

A Student-t based sparsity enforcing hierarchical prior for linear inverse problems and its efficient Bayesian computation for 2D and 3D Computed Tomography

Ali Mohammad-Djafari, Li Wang, Nicolas Gac, Folkert Bleichrodt

▶ **To cite this version:**

Ali Mohammad-Djafari, Li Wang, Nicolas Gac, Folkert Bleichrodt. A Student-t based sparsity enforcing hierarchical prior for linear inverse problems and its efficient Bayesian computation for 2D and 3D Computed Tomography. International traveling workshop on interactions between sparse models and technology, Aug 2016, AALBORG, Denmark. hal-01403787

HAL Id: hal-01403787
https://hal.science/hal-01403787

Submitted on 27 Nov 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.

A Student-t based sparsity enforcing hierarchical prior for linear inverse problems and its efficient Bayesian computation for 2D and 3D Computed Tomography

Ali Mohammad-Djafari[1], Li Wang[1], Nicolas Gac[1] and Folkert Bleichrodt[2].

[1] Laboratoire des signaux et systèmes, CNRS, Centralesupélec-Univ Paris Saclay, Gif-sur-Yvette, France

[2] Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands

***Abstract*— In many imaging systems and in particular in X ray Computed Tomography (CT) the reconstruction problem can be written as a linear inverse problem. In these problems, one property which can often be exploited is sparsity of the solution in an appropriate basis. In this work we consider the Student-t model in its hierarchical Normal-Inverse Gamma with an appropriate dictionary based coefficient. Then, thanks to the hierarchical generative model of the observation, we derive the expression of the joint posterior law of all the unknowns and an alternate optimisation algorithm for obtaining the joint MAP solution. We then detail the implementation issues of this algorithms for parallel computation and show the results on real size 2D and 3D phantoms.**

1 Introduction

In many imaging systems, very often the reconstruction problem can be written as a linear inverse problem. This is the case in X ray Computed Tomography (CT) [1, 2]. In these problems, the solution often has a sparse representation in a suitable basis. This sparsity property can be exploited in the reconstruction algorithm. For example, for a piecewise continuous or constant image, using Haar Transform (HT) gives rise to very sparse coefficients. To impose sparsity, three great categories of priors have been used: i) Generalized Gaussian (GG), ii) mixture of two Gaussian and iii) the heavy tailed probability density functions such as Cauchy and its generalisation a Student-t [3, 4, 5, 6].

In this work we consider Student-t model in its hierarchical Normal-Inverse Gamma with an appropriate dictionary based coefficient. Then, thanks to the hierarchical generative model of the observation, we derive the expression of the joint posterior law of all the unknowns and an alternate optimisation algorithm for obtaining the joint MAP solution.

We then detail the implementation issues of this algorithms for parallel computation and show the results on real size 2D and 3D phantoms.

2 Proposed model

We consider a linear model

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

where $\boldsymbol{f}$ is the object to be reconstructed, $\boldsymbol{H}$ is the forward model, $\boldsymbol{g}$ the observed quantities and $\boldsymbol{\epsilon}$ represents the errors. We choose an appropriate basis or dictionary $\boldsymbol{D}$ for the object $\boldsymbol{f}$ in such a way that we can write

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

and such that $\boldsymbol{z}$ is sparse. For imposing the sparsity of $\boldsymbol{z}$ we propose to use the generalized Student-t prior for its elements z_j:

$$\mathcal{S}t(z_j|\alpha,\beta) = \int_0^\infty \mathcal{N}(z_j|,0,v_{z_j})\,\mathcal{IG}(v_{z_j}|\alpha,\beta)\,\mathrm{d}v_{z_j} \tag{3}$$

which gives us the possibility to write

$$\begin{cases} p(z_j|v_{z_j}) = \mathcal{N}(z_j|0,v_{z_j}) \propto {v_{z_j}}^{-1/2}\exp\left[-\frac{1}{2}\sum_j \frac{z_j^2}{v_{z_j}}\right] \\ p(v_{z_j}) = \mathcal{IG}(v_{z_j}|\alpha_{z_0},\beta_{z_0}) \propto {v_{z_j}}^{-(\alpha_{z_0}+1)}\exp\left[-\frac{\beta_{z_0}}{v_{z_j}}\right] \end{cases} \tag{4}$$

and where
$p(\boldsymbol{z}|\boldsymbol{v}_z) = \prod_j p(z_j|v_{z_j})$, $p(\boldsymbol{v}_z|\alpha_{z_0},\beta_{z_0}) = \prod_j p(v_{z_j}|\alpha_{z_0},\beta_{z_0})$
and $p(\boldsymbol{z},\boldsymbol{v}_z|\alpha_{z_0},\beta_{z_0}) = p(\boldsymbol{z}|\boldsymbol{v}_z)p(\boldsymbol{v}_z|\alpha_{z_0},\beta_{z_0})$.

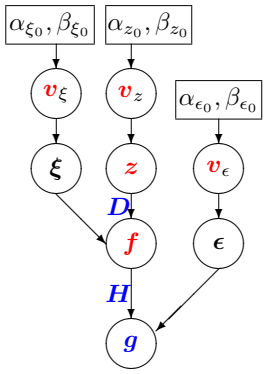
In the forward model (1) and the prior model (2) we also use Student-t for the elements of ϵ_i and ξ_j which give rise to the following relations:

$$\begin{cases} p(\boldsymbol{g}|\boldsymbol{f},v_\epsilon) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f},\boldsymbol{V}_\epsilon) \text{ with } \boldsymbol{V}_\epsilon = \mathrm{diag}\,[\boldsymbol{v}_\epsilon] \\ p(\boldsymbol{v}_\epsilon) = \prod_i \mathcal{IG}(v_{\epsilon_i}|\alpha_{\epsilon_0},\beta_{\epsilon_0}) \end{cases} \tag{5}$$

and

$$\begin{cases} p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{v}_\xi) = \mathcal{N}(\boldsymbol{f}|\boldsymbol{D}\boldsymbol{z},\boldsymbol{V}_\xi) \text{ with } \boldsymbol{V}_\xi = \mathrm{diag}\,[\boldsymbol{v}_\xi] \\ p(\boldsymbol{v}_\xi) = \prod_j \mathcal{IG}(v_{\xi_j}|\alpha_{\xi_0},\beta_{\xi_0}) \end{cases} \tag{6}$$

With these prior models, we can show the graphical generative model of the data as follows:



$$\begin{cases} \boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon} \\ \boldsymbol{f} = \boldsymbol{D}\boldsymbol{z} + \boldsymbol{\xi}, \quad \boldsymbol{z} \text{ sparse} \end{cases}$$
$$\begin{cases} p(\boldsymbol{g}|\boldsymbol{f},v_\epsilon) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f},\boldsymbol{V}_\epsilon) \\ p(\boldsymbol{f}|\boldsymbol{z}) = \mathcal{N}(\boldsymbol{f}|\boldsymbol{D}\boldsymbol{z},\boldsymbol{V}_\xi) \\ p(\boldsymbol{z}|\boldsymbol{v}_z) = \mathcal{N}(\boldsymbol{z}|0,\boldsymbol{V}_z) \\ p(v_\epsilon) = \prod_i \mathcal{IG}(v_{\epsilon_i}|\alpha_{\epsilon_0},\beta_{\epsilon_0}) \\ p(\boldsymbol{v}_z) = \prod_j \mathcal{IG}(v_{z_j}|\alpha_{z_0},\beta_{z_0}) \\ p(v_\xi) = \prod_j \mathcal{IG}(v_{\xi_j}|\alpha_{\xi_0},\beta_{\xi_0}) \end{cases}$$

Figure 1: The graphical model of the generative forward model for linear inverse problems and hierarchical sparsity enforcing model.

With these relations and equations, the expression of the posterior law of all the unknowns writes:

$$p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{v}_\epsilon,\boldsymbol{v}_z,\boldsymbol{v}_\xi|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{v}_\epsilon)\,p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{v}_\xi)\,p(\boldsymbol{z}|\boldsymbol{v}_z)\,p(\boldsymbol{v}_\epsilon)\,p(\boldsymbol{v}_z)\,p(\boldsymbol{v}_\xi) \tag{7}$$

We can then use it for inferring these unknowns. Classically, there are two major estimators: JMAP and Posterior Means (PM). The JMAP writes:

$$(\widehat{\boldsymbol{f}},\widehat{\boldsymbol{z}},\widehat{\boldsymbol{v}}_\epsilon,\widehat{\boldsymbol{v}}_z,\widehat{\boldsymbol{v}}_\xi) = \underset{(\boldsymbol{f},\boldsymbol{z},\boldsymbol{v}_\epsilon,\boldsymbol{v}_z,\boldsymbol{v}_\xi)}{\arg\max} \{p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{v}_\epsilon,\boldsymbol{v}_z,\boldsymbol{v}_\xi|\boldsymbol{g})\} \tag{8}$$

for which, the easiest optimisation is alternate optimisation. Hopefully, with the proposed hierarchical structure and conjugate priors, in any of these steps we have analytical solutions:

$$\begin{cases} \boldsymbol{f} \leftarrow \arg\min_{\boldsymbol{f}} \left\{ J(\boldsymbol{f}) = \|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2_{\boldsymbol{V}_\epsilon} + \|\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\|^2_{\boldsymbol{V}_\xi}) \right\} \\ \boldsymbol{z} \leftarrow \arg\min_{\boldsymbol{z}} \left\{ J(\boldsymbol{z}) = \|\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\|^2_{\boldsymbol{V}_\xi} + \|\boldsymbol{z}\|^2_{\boldsymbol{V}_z} \right\} \\ v_{\epsilon_i} \leftarrow (\beta_{\epsilon_0} + \frac{1}{2}(g_i - [\boldsymbol{H}\boldsymbol{f}]_i)^2 \;/\; (\alpha_{\epsilon_0} + 3/2) \\ v_{\xi_j} \leftarrow (\beta_{\xi_0} + \frac{1}{2}(f_j - [\boldsymbol{D}\boldsymbol{z}]_j)^2 \;/\; (\alpha_{\xi_0} + 3/2) \\ v_{z_j} \leftarrow (\beta_{z_0} + \frac{1}{2}|z_j|^2) \;/\; (\alpha_{z_0} + 3/2) \end{cases} \tag{9}$$

As we can see, the main part of the algorithm is two optimizations of quadratic criteria which can also be written as

$$\begin{cases} J(\boldsymbol{f}) &= \|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2_{\boldsymbol{V}_\epsilon} + \|\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\|^2_{\boldsymbol{V}_\xi} \\ &= \|\boldsymbol{V}_\epsilon^{-1/2}(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})\|_2^2 + \|\boldsymbol{V}_\xi^{-1/2}(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z})\|_2^2 \\ J(\boldsymbol{z}) &= \|\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\|^2_{\boldsymbol{V}_\xi} + \|\boldsymbol{z}\|^2_{\boldsymbol{V}_z} \\ &= \|\boldsymbol{V}_\xi^{-1/2}(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z})\|_2^2 + \|\boldsymbol{V}_z{}^{-1/2}\boldsymbol{z}\|_2^2 \end{cases} \tag{10}$$

For implementation of the gradient based optimisation algorithms we need their gradients:

$$\begin{cases} \nabla J(\boldsymbol{f}) = -2\boldsymbol{H}'\boldsymbol{V}_\epsilon{}^{-1}(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}) + 2\boldsymbol{V}_\xi{}^{-1}(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}) \\ \nabla J(\boldsymbol{z}) = -2\boldsymbol{D}'\boldsymbol{V}_\xi{}^{-1}(\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}) + 2\boldsymbol{V}_z{}^{-1}\boldsymbol{z} \end{cases} \tag{11}$$

The algorithm we propose here which can be implemented efficiently and scale up for real applications and Big Data is as follows

$$\begin{cases} \boldsymbol{f} \leftarrow \boldsymbol{f} - \gamma_f \nabla J(\boldsymbol{f}) \\ \boldsymbol{z} \leftarrow \boldsymbol{z} - \gamma_z \nabla J(\boldsymbol{z}) \\ v_{\epsilon_i} \leftarrow (\beta_{\epsilon_0} + \frac{1}{2}(g_i - [\boldsymbol{H}\boldsymbol{f}]_i)^2 \;/\; (\alpha_{\epsilon_0} + 3/2) \\ v_{\xi_j} \leftarrow (\beta_{\xi_0} + \frac{1}{2}(f_j - [\boldsymbol{D}\boldsymbol{z}]_j)^2 \;/\; (\alpha_{\xi_0} + 3/2) \\ v_{z_j} \leftarrow (\beta_{z_0} + \frac{1}{2}|z_j|^2) \;/\; (\alpha_{z_0} + 3/2) \end{cases} \tag{12}$$

where γ_f and γ_z have to be adapted and updated at each iteration.

3 Simulation results

To show the effectiveness of the proposed method, we show here two examples of X ray image reconstruction from 32 projections. The images have 128x128 and 512x512 pixels and in both cases we simulated 32 projections uniformly distributed between 0 and 180 degrees. A Gaussian noise is added in such a way to have a SNR of 20dB.

We will show examples of 3D reconstruction results in the final paper. These results are obtained with objects of volume 256x256x256 and 32 projections of size 256x256.

4 Conclusion

We proposed a hierarchical Normal-Inverse Gamma prior for modelling the sparsity of both the error terms of the data-forward model and the error terms of dictionary based decomposition of the unknown images. With these priors we obtain an expression for the joint posterior law of all the unknowns (image itself $\boldsymbol{f}$, the coefficients of the decomposition $\boldsymbol{z}$ and their respective hidden variances $\boldsymbol{v}_\epsilon$, $\boldsymbol{v}_\xi$ and $\boldsymbol{v}_z$). An approximate Bayesian Computation (ABC) based on the alternate optimization of this joint posterior law with respect its arguments gives an algorithm which can be implemented in an efficient way which can scales up for real applications.

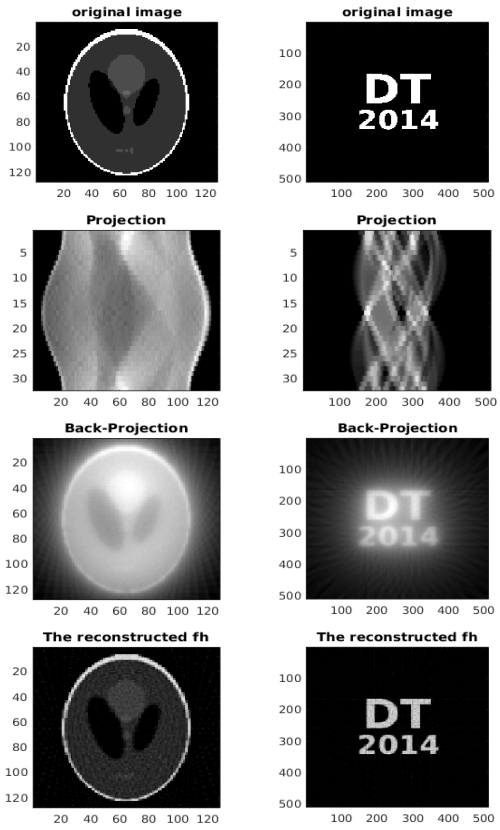


Figure 2: X ray CT: Originals, Projections, Initialization by Back-projection (BP), Reconstruction results.

References

[1] A. Mohammad-Djafari, *Inverse Problems in Vision and 3D Tomography*. ISTE-WILEY, 2010.

[2] K. Batenburg, S. Bals, J. Sijbers, C. Kubel, P. Midgley, J. Hernandez, U. Kaiser, E. Encina, E. Coronado, and G. Tendeloo, "3d imaging of nanomaterials by discrete tomography," *Ultramicroscopy*, vol. 109, no. 6, pp. 730–740, 2009.

[3] J.-C. Baritaux, K. Hassler, and M. Unser, "An efficient numerical method for general l_p regularization in fluorescence molecular tomography," *IEEE Transactions on Medical Imaging*, vol. 29, pp. 1075–1087, April 2010.

[4] F. Luisier, C. Vonesch, T. Blu, and M. Unser, "Fast Haar-wavelet denoising of multidimensional fluorescence microscopy data," in *Proceedings of the Sixth IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'09)*, (Boston MA, USA), pp. 310–313, June 28-July 1 2009.

[5] A. Mohammad-Djafari, "Bayesian approach with prior models which enforce sparsity in signal and image processing," *EURASIP Journal on Advances in Signal Processing*, vol. Special issue on Sparse Signal Processing, p. 2012:52, 2012.

[6] R. Leary, Z. Saghi, P. Midgley, and D. Holland, "Compressed sensing electron tomography," *Ultramicroscopy*, vol. 131, pp. 70–91, 2013.