investigated in this paper lead to an algorithm acceleration by Variational Bayesian Approximation (VBA) [1] and hardware acceleration thanks to projection and back-projection operators paralleled on many core processors like GPU [2].

In this paper, we will consider a Student-t prior on the gradient of the image implemented in a hierarchical way [3, 4, 1]. Operators H (forward or projection) and H^t (adjoint or back-projection) implanted in multi-GPU [2] have been used in this study. Different methods will be evalued on synthetic volume "Shepp and Logan" in terms of quality and time of reconstruction. We used several simple regularizations of order 1 and order 2. Other prior models also exists [5]. Sometimes for a discrete image, we can do the segmentation and reconstruction at the same time, then the reconstruction can be done with less projections.

Keywords: Computed Tomography, Limited projections, Non Destructive Testing (NDT), Hierarchical Model, Bayesian JMAP, Variational Bayesian Approximation (VBA), Gaussian, Mixture of Gaussians (MoG) and Student-t prior models

INTRODUCTION OF COMPUTED TOMOGRAPHY

Computed Tomography

X-ray computed tomography (X-ray CT) is a technology that uses computer-processed X-ray to produce tomographic images of specific areas of a scanned object, allowing the users to see what is inside it without cutting it. Digital geometry processing is used to generate a three-dimensional image of the inside of an object from a large series of two-dimensional radiographic images taken around a single axis of rotation. X-ray tomographic image reconstruction consist of determining an object function from its projections. The main forward problem of the tomography is the Radon transform.

When considering a practical problem, a discretization is necessary. If we discretize

---

[1] China Scholarship Council

$f(x,y)$ into pixels and put all the pixels in a vector $\boldsymbol{f}$ and put all the data $g(\boldsymbol{\phi},\boldsymbol{r})$ of different angles $\boldsymbol{\phi}$ in a vector $\boldsymbol{g}$, we then obtain:

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{\varepsilon}$ represents the error, and $\boldsymbol{H}$ is the projection operator in which the element $\boldsymbol{H}_{ij}$ represents the length of the ray i in the pixel j.

BAYESIAN APPROACH

From this point, the main objective is to infer on $\boldsymbol{f}$ given the data $\boldsymbol{g}$ assuming the forward model $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\varepsilon}$. By being Bayesian, we mean to use the Bayes rule:

$$p(\boldsymbol{f} \mid \boldsymbol{g}) = \frac{p(\boldsymbol{g} \mid \boldsymbol{f})p(\boldsymbol{f})}{p(\boldsymbol{g})} \propto p(\boldsymbol{g} \mid \boldsymbol{f})p(\boldsymbol{f}) \tag{2}$$

to obtain what is called the posterior law $p(\boldsymbol{f}|\boldsymbol{g})$ from the likelihood $p(\boldsymbol{g}|\boldsymbol{f})$ and the prior $p(\boldsymbol{f})$. To be able to use the Bayesian approach, first we need to assign $p(\boldsymbol{g}|\boldsymbol{f})$ and $p(\boldsymbol{f})$. Then, we can obtain the expression of the posterior law. Finally, we can infer on $\boldsymbol{f}$ using this posterior law.

Markov model. In the Markov model, the value of f_j has a relationship with the values of its neighbours. For example, in the case of 1D, $f_j = F(f_{j1}, f_{j2}, \cdots)$, and in an image, the value of a pixel depends on the values of its neighbours pixels.

When the value of f_j depends only on the values of the neighbours of distance 1, the model is called to have order 1, and for a Gaussian model we have:

$$p(f_j \mid f_{j-1}, \sigma^2) = \mathcal{N}\left(f_j \mid f_{j-1}, \sigma^2\right) \propto \exp\left[-\frac{1}{2}\frac{(f_j - f_{j-1})^2}{\sigma_f^2}\right] \tag{3}$$

From this we can write:

$$p(\boldsymbol{f}) \propto \exp\left\{-\gamma\left|[\boldsymbol{D}\boldsymbol{f}]_j\right|^{\beta}\right\}$$

where $\boldsymbol{D}\boldsymbol{f}$ represents a suitable defined gradient of the image, γ is a scale parameter and β a shape parameter. For $\beta = 2$ a Gauss-Markov model is obtained.

For the noise term $\boldsymbol{\varepsilon}$ we choose a Gaussia prior law. This leads to

$$p(\boldsymbol{g} \mid \boldsymbol{f}) \propto \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right\}$$

and so

$$p(\boldsymbol{f} \mid \boldsymbol{g}, \sigma_f^2, \sigma_\varepsilon^2) \propto \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\boldsymbol{J}(\boldsymbol{f})\right\}$$

with $\boldsymbol{J}(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{D}\boldsymbol{f}\|_\beta^\beta$, where λ is the parameter of the regularization term. When $\beta = 2$ (Gauss-Markov case), we got the analytical solution:

$$\hat{\boldsymbol{f}} = \left(\boldsymbol{H}^t\boldsymbol{H} + \lambda\boldsymbol{D}^t\boldsymbol{D}\right)^{-1}\boldsymbol{H}^t\boldsymbol{g}$$

We have considered different prior laws with Markov model: the Gaussian law, the Generalized Gaussian law, the Cauchy law and the Huber law.

UNSUPERVISED BAYESIAN

In previous section, there are parameters σ_ε^2 and σ_f^2 as well as β which have to be assigned. These are called hyper parameters. For a practical applications we need to estimate them also. In the Bayesian approach this can be done via the joint posterior:

$$p(\boldsymbol{f}, \boldsymbol{\theta} \mid \boldsymbol{g}, \theta_0) = \frac{p(\boldsymbol{g} \mid \boldsymbol{f}, \theta_1) p(\boldsymbol{f} \mid \theta_2) p(\boldsymbol{\theta} \mid \theta_0)}{p(\boldsymbol{g} \mid \theta_0)} \tag{4}$$

where $\boldsymbol{\theta} = [\theta_1, \theta_2]$ and θ_0 is a parameter of θ_1 and θ_2.

In this paper we set $\beta = 2$ but we want to estimate $\theta_1 = v_\varepsilon = \sigma_\varepsilon^2$ and $\theta_2 = v_f = \sigma_f^2$. As θ_1 and θ_2 are variances, we use a conjugate prior for them. Here we choose the inverse Gamma law as the conjugate prior:

$$\begin{cases} p(v_\varepsilon \mid \alpha_\varepsilon, \beta_\varepsilon) = \mathscr{IG}(v_\varepsilon \mid \alpha_\varepsilon, \beta_\varepsilon) \\ p(v_f \mid \alpha_f, \beta_f) = \mathscr{IG}(v_f \mid \alpha_f, \beta_f) \end{cases} \tag{5}$$

One way to estimate the unknowns of our model is to compute the Joint Maximum A Posteriori (JMAP) [6].

$$\left(\hat{\boldsymbol{f}}, \hat{\boldsymbol{\theta}}\right) = \underset{(\boldsymbol{f}, \boldsymbol{\theta})}{\arg\max}\, p\left(\boldsymbol{f}, \boldsymbol{\theta} \mid \boldsymbol{g}, \theta_0\right) \tag{6}$$

In the case where the only hyper parameters are v_ε and v_f, we can apply the following iterative algorithm:

$$\begin{cases} \hat{\boldsymbol{f}} = \arg\max_f p\left(\hat{\boldsymbol{f}}, v_\varepsilon, v_f \mid \boldsymbol{g}, \theta_0\right) = \left(\boldsymbol{H}^t \boldsymbol{H} + \frac{v_\varepsilon}{v_f} \boldsymbol{D}^t \boldsymbol{D}\right)^{-1} \boldsymbol{H}^t \boldsymbol{g} \\ \hat{v_\varepsilon} = \frac{\frac{1}{2}\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \beta_\varepsilon}{\alpha_\varepsilon + \frac{M}{2} + 1} \\ \hat{v_f} = \frac{\beta_f + \frac{1}{2}\|\boldsymbol{D}\boldsymbol{f}\|^2}{\alpha_f + \frac{N}{2} + 1} \end{cases} \tag{7}$$

JMAP AND VBA WITH STUDENT-T PRIOR

In this section, we will consider the hierarchical model which uses a Student-t distribution for modeling the distribution of sparse signals or images.

For an image in which most parts are homogeneous, the gradient of the image is sparse. To enforce sparsity, we propose to use a heavy tailed prior law, for example, the Generalized Gaussian law and the Student-t law. Here we propose to use the Student-t distribution. For the Student-t prior law, we have the property:

$$\boldsymbol{St}\left(f_j \mid \nu, \tau\right) = \int_0^\infty \mathscr{N}\left(f_j \mid 0, \frac{1}{z_j}\right) \mathscr{G}\left(z_j \mid \frac{\nu}{2}, \frac{\nu\tau^2}{2}\right) dz_j$$

where the z_j is a hidden variable which represents the inverse variance of f_j [4, 1].

This property of Infinite Gaussian Mixture gives the possibility to propose the following hierarchical model:

$$\begin{cases} p(f_j \mid z_j) = \mathcal{N}\left(f_j \mid f_{j-1}, \frac{v_f}{z_j}\right) \\ p(z_j \mid \alpha_{z_j}, \beta_{z_j}) = \mathcal{G}\left(z_j \mid \alpha_{z_j}, \beta_{z_j}\right) \end{cases}$$

JMAP. With the hierarchical model, we can obtain the expression of the joint a posterior:

$$p\left(\boldsymbol{f}, \boldsymbol{Z}, \boldsymbol{\theta} \mid \boldsymbol{g}\right) \propto p\left(\boldsymbol{g} \mid \boldsymbol{f}, v_\varepsilon\right) p\left(\boldsymbol{f} \mid \boldsymbol{Z}, v_f\right) p\left(\boldsymbol{Z} \mid \alpha_{\boldsymbol{Z}}, \beta_{\boldsymbol{Z}}\right) p\left(v_\varepsilon\right) p\left(v_f\right)$$

where $p\left(\boldsymbol{f} \mid \boldsymbol{Z}, v_f\right) = \prod_j p\left(f_j \mid z_j, v_f\right)$ and $p\left(\boldsymbol{Z}\right) = \prod_j p\left(z_j\right)$ and for $p\left(v_\varepsilon\right)$ and $p\left(v_f\right)$ we use Inverse Gamma law which is given in (6).

An alternating optimisation of this JMAP criterion results to the following algorithm:

$$\begin{cases} \hat{\boldsymbol{f}}^{(k+1)} = \left(\boldsymbol{H}^t\boldsymbol{H} + \frac{v_\varepsilon^{(k)}}{v_f^{(k)}}\boldsymbol{Z}^{(k)}\right)^{-1}\boldsymbol{H}^t\boldsymbol{g} \\ \hat{z}_j^{(k+1)} = \dfrac{\alpha_{z_j} - \frac{1}{2}}{\beta_{z_j} + \frac{\left(f_j^{(k+1)}\right)^2}{2v_f^{(k)}}} \\ \hat{v}_\varepsilon^{(k+1)} = \dfrac{\frac{1}{2}\left\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}^{(k+1)}\right\|^2 + \beta_\varepsilon}{\alpha_\varepsilon + \frac{M}{2} + 1} \\ \hat{v}_f^{(k+1)} = \dfrac{\frac{1}{2}\sum_{j=1}^N z_j^{(k+1)}\left(f_j^{(k+1)}\right)^2 + \beta_f}{\alpha_f + \frac{N}{2} + 1} \end{cases} \tag{8}$$

VBA. As we will see, the main inconvenience of the JMAP approach is that we are summarizing the joint posterior law $p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})$ by its mode only. Also, for obtaining this mode, in general, an iterative alternating optimization is used, where at each iteration only the values of the estimates at previous iterations are used without accounting for their corresponding uncertainties.

In the VBA approach, the main idea here is to approximate $p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})$ by a separable law $q(\boldsymbol{f}, \boldsymbol{\theta}) = q_1(\boldsymbol{f})q_2(\boldsymbol{\theta}) = \prod_j q_{1j}(f_j)q_2(\boldsymbol{\theta})$ which can then be used for inferring on $\boldsymbol{f}$ or $\boldsymbol{\theta}$. The main criterion used is the Kullback-Leibler divergence[7, 8, 3]. As we will see, $\boldsymbol{f}$ depends on $q_2(\boldsymbol{\theta})$ and $q_2(\boldsymbol{\theta})$ depends on $q_1(\boldsymbol{f})$, thus accounting for uncertainties in both steps.

We assume that $f_j \mid \mu_j, v_{f_j}, z_j \sim \mathcal{N}(f_j \mid \mu_j, \frac{\sigma_{f_j}}{z_j})$, $z_j \mid \alpha_{z_j}, \beta_{z_j} \sim \mathcal{G}(z_j \mid \alpha_{z_j}, \beta_{z_j})$, $v_\varepsilon \mid \alpha_\varepsilon, \beta_\varepsilon \sim \mathcal{IG}(v_\varepsilon \mid \alpha_\varepsilon, \beta_\varepsilon)$ and $v_f \mid \alpha_f, \beta_f \sim \mathcal{IG}(v_f \mid \alpha_f, \beta_f)$. And we are now considering all the hyper parameters, parameters and unknowns. The alternate optimization

of VBA is then given by:

$$\begin{cases}
\tilde{\lambda}_j^{(k+1)} = \frac{\tilde{\alpha}_f^{(k)}}{\tilde{\beta}_f^{(k)}} \frac{\tilde{\beta}_\varepsilon^{(k)}}{\tilde{\alpha}_\varepsilon^{(k)}} \frac{\tilde{\alpha}_{zj}^{(k)}}{\tilde{\beta}_{zj}^{(k)}} \frac{1}{\mathscr{A}} \\
\tilde{\mu_j}^{(k+1)} = -\frac{\mathscr{B}}{\mathscr{A}\left(1+\tilde{\lambda}_j^{(k+1)}\right)} \\
\tilde{\sigma_{f_j}}^{(k+1)} = \frac{1}{\frac{\tilde{\alpha_\varepsilon}^{(k)}}{\tilde{\beta_\varepsilon}^{(k)}} \mathscr{A}\left(1+\tilde{\lambda}_j^{(k+1)}\right)} \\
\tilde{\alpha}_{z_j}^{(k+1)} = \alpha_{z_j}^{(k)} + \frac{1}{2} \\
\tilde{\beta}_{z_j}^{(k+1)} = \beta_{z_j}^{(k)} + \frac{1}{2}\mu_j^{(k+1)^2} \frac{\tilde{\alpha}_f^{(k)}}{\tilde{\beta}_f^{(k)}} \\
\tilde{z}_j^{(k+1)} = \frac{\tilde{\alpha}_{z_j}^{(k+1)}}{\tilde{\beta}_{z_j}^{(k+1)}} \\
\tilde{\alpha}_\varepsilon^{(k+1)} = \alpha_\varepsilon^{(k)} + \frac{M}{2} \\
\tilde{\beta}_\varepsilon^{(k+1)} = \beta_\varepsilon^{(k)} + \frac{1}{2}\left(\left\|\boldsymbol{g} - \boldsymbol{H}\tilde{\boldsymbol{\mu}}^{(k+1)}\right\|^2 + \sum_{j=1}^{N}\left(\boldsymbol{H}^t\boldsymbol{H}\right)_{jj}\tilde{\sigma}_{fj}^{(k+1)}\right) \\
\tilde{\alpha}_f^{(k+1)} = \alpha_f^{(k)} + \frac{N}{2} \\
\tilde{\beta}_f^{(k+1)} = \beta_f^{(k)} + \frac{1}{2}\sum_{j=1}^{N} \frac{\tilde{\alpha}_{zj}^{(k+1)}}{\tilde{\beta}_{zj}^{(k+1)}}\left((\tilde{\mu}_j^{(k+1)})^2 + \tilde{\sigma}_{f_j}^{(k+1)}\right)
\end{cases} \tag{9}$$

where N is the length of the vector $\boldsymbol{f}$, M is the length of the vector $\boldsymbol{g}$, and the definition of $\mathscr{A}$ and $\mathscr{B}$ are:

$$\begin{cases}
\mathscr{A} = \left(\boldsymbol{H}^t\boldsymbol{H}\right)_{jj} \\
\mathscr{B} = -\left(\boldsymbol{H}^t\boldsymbol{g}\right)_j + \left(\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{\mu}\right)_j - \left(\boldsymbol{H}^t\boldsymbol{H}\right)_{jj}\mu_j
\end{cases}$$

Here we need to determine the diagonals elements of the matrix $\boldsymbol{H}^t\boldsymbol{H}$ to optimize the unknowns.

IMPLEMENTATION

In the forward model $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\varepsilon}$, the vector $\boldsymbol{f}$ correspond to the object, $\boldsymbol{H}$ is projection operator and $\boldsymbol{g}$ represents the projections. Here in our simulation, we use the synthetic volume "Shepp and Logan" with size $256 \times 256 \times 256$ voxels. In this report, the projection has been done in 256 angles, and under each angle the screen of detectors will receive an image of size 256×256 pixels. The matrix $\boldsymbol{H}$ corresponds to the matrix of projection, of which the most component are zero. In the problem 3D, we don't know the matrix $\boldsymbol{H}$, but the projection $\boldsymbol{H}\boldsymbol{f}$ and back-projection $\boldsymbol{H}'\boldsymbol{g}$ can be accessed. The term $\boldsymbol{\varepsilon}$ represents the noise which follows a Gaussian law. In the algorithm, the most costly parts of the computation are operations on $\boldsymbol{H}\boldsymbol{f}$ (projection) and the $\boldsymbol{H}^t\boldsymbol{g}$ (back-projection). These two operations are implemented using a GPU.

NUMERICAL RESULTS

In the numerical implementation part, we have compared the JMAP method with the results of differents reconstruction methods: simple back-projection, filtered back-projection, lest square method and Bayes method (MAP) with different regularizations (Simple Gaussian, Generalized Gaussian, Cauchy and Huber). The middle slice after reconstruction of different methods (after 200 iterations) is shown in ***FIGURE* 1**, and we can compare the relatif error of different method, where $\delta_f = \frac{\|\hat{\boldsymbol{f}}-\boldsymbol{f}\|^2}{\|\hat{\boldsymbol{f}}\|^2}$ and $\delta_g = \frac{\|\hat{\boldsymbol{g}}-\boldsymbol{g}\|^2}{\|\hat{\boldsymbol{g}}\|^2}$.

With the theoretical projection $\boldsymbol{H}$ and back-projection $\boldsymbol{H}^t$, the object obtained has a large error, and we can not distinguish the details in the image. With the Feldkamp back-projection, we can obtain an image which is already clear enough to distinguish the different materials, but the boundaries are blurred. When we apply the optimisation with different prior models, we can see that the boundaries are easier to distinguish and different materials are clearly separated. In the JMAP method, the boundaries of different materials are more clear than the others methods.

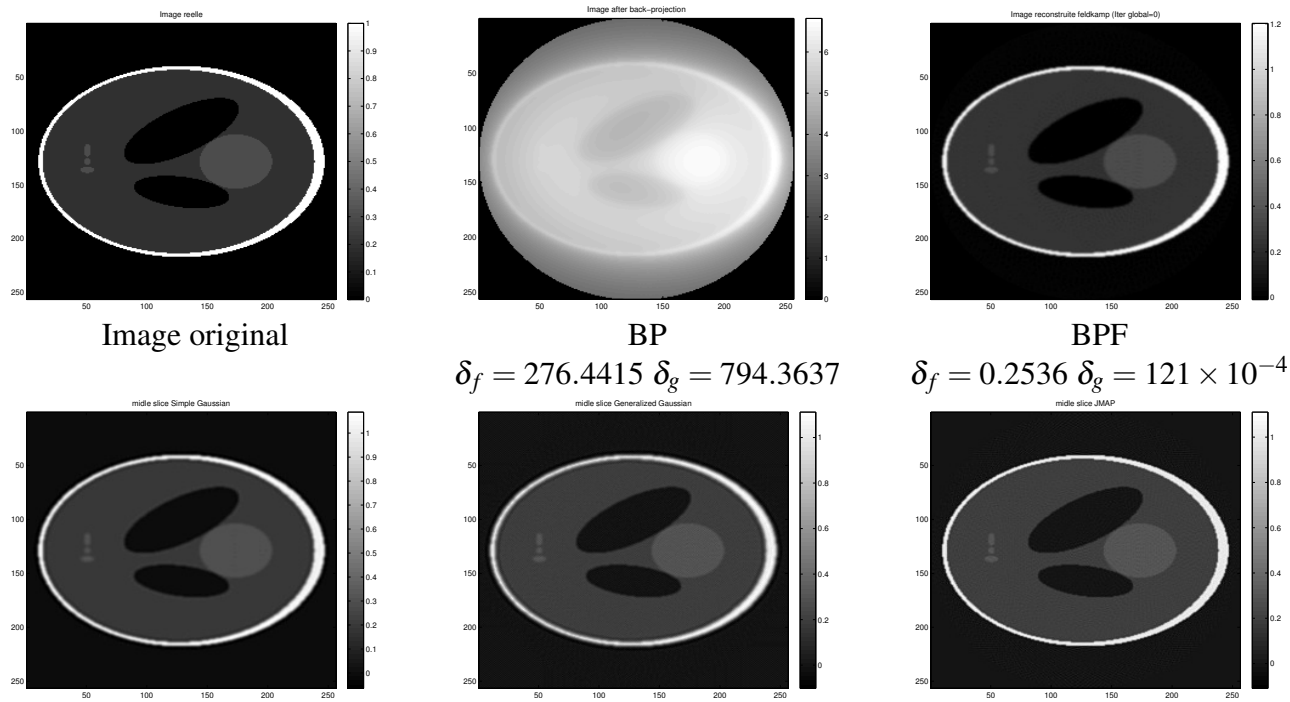

Image original | BP $\delta_f = 276.4415$ $\delta_g = 794.3637$ | BPF $\delta_f = 0.2536$ $\delta_g = 121 \times 10^{-4}$

Simple Gaussian $\delta_f = 0.0599$ $\delta_g = 4.7671 \times 10^{-4}$ | Generalized Gaussian $\delta_f = 0.0790$ $\delta_g = 7.5320 \times 10^{-4}$ | JMAP $\delta_f = 0.0496$ $\delta_g = 3.5128 \times 10^{-4}$

FIGURE 1. Middle slice obtained with different methods.

A more accurate compare is showed in ***FIGURE* 2**, the error of the JMAP method is smaller than the other non-hirarchical methods.

We can see that the method JMAP has the lowest error among the methods that we considered. With the JMAP method, the borders of two different materials are more distinct.

The VBA method for the big data 3D problem is more complicated than the JMAP

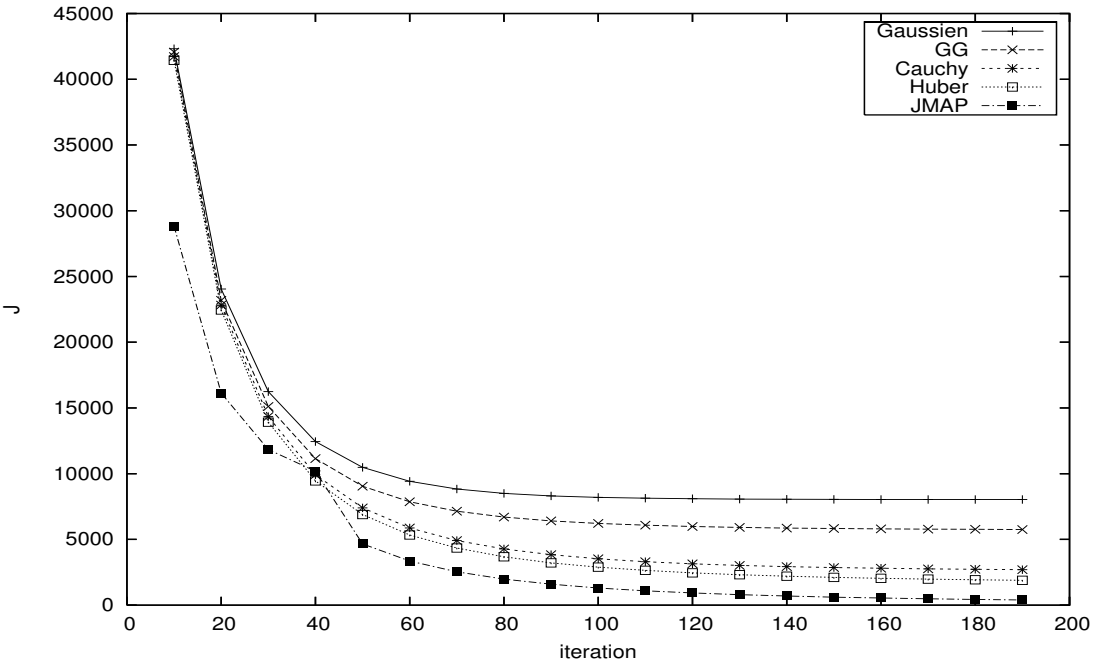

FIGURE 2. The criterion of different method of reconstruction.

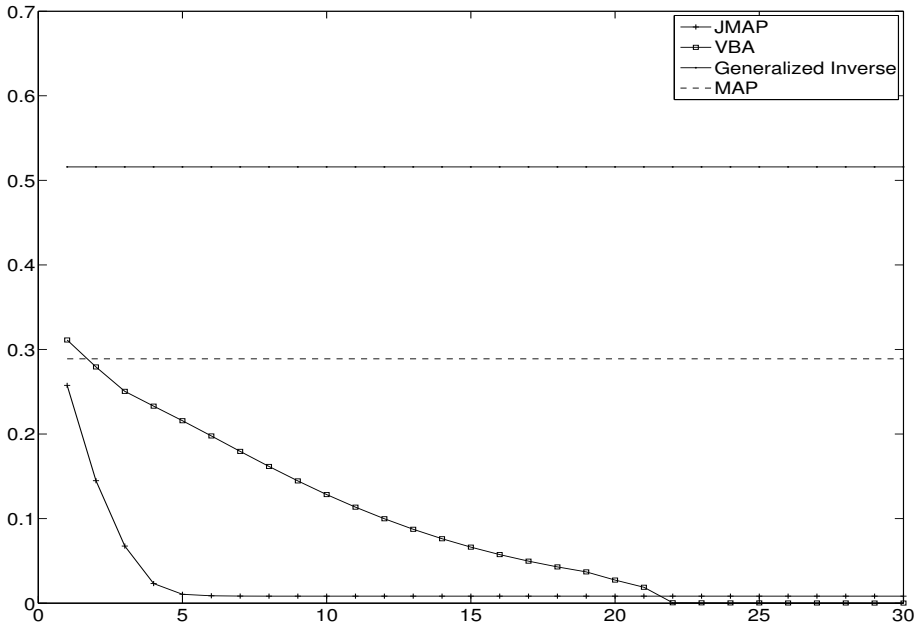

FIGURE 3. The criterion of different method of reconstruction.

method. The calculate of the diagonals elements of the matrix $\boldsymbol{H}^t\boldsymbol{H}$ for the case 3D is very difficult because of the unknown huge size matrix $\boldsymbol{H}$ and the time needed for the projection and back-projection. Thus we have compared the VBA method and other methods for a case 1D where the $\boldsymbol{f}$ is of size $N = 29$, $\boldsymbol{g}$ is of size $M = 96$ and the matrix $\boldsymbol{H}$ is known. The relative error of different methods in this case is showed in $\boldsymbol{figure}$ **3**. From the compare of different method we can see that the VBA method works better than the others.

CONCLUSIONS

In this paper, JMAP method and VBA method are proposed for doing Bayesian computations for inverse problems where a hierarchical prior modeling is used for the unknowns. A Student-t prior model, which can be written via hidden variables, is considered, and it gives the model a hierarchical structure. En comparing the different reconstruction models, we can say that the method JMAP, which considers a hierarchical problem, has a better property than the other methods. The VBA method also considers the hierarchical problem. The difference is that the VBA method will take into account not only the unknown parameters, but also the uncertainties of the unknowns. The main problem of VBA is to calculate the diagonal elements of the matrix $\boldsymbol{H}^t\boldsymbol{H}$. For a cube of $512 \times 512 \times 512$ voxels, it takes more than 10 days to calculate the diagonals elements. The next step is to optimise the coding of projection and back-projection to reduce the time of calculation and compare these two methods.

REFERENCES

1. A. Mohammad-Djafari, "Bayesian inference with hierarchical prior models for inverse problems in imaging systems," in *Systems, Signal Processing and their Applications (WoSSPA), 2013 8th International Workshop on*, IEEE, 2013, pp. 7–18.
2. N. Gac, A. Vabre, A. Mohammad-Djafari, et al., "Multi GPU parallelization of 3D bayesian CT algorithm and its application on real foam reonconstruction with incomplete data set" *Proceedings FVR 2011* pp. 35–38 (2011).
3. R. Molina, A. López, J. M. Martín, and A. K. Katsaggelos, "Variational posterior distribution approximation in bayesian emission tomography reconstruction using a gamma mixture prior.," in *VISAPP (Special Sessions)*, 2007, pp. 165–176.
4. A. Mohammad-Djafari, and L. Robillard, "Hierarchical Markovian models for 3D computed tomography in non destructive testing applications" *EUSIPCO, Florence, Italy* (2006).
5. H. Zou, and T. Hastie,"Regularization and variable selection via the elastic net", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**, 301–320 (2005).
6. A. Mohammad-Djafari,"Gauss-Markov-Potts Priors for Images in Computer Tomography Resulting to Joint Optimal Reconstruction and segmentation" *International Journal of Tomography & Simulation* **11**, 76–92 (2009).
7. C. M. Bishop, and M. E. Tipping, "Variational relevance vector machines," in *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc., 2000, pp. 46–53.
8. A. Miyamoto, K. Watanabe, K. Ikeda, and M.-a. Sato, "Phase diagrams of a variational Bayesian approach with ARD prior in NIRS-DOT," in *Neural Networks (IJCNN), The 2011 International Joint Conference on*, IEEE, 2011, pp. 1230–1236.