

# 3D X-ray Computed Tomography reconstruction using sparsity enforcing Hierarchical Model based on Haar Transformation

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

**Abstract**—In this paper, we consider the 3D X-ray CT reconstruction problem by using the Bayesian approach with a hierarchical prior model. A generalized Student-t distributed prior model is used to enforce the sparse structure of the multilevel Haar Transformation of the image. Comparisons with some state of the art methods are presented, showing that the proposed method gives more accurate reconstruction results and a faster convergence. Simulation results are also provided to show the effectiveness of the proposed hierarchical model for a reconstruction with more limited projections.

**Index Terms**—Computed Tomography (CT), Bayesian Approach, Hierarchical Model, Generalized Student-t distribution, Joint Maximum A Posterior (JMAP).

## I. INTRODUCTION

An introduction of the development of Tomography and several analytic and algebraic state of the art reconstruction methods are presented in [1]. In X-ray CT, the intensity of X radiation is attenuated when passing through the object, and the parameters to be reconstructed are the linear attenuation coefficients inside the object under the test. The Radon Transform (RT), presented in detail in [1], is one of the most commonly used forward modeling when treating the 3D X-ray CT projection problem, with the expression:

$$g(r, \phi) = \mathcal{R}f(x, y) = \int_{L_{r, \phi}} f(x, y) dl \quad (1)$$

where  $f(x, y)$  represents the attenuation coefficient,  $r$  is the perpendicular length from center point of coordinate and  $\phi$  is the considered X ray angle.  $L_{r, \phi}$  is the length of ray  $(r, \phi)$  passing through the object.

There have been different analytic methods to solve the reconstruction problem, for example the Back-Projection (BP) [2], the Filtered Back-Projection (FBP) [2], [3], etc. We may mention here the FBP method, which has a very good performance when dealing with sufficient detected data. It can be summarized as  $\hat{f} = \mathcal{B}\mathbf{F}^{-1}|\Omega|\mathbf{F}g$  where  $\mathbf{F}$  and  $\mathbf{F}^{-1}$  represent the direct and the inverse Fourier Transform (FT),  $|\Omega|$  a modulated filter and  $\mathcal{B}$  the Back Projection.

### A. Deterministic Regularization based methods

Comparing with the conventional FBP method, the regularization and statistical methods give more satisfactory reconstruction results [3]–[5]. The statistical methods can also give more robust and precise results comparing with FBP reconstruction method, especially in the low dose and time limit situations where insufficient data are given.

In many algebraic and statistical methods, one consider the linear forward model:

$$g = \mathbf{H}\mathbf{f} + \epsilon \quad (2)$$

where  $g \in \mathbb{R}^{M \times 1}$  represents the projection data,  $f \in \mathbb{R}^{N \times 1}$  the object and  $\epsilon \in \mathbb{R}^{M \times 1}$  the additive noise. Matrix  $\mathbf{H} \in \mathbb{R}^{M \times N}$  corresponds to the linear projection system. In the regularization methods, the result of reconstruction is obtained by optimizing the criterion which combines a data-model term and the regularization terms.

A general criterion of the regularization method is  $\mathcal{J}(f) = \|g - \mathbf{H}\mathbf{f}\|^2 + \lambda\mathbf{R}(f)$  where  $\mathbf{R}(f)$  is the regularization term and  $\lambda$  is called the regularization parameter. We list out some conventional regularization reconstruction methods: (a) the Least Square (LS) method with  $\mathbf{R}(f) = 0$  where the estimating result minimize the disparity but is not guaranteed to be precise because of the ill-posedness; (b) the Quadratic Regularization (QR) method [4] with  $\mathbf{R}(f) = \|\mathbf{D}\mathbf{f}\|^2$  which enforces the global smoothness of the estimated result and (c) the Total Variation (TV) method [6], where the difference of neighbour pixels are enforced to be sparse with  $\mathbf{R}(f) = \|\mathbf{D}\mathbf{f}\|_1$ . By using  $l_1$  norm, the sparsity of the penalty term is enforced. The appearance of the non-differentiable  $l_1$  term leads to difficulties for the computation of the gradient of constraints. Many methods have been studied in order to solve this  $l_1$  norm optimization problem, for example the Newton's method [7] and the Split Bregman method [6].

Note that in all the above mentioned regularization methods, there is a parameter  $\lambda$  controlling the trade-off of the disparity and the regularization penalty, which need to be fixed. There have been different methods to choose a suitable value for  $\lambda$ , for example the Cross Validation (CV) and L-curve methods, with details presented in [8], [9]. However, the computation for choosing this parameter should be done for each simulation and different dataset, which is not practical. Bayesian methods, therefore, are often used to estimate the parameters and variables simultaneously.

### B. Bayesian Method

As mentioned above, the Bayesian methods solves the reconstruction problem as well as the estimation of the parameters at the same time thanks to appropriate prior modelling. In short, by using the Bayes formula, the complete solution of the inverse problem is provided by the posterior distribution:

$$p(\mathbf{f}, \boldsymbol{\theta} | g) = \frac{p(g | \mathbf{f}, \boldsymbol{\theta})p(\mathbf{f} | \boldsymbol{\theta})p(\boldsymbol{\theta})}{p(g)}, \quad (3)$$

Which is then used to infer both on  $\mathbf{f}$  and on  $\theta$ .

## II. THE SPARSITY ENFORCING HIERARCHICAL HAAR TRANSFORM BASED METHOD (HHBM)

The HHBM method presented in this paper is an extension of the work presented in [10]. Instead of estimating only the hidden variable  $\mathbf{z}$  and obtaining the final reconstruction result  $\mathbf{f}$  by a post-processing, here a more general model is proposed, and the method is tested on 3D object reconstruction simulations.

Typically the sparse property of the Haar transformation coefficient is enforced by using three kinds of distributions: the Generalized Gaussian distributions, the Gaussian Mixture distributions and the heavy tailed distributions.

The standard Student-t distribution is heavy tailed, but from the definition of its variance we easily figure out that there is a lower limit of the variance value of Student-t distribution:  $\text{Var}[f] = \frac{\nu}{\nu-2} > 1$ , ( $\nu > 2$ ). This limit implies that this heavy-tailed distribution can't have a small variance, therefore the sparsity couldn't be intensively enforced. In this paper we use a generalization of Student-t distribution (Stg) which can be obtained by using the marginalization of the Normal-Inverse Gamma bi-variate distribution:

$$\text{St}_g(f|\alpha, \beta) = \int \mathcal{N}(f|0, v) \mathcal{IG}(v|\alpha, \beta) dv. \quad (4)$$

This generalization of Student-t distribution adds a supplementary parameter compared to the standard one, and hence more capable to control the level of sparsity of the prior distribution.

### A. Hierarchical Bayesian Model of HHBM

The additive noise is defined belonging to an i.i.d. Gaussian law with mean equal to zero and unknown variance vector  $\mathbf{v}_\epsilon$ :  $p(\epsilon|\mathbf{v}_\epsilon) = \mathcal{N}(\epsilon|\mathbf{0}, \mathbf{V}_\epsilon)$ , where  $\mathbf{V}_\epsilon = \text{diag}[\mathbf{v}_\epsilon]$  is a diagonal matrix. According to the linear forward model shown in Eq.(2), we obtain an expression for the likelihood:  $p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon) = \mathcal{N}(\mathbf{g}|\mathbf{H}\mathbf{f}, \mathbf{V}_\epsilon)$ .

In the considered applications, generally the object to be reconstructed is piece-wise homogeneous, considering that it consists of several different materials. With this property, an information which can be considered as a prior knowledge is the sparseness of the contours.

The sparse representation of the prior piece-wise continuous image used in this paper is the multilevel Haar Transformation. A vector  $\mathbf{z}$  is used to present the  $l$ -level Haar transformation coefficients of  $\mathbf{f}$ . Prior distribution of  $\mathbf{f}$  depends on  $\mathbf{z}$ :  $p(\mathbf{f}|\mathbf{z}, \mathbf{v}_\xi) = \mathcal{N}(\mathbf{f}|\mathbf{D}\mathbf{z}, \mathbf{V}_\xi)$  where  $\mathbf{D}$  represents the inverse Multilevel Haar transformation operator, and the additive noise  $\xi$  is considered to be i.i.d. Gaussian distributed. Definition of diagonal matrix:  $\mathbf{V}_\xi = \text{diag}[\mathbf{v}_\xi]$ .

The vector  $\mathbf{z} = [z_1, z_2, \dots, z_N]$ , therefore, is sparse. As mentioned above, the generalized Student-t distribution is used in order to enforce the sparsity. By using the Normal-Inverse Gamma property given in Eq.(4), the prior distribution for  $\mathbf{z}$  is:

$$\begin{cases} p(\mathbf{z}|\mathbf{v}_z) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{V}_z) \text{ where } \mathbf{V}_z = \text{diag}[\mathbf{v}_z], \\ p(\mathbf{v}_z|\alpha_{z_0}, \beta_{z_0}) = \prod_j^N \mathcal{IG}(v_{z_j}|\alpha_{z_0}, \beta_{z_0}), \end{cases} \quad (5)$$

where the elements of vector  $\mathbf{v}_z$  are supposed to be i.i.d.

On the other hand, when considering the variance of two noises:  $\mathbf{v}_\epsilon$  and  $\mathbf{v}_\xi$ , by knowing that the variance is positive, and the fact that the majority of the values are small, we choose the Inverse Gamma distribution to model them:  $p(\mathbf{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \prod_i^M \mathcal{IG}(v_{\epsilon_i}|\alpha_{\epsilon_0}, \beta_{\epsilon_0})$  and  $p(\mathbf{v}_\xi|\alpha_{\xi_0}, \beta_{\xi_0}) = \prod_j^N \mathcal{IG}(v_{\xi_j}|\alpha_{\xi_0}, \beta_{\xi_0})$ .

With all the proposed prior distributions, the model containing all the variables, parameters and hyper-parameters is:

$$\begin{cases} p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon) \propto |\mathbf{V}_\epsilon|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\mathbf{g} - \mathbf{H}\mathbf{f})^T \mathbf{V}_\epsilon^{-1} (\mathbf{g} - \mathbf{H}\mathbf{f}) \right], \\ p(\mathbf{f}|\mathbf{z}, \mathbf{v}_\xi) \propto |\mathbf{V}_\xi|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\mathbf{f} - \mathbf{D}\mathbf{z})^T \mathbf{V}_\xi^{-1} (\mathbf{f} - \mathbf{D}\mathbf{z}) \right], \\ p(\mathbf{z}|\mathbf{v}_z) \propto |\mathbf{V}_z|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \mathbf{z}^T \mathbf{V}_z^{-1} \mathbf{z} \right], \\ p(\mathbf{v}_z|\alpha_{z_0}, \beta_{z_0}) \propto \prod_j^N v_{z_j}^{-(\alpha_{z_0}+1)} \exp \left[ -\beta_{z_0} v_{z_j}^{-1} \right], \\ p(\mathbf{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \propto \prod_i^M v_{\epsilon_i}^{-(\alpha_{\epsilon_0}+1)} \exp \left[ -\beta_{\epsilon_0} v_{\epsilon_i}^{-1} \right], \\ p(\mathbf{v}_\xi|\alpha_{\xi_0}, \beta_{\xi_0}) \propto \prod_j^N v_{\xi_j}^{-(\alpha_{\xi_0}+1)} \exp \left[ -\beta_{\xi_0} v_{\xi_j}^{-1} \right]. \end{cases} \quad (6)$$

The corresponding directed acyclic graph (DAG) of the proposed model is shown in Fig.(1).

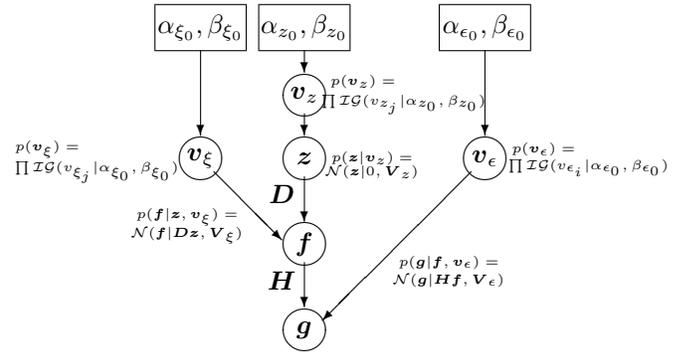


Fig. 1: DAG of proposed model.

### B. Bayesian Inference

Via the Bayesian inference, the posterior law is obtained from the likelihood and priors:

$$p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\xi, \mathbf{v}_z|\mathbf{g}) \propto p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon) p(\mathbf{f}|\mathbf{z}, \mathbf{v}_\xi) p(\mathbf{z}|\mathbf{v}_z) \cdot p(\mathbf{v}_\epsilon) p(\mathbf{v}_\xi) p(\mathbf{v}_z) \quad (7)$$

From the posterior distribution obtained, different estimation methods can be used. Mainly there are two options: Posterior Mean (PM) and Joint Maximum A Posterior (JMAP) [11]. The first one can be computed either by MCMC methods [12], Variational Bayesian Approximation (VBA) [13] or any other approximation methods. However, the computational costs of these methods are too high for 3D applications. This is the reason why we use the JMAP method to estimate all the variables iteratively in this paper.

### C. Estimation

The JMAP computation aims at iteratively and alternately estimate the variables and parameters by maximizing the posterior:

$$(\hat{\mathbf{f}}, \hat{\mathbf{z}}, \hat{\mathbf{v}}_\epsilon, \hat{\mathbf{v}}_\xi, \hat{\mathbf{v}}_z) = \arg \max_{\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\xi, \mathbf{v}_z} \{p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\xi, \mathbf{v}_z|\mathbf{g})\} \quad (8)$$

30iter	phantom size: 64*64*64											
	64 projections						32 projections					
	40dB			20dB			40dB			20dB		
	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM
$err_f$	0.1138	0.0598	0.0228	0.1354	0.0636	0.0739	0.1537	0.1296	0.0696	0.1799	0.1333	0.1080
$err_g$	0.0020	0.0013	5.1537*e-4	0.0046	0.0060	0.0105	0.0015	0.0030	0.0019	0.0030	0.0077	0.0102
Tc	0.2561	0.7681	1.8336	0.2346	0.7880	2.1800	0.2287	0.7526	1.6044	0.2121	0.7185	1.5533

TABLE I: Comparison of NMSE of reconstructed phantom with 30 iterations and the computation time of each iteration by using different methods.

Because of the huge size of data, the descend gradient algorithm is used since the computation of big size matrix inversion is too expensive. The iterative updating rule is:

$$\begin{aligned}
\text{iter} : \hat{\mathbf{f}}^{(k+1)} &= \hat{\mathbf{f}}^{(k)} - \hat{\gamma}_f^{(k)} \nabla \mathcal{J}(\hat{\mathbf{f}}^{(k)}); \\
\text{iter} : \hat{\mathbf{z}}^{(k+1)} &= \hat{\mathbf{z}}^{(k)} - \hat{\gamma}_z^{(k)} \nabla \mathcal{J}(\hat{\mathbf{z}}^{(k)}); \\
\hat{v}_{z_j} &= \left( \beta_{z_0} + \frac{1}{2} \hat{z}_j^2 \right) / (\alpha_{z_0} + 3/2), \quad \forall j \in [1, N]; \\
\hat{v}_{\epsilon_i} &= \left( \beta_{\epsilon_0} + \frac{1}{2} \left( g_i - [\mathbf{H}\hat{\mathbf{f}}]_i \right)^2 \right) / (\alpha_{\epsilon_0} + 3/2), \quad \forall i \in [1, M]; \\
\hat{v}_{\xi_j} &= \left( \beta_{\xi_0} + \frac{1}{2} \left( \hat{f}_j - [\mathbf{D}\hat{\mathbf{z}}]_j \right)^2 \right) / (\alpha_{\xi_0} + 3/2), \quad \forall j \in [1, N],
\end{aligned} \tag{9}$$

where

$$\begin{aligned}
\mathcal{J}(\mathbf{f}) &= \frac{1}{2} \left\| \mathbf{V}_\epsilon^{-\frac{1}{2}} (\mathbf{g} - \mathbf{H}\mathbf{f}) \right\|^2 + \frac{1}{2} \left\| \mathbf{V}_\xi^{-\frac{1}{2}} (\mathbf{f} - \mathbf{D}\mathbf{z}) \right\|^2; \\
\mathcal{J}(\mathbf{z}) &= \frac{1}{2} \left\| \mathbf{V}_\epsilon^{-\frac{1}{2}} (\mathbf{f} - \mathbf{D}\mathbf{z}) \right\|^2 + \frac{1}{2} \left\| \mathbf{V}_z^{-\frac{1}{2}} \mathbf{z} \right\|^2; \\
\hat{\gamma}_f^{(k)} &= \left( \left\| \nabla \mathcal{J}(\hat{\mathbf{f}}^{(k)}) \right\|^2 \right) / \left( \left\| \hat{\mathbf{Y}}_\epsilon \mathbf{H} \nabla \mathcal{J}(\hat{\mathbf{f}}^{(k)}) \right\|^2 + \left\| \hat{\mathbf{Y}}_\xi \nabla \mathcal{J}(\hat{\mathbf{f}}^{(k)}) \right\|^2 \right); \\
\hat{\gamma}_z^{(k)} &= \left( \left\| \nabla \mathcal{J}(\hat{\mathbf{z}}^{(k)}) \right\|^2 \right) / \left( \left\| \hat{\mathbf{Y}}_\epsilon \mathbf{D} \nabla \mathcal{J}(\hat{\mathbf{z}}^{(k)}) \right\|^2 + \left\| \hat{\mathbf{Y}}_z \nabla \mathcal{J}(\hat{\mathbf{z}}^{(k)}) \right\|^2 \right),
\end{aligned} \tag{10}$$

where  $\nabla \mathcal{J}(\cdot)$  is the gradient of  $\mathcal{J}(\cdot)$ .  $\hat{\gamma}_f$  and  $\hat{\gamma}_z$  are obtained by using optimized step length strategy, see [14]. The initialization of the hyper parameters  $\alpha$ s and  $\beta$ s are:  $\alpha_{z_0} = \alpha_{\epsilon_0} = \alpha_{\xi_0} = 2 + \epsilon_1$ ,  $\beta_{z_0} = \beta_{\epsilon_0} = \beta_{\xi_0} = \epsilon_2$  where  $\epsilon_1$  and  $\epsilon_2$  are small values close to zero.

### III. SIMULATION RESULTS

In the simulations, the Shepp Logan phantom with size  $64^3$  is used as the original object. It consists of several homogeneous zones, each of which corresponds to a different material. Projections are applied in angles uniformly distributed from  $0^\circ$  to  $180^\circ$ . The reconstruction performance is measured in terms of the Normalize Mean Square Error (NMSE), or the relative error  $\delta_f$ , defined as  $\delta_f = \text{NMSE} = \left\| \hat{\mathbf{f}} - \mathbf{f} \right\|^2 / \left\| \mathbf{f} \right\|^2$ , where  $\hat{\mathbf{f}}$  is the reconstructed result.

The comparison is between a) the proposed method, b) the Quadratic Regularization (QR) method and c) the Total Variation (TV) method. The simulation of TV method is realized by using the Split Bregman method [6]. The descend gradient optimization algorithm is used in HHBM method as mentioned in Eq.(9).

In order to deal with 3D data, considering the big data size constraint, a GPU processor is used. Several toolbox are accessible for usage of GPU in MATLAB. Here we used the

ASTRA toolbox [15], which facilitates the coding with GPU in MATLAB.

The middle slice of the reconstructed phantom is shown in Fig.(2). The left figures show the reconstructed middle slices of the 3D Shepp-Logan phantom, and on the right there are the details of the corresponding middle slice.

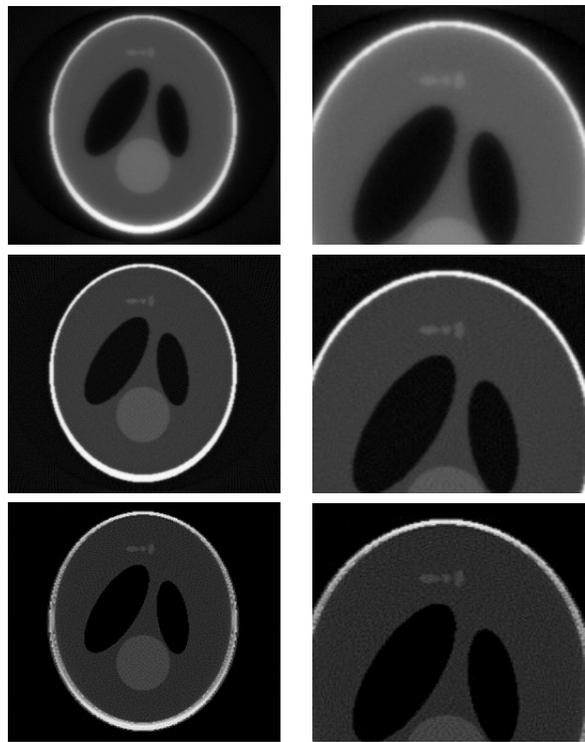


Fig. 2: The middle slices and zones of the reconstructed 3D Shepp-Logan phantom by using different methods with 1/4 of total projections with noise of SNR=40dB. Top: QR, middle: TV, bottom: HHBM.

The NMSE by using different methods are compared. Fig.(3) shows the  $\delta_f$  of reconstructed Shepp Logan phantom of size  $64^3$  by using 64 projections, when SNR=40dB and 20dB. It shows that when the noise is at a low level, the HHBM method converges much faster and gives a more precise reconstructed result, when with a more important noise, the TV method keeps robust, but the HHBM method still converges faster than the others. Fig.(4) compares the  $\delta_f$  of the reconstructed phantom of size  $64^3$  with 32 projections. It shows that when there are insufficient projection data, the HHBM method remains robust, and the convergence is efficient. This property of the HHBM method is important in

the applications where the reconstruction time is important and the results need to be obtained within a few iterations.

More results are given in Tab.(I), where  $err_f$  and  $err_g$  are respectively the NMSE of reconstructed phantom and reconstructed projections after 30 iterations. The value  $T_c$  shows the computation time for each iteration. From the table and the figures we can see that, even though the computation time for one iteration of HHBM is longer than TV method (about 2 times), the number of iterations needed for reaching convergence by using HHBM is much smaller than TV method.

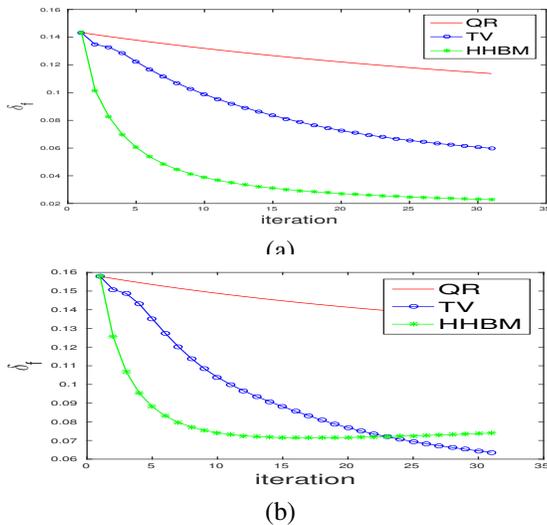


Fig. 3: Comparison of NMSE of reconstructed Shepp Logan phantom of size  $64^3$  by using different methods with 64 projections and noise of (a) 40dB and (b) 20dB.

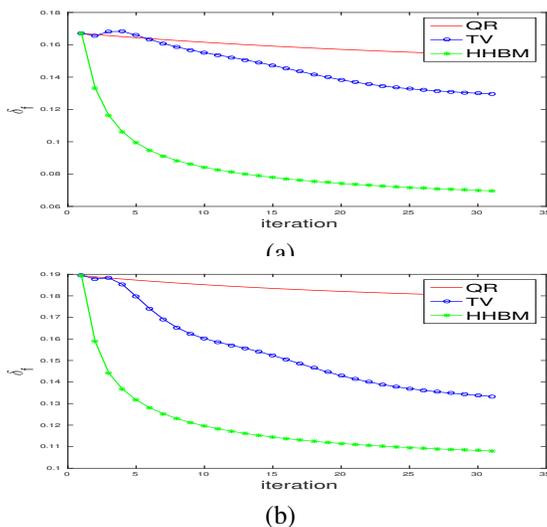


Fig. 4: Comparison of NMSE of reconstructed Shepp Logan phantom of size  $64^3$  by using different methods with 32 projections and noise of (a) 40dB and (b) 20dB.

## IV. CONCLUSION

We proposed to enforce sparsity of the multi-level Haar Transformation of the piece-wise continuous images by using a Generalized Student-t distribution, and use the Bayesian method to estimated the image and the transformation coefficients simultaneously. We compared the proposed HHBM method with the Quadratic Regularization and the Total Variation method. We conclude from the simulation results that the HHBM method is more robust for cases with less number of projections and higher noise level. The convergence speed is faster than the state of the art methods. As future work, we search for better sparse representation bases, and a more general Bayesian estimation going to Posterior Mean (PM) computation via Variational Bayesian Approximation (VBA).

## REFERENCES

- [1] Rolf Clackdoyle and Michel Defrise, "Tomographic reconstruction in the 21st century," *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 60–80, 2010.
- [2] Stanley R Deans, *The Radon transform and some of its applications*, Courier Corporation, 2007.
- [3] Nabil Chetih and Zoubeida Messali, "Tomographic image reconstruction using Filtered Back Projection (FBP) and Algebraic Reconstruction Technique (ART)," in *Control, Engineering & Information Technology (CEIT), 2015 3rd International Conference on*. IEEE, 2015, pp. 1–6.
- [4] Simon Rit, David Sarrut, and Laurent Desbat, "Comparison of analytic and algebraic methods for motion-compensated cone-beam CT reconstruction of the thorax," *IEEE Transactions on Medical Imaging*, vol. 28, no. 10, pp. 1513–1525, 2009.
- [5] Jang Hwan Cho and Jeffrey A Fessler, "Regularization designs for uniform spatial resolution and noise properties in statistical image reconstruction for 3-D X-ray CT," *IEEE Transactions on Medical Imaging*, vol. 34, no. 2, pp. 678–689, 2015.
- [6] Tom Goldstein and Stanley Osher, "The Split Bregman method for L1-regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.
- [7] Tony F Chan, Gene H Golub, and Pep Mulet, "A nonlinear primal-dual method for Total Variation-based image restoration," *SIAM Journal on Scientific Computing*, vol. 20, no. 6, pp. 1964–1977, 1999.
- [8] Sathish Ramani, Zhihao Liu, Jeffrey Rosen, Jon-Fredrik Nielsen, and Jeffrey A Fessler, "Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3659–3672, 2012.
- [9] Nikolas P Galatsanos and Aggelos K Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Transactions on Image Processing*, vol. 1, no. 3, pp. 322–336, 1992.
- [10] 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 International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 883–887.
- [11] Jinyi Qi and Richard M Leahy, "Resolution and noise properties of MAP reconstruction for fully 3-D PET," *IEEE Transactions on Medical Imaging*, vol. 19, no. 5, pp. 493–506, 2000.
- [12] Walter R Gilks, *Markov chain monte carlo*, Wiley Online Library.
- [13] Dimitris G Tzikas, Aristidis C Likas, and Nikolaos P Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 131–146, 2008.
- [14] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge university press, 2004.
- [15] Willem Jan Palenstijn, K Joost Batenburg, and Jan Sijbers, "The ASTRA tomography toolbox," in *13th International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE*, 2013, vol. 2013.