

Comparaison des performances d’algorithmes itératifs bayésiens basés sur trois classes de modèles *a priori* parcimonieux appliqués à la reconstruction tomographique

Mircea DUMITRU, Li WANG, Nicolas GAC, Ali MOHAMMAD-DJAFARI

Laboratoire des signaux et systèmes
CentraleSupélec, Gif-sur-Yvette, 91192, France, France

Mircea.Dumitru@lss.supelec.fr, Li.Wang@lss.supelec.fr
nicolas.gac@lss.supelec.fr, djafari@lss.supelec.fr

Résumé – La reconstruction d’images constantes par morceaux ou homogènes dans le contexte de la tomographie à rayons X est considérée dans le cadre d’une approche bayésienne. Plus précisément, la transformation parcimonieuse de ces images est modélisée par des distributions à queue lourde exprimées sous forme hiérarchique de mélanges de variance normale. Les algorithmes itératifs dérivés (via Joint Maximum A Posteriori) ont des expressions de mise à jour identiques, à l’exception des variances estimées. Nous montrons que le comportement de chaque algorithme est différent en termes de sensibilité à la sélection du modèle et de qualité des reconstructions en tomographie X.

Abstract – The piecewise constant or homogeneous image reconstruction in the context of X-ray Computed Tomography is considered within a Bayesian approach. More precisely, the sparse transformation of such images is modelled with heavy tailed distributions expressed as Normal variance mixtures marginals. The derived iterative algorithms (via Joint Maximum A Posteriori) have identical updating expressions, except for the estimated variances. We show that the behaviour of the each algorithm is different in terms of sensibility to the model selection and reconstruction performances when applied in Computed Tomography.

1 Introduction

Nous considérons le modèle suivant :

$$\begin{cases} \mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon} \\ \mathbf{f} = \mathbf{D}\mathbf{z} + \boldsymbol{\xi} \end{cases}, \quad (1)$$

où \mathbf{g} représente les données observées, \mathbf{H} représente l’opérateur de projection linéaire, \mathbf{f} représente l’image à reconstruire et $\boldsymbol{\epsilon}$ compte pour le bruit et les incertitudes du modèle [12]. L’image constante par morceaux \mathbf{f} est exprimée par une transformation \mathbf{D} (telle que la transformée en ondelettes de Haar [9], [2]) permettant de mettre en évidence les contours et par la suite de favoriser par des lois spécifiques la parcimonie de ces hautes fréquences spatiales. La distribution *a priori* de la structure parcimonieuse \mathbf{z} est modélisée par des distributions à queue lourde exprimées comme la marginale d’une loi *Normal variance mixture* [5], [1]. Des lois *a priori* conjuguées sont considérées comme les lois de mélange (\mathcal{M}). Trois lois sont considérées : i) la loi Inverse-Gamma (\mathcal{IG}), donnant lieu à la loi de Student-t (*St*) *a priori* [6], [11] ii) la loi Inverse-Gaussienne Généralisée (\mathcal{GIG}) donnant lieu à la loi normale inverse-gaussienne (\mathcal{NIG}) *a priori* [8], [7] et iii) la loi Gamma (\mathcal{G}) donnant lieu à la loi de Variance-Gamma (\mathcal{VG}), *a priori* [10].

Pour $\boldsymbol{\epsilon}$ on considère une loi gaussienne séparable mais stationnaire dont les variances seront à estimer. Fig. (1) montre la structure du modèle hiérarchique et la vraisemblance et les lois *a priori* proposées. Dans ce contexte, quelque soit la loi *a*

priori, le modèle hiérarchique correspondant a les mêmes lois normales multidimensionnelles ($\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon$), ($\mathbf{f}|\mathbf{z}, \mathbf{v}_\xi$) et ($\mathbf{z}|\mathbf{v}_z$). Par conséquent, la distribution *a posteriori*

$$p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\xi, \mathbf{v}_z|\mathbf{g}) \propto p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon)p(\mathbf{f}|\mathbf{z}, \mathbf{v}_\xi)p(\mathbf{z}|\mathbf{v}_z)p(\mathbf{v}_\epsilon|\boldsymbol{\theta}_\epsilon)p(\mathbf{v}_\xi|\boldsymbol{\theta}_\xi)p(\mathbf{v}_z|\boldsymbol{\theta}_z). \quad (2)$$

aura une partie commune (formée par le produit des trois lois normales) indépendamment de la loi *a priori* choisie pour imposer la parcimonie. Ceci conduit à des expressions analytiques identiques pour les estimations de \mathbf{f} (respectivement \mathbf{z}) lorsqu’on considère l’estimation conjointe Maximum A Posteriori (JMAP). Selon la loi du mélange, des estimations différentes correspondent à \mathbf{v}_ϵ , \mathbf{v}_ξ et \mathbf{v}_z .

Nous présentons les similitudes et les différences théoriques entre les algorithmes JMAP correspondant à différentes distributions \mathcal{M} . Ensuite, une étude pratique des performances de l’estimation JMAP dans le contexte de la reconstruction par tomographie par rayons X est réalisée pour différents nombres de projections et différents niveaux de bruit. Nous étudions le comportement des algorithmes en termes de stabilité (par rapport aux valeurs des hyperparamètres) et de taux de convergence.

Le reste de l’article est organisé comme suit. La Sec. (2) présente un bref exposé des lois \mathcal{M} considérées ainsi que des lois *a priori* correspondantes (à queue lourde). La Sec. (3) présente et compare les algorithmes itératifs JMAP correspondant aux

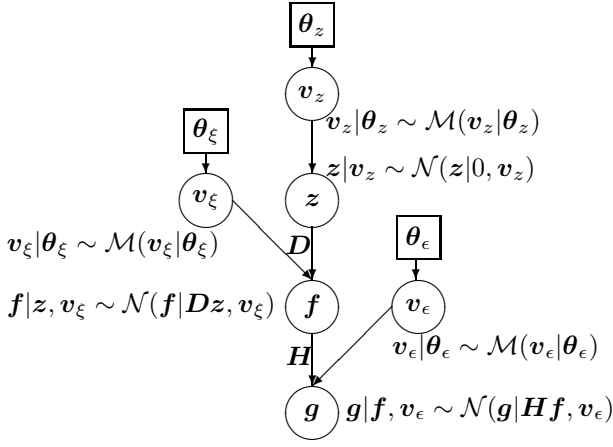


FIGURE 1 – Modèle hiérarchique pour le modèle direct

distributions \mathcal{M} considérées. Les résultats des simulations et les évaluations empiriques sont présentés dans la Sec. (4). Les conclusions sont tirées dans la Sec. (5).

2 Lois de mélange et modèles *a priori*

La structure parcimonieuse de z est prise en compte par un modèle *a priori* utilisant des distributions à queue lourde pour chaque pixel z_j . Nous nous concentrons sur de telles distributions qui peuvent être exprimées par :

$$\text{Modèle } a \text{ priori} : \begin{cases} p(z_j | 0, v_{z_j}) = \mathcal{N}(z_j | 0, v_{z_j}) \\ p(v_{z_j} | \theta_z) = \mathcal{M}(v_{z_j} | \theta_z) \end{cases} \quad (3)$$

Selon le choix de la distribution \mathcal{M} , on obtient une loi *a priori* différente pour $p(z_j | 0, \theta_z)$. Par la suite nous considérons trois classes de telles lois.

2.1 Distribution de mélange \mathcal{IG}

La loi Inverse-Gamma est considérée comme la loi de mélange dans l'éq. (3), $\mathcal{M}(v_{z_j} | \theta_z) = \mathcal{IG}(v_{z_j} | \alpha_z, \beta_z)$. La loi marginale correspondant de la loi de probabilité conjointe $(v_{z_j} | \alpha_z, \beta_z)$ est une loi St avec deux paramètres :

$$p(z_j | \alpha_z, \beta_z) = \frac{\Gamma(\alpha_z + \frac{1}{2})}{\sqrt{2\pi\beta_z}\Gamma(\alpha_z)} \left(1 + \frac{z_j^2}{2\beta_z}\right)^{-(\alpha_z + \frac{1}{2})} \quad (4)$$

Remarque : Imposer $\alpha_z = \beta_z = \nu_z/2$ conduit à la forme standard St avec un paramètre [3].

2.2 Distribution de mélange \mathcal{GIG}

Dans l'éq. (3), la loi inverse gaussienne généralisée est considérée, $\mathcal{M}(v_{z_j} | \theta_z) = \mathcal{GIG}(v_{z_j} | \gamma_z^2, \delta_z^2, -1/2)$. La loi marginale correspondant de la loi conjointe $(v_{z_j} | \gamma_z^2, \delta_z^2)$ est une loi \mathcal{NIG} avec les paramètres de localisation et d'asymétrie égaux à zéro :

$$p(z_j | \gamma_z, \delta_z) = \frac{\gamma_z \delta_z \mathcal{K}_1(\gamma_z \sqrt{\delta_z^2 + z_j^2})}{\pi \sqrt{\delta_z^2 + z_j^2}} \exp\{\gamma_z \delta_z\}, \quad (5)$$

où \mathcal{K}_1 désigne la fonction de Bessel modifiée du second type.

2.3 Distribution de mélange \mathcal{G}

Dans l'éq. (3) la loi Gamma est considérée comme la distribution de mélange, $\mathcal{M}(v_{z_j} | \theta_z) = \mathcal{G}(v_{z_j} | K_z, \theta_z)$. La loi marginale correspondant est une loi \mathcal{VG} avec les paramètres de localisation et d'asymétrie égaux à zéro :

$$p(z_j | k_z, \theta_z) = \frac{\theta_z^{2k_z} |z_j|^{k_z - \frac{1}{2}} \mathcal{K}_{k_z - \frac{1}{2}}(k_z |z_j|)}{\sqrt{\pi} \Gamma(k_z) (2\theta_z)^{k_z - \frac{1}{2}}} \quad (6)$$

Remarque : Un cas particulier de la loi *a priori* \mathcal{VG} est la loi de Laplace, correspondant à la loi de mélange exponentielle [4].

Remarque : $\mathcal{GIG}(x | 2\theta, \delta \searrow 0, k) = \mathcal{G}(x | k, \theta)$; la loi *a priori* \mathcal{VG} peut également être considérée comme un mélange de la loi normale avec la loi \mathcal{GIG} [13].

3 Estimation du Maximum *a posteriori*

JMAP estime les inconnues du modèle hiérarchique en maximisant la distribution *a posteriori*, Eq. (2) ou en minimisant le critère \mathcal{L} , défini par

$$\mathcal{L}(\mathbf{f}, \mathbf{z}, \mathbf{v}_\xi, \mathbf{v}_\epsilon, \mathbf{v}_z) = -\ln p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\xi, \mathbf{v}_\epsilon, \mathbf{v}_z | \mathbf{g}) \quad (7)$$

Pour les lois *a priori* mélangées de variance normale, la structure du modèle hiérarchique est la même pour $p(\mathbf{g} | \mathbf{f}, \mathbf{v}_\epsilon)$, $p(\mathbf{f} | \mathbf{z}, \mathbf{v}_\xi)$ et $p(\mathbf{z} | \mathbf{v}_z)$ alors la première partie de la distribution *a posteriori* (et le critère \mathcal{L}) est la même, quel que soit le choix de \mathcal{M} . Par conséquent, les estimations $\hat{\mathbf{f}}$ et $\hat{\mathbf{z}}$ ont les mêmes expressions analytiques. La minimisation du critère \mathcal{L} par optimisation alternée par rapport à chaque inconnue donne l'expression analytique :

$$\begin{aligned} \hat{\mathbf{f}} &= \left(\mathbf{H}^T \hat{\mathbf{V}}_\epsilon^{-1} \mathbf{H} + \hat{\mathbf{V}}_\xi^{-1} \right)^{-1} \left(\mathbf{H}^T \hat{\mathbf{V}}_\epsilon^{-1} \mathbf{g} + \hat{\mathbf{V}}_\xi^{-1} \mathbf{D} \hat{\mathbf{z}} \right) \\ \hat{\mathbf{z}} &= \left(\mathbf{D}^T \hat{\mathbf{V}}_\xi^{-1} \mathbf{D} + \hat{\mathbf{V}}_z^{-1} \right)^{-1} \mathbf{D}^T \hat{\mathbf{V}}_\xi^{-1} \hat{\mathbf{f}}, \end{aligned} \quad (8)$$

où

$$\begin{aligned} \hat{\mathbf{v}}_\xi &= [\dots \hat{v}_{\xi_j} \dots]; \hat{\mathbf{v}}_\epsilon = [\dots \hat{v}_{\epsilon_i} \dots]; \hat{\mathbf{v}}_z = [\dots \hat{v}_{z_j} \dots] \\ \hat{\mathbf{V}}_\xi &= \text{diag}[\hat{v}_{\xi_j}]; \hat{\mathbf{V}}_\epsilon = \text{diag}[\hat{v}_{\epsilon_i}]; \hat{\mathbf{V}}_z = \text{diag}[\hat{v}_{z_j}] \end{aligned} \quad (9)$$

En introduisant les notations

$$d(\xi_j) = \hat{f}_j - \mathbf{D}_j \mathbf{z}; d(\epsilon_i) = g_i - \mathbf{H}_i \mathbf{f}; d(z_j) = z_j, \quad (10)$$

où f_j désigne l'élément j^{eme} de \mathbf{f} , \mathbf{D}_j indique la ligne j^{eme} de \mathbf{D} , g_i indique l'élément i^{eme} de \mathbf{g} et \mathbf{H}_i représente la ligne i^{eme} de \mathbf{H} , les expressions analytiques des estimations de variance selon la distribution du mélange sont :

$$\text{St} : \hat{v}_{\zeta_k} = \frac{\beta_\zeta + \frac{1}{2} d^2(\zeta_k)}{\alpha_\zeta + \frac{3}{2}} \quad (11)$$

$$\text{NIG} : \hat{v}_{\zeta_k} = \frac{\sqrt{4 + \gamma_\zeta^2 (\delta_\zeta^2 + d^2(\zeta_k))} - 2}{\gamma_\zeta^2} \quad (12)$$

$$\text{VG} : \hat{v}_{\zeta_k} = \frac{\sqrt{(\frac{3}{2} - k_\zeta)^2 + 2 \frac{1}{\theta_\zeta} d^2(\zeta_k)} - (\frac{3}{2} - k_\zeta)}{2 \frac{1}{\theta_\zeta}} \quad (13)$$

où $\zeta = \{\xi, \epsilon, z\}$. Les algorithmes itératifs JMAP sont présentés dans la Fig (2).

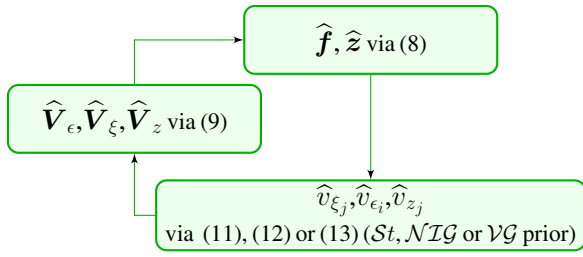


FIGURE 2 – Algorithme itératif Joint MAP

4 Résultats de simulations

2D RSB Prior	phantom size : 256*256								
	40dB			30dB			20dB		
	<i>St</i>	<i>NIG</i>	<i>VG</i>	<i>St</i>	<i>NIG</i>	<i>VG</i>	<i>St</i>	<i>NIG</i>	<i>VG</i>
128proj	0.0124	0.0118	0.0127	0.0271	0.0311	0.0279	0.0833	0.1923	0.1561
64proj	0.0247	0.0227	0.0231	0.0373	0.0375	0.0343	0.1099	0.1724	0.1368
32proj	0.0396	0.0351	0.0358	0.0472	0.0454	0.0469	0.1163	0.1415	0.1186

TABLE 1 – cas 2D : EMQN du fantôme reconstruit pour 50 itérations correspondant aux trois lois *a priori*.

3D RSB Prior	phantom size : 256*256*256								
	40dB			30dB			20dB		
	<i>St</i>	<i>NIG</i>	<i>VG</i>	<i>St</i>	<i>NIG</i>	<i>VG</i>	<i>St</i>	<i>NIG</i>	<i>VG</i>
128proj	0.0744	0.0112	0.0326	0.0330	0.0309	0.0580	0.1031	0.1618	0.1451
64proj	0.0783	0.0175	0.0379	0.0552	0.0352	0.0623	0.1086	0.1517	0.1424
32proj	0.1283	0.0275	0.0357	0.1059	0.0406	0.0575	0.1219	0.1274	0.1258

TABLE 2 – 3D case : EMQN du fantôme reconstruit pour 50 itérations correspondant aux trois lois *a priori*.

Le problème de la reconstruction des images constantes par morceaux par rayons X est abordé dans cette section. Les résultats sont présentés pour les simulations 2D et 3D. Le fantôme Shepp Logan (256^2 respectivement 256^3) est utilisé comme image d'origine. Les projections sont simulées uniformément entre 0° et 180° . La représentation parcimonieuse de l'image est obtenue via la transformée Haar multiniveaux. Pour les simulations 3D, le GPU est utilisé. L'erreur quadratique moyenne normalisée (EMQN) est utilisée pour estimer les écarts de reconstruction. Les résultats correspondent à 50 itérations. Dans le contexte de la tomographie par rayons X, la matrice H n'est pas considérée explicitement mais par la projection Hf et la retroprojection $H^t g$. Les estimations correspondant à f et z , Eq. (8) sont calculées à l'aide de l'algorithme de descente du gradient et le pas de descente est obtenue à l'aide d'une stratégie d'optimisation de longueur du pas.

Nous rapportons des simulations pour des projections de 128, 64, 32 et des niveaux de bruit correspondant aux RSB = 40, 30, 20, 10 dB pour lesquels la EMQN est calculée. Pour le cas 2D, la sensibilité des algorithmes aux hyperparamètres est étudiée pour tous les cas de niveaux de bruit et de nombre de projections en considérant pour chaque hyperparamètre 0.001, 0.01, 0.1, 1, 2.1, 10, 100.

La Fig (3) montre l'évolution du EMQN au cours des itérations pour les trois algorithmes JMAP en fonction des hyperparamètres, pour 40 dB (colonne gauche) et 30 dB (colonne

droite).

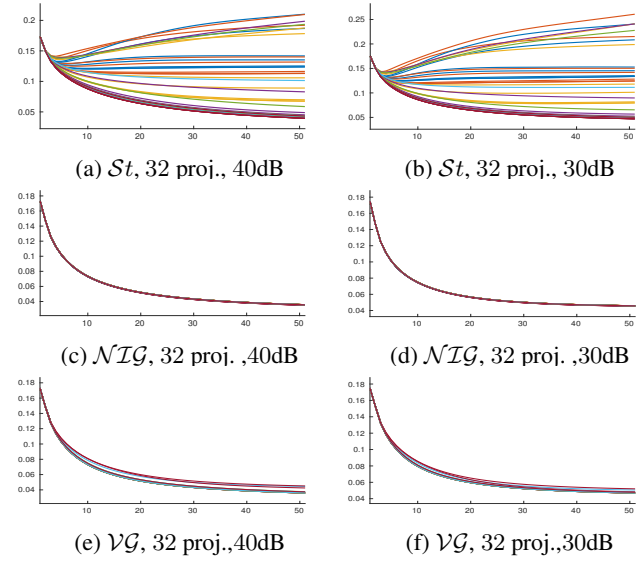


FIGURE 3 – Etude 2D : Performances de la reconstruction des algorithmes JMAP (EMQN) pendant les itérations en fonction des hyperparamètres.

La loi *a priori* *St* présente une forte sensibilité vis-à-vis des hyperparamètres, Fig. (3a) et (3b), avec des performances de reconstruction très différentes. Pour la loi *a priori* *VG*, l'influence des hyperparamètres est faible, Fig. (3c) et (3d). Pour la loi *a priori* *NIG*, les performances de reconstruction en fonction des hyperparamètres sont presque superposées, Fig. (3e) et (3f). Une comparaison entre les trois EMQN de reconstruction au cours des itérations, correspondant aux hyperparamètres optimaux expérimentaux, est présentée pour 40 dB (à gauche) et 30 dB (droite) pour 64 et 32 projections. Les performances

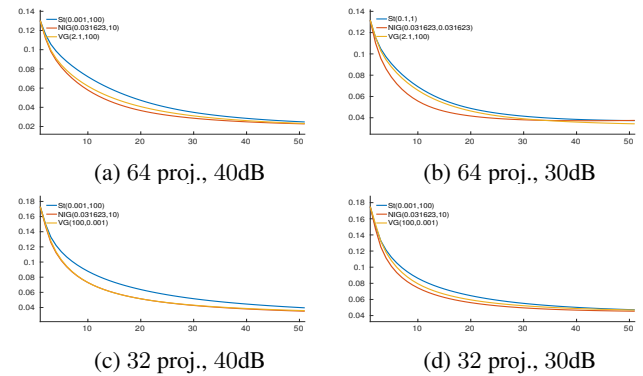


FIGURE 4 – Etude 2D : comportement des algorithmes JMAP pour les hyperparamètres optimaux, 40dB et 30dB.

de reconstruction en 2D pour tous les niveaux de bruit et le nombre de projections considérées, correspondant aux hyperparamètres optimaux, sont présentées dans le Tab. (1). Un comportement différent des lois *a priori* est signalé en fonction du niveau de bruit. Pour 40dB et 30dB les performances de reconstruction sont très similaires. En particulier, le *NIG* semble

donner des performances légèrement meilleures pour 40dB, pour tous les cas de nombre de projections considérés. Cependant, pour des niveaux de bruit importants, 20dB, les performances de reconstruction semblent être nettement meilleures pour le St . Le fantôme 2D Shepp Logan reconstitué correspondant aux trois priors est présenté dans Fig. (5) : 64 proj et 40dB dans Fig. (5a) et 32 proj et 20 dB dans Fig. (5b).

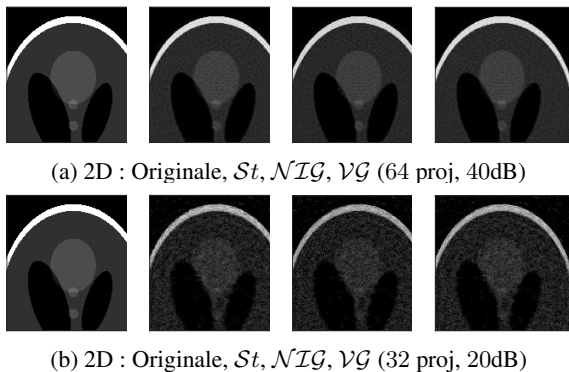


FIGURE 5 – Reconstructions 2D pour St , \mathcal{NIG} et \mathcal{VG} .

Les performances de reconstruction 3D pour tous les niveaux de bruit et le nombre de projections considérées, correspondant aux hyperparamètres optimaux, sont présentées dans le Tab. (2). Dans le cas 3D, le \mathcal{NIG} semble donner de meilleures performances de reconstruction pour 40 dB et 30 dB alors que le St prior semble être plus approprié pour 20 dB. La tranche centrale du fantôme de Shepp Logan 3D reconstitué correspondant aux trois prières est présentée dans la Fig. (6) : 64 proj et 40 dB dans Fig. (6a) et 32 proj et 20 dB dans la Fig. (6b).

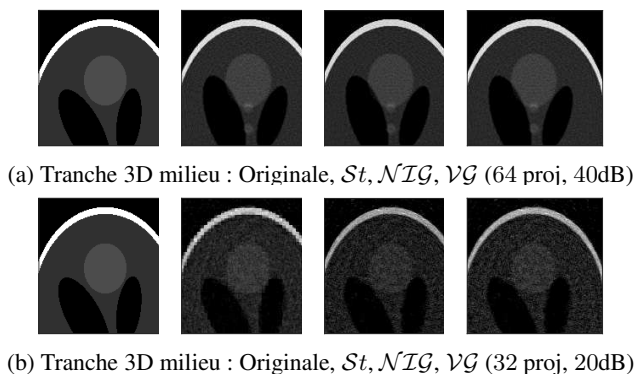


FIGURE 6 – Reconstructions 3D pour St , \mathcal{NIG} et \mathcal{VG} .

5 Conclusion

Dans cet article, nous avons présenté les algorithmes JMAP correspondant à trois lois *a priori* différentes dans le contexte de parcimonie. Nous avons considéré les lois à queue lourde exprimées comme mélanges de variance normale, montrant que les équations de mise à jour ont essentiellement la même expression analytique, à l'exception des variances apparaissant

dans le modèle hiérarchique. D'abord nous avons montré des comportements différents des lois *a priori* en termes de sensibilité aux hyperparamètres. Nous avons ensuite examiné les performances de reconstruction pour 2D et 3D, en signalant que pour différents niveaux de bruit, différents priors semblent donner de meilleures performances. Cependant, une étude plus détaillée, prenant en considération le taux de sparsité associé à chaque loi *a priori*, en fonction des hyperparamètres, doit être développée afin d'étudier les résultats de reconstruction correspondant à ces hyperparamètres. L'estimation de la moyenne *a posteriori* par l'approximation bayésienne variationnelle (ABV) doit être étudiée et comparée. La conclusion principale de cette étude est que pour des valeurs de RSB élevées, la loi \mathcal{NIG} *a priori* est très stable et donne des performances légèrement meilleures tandis que pour des valeurs RSB réduites le modèle St a de meilleures performances en termes de reconstruction.

Références

- [1] S. Derin Babacan, Shinichi Nakajima, and Minh N. Do. Bayesian Group-Sparse Modeling and Variational Inference. *IEEE Trans. Signal Process.*, 14 :2906–2921, 2014.
- [2] Ronald A DeVore, Björn Jawerth, and Bradley J Lucier. Image compression through wavelet transform coding. *IEEE Transactions on information theory*, 38(2) :719–746, 1992.
- [3] Mircea Dumitru. *A Bayesian approach for periodic components estimation for chronobiological signals*. PhD thesis, Université Paris-Saclay, Mar 2016.
- [4] Torbjorn Eltoft, Taesu Kim, and Te-Won Lee. On the multivariate Laplace distribution. *IEEE Signal Processing Letters*, 13 :300–303, 2006.
- [5] Ali Mohammad-Djafari. Bayesian approach with prior models which enforce sparsity in signal and image processing. *EURASIP Journal on Advances in Signal Processing*, 52, 2012.
- [6] Ali Mohammad-Djafari and Mircea Dumitru. Bayesian sparse solutions to linear inverse problems with non-stationary noise with Student-t priors. *Digital Signal Processing*, 47 :128–156, 2015.
- [7] Tor Arne Oigard, Alfred Hanssen, and Roy Edgar Hansen. The multivariate normal inverse Gaussian distribution : EM-estimation and analysis of synthetic aperture sonar data. In *2004 12th European Signal Processing Conference*, pages 1433–1436. IEEE, 2004.
- [8] Hamidreza Sadreazami, M. Omair Ahmad, and M. N. Shanmukha Swamy. Contourlet domain image denoising using Normal Inverse Gaussian distribution. In *2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE)*. IEEE, 2014.
- [9] Radomir S Stanković and Bogdan J Falkowski. The Haar wavelet transform : its status and achievements. *Computers & Electrical Engineering*, 29(1) :25–44, 2003.
- [10] Rasool Tahmasbi and Sadegh Rezaei. A Soft Voice Activity Detection Using GARCH Filter and Variance Gamma Distribution. *IEEE Transactions on Audio, Speech and Language Processing*, 15 :1129–1134, 2007.
- [11] Konstantinos E. Themelis, Athanasios A. Rontogiannis, and Konstantinos D. Koutroumbas. A Variational Bayes Framework for Sparse Adaptive Estimation. *IEEE Transactions on Signal Processing*, 62 :4723–4736, 2014.
- [12] 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.
- [13] Darren Wraith and Florence Forbes. Location and scale mixtures of Gaussians with flexible tail behaviour : Properties, inference and application to multivariate clustering. *Computational Statistics & Data Analysis*, 90 :61–73, 2015.