

PERFORMANCE COMPARISON OF BAYESIAN ITERATIVE ALGORITHMS FOR THREE CLASSES OF SPARSITY ENFORCING PRIORS WITH APPLICATION IN COMPUTED TOMOGRAPHY

Mircea Dumitru* Li Wang* Nicolas Gac* Ali Mohammad-Djafari*

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

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.

Index Terms— Normal variance mixtures, Bayesian inference, sparsity enforcing priors, image reconstruction.

1. INTRODUCTION

The linear forward model considered in this paper is given by

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

where \mathbf{g} represents the observed data, \mathbf{H} represents the linear projection operator, \mathbf{f} represents the image to be reconstructed, ϵ accounts for errors and model uncertainties [1]. The piecewise continuous image \mathbf{f} is expressed via a transformation \mathbf{D} (such as Haar wavelet transform [2], [3]) of a sparse structure \mathbf{z} accounting for the uncertainties ξ . The prior distribution of the sparse structure \mathbf{z} is modelled by using heavy tailed distributions expressed as marginals of Normal (\mathcal{N}) variance mixtures [4], [5]. Conjugate priors are considered as the mixing (\mathcal{M}) distributions. Three categories of distributions are considered: a) Inverse Gamma (\mathcal{IG}) distribution, (Student-t (\mathcal{St}) prior) [6], [7], [8], [9], b) a particular case of the generalized inverse Gaussian (\mathcal{GIG}) distribution (Normal inverse Gaussian (\mathcal{NIG}) prior) [10], [11], [12] and c) Gamma (\mathcal{G}) distribution (Variance-Gamma (\mathcal{VG}) prior) [13], [14].

Non-stationary independent Gaussian uncertainties model is assumed with conjugate priors modelling their corresponding variances. Figure (1) shows the structure of the hierarchical model and the likelihood and main prior laws proposed. In

this context, regardless of the prior, the corresponding hierarchical model has the same multivariate Normal distributions assigned for $(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon)$, $(\mathbf{f}|\mathbf{z}, \mathbf{v}_\xi)$ and $(\mathbf{z}|\mathbf{v}_z)$. Consequently, the posterior distribution

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

will have a common part (formed by the product of the three multivariate Normal distributions) regardless of the prior chosen to enforce sparsity. This leads to identical analytical expressions for the estimates of \mathbf{f} (respectively \mathbf{z}) when Joint Maximum A Posteriori (JMAP) estimation is considered. Depending on the mixing distribution, different estimates correspond to \mathbf{v}_ϵ , \mathbf{v}_ξ and \mathbf{v}_z . In this paper, first the theoretical sim-

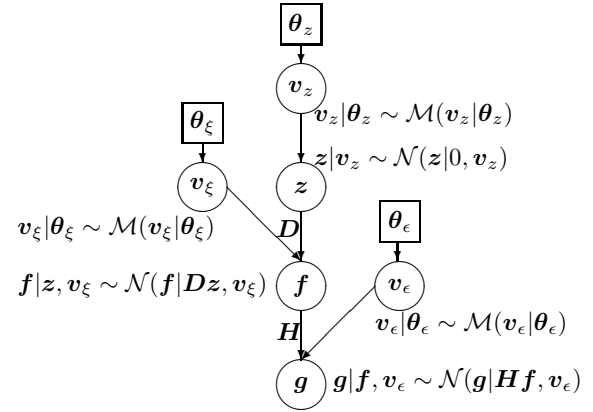


Fig. 1: Hierarchical model for forward model, Eq. (1).

ilarities and differences between the JMAP algorithms corresponding to different \mathcal{M} distributions are presented. Then, a practical study of the JMAP estimation performances in the context of X-ray Computed Tomography reconstruction for different number of projections and different levels of noise. Also, we investigate the behaviour of the algorithms in terms of stability (relative to prior hyperparameters values) and convergence rate.

The rest of the paper is organized as follows. Sec. (2) provides a brief presentations of the \mathcal{M} distributions considered and their corresponding (heavy tailed) priors. Sec. (3) presents

and compares the JMAP iterative algorithms corresponding the \mathcal{M} distributions considered. Simulations results and empirical evaluations are presented in Sec. (4). Conclusions are drawn in Sec. (5).

2. MIXING DISTRIBUTIONS AND PRIOR MODELS

The sparse structure of \mathbf{z} is accounted by a prior model using heavy-tailed distributions for each pixel z_j . We focus on such distributions that can be expressed as marginals of Normal variance mixtures. In this context, the general prior model is given by Eq. (3):

$$\text{Prior Model: } \begin{cases} p(z_j|0, v_{z_j}) = \mathcal{N}(z_j|0, v_{z_j}) \\ p(v_{z_j}|\boldsymbol{\theta}_z) = \mathcal{M}(v_{z_j}|\boldsymbol{\theta}_z) \end{cases} \quad (3)$$

Depending on the choice of the \mathcal{M} distribution, a different prior corresponds to $p(z_j|0, \boldsymbol{\theta}_z)$. In the following, a brief presentation of the \mathcal{M} distributions considered and their corresponding heavy-tailed prior distribution is given.

2.1. \mathcal{IG} mixing distribution – St prior

Inverse Gamma is considered as the mixing distribution in Eq. (3), $\mathcal{M}(v_{z_j}|\boldsymbol{\theta}_z) = \mathcal{IG}(v_{z_j}|\alpha_z, \beta_z)$. The corresponding hyperparameters are in this case $\boldsymbol{\theta}_z = (\alpha_z, \beta_z)$ and the corresponding marginal of the joint probability distribution $(v_{z_j}|\alpha_z, \beta_z)$ is a two-parameter St distribution:

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

Observation: Imposing $\alpha_z = \beta_z = v_z/2$ in the \mathcal{IG} mixing distribution leads to the standard St form with one parameter [15].

2.2. \mathcal{GIG} mixing distribution – \mathcal{NIG} prior

In Eq. (3) the generalized inverse Gaussian is considered as the mixing distribution, $\mathcal{M}(v_{z_j}|\boldsymbol{\theta}_z) = \mathcal{GIG}(v_{z_j}|\gamma_z^2, \delta_z^2, -1/2)$. The hyperparameters are $\boldsymbol{\theta}_z = (\gamma_z^2, \delta_z^2)$ and the corresponding marginal of the joint probability distribution $(v_{z_j}|\gamma_z^2, \delta_z^2)$ is a zero location and asymmetry parameter \mathcal{NIG} distribution:

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

where \mathcal{K}_1 denotes the modified Bessel function of the second kind.

2.3. \mathcal{G} mixing distribution – \mathcal{VG} prior

In Eq. (3) Gamma is considered as the mixing distribution, $\mathcal{M}(v_{z_j}|\boldsymbol{\theta}_z) = \mathcal{G}(v_{z_j}|k_z, \theta_z)$. In this case $\boldsymbol{\theta}_z = (k_z, \theta_z)$ and

the joint probability distribution marginal is a zero location and asymmetry parameter Variance-Gamma distribution:

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

Observation: A particular case of the \mathcal{VG} prior is the Laplace prior, corresponding to Exponential mixing distribution [16].

Observation: $\mathcal{GIG}(x|2\theta, \delta \searrow 0, k) = \mathcal{G}(x|k, \theta)$; \mathcal{VG} prior can also be viewed as a Normal variance mixture with \mathcal{GIG} mixing distribution [17].

3. JMAP ESTIMATION

JMAP estimates the unknowns of the hierarchical model by maximizing the posterior distribution, Eq. (2) or minimizing the criterion \mathcal{L} , defined by

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

For Normal variance mixture priors, the structure of the hierarchical model, Fig. (1) is the same for $p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon)$, $p(\mathbf{f}|\mathbf{z}, \mathbf{v}_\xi)$ and $p(\mathbf{z}|\mathbf{v}_z)$ so the first part of the posterior distribution (and the criterion \mathcal{L}) is the same, regardless of the choice of the \mathcal{M} . Consequently, the estimates $\hat{\mathbf{f}}$ and $\hat{\mathbf{z}}$ have the same analytical expressions. The minimization of the criterion \mathcal{L} by alternate optimization with respect to each unknown gives the analytical expression:

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

where

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

Introducing the notations

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

where f_j denotes the j^{th} element of \mathbf{f} , \mathbf{D}_j denotes the j^{th} line of \mathbf{D} , g_i denotes the i^{th} element of \mathbf{g} and \mathbf{H}_i denotes the i^{th} line of \mathbf{H} , the analytical expressions of variance estimates depending on mixing distribution are:

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

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

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

where $\zeta = \{\xi, \epsilon, z\}$. The JMAP iterative algorithms are presented in Fig (2).

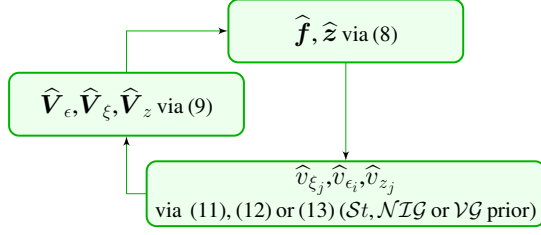


Fig. 2: Joint MAP iterative algorithm

4. SIMULATION RESULTS

X-ray CT piecewise constant images reconstruction problem is addressed in this section. Results are reported for 2D and 3D simulations. The Shepp Logan phantom (256^2 respectively 256^3) is used as the original image. The projections are simulated uniformly between 0° and 180° . The sparse representation of the image is obtained via the multilevel Haar transform. For 3D simulations, GPU is used. The normalised mean squared error (NMSE) is used for estimating the reconstruction deviations. The results correspond for 50 iteration. In the context of X-ray CT, the matrix \mathbf{H} is not considered explicitly but via the forward projection $\mathbf{H}\mathbf{f}$ and the back-projection $\mathbf{H}^t\mathbf{g}$ operators. The estimations corresponding to \mathbf{f} and \mathbf{z} Eq. (8) are computed using the gradient descent algorithm, and the descent step length is obtained using an optimized step length strategy [18].

We report simulations for 128, 64, 32 projections and noise levels corresponding to $\text{SNR}=\{40, 30, 20, 10\}\text{dB}$, for which the normalized mean reconstruction error (NMSE) $\frac{\|\mathbf{f}-\hat{\mathbf{f}}\|}{\|\mathbf{f}\|}$ is computed (Table (1) and (2)). For the 2D case, the sensibility of the algorithms to the hyperparameters is investigated for all cases of levels of noise and numbers of projections by considering for each hyperparameter $\{0.001, 0.01, 0.1, 1, 2.1, 10, 100\}$.

Figure (3) presents the evolution of the NMSE during the iterations for the three JMAP algorithms depending on the hyperparameters, for 40dB (left column) and 30dB (right column). The St prior presents a strong sensibility with respect to the hyperparameters, Figure (3a) and (3b), with very different reconstruction performances. For the VG prior, the influence of the hyperparameters is weak, Figure (3c) and (3d). For the NIG prior the reconstruction performances depending on the hyperparameters are almost superposed, Figure (3e) and (3f). A comparison between the three priors reconstruction NMSEs during iterations, corresponding to experimental optimal hyperparameters is presented in Figure (4) for 40dB (left) and 30dB (right) for 64 and 32 projections.

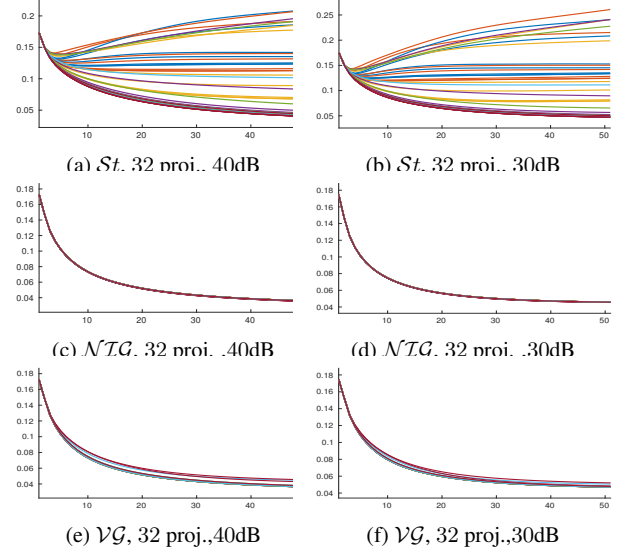


Fig. 3: 2D study: JMAP algorithms reconstruction error (NMSE) (y axis) during iterations (x axis) depending on the hyperparameters.

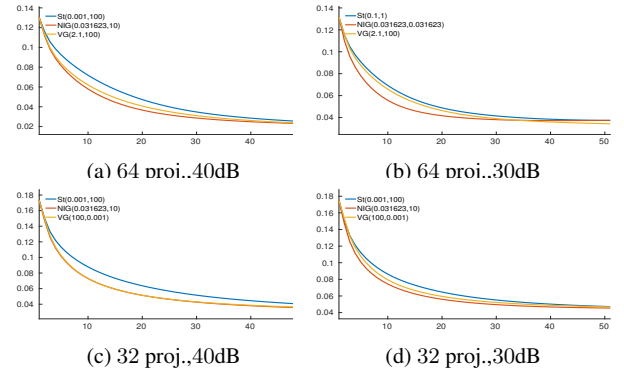


Fig. 4: 2D study: JMAP algorithms NMSE (y axis) for the optimal hyperparameters during iterations (x axis), 40dB and 30dB.

The 2D reconstruction performances for all levels of noise and number of projections considered, corresponding to the optimal hyperparameters are presented in Table (1). A different behaviour of the priors is reported depending on the level of noise. For 40dB and 30dB the reconstruction performances are very similar. In particular, the NIG seems to give slightly better performances for 40dB, for all cases of number of projections considered. However, for important levels of noise, 20dB, the reconstruction performances seems to be significantly better for the St prior. The reconstructed 2D Shepp Logan phantom corresponding to the three priors is presented in Figure (5): 64 proj and 40dB in Figure (5a) and 32 proj and 20dB in Figure (5b).

2D SNR Prior	phantom size: 256*256								
	40dB			30dB			20dB		
	St	\mathcal{NIG}	\mathcal{VG}	St	\mathcal{NIG}	\mathcal{VG}	St	\mathcal{NIG}	\mathcal{VG}
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: 2D case: NMSE of reconstructed phantom with 50 iterations corresponding to the three priors.

3D SNR Prior	phantom size: 256*256*256								
	40dB			30dB			20dB		
	St	\mathcal{NIG}	\mathcal{VG}	St	\mathcal{NIG}	\mathcal{VG}	St	\mathcal{NIG}	\mathcal{VG}
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: NMSE of reconstructed phantom with 50 iterations and computation time (s) per iteration.

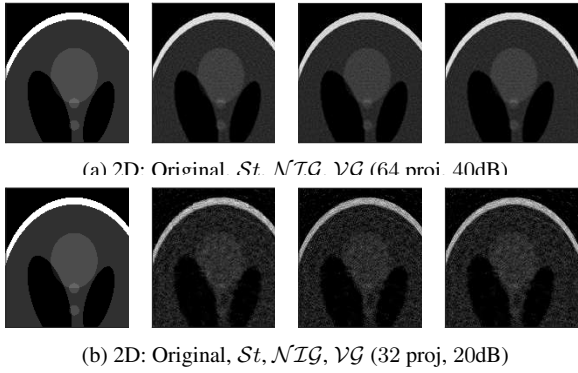


Fig. 5: 2D reconstructions corresponding to St , \mathcal{NIG} and \mathcal{VG}

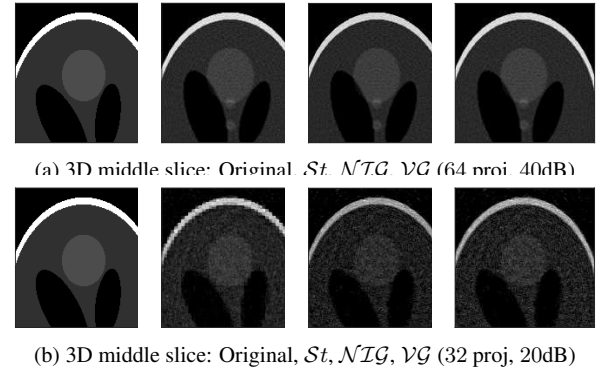


Fig. 6: 3D reconstructions corresponding to St , \mathcal{NIG} and \mathcal{VG}

The 3D reconstruction performances for all levels of noise and number of projections considered, corresponding to the optimal hyperparameters are presented in Table (2). In the 3D case, the \mathcal{NIG} seems to give better reconstruction performances for 40dB and 30dB while the St prior seems to be more appropriate for 20dB. The middle slice of reconstructed 3D Shepp Logan phantom corresponding to the three priors is presented in Figure (6): 64 proj and 40dB in Figure (6a) and 32 proj and 20dB in Figure (6b).

5. CONCLUSION

In this paper we have presented the JMAP algorithms corresponding to three different prior models in the sparsity enforcing context. We have considered heavy tailed distribution expressed as Normal variance mixture, showing that essentially the updating equations have the same analytical expression, except for the variances appearing in the hierarchical model. First we have showed different behaviours of the priors in terms of sensibility to the hyperparameters. Then we have

examined the reconstruction performances for 2D and 3D, Table (1) and (2) reporting that for different levels of noise different priors seem to give better performances. A study considering simulations results corresponding to other phantoms and real tomographic data is currently in progress. However, a more detailed study considering the sparsity rate associated with each prior depending on the hyperparameters needs to be developed in order to investigate the reconstruction results corresponding to such hyperparameters. Also, the very similar posterior mean estimation via Variational Bayesian Approximation (VBA) are to be investigated and compared. The main conclusion of this study is that for high SNR values the \mathcal{NIG} prior is very stable and gives slightly better performances while for lower SNR values the St model has better performances in terms of reconstruction.

6. REFERENCES

- [1] Li Wang, Ali Mohammad-Djafari, Nicolas Gac, and Mircea Dumitru, "Computed tomography reconstruc-

- tion based on a hierarchical model and variational Bayesian method,” in *2016 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 883–887.
- [2] Radomir S Stanković and Bogdan J Falkowski, “The Haar wavelet transform: its status and achievements,” *Computers & Electrical Engineering*, vol. 29, no. 1, pp. 25–44, 2003.
 - [3] Ronald A DeVore, Björn Jawerth, and Bradley J Lucier, “Image compression through wavelet transform coding,” *IEEE Transactions on information theory*, vol. 38, no. 2, pp. 719–746, 1992.
 - [4] Ali Mohammad-Djafari, “Bayesian approach with prior models which enforce sparsity in signal and image processing,” *EURASIP Journal on Advances in Signal Processing*, vol. 52, 2012.
 - [5] S. Derin Babacan, Shinichi Nakajima, and Minh N. Do, “Bayesian Group-Sparse Modeling and Variational Inference,” *IEEE Trans. Signal Process*, vol. 14, pp. 2906–2921, 2014.
 - [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*, vol. 47, pp. 128–156, 2015.
 - [7] Konstantinos E. Themelis, Athanasios A. Rontogiannis, and Konstantinos D. Koutroumbas, “A Variational Bayes Framework for Sparse Adaptive Estimation,” *IEEE Transactions on Signal Processing*, vol. 62, pp. 4723–4736, 2014.
 - [8] Samuel Kotz and Saralees Nadarajah, *Multivariate T-Distributions and Their Applications*, Cambridge University Press, 2010.
 - [9] Mircea Dumitru, Ali Mohammad-Djafari, and Baghai Simona S., “Precise periodic components estimation for chronobiological signals through Bayesian Inference with sparsity enforcing prior,” *EURASIP Journal on Bioinformatics and Systems Biology, Springer*, vol. 3, 2016.
 - [10] 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.
 - [11] Tor Arne Oigard and Alfred Hanssen, “The multivariate normal inverse Gaussian heavy-tailed distribution; simulation and estimation,” in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2002, pp. 1489–1492.
 - [12] 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*. IEEE, 2004, pp. 1433–1436.
 - [13] Saman Atapattu, Chintha Tellambura, and Hai Jiang, “A Mixture Gamma Distribution to Model the SNR of Wireless Channels,” *IEEE Transactions on Wireless Communications*, vol. 10, pp. 4193–4203, 2011.
 - [14] Rasool Tahmasbi and Sadeq Rezaei, “A Soft Voice Activity Detection Using GARCH Filter and Variance Gamma Distribution,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, pp. 1129–1134, 2007.
 - [15] Mircea Dumitru, *A Bayesian approach for periodic components estimation for chronobiological signals*, Ph.D. thesis, Université Paris-Saclay, Mar 2016.
 - [16] Torbjorn Eltoft, Taesu Kim, and Te-Won Lee, “On the multivariate Laplace distribution,” *IEEE Signal Processing Letters*, vol. 13, pp. 300–303, 2006.
 - [17] 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*, vol. 90, pp. 61–73, 2015.
 - [18] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge university press, 2004.