Unsupervised sparsity enforcing iterative algorithms for 3D image reconstruction in X-ray Computed Tomography

Mircea Dumitru, Nicolas Gac, Li Wang and Ali Mohammad-Djafari

***Abstract*—Unsupervised iterative reconstruction algorithms based on a Bayesian approach for piecewise constant images are presented in this paper. Such images can be expressed via a sparse representation and the reconstruction problem can be addressed using sparsity enforcing priors. We focus on sparsity enforcing priors expressed as Normal variance mixture, considering three mixing distributions: Inverse Gamma distribution, corresponding to Student-t prior, general inverse Gaussian distribution with the real parameter fixed, corresponding to Normal-inverse Gaussian prior and Gamma distribution corresponding to Variance-Gamma prior. We present and discuss the corresponding iterative algorithms considering the Joint Maximum A Posteriori estimation showing simulations results for 3D X-ray Computed Tomography.**

Keywords*—*sparsity, Normal variance mixtures, Bayesian inference, Computed Tomography

I. Introduction

A widely used discretized linear forward model in the inverse problem approach for image reconstruction is given by

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

where $\boldsymbol{g}$ represents the $N \times 1$ observed data, $\boldsymbol{H}$ represents the $N \times M$ linear projection operator, $\boldsymbol{f}$ represents the $M \times 1$ image to be reconstructed and $\boldsymbol{\epsilon}$ accounts for measurement errors and model uncertainties. Piecewise continuous images $\boldsymbol{f}$ can be expressed as a transformation applied on a sparse structure $\boldsymbol{z}$ accounting for the uncertainties,

$$\boldsymbol{f} = \boldsymbol{D}\boldsymbol{z} + \boldsymbol{\xi}. \tag{2}$$

Equations (1) and (2) represent the linear forward model considered during this paper. For Equation (2) different sparse representations can be considered. In the sparsity context, two main approaches have been successfully used for estimating the unknowns from Equations (1) and (2): deterministic methods [5], [3] and the Bayesian approach [1]. For the deterministic methods, an important issue is the choice of the regularization parameter.

We consider an inversion based on a Bayesian approach, building an hierarchical model, Subsection (II-A) accounting for the sparse structure of $\boldsymbol{z}$, using sparsity enforcing priors. Sparsity enforcing priors expressed via conjugate distributions laws lead to analytical expressions for the unknowns of the hierarchical model therefore we focus on heavy tailed Normal ($\mathcal{N}$) variance mixtures: Student-t ($\mathcal{S}t$) prior, corresponding to Inverse Gamma ($\mathcal{IG}$) mixing distribution, Normal-inverse Gaussian ($\mathcal{NIG}$) distribution, corresponding to generalized inverse Gaussian ($\mathcal{GIG}$) mixing distribution with real parameter fixed, $p = -1/2$ and Variance-Gamma ($\mathcal{VG}$) distribution corresponding to Gamma mixing distribution (with particular case Laplace ($\mathcal{L}$) prior corresponding to Exponential ($\mathcal{E}$) mixing distribution), Subsection (II-B). The likelihood is obtained via the considered linear forward model and the distributions chosen to model the uncertainties $\boldsymbol{\epsilon}$ and $\boldsymbol{\xi}$. The hyperparameters (i.e. the parameters of the assigned variance distributions, $\boldsymbol{\theta}_\epsilon$, $\boldsymbol{\theta}_\xi$ and $\boldsymbol{\theta}_z$) are estimated along with the unknowns of the forward model $\boldsymbol{f}$ and $\boldsymbol{z}$ and the corresponding variances $\boldsymbol{v}_\epsilon$, $\boldsymbol{v}_\xi$ and $\boldsymbol{v}_z$ from the posterior distribution obtained via the proportionality given by the Bayes rule

$$\begin{aligned} p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\xi, \boldsymbol{v}_z|\boldsymbol{g}) \propto & p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{v}_\epsilon)p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{v}_\xi)p(\boldsymbol{z}|\boldsymbol{v}_z)p(\boldsymbol{v}_\epsilon|\boldsymbol{\theta}_\epsilon) \\ & p(\boldsymbol{v}_\xi|\boldsymbol{\theta}_\xi)p(\boldsymbol{v}_z|\boldsymbol{\theta}_z). \end{aligned} \tag{3}$$

The corresponding JAMP iterative algorithms are derived and compared in Section (III). Details of the computations can be found in [2] Simulation results for X-ray Computed Tomography (CT) reconstruction are presented in Section (IV) and conclusions are drawn in Section (V).

II. Hierarchical model and Normal variance mixtures

First, we briefly introduce the construction of the hierarchical model corresponding to the linear forward model Equations (1) and (2), in the context of Normal variance mixtures priors, Subsection (II-A).Then we introduce the sparsity enforcing priors discussed, both corresponding to Normal variance mixtures, Subsection (II-B).

A. Hierarchical model

The construction of the hierarchical model, Figure (1), corresponding to the linear forward model, Equations (1) and (2) is done using the assigned distributions. As mentioned, we consider two Normal variance mixtures: therefore, in both cases $\boldsymbol{f}$, $\boldsymbol{z}$ are modelled as zero mean multivariate $\mathcal{N}$ distribution, deriving the likelihoods via the linear forward model. For the variances corresponding to the uncertainties we consider the $\mathcal{IG}$ and the $\mathcal{GIG}$ distributions.

M. Dumitru, N. Gac, L. Wang and A. Mohammad-Djafari are with Laboratoire des signaux et systèmes, CentraleSupélec, Gif-sur-Yvette, 91192, France, e-mail: surname.name@lss.supelec.fr

DOI: 10.12059/Fully3D.2017-11-3201017

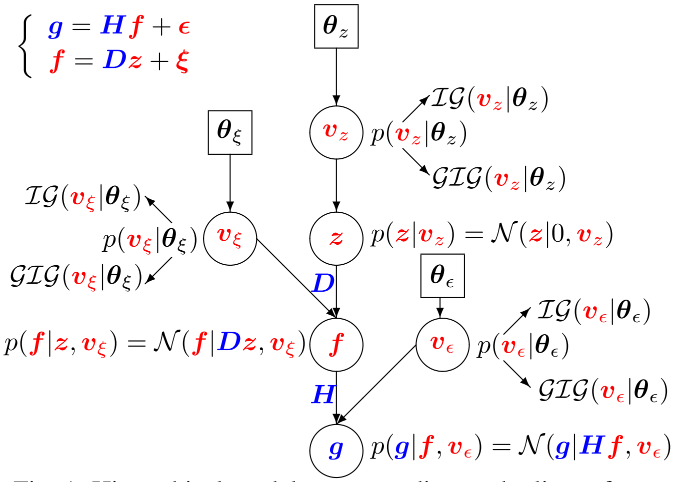
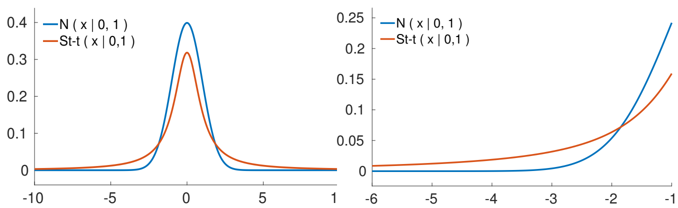


Fig. 1: Hierarchical model corresponding to the linear forward model, Eq. (1) and Eq. (2)

B. Normal variance mixtures

a) Student-t: The $\mathcal{S}t$ distribution is a sparsity enforcing distribution because of its heavy-tailed form. It can be expressed as a Normal variance mixture, with the mixing distribution an $\mathcal{IG}$ distribution. The comparison between the standard $\mathcal{N}$ distribution $\mathcal{N}(x|0,1)$ and the $\mathcal{S}t$ distribution $\mathcal{S}t-t(x|0,1)$ with one degree of freedom is presented in Figure (2). The Student-t Prior Model (StPM), Equation (4),



(a) $\mathcal{N}$ vs. $\mathcal{S}t$ (b) $\mathcal{S}t$ heavy tailed property

Fig. 2: Comparison between $\mathcal{N}$ and $\mathcal{S}t$ distribution.

considers a zero-mean $\mathcal{N}$ distribution for $z_j|v_{z_j}$ and an $\mathcal{IG}$ distribution for the variance $v_{z_j}|\alpha_z,\beta_z$, with the corresponding shape and scale parameters, α_z and β_z:

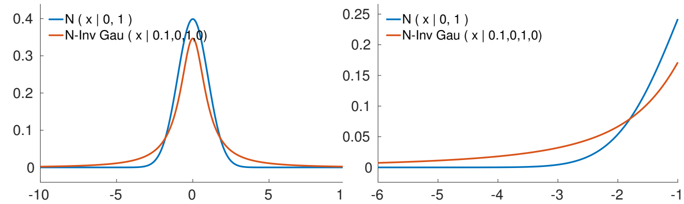
$$\textbf{StPM:}\ \begin{cases} p(z_j|0,v_{z_j}) = \mathcal{N}(z_j|0,\ v_{z_j}) \\ p(v_{z_j}|\alpha_z,\beta_z) = \mathcal{IG}(v_{z_j}|\alpha_z,\beta_z) \end{cases} \tag{4}$$

The marginal of the joint probability distribution $p(z_j|\alpha_z,\beta_z)$ is a two parameters Student-t distribution with the probability density:

$$p(z_j|\alpha_z,\beta_z) = \frac{\Gamma(\alpha_z+\frac{1}{2})}{\sqrt{2\pi\beta_z}\Gamma(\alpha_z)}\left(1+\frac{{z_j}^2}{2\beta_z}\right)^{-(\alpha_z+\frac{1}{2})}. \tag{5}$$

Observation: The standard $\mathcal{S}t$ probability density function is obtained by imposing the equality between the shape and scale parameters $\alpha_z=\beta_z=\nu_z/2$ for the $\mathcal{IG}$ mixing distribution. The form presented in Equation (5) allows the corresponding variance to take any positive values, which is important in the sparsity context.

b) Normal-inverse Gaussian: The $\mathcal{NIG}$ distribution is expressed as a Normal variance mixture with the mixing distribution a $\mathcal{GIG}$ distribution with the real parameter fixed $p=-1/2$. The comparison between the standard $\mathcal{N}$ distribution $\mathcal{N}(x|0,1)$ and the $\mathcal{NIG}$ distribution $\mathcal{NIG}(x|0.1,1)$ is presented in Figure (3). The Normal-Inverse Gaussian Prior



(a) $\mathcal{N}$ vs. $\mathcal{NIG}$ (b) $\mathcal{NIG}$ heavy tailed property

Fig. 3: Comparison between $\mathcal{N}$ and $\mathcal{NIG}$ distribution.

Model (NIGPM), Equation (6), considers zero-mean $\mathcal{N}$ distributions for $z_j|v_{z_j}$, and generalized inverse Gaussian distributions for the variances $v_{z_j}|\gamma_z^2,\delta_z^2$, with the corresponding parameters γ_z^2, δ_z^2 and $p_z=-1/2$:

$$\textbf{NIGPM:}\ \begin{cases} p(z_j|0,v_{z_j}) = \mathcal{N}(z_j|0,\ v_{z_j}) \\ p(v_{z_j}|\gamma_z^2,\delta_z^2) = \mathcal{GIG}(v_{z_j}|\gamma_z^2,\delta_z^2,p_z=-\frac{1}{2}) \end{cases} \tag{6}$$

The marginal of the joint probability distribution $p(z_j,|\gamma_z,\delta_z)$ is a $\mathcal{NIG}$ distribution with zero location and asymmetry β parameters:

$$p(z_j|\gamma_z,\delta_z) = \frac{\gamma_z\delta_z\mathcal{K}_1\left(\gamma_z\sqrt{\delta_z^2+{z_j}^2}\right)}{\pi\sqrt{\delta_z^2+{z_j}^2}}\exp\{\gamma_z\delta_z\}, \tag{7}$$

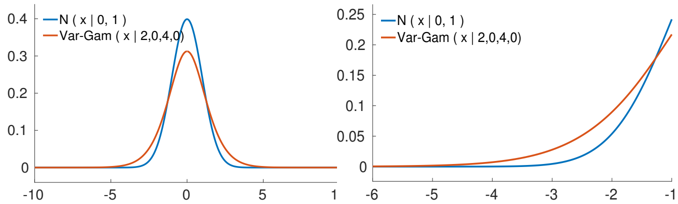
where $\mathcal{K}_p$ denotes the modified Bessel function of the second kind.

Observation: A great interest in the use of the Normal-inverse Gaussian in the sparsity context is the possibility to control the tail heaviness via the parameter γ_z.

c) Variance Gamma: Considering the Gamma distribution as the mixing distribution for a Normal variance mixture, leads to the Variance-Gamma distribution as prior. The comparison between the standard Normal distribution $\mathcal{N}(x|0,1)$ and the Variance-Gamma distribution $\mathcal{VG}(x|0,1)$ is presented in Figure (4). The Variance-Gamma Prior Model (VGPM), Equation (8), considers zero-mean Normal distributions for $z_j|v_{z_j}$, and Gamma distributions with the corresponding shape and scale parameters α_z and β_z for the variances $v_{z_j}|\alpha_z,\beta_z$:

$$\textbf{VGPM:}\ \begin{cases} p(z_j|0,v_{z_j}) = \mathcal{N}(z_j|0,\ v_{z_j}) \\ p(v_{z_j}|\alpha_z,\beta_z) = \mathcal{G}(v_{z_j}|\alpha_z,\beta_z) \end{cases} \tag{8}$$

The marginal of the joint probability distribution $p(z_j|\alpha_z,\beta_z)$ is a $\mathcal{VG}$ distribution with the zero location and asymmetry



(a) $\mathcal{N}$ vs. $\mathcal{VG}$ (b) $\mathcal{VG}$ heavy tailed property

Fig. 4: Comparison between the $\mathcal{N}$ and $\mathcal{VG}$ distribution.

parameters:

$$\mathcal{VG}(x|\alpha,\beta) = \frac{\beta^{2\alpha}|x|^{\alpha-\frac{1}{2}}\mathcal{K}_{\alpha-\frac{1}{2}}(\alpha|x|)}{\sqrt{\pi}\Gamma(\alpha)(2\beta)^{\alpha-\frac{1}{2}}}. \quad (9)$$

Observation: Evidently, $\mathcal{GIG}(x|2\beta,\delta\searrow 0,\alpha) = \mathcal{G}(x|\alpha,\beta)$ so the $\mathcal{VG}$ prior can also be viewed as a Normal variance mixture with $\mathcal{GIG}$ mixing distribution. Also, we mention that the a particular case of the $\mathcal{VG}$ obtained via $\mathcal{G}$ mixing distribution is $\mathcal{L}$ distribution, obtained via $\mathcal{E}$ mixing distribution.

III. Iterative algorithms

In order the favour a sparse solution for $\boldsymbol{z}$, Equation (2), we consider the following sparsity enforcing priors, expressed as Normal variance mixtures:

1) $\mathcal{S}t$ via StPM, Equation (4), $\mathcal{IG}$ mixing distribution
2) $\mathcal{NIG}$ via NIGPM, Equation (6), $\mathcal{GIG}$ mixing distribution (with fixed real parameter, $p_z = 1/2$)
3) $\mathcal{VG}$ via VGPM, Equation (8), $\mathcal{G}$ mixing distribution

For JMAP estimation, the posterior distribution, Equation (3), is used. Because of the use of the sparsity enforcing priors expressed as Normal variance mixtures, for all prior models considered the first part of the posterior distribution is the same (i.e. $p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{v}_\epsilon)$, $p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{v}_\xi)$, $p(\boldsymbol{z}|\boldsymbol{v}_z)$). The differences between the resulting algorithms corresponds to the choice of the mixing distribution: $\mathcal{IG}$, $\mathcal{GIG}$ or $\mathcal{G}$. Consequently, the analytical expressions corresponding to $\boldsymbol{f}$ and $\boldsymbol{z}$ are the same. Numerically, the differences are induced during the iterations because of different expressions corresponding to variance estimates. Furthermore, different behaviours are associated with each prior model in terms of sensibility to the hyperparameters or rate of convergence. For JMAP, the unknowns are estimated by maximizing the posterior distribution, or minimizing the criterion $\mathcal{L}$:

$$\left(\widehat{\boldsymbol{f}},\,\widehat{\boldsymbol{z}},\,\widehat{\boldsymbol{v}}_\xi,\,\widehat{\boldsymbol{v}}_\epsilon,\,\widehat{\boldsymbol{v}}_z\right) = \underset{(\boldsymbol{f},\,\boldsymbol{z},\,\boldsymbol{v}_\xi,\,\boldsymbol{v}_\epsilon,\,\boldsymbol{v}_z)}{\arg\min}\ \mathcal{L}(\boldsymbol{f},\,\boldsymbol{z},\,\boldsymbol{v}_\xi,\,\boldsymbol{v}_\epsilon,\,\boldsymbol{v}_z), \quad (10)$$

where the criterion $\mathcal{L}$ is defined as:

$$\mathcal{L}(\boldsymbol{f},\,\boldsymbol{z},\,\boldsymbol{v}_\xi,\,\boldsymbol{v}_\epsilon,\,\boldsymbol{v}_z) = -\ln p(\boldsymbol{f},\,\boldsymbol{z},\,\boldsymbol{v}_\xi,\,\boldsymbol{v}_\epsilon,\,\boldsymbol{v}_z|\boldsymbol{g}) \quad (11)$$

The optimisation algorithm considered is an alternate optimization of the criterion $\mathcal{L}(\boldsymbol{f},\,\boldsymbol{z},\,\boldsymbol{v}_\xi,\,\boldsymbol{v}_\epsilon,\,\boldsymbol{v}_z)$ with respect to the each unknown, leading to an interative algorithm,

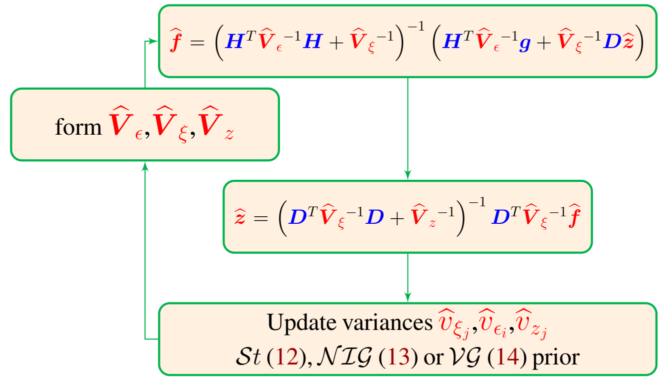


Fig. 5: Joint MAP iterative algorithm

Figure (5). Depending on the mixing distribution, the update of the variances is done as it follows:

1) For $\mathcal{IG}$ mixing distribution ($\mathcal{S}t$ prior)

$$\begin{cases} \widehat{v}_{\xi_j} = \dfrac{\beta_\xi + \frac{1}{2}\left(\widehat{f}_j - \boldsymbol{D}_j\widehat{\boldsymbol{z}}\right)^2}{\alpha_\xi + \frac{3}{2}} \\ \widehat{v}_{\epsilon_i} = \dfrac{\beta_\epsilon + \frac{1}{2}\left(g_i - \boldsymbol{H}_i\widehat{\boldsymbol{f}}\right)^2}{\alpha_\epsilon + \frac{3}{2}} \\ \widehat{v}_{z_j} = \dfrac{\beta_z + \frac{1}{2}{z_j}^2}{\alpha_z + \frac{3}{2}} \end{cases} \quad (12)$$

2) For $\mathcal{GIG}$ mixing distribution ($\mathcal{NIG}$ prior)

$$\begin{cases} \widehat{v}_{\xi_j} = \dfrac{2 + \sqrt{4 + \gamma_\xi^2\delta_\xi^2\left(\widehat{f}_j - \boldsymbol{D}_j\widehat{\boldsymbol{z}}\right)^2}}{\gamma_\xi^2} \\ \widehat{v}_{\epsilon_i} = \dfrac{2 + \sqrt{4 + \gamma_\epsilon^2\delta_\epsilon^2\left(g_i - \boldsymbol{H}_i\widehat{\boldsymbol{f}}\right)^2}}{\gamma_\epsilon^2} \\ \widehat{v}_{z_j} = \dfrac{2 + \sqrt{4 + \gamma_z^2\delta_z^2{z_j}^2}}{\gamma_z^2} \end{cases} \quad (13)$$

3) For $\mathcal{G}$ mixing distribution ($\mathcal{VG}$ prior)

$$\begin{cases} \widehat{v}_{\xi_j} = \dfrac{\left(\alpha_\epsilon - \frac{3}{2}\right) + \sqrt{\left(\alpha_\epsilon - \frac{3}{2}\right)^2 + 2\beta_\epsilon\left(\widehat{f}_j - \boldsymbol{D}_j\widehat{\boldsymbol{z}}\right)^2}}{\beta_\epsilon} \\ \widehat{v}_{\epsilon_i} = \dfrac{\left(\alpha_\xi - \frac{3}{2}\right) + \sqrt{\left(\alpha_\xi - \frac{3}{2}\right)^2 + 2\beta_\xi\left(g_i - \boldsymbol{H}_i\widehat{\boldsymbol{f}}\right)^2}}{\beta_\xi} \\ \widehat{v}_{z_j} = \dfrac{\left(\alpha_z - \frac{3}{2}\right) + \sqrt{\left(\alpha_z - \frac{3}{2}\right)^2 + 2\beta_z{z_j}^2}}{\beta_z} \end{cases} \quad (14)$$

$\widehat{f}_j$ denotes the element i of vector $\widehat{\boldsymbol{f}}$, g_i denotes the element j of vector $\boldsymbol{g}$, $\boldsymbol{H}_i$ denotes the line i of matrix $\boldsymbol{H}$, $\boldsymbol{D}_j$ denotes the line j of matrix $\boldsymbol{D}$. For all three prior models, the variance matrices are given by:

$$\begin{array}{c} \widehat{\boldsymbol{v}}_\xi = \left[\ldots\widehat{v}_{\xi_j}\ldots\right]\ ;\ \widehat{\boldsymbol{v}}_\epsilon = \left[\ldots\widehat{v}_{\epsilon_i}\ldots\right]\ ;\ \widehat{\boldsymbol{v}}_z = \left[\ldots\widehat{v}_{z_j}\ldots\right] \\ \widehat{\boldsymbol{V}}_\xi = \text{diag}\left[\widehat{\boldsymbol{v}}_\xi\right]\ ;\ \widehat{\boldsymbol{V}}_\epsilon = \text{diag}\left[\widehat{\boldsymbol{v}}_\epsilon\right]\ ;\ \widehat{\boldsymbol{V}}_z = \text{diag}\left[\widehat{\boldsymbol{v}}_z\right] \end{array} \quad (15)$$

IV. Simulation results

For simulations the reconstruction of X-ray CT piecewise constant images problem is considered. The Shepp Logan phantom (256^3) is used as the original image. 128 projections are simulated uniformly between $0°$ and $180°$. For the sparse representation of the image the multilevel Haar transform is used as in [4]. Figure (6) presents a typical example result with the slice comparison between the original volume and the reconstructed volumes corresponding to the three prior models. Figure (7) presents the comparison of the corresponding normalized mean squared error (NMSE) during iterations for the three prior models: The JMAP iterative algorithms corresponding to the three prior models have similar performances in terms of reconstruction accuracy. However, different rates of convergence correspond to the three prior models. Figure (8) shows how the NMSE depends on the prior hyperparameters for the three models. Very interestingly we see that the two last prior models are less sensitive to their corresponding hyperparameters.

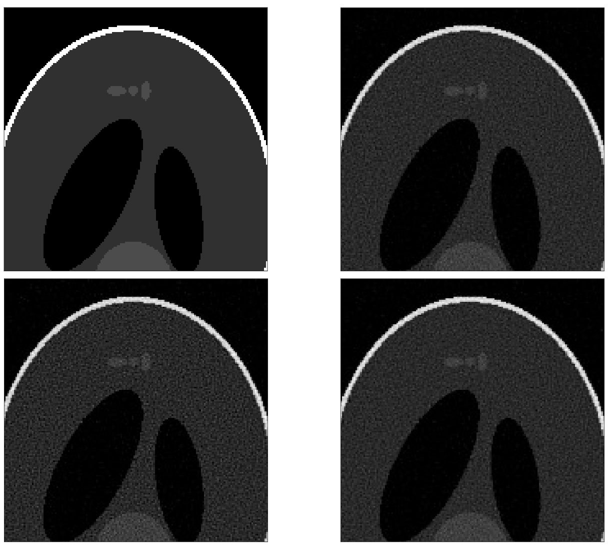


Fig. 6: Slice comparison between the original volume (top left) and JMAP reconstructed volumes $\mathcal{S}t$ (top right), $\mathcal{NIG}$ (bottom left), $\mathcal{VG}$ (bottom right). 64^3 phantom size, 128 projections, SNR=30dB

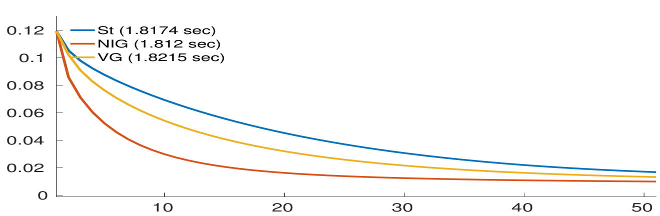


Fig. 7: NMSE during iterations: $\mathcal{S}t$, $\mathcal{NIG}$ and $\mathcal{VG}$

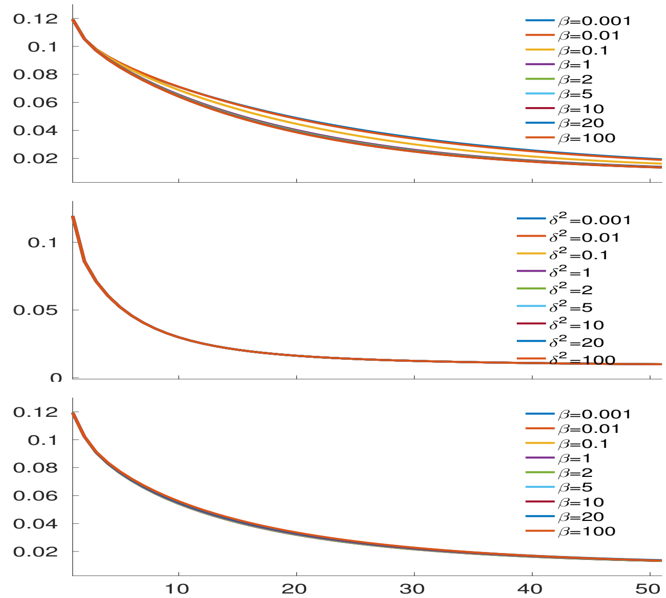


Fig. 8: Influence of the prior hyperparameters: NMSE vs. iterations for $\mathcal{S}t$ (top) $\mathcal{NIG}$ (middle) $\mathcal{VG}$ (bottom) priors

V. Conclusion

Sparsity enforcing iterative algorithms via a Bayesian approach were considered in this paper. Heavy tailed distributions expressed as Normal variance mixtures were considered in order to obtain analytical expressions for the unknowns of the model. The prior models and their corresponding JMAP iterative algorithms were developed and compared: the reconstruction accuracy is similar for the three prior models considered, but the rate of convergence is different and the sensibility to the prior hyperparameters is different. This results are encouraging. However we are now investigating methods to compute the Posterior mean via Variational Bayesian Approximation (VBA). The structure of the algorithms is the same but the computations need the diagonal elements of the covariance matrices which are too costly for 3D applications.

References

[1] N. Dobigeon, A. O. Hero, and J-Y Tourneret. Hierarchical bayesian sparse image reconstruction with application to mrfm. *IEEE Trans. on Image Process*, 18:2059 – 2070, 2009.

[2] M. Dumitru, A. Mohammad-Djafari, and Baghai S. S. Precise periodic components estimation for chronobiological signals through bayesian inference with sparsity enforcing prior. *EURASIP Journal on Bioinformatics and Systems Biology, Springer*, 3, 2016.

[3] P. C. Hansen and D. P. O'Leary. The use of the l-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.*, 14:1487–1503, 1993.

[4] Li Wang, Ali Mohammad-Djafari, Nicolas Gac, and Mircea Dumitru. Computed tomography reconstruction based on a hierarchical model and variational Bayesian method. In *2016 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 883–887. IEEE, 2016.

[5] David Wipf and Srikantan Nagarajan. Iterative reweighted l_1 and l_2 methods for finding sparse solutions. *IEEE J. Sel. Topics Signal Process*, 4:317–329, 2010.