



HAL
open science

Bayesian method with sparsity enforcing prior of dual-tree complex wavelet transform coefficients for X-ray CT image reconstruction

Li Wang, Ali Mohammad-Djafari, Nicolas Gac

► **To cite this version:**

Li Wang, Ali Mohammad-Djafari, Nicolas Gac. Bayesian method with sparsity enforcing prior of dual-tree complex wavelet transform coefficients for X-ray CT image reconstruction. 25th European Signal Processing Conference (EUSIPCO 2017), Aug 2017, Kos island, Greece. 10.23919/eusipco.2017.8081253 . hal-01567875

HAL Id: hal-01567875

<https://hal.science/hal-01567875v1>

Submitted on 24 Jul 2017

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.

Bayesian Method With Sparsity Enforcing Prior of Dual-Tree Complex Wavelet Transform Coefficients for X-ray CT Image Reconstruction

Li Wang

Laboratoire des signaux et système
Université Paris Saclay
Email: li.wang@lss.supelec.fr

Ali Mohammad-Djafari

Laboratoire des signaux et système
CNRS, France
Email: djafari@lss.supelec.fr

Nicolas Gac

Laboratoire des signaux et système
Université Paris Saclay
Email: nicolas.gac@lss.supelec.fr

Abstract—In this paper, a Bayesian method with a hierarchical sparsity enforcing prior model for Dual-Tree Complex Wavelet Transform (DT-CWT) coefficients is proposed. This model is used for X-ray Computed Tomography (CT) image reconstruction. A generalized Student-t distributed prior model is used to enforce the sparse structure of the DT-CWT coefficient of the image. The joint Maximum A Posterior algorithm (JMAP) is used in this Bayesian context. Comparisons with the conventional and other state-of-the-art methods are presented, showing that the proposed method gives more accurate and robust reconstruction results while the dataset is insufficient.

Index Terms—Computed Tomography (CT), Bayesian Approach, Hierarchical Model, Dual-Tree Complex Wavelet Transformation (DT-CWT), Generalized Student-t distribution, Joint Maximum A Posterior (JMAP).

I. INTRODUCTION

In X-ray CT, the intensity of X-ray 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 the most commonly used forward modeling for X-ray CT. There have been different analytic methods to solve the reconstruction problem, for example the Back-Projection (BP) [2], the Filtered Back-Projection (FBP) [2], etc. Iterative methods are developed with a lot of efforts in order to solve the ill-posed CT reconstruction, concerning fewer projection numbers or less exposed time during the projection. Accounting for the modeling uncertainties and measurement noise, the forward system is modeled by:

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon \quad (1)$$

where $\mathbf{g} \in \mathbb{R}^{M \times 1}$ represents the projection data, $\mathbf{f} \in \mathbb{R}^{N \times 1}$ 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.

The regularization methods [3], [4] are often used for the ill-posed inverse problems. A general criterion of the regularization method is $\mathcal{J}(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \lambda\mathbf{R}(\mathbf{f})$ where $\mathbf{R}(\mathbf{f})$ is the regularization term and λ is called the regularization parameter. We list out some conventional regularization reconstruction methods: (a) the Least Square (LS) method [4] with $\mathbf{R}(\mathbf{f}) = 0$ where the obtained result minimizes the

disparity but is not guaranteed to be precise because of the ill-posedness; (b) the Quadratic Regularization (QR) method [5] with $\mathbf{R}(\mathbf{f}) = \|\mathbf{D}\mathbf{f}\|_2^2$ 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 $\mathbf{R}(\mathbf{f}) = \|\mathbf{D}\mathbf{f}\|_1$. 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]. Another class of the regularization method, called "Synthesis", considers a linear sparse transformation $\mathbf{f} = \mathbf{D}\mathbf{z}$ and minimizes a criterion: $\mathcal{J}(\mathbf{z}) = \|\mathbf{g} - \mathbf{H}\mathbf{D}\mathbf{z}\|_2^2 + \lambda\mathbf{R}(\mathbf{z})$ for example $\mathbf{R}(\mathbf{z}) = \|\mathbf{z}\|_1$ [8]. When $\hat{\mathbf{z}}$ obtained the object is reconstructed by $\hat{\mathbf{f}} = \mathbf{D}\hat{\mathbf{z}}$.

Note that in all the above mentioned regularization methods, there is a parameter λ controlling the trade-off of the disparity and the regularization penalty, and it needs to be fixed. There have been different methods to choose a suitable value for λ , for example the Cross Validation (CV) and L-curve methods, with details presented in [9], [10]. However, the computation for choosing this parameter should be done for each situation and different dataset, which is very costly in computation. Bayesian methods [11], [12], therefore, are often used to estimate the parameters and variables simultaneously [13].

The choice of prior distributions is crucial in Bayesian methods. In a previous work [14], a hierarchical structured prior distribution is defined via the Haar transformation, with the direct model $\mathbf{f} = \mathbf{D}\mathbf{z} + \boldsymbol{\xi}$, with coefficients presented by \mathbf{z} and transform operator by \mathbf{D} , and $\boldsymbol{\xi}$ defines the uncertainties which relax the transformation relationship in this hierarchical model. $\boldsymbol{\xi}$ is then modelled through a heavy tailed distribution to enforce its sparsity. In particular a generalized form of Student-t distribution (details presented in Section II) has been used and can be presented with a hierarchical model by defining $p(\boldsymbol{\xi}|\mathbf{v}_\xi)$ as a normal distribution and $p(\mathbf{v}_\xi)$ as an Inverse Gamma distribution, where \mathbf{v}_ξ is the variance of noise $\boldsymbol{\xi}$. In this way, one can obtain $p(\mathbf{f}|\mathbf{z}, \mathbf{v}_\xi)$, $p(\mathbf{z}|\mathbf{v}_z)$ and $p(\mathbf{v}_\xi)$ and by combining with the likelihood $p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon)$ and

$p(v_\epsilon)$, where v_ϵ is the variance of the noise ϵ , the posterior law $p(\mathbf{f}, \mathbf{z}, v_\epsilon, v_\xi, v_z | \mathbf{g})$ is obtained and is used for doing inference. In this paper we focus on the choice of the linear transformation \mathbf{D} . We had used the Haar Transform before [14] for piecewise constant object. This time we focus on DT-CWT which can be used in more general object types as for example piecewise continuous ones. We give some arguments on the appropriateness of DT-CWT and show its performances.

II. THE HIERARCHICAL BAYESIAN METHOD

In our previous work [14], the discrete Haar transformation coefficient is used as the sparse representation of the piecewise continuous image. There are also variety of works on X-ray CT based on discrete transformations or dictionary bases [8], [15], [16]. In this paper, the Dual Tree-Complex Wavelet Transformation (DT-CWT) [17] is used.

A. Dual-tree Complex Wavelet Transformation (DT-CWT)

Wavelets provide an optimal representation for many signals containing singularities (jumps and spikes), for example a piecewise smooth function. The wavelet representation is optimally sparse for such signals. A most basic discrete wavelet transformation is the Haar transformation [18]. However, the typical wavelet transform suffers from four fundamental shortcomings [17]: 1) the wavelet coefficients tend to oscillate positive and negative around singularities, which complicates wavelet-based processing, making singularity extraction and signal modeling very challenging; 2) they are shift variant, thus a small shift of the signal greatly perturbs the wavelet coefficient oscillation pattern around singularities; 3) the wavelet coefficients results in substantial aliasing, which can be cancelled by the inverse DWT only if the wavelet and scaling coefficients are not changed, and any wavelet coefficient processing will lead to artifacts in the reconstructed signal; 4) the lack of directional selectivity complicates modeling and processing of geometric image features like ridges and edges. To overcome these problems, the DT-CWT, first introduced by Kingsbury in 1998 [19], is proposed, which is approximately shift-invariant, directionally selective, and computationally efficient.

Also, it was pointed out in [15] that the dictionary redundancy improves the sparsity of representation. In the discrete Haar transform the size of the coefficient is the same as the size of image. In DT-CWT, the size of coefficients is two times the size of image (see Fig.1). So, following the above mentioned idea, DT-CWT will result to much sparser coefficient than the Haar transformation. Dual-tree of wavelet filters is used to obtain the real and imaginary parts of complex wavelet coefficients. So the coefficients can be divided into 2 groups, which correspond to the real and imaginary parts respectively.

In this paper, a Bayesian hierarchical structured method based on DT-CWT transformation (BH-DTCWT) is proposed.

B. Sparsity Enforcing prior distribution

The sparse property of the DT-CWT transformation coefficient can be enforced by using three kinds of distributions:

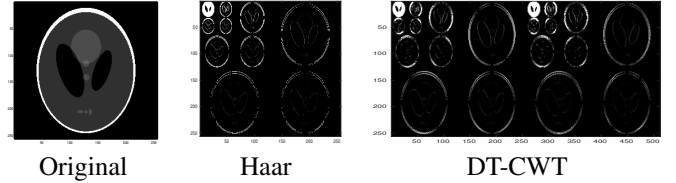


Fig. 1: The original phantom and the 3-level Haar transform coefficients and 3-level DT-CWT transform coefficients with scale adapted. The scales are adapted in order to be more visually clearly.

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:

$$St_g(f|\alpha, \beta) = \int \mathcal{N}(f|0, v) \mathcal{IG}(v|\alpha, \beta) dv. \quad (2)$$

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 [13].

C. Hierarchical Model of the proposed HB-DTCWT

The additive noise is assumed zero mean but non-stationary Gaussian distributed with unknown variances: $p(\epsilon|v_\epsilon) = \mathcal{N}(\epsilon|\mathbf{0}, \mathbf{V}_\epsilon)$, where $\mathbf{V}_\epsilon = \text{diag}[v_\epsilon]$ is a diagonal matrix and $v_\epsilon = [\dots, v_{\epsilon_i}, \dots]^T$. According to the linear forward model shown in Eq.(1), we obtain an expression for the likelihood: $p(\mathbf{g}|\mathbf{f}, v_\epsilon) = \mathcal{N}(\mathbf{g}|\mathbf{H}\mathbf{f}, \mathbf{V}_\epsilon)$.

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 DT-CWT coefficient.

As presented previously, we use the DT-CWT coefficient as the sparse representation of the piece-wise continuous image. \mathbf{z} is used to present the l -level DT-CWT coefficients of the image \mathbf{f} . Prior distribution of \mathbf{f} depending on \mathbf{z} is: $p(\mathbf{f}|\mathbf{z}, v_\xi) = \mathcal{N}(\mathbf{f}|\mathbf{D}\mathbf{z}, \mathbf{V}_\xi)$ where \mathbf{D} represents the inverse DT-CWT operator, and the additive noise ξ is considered to be zero mean Gaussian distributed and $\mathbf{V}_\xi = \text{diag}[v_\xi]$, $v_\xi = [\dots, v_{\xi_j}, \dots]^T$.

Vector $\mathbf{z} = [z_1, z_2, \dots, z_{2N}]^T$ is sparse. As mentioned above, the generalized Student-t distribution is used in order to enforce the sparsity. By using the Normal-Inverse Gamma

marginalization property given in Eq.(2), the prior distribution for \mathbf{z} can be realized by:

$$\begin{cases} p(\mathbf{z}|\mathbf{v}_z) = \mathcal{N}(\mathbf{z}|0, \mathbf{V}_z) \text{ where } \mathbf{V}_z = \text{diag}[\mathbf{v}_z], \\ p(\mathbf{v}_z|\alpha_{z_0}, \beta_{z_0}) = \prod_{k=1}^{2N} \mathcal{IG}(v_{z_k}|\alpha_{z_0}, \beta_{z_0}), \end{cases} \quad (3)$$

where the elements of vector $\mathbf{v}_z = [\dots, v_{z_k}, \dots]^T$ are supposed to be i.i.d.

On the other hand, when considering the variance of two noises: \mathbf{v}_ϵ and \mathbf{v}_ξ , 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: $p(\mathbf{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \prod_{i=1}^M \mathcal{IG}(v_{\epsilon_i}|\alpha_{\epsilon_0}, \beta_{\epsilon_0})$ and $p(\mathbf{v}_\xi|\alpha_{\xi_0}, \beta_{\xi_0}) = \prod_{j=1}^N \mathcal{IG}(v_{\xi_j}|\alpha_{\xi_0}, \beta_{\xi_0})$.

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

$$\begin{cases} p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon) \propto |\mathbf{V}_\epsilon|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{g} - \mathbf{H}\mathbf{f})^T \mathbf{V}_\epsilon^{-1}(\mathbf{g} - \mathbf{H}\mathbf{f})\right], \\ p(\mathbf{f}|\mathbf{z}, \mathbf{v}_\xi) \propto |\mathbf{V}_\xi|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{f} - \mathbf{D}\mathbf{z})^T \mathbf{V}_\xi^{-1}(\mathbf{f} - \mathbf{D}\mathbf{z})\right], \\ p(\mathbf{z}|\mathbf{v}_z) \propto |\mathbf{V}_z|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\mathbf{z}^T \mathbf{V}_z^{-1}\mathbf{z}\right], \\ p(\mathbf{v}_z|\alpha_{z_0}, \beta_{z_0}) \propto \prod_{k=1}^{2N} v_{z_k}^{-(\alpha_{z_0}+1)} \exp[-\beta_{z_0}v_{z_k}^{-1}], \\ p(\mathbf{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \propto \prod_{i=1}^M v_{\epsilon_i}^{-(\alpha_{\epsilon_0}+1)} \exp[-\beta_{\epsilon_0}v_{\epsilon_i}^{-1}], \\ p(\mathbf{v}_\xi|\alpha_{\xi_0}, \beta_{\xi_0}) \propto \prod_{j=1}^N v_{\xi_j}^{-(\alpha_{\xi_0}+1)} \exp[-\beta_{\xi_0}v_{\xi_j}^{-1}]. \end{cases} \quad (4)$$

The corresponding directed acyclic graph (DAG) of the proposed model is shown in Fig.(2).

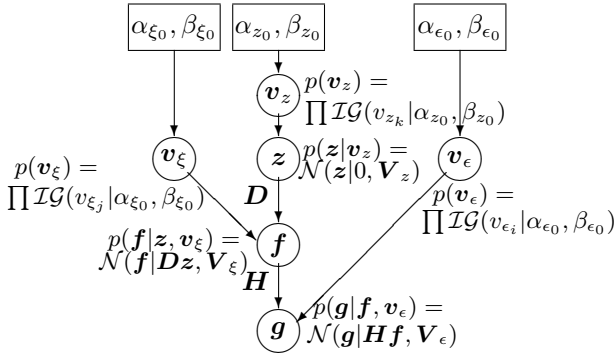


Fig. 2: DAG of proposed model.

D. Bayesian Inference

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

$$p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\xi, \mathbf{v}_z|\mathbf{g}) \propto p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon)p(\mathbf{f}|\mathbf{z}, \mathbf{v}_\xi)p(\mathbf{z}|\mathbf{v}_z) \cdot p(\mathbf{v}_\epsilon)p(\mathbf{v}_\xi)p(\mathbf{v}_z) \quad (5)$$

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). The first one can be computed either by MCMC methods [20], Variational Bayesian Approximation (VBA) [21] or any other approximation methods. In this paper, we use the JMAP method to estimate all the variables iteratively in this paper.

E. Bayesian Point Estimation

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

$$(\hat{\mathbf{f}}, \hat{\mathbf{z}}, \hat{\mathbf{v}}_\epsilon, \hat{\mathbf{v}}_\xi, \hat{\mathbf{v}}_z) = \arg \max_{\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\xi, \mathbf{v}_z} \{p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\xi, \mathbf{v}_z|\mathbf{g})\} \quad (6)$$

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

$$\begin{aligned} \text{iter} : \hat{\mathbf{f}}^{(n+1)} &= \hat{\mathbf{f}}^{(n)} - \hat{\gamma}_f^{(n)} \nabla \mathcal{J}(\hat{\mathbf{f}}^{(n)}); \\ \text{iter} : \hat{\mathbf{z}}^{(n+1)} &= \hat{\mathbf{z}}^{(n)} - \hat{\gamma}_z^{(n)} \nabla \mathcal{J}(\hat{\mathbf{z}}^{(n)}); \\ \hat{v}_{z_k} &= \left(\beta_{z_0} + \frac{1}{2}\hat{z}_k^2\right) / (\alpha_{z_0} + 3/2), \quad \forall k \in [1, N]; \\ \hat{v}_{\epsilon_i} &= \left(\beta_{\epsilon_0} + \frac{1}{2}\left(g_i - [\mathbf{H}\hat{\mathbf{f}}]_i\right)^2\right) / (\alpha_{\epsilon_0} + 3/2), \quad \forall i \in [1, M]; \\ \hat{v}_{\xi_j} &= \left(\beta_{\xi_0} + \frac{1}{2}\left(\hat{f}_j - [\mathbf{D}\hat{\mathbf{z}}]_j\right)^2\right) / (\alpha_{\xi_0} + 3/2), \quad \forall j \in [1, N], \end{aligned} \quad (7)$$

where

$$\begin{aligned} \mathcal{J}(\mathbf{f}) &= \frac{1}{2} \left\| \mathbf{V}_\epsilon^{-\frac{1}{2}}(\mathbf{g} - \mathbf{H}\mathbf{f}) \right\|_2^2 + \frac{1}{2} \left\| \mathbf{V}_\xi^{-\frac{1}{2}}(\mathbf{f} - \mathbf{D}\mathbf{z}) \right\|_2^2; \\ \mathcal{J}(\mathbf{z}) &= \frac{1}{2} \left\| \mathbf{V}_\xi^{-\frac{1}{2}}(\mathbf{f} - \mathbf{D}\mathbf{z}) \right\|_2^2 + \frac{1}{2} \left\| \mathbf{V}_z^{-\frac{1}{2}}\mathbf{z} \right\|_2^2; \\ \hat{\gamma}_f^{(n)} &= \left(\left\| \nabla \mathcal{J}(\hat{\mathbf{f}}^{(n)}) \right\|_2^2 \right) / \left(\left\| \hat{\mathbf{Y}}_\epsilon \mathbf{H} \nabla \mathcal{J}(\hat{\mathbf{f}}^{(n)}) \right\|_2^2 + \left\| \hat{\mathbf{Y}}_\xi \nabla \mathcal{J}(\hat{\mathbf{f}}^{(n)}) \right\|_2^2 \right); \\ \hat{\gamma}_z^{(n)} &= \left(\left\| \nabla \mathcal{J}(\hat{\mathbf{z}}^{(n)}) \right\|_2^2 \right) / \left(\left\| \hat{\mathbf{Y}}_\epsilon \mathbf{D} \nabla \mathcal{J}(\hat{\mathbf{z}}^{(n)}) \right\|_2^2 + \left\| \hat{\mathbf{Y}}_z \nabla \mathcal{J}(\hat{\mathbf{z}}^{(n)}) \right\|_2^2 \right), \end{aligned} \quad (8)$$

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 [22]. The initialization of the hyper parameters α s and β 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 ϵ_1 and ϵ_2 are small values close to zero.

III. SIMULATIONS RESULTS

The simulated Shepp-Logan image of size 256^2 , shown in Fig.(3)(a), is used as the original image, and the value of the pixels are normalized to $[0, 1]$. Parallel projections are used as the dataset. For each projection, 256^2 detectors are used, and an image of size 256×256 is obtained. The red block is the zone that we choose to show the results in Fig.(4). The uncontaminated sinogram by applying 128 projections is shown in Fig.(3)(b). The sinogram with 128 projections and SNR=20dB is shown in Fig.(3)(c). The projections are distributed uniformly from 0 to π . The signal-to-noise ratio of the dataset is defined as:

$$\text{SNR} = \|\mathbf{g}_0\|_2^2 / \|\epsilon\|_2^2 = \|\mathbf{g} - \epsilon\|_2^2 / \|\epsilon\|_2^2, \quad (9)$$

where \mathbf{g}_0 is the unbiased sinogram and \mathbf{g} is the contaminated sinogram with the additive noise represented by ϵ in Eq.(1). A lower level of SNR refers to the dataset with more critical additive noise ϵ .

The proposed BH-DTCWT method is used for the reconstructed and is compared with the conventional TV method and the HHBM method in our previous work [14]. In the

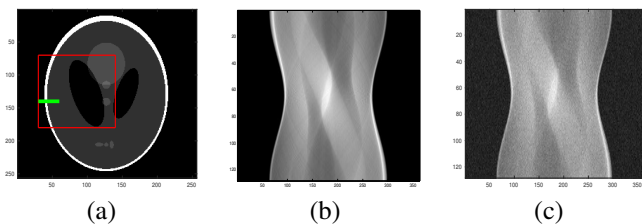


Fig. 3: Original Shepp Logan image of size 256^2 and its (b) sinogram with 128 projections and (c) the same data with additive noise with SNR=20dB.

TV method, the Split-Bregman algorithm [6] is used. The regularization parameter is chosen empirically, which should have a bigger value when the data has a lower SNR value, namely the level of noise is higher. In both HHBM and BH-DTCWT methods, the multi-level transformations are chosen to have 5 levels.

The reconstructed images are shown in Fig.(4), represented by zones of them corresponding to Fig.(3)(a). The red curve represents the profile of the image at the position of the blue line. The first column shows the results reconstructed from dataset with 128 projections, by using TV, HHBM and BH-DTCWT methods respectively. Visually, the TV method give a better result with smoother homogeneous zones. In the right column, images are reconstructed from dataset with 64 projections, and they show the convenient of the HHBM and BH-DTCWT methods: they obtain a better result than the TV method with less number of projections.

In order to show more clearly the effect at the edges of different methods, we compare a zone of the profiles in Fig.(4), which is at the position of the thick green line shown in Fig.(3)(a). Comparisons are demonstrated in Fig.(5). In the figure on top are the profiles obtained with 64 projection, and in the bottom are those from 32 projections. As we can see, the proposed method stays more robust at the edges while there are insufficient number of distributions.

Both HHBM and BH-DTCWT methods have a good performance when dealing with the reconstruction with less number of projections. Fig.(6)-Fig.(8) show the comparison of the normalized Mean Square Error (NMSE), represent by δ_f in the ordinates in Fig.(6)-Fig.(8), with the definition:

$$\text{NMSE} = \frac{\|\hat{f} - f\|_2^2}{\|f\|_2^2}. \quad (10)$$

The performance of TV is very good when there are sufficient number of projections, but when the number of projections become insufficient, the HHBM and BH-DTCWT methods are more robust than the TV method. The BH-DTCWT method is not strictly better than the HHBM method, but is better in most of the cases in the reconstruction of Shepp Logan phantom.

IV. CONCLUSION

We proposed a Bayesian method based on a hierarchical prior model in which the Dual-Tree Complex Wavelet Transformation is used. Generalized Student-t distribution is used

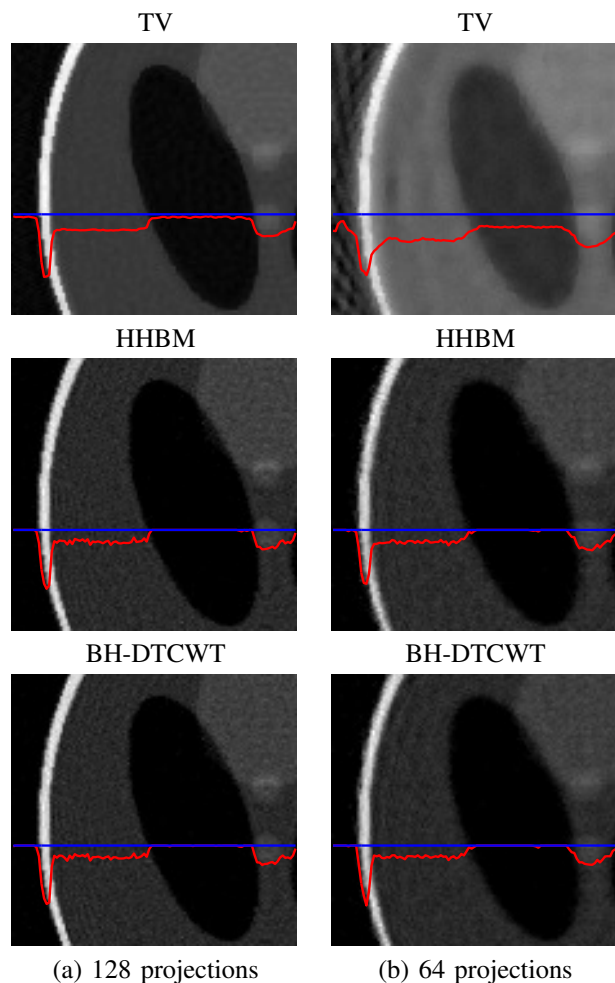


Fig. 4: Reconstructed Shepp-Logan image by using different reconstruction methods. Top: the TV method, middle: the HHBM method and bottom the BH-DTCWT method with a dataset of (a) 128 projections and (b) 64 projections and a high SNR=40dB.

to enforce the sparsity of the transformation coefficient. Comparing with the state-of-the-art methods, the proposed method remain robust when dealing with reconstruction problems with less number of projections, namely insufficient dataset, in the CT reconstruction of the Shepp Logan image. Another advantage is that by using the Bayesian method we have no need to choose a value for the parameters, such as the regularization parameter in TV, and they are automatically estimated in the simulations. In the future work we will complete the comparisons by using different types of phantoms, for example phantoms with periodical textures, in order to analyse the influence of the transformation type.

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.

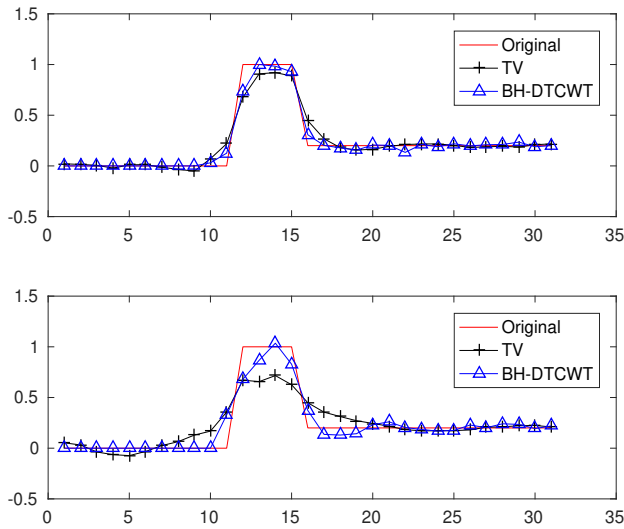


Fig. 5: Comparison of a zone of profile with TV method, from dataset of 64 projections on the top, and 32 projections on the bottom.

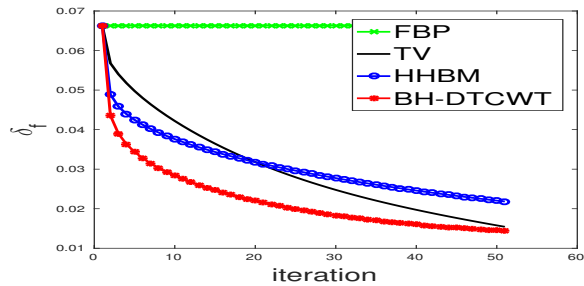


Fig. 6: NMSE of reconstructed image along iterations from sinogram with 128 projections and a high SNR=40dB.

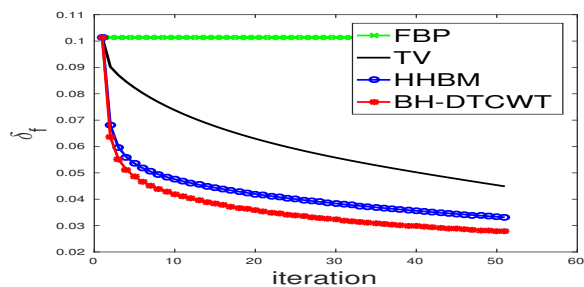


Fig. 7: NMSE of reconstructed image along iterations from sinogram with 64 projections and a high SNR=40dB.

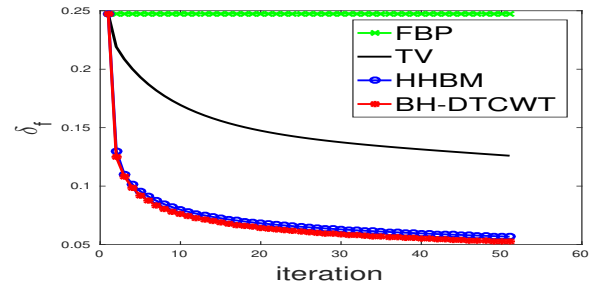


Fig. 8: NMSE of reconstructed image along iterations from sinogram with 32 projections and a high SNR=40dB.

[3] 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.

[4] Gene H Golub, Per Christian Hansen, and Dianne P O'Leary, "Tikhonov regularization and total least squares," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 1, pp. 185–194, 1999.

[5] 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.

[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] Ignace Loris, Guust Nolet, Ingrid Daubechies, and FA Dahlen, "Tomographic inversion using ℓ_1 -norm regularization of wavelet coefficients," *Geophysical Journal International*, vol. 170, no. 1, pp. 359–370, 2007.

[9] 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.

[10] 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.

[11] Ali Mohammad-Djafari, "A full Bayesian approach for inverse problems," in *Maximum entropy and Bayesian methods*, pp. 135–144. Springer, 1996.

[12] Mark E Glickman and David A van Dyk, "Basic Bayesian methods," *Topics in Biostatistics*, pp. 319–338, 2007.

[13] Mircea Dumitru, *A Bayesian approach for periodic components estimation for chronobiological signals*, Ph.D. thesis, Paris Saclay, 2016.

[14] 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.

[15] Qiong Xu, Hengyong Yu, Xuanqin Mou, Lei Zhang, Jiang Hsieh, and Ge Wang, "Low-dose X-ray CT reconstruction via dictionary learning," *IEEE Transactions on Medical Imaging*, vol. 31, no. 9, pp. 1682–1697, 2012.

[16] Farrokh Rashid-Farrokhi, KJ Ray Liu, Carlos Alberto Berenstein, and David Walnut, "Wavelet-based multiresolution local tomography," *IEEE Transactions on Image Processing*, vol. 6, no. 10, pp. 1412–1430, 1997.

[17] Ivan W Selesnick, Richard G Baraniuk, and Nick C Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, 2005.

[18] 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.

[19] Nick G Kingsbury, "The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters," in *IEEE Digital Signal Processing Workshop*. Citeseer, 1998, vol. 86, pp. 120–131.

[20] Walter R Gilks, "Markov chain monte carlo," *Encyclopedia of Biostatistics*, 2005.

[21] 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.

[22] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge university press, 2004.