



**HAL**  
open science

# Bayesian 3D X-ray Computed Tomography with a Hierarchical Prior model for Sparsity in Haar Transform domain

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

## ► To cite this version:

Li Wang, Ali Mohammad-Djafari, Nicolas Gac, Mircea Dumitru. Bayesian 3D X-ray Computed Tomography with a Hierarchical Prior model for Sparsity in Haar Transform domain. Entropy, 2018, Special Issue "Probabilistic Methods for Inverse Problems", <10.3390/e20120977>. <hal-01950706>

**HAL Id: hal-01950706**

**<https://hal.science/hal-01950706v1>**

Submitted on 19 Jan 2020

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.



HAL Authorization

Article

# Bayesian 3D X-ray Computed Tomography with a Hierarchical Prior model for Sparsity in Haar Transform domain

Li Wang <sup>1</sup>, Ali Mohammad-Djafari <sup>1</sup>, Nicolas Gac <sup>1</sup> and Mircea Dumitru <sup>1</sup>

<sup>1</sup> Laboratoire des signaux et système, Centralesupelec, CNRS, 3 Rue Joliot Curie, 91192 Gif sur Yvette; firstname.surname@l2s.centralesupelec.fr

Academic Editor: name

Version December 3, 2018 submitted to Entropy; Typeset by L<sup>A</sup>T<sub>E</sub>X using class file mdpi.cls

**Abstract:** In this paper, a hierarchical prior model based on the Haar transformation and an appropriate Bayesian computational method for X-ray CT reconstruction are presented. Given the piece-wise continuous property of the object, a multilevel Haar transformation is used to associate a sparse representation for the object. The sparse structure is enforced via a generalized Student-t distribution ( $\mathcal{St}_g$ ), expressed as the marginal of a Normal-Inverse Gamma distribution. The proposed model and corresponding algorithm are designed to adapt to specific 3D data sizes and to be used in both medical and industrial Non Destructive Testing (NDT) applications. In the proposed Bayesian method, a hierarchical structured prior model is proposed, and the parameters are iteratively estimated. The initialization of the iterative algorithm uses the parameters of the prior distributions. A novel strategy for the initialization is presented and proved experimentally. We compare the proposed method with two state-of-the-art approaches, showing that our method has better reconstruction performance when fewer projections are considered and when projections are acquired from limited angles.

**Keywords:** X-ray Computed Tomography; inverse problem; sparsity; hierarchical structure; Generalized Student-t distribution; Haar transformation.

## 1. Introduction

Computed Tomography (CT) has been developed and widely used in medical diagnosis [1] and industrial Non Destructive Testing (NDT) [2] in recent decades. In CT, objects are observed using different techniques, for example X-rays [3], ultrasound [4], microwaves [5], infra-red [6]. X-ray CT employs the absorption of X-rays by the organs in a body or by the materials in industrial components to reconstruct the internal structure of the imaged object. When performing X-ray CT, a set of X-ray images of the measured parts are acquired. The intensity measured by the X-ray images corresponds to the intensity of the radiation passing through and attenuated by the object. CT reconstruction is typically treated as an inverse problem.

The conventional analytical techniques for CT reconstruction are based on the Radon Transform [7]. Filtered Back-Projection (FBP) [8] is the most frequently used analytical method in practical applications. FBP performs well when reconstructing from sufficient data with a high signal-to-noise ratio (SNR), but it suffers from artifacts when reconstructing from insufficient data or with noise.

Owing to considerations regarding patients' health in medical CT, and in order to reduce acquisition time in industrial applications, reconstruction with insufficient datasets is increasingly attracting the attention of researchers. Reconstruction from fewer projections is an ill-posed inverse problem [9,10]. In this case, conventional analytical reconstruction methods provide unsatisfactory results and iterative methods can be used to improve the reconstruction performance. The Algebraic

Reconstruction Technique (ART) [11,12], the Simultaneous algebraic reconstruction technique (SART) [13] and the Simultaneous Iterative Reconstruction Technique (SIRT) [14,15] are some of the iterative methods proposed initially. These methods consider the discretized forward system model:  $\mathbf{g} = \mathbf{H}\mathbf{f}$ , where  $\mathbf{f} \in \mathbb{R}^{N \times 1}$  represents the object,  $\mathbf{g} \in \mathbb{R}^{M \times 1}$  represents the observed dataset and matrix  $\mathbf{H} \in \mathbb{R}^{M \times N}$  is the linear projection operator, mainly based on the geometry of acquisition (e.g. parallel beam, cone beam, etc) [16–18]. Typically, the system of equations is under-determined, i.e.  $N > M$ . In this context, regularization methods are frequently used and the forward system is modeled as:

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

where  $\boldsymbol{\epsilon} \in \mathbb{R}^{M \times 1}$  represents the additive noise applied to the projection system. The regularization methods estimate the unknowns by minimizing a penalty criterion, which generally consists of two terms:

$$J(\mathbf{f}) = Q(\mathbf{g}, \mathbf{f}) + \lambda R(\mathbf{f}). \quad (2)$$

29 The loss function  $Q(\mathbf{g}, \mathbf{f})$  describes discrepancies in the observed data, such as the quadratic ( $L_2$ )  
 30 loss  $Q(\mathbf{g}, \mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2$  or  $L_q$  loss  $Q(\mathbf{g}, \mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_q^q$  with  $1 \leq q < 2$ . Other expressions  
 31 such as the Huber function are also reported. The regularization term  $R(\mathbf{f})$  is a penalty on the  
 32 complement criterion of  $\mathbf{f}$ , such as restriction for smoothness  $\|\Phi(\mathbf{f})\|_2^2$  or for sparsity  $\|\Phi(\mathbf{f})\|_1$ , where  
 33  $\Phi(\mathbf{f})$  represents a linear function of  $\mathbf{f}$ . The parameter  $\lambda$  is known as the regularization parameter,  
 34 which controls the trade-off between the forward model discrepancy and the penalty term.

35 By choosing different regularization functions  $R(\mathbf{f})$ , different regularization methods can be  
 36 implemented.  $R(\mathbf{f}) = 0$  refers to the Least-Squares (LS) method [19], with the drawback that the  
 37 reconstruction is sensitive to the noise due to the ill-posedness of the problem and the ill-conditioning  
 38 of the operator  $\mathbf{H}$ . Quadratic Regularization (QR), also known as the Tikhonov method [20], is given  
 39 by  $R(\mathbf{f}) = \|\Phi(\mathbf{f})\|_2^2$ , where the linear operator  $\Phi(\cdot)$  is the derivation operator in most cases. The  
 40 well-known Total Variation (TV) method [21–23] is defined by  $R(\mathbf{f}) = \|\mathbf{D}_{TV}\mathbf{f}\|_{TV}$  where  $\mathbf{D}_{TV}$  is the  
 41 gradient operator.  $\mathbf{D}_{TV}$  is equal to  $\|\mathbf{D}_x\mathbf{f}\|_1 + \|\mathbf{D}_y\mathbf{f}\|_1 + \|\mathbf{D}_z\mathbf{f}\|_1$  for a 3D object in an anisotropic form,  
 42 where  $\mathbf{D}_x$ ,  $\mathbf{D}_y$  and  $\mathbf{D}_z$  are respectively the gradient operators in the  $x$ ,  $y$  and  $z$  directions. The  $L_1$  norm  
 43 is used in TV for sparse estimations, which enforces the sparsity of  $\mathbf{D}_{TV}\mathbf{f}$ . The appearance of the  
 44 non-differentiable  $L_1$  term leads to difficulties for the implementation of optimization algorithms.  
 45 Many optimization algorithms have been proposed to solve this  $L_1$  norm optimization problem,  
 46 for example the primal-dual method [24], the Split Bregman method [22], etc. In regularization  
 47 optimization, due to the large projection data size and the great number of voxels, the explicit  
 48 expression of the solution can not be used directly because of the impossibility of inverting the large  
 49 size matrix such as  $(\mathbf{H}^T\mathbf{H} + \lambda\mathbf{D}^T\mathbf{D})^{-1}$ . Hence, optimization algorithms such as gradient descent or  
 50 conjugate gradient are often used.

More general regularization methods have been developed based on the constrained and dual-variable regularization method:

$$J(\mathbf{f}, \mathbf{z}) = Q_1(\mathbf{g}, \mathbf{f}) + \eta Q_2(\mathbf{f}, \mathbf{z}) + \lambda R(\mathbf{z}), \quad (3)$$

51 which corresponds to the maximum a posterior optimization of a hierarchical structured model where  
 52 both  $\mathbf{f}$  and  $\mathbf{z}$  are unknown variables. In such a model, the penalty regularization term is set on  $\mathbf{z}$ ,  
 53 which is associated with  $\mathbf{f}$  via a linear transformation. The loss functions  $Q_1(\mathbf{g}, \mathbf{f})$  and  $Q_2(\mathbf{f}, \mathbf{z})$  are  
 54 for example quadratic ( $L_2$ ), i.e.  $Q_1(\mathbf{g}, \mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2$  and  $Q_2(\mathbf{f}, \mathbf{z}) = \|\mathbf{f} - \mathbf{D}\mathbf{z}\|_2^2$  where  $\mathbf{D}$  is a linear  
 55 transform operator such as a wavelet transformation.

56 Among the methods treating this type of regularization problem, we mention here the  
 57 Alternating Direction Method of Multipliers (ADMM) [25]. It minimizes  $\Phi(\mathbf{f}) + \Psi(\mathbf{z})$  subject to  
 58  $\mathbf{A}\mathbf{f} + \mathbf{B}\mathbf{z} = \mathbf{C}$ , and it covers a large number of estimation forms. One example is when  $\Phi(\mathbf{f}) =$

59  $\|g - Hf\|^2$ ,  $\Psi(z) = R(z)$ ,  $A = I$ ,  $B = -D$  and  $C = \mathbf{0}$ , and refers to the above-mentioned bi-variable  
60 regularization method corresponding to Eq. (3).

61 In the above-mentioned regularization methods, there is always a regularization parameter  $\lambda$   
62 to be fixed. Sometimes the regularization term consists of more than one parts and each of them  
63 are weighted by a regularization parameter, for example the Elastic-net regularizer [26]. In these  
64 cases, two or even more regularization parameters are needed to be fixed. Cross Validation (CV) and  
65 the L-curve method [20,27,28] are conventional methods used to determine suitable values for these  
66 parameters. However, this work must be repeated for different simulated datasets and is therefore  
67 very costly. Statistical methods, therefore, have been developed and used to solve this problem.

68 From the probabilistic point of view, a Gaussian model for the additive noise in the forward  
69 model, Eq. (1), leads to the quadratic expression  $\|g - Hf\|^2$  in the corresponding regularization  
70 criterion. But in some types of tomography, for example Positron Emission Tomography (PET)  
71 or X-ray tomography with very low number of phantoms, the noise is modeled by a Poisson  
72 distribution. In order to account for a more precise modeling of the noise and the other variables  
73 and parameters, statistical methods are used [29]. The Maximum Likelihood (ML) methods [30]  
74 and different estimation algorithms such as the Expectation Maximization (EM) algorithms [31],  
75 the stochastic EM (SEM) [32], the Ordered subsets-EM (OS-EM) [33] are commonly used in PET-CT  
76 reconstruction problems.

Another widely used type of probabilistic method for PET or X-ray CT reconstruction is the Bayesian inference [34–37]. The prior knowledge is translated by the prior probability model and is used to obtain the expression of the posterior distribution. The basic Bayesian formula is:

$$p(f|g, \theta) = \frac{p(g|f, \theta_1)p(f|\theta_2)}{p(g|\theta)}, \text{ with } p(g|\theta) = \int p(g|f, \theta_1)p(f|\theta_2) df, \quad (4)$$

where  $p(g|f, \theta_1)$  is the likelihood,  $p(f|\theta_2)$  is the prior distribution,  $p(f|g, \theta)$  is the posterior distribution,  $\theta = (\theta_1, \theta_2)$  are the parameters of these different distributions and  $p(g|\theta)$  is the evidence of the parameters in the data  $g$ . By using Maximum A Posterior (MAP) estimator  $\hat{f} = \arg \max_f \{p(f|g, \theta)\} = \arg \min_f \{-\ln p(f|g, \theta)\}$ , links between the Bayesian method and almost all the regularization methods can be illustrated. A Gaussian prior for  $p(f)$  in Eq. (4) leads to the Quadratic ( $L_2$ ) Regularization method, while a Laplacian prior in Eq. (4) leads to the  $L_1$  (LASSO or TV) regularization method. The regularization parameter can be related to  $\theta_1$  and  $\theta_2$ . One advantage of the Bayesian method is having some explanation for the regularization parameter via its link with  $\theta_1$  and  $\theta_2$ . For example when  $p(g|f, \theta_1)$  and  $p(f|\theta_2)$  are Gaussian with  $\theta_1$  and  $\theta_2$  respectively the variances of the noise and the variance of the prior, then the regularization parameter is  $\lambda = \theta_1/\theta_2$ . Another advantage of the Bayesian method is that these parameters can also be estimated to achieve unsupervised or semi-supervised methods. This is achieved by obtaining the expression of the joint posterior probability law:

$$p(f, \theta|g) = \frac{p(g|f, \theta_1)p(f|\theta_2)p(\theta)}{p(g)}, \quad (5)$$

where  $p(\theta)$  is an appropriate prior on  $\theta$ . For a hierarchical structured model where a hidden variable  $z$  appears in the prior model, we have:

$$p(f, z, \theta|g) = \frac{p(g|f, \theta_1)p(f|z, \theta_2)p(z|\theta_3)p(\theta)}{p(g)}, \quad (6)$$

77 where  $\theta = [\theta_1, \theta_2, \theta_3]$ .

78 With the posterior distribution obtained from an unsupervised Bayesian inference as in Eq. (5),  
79 we distinguish three estimation methods. The first method consists of integrating out  $\theta$  from  $p(f, \theta|g)$   
80 to obtain  $p(f|g)$  and then use  $p(f|g)$  to infer on  $f$ . The second approach is firstly to marginalize  
81  $p(f, \theta|g)$  with respect to  $f$  to obtain  $p(\theta|g) = \int p(f, \theta|g) df$  and estimate  $\hat{\theta} = \arg \max_{\theta} \{p(\theta|g)\}$ ,

82 then use  $\hat{\theta}$  as it was known. Unfortunately, these approaches do not often give explicit expressions  
83 for  $p(f|g)$  or  $p(\theta|g)$ . The third and easiest algorithm to implement is the joint optimization, which  
84 estimates variable  $f$  and parameter  $\theta$  iteratively and alternately. Bayesian point estimators such  
85 as joint maximum a posteriori (JMAP) [38] and Posterior Mean (PM) [39] via Variational Bayesian  
86 Approximation (VBA) methods [40–42] are often used.

87 In order to distinguish the details of a reconstructed object, a high-resolution image is expected.  
88 In industrial applications, especially for the NDT of a large-size object, the size of the projection (1000  
89 images of  $1000^2$  pixels) and the number of the voxels ( $1000^3$  voxels) become critical, and so does the  
90 projection and back-projection operators in CT. It is necessary to account for some computational  
91 aspects, for example the GPU processor [43,44].

92 In our previous work [45], we proposed to use the Bayesian method via a synthesis model, in  
93 which the multilevel Haar transformation coefficient  $z$  of the image is first estimated, and then the  
94 final image reconstruction result is obtained from post processing:  $\hat{f} = D\hat{z}$ . In this case, when using  
95 a Laplacian prior model and the MAP estimator, the problem becomes equivalent to the optimization  
96 of  $J(z) = \|g - HDz\|^2 + \lambda \|z\|_1$  which is a typical  $L_1$  regularization method. The particularity of our  
97 work was to use a generalized Student-t ( $St_g$ ) prior model [46] in place of the Laplacian model.

98 In this paper, we present a Hierarchical Haar transform-based Bayesian method (HHBM), first  
99 proposed in [47], in which the object to be reconstructed,  $f$ , is related to the Haar transformation  
100 coefficient  $z$  by  $f = Dz + \zeta$  where  $\zeta$  represents the modelization uncertainties.  $f$  and  $z$  are  
101 estimated simultaneously. Wavelets provide an optimal representation for a piecewise continuous  
102 function consisting of homogeneous blocs separated by jump discontinuities (the contours), as the  
103 wavelet representation is sparse for such signals. Transformations used are, for example, the Haar  
104 transformation [48], the curvelet transformation (CVT) [49], the contourlet transformation (CT) [50],  
105 Dual-Tree Complex Wavelet Transform (DT-CWT) [51], etc. As long as the object under consideration,  
106  $f$ , is piecewise continuous or constant, the Haar transform is appropriate, with the advantage that: 1)  
107 the transform coefficients are sparse, 2) the transformation operator is orthogonal so that the inverse  
108 operator and the transpose are identical, and 3) the computation of this transformation consists of  
109 only additions and subtractions and the cost of computation is only  $O(\sqrt{N})$  where  $N$  is the size of the  
110 object  $f$ .

111 The sparsity of the transformation coefficient is generally defined by three classes of  
112 distributions: the Generalized Gaussian distributions [52], the mixture distributions [53] and the  
113 heavy-tailed distributions [54]. In this paper, we use a generalization of Student-t distribution ( $St_g$ ),  
114 which belongs to the heavy-tailed family and has many advantages when enforcing the sparsity of  
115 variables [46].

116 In this paper, we extend extensively the previous work by: 1) adapting the forward model and  
117 prior models to the 3D case, which is more appropriate for real 3D large data size applications; 2)  
118 comparing the RMSE of the phantom reconstructed by HHBM method with those by conventional  
119 QR and TV methods, we show the advantages of the semi-supervised property of the HHBM method  
120 and that the HHBM method outperforms the TV method when insufficient data is estimated; 3)  
121 proposing new ideas for fixing the hyper-parameters in the proposed model; and 4) evaluating the  
122 performance of the proposed method in the situations when the number or the angle distribution of  
123 the projections are limited.

124 The rest of this paper is organized as follows: section 2 presents the proposed  
125 hierarchically-structured Bayesian method; section 3 gives the details of the implementations and  
126 the choice of hyperparameters, as well as the simulation results; some points on the initialization of  
127 hyper-parameters are discussed in section 4; conclusions are drawn and prospective future research  
128 is presented in section 5.

## 129 2. The Semi-supervised Hierarchical Model

130 The Hierarchical Haar-based Bayesian Method (HHBM) is presented, in which the object  $f$   
 131 and its multilevel Haar transform coefficient  $z$  are considered jointly. A sparse enforcing prior  
 132 is defined on  $z$ . The wavelet transformation has been used for the reconstruction of tomography  
 133 images in some state-of-the-art works [55–58], using both regularization and Bayesian methods. In  
 134 these state-of-the-art methods, the phantom  $f$  is obtained by a post-processing from reconstructed  
 135 coefficient  $z$ . In this paper, the phantom  $f$  and the coefficient  $z$  are simultaneously estimated.

### 136 2.1. Forward system model and likelihood

In the proposed method, the forward model introduced in Eq. (1) is considered. Generally,  
 the noise in tomography is modeled by Poisson distribution [59], but, in X-ray CT, the Gaussian  
 approximation is often used. We adopt the Gaussian approximation and propose to use a zero mean  
 and non-stationary model where the variance is considered to be unknown, belonging to an Inverse  
 Gamma distribution given that this distribution provides a good adaptation of the positivity property  
 of the variances  $v_{\epsilon_i}$ :

$$p(\epsilon|\mathbf{v}_\epsilon) = \mathcal{N}(\epsilon|\mathbf{0}, \mathbf{V}_\epsilon), \mathbf{V}_\epsilon = \text{diag}[v_\epsilon], \text{ where } v_\epsilon = [v_{\epsilon_1}, \dots, v_{\epsilon_M}]' \in \mathbb{R}^{M \times 1} \quad (7)$$

$$p(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \prod_{i=1}^M \mathcal{IG}(v_{\epsilon_i}|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \quad (8)$$

137 The vector  $v_\epsilon$  is considered in order to account for the difference of sensitivity to noise for each  
 138 detector in each projection direction.

According to the forward model of the linear system, Eq. (1), and the prior model of the noise,  
 Eq. (7), the likelihood of this model system is:

$$p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon) = \mathcal{N}(\mathbf{g}|\mathbf{H}\mathbf{f}, \mathbf{V}_\epsilon). \quad (9)$$

139 In Bayesian inference, the likelihood is combined with the prior distributions to determine the  
 140 posterior distribution.

### 141 2.2. Hierarchical prior model and prior distributions

142 Typically the objects considered in medical and industrial X-ray CT are piecewise continuous.  
 143 In this paper, a hierarchical prior model is used to define the piecewise continuous property. In  
 144 this hierarchical prior model, a sparsity enforcing model is defined for the wavelet transformation  
 145 coefficients of the image. A large number of methods accounting for sparse structure of the solution  
 146 have been proposed in the literature. Among them, the L1 regularization method is most frequently  
 147 used, which minimizes the criterion  $J(f) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \lambda \|\mathbf{\Phi}\mathbf{f}\|_1$  where  $\mathbf{\Phi}$  is a linear operator,  
 148 for example the gradient in the TV method. Another class of methods, known as "synthesis" [60],  
 149 minimizes  $J(z) = \|\mathbf{g} - \mathbf{H}\mathbf{D}\mathbf{z}\|_2^2 + \lambda \|\mathbf{z}\|_1$  where  $\mathbf{f} = \mathbf{D}\mathbf{z}$ , and  $\mathbf{z} = \mathbf{D}^{-1}\mathbf{f}$  is for example a wavelet  
 150 transform.

In this paper, we propose to use the multilevel Haar transformation as the sparse dictionary. The  
 transformation is modeled using a discretized forward model:

$$\mathbf{f} = \mathbf{D}\mathbf{z} + \boldsymbol{\zeta}, \quad (10)$$

151 where  $\mathbf{D} \in \mathbb{R}^{N \times N}$  represents the inverse multilevel Haar transformation and  $\boldsymbol{\zeta} \in \mathbb{R}^{N \times 1}$  represents the  
 152 uncertainties of the transformation, which is introduced to relax the exact relation of the transform  
 153 operator  $\mathbf{D}$ .  $\boldsymbol{\zeta}$  is supposed to be sparse. Unlike the gradient operator used in the TV method,  
 154 the multilevel Haar transformation is orthogonal, i.e.  $\mathbf{D}^{-1} = \mathbf{D}^T$ . This property provides certain

155 advantages during optimization, especially for the big data size problems, as the inversion and  
156 transpose of the operator are identical and can be replaced by each other for different types of  $\mathbf{D}$ .

$\boldsymbol{\zeta}$  is modeled by a Gaussian distribution,  $p(\boldsymbol{\zeta}) = \mathcal{N}(\boldsymbol{\zeta}|0, \mathbf{V}_\zeta)$ , where  $\mathbf{V}_\zeta = \text{diag}[v_\zeta]$  and  $v_\zeta = [v_{\zeta_1}, \dots, v_{\zeta_N}]'$ .  $v_\zeta$  is considered an unknown variance. It is modeled in order to realize a semi-supervised system where the variance is estimated. Here  $v_\zeta$  is modeled by an Inverse Gamma distribution, with the same consideration as the model of  $v_\epsilon$ . The Gaussian model with an Inverse Gamma distributed variance,  $p(v_\zeta|\alpha_{\zeta_0}, \beta_{\zeta_0}) = \prod_{j=1}^N \mathcal{IG}(v_{\zeta_j}|\alpha_{\zeta_0}, \beta_{\zeta_0})$ , leads to a generalized Student-t ( $St_g$ ) distribution [46]. Consequently, a  $St_g$  distribution is derived for  $\boldsymbol{\zeta}$ , and the sparse property of  $\boldsymbol{\zeta}$  can be guaranteed. From Eq. (10), the conditional distribution  $p(\mathbf{f}|\mathbf{z}, v_\zeta)$  is derived:

$$p(\mathbf{f}|\mathbf{z}, v_\zeta) = \mathcal{N}(\mathbf{f}|\mathbf{D}\mathbf{z}, \mathbf{V}_\zeta), \quad (11)$$

with

$$p(v_\zeta|\alpha_{\zeta_0}, \beta_{\zeta_0}) = \prod_j^N \mathcal{IG}(v_{\zeta_j}|\alpha_{\zeta_0}, \beta_{\zeta_0}). \quad (12)$$

157 For practical applications where these parameters are not known or difficult to obtain, we use a  
158 semi-supervised method in which the variances of noises,  $v_\epsilon$  and  $v_\zeta$ , are unknown. In HHBM, Inverse  
159 Gamma distribution is used to model  $v_\epsilon$  and  $v_\zeta$ ,  $p(v_\epsilon) = \mathcal{IG}(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0})$  and  $p(v_\zeta) = \mathcal{IG}(v_\zeta|\alpha_{\zeta_0}, \beta_{\zeta_0})$ .  
160 Consequently, both  $\boldsymbol{\zeta}$  and  $\epsilon$  are modeled by a  $St_g$  distribution.

Vector  $\mathbf{z} = [z_1, \dots, z_N]'$  represents the multilevel Haar transformation coefficient of piece-wise continuous  $f$ . As mentioned above,  $\mathbf{z}$  is sparse. In this paper, the generalized Student-t distribution ( $St_g$ ) [46] is used to enforce the sparsity structure of  $\mathbf{z}$ . The  $St_g$  distribution can be expressed as the marginal of a bivariate Normal-Inverse Gamma distribution:

$$St_g(\mathbf{z}|\alpha, \beta) = \int \mathcal{N}(\mathbf{z}|0, v) \mathcal{IG}(v|\alpha, \beta) dv. \quad (13)$$

161 Thanks to the fact that Normal and Inverse Gamma are conjugate distributions, the use of the  $St_g$  via  
162 Eq. (13) simplifies the computations when using the Bayesian point estimators such as the Posterior  
163 Mean via the Variational Bayesian Approximation (VBA) method [42].

164 From Eq. (13), the  $St_g$  prior distribution modeling  $\mathbf{z}$  is expressed as the following model:

$$p(\mathbf{z}|v_z) = \mathcal{N}(\mathbf{z}|0, \mathbf{V}_z), \quad \text{where } \mathbf{V}_z = \text{diag}[v_z], \quad v_z = [v_{z_1}, v_{z_2}, \dots, v_{z_N}] \quad (14)$$

$$p(v_z|\alpha_{z_0}, \beta_{z_0}) = \prod_j^N \mathcal{IG}(v_{z_j}|\alpha_{z_0}, \beta_{z_0}), \quad (15)$$

165 where  $v_{z_j}, \forall j = 1 : N$  are i.i.d. distributed. The difference between the standard  $St$  distribution,  
166  $St(\mathbf{z}|v) = \int \mathcal{N}(\mathbf{z}|0, v) \mathcal{IG}(v|\frac{v}{2}, \frac{v}{2}) dv$ , and the generalized  $St_g$ , given in Equation (13), is that  $St(\mathbf{z}|v)$   
167 is governed by one parameter  $v$  but  $St_g(\mathbf{z}|\alpha, \beta)$  is governed by two parameters  $(\alpha, \beta)$ . With these two  
168 parameters, the  $St_g$  does not only enforce the sparsity of the variable, but also control the sparsity  
169 rate [46]. By changing the values of the two hyper-parameters  $\alpha_{z_0}$  and  $\beta_{z_0}$ , we can obtain either a  
170 heavy-tailed distribution with a narrow peak, or a distribution approaching a Gaussian distribution.

## 171 2.3. The HHBM method

The prior models of the proposed Bayesian method based on the forward model of Eq. (1) and the prior model of Eq. (10) are:

$$p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon) \propto |\mathbf{V}_\epsilon|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\mathbf{g} - \mathbf{H}\mathbf{f})^T \mathbf{V}_\epsilon^{-1} (\mathbf{g} - \mathbf{H}\mathbf{f}) \right], \quad (16)$$

$$p(\mathbf{f}|\mathbf{z}, \mathbf{v}_\xi) \propto |\mathbf{V}_\xi|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\mathbf{f} - \mathbf{D}\mathbf{z})^T \mathbf{V}_\xi^{-1} (\mathbf{f} - \mathbf{D}\mathbf{z}) \right], \quad (17)$$

$$p(\mathbf{z}|\mathbf{v}_z) \propto |\mathbf{V}_z|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \mathbf{z}^T \mathbf{V}_z^{-1} \mathbf{z} \right], \quad (18)$$

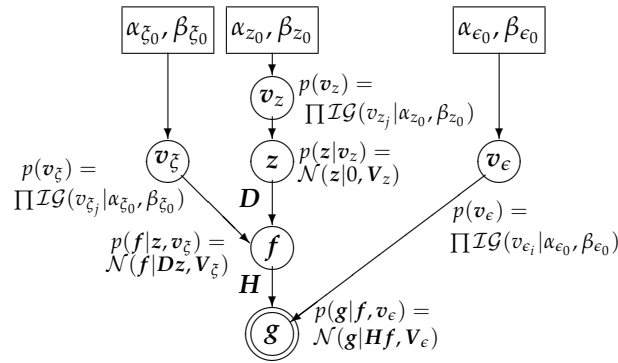
$$p(\mathbf{v}_z|\alpha_{z_0}, \beta_{z_0}) \propto \prod_j^N v_{z_j}^{-(\alpha_{z_0}+1)} \exp \left[ -\beta_{z_0} v_{z_j}^{-1} \right], \quad (19)$$

$$p(\mathbf{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \propto \prod_i^M v_{\epsilon_i}^{-(\alpha_{\epsilon_0}+1)} \exp \left[ -\beta_{\epsilon_0} v_{\epsilon_i}^{-1} \right], \quad (20)$$

$$p(\mathbf{v}_\xi|\alpha_{\xi_0}, \beta_{\xi_0}) \propto \prod_j^N v_{\xi_j}^{-(\alpha_{\xi_0}+1)} \exp \left[ -\beta_{\xi_0} v_{\xi_j}^{-1} \right]. \quad (21)$$

172

173 Fig.1 shows the generative graph of the proposed model in which the hyperparameters in the  
174 rectangles need to be initialized:



**Figure 1.** Generative graph of the proposed model illustrating all the unknowns (circles), hyperparameters (boxes) and data (double circles).

175

Via the Bayes rule, Eq. (6), the joint posterior distribution of all the unknowns given in the data is derived:

$$\begin{aligned} p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\xi, \mathbf{v}_z | \mathbf{g}) &= \frac{p(\mathbf{g}, \mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\xi, \mathbf{v}_z)}{p(\mathbf{g})} \\ &= \frac{p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon) p(\mathbf{f}|\mathbf{z}, \mathbf{v}_\xi) p(\mathbf{z}|\mathbf{v}_z) p(\mathbf{v}_z) p(\mathbf{v}_\epsilon) p(\mathbf{v}_\xi)}{p(\mathbf{g})} \\ &\propto p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon) p(\mathbf{f}|\mathbf{z}, \mathbf{v}_\xi) p(\mathbf{z}|\mathbf{v}_z) p(\mathbf{v}_z) p(\mathbf{v}_\epsilon) p(\mathbf{v}_\xi). \end{aligned} \quad (22)$$

176

177

178

Bayesian point estimators are often used for estimation via the a posteriori distribution. In this paper, we focus on the JMAP estimation, given that in the case of the large data size of the 3D object, the computational costs for the VBA algorithm is too expensive.

179 **2.4. Joint Maximum A Posteriori (JMAP) Estimation**

The negative logarithm of the posterior distribution is used as the criterion of optimization in order to simplify the exponential terms. The maximization of posterior distribution becomes a minimization of the criterion:

$$\begin{aligned} (\mathbf{f}, \mathbf{z}, \mathbf{v}_z, \mathbf{v}_\epsilon, \mathbf{v}_{\tilde{\zeta}}) &= \arg \max \{p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_{\tilde{\zeta}}, \mathbf{v}_z | \mathbf{g})\} \\ &= \arg \min \{-\ln p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_{\tilde{\zeta}}, \mathbf{v}_z | \mathbf{g})\} \\ &= \arg \min J(\mathbf{f}, \mathbf{z}, \mathbf{v}_z, \mathbf{v}_\epsilon, \mathbf{v}_{\tilde{\zeta}}). \end{aligned} \quad (23)$$

We substitute the distribution formulas and obtain:

$$\begin{aligned} J(\mathbf{f}, \mathbf{z}, \mathbf{v}_z, \mathbf{v}_\epsilon, \mathbf{v}_{\tilde{\zeta}}) &= -\ln p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_{\tilde{\zeta}}, \mathbf{v}_z | \mathbf{g}) \\ &= \frac{1}{2} \sum_i^M \ln v_{\epsilon_i} + \frac{1}{2} (\mathbf{g} - \mathbf{H}\mathbf{f})^T \mathbf{V}_\epsilon^{-1} (\mathbf{g} - \mathbf{H}\mathbf{f}) \\ &\quad + \frac{1}{2} \sum_j^N \ln v_{\tilde{\zeta}_j} + \frac{1}{2} (\mathbf{f} - \mathbf{D}\mathbf{z})^T \mathbf{V}_{\tilde{\zeta}}^{-1} (\mathbf{f} - \mathbf{D}\mathbf{z}) \\ &\quad + \frac{1}{2} \sum_j^N \ln v_{z_j} + \frac{1}{2} \mathbf{z}^T \mathbf{V}_z^{-1} \mathbf{z} + (\alpha_{z_0} + 1) \sum_j^N \ln v_{z_j} \\ &\quad + \beta_{z_0} \sum_j^N v_{z_j}^{-1} + (\alpha_{\epsilon_0} + 1) \sum_i^M \ln v_{\epsilon_i} + \beta_{\epsilon_0} \sum_i^M v_{\epsilon_i}^{-1} \\ &\quad + (\alpha_{\tilde{\zeta}_0} + 1) \sum_j^N \ln v_{\tilde{\zeta}_j} + \beta_{\tilde{\zeta}_0} \sum_j^N v_{\tilde{\zeta}_j}^{-1}. \end{aligned} \quad (24)$$

180 The unknown variables are determined by obtaining the expressions of the alternate  
181 minimization in Eq. (24):

$$\hat{\mathbf{f}} = \left( \mathbf{H}^T \hat{\mathbf{V}}_\epsilon^{-1} \mathbf{H} + \hat{\mathbf{V}}_{\tilde{\zeta}}^{-1} \right)^{-1} \left( \mathbf{H}^T \hat{\mathbf{V}}_\epsilon^{-1} \mathbf{g} + \hat{\mathbf{V}}_{\tilde{\zeta}}^{-1} \mathbf{D}\hat{\mathbf{z}} \right), \quad (25)$$

$$\hat{\mathbf{z}} = \left( \mathbf{D}^T \hat{\mathbf{V}}_{\tilde{\zeta}}^{-1} \mathbf{D} + \hat{\mathbf{V}}_z^{-1} \right)^{-1} \mathbf{D}^T \hat{\mathbf{V}}_{\tilde{\zeta}}^{-1} \hat{\mathbf{f}}, \quad (26)$$

$$\hat{v}_{z_j} = \left( \beta_{z_0} + \frac{1}{2} \hat{z}_j^2 \right) / (\alpha_{z_0} + 3/2), \quad (27)$$

$$\hat{v}_{\epsilon_i} = \left( \beta_{\epsilon_0} + \frac{1}{2} \left( g_i - [\mathbf{H}\hat{\mathbf{f}}]_i \right)^2 \right) / (\alpha_{\epsilon_0} + 3/2), \quad (28)$$

$$\hat{v}_{\tilde{\zeta}_j} = \left( \beta_{\tilde{\zeta}_0} + \frac{1}{2} \left( \hat{f}_j - [\mathbf{D}\hat{\mathbf{z}}]_j \right)^2 \right) / (\alpha_{\tilde{\zeta}_0} + 3/2), \quad (29)$$

182  $\forall i \in [1, M]$  and  $\forall j \in [1, N]$ .

183 In 3D X-ray CT, the inversion of matrix  $\left( \mathbf{H}^T \hat{\mathbf{V}}_\epsilon^{-1} \mathbf{H} + \hat{\mathbf{V}}_{\tilde{\zeta}}^{-1} \right)^{-1}$  and  $\left( \mathbf{D}^T \hat{\mathbf{V}}_{\tilde{\zeta}}^{-1} \mathbf{D} + \hat{\mathbf{V}}_z^{-1} \right)^{-1}$  in  
184 Eq.(25) and Eq.(26) are impossible due to the large data size. First-order optimization methods are  
185 generally used in this case. In this paper we use the gradient descent algorithm:

$$\text{for } k = 1 \rightarrow I_G : \hat{\mathbf{f}}^{(k+1)} = \hat{\mathbf{f}}^{(k)} - \hat{\gamma}_f^{(k)} \nabla J_f(\hat{\mathbf{f}}^{(k)}), \quad (30)$$

$$\text{for } k = 1 \rightarrow I_G : \hat{\mathbf{z}}^{(k+1)} = \hat{\mathbf{z}}^{(k)} - \hat{\gamma}_z^{(k)} \nabla J_z(\hat{\mathbf{z}}^{(k)}), \quad (31)$$

186 where  $I_G$  is the number of iterations for the gradient descent estimation,  $\nabla J_f(\cdot)$  and  $\nabla J_z(\cdot)$  are the  
 187 derivatives of the criterion (24) regarding  $f$  and  $z$  respectively.  $\hat{\gamma}_f(\cdot)$  and  $\hat{\gamma}_z(\cdot)$  are the corresponding  
 188 descent step lengths which are obtained by using an optimized step length strategy [61]:

$$\nabla J(f) = -\mathbf{H}^T \mathbf{V}_\epsilon^{-1} (\mathbf{g} - \mathbf{H}f) + \mathbf{V}_\xi^{-1} (f - \mathbf{D}z), \quad (32)$$

$$\nabla J(z) = -\mathbf{D}^T \mathbf{V}_\xi^{-1} (f - \mathbf{D}z) + \mathbf{V}_z^{-1} z, \quad (33)$$

$$\hat{\gamma}_f^{(k)} = \frac{\|\nabla J(\hat{f}^{(k)})\|^2}{\|\hat{\mathbf{Y}}_\epsilon \mathbf{H} \nabla J(\hat{f}^{(k)})\|^2 + \|\hat{\mathbf{Y}}_\xi \nabla J(\hat{f}^{(k)})\|^2}, \quad (34)$$

$$\hat{\gamma}_z^{(k)} = \frac{\|\nabla J(\hat{z}^{(k)})\|^2}{\|\hat{\mathbf{Y}}_\xi \mathbf{D} \nabla J(\hat{z}^{(k)})\|^2 + \|\hat{\mathbf{Y}}_z \nabla J(\hat{z}^{(k)})\|^2}, \quad (35)$$

189 where  $\mathbf{Y}_\epsilon = \mathbf{V}_\epsilon^{-\frac{1}{2}}$ ,  $\mathbf{Y}_\xi = \mathbf{V}_\xi^{-\frac{1}{2}}$  and  $\mathbf{Y}_z = \mathbf{V}_z^{-\frac{1}{2}}$ .

190 The algorithm concerning the optimization of all the unknowns is given in Algorithm 1.

---

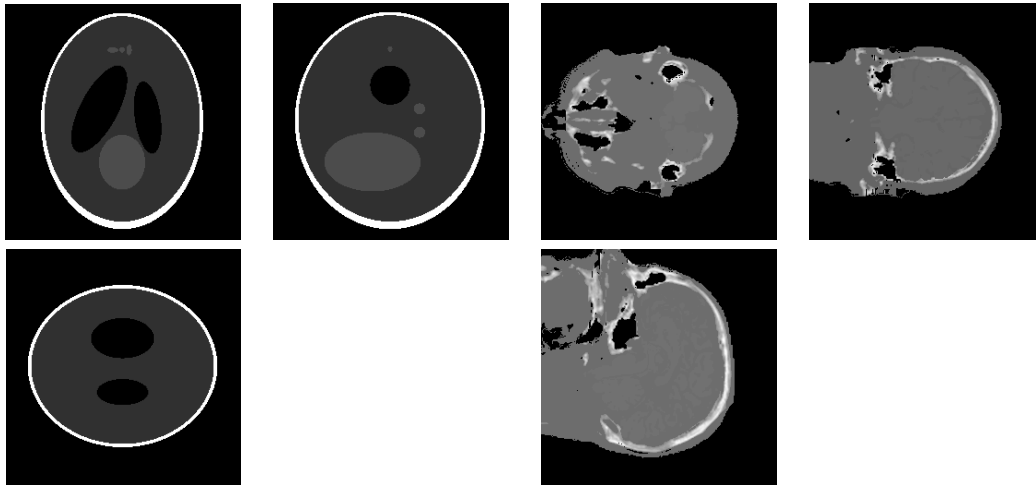
**Algorithm 1** The JMAP algorithm for HHBM method
 

---

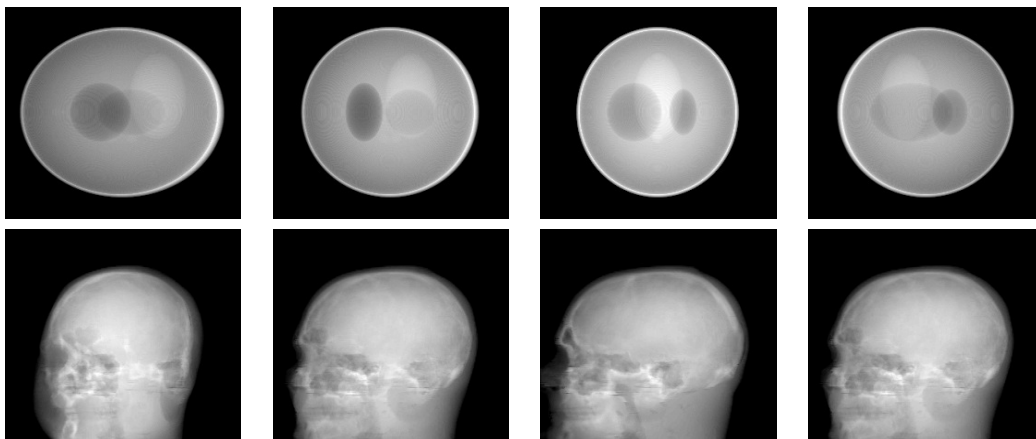
- 1: Fix parameters  $\alpha_{z_0}, \beta_{z_0}, \alpha_{\epsilon_0}, \beta_{\epsilon_0}, \alpha_{\xi_0}, \beta_{\xi_0}, l$
  - 2: **Input:**  $\mathbf{H}, \mathbf{D}, \mathbf{g}$
  - 3: **Output:**  $\hat{f}, \hat{z}, \hat{v}_z, \hat{v}_\epsilon, \hat{v}_\xi$
  - 4: **Initialization:**
  - 5:  $\hat{f} \leftarrow$  normalized FBP
  - 6:  $\hat{z} \leftarrow l$ -level Haar transformation of  $\hat{f}$
  - 7: **for**  $k = 1$  to  $I_{max}$  **do**
  - 8:    $\hat{f}^{(0)} = \hat{f}$
  - 9:   **for**  $k = 1$  to  $I_G$  **do**
  - 10:     Calculate  $\nabla J(\hat{f}^{(k-1)})$  according to Eq. (32)
  - 11:     Update  $\hat{\gamma}_f^{(k)}$  according to Eq. (34)
  - 12:     Update  $\hat{f}^{(k)} = \hat{f}^{(k-1)} - \hat{\gamma}_f^{(k)} \nabla J(\hat{f}^{(k-1)})$
  - 13:   **end for**
  - 14:    $\hat{f} = \hat{f}^{(I_G)}$
  - 15:    $\hat{z}^{(0)} = \hat{z}$
  - 16:   **for**  $k = 1$  to  $I_G$  **do**
  - 17:     Calculate  $\nabla J(\hat{z}^{(k-1)})$  according to Eq. (33)
  - 18:     Update  $\hat{\gamma}_z^{(k)}$  according to Eq. (35)
  - 19:     Update  $\hat{z}^{(k)} = \hat{z}^{(k-1)} - \hat{\gamma}_z^{(k)} \nabla J(\hat{z}^{(k-1)})$
  - 20:   **end for**
  - 21:    $\hat{z} = \hat{z}^{(I_G)}$
  - 22:   Optimize  $v_z$  according to Eq. (27)
  - 23:   Optimize  $v_\epsilon$  according to Eq. (28)
  - 24:   Optimize  $v_\xi$  according to Eq. (29)
  - 25: **end for**
-

### 191 3. Initialization and Experimental Results

192 For the simulations, the 3D simulated "Shepp-Logan" phantom, shown in Fig. 2 on left, Fig.3 on  
 193 top and the 3D real "Head" object, shown in Fig.2 on right, Fig.3 on bottom, both of size  $256^3$ , are used  
 194 as the object of interest to compare the performance of the proposed method to the performance of  
 195 the other state-of-the-art methods. Both Shepp-Logan and Head phantoms consist of several different  
 196 homogeneous areas, so both are piecewise continuous. The voxel values of the original objects are  
 197 normalized to  $[0, 1]$ . The projection directions are uniformly distributed, each projection consists of  
 198  $256^2$  detectors corresponding to a  $256^2$  size image.



**Figure 2.** The three figure on left show three middle slice views of the 3 dimensional SheppLogan phantom, and the three figures on right show the 3 dementional Head phantom.



**Figure 3.** Top: four projections of the 3D SheppLogan phantom from 30, 60, 90, 120 degrees (left to right, respectively); bottom: four projections of the 3D Head phantom from 30, 60, 90, 120 degrees (left to right, respectively).

199 The proposed HHBM method is compared with the conventional Quadratic Regularization (QR)  
 200 and Total Variation (TV) methods. For the QR method, the gradient descent algorithm is used for the  
 201 3D large data size problem. For the TV method, the Split-Bregman method [22] is used to solve the  
 202  $L_1$  norm minimization problem.

203 To evaluate the proposed method and compare it with the state-of-the-art methods, four different  
 204 metrics are used:

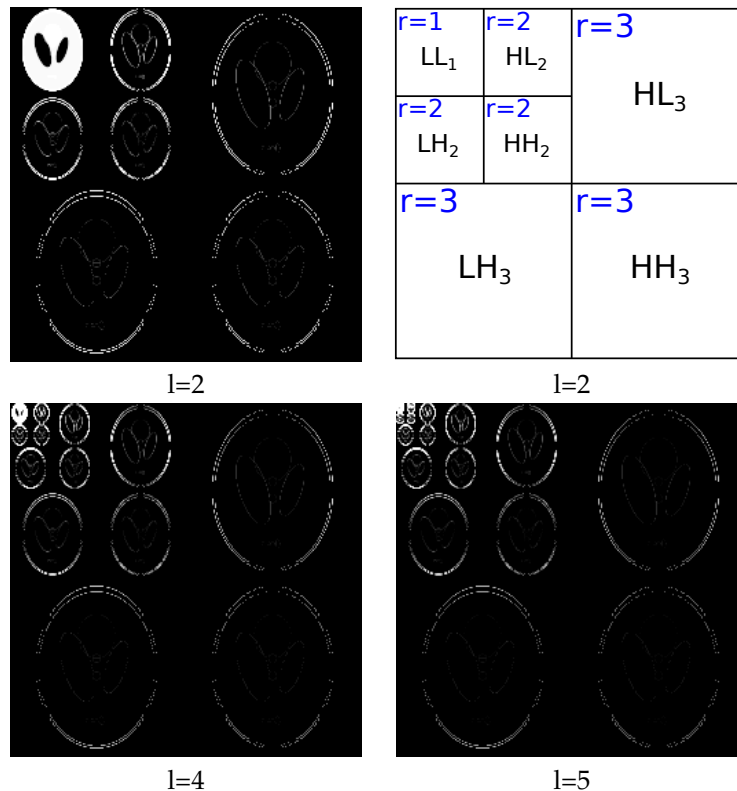
- 205 • the Relative Mean Square Error (RMSE),  $\text{RMSE} = \frac{\|f - \hat{f}\|^2}{\|f\|^2}$  which shows a relative error  
 206 of results;  
 207 • the Improvement of Signal-to-Noise Ratio (ISNR) which measures the improvement during  
 208 iterations;  
 209 • the Peak Signal-to-Noise-Ratio (PSNR) which presents the SNR relative to the peak data value;  
 210 • the Structural Similarity of IMage (SSIM) [62] which evaluates the quality of result approaching  
 211 human visualizing.

212 In 3D X-ray CT, the projection matrix  $H$  is very large and is not accessible. For the simulations,  
 213 only the projection operator  $Hf$  and the back-projection operator  $H^T g$  are used. Considering that the  
 214 costly projection and back-projection operators are computed in every iteration, the GPU processor is  
 215 used via the ASTRA toolbox [63] to accelerate the computation.

### 216 3.1. Initializations

217 The initialization for the variables  $f$  and  $z$  as well as the hyperparameters  $\alpha_{\epsilon_0}$ ,  $\beta_{\epsilon_0}$ ,  $\alpha_{\zeta_0}$ ,  $\beta_{\zeta_0}$ ,  $\alpha_{z_0}$   
 218 and  $\beta_{z_0}$  are discussed in this section.

219 The reconstructed phantom obtained by using the Filtered Back Projection (FBP) method is  
 220 considered to be initial value  $\hat{f}_{ini}$ . The initialization of coefficient  $\hat{z}_{ini}$  is the multilevel Haar  
 221 transformation of  $\hat{f}_{ini}$ :  $\hat{z}_{ini} = D^{-1}\hat{f}_{ini}$ . In this article, we choose the level of transformation such  
 222 that  $z$  has a sparse structure. As shown in Fig.4, when the transform level is small, for example  
 223 2–levels, the coefficient  $z$  is not sparse; when the transform level is sufficiently large, the coefficient  
 224 is sparse. In this paper, we set  $z$  as a 5–level Haar transform coefficient.



**Figure 4.** Slice of 3D multilevel Haar transformation coefficient  $z$  of the  $256^3$  Shepp-Logan phantom with: 2 levels (top-left), 4 levels (bottom-left) and 5 levels (bottom-right). The figure on the top-right shows the ranks of coefficients for a 2-level transformation.

225 The initialization for  $\alpha_{z_0}$  and  $\beta_{z_0}$  is based on the sparse structure of the variable  $z$ . In Figure 4,  
 226 we can see that the sparsity rate depends on the rank of transform coefficient  $r$ , where  $r \in [1, l + 1]$ .

227 For example when  $l = 2$ , shown in Figure 4, the coefficient has 3 ranks,  $r \in [1, 2, 3]$ . The first rank  
 228  $r = 1$  corresponds to the low frequency components in the transform coefficient.

229 The variable  $z$  is modeled by a  $St_g$  distribution, with variance equal to  $\text{Var}[z_j | \alpha_{z_0}, \beta_{z_0}] =$   
 230  $\beta_{z_0} / (\alpha_{z_0} - 1)$ ,  $\forall j \in [0, N]$ . In this article, we fix the value  $\alpha_{z_0} = 2.1$  in order to have  $\text{Var}[z_j | \alpha_{z_0} =$   
 231  $2.1, \beta_{z_0}] \approx \beta_{z_0}$ ,  $\forall j \in [0, N]$ . The sparsity rate of  $z$  is defined by initializing different values for  $\beta_{z_0}$ ,  
 232 with a sparser structure when  $\beta_{z_0}$  is smaller.  $\beta_{z_0}$  is initialized as  $\beta_{z_0} = 10^{-r+1}$ . When  $l = 5$ , we have  
 233  $r = [1, 2, 3, 4, 5, 6]$ , and the hyperparameters  $\beta_{z_0} = [10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}]$  respectively.

234 The initialization for the hyperparameters,  $\alpha_{\epsilon_0}$ ,  $\beta_{\epsilon_0}$ ,  $\alpha_{\xi_0}$  and  $\beta_{\xi_0}$ , is based on the prior models of  
 235 the variances  $v_\epsilon$  and  $v_\xi$  we have chosen. In the proposed method, we consider the background of  
 236 the generalized Student-t distribution, in which both  $\epsilon$  and  $\xi$  are modeled by a Gaussian distribution  
 237 with Inverse Gamma distributed variance, i.e. the  $St_g$  distribution according to Eq. (13).

The noise  $\epsilon$  depends on the SNR of the dataset. In order to exploit this information in the  
 initialization, we express the biased dataset as the sum of uncontaminated dataset  $g_0$  and the additive  
 noise  $\epsilon$ :

$$g = g_0 + \epsilon. \quad (36)$$

As the noise  $\epsilon$  and the uncontaminated data  $g_0$  are supposed to be independent, we have:

$$\|g\|^2 = \|g_0\|^2 + \|\epsilon\|^2. \quad (37)$$

And the SNR of the dataset is:

$$\text{SNR} = 10 \log \frac{\|g_0\|^2}{\|\epsilon\|^2} = 10 \log \frac{\|g\|^2 - \|\epsilon\|^2}{\|\epsilon\|^2}. \quad (38)$$

With  $\mathbb{E}[\epsilon] = 0$ , we have:

$$v_\epsilon = \mathbb{E}[\epsilon^2] \approx \frac{\|\epsilon\|^2}{M} = \frac{\|g\|^2}{M} \times \frac{1}{1 + 10^{\text{SNR}/10}}. \quad (39)$$

The mean of variance  $v_\epsilon$  of the noise  $\epsilon$  is  $\mathbb{E}[v_{\epsilon_i} | \alpha_{\epsilon_0}, \beta_{\epsilon_0}] = \beta_{\epsilon_0} / (\alpha_{\epsilon_0} - 1)$ ,  $\forall i \in [0, M]$ , so we obtain:

$$\beta_{\epsilon_0} = \frac{\|g\|^2}{M} \times \frac{1}{1 + 10^{\text{SNR}/10}} \times (\alpha_{\epsilon_0} - 1). \quad (40)$$

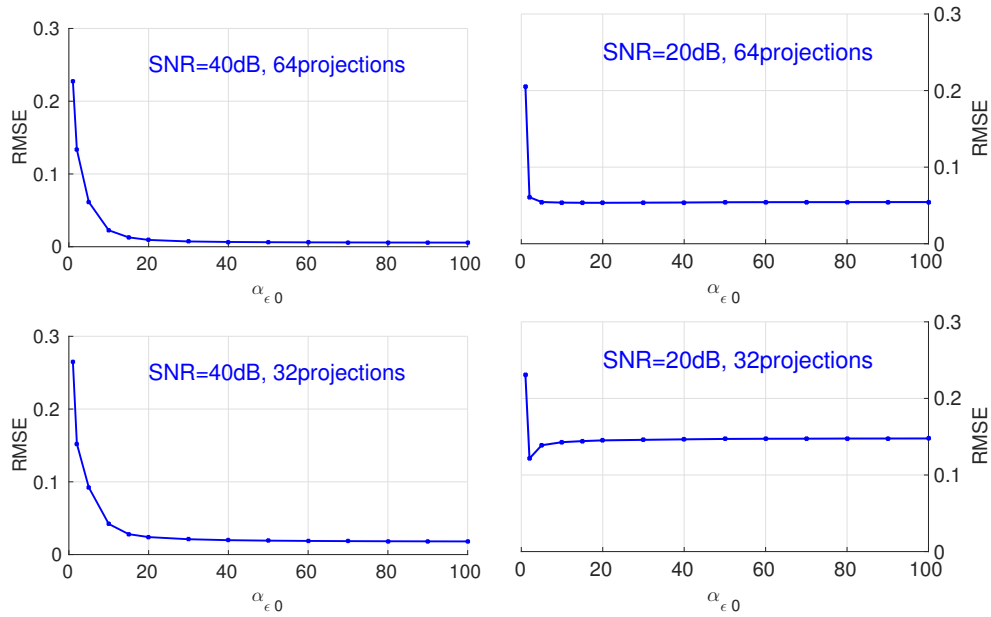
238 The two hyperparameters  $\alpha_{\epsilon_0}$  and  $\beta_{\epsilon_0}$  are combined according to Eq. (40); hence, initialization  
 239 for one of them is sufficient. In real applications, the SNR of the dataset is unknown, but we can use  
 240 the projection of an empty object, i.e.  $f = \mathbf{0}$ , to obtain a rough value of the variance of noise  $v_\epsilon$ .

241 Fig.5 shows the influence of the value of  $\alpha_{\epsilon_0}$  on the reconstruction. According to the results, a  
 242 bigger value for  $\alpha_{\epsilon_0}$  results in a smaller value on RMSE for different numbers of projections and SNR  
 243 of the dataset. This monotonous property facilitates the initialization of this hyperparameter, as a  
 244 large value for  $\alpha_{\epsilon_0}$  satisfies all cases. When  $\alpha_{\epsilon_0}$  is greater than a threshold value, the RMSE does not  
 245 change with different initialization values for  $\alpha_{\epsilon_0}$ .

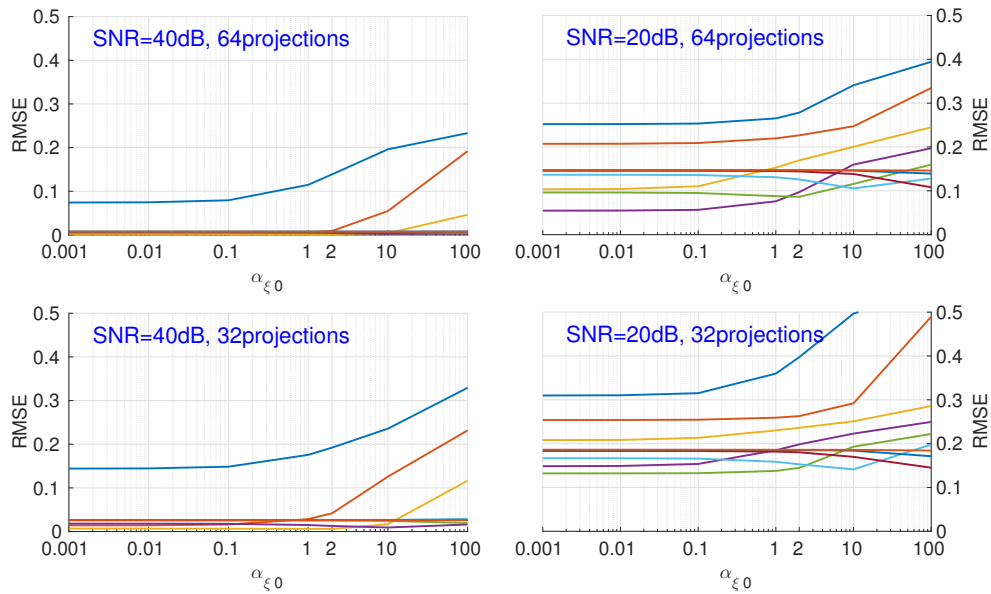
246 For  $\xi$ , both  $\alpha_{\xi_0}$  and  $\beta_{\xi_0}$  are analyzed for the influence of the reconstruction results.

247 Fig.6 shows the influence of the hyperparameter  $\alpha_{\xi_0}$ . Different colors represent an initialization  
 248 with a different value of  $\beta_{\xi_0}$ . As here we focus on the analysis of  $\alpha_{\xi_0}$  for all cases of  $\beta_{\xi_0}$  values, we  
 249 don't show the corresponding  $\beta_{\xi_0}$  value for each different color. For different noise levels, different  
 250 numbers of projections and different  $\beta_{\xi_0}$  values, the RMSE has an upward trend when the value of  
 251  $\alpha_{\xi_0}$  becomes larger.

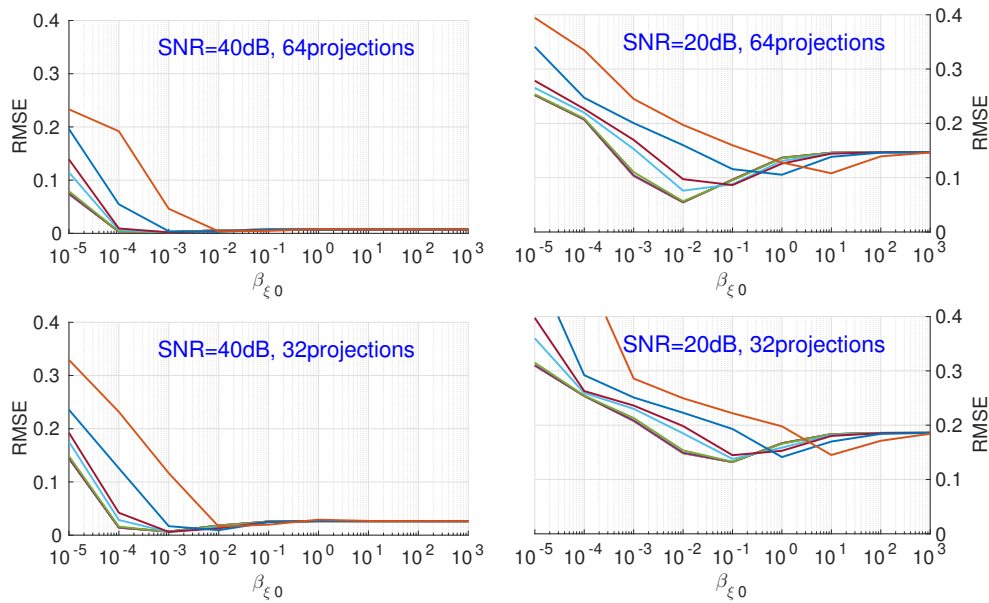
252 Fig.7 shows the influence of the hyperparameter  $\beta_{\xi_0}$ . Different colors represent an initialization  
 253 with a different hyperparameter  $\alpha_{\xi_0}$ . For different noise levels, different numbers of projections and  
 254 different values for  $\alpha_{\xi_0}$ , when the value of  $\beta_{\xi_0}$  increases, the RMSE decreases; then, after a slight  
 255 increase, levels out.



**Figure 5.** Influence of hyperparameter  $\alpha_{\epsilon_0}$  on RMSE of final reconstruction results for different numbers of projections and noise.



**Figure 6.** Influence of hyperparameter  $\alpha_{\xi_0}$ , with different fixed values of  $\beta_{\xi_0}$ , on RMSE of reconstruction results for different number of projections and noise. Each color corresponds to a different initialization of  $\beta_{\xi_0}$ .



**Figure 7.** Influence of hyperparameter  $\beta_{\xi_0}$ , with different fixed value of  $\alpha_{\xi_0}$ , on RMSE of reconstruction results for different number of projections and noise. Each color corresponds to a different initialization of  $\alpha_{\xi_0}$ .

256 In [46], it is pointed out that when  $\alpha$  and  $\beta$  of the  $St_g$  distribution are both large, the  $St_g$   
 257 distribution approaches a Gaussian distribution, which is the case for the additive noise  $\epsilon$ . If  $\alpha$  and  $\beta$   
 258 are very small (approaching 0), the  $St_g$  distribution becomes a non-informative distribution (Jeffreys  
 259 distribution); when  $\alpha$  and  $\beta$  are both small, the  $St_g$  has the sparsity enforcing property, which is  
 260 the case for the sparse  $\xi$ . Consequently, the initialization of the hyperparameters are theoretically  
 261 supported, and they can be initialized with respect to these properties in other simulations.

**Table 1.** Comparison of RMSE, ISNR, PSNR and SSIM of reconstructed phantom with 50 global iterations (10 gradient descent iterations in each global iteration). The values of the regularization parameters are respectively  $\lambda_{QR} = 10$  and  $\lambda_{TV} = 50$  for SNR=40dB,  $\lambda_{QR} = 600$  and  $\lambda_{TV} = 100$  for SNR=20dB.

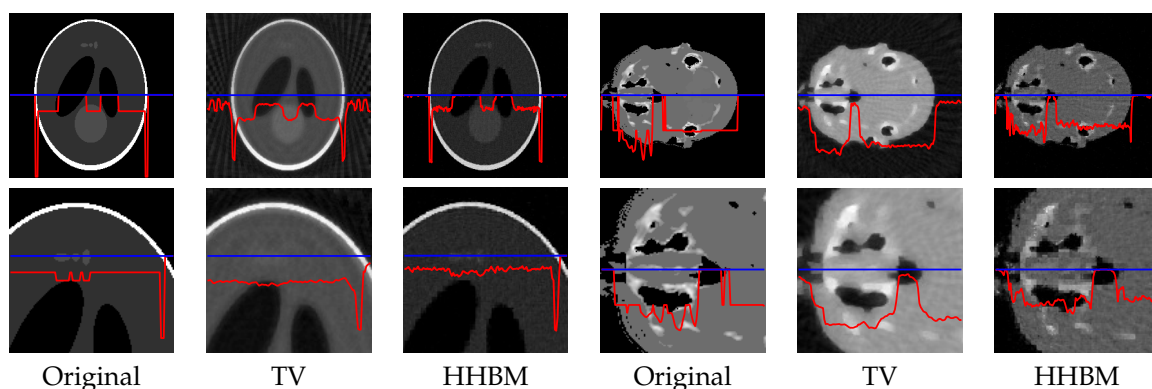
	256 × 256 × 256 Shepp-Logan phantom											
	180 projections						90 projections					
	40dB			20dB			40dB			20dB		
	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM
<b>RMSE</b>	0.0236	0.0114	<b>0.0069</b>	0.1309	<b>0.0209</b>	0.0755	0.0401	0.0212	<b>0.0092</b>	0.1558	<b>0.0491</b>	0.1117
<b>ISNR</b>	5.5584	8.7217	<b>10.9346</b>	7.2024	<b>15.1775</b>	10.2162	6.6136	9.3832	<b>12.9973</b>	8.4583	<b>13.4765</b>	9.9056
<b>PSNR</b>	30.0675	33.2308	<b>35.4437</b>	22.6318	<b>30.6069</b>	25.0209	27.7743	30.5439	<b>34.1579</b>	21.8754	<b>26.8937</b>	23.3227
<b>SSIM</b>	0.9999	0.9999	<b>1.0000</b>	0.9992	<b>0.9999</b>	0.9995	0.9997	0.9999	<b>0.9999</b>	0.9990	<b>0.9997</b>	0.9993
	60 projections						45 projections					
	40dB			20dB			40dB			20dB		
	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM
<b>RMSE</b>	0.0636	0.0321	<b>0.0107</b>	0.1656	<b>0.0753</b>	0.1293	0.0904	0.0474	<b>0.0132</b>	0.1854	0.0901	<b>0.1414</b>
<b>ISNR</b>	9.3826	12.3480	<b>17.1346</b>	9.1492	<b>12.5701</b>	10.2226	10.3301	13.1308	<b>18.6839</b>	10.0137	13.1476	<b>11.1916</b>
<b>PSNR</b>	25.7693	28.7347	<b>33.5214</b>	21.6116	<b>25.0325</b>	22.6849	24.2404	27.0412	<b>32.5942</b>	21.1195	24.2535	<b>22.2974</b>
<b>SSIM</b>	0.9996	0.9995	<b>0.9999</b>	0.9990	<b>0.9995</b>	0.9992	0.9994	0.9997	<b>0.9999</b>	0.9988	0.9994	<b>0.9991</b>
	36 projections						18 projections					
	40dB			20dB			40dB			20dB		
	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM
<b>RMSE</b>	0.1177	0.0680	<b>0.0169</b>	0.1957	<b>0.1116</b>	0.1500	0.2581	0.2104	<b>0.0574</b>	0.2907	0.2313	<b>0.2014</b>
<b>ISNR</b>	10.6591	13.0424	<b>19.0933</b>	10.8633	<b>13.3032</b>	12.0187	10.7122	11.5992	<b>17.2373</b>	10.8088	11.8022	<b>12.4036</b>
<b>PSNR</b>	23.0949	25.4783	<b>31.5292</b>	20.8865	<b>23.3264</b>	22.0420	19.6263	20.5133	<b>26.1514</b>	19.1085	20.1020	<b>20.7033</b>
<b>SSIM</b>	0.9993	0.9996	<b>0.9999</b>	0.9988	<b>0.9993</b>	0.9990	0.9983	0.9987	<b>0.9996</b>	0.9981	0.9985	<b>0.9987</b>

### 262 3.2. Simulation results with limited number of projections

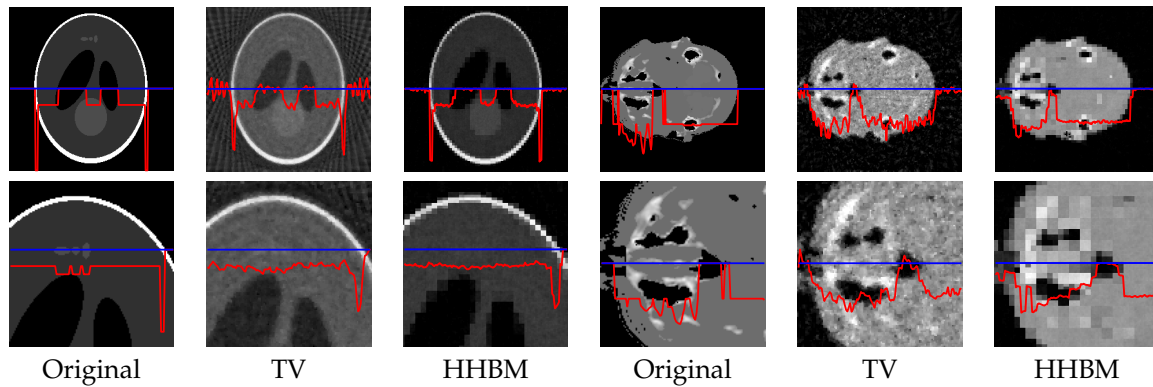
263 We apply 180, 90, 60, 45, 36 and 18 projections evenly distributed in  $[0, 180]$  degrees for the  
 264 reconstruction of the 3D Shepp-Logan phantom of size  $256^3$ ; each projection contains  $256^2$  detectors.  
 265 The number of projections are chosen such that there is respectively one projection every 1, 2, 3, 4, 5  
 266 and 10 degrees.

267 In Table 1, different evaluation metrics of the reconstructed  $256^3$  Shepp-Logan phantom are  
 268 compared. It is shown that the HHBM method does not always perform better than the TV method,  
 269 especially when there are sufficient numbers of projections. But when there is insufficient projection  
 270 data, the HHBM method is more robust than the TV method. On the other hand, as it is known  
 271 that the choice of regularization parameter plays an important role in the regularization methods  
 272 like QR or TV, and the value for the regularization parameter should be selected for each different  
 273 system settings, the HHBM method is much more robust on the initialization of hyper-parameters.  
 274 As we can see from Fig.5, Fig.6 and Fig.7, once we have chosen the hyperparameters in a certain  
 275 interval, which is not difficult to fix according to the properties of the prior model, we can obtain the  
 276 appropriate reconstruction results. More importantly, in the Bayesian approach, the prior model can  
 277 be chosen from a variety of other suitable distributions, which gives more possibilities for the models  
 278 than the conventional regularization methods. We may also choose different point estimators from  
 279 the posterior distribution, for example the Posterior Mean, etc.

280 Figure 8 and Figure 9 show the reconstructed middle slice of the "Shepp-Logan" phantom  
 281 and "Head" object by using the TV and HHBM methods from 36 projections with SNR=40dB and  
 282 SNR=20dB. The red curve illustrates the profile of the blue line position. In the reconstructed  
 283 Shepp-Logan phantom obtained using the TV method, the three small circles on the top of the slice are  
 284 not evident. By using the HHBM method, we can distinguish these three small circles. By comparing  
 285 the profiles of the slice of reconstructed Shepp Logan phantom, we can see that by using the HHBM  
 286 method, the contour positions on the profile are closer to the original profile than those obtained  
 287 using the TV method. In the reconstructed Head object, there are more details than the simulated  
 288 Shepp-Logan phantom, especially in the zoom area in the second line in Figure 8. By comparing  
 289 the results, we can see that for the type of object which contains some small details, the TV method  
 290 derives a result with smoother homogeneous areas but with fewer details in the contour areas than  
 291 the HHBM method. Some of the white material, which is dispersed into discontinuous small blocks  
 292 in the Head object, is connected in the results of the TV method. From these images we conclude that  
 293 with an insufficient number of projections the proposed method gives results with clearer contours  
 and details.

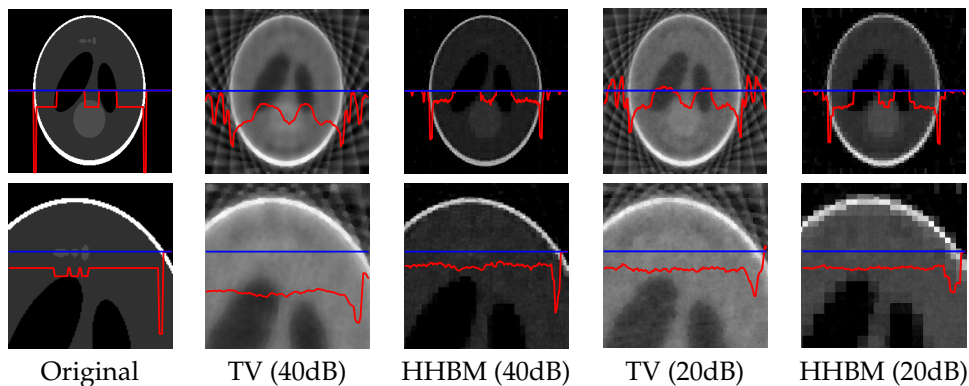


**Figure 8.** Reconstructed "Shepp-Logan" phantom and "Head" object of size  $256^3$ , with a dataset of **36 projections** and **SNR=40dB**, by using the TV and HHBM methods. Bottom figures are zones of the corresponding top figures. The red curves are the profiles at the position of the corresponding blue lines.



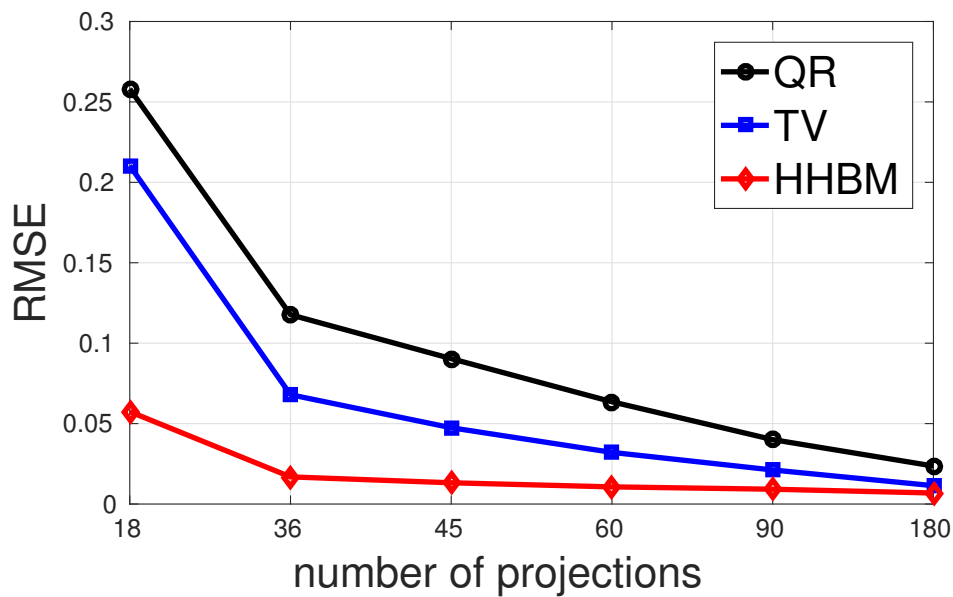
**Figure 9.** Reconstructed "Shepp-Logan" phantom and "Head" object of size  $256^3$ , with a dataset of **36 projections** and **SNR=20 dB**, by using the TV and HHBM methods. Bottom figures are zones of the corresponding top figures. The red curves are the profiles at the position of the corresponding blue lines.

295 Fig.10 shows the reconstructed Shepp-Logan phantom from 18 projections of SNR=40dB and  
 296 20dB. In this very underdetermined case, the HHBM method can still obtain a result which is clear  
 enough to distinguish the primary zones and contours of the object.

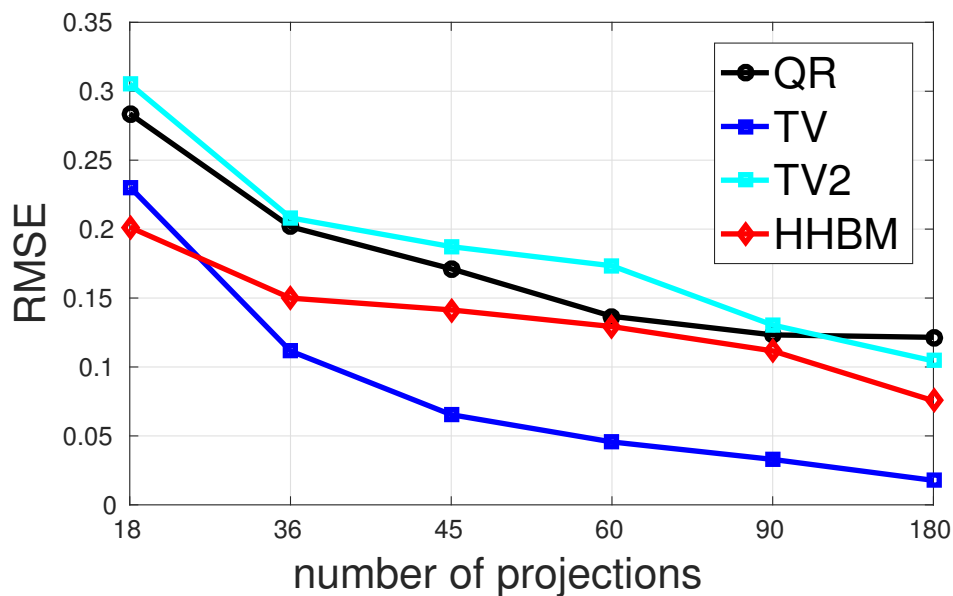


**Figure 10.** Reconstructed Shepp-Logan phantom of size  $256^3$ , with a dataset of **18 projections** of **SNR=40dB** and **20dB**, by using the TV and HHBM methods. The red curves are the profiles at the position of the corresponding blue lines.

297 Fig.11 and Fig.12 show the comparison between the QR, TV and HHBM methods with a high  
 298 SNR=40dB and a low SNR=20dB dataset respectively. The abscissa corresponds to the number of  
 299 projections evenly distributed from  $0^\circ$  to  $180^\circ$ , and the ordinate is the RMSE after 50 iterations. In the  
 300 simulations we used a SNR=40dB to represent a weak noise case and SNR=20dB for a strong noise  
 301 case. When SNR=40dB, the HHBM method outperforms both quadratic regularization and the TV  
 302 method. When SNR=20dB, the TV method with the optimal regularization parameter outperforms  
 303 the HHBM method. However, in Fig.12, we show another curve in light green (TV2) showing the  
 304 TV reconstruction with some random regularization parameters, which are chosen as a value not far  
 305 from the optimal regularization parameters. From these results we can see that the HHBM method  
 306 is more robust than the TV reconstruction method with respect to the regularization parameter, the  
 307 optimal value of which is, on the other hand, difficult to determine in the real applications where we  
 308 cannot evaluate the estimation quality.  
 309



**Figure 11.** The performances of different methods for reconstructing Shepp-Logan phantom in terms of RMSE with different numbers of projections evenly distributed in  $[0^\circ, 180^\circ]$  and a high  $\text{SNR}=40\text{dB}$ .

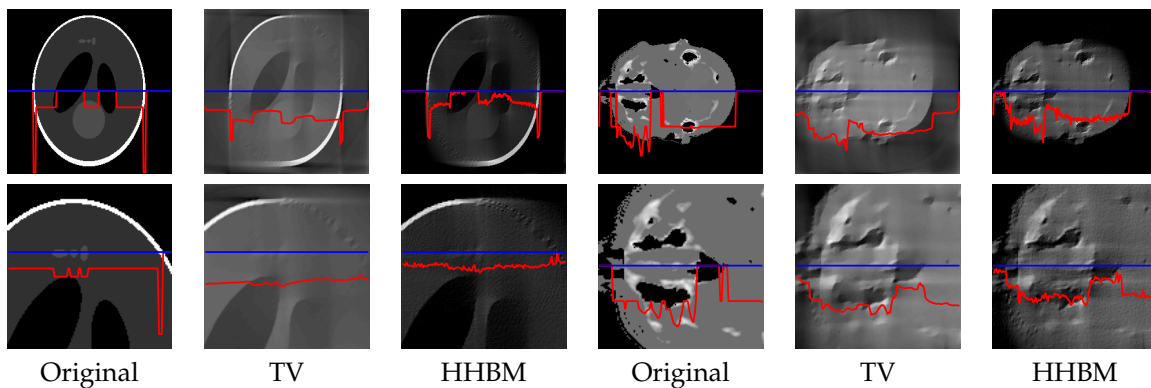


**Figure 12.** The performances of different methods for reconstructing Shepp-Logan phantom in terms of RMSE with different numbers of projections evenly distributed in  $[0^\circ, 180^\circ]$  and a low  $\text{SNR}=20\text{dB}$ .

310 *3.3. Simulation results with limited angle of projections*

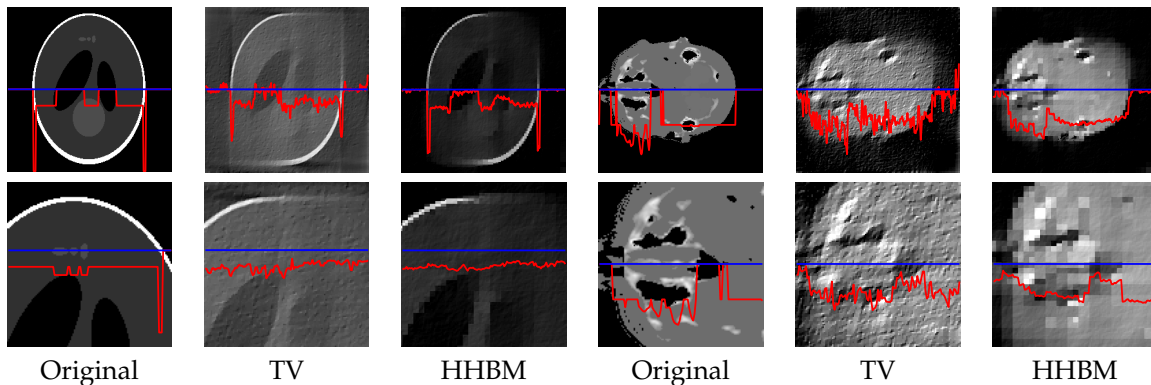
311 In both medical and industrial X-ray CT, another common challenge is the limit of projection  
 312 angles. In this part of the simulation, we use evenly distributed projections in a limited range of  
 313 angles for the simulated 3D "Shepp-Logan" phantom and the 3D "Head" object, both of which have a  
 314 size of  $256^3$ .

315 Figure 13 and Figure 14 show the middle slice of the reconstructed Shepp-Logan phantom and  
 316 the Head object, from 90 projections distributed between  $0^\circ$  and  $90^\circ$ , with projection SNR of 40dB and  
 317 20dB. By using the TV method, the reconstructed object is blurry along the the diagonal direction for  
 318 which there is no projection data, and there is a square corner where the object should have a rounded  
 319 edge. By using the HHBM method, we get results that are more consistent with the true shape and  
 clearer contours.



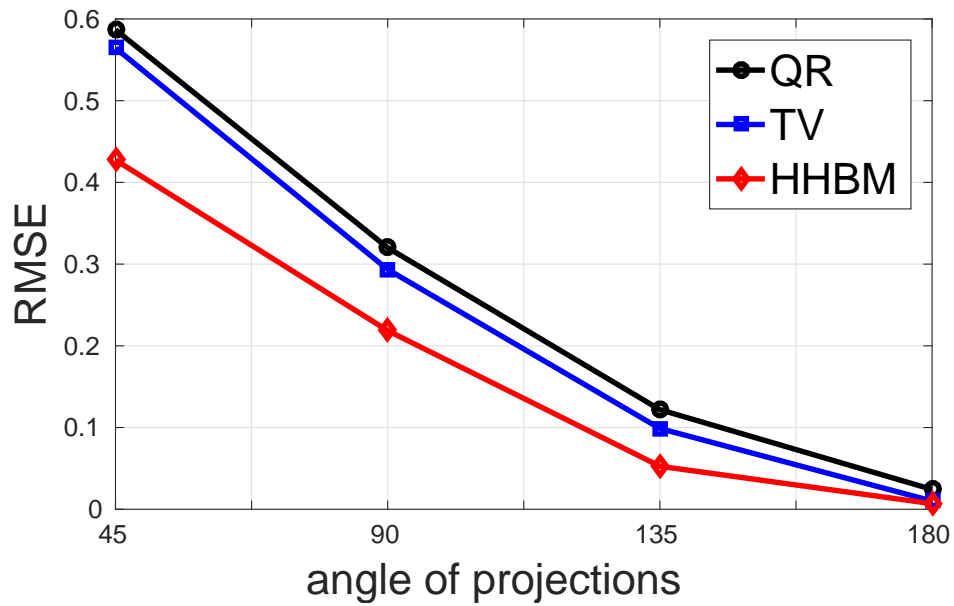
**Figure 13.** Slice of reconstructed 3D Shepp-Logan phantom and 3D Head object, with **90 projections** evenly distributed in  $[0^\circ, 90^\circ]$ , **SNR=40dB**. Bottom figures are parts of the corresponding top figures. The red curves are the profiles at the position of the corresponding blue lines.

320

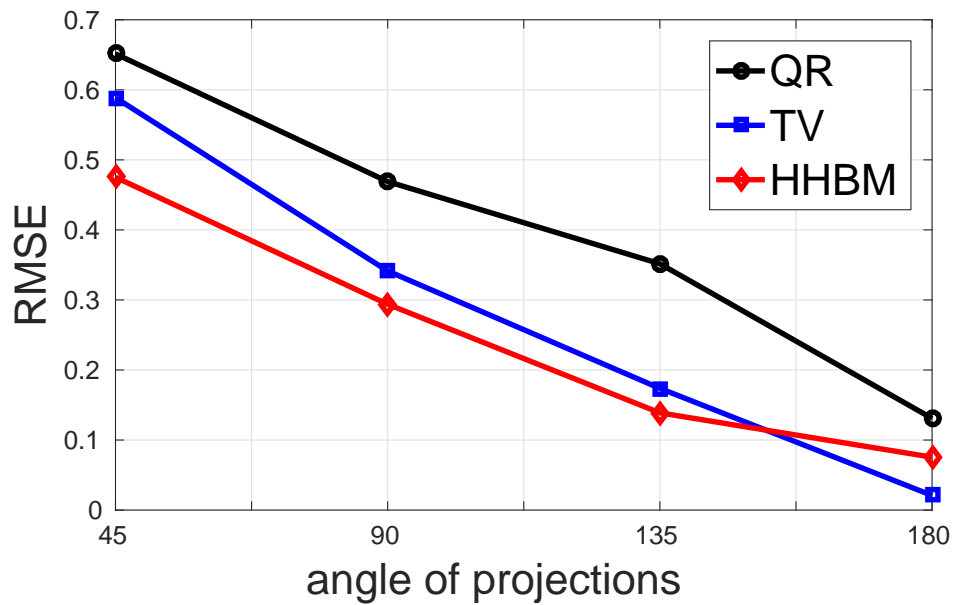


**Figure 14.** Slice of reconstructed 3D Shepp-Logan phantom and 3D Head object, with **90 projections** evenly distributed in  $[0^\circ, 90^\circ]$ , **SNR=20dB**. Bottom figures are parts of the corresponding top figures. The red curves are the profiles at the position of the corresponding blue lines.

321 Figure 15 and Figure 16 show the comparison of the performance in terms of RMSE of different  
 322 methods with a high SNR=40dB and a low SNR=20dB. In this comparison, four cases of limited  
 323 projection angles are considered and they are: 45, 90, 135 and 180 projections evenly distributed in  
 324  $[0^\circ, 45^\circ]$ ,  $[0^\circ, 90^\circ]$ ,  $[0^\circ, 135^\circ]$  and  $[0^\circ, 180^\circ]$  respectively. There is one projection every  $1^\circ$ . From these  
 325 two figures, we conclude that the proposed HHBM method remains more robust than the other two  
 326 conventional methods when there are limited projection angles.



**Figure 15.** The performance of different methods for reconstructing Shepp-Logan phantom in terms of RMSE with different **limited projection angles** and a high SNR=40dB.

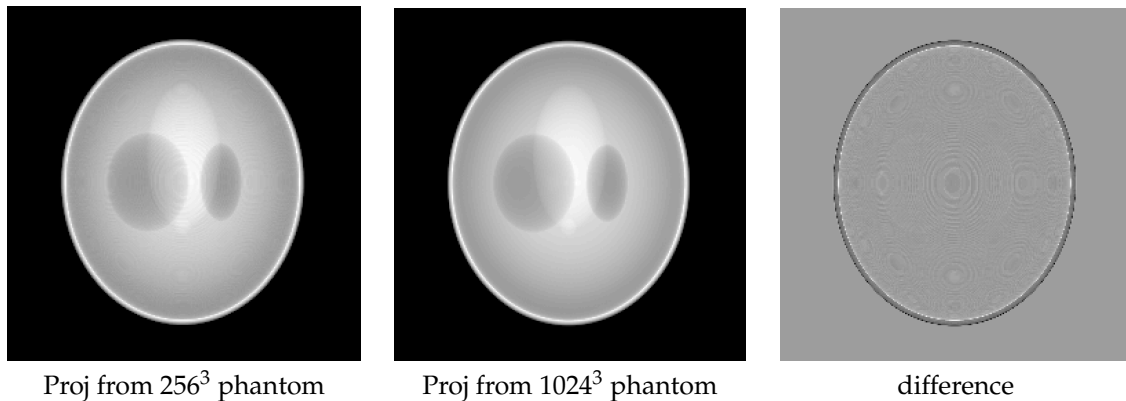


**Figure 16.** The performance of different methods for reconstructing Shepp-Logan phantom in terms of RMSE with different **limited projection angles** and a low SNR=20dB.

327 *3.4. Simulation with a different forward model*

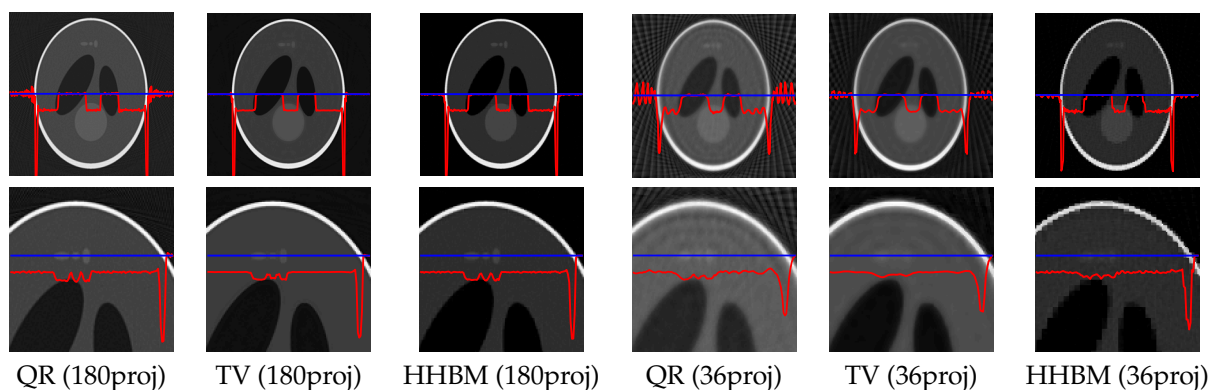
328 In order to avoid the inverse crime, we apply a slightly different forward model. During the  
 329 projection, the projector is applied on the Shepp-Logan phantom of size  $1024^3$ , with the detector size  
 330 of  $256^2$  for each projection direction.

331 In Figure 17, we show the results of the two projection data obtained as well as their difference  
 332 which can be considered as the forward modelling error. We then use the data obtained from  $1024^3$   
 333 phantom to reconstruct an object of size  $256^3$ .



**Figure 17.** The projection image in angle 90 degree by using two different forward model: projection from  $256^3$  phantom (left) and projection from  $1024^3$  phantom (middle). The difference between them is on the right.

334 In Figure 18, the middle slice of the reconstructed Shepp-Logan phantom by using QR, TV and  
 335 HHBM methods are presented, by using 180 and 36 projections respectively. From the figures, we  
 336 can see that when there are 180 projections, all three methods performs well, and the TV method  
 337 detects better the contours while HHBM method has more noise at the contour areas. When there are  
 338 insufficient projection numbers (36 projections in this simulation), the HHBM method outperforms  
 339 the QR and TV methods for reconstructing the details in the phantom, for example the three small  
 circles in the top of the phantom.



**Figure 18.** Reconstructed phantom with different forward model and **180 projections** and **36 projections** by using QR (left), TV (middle) and HHBM (right) methods.

340 In Table 2, the RMSE of the reconstructed phantom by using the different methods are compared.  
 341 We can conclude that, when the projector model is different than the reconstruction one, all these  
 342 three methods (QR, TV and HHBM) have good performance when there is 180 projections. When  
 343 the projection number decreases, TV and HHBM methods outperforms QR method. Comparing with  
 344

345 the TV method, HHBM method is more robust to the number of projections. When the projection  
 number is smaller than 60, HHBM method outperforms TV method.

**Table 2.** The RMSE of reconstructed  $256^3$  Shepp-Logan phantom by using projection obtained from a  
 $1024^3$  Shepp-Logan phantom.

RMSE	QR	TV	HHBM
180 proj	0.0581	0.0540	0.0558
90 proj	0.0655	0.0573	0.0610
60 proj	0.0846	0.0675	0.0690
45 proj	0.1079	0.0830	0.0783
36 proj	0.1326	0.1027	0.0882

346

347

All the MATLAB codes for the simulations in this paper can be found on github [64].

#### 348 4. Discussion

349 One advantage of the Bayesian approach is the estimation of the parameters along with  
 350 the estimation of unknown variables of the forward model at each iteration. However, like in  
 351 regularization methods, the hyper-parameters need to be initialized.

352 While the parameters in the regularization methods play an important role in the final results and  
 353 they are costly to fix, the hyper-parameters in HHBM can be initialized based on the prior information  
 354 (the sparse structure of  $z$ ) and the prior model (the Student-t distribution). In this article, we have  
 355 shown that once the hyper-parameters are fixed in a certain appropriate interval, which is not difficult  
 356 to obtain, the corresponding algorithm is robust. In this work, the hyper-parameters are not fixed  
 357 via the classical approach using non-informative prior laws (i.e. considering the Inverse Gamma  
 358 corresponding parameters such that it approaches Jeffreys), [65].

#### 359 5. Conclusions

360 In this paper, we propose a Bayesian method with a hierarchical structured prior model based on  
 361 multilevel Haar transformation (HHBM) for 3D X-ray CT reconstruction. Simulation results indicate  
 362 that for a limited number of projections or limited projection angles, the proposed method is more  
 363 robust to noise and to regularization parameters than the classical QR and TV methods.

364 Indeed, we observe a relatively weak influence of the hyper-parameters in the behavior of the  
 365 corresponding iterative algorithm. The interest of this weak dependency is that it offers a practical  
 366 way to ensure the initialization of the algorithm which typically is not-trivial.

367 In this context, as future work, we are investigating the causes of the relative weak influence of  
 368 the hyper-parameters and the theoretical foundation of the corresponding robust interval, extending  
 369 the discussion to the same approach using sparsity enforcing priors expressed as Normal variance  
 370 mixtures, but for other mixing distributions (Gamma, generalized inverse Gaussian), [66].

371 Another extension of this work is to consider the Posterior Mean as an estimator. This can be  
 372 done via the Variational Bayesian Approach (VBA) but a practical implementation requires a method  
 373 of accessing the diagonal elements of the large matrix  $\mathbf{H}^T \mathbf{H}$ , which is being studied by our group.

374 **Author Contributions:** This work has been done during the PhD preparation of the first author. Ali  
 375 Mohammad-Djafari and Nicolas Gac have been the co-supervisors of the PhD work. Mircea Dumitru was a  
 376 Post-Doc who also had finished his PhD under the supervision of Ali Mohammad-Djafari and was continuing  
 377 to work on this project.

378 **Conflicts of Interest:** The authors declare no conflict of interest.

#### 379 Bibliography

- 380 1. Arridge, S.R. Optical tomography in medical imaging. *Inverse Problems* **1999**, *15*, R41.

- 381 2. Hanke, R.; Fuchs, T.; Uhlmann, N. X-ray based methods for non-destructive testing and material  
382 characterization. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers,*  
383 *Detectors and Associated Equipment* **2008**, *591*, 14–18.
- 384 3. Kalender, W.A. X-ray Computed Tomography. *Physics in medicine and biology* **2006**, *51*, R29.
- 385 4. Natterer, F.; Wubbeling, F. A propagation-backpropagation method for ultrasound tomography. *Inverse*  
386 *Problems* **1995**, *11*, 1225.
- 387 5. Bolomey, J.C.; Pichot, C. Microwave tomography: from theory to practical imaging systems. *International*  
388 *Journal of Imaging Systems and Technology* **1990**, *2*, 144–156.
- 389 6. Nelson, J.; Milner, T.; Tanenbaum, B.; Goodman, D.; Van Gemert, M. Infra-red tomography of  
390 port-wine-stain blood vessels in human skin. *Lasers in Medical Science* **1996**, *11*, 199–204.
- 391 7. Deans, S.R. *The Radon transform and some of its applications*; Courier Corporation, 2007.
- 392 8. Feldkamp, L.; Davis, L.; Kress, J. Practical cone-beam algorithm. *JOSA A* **1984**, *1*, 612–619.
- 393 9. Alifanov, O.M.; Artioukhine, E.A.; Rummyantsev, S.V. *Extreme methods for solving ill-posed problems with*  
394 *applications to inverse heat transfer problems*; Begell house New York, 1995.
- 395 10. O’Sullivan, F. A statistical perspective on ill-posed inverse problems. *Statistical Science* **1986**, pp. 502–518.
- 396 11. Gordon, R. A tutorial on ART (algebraic reconstruction techniques). *IEEE Transactions on Nuclear Science*  
397 **1974**, *21*, 78–93.
- 398 12. Gordon, R.; Bender, R.; Herman, G.T. Algebraic reconstruction techniques (ART) for three-dimensional  
399 electron microscopy and X-ray photography. *Journal of Theoretical Biology* **1970**, *29*, 471–481.
- 400 13. Andersen, A.H.; Kak, A.C. Simultaneous algebraic reconstruction technique (SART): a superior  
401 implementation of the ART algorithm. *Ultrasonic Imaging* **1984**, *6*, 81–94.
- 402 14. Trampert, J.; Leveque, J.J. Simultaneous iterative reconstruction technique: physical interpretation based  
403 on the generalized least squares solution. *J. geophys. Res* **1990**, *95*, 553–9.
- 404 15. Bangliang, S.; Yiheng, Z.; Lihui, P.; Danya, Y.; Baofen, Z. The use of simultaneous iterative reconstruction  
405 technique for electrical capacitance tomography. *Chemical Engineering Journal* **2000**, *77*, 37–41.
- 406 16. Coric, S.; Leeser, M.; Miller, E.; Trepanier, M. Parallel-beam backprojection: an FPGA implementation  
407 optimized for medical imaging. Proceedings of the 2002 ACM/SIGDA tenth international symposium  
408 on Field-programmable gate arrays. ACM, 2002, pp. 217–226.
- 409 17. Ay, M.R.; Zaidi, H. Development and validation of MCNP4C-based Monte Carlo simulator for fan-and  
410 cone-beam x-ray CT. *Physics in Medicine and Biology* **2005**, *50*, 4863.
- 411 18. Mozzo, P.; Procacci, C.; Tacconi, A.; Tinazzi Martini, P.; Bergamo Andreis, I. A new volumetric CT machine  
412 for dental imaging based on the cone-beam technique: preliminary results. *European Radiology* **1998**,  
413 *8*, 1558–1564.
- 414 19. Ben-Israel, A.; Greville, T.N. *Generalized inverses: theory and applications*; Vol. 15, Springer Science &  
415 Business Media, 2003.
- 416 20. Calvetti, D.; Morigi, S.; Reichel, L.; Sgallari, F. Tikhonov regularization and the L-curve for large discrete  
417 ill-posed problems. *Journal of Computational and Applied Mathematics* **2000**, *123*, 423–446.
- 418 21. Chambolle, A.; Lions, P.L. Image recovery via total variation minimization and related problems.  
419 *Numerische Mathematik* **1997**, *76*, 167–188.
- 420 22. Goldstein, T.; Osher, S. The split Bregman method for L1-regularized problems. *SIAM Journal on Imaging*  
421 *Sciences* **2009**, *2*, 323–343.
- 422 23. Sidky, E.Y.; Pan, X. Image reconstruction in circular cone-beam computed tomography by constrained,  
423 total-variation minimization. *Physics in Medicine and Biology* **2008**, *53*, 4777.
- 424 24. Chan, T.F.; Golub, G.H.; Mulet, P. A nonlinear primal-dual method for total variation-based image  
425 restoration. *SIAM journal on scientific computing* **1999**, *20*, 1964–1977.
- 426 25. Wahlberg, B.; Boyd, S.; Annergren, M.; Wang, Y. An ADMM algorithm for a class of total variation  
427 regularized estimation problems. *IFAC Proceedings Volumes* **2012**, *45*, 83–88.
- 428 26. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical*  
429 *Society: Series B (Statistical Methodology)* **2005**, *67*, 301–320.
- 430 27. Ramani, S.; Liu, Z.; Rosen, J.; Nielsen, J.F.; Fessler, J.A. Regularization parameter selection for nonlinear  
431 iterative image restoration and MRI reconstruction using GCV and SURE-based methods. *IEEE*  
432 *Transactions on Image Processing* **2012**, *21*, 3659–3672.

- 433 28. Galatsanos, N.P.; Katsaggelos, A.K. Methods for choosing the regularization parameter and estimating  
434 the noise variance in image restoration and their relation. *IEEE Transactions on Image Processing* **1992**,  
435 *1*, 322–336.
- 436 29. Wang, G.; Schultz, L.; Qi, J. Statistical image reconstruction for muon tomography using a Gaussian scale  
437 mixture model. *IEEE Transactions on Nuclear Science* **2009**, *56*, 2480–2486.
- 438 30. Redner, R.A.; Walker, H.F. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*  
439 **1984**, *26*, 195–239.
- 440 31. Moon, T.K. The expectation-maximization algorithm. *IEEE Signal Processing Magazine* **1996**, *13*, 47–60.
- 441 32. Tregouet, D.; Escolano, S.; Turet, L.; Mallet, A.; Golmard, J. A new algorithm for haplotype-based  
442 association analysis: the Stochastic-EM algorithm. *Annals of human genetics* **2004**, *68*, 165–177.
- 443 33. Hudson, H.M.; Larkin, R.S. Accelerated image reconstruction using ordered subsets of projection data.  
444 *IEEE Transactions on Medical Imaging* **1994**, *13*, 601–609.
- 445 34. Gelman, A.; Carlin, J.B.; Stern, H.S.; Rubin, D.B. *Bayesian data analysis*; Vol. 2, Chapman & Hall/CRC Boca  
446 Raton, FL, USA, 2014.
- 447 35. Bali, N.; Mohammad-Djafari, A. Bayesian approach with hidden Markov modeling and mean field  
448 approximation for hyperspectral data analysis. *IEEE Transactions on Image Processing* **2008**, *17*, 217–225.
- 449 36. Mohammad-Djafari, A. Joint estimation of parameters and hyperparameters in a Bayesian approach of  
450 solving inverse problems. Image Processing, 1996. Proceedings., International Conference on. IEEE, 1996,  
451 Vol. 2, pp. 473–476.
- 452 37. Kolehmainen, V.; Vanne, A.; Siltanen, S.; Jarvenpaa, S.; Kaipio, J.P.; Lassas, M.; Kalke, M. Parallelized  
453 Bayesian inversion for three-dimensional dental X-ray imaging. *IEEE Transactions on Medical Imaging*  
454 **2006**, *25*, 218–228.
- 455 38. Qi, J.; Leahy, R.M. Resolution and noise properties of MAP reconstruction for fully 3-D PET. *IEEE*  
456 *Transactions on Medical Imaging* **2000**, *19*, 493–506.
- 457 39. Ericson, W.A. A note on the posterior mean of a population mean. *Journal of the Royal Statistical Society.*  
458 *Series B (Methodological)* **1969**, pp. 332–334.
- 459 40. Fox, C.W.; Roberts, S.J. A tutorial on variational Bayesian inference. *Artificial intelligence review* **2012**,  
460 *38*, 85–95.
- 461 41. Tzikas, D.G.; Likas, A.C.; Galatsanos, N.P. The variational approximation for Bayesian inference. *IEEE*  
462 *Signal Processing Magazine* **2008**, *25*, 131–146.
- 463 42. Ayasso, H.; Mohammad-Djafari, A. Joint NDT image restoration and segmentation using  
464 Gauss–Markov–Potts prior models and variational Bayesian computation. *IEEE Transactions on Image*  
465 *Processing* **2010**, *19*, 2265–2277.
- 466 43. Noël, P.B.; Walczak, A.M.; Xu, J.; Corso, J.J.; Hoffmann, K.R.; Schafer, S. GPU-based cone beam computed  
467 tomography. *Computer methods and programs in biomedicine* **2010**, *98*, 271–277.
- 468 44. Gac, N.; Mancini, S.; Desvignes, M.; Houzet, D. High speed 3D tomography on CPU, GPU, and FPGA.  
469 *EURASIP Journal on Embedded Systems* **2008**, *2008*, 5.
- 470 45. Wang, L.; Mohammad-Djafari, A.; Gac, N.; Dumitru, M. Computed tomography reconstruction based on  
471 a hierarchical model and variational Bayesian method. 2016 IEEE International Conference on Acoustics,  
472 Speech and Signal Processing (ICASSP). IEEE, 2016, pp. 883–887.
- 473 46. Dumitru, M. A Bayesian approach for periodic components estimation for chronobiological signals. PhD  
474 thesis, Paris Saclay, 2016.
- 475 47. Wang, L.; Mohammad-Djafari, A.; Gac, N. X-ray Computed Tomography using a sparsity enforcing  
476 prior model based on Haar transformation in a Bayesian framework. *Fundamenta Informaticae* **2017**,  
477 *155*, 449–480.
- 478 48. Stanković, R.S.; Falkowski, B.J. The Haar wavelet transform: its status and achievements. *Computers &*  
479 *Electrical Engineering* **2003**, *29*, 25–44.
- 480 49. Starck, J.L.; Candès, E.J.; Donoho, D.L. The curvelet transform for image denoising. *IEEE Transactions on*  
481 *Image Processing* **2002**, *11*, 670–684.
- 482 50. Do, M.N.; Vetterli, M. The contourlet transform: an efficient directional multiresolution image  
483 representation. *IEEE Transactions on Image Processing* **2005**, *14*, 2091–2106.
- 484 51. Selesnick, I.W.; Baraniuk, R.G.; Kingsbury, N.C. The dual-tree complex wavelet transform. *IEEE Signal*  
485 *Processing Magazine* **2005**, *22*, 123–151.

- 486 52. Nadarajah, S. A generalized normal distribution. *Journal of Applied Statistics* **2005**, *32*, 685–694.
- 487 53. Rasmussen, C.E. The infinite Gaussian mixture model. NIPS, 1999, Vol. 12, pp. 554–560.
- 488 54. Klebanov, L.B. *Heavy tailed distributions*; Matfyzpress, 2003.
- 489 55. Frese, T.; Bouman, C.A.; Sauer, K. Adaptive wavelet graph model for Bayesian tomographic  
490 reconstruction. *IEEE Transactions on Image Processing* **2002**, *11*, 756–770.
- 491 56. Delaney, A.H.; Bresler, Y. Multiresolution tomographic reconstruction using wavelets. *IEEE Transactions*  
492 *on Image Processing* **1995**, *4*, 799–813.
- 493 57. Rantala, M.; Vanska, S.; Jarvenpaa, S.; Kalke, M.; Lassas, M.; Moberg, J.; Siltanen, S. Wavelet-based  
494 reconstruction for limited-angle X-ray tomography. *IEEE Transactions on Medical Imaging* **2006**,  
495 *25*, 210–217.
- 496 58. Loris, I.; Nolet, G.; Daubechies, I.; Dahlen, F. Tomographic inversion using l1-norm regularization of  
497 wavelet coefficients. *Geophysical Journal International* **2007**, *170*, 359–370.
- 498 59. Sauer, K.; Bouman, C. A local update strategy for iterative reconstruction from projections. *IEEE*  
499 *Transactions on Signal Processing* **1993**, *41*, 534–548.
- 500 60. Elad, M.; Milanfar, P.; Rubinstein, R. Analysis versus synthesis in signal priors. *Inverse Problems* **2007**,  
501 *23*, 947.
- 502 61. Boyd, S.; Vandenberghe, L. *Convex optimization*; Cambridge university press, 2004.
- 503 62. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: from error visibility to  
504 structural similarity. *IEEE Transactions on Image Processing* **2004**, *13*, 600–612.
- 505 63. Palenstijn, W.J.; Batenburg, K.J.; Sijbers, J. The ASTRA tomography toolbox. 13th International  
506 Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE, 2013,  
507 Vol. 2013.
- 508 64. Dumitru, M.; Wang, L.; Gac, N.; Mohammad-Djafari, A. iterTomoGPI-GPL, Gitlab repository, Apr2018.  
509 [Online; [https://github.com/nicolasgac/TomoGPI\\_for\\_Astra](https://github.com/nicolasgac/TomoGPI_for_Astra)].
- 510 65. Figueiredo, M.A.; Nowak, R.D. Wavelet-based image estimation: an empirical Bayes approach using  
511 Jeffrey’s noninformative prior. *IEEE Transactions on Image Processing* **2001**, *10*, 1322–1331.
- 512 66. Dumitru, M.; Wang, L.; Gac, N.; Mohammad-Djafari, A. Performance Comparison Of Bayesian Iterative  
513 Algorithms For Three Classes Of Sparsity Enforcing Priors With Application In Computed Tomography.  
514 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017.