



**HAL**  
open science

# X-ray Computed Tomography using a sparsity enforcing prior model based on Haar transformation in a Bayesian framework

Li Wang, Ali Mohammad-Djafari, Nicolas Gac

► **To cite this version:**

Li Wang, Ali Mohammad-Djafari, Nicolas Gac. X-ray Computed Tomography using a sparsity enforcing prior model based on Haar transformation in a Bayesian framework. *Fundamenta Informaticae*, 2017, 155 (4), pp.449-480. 10.3233/FI-2017-1594 . hal-01490523

**HAL Id: hal-01490523**

**<https://hal.science/hal-01490523>**

Submitted on 15 Mar 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.

# X-ray Computed Tomography using a sparsity enforcing prior model based on Haar transformation in a Bayesian framework

**Li Wang\***

*Laboratoire des Signaux et Système*

*CentraleSupélec, CNRS, Université Paris Sud, Université Paris-Saclay*

*3, Rue Joliot-Curie, 91192 Gif-sur-Yvette, France*

*li.wang@lss.supelec.fr*

**Ali Mohammad-Djafari**

*Laboratoire des Signaux et Système*

*CentraleSupélec, CNRS, Université Paris Sud, Université Paris-Saclay*

*djafari@lss.supelec.fr*

**Nicolas Gac**

*Laboratoire des Signaux et Système*

*CentraleSupélec, CNRS, Université Paris Sud, Université Paris-Saclay*

*nicolas.gac@lss.supelec.fr*

---

**Abstract.** X-ray Computed Tomography (CT) has become a hot topic in both medical and industrial applications in recent decades. Reconstruction by using a limited number of projections is a significant research domain. In this paper, we propose to solve the X-ray CT reconstruction problem by using the Bayesian approach with a hierarchical structured prior model basing on the multilevel Haar transformation. In the proposed model, the multilevel Haar transformation is used as the sparse representation of a piecewise continuous image, and a generalized Student-t distribution is used to enforce its sparsity. The simulation results compare the performance of the proposed method with some state-of-the-art methods.

**Keywords:** Computed Tomography (CT), Bayesian Approach, Hierarchical Model, Generalized Student-t distribution, Joint Maximum A Posterior (JMAP)

## 1. Introduction

Computed Tomography is a technique widely used to achieve medical decisions [1, 2] and to conduct industrial Non-Destructive Testing (NDT) [3, 4]. In a tomography system, objects (2D or 3D) are observed by using different techniques: X-rays [5], ultrasound [6], microwave [7], infra-red [8], etc. Each observation process provides different type of information on the object under study. In [9], a pioneering introduction of the reconstruction of a three-dimensional density distribution from its projection data is given. The problem of reconstructing an object from its projections then appeared in various fields of research [5]. Other reconstruction methods are also proposed, for example the direct reconstruction methods proposed by Bracewell & Riddle [10] and Berry & Gibbs [11].

In X-ray CT, the measured data provides information on the structure of the density inside the object [12]. The main analytical method for CT reconstruction is the Filtered Back-Projection (FBP) [13, 14] based on the Radon transform [13, 15, 16], with its extension in 3D known as Feldkamp [17]. It performs fairly well when dealing with problems of reconstruction from sufficient data with a high SNR.

In order to use the least possible radiation dose in medical applications with respect to patients' care, or to spend less time for detection in industrial NDT, reconstruction is expected to be done with ever fewer projections. Then reconstruction becomes an ill-posed inverse problem [18, 19]. In this case, the conventional analytical reconstruction methods always result artefacts to the reconstructed image. Algebraic methods are used to solve this problem. Here we mention two classical methods: the Algebraic Reconstruction Technique (ART) [20, 21] and the Simultaneous Iterative Reconstruction Technique (SIRT) [22, 23]. These methods as well as many other related algorithms consider the discretized model:  $\mathbf{g} = \mathbf{H}\mathbf{f}$ , where  $\mathbf{f} \in \mathbb{R}^{N \times 1}$  represents the image to be reconstructed and  $\mathbf{g} \in \mathbb{R}^{M \times 1}$  represents the measured data. Matrix  $\mathbf{H} \in \mathbb{R}^{M \times N}$  is the linear projection operator, based either on a line integral or strip or fan beam or cone beam forward operator.

As in general the above-mentioned equation system is underdetermined, there is a need for the regularization methods. These methods approach the final result by considering the discretized model:

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

where  $\epsilon \in \mathbb{R}^{M \times 1}$  corresponds to the noise and the uncertainties of the system. These methods minimize a criterion consisting of two terms: a discrepancy of the observed data to the output of the forward model, very often  $\|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2$ , and a penalty term  $\Omega(\mathbf{f})$ . The solution is then defined as:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \left\{ \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda \Omega(\mathbf{f}) \right\}, \quad (2)$$

where  $\lambda$  is called the regularization parameter. Different regularization methods differ on the choice of  $\Omega(\mathbf{f})$ . When  $\Omega(\mathbf{f}) = 0$ , it refers to the Least-Squares (LS) method [24], with the drawback that the obtained solution may be too sensitive to the noise owing to the ill-posedness of the problem and the ill-conditioning of the operator  $\mathbf{H}$ . When  $\Omega(\mathbf{f}) = \|\mathbf{D}\mathbf{f}\|^2$ , the Quadratic Regularization (QR) method [25] is implied, where the linear operator  $\mathbf{D}$  is often a derivation operator. The famous Total Variation (TV) method [26, 27] is defined by restricting  $\Omega(\mathbf{f}) = \|\mathbf{D}_{TV}\mathbf{f}\|_1$  where  $\mathbf{D}_{TV}$  is the gradient operator. By using the  $l_1$  norm, the sparsity of the  $\mathbf{D}_{TV}\mathbf{f}$  term is enforced. The appearance of the non-differentiable  $l_1$  term leads to difficulties for the implementation of the optimization algorithm. A large number of methods have been proposed to solve this  $l_1$  norm optimization problem, for example the Newton's method [28], the Split Bregman method [27], etc. Many other efforts have been made to develop fast convergence algorithms to reach the solution of Eq.(2). We may mention here ADMM [29], ISTA [30], FISTA [31], etc. Other constraints exist to enforce the compact regions, for example the Mumford-Shah functional [32].

It should be noted that in the above-mentioned regularization methods, there is a parameter,  $\lambda$ , which controls the trade-off between the data-model adequacy and the regularization penalty. This parameter

needs to be fixed. There have been various methods for choosing suitable values for  $\lambda$ , for example the Cross Validation (CV) method and the L-curve method [33, 34]. However, the computation for choosing this regularization parameter needs to be done for each simulation and each different set of data, which is expensive and not practical. In some regularization methods, for example the Elastic Net method [35], there are two or even more such parameters to be fixed.

The data-model adequacy term  $\|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2$  can also take other different expressions, which depend on the noise model. When a Gaussian distribution is used to model the noise, the quadratic expression  $\|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2$  is obtained naturally. In Computed Tomography, in particular in Positrons Emission Tomography when the number of detected photons is very low, the noise may be better modelled by a Poisson distribution. To account for a more precise modelling of the noise, statistical methods have been proposed [36, 37]. The Maximum Likelihood (ML) methods [38] and their different algorithms such as the Expectation Maximization (EM) algorithms [39, 40], the stochastic EM (SEM) [41], the Ordered subsets-EM (OS-EM) [42] are used commonly in the PET and SPECT reconstruction problems.

In classic ML methods, it is difficult to account for any prior modelling. The Bayesian methods are then widely used [43]. The main strategy of the Bayesian inference is to assign a prior probability model to translate our prior knowledge and then to obtain the expression of the posterior distribution. The basic Bayesian formula is:

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

where  $p(\mathbf{g}|\mathbf{f})$  is the likelihood,  $p(\mathbf{f})$  the law a priori and  $p(\mathbf{f}|\mathbf{g})$  the law a posteriori. Once the posterior distribution is obtained, we can use it for any inference. In practice, however, we may be obliged to use point estimators such as maximum a posteriori (MAP) [44] or Posterior Mean (PM) [45]. Using a MAP estimator  $\hat{\mathbf{f}} = \arg \max_{\mathbf{f}} \{p(\mathbf{f}|\mathbf{g})\} = \arg \min_{\mathbf{f}} \{-\ln p(\mathbf{f}|\mathbf{g})\}$  can make the link between this estimator and almost all the regularization methods. A Gaussian prior for  $p(\mathbf{f})$  results in a Quadratic Regularization method, while a double exponential prior results in a Total Variation regularization method.

When applying the Bayesian approach to real applications, we may need to determine the parameters of the likelihood and the prior models. These parameters, often called hyper-parameters, also have to be estimated in order to achieve unsupervised or semi-supervised methods. Fortunately the Bayesian approach offers this possibility by obtaining the joint posterior probabilistic law:

$$p(\mathbf{f}, \boldsymbol{\theta}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f}, \boldsymbol{\theta}_1)p(\mathbf{f}|\boldsymbol{\theta}_2)p(\boldsymbol{\theta})}{p(\mathbf{g})}, \quad (4)$$

where the vector  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$  contains all the hyper-parameters to be estimated. This joint Posterior law can then be used in different ways to infer on  $\mathbf{f}$  and  $\boldsymbol{\theta}$ . One can first marginalize it to obtain  $p(\boldsymbol{\theta}|\mathbf{g}) = \int p(\mathbf{f}, \boldsymbol{\theta}|\mathbf{g}) d\mathbf{f}$  and estimate  $\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \{p(\boldsymbol{\theta}|\mathbf{g})\}$ , then use it as it was known. Another approach is the joint optimization, which estimates the variable  $\mathbf{f}$  and hyper-parameters  $\boldsymbol{\theta}$  simultaneously. The Variational Bayesian Approaching (VBA) method [46, 47] is often used during the point estimations.

Often, in order to distinguish the details of an object, a high resolution image is needed. In industrial applications, especially for the NDT of a large-sized object, the size of data becomes critical, and the projection and back-projection operators in CT become expensive. It is then necessary to account for the computational aspects, for example the use of the GPU processor [48–50].

In the previous work [51], we proposed using the Bayesian method via a synthesis model  $\mathbf{f} = \mathbf{D}\mathbf{z}$ , in which the Haar transformation coefficient  $\mathbf{z}$  of the image is first estimated, and then the final image reconstruction result is obtained from the inversion of the Haar transformation of the coefficients by  $\hat{\mathbf{f}} = \mathbf{D}\hat{\mathbf{z}}$ . When using a Double Exponential (DE) prior model for  $\mathbf{z}$  and the MAP estimator, the problem becomes equivalent to the optimization of  $\mathcal{J}(\mathbf{z}) = \|\mathbf{g} - \mathbf{H}\mathbf{D}\mathbf{z}\|^2 + \lambda \|\mathbf{z}\|_1$  which has been used very often in  $l_1$  regularization methods. The specificity of our work was to use a generalized Student-t model in place of DE. In this paper, we present a method which estimate the image and the transformation

coefficient simultaneously. The generalization of Student-t distribution is used to enforce the sparsity of the Haar transformation coefficient of the image. In this way, the edges of the reconstructed image is preserved compared with the previous work. From the reconstructed Haar transformation coefficient, we also obtain the high-frequency component (the edge) information of the image.

The rest of this paper is organized as follows. First, we present the Multilevel Haar transformation which is used in the proposed model. Then some sparsity-enforcing distributions and a generalization of Student-t distribution is presented. The proposed Bayesian method is then introduced with details. Finally, we give our simulation results and draw our conclusions.

*Notation:* Column vectors are represented as boldface lower-case letters, e.g.,  $\mathbf{x}$ , and matrices as boldface upper-case letters, e.g.,  $\mathbf{X}$ , while the  $i$ -th component of vector  $\mathbf{x}$  is denoted by  $x_i$  the  $ij$ -th element of matrix  $\mathbf{X}$  by  $X_{ij}$ . Moreover,  $(\cdot)^T$  denotes transposition,  $\|\cdot\|_1$  stands for the  $l_1$ -norm,  $\|\cdot\|$  stands for the standard  $l_2$ -norm,  $|\cdot|$  denotes the determinant of a matrix or absolute value in the case of a scalar,  $\mathcal{N}(\cdot)$  is the Gaussian distribution,  $\mathcal{IG}(\cdot)$  is the Inverse-Gamma (IG) distribution,  $\text{diag}[\mathbf{x}]$  denotes a diagonal matrix whose diagonal entries are the elements of  $\mathbf{x}$ .

## 2. The Hierarchical Haar transform-Based Method (HHBM)

In this paper, we propose a HHBM method, which considers to reconstruct X-ray CT images by using a Bayesian method. This method can be used to solve the CT reconstruction problems for piecewise continuous images consisting of several homogeneous blocs. By using this proposed method, we impose a hidden variable: the multilevel Haar transformation coefficient of the image. This transformation coefficient has a sparse property, which is defined in this method by a generalized Student-t distribution. With the simultaneous estimations of the coefficient and the image, the edges of the image are preserved.

### 2.1. Multilevel Haar Transformation

In a typical X-ray CT problem, the object under consideration consists of a finite number of homogeneous materials, leading to the fact that the images to be reconstructed are piecewise continuous. In each homogeneous bloc which corresponds to one material, the gradient of the image is zero. There are non-zeros values only at the edges between different blocs. In order to take into account this property, we use a sparsity-enforcing law to define a sparse transformation coefficient of the object.

The piece-wise continuous objects [52, 53] can be represented by different kinds of sparse transformation coefficients. In this paper, we use the multilevel Haar transformation, shown in Fig.(1), with the following properties:

- It is an orthogonal transformation [54, 55], meaning that if we note by  $\mathbf{f} = \mathbf{D}\mathbf{z}$  we have  $\mathbf{z} = \mathbf{D}^{-1}\mathbf{f} = \mathbf{D}^T\mathbf{f}$ , with the property  $\mathbf{D}^T = \mathbf{D}^{-1}$ .
- It has a multilevel structure that, after the transformation, only the low-pass component, which corresponds to the top-left small part in the coefficient image shown in Fig.(1), is not sparse.
- Both the forward and inverse operators of the transformation can be implemented easily for 2D and 3D cases. For an object of size  $N$ , the cost of computation is only  $2 \times o(\log N)$  or  $3 \times o(\log N)$  for 2D or 3D cases respectively.

The transformation operator can be represented by:

$$\mathbf{f} = \mathbf{D}\mathbf{z} + \boldsymbol{\xi}, \quad (5)$$

where  $\boldsymbol{\xi}$  represents the noise of this transformation.

For the multilevel Haar transformation, we define the transform level by  $l$ :  $f_{\mathcal{H}}^{(l)} = \mathcal{H}_l(f)$ . The definition of this operator is:

$$f_{\mathcal{H}}^{(i)} = \mathcal{H}_i(f) = \mathcal{H}aar \left( f_{\mathcal{H}}^{(i-1)} \left( 1 : \frac{N}{2^{i-1}}, 1 : \frac{N}{2^{i-1}} \right) \right), \quad (6)$$

$$f_{\mathcal{H}}^{(1)} = \mathcal{H}_1(f) = \mathcal{H}aar (f (1 : N, 1 : N)), \quad (7)$$

where  $\mathcal{H}aar(\cdot)$  represents the Haar transformation [56].

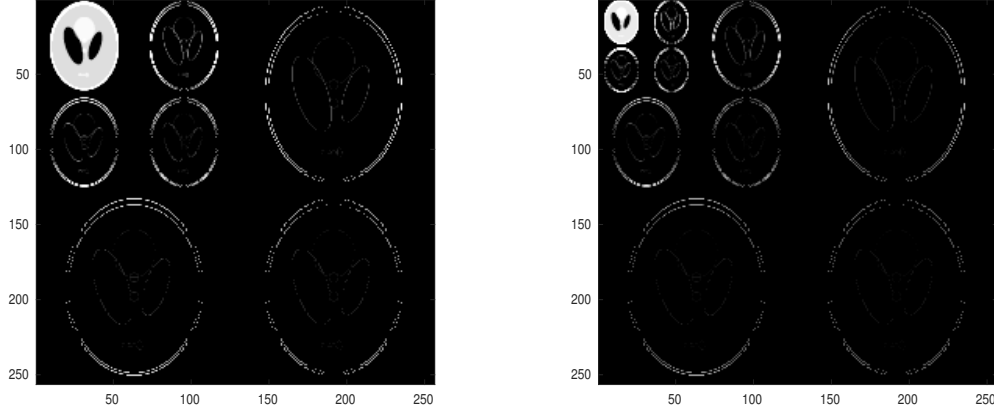


Figure 1: Multi-level Haar Transformation of the Shepp Logan image. Left: 2-level Haar transformation; right: 3-level Haar transformation.

## 2.2. Generalized Student-t Distribution

The sparsity of the multilevel Haar transformation coefficient  $z$  need to be defined by a statistical model. Typically the sparse property of a statistical variable is enforced by using three kinds of distributions: the Generalized Gaussian distributions [57], the Gaussian Mixture distributions [58, 59] and the heavy tailed distributions [60, 61]. The Student-t distribution [62, 63] is heavy tailed. We present here a generalization of this distribution [64], which has some good properties for the definition of the prior model.

The standard Student-t distribution has the probability density function (pdf):

$$\mathcal{S}t(x|v) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\Gamma(\frac{v}{2})} \left( 1 + \frac{x^2}{v} \right)^{-\frac{v+1}{2}}, \quad (8)$$

of which the variance depends only on the degree of freedom  $v$ :

$$\text{Var}[x] = \begin{cases} \frac{v}{v-2} & \text{for } v > 2 \\ \infty & \text{for } 1 < v < 2 \\ \text{undefined} & \text{otherwise} \end{cases}. \quad (9)$$

From the definition we easily determine that there is a limit for the variance of the Student-t distribution:  $\text{Var}[x] = \frac{v}{v-2} = 1 + \frac{2}{v-2} > 1$ . This limit leads to the consequence that this heavy-tailed distribution cannot have a small variance, therefore the sparsity could not be intensively enforced.

We introduce the generalized Student-t distribution (Stg). It comes from the Normal-Gamma marginalization property:  $\mathcal{S}t_g(x|\alpha, \beta) = \int \mathcal{N}(x|0, 1/v) \mathcal{G}(v|\alpha, \beta) dv$  or the Normal-Inverse Gamma marginalization property:  $\mathcal{S}t_g(x|\alpha, \beta) = \int \mathcal{N}(x|0, v) \mathcal{IG}(v|\alpha, \beta) dv$  of the Student-t distribution

**Theorem 2.1.** For two random variables  $x$  and  $v$  following a bi-variate Normal-Inverse Gamma distribution:  $(x, v) \sim \mathcal{N}(x|0, v) \mathcal{IG}(v|\alpha, \beta)$ , the marginal distribution corresponding to  $x$  is a Student-t like distribution, generalizing the classical form, denoted  $St_g$ , depending on the two parameters  $\alpha$  and  $\beta$  and having the following pdf:

$$\begin{aligned} St_g(x|\alpha, \beta) &= \int \mathcal{N}(x|0, v) \mathcal{IG}(v|\alpha, \beta) dv \\ &= \frac{\Gamma(\alpha + \frac{1}{2})}{\sqrt{2\pi\beta}\Gamma(\alpha)} \left(1 + \frac{x^2}{2\beta}\right)^{-(\alpha + \frac{1}{2})}. \end{aligned} \quad (10)$$

This generalization of the Student-t distribution adds a supplementary parameter to the standard one. More possibilities of the distribution shape are therefore available. The shape of the peak and tail of the distribution can be controlled by changing the values of these two parameters. We can either get a super heavy-tailed distribution with a super narrow peak, or a distribution that approaches the Normal distribution, by setting different pair of parameters  $\alpha$  and  $\beta$ .

**Remark 2.2.** If  $x$  belong to a generalized Student-t distribution, then it has the expectation and variance:

$$E[x] = \begin{cases} 0 & \text{for } \alpha > \frac{1}{2} \\ \text{undefined} & \text{otherwise} \end{cases}, \quad (11)$$

and

$$\text{Var}[x] = \begin{cases} \frac{\beta}{\alpha-1} & \text{for } \alpha > 1 \\ \infty & \text{for } \frac{1}{2} < \alpha \leq 1 \\ \text{undefined} & \text{for } 0 < \alpha \leq \frac{1}{2} \end{cases}. \quad (12)$$

With this property, we can use the  $St_g$  distribution to define the hidden variables, the multilevel Haar transformation coefficient in the HHBM method, or other sparse representations of the image.

In this paper, we use this generalization of Student-t distribution to define the sparse coefficient. First of all, by commanding the two hyper-parameters, we can control the sparsity rate of the coefficient. What's more, thanks to the fact that it is a marginalization of two conjugate distributions, it is more convenient for us to apply different estimation methods, for example the Posterior Mean (PM) via Variational Bayesian Approximation (VBA).

### 2.3. Hierarchical Bayesian Model of HHBM

The Bayesian method attempts to solve the reconstruction problem statistically based on the forward model presented in Eq.(1). By using the Bayesian method, first of all we define the probabilistic model for all the prior information.

The additive noise  $\epsilon$  in this model is defined as belonging to a Gaussian distribution, with mean equal to zero and variance  $\mathbf{v}_\epsilon$ :

$$p(\epsilon|\mathbf{v}_\epsilon) = \mathcal{N}(\epsilon|\mathbf{0}, \mathbf{V}_\epsilon), \quad (13)$$

where  $\mathbf{V}_\epsilon = \text{diag}[\mathbf{v}_\epsilon]$  is a diagonal matrix. According to the linear forward model, the distribution of the additive noise provides the likelihood:

$$p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon) = \mathcal{N}(\mathbf{g}|\mathbf{H}\mathbf{f}, \mathbf{V}_\epsilon). \quad (14)$$

For the multilevel Haar transformation of the image, the likelihood corresponding to the forward model in Eq.(5) is given below:

$$p(\mathbf{f}|\mathbf{z}, \mathbf{v}_\xi) = \mathcal{N}(\mathbf{f}|\mathbf{D}\mathbf{z}, \mathbf{V}_\xi), \quad (15)$$

where  $\mathbf{D}$  represents the inverse Multilevel Haar transformation operator, and the additive noise  $\xi$  is considered to belong to a Gaussian distribution, with mean equal to zero and variance  $\mathbf{v}_\xi$ . The definition of the covariance matrix is:  $\mathbf{V}_\xi = \text{diag}[\mathbf{v}_\xi]$ .

The vector  $\mathbf{z} = [z_1, z_2, \dots, z_N]$ , as mentioned above, is sparse. As discussed in Section (2.2), the generalized Student-t distribution is used to define  $\mathbf{z}$  in order to enforce the sparsity property. By using the Normal-Inverse Gamma property shown in Eq.(10), the prior distribution for  $\mathbf{z}$  can be realized by a hierarchical structured Normal-Inverse Gamma distribution, where  $\mathbf{z}$  belongs to a Normal distribution, the variance of which belongs to an Inverse Gamma distribution depending on two hyper-parameters  $\alpha_{z_0}$  and  $\beta_{z_0}$ :

$$p(\mathbf{z}|\mathbf{v}_z) = \mathcal{N}(\mathbf{z}|0, \mathbf{V}_z), \quad (16)$$

$$p(\mathbf{v}_z|\alpha_{z_0}, \beta_{z_0}) = \prod_{j=1}^N \mathcal{IG}(v_{z_j}|\alpha_{z_0}, \beta_{z_0}), \quad (17)$$

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

Moreover, we need to decide the probabilistic model for the variances of the two noises:  $\mathbf{v}_\epsilon$  and  $\mathbf{v}_\xi$  in this unsupervised model. Knowing that the variance is always positive, and that typically the value of the variance is small, we choose the Inverse Gamma distribution to model them:

$$p(\mathbf{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \prod_{i=1}^M \mathcal{IG}(v_{\epsilon_i}|\alpha_{\epsilon_0}, \beta_{\epsilon_0}), \quad (18)$$

$$p(\mathbf{v}_\xi|\alpha_{\xi_0}, \beta_{\xi_0}) = \prod_{j=1}^N \mathcal{IG}(v_{\xi_j}|\alpha_{\xi_0}, \beta_{\xi_0}), \quad (19)$$

where the elements in each vector,  $\mathbf{v}_\epsilon$  and  $\mathbf{v}_\xi$ , are supposed to be i.i.d.  $\alpha_{\epsilon_0}$ ,  $\beta_{\epsilon_0}$ ,  $\alpha_{\xi_0}$  and  $\beta_{\xi_0}$  are hyper-parameters for these distributions respectively.

As long as we have defined the models for all the prior distributions, the model of the hierarchical system containing all the variables, parameters and hyper-parameters is:

$$p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon) = \mathcal{N}(\mathbf{g}|\mathbf{H}\mathbf{f}, \mathbf{V}_\epsilon) \propto |\mathbf{V}_\epsilon|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\mathbf{g} - \mathbf{H}\mathbf{f})^T \mathbf{V}_\epsilon^{-1} (\mathbf{g} - \mathbf{H}\mathbf{f}) \right] \quad (20)$$

$$p(\mathbf{f}|\mathbf{z}, \mathbf{v}_\xi) = \mathcal{N}(\mathbf{f}|\mathbf{D}\mathbf{z}, \mathbf{V}_\xi) \propto |\mathbf{V}_\xi|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\mathbf{f} - \mathbf{D}\mathbf{z})^T \mathbf{V}_\xi^{-1} (\mathbf{f} - \mathbf{D}\mathbf{z}) \right] \quad (21)$$

$$p(\mathbf{z}|\mathbf{v}_z) = \mathcal{N}(\mathbf{z}|0, \mathbf{V}_z) \propto |\mathbf{V}_z|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \mathbf{z}^T \mathbf{V}_z^{-1} \mathbf{z} \right] \quad (22)$$

$$p(\mathbf{v}_z|\alpha_{z_0}, \beta_{z_0}) = \prod_{j=1}^N \mathcal{IG}(v_{z_j}|\alpha_{z_0}, \beta_{z_0}) \propto \prod_{j=1}^N v_{z_j}^{-(\alpha_{z_0}+1)} \exp \left[ -\beta_{z_0} v_{z_j}^{-1} \right] \quad (23)$$

$$p(\mathbf{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \prod_{i=1}^M \mathcal{IG}(v_{\epsilon_i}|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \propto \prod_{i=1}^M v_{\epsilon_i}^{-(\alpha_{\epsilon_0}+1)} \exp \left[ -\beta_{\epsilon_0} v_{\epsilon_i}^{-1} \right] \quad (24)$$

$$p(\mathbf{v}_\xi|\alpha_{\xi_0}, \beta_{\xi_0}) = \prod_{j=1}^N \mathcal{IG}(v_{\xi_j}|\alpha_{\xi_0}, \beta_{\xi_0}) \propto \prod_{j=1}^N v_{\xi_j}^{-(\alpha_{\xi_0}+1)} \exp \left[ -\beta_{\xi_0} v_{\xi_j}^{-1} \right] \quad (25)$$

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



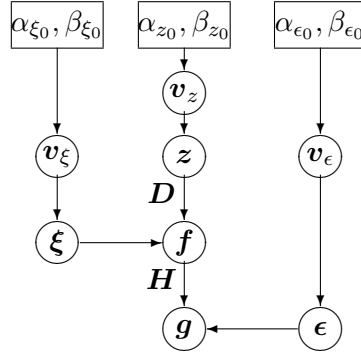


Figure 2: DAG of the proposed HHBM model.

distribution is then obtained from Equations Eq.(20)-Eq.(25) via unsupervised Bayesian formula:

$$\begin{aligned}
p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\xi, \mathbf{v}_z | \mathbf{g}) &\propto p(\mathbf{g} | \mathbf{f}, \mathbf{v}_\epsilon) p(\mathbf{f} | \mathbf{z}, \mathbf{v}_\xi) p(\mathbf{z} | \mathbf{v}_z) p(\mathbf{v}_\epsilon) p(\mathbf{v}_\xi) p(\mathbf{v}_z) \\
&= |\mathbf{V}_\epsilon|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\mathbf{g} - \mathbf{H}\mathbf{f})^T \mathbf{V}_\epsilon^{-1} (\mathbf{g} - \mathbf{H}\mathbf{f}) \right] |\mathbf{V}_\xi|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\mathbf{f} - \mathbf{D}\mathbf{z})^T \mathbf{V}_\xi^{-1} (\mathbf{f} - \mathbf{D}\mathbf{z}) \right] \\
&\quad |\mathbf{V}_z|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \mathbf{z}^T \mathbf{V}_z^{-1} \mathbf{z} \right] \prod_{j=1}^N \left\{ v_{z_j}^{-(\alpha_{z_0}+1)} \exp \left[ -\beta_{z_0} v_{z_j}^{-1} \right] \right\} \prod_{i=1}^M \left\{ v_{\epsilon_i}^{-(\alpha_{\epsilon_0}+1)} \exp \left[ -\beta_{\epsilon_0} v_{\epsilon_i}^{-1} \right] \right\} \\
&\quad \prod_{j=1}^N \left\{ v_{\xi_j}^{-(\alpha_{\xi_0}+1)} \exp \left[ -\beta_{\xi_0} v_{\xi_j}^{-1} \right] \right\}
\end{aligned} \tag{26}$$

Once the posterior distribution has been obtained, estimation methods, for example the MAP algorithm [44] or the posterior mean (PM) via variational Bayesian approach (VBA) [46,47] method, are then used to approach the reconstruction result. In this paper we use the Joint MAP method which iteratively estimates all the variables and parameters.

### 3. Joint Maximum A Posterior (JMAP) Algorithm for HHBM

The JMAP algorithm aims at iteratively and alternately estimating all the variables and parameters by maximizing the posterior distribution, or minimizing the negative logarithm of the posterior law:

$$\begin{aligned}
(\hat{\mathbf{f}}, \hat{\mathbf{z}}, \hat{\mathbf{v}}_\epsilon, \hat{\mathbf{v}}_\xi, \hat{\mathbf{v}}_z) &= \arg \max_{\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\xi, \mathbf{v}_z} \{p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\xi, \mathbf{v}_z | \mathbf{g})\} \\
&= \arg \min_{\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\xi, \mathbf{v}_z} \{-\ln p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\xi, \mathbf{v}_z | \mathbf{g})\}.
\end{aligned} \tag{27}$$

The criterion of the minimization,  $\mathcal{J}(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\xi, \mathbf{v}_z) = -\ln p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\xi, \mathbf{v}_z | \mathbf{g})$ , is then obtained

from Eq.(26):

$$\begin{aligned}
\mathcal{J}(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\xi, \mathbf{v}_z) &= \frac{1}{2} \sum_{i=1}^M \ln v_{\epsilon_i} + \frac{1}{2} (\mathbf{g} - \mathbf{H}\mathbf{f})^T \mathbf{V}_\epsilon^{-1} (\mathbf{g} - \mathbf{H}\mathbf{f}) + \frac{1}{2} \sum_{j=1}^N \ln v_{\xi_j} \\
&+ \frac{1}{2} (\mathbf{f} - \mathbf{D}\mathbf{z})^T \mathbf{V}_\xi^{-1} (\mathbf{f} - \mathbf{D}\mathbf{z}) + \frac{1}{2} \sum_{j=1}^N \ln v_{z_j} + \frac{1}{2} \mathbf{z}^T \mathbf{V}_z^{-1} \mathbf{z} \\
&+ (\alpha_{z_0} + 1) \sum_{j=1}^N \ln v_{z_j} + \sum_{j=1}^N \beta_{z_0} v_{z_j}^{-1} + (\alpha_{\epsilon_0} + 1) \sum_{i=1}^M \ln v_{\epsilon_i} + \sum_{i=1}^M \beta_{\epsilon_0} v_{\epsilon_i}^{-1} \\
&+ (\alpha_{\xi_0} + 1) \sum_{j=1}^N \ln v_{\xi_j} + \sum_{j=1}^N \beta_{\xi_0} v_{\xi_j}^{-1}.
\end{aligned} \tag{28}$$

By using the JMAP algorithm, we estimate all the variables by minimizing the corresponding terms in the constraint equation Eq.(28). While considering one variable, we keep all the other variables fixed, equal to the estimated value of the previous iteration.

$$\hat{\mathbf{f}}^{(k)} = \arg \min_{\mathbf{f}} \left\{ \mathcal{J}(\mathbf{f}, \hat{\mathbf{z}}^{(k-1)}, \hat{\mathbf{v}}_z^{(k-1)}, \hat{\mathbf{v}}_\epsilon^{(k-1)}, \hat{\mathbf{v}}_\xi^{(k-1)}) \right\}, \tag{29}$$

$$\hat{\mathbf{z}}^{(k)} = \arg \min_{\mathbf{z}} \left\{ \mathcal{J}(\hat{\mathbf{f}}^{(k)}, \mathbf{z}, \hat{\mathbf{v}}_z^{(k-1)}, \hat{\mathbf{v}}_\epsilon^{(k-1)}, \hat{\mathbf{v}}_\xi^{(k-1)}) \right\}, \tag{30}$$

$$\hat{\mathbf{v}}_z^{(k)} = \arg \min_{\mathbf{v}_z} \left\{ \mathcal{J}(\hat{\mathbf{f}}^{(k)}, \hat{\mathbf{z}}^{(k)}, \mathbf{v}_z, \hat{\mathbf{v}}_\epsilon^{(k-1)}, \hat{\mathbf{v}}_\xi^{(k-1)}) \right\}, \tag{31}$$

$$\hat{\mathbf{v}}_\epsilon^{(k)} = \arg \min_{\mathbf{v}_\epsilon} \left\{ \mathcal{J}(\hat{\mathbf{f}}^{(k)}, \hat{\mathbf{z}}^{(k)}, \hat{\mathbf{v}}_z^{(k)}, \mathbf{v}_\epsilon, \hat{\mathbf{v}}_\xi^{(k-1)}) \right\}, \tag{32}$$

$$\hat{\mathbf{v}}_\xi^{(k)} = \arg \min_{\mathbf{v}_\xi} \left\{ \mathcal{J}(\hat{\mathbf{f}}^{(k)}, \hat{\mathbf{z}}^{(k)}, \hat{\mathbf{v}}_z^{(k)}, \hat{\mathbf{v}}_\epsilon^{(k)}, \mathbf{v}_\xi) \right\}, \tag{33}$$

where  $\cdot^{(k)}$  represents the estimated value of the  $k$ th iteration. The variables with the hat symbol mean that in the current step they are considered as knowns. We then get the expressions for the minimization constraints as:

$$\hat{\mathbf{f}}^{(k+1)} = \arg \min_{\mathbf{f}} \left\{ \frac{1}{2} (\mathbf{g} - \mathbf{H}\mathbf{f})^T (\widehat{\mathbf{V}}_\epsilon^{(k)})^{-1} (\mathbf{g} - \mathbf{H}\mathbf{f}) + \frac{1}{2} (\mathbf{f} - \mathbf{D}\hat{\mathbf{z}}^{(k)})^T (\widehat{\mathbf{V}}_\xi^{(k)})^{-1} (\mathbf{f} - \mathbf{D}\hat{\mathbf{z}}^{(k)}) \right\}, \tag{34}$$

$$\hat{\mathbf{z}}^{(k+1)} = \arg \min_{\mathbf{z}} \left\{ \frac{1}{2} (\hat{\mathbf{f}}^{(k+1)} - \mathbf{D}\mathbf{z})^T (\widehat{\mathbf{V}}_\xi^{(k)})^{-1} (\hat{\mathbf{f}}^{(k+1)} - \mathbf{D}\mathbf{z}) + \frac{1}{2} \mathbf{z}^T (\widehat{\mathbf{V}}_z^{(k)})^{-1} \mathbf{z} \right\}, \tag{35}$$

$$\hat{\mathbf{v}}_z^{(k+1)} = \arg \min_{\mathbf{v}_z} \left\{ \frac{1}{2} \sum_{j=1}^N \ln v_{z_j} + \frac{1}{2} (\hat{\mathbf{z}}^{(k+1)})^T \mathbf{V}_z^{-1} \hat{\mathbf{z}}^{(k+1)} + (\alpha_{z_0} + 1) \sum_{j=1}^N \ln v_{z_j} + \sum_{j=1}^N \beta_{z_0} v_{z_j}^{-1} \right\}, \tag{36}$$

$$\hat{\mathbf{v}}_\epsilon^{(k+1)} = \arg \min_{\mathbf{v}_\epsilon} \left\{ \frac{1}{2} \sum_{i=1}^M \ln v_{\epsilon_i} + \frac{1}{2} (\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}^{(k+1)})^T \mathbf{V}_\epsilon^{-1} (\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}^{(k+1)}) + (\alpha_{\epsilon_0} + 1) \sum_{i=1}^M \ln v_{\epsilon_i} + \sum_{i=1}^M \beta_{\epsilon_0} v_{\epsilon_i}^{-1} \right\}, \tag{37}$$

$$\hat{\mathbf{v}}_\xi^{(k+1)} = \arg \min_{\mathbf{v}_\xi} \left\{ \frac{1}{2} \sum_{j=1}^N \ln v_{\xi_j} + \frac{1}{2} (\hat{\mathbf{f}}^{(k+1)} - \mathbf{D}\hat{\mathbf{z}}^{(k+1)})^T \mathbf{V}_\xi^{-1} (\hat{\mathbf{f}}^{(k+1)} - \mathbf{D}\hat{\mathbf{z}}^{(k+1)}) + (\alpha_{\xi_0} + 1) \sum_{j=1}^N \ln v_{\xi_j} + \sum_{j=1}^N \beta_{\xi_0} v_{\xi_j}^{-1} \right\} \tag{38}$$

### 3.1. Estimate $\mathbf{f}$ and $\mathbf{z}$

For the estimations of the object  $\mathbf{f}$  and the transformation coefficient  $\mathbf{z}$ , the updating equations obtained from the minimization in Eq.(34) and Eq.(35) are:

$$\hat{\mathbf{f}} = \left( \mathbf{H}^T \mathbf{V}_\epsilon^{-1} \mathbf{H} + \mathbf{V}_\xi^{-1} \right)^{-1} \left( \mathbf{H}^T \mathbf{V}_\epsilon^{-1} \mathbf{g} + \mathbf{V}_\xi^{-1} \mathbf{D} \mathbf{z} \right), \quad (39)$$

$$\hat{\mathbf{z}} = \left( \mathbf{D}^T \mathbf{V}_\xi^{-1} \mathbf{D} + \mathbf{V}_z^{-1} \right)^{-1} \mathbf{D}^T \mathbf{V}_\xi^{-1} \hat{\mathbf{f}}. \quad (40)$$

As we have already said, in X-ray CT, the size of the object  $\mathbf{f}$  and the size of its transformation coefficient  $\mathbf{z}$  are very large. Hence the computation of the matrix inversion in Eq.(39) and Eq.(40) is too expensive and even impossible to compute. In this paper we present the gradient descent algorithm for these estimations.

$$\text{for } k = 1 \rightarrow \text{Iter2} : \hat{\mathbf{f}}^{(k+1)} = \hat{\mathbf{f}}^{(k)} - \hat{\gamma}_f^{(k)} \nabla \mathcal{J}(\hat{\mathbf{f}}^{(k)}), \quad (41)$$

$$\text{for } k = 1 \rightarrow \text{Iter2} : \hat{\mathbf{z}}^{(k+1)} = \hat{\mathbf{z}}^{(k)} - \hat{\gamma}_z^{(k)} \nabla \mathcal{J}(\hat{\mathbf{z}}^{(k)}), \quad (42)$$

where *Iter2* is the number of iterations for the descent gradient step,  $\nabla \mathcal{J}(\cdot)$  is the corresponding gradient of the constraint equation, and  $\hat{\gamma}_{(\cdot)}$  is the corresponding descent step length:

$$\nabla \mathcal{J}(\mathbf{f}) = -\mathbf{H}^T \mathbf{V}_\epsilon^{-1} (\mathbf{g} - \mathbf{H} \mathbf{f}) + \mathbf{V}_\xi^{-1} (\mathbf{f} - \mathbf{D} \mathbf{z}), \quad (43)$$

$$\nabla \mathcal{J}(\mathbf{z}) = -\mathbf{D}^T \mathbf{V}_\xi^{-1} (\mathbf{f} - \mathbf{D} \mathbf{z}) + \mathbf{V}_z^{-1} \mathbf{z}, \quad (44)$$

$$\hat{\gamma}_f^{(k)} = \left( \left\| \nabla \mathcal{J}(\hat{\mathbf{f}}^{(k)}) \right\|^2 \right) / \left( \left\| \hat{\mathbf{Y}}_\epsilon \mathbf{H} \nabla \mathcal{J}(\hat{\mathbf{f}}^{(k)}) \right\|^2 + \left\| \hat{\mathbf{Y}}_\xi \nabla \mathcal{J}(\hat{\mathbf{f}}^{(k)}) \right\|^2 \right), \quad (45)$$

$$\hat{\gamma}_z^{(k)} = \left( \left\| \nabla \mathcal{J}(\hat{\mathbf{z}}^{(k)}) \right\|^2 \right) / \left( \left\| \hat{\mathbf{Y}}_\xi \mathbf{D} \nabla \mathcal{J}(\hat{\mathbf{z}}^{(k)}) \right\|^2 + \left\| \hat{\mathbf{Y}}_z \nabla \mathcal{J}(\hat{\mathbf{z}}^{(k)}) \right\|^2 \right). \quad (46)$$

The descent step length  $\hat{\gamma}_f$  and  $\hat{\gamma}_z$  are obtained by using an optimized step length strategy, see [65]. The calculation details are given in Appendix A. The computation algorithms for updating  $\mathbf{f}$  and  $\mathbf{z}$  are given in Algo.(1) and Algo.(2) respectively.

---

#### Algorithm 1 Gradient descent algorithm for $\hat{\mathbf{f}}$

---

- 1: **Input:**  $\mathbf{H}, \mathbf{D}, \mathbf{g}, \hat{\mathbf{z}}$
  - 2: **Output:**  $\hat{\mathbf{f}}$
  - 3: **Initialization:**  $\hat{\mathbf{f}}^{(0)}$
  - 4: **for**  $it = 1$  to *Iter2* **do**
  - 5:     Set  $k = it$
  - 6:     Calculate  $\nabla \mathcal{J}(\hat{\mathbf{f}}^{(k-1)})$  according to Eq.(43)
  - 7:     Update  $\hat{\gamma}_f^{(k)}$  according to Eq.(45)
  - 8:     Update  $\hat{\mathbf{f}}^{(k)} = \hat{\mathbf{f}}^{(k-1)} - \hat{\gamma}_f^{(k)} \nabla \mathcal{J}(\hat{\mathbf{f}}^{(k-1)})$
  - 9: **end for**
-

**Algorithm 2** Gradient descent algorithm for  $\hat{\mathbf{z}}$ 

- 
- 1: **Input:**  $H, D, g, \hat{\mathbf{f}}$
  - 2: **Output:**  $\hat{\mathbf{z}}$
  - 3: **Initialization:**  $\hat{\mathbf{z}}^{(0)}$
  - 4: **for**  $it = 1$  to  $Iter2$  **do**
  - 5:     Set  $k = it$
  - 6:     Calculate  $\nabla \mathcal{J}(\hat{\mathbf{z}}^{(k-1)})$  according to Eq.(44)
  - 7:     Update  $\hat{\gamma}_z^{(k)}$  according to Eq.(46)
  - 8:     Update  $\hat{\mathbf{z}}^{(k)} = \hat{\mathbf{z}}^{(k-1)} - \hat{\gamma}_z^{(k)} \nabla \mathcal{J}(\hat{\mathbf{z}}^{(k-1)})$
  - 9: **end for**
- 

**3.2. Estimation of the other variables**

The iterative updating rules for the other variables are obtained by calculating the minimizations in Eq.(36)-Eq.(38) and are given as the following formulas:

$$\hat{v}_{z_j} = \left( \beta_{z_0} + \frac{1}{2} \hat{z}_j^2 \right) / (\alpha_{z_0} + 3/2), \quad \forall j \in [1, N], \quad (47)$$

$$\hat{v}_{\epsilon_i} = \left( \beta_{\epsilon_0} + \frac{1}{2} \left( g_i - [\mathbf{H}\hat{\mathbf{f}}]_i \right)^2 \right) / (\alpha_{\epsilon_0} + 3/2), \quad \forall i \in [1, M], \quad (48)$$

$$\hat{v}_{\xi_j} = \left( \beta_{\xi_0} + \frac{1}{2} \left( \hat{f}_j - [\mathbf{D}\hat{\mathbf{z}}]_j \right)^2 \right) / (\alpha_{\xi_0} + 3/2), \quad \forall j \in [1, N]. \quad (49)$$

**3.3. Summary of the algorithm**

Now we have obtained all the updating equations for all the variables. In the computation, we still need to initialize  $\mathbf{f}$  and  $\mathbf{z}$ . For the initialization of  $\mathbf{f}$ , we use the least square estimation:

$$\hat{\mathbf{f}}_{ini} = \arg \min_{\mathbf{f}} \left\{ \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 \right\} \quad (50)$$

with a very limited number of iterations.

Then we initialize  $\mathbf{z}$  as the  $l$ -level Haar transformation coefficient of  $\hat{\mathbf{f}}_{ini}$ :  $\hat{\mathbf{z}}_{ini} = \mathbf{D}^{-1} \hat{\mathbf{f}}_{ini}$ . The initialization of the other unknowns is obtained from Eq.(36)-Eq.(38) by using  $\hat{\mathbf{f}}_{ini}$  and  $\hat{\mathbf{z}}_{ini}$ . The Algo.(3) gives the details of the JMAP algorithm for the HHBM method.

**4. Simulations**

In the simulations, the 2D Shepp Logan phantom of size  $64 \times 64$  and  $256 \times 256$  are used as the original images. Fig.(3)(a) shows the image of size  $256 \times 256$ . It consists of several homogeneous blocs, each of which corresponds to a different material. The projections are obtained from angles uniformly distributed from  $0^\circ$  to  $180^\circ$ , no matter how many number of projections we apply.

The comparison is between a) the proposed HHBM method, b) the Quadratic Regularization (QR) method [25], which is a basic regularization method and c) the Total Variation (TV) method [26], which performs fairly well when treating piecewise continuous image reconstruction problems. The reconstruction by using the TV method is realized by using the previously mentioned Split Bregman algorithm [27]. The gradient descent optimization algorithm is used in the HHBM method as mentioned in Eq.(41) and Eq.(42).

**Algorithm 3** Summary of 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:**  $H, D, g$
  - 3: **Output:**  $\hat{f}, \hat{z}, \hat{v}_z, \hat{v}_\epsilon, \hat{v}_\xi$
  - 4: **Initialization:**
  - 5:  $\hat{f}^{(0)} \leftarrow \arg \min_f \{ \|g - Hf\|^2 \}$  with a few iterations
  - 6:  $\hat{z}^{(0)} \leftarrow l$ -level Haar transformation of  $\hat{f}^{(0)}$
  - 7: **for**  $it = 1$  to  $Itmax$  **do**
  - 8:     Update  $\hat{f}$  according to Algo (1)
  - 9:     Update  $\hat{z}$  according to Algo (2)
  - 10:    Compute  $v_z, v_\epsilon, v_\xi$  according to Eq.(47)-Eq.(49)
  - 11: **end for**
- 

In the simulations, we have tested with 4 and 5 and 6 levels of transformation, and we found that this transformation level doesn't have big influence on the cost of computation and the final performance. In the simulations, we use a 5-level Haar transformation as the sparse coefficient of the image, shown in Fig.(3)(b). The sinogram of the projection of the 2D  $256 \times 256$  Shepp Logan phantom with 128 projections and a high SNR=40dB is shown in Fig.(3)(c), each vertical line of which corresponds to the detected data of one projection direction. In Fig.(3)(d), the sinogram of projection with a low SNR=20dB is displayed.

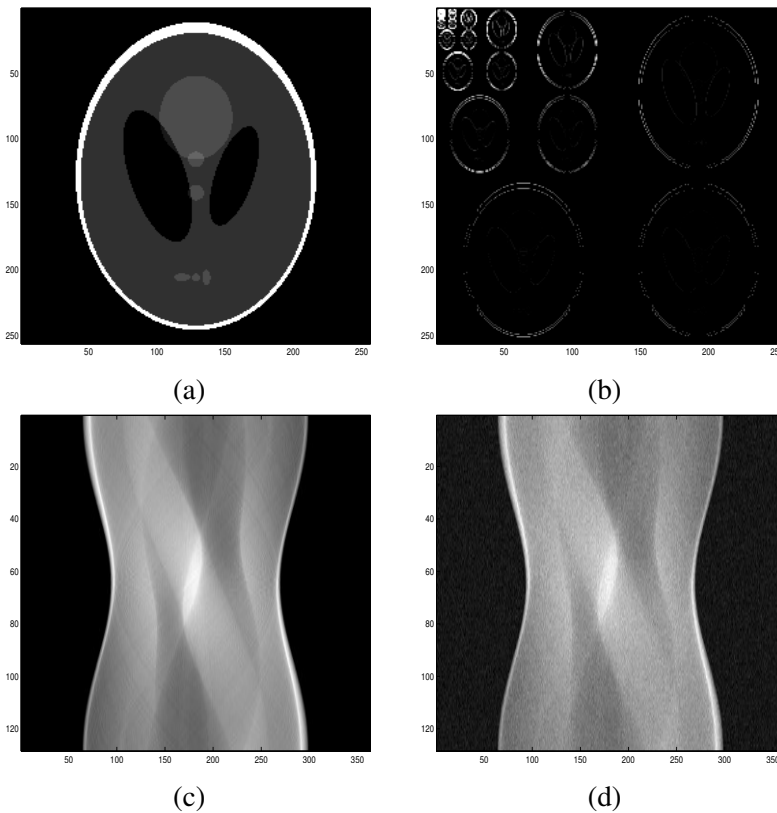


Figure 3: (a) original 2D  $256 \times 256$  Shepp Logan phantom, (b) its 5-level Haar transformation, (c) its sinogram of projection with 128 projections and a high SNR=40dB and (d) its sinogram of projection with 128 projection and a low SNR=20dB.

## 4.1. Measurement of performance

A large number of measurements can be used for the evaluation of the reconstruction performance of the different methods. In this paper, we consider three evaluations: the relative error  $\delta_f$ , the improvement of the signal-to-noise ratio (ISNR), and the peak signal-to-noise ratio (PSNR). They show respectively the quality of the entire reconstructed image, the improvement of the estimation iterations from the initialized image, and the signal-to-noise ratio (SNR) of the result corresponding to the biggest value of the image. The definitions of these evaluations are:

- $\delta_f = \frac{\|\mathbf{f} - \hat{\mathbf{f}}\|^2}{\|\mathbf{f}\|^2}$ , where  $\mathbf{f}$  represents the original image and  $\hat{\mathbf{f}}$  represents the reconstructed image.
- $ISNR = 10 \log_{10} \frac{\|\mathbf{f} - \hat{\mathbf{f}}_{(0)}\|^2}{\|\mathbf{f} - \hat{\mathbf{f}}\|^2}$ , where  $\hat{\mathbf{f}}_{(0)}$  is the initialized image for the iterative algorithms.
- $PSNR = 10 \log_{10} \frac{\max(\mathbf{f})^2}{\frac{1}{N} \sum_j (f_j - \hat{f}_j)^2}$ , where  $N$  is the length of the vector  $\mathbf{f}$ .

## 4.2. Simulation results

In the Table (1) and Table (2), different evaluations of the results are given, corresponding to the reconstruction results of Shepp Logan phantom of size  $64 \times 64$  and  $256 \times 256$  respectively. The evaluations are estimated with the reconstructed image within 50 iterations, using data with  $SNR = 20dB$  or  $SNR = 40dB$ .

Table 1: Different measurement of performance. The evaluations are estimated with 50 iterations for each method and each different case, for the Shepp Logan image of size  $64 \times 64$ .  $\lambda = 20$  in QR and TV methods when there is a high  $SNR=40dB$  noise.  $\lambda = 300$  in QR and  $\lambda = 40$  in TV when there is a low  $SNR=20dB$  noise.

	64 projection						32 projection					
	Data SNR=40dB			Data SNR=20dB			Data SNR=40dB			Data SNR=20dB		
	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM
$\delta_f$	0.1164	0.1046	<b>0.0255</b>	0.1324	0.1146	<b>0.1130</b>	0.1812	1675	<b>0.1114</b>	0.1958	0.1761	0.2262
ISNR	2.1171	2.5803	<b>8.7153</b>	1.7312	2.3562	<b>2.4175</b>	0.6862	1.0266	<b>2.8000</b>	0.6252	1.0843	-0.0015
PSNR	21.4654	21.9286	<b>28.0635</b>	20.9068	21.5319	<b>21.5931</b>	19.5432	19.8836	<b>21.6570</b>	19.2071	19.6662	18.5804

Table 2: Different measurement of performance. The evaluations are estimated with 50 iterations for each method and each different case, for the Shepp Logan image of size  $256 \times 256$ .  $\lambda = 20$  in QR and TV methods when there is a high  $SNR=40dB$  noise.  $\lambda = 300$  in QR and  $\lambda = 40$  in TV when there is a low  $SNR=20dB$  noise.

	128 projection						64 projection					
	Data SNR=40dB			Data SNR=20dB			Data SNR=40dB			Data SNR=20dB		
	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM	QR	TV	HHBM
$\delta_f$	0.0318	0.0147	<b>0.0120</b>	0.0839	0.0649	0.0715	0.0666	0.0443	<b>0.0376</b>	0.1323	0.0996	0.1110
ISNR	3.0647	6.4358	<b>7.3075</b>	1.3952	2.5113	2.0895	1.9484	3.7192	<b>4.4326</b>	1.7058	2.9398	2.4657
PSNR	27.1421	30.5133	<b>31.3850</b>	22.9369	24.0530	<b>31.3850</b>	23.9383	25.7091	<b>26.4225</b>	20.9580	22.1920	21.7179

From the results in these two tables, we find that for the reconstruction of the  $64 \times 64$  Shepp Logan image, the proposed HHBM method performs quite well compared to the other two state-of-the-art methods,

in all respects of the evaluation (relative error, ISNR, PSNR). In the reconstruction of the  $256 \times 256$  Shepp Logan image, the proposed HHBM method has a better reconstruction result within 50 iterations when the data is noisy with a high SNR, with a sufficient or insufficient number of projections. When the data is noisy with a low SNR=20dB, the proposed HHBM performs better with respect to the PSNR evaluation, but the TV method surpasses the HHBM method with respect to the relative error value. Discussion of the results will be given after, in the subsection 4.3.

Table (3) shows different metrics of the  $256 \times 256$  Shepp Logan image reconstructed from 128, 64 and 32 projections data set when the following convergence criterion is reached:

$$\left| \delta_f^{(k+1)} - \delta_f^{(k)} \right| < \varepsilon, \quad (51)$$

where  $\varepsilon$  is a small fixed value,  $\delta_f^{(k)}$  represents the relative error after  $k$  iterations, and in this simulation we set  $\varepsilon = 0.0001$ . We also estimated the structural similarity (SSIM) metric [66] in the comparison, with the definition:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (52)$$

where  $C_1$  and  $C_2$  are fixed constant,  $\mu$  is the mean and  $\sigma$  is the standard deviation. We can conclude from the table that the difference of the SSIM of final results by using different methods are not evident.

Table 3: Different measurement of performance. The evaluations are estimated until the convergence criterion is reached, for the Shepp Logan image of size  $256 \times 256$ .  $\lambda = 20$  in QR and TV methods when there is a high SNR=40dB noise.  $\lambda = 300$  in QR and  $\lambda = 40$  in TV when there is a low SNR=20dB noise.

	128 projections					
	Data SNR=40dB			Data SNR=20dB		
	QR	TV	HHBM	QR	TV	HHBM
$\delta_f$	0.0386	0.0208	<b>0.0174</b>	0.0932	0.0889	<b>0.0766</b>
ISNR	2.2345	4.9143	<b>5.6945</b>	0.9380	1.1393	<b>1.7866</b>
PSNR	26.3199	28.9918	<b>29.7719</b>	22.4797	22.6811	<b>23.3283</b>
SSIM	0.9996	<b>0.9998</b>	<b>0.9998</b>	0.9991	0.9992	<b>0.9993</b>
	64 projections					
	Data SNR=40dB			Data SNR=20dB		
	QR	TV	HHBM	QR	TV	HHBM
$\delta_f$	0.0729	0.0472	<b>0.0387</b>	0.1281	0.1057	<b>0.0946</b>
ISNR	0.5528	3.4429	<b>4.3076</b>	1.8463	2.6800	<b>3.1635</b>
PSNR	23.5427	25.4328	<b>26.2975</b>	21.0985	21.9321	<b>22.4157</b>
SSIM	0.9993	<b>0.9996</b>	<b>0.9996</b>	0.9988	0.9990	<b>0.9991</b>
	32 projections					
	Data SNR=40dB			Data SNR=20dB		
	QR	TV	HHBM	QR	TV	HHBM
$\delta_f$	0.1521	0.1290	0.0784	0.1925	0.1660	0.1254
ISNR	2.0943	2.8105	<b>4.9711</b>	2.5575	3.1988	<b>4.4188</b>
PSNR	20.3511	21.0673	<b>23.2280</b>	19.3292	19.9706	<b>21.1905</b>
SSIM	0.9986	0.9988	<b>0.9993</b>	0.9982	0.9985	<b>0.9989</b>

In order to show clearly the reconstructed Shepp Logan phantom, Fig.(5)-(8) show two zones of the image. The two zones are correspondingly the red and blue areas shown in Fig.(4). In the results, an example of profile is displayed in the images in order to show the change at the contour positions.

The comparison of the reconstructed 2D  $256 \times 256$  images using different HHBM methods, with **128 projections** and a high **SNR=40dB**, is shown in Fig.(5). In the images, we see that in the reconstructed image using the QR method, the shape of different blocs is not clear. In the result of the TV method, we find that the homogeneous areas of the reconstructed image are very smooth, but compared with the result of the

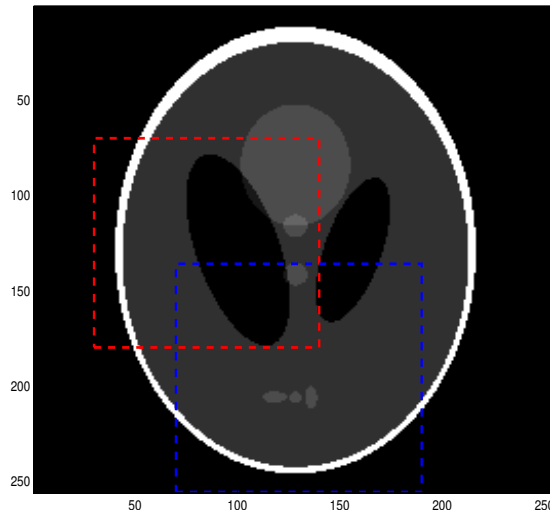


Figure 4: The original  $256 \times 256$  Shepp Logan phantom and two zones used to demonstrate the results.

proposed HHBM method, the shape of the blocs is not clear enough. In the HHBM reconstructed image, there are a few artefacts in the homogeneous blocs, but the shape of the image is clearer than these two state-of-the-art methods. In the figures, the blue line is the position of the profile which is displayed in red.

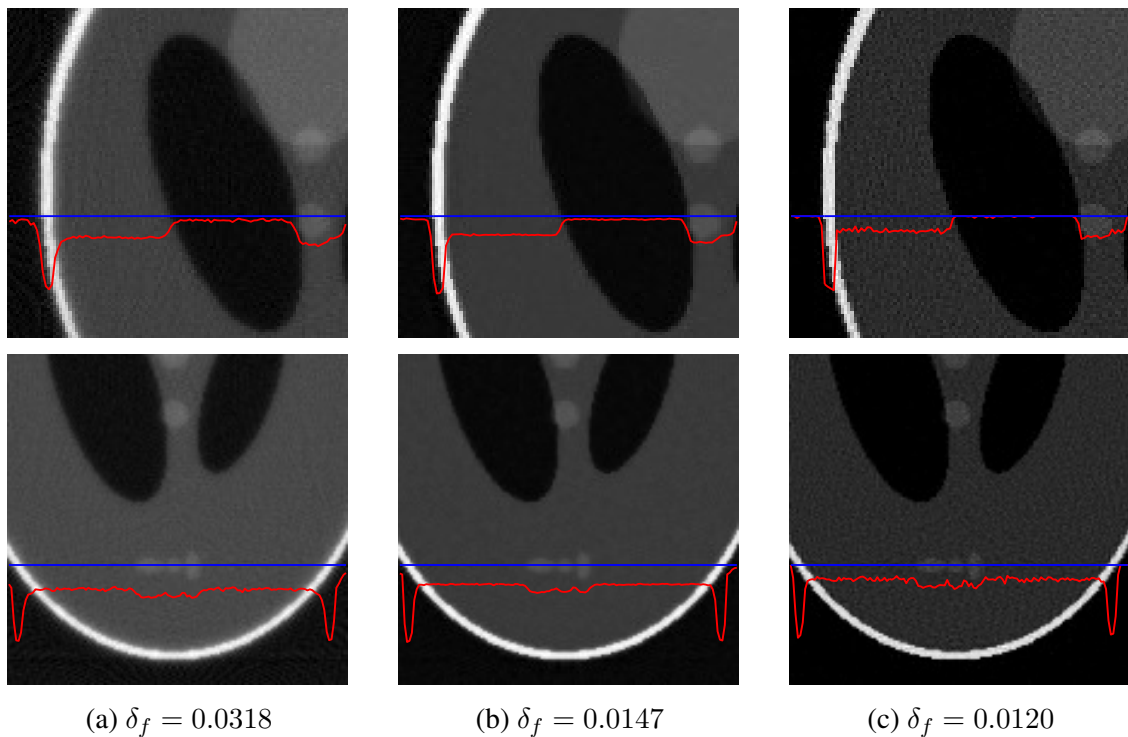


Figure 5: Comparison of zones of reconstructed  $256 \times 256$  Shepp Logan images by using different methods with **128 projections** and a high **SNR=40dB** by using: (a) QR, (b) TV and (c) proposed HHBM method. The red curve shows the profile at the blue line position.

The comparison of the reconstructed 2D  $256 \times 256$  images using different methods, with **128 projec-**



tions and a low **SNR=20dB**, is shown in Fig.(6). In the low SNR case, we see that all these methods have a result with noise in the homogeneous areas. Comparing these methods, the proposed HHBM method has clearer contours.

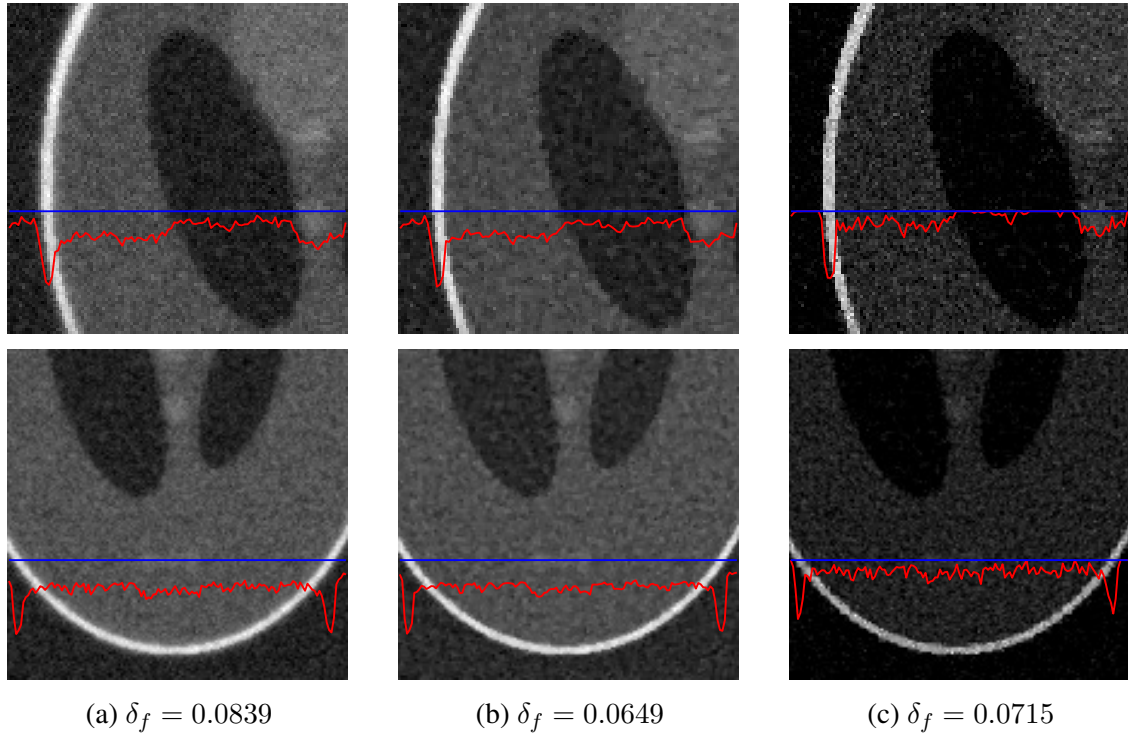


Figure 6: Comparison of reconstructed  $256 \times 256$  Shepp Logan images by using different methods with **128 projections** and a low **SNR=20dB** by using: (a) QR, (b) TV and (c) proposed HHBM method. The red curve shows the profile at the blue line position.

The comparison of the reconstructed 2D  $256 \times 256$  images using different methods, with **64 projections** and a high **SNR=40dB**, is shown in Fig.(7). In this case, the reconstruction is done by using only a quarter of the projections data. In this case, the TV method begins to have artefacts in the homogeneous areas, and the contours are not very clear. The result of the HHBM method has clearer contours.

The comparison of the reconstructed 2D  $256 \times 256$  images using different methods, with **32 projections** and a high **SNR=40dB**, is shown in Fig.(8). In this case, the reconstruction is done by using only  $1/8$  of the projection data. The result of the HHBM method still has clearer contours.

In Fig.(9), Fig.(10) and Fig.(11), the comparisons of the convergence of the relative error are shown, for the simulations with 128 projections and a high **SNR=40dB**, with 128 projections and a low **SNR=20dB**, with 64 projections and a high **SNR=40dB**, respectively. From these comparisons, we see that by using the HHBM method, the convergence is faster than the other state-of-the-art methods. Within just a few iterations, the reconstructed image already has a very good quality. But in the simulation with a low SNR, the relative error of the final results of HHBM is not very small.

In Fig.(12), we compare the convergence performance of the reconstruction with only 32 projections for a reconstruction of the Shepp Logan phantom of size  $256 \times 256$ . In this comparison, we see that the HHBM method performs better than the FBP, QR and TV methods.

The comparison of the profiles of the reconstructed images is given in Fig.(13) and Fig.(14), corresponding to a high **SNR=40dB** and a low **SNR=20dB** respectively. The top figures show the profiles of the reconstructed Shepp Logan image, and the bottom figures show a zone of the top figures, in order to

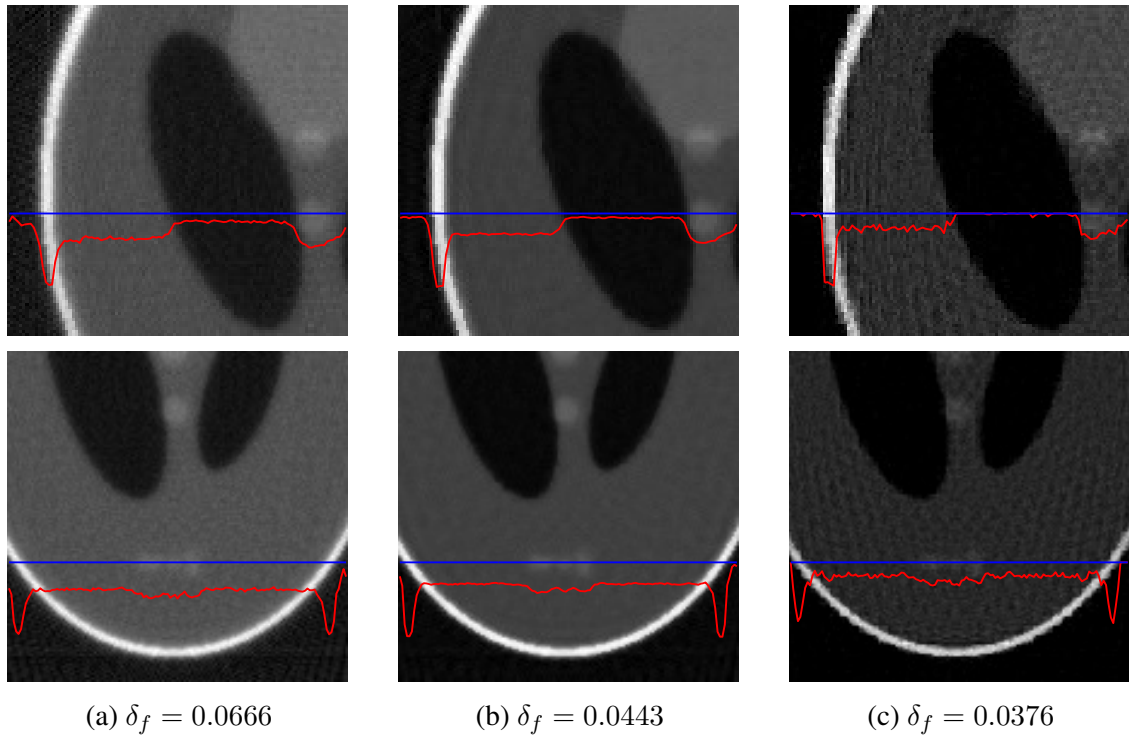


Figure 7: Comparison of reconstructed  $256 \times 256$  Shepp Logan images by using different methods with **64 projections** and a high **SNR=40dB** by using: (a) QR, (b) TV and (c) proposed HHBM method. The red curve shows the profile at the blue line position.

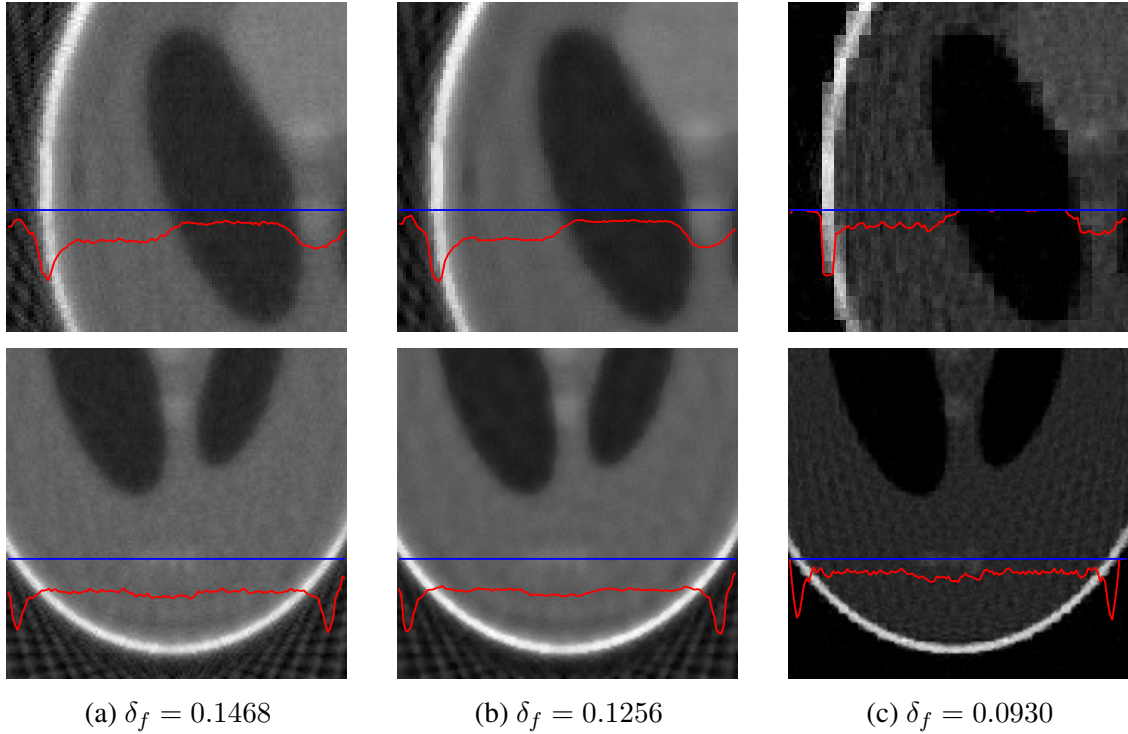


Figure 8: Comparison of reconstructed  $256 \times 256$  Shepp Logan images by using different methods with **32 projections** and a high **SNR=40dB** by using: (a) QR, (b) TV and (c) proposed HHBM method. The red curve shows the profile at the blue line position.

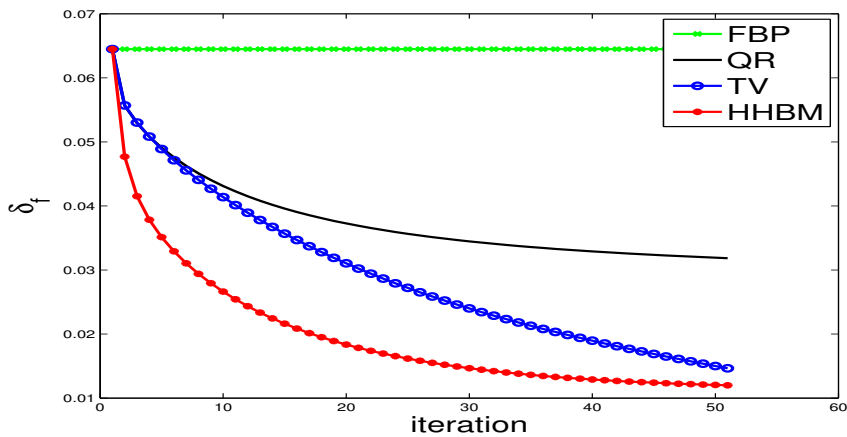


Figure 9: The performance of convergence of relative error  $\delta_f$  for phantom of size  $256 \times 256$  by using different methods with 128 projections and a high SNR=40dB.

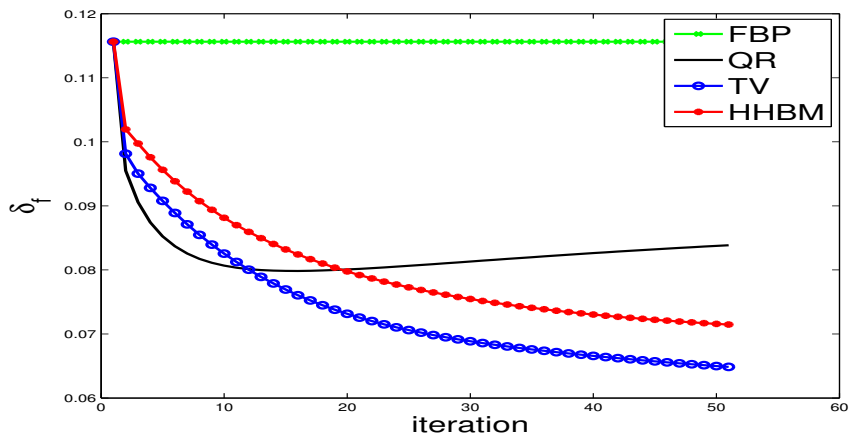


Figure 10: The performance of convergency of relative error  $\delta_f$  for phantom of size  $256 \times 256$  by using different methods with 128 projections and a low SNR=20dB.

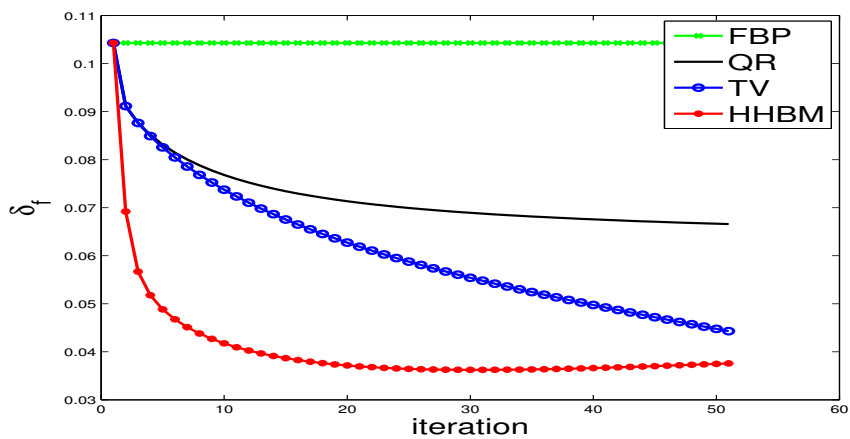


Figure 11: The performance of convergency of relative error  $\delta_f$  for phantom of size  $256 \times 256$  by using different methods with 64 projections and a high SNR=40dB.

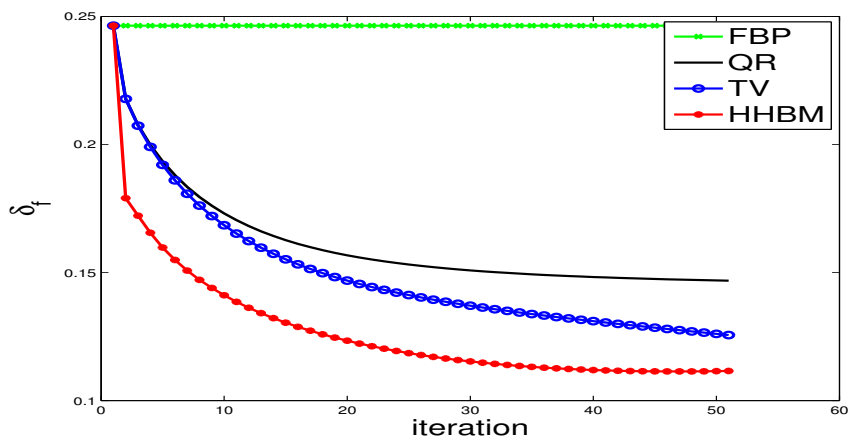


Figure 12: The performance of convergency of relative error  $\delta_f$  for phantom of size  $256 \times 256$  by using different methods with 32 projections and a high SNR=40dB.

see more clearly the details. In these profile figures, we see that by using the HHBM method, the profile is closer to the original profile at the contour positions, thus at the areas where there are major changes.

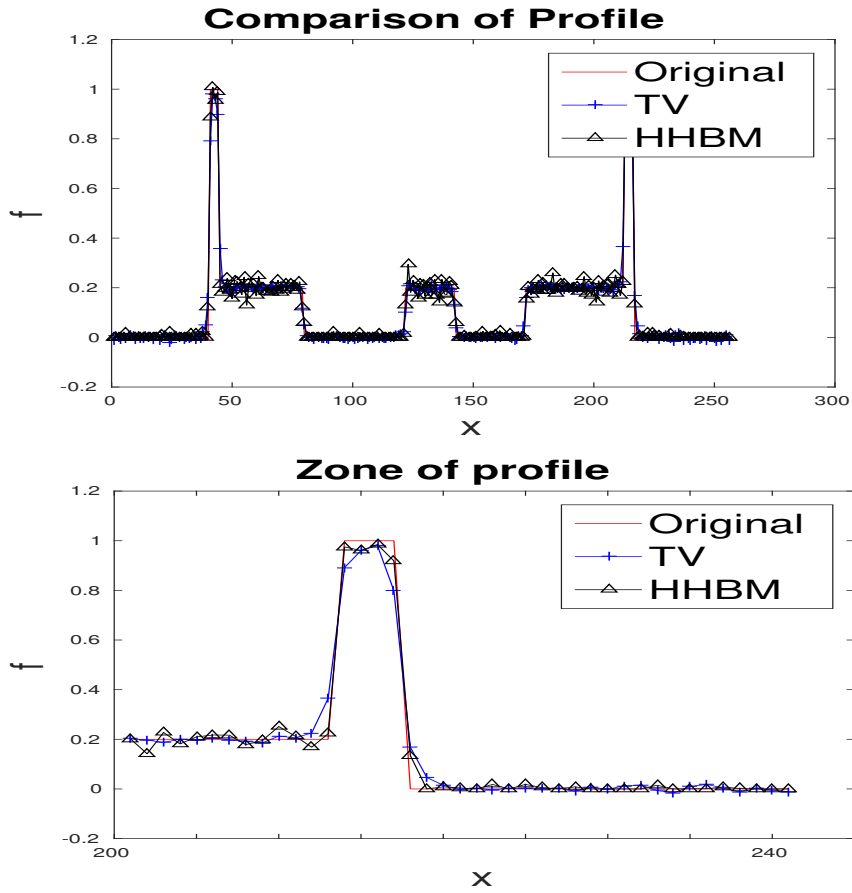


Figure 13: Comparison of profiles of reconstructed Shepp Logan image using different methods with 128 projections and a high SNR=40dB, after 50 iterations. The top image is the profile of original and reconstructed Shepp Logan image, the bottom image is a zone of the top image.

Then we analyze the performance of different methods when using a different number of projections and different SNR. The comparisons are given in Fig.(15) and Fig.(16). In these two figures, the left one analyzes the influence of the number of projections, and the right one analyzes the influence of the value of SNR. In all these figures, we use the relative error to evaluate the results. The evaluation is measured after 50 iterations for each method and each simulation. We can see from the results that all the methods have better results when dealing with (1) larger numbers of projections, and (2) data with a higher SNR. From the results, we notice that, when reconstructing from the data with a high SNR, the HHBM method is robust, even when the projection number is very small. When the size of phantom is smaller, the TV method performs better when there is a very low SNR (10dB). When the size of phantom becomes bigger, the noise has a significant influence on the TV method.

### 4.3. Discussion

The main drawback of the proposed HHBM method is that the noise in the homogeneous areas cannot be removed efficiently when the data has a low SNR. The problem comes from the hierarchical structured multilevel Haar transformation. In this transformation, the low-frequency component of the image, cor-

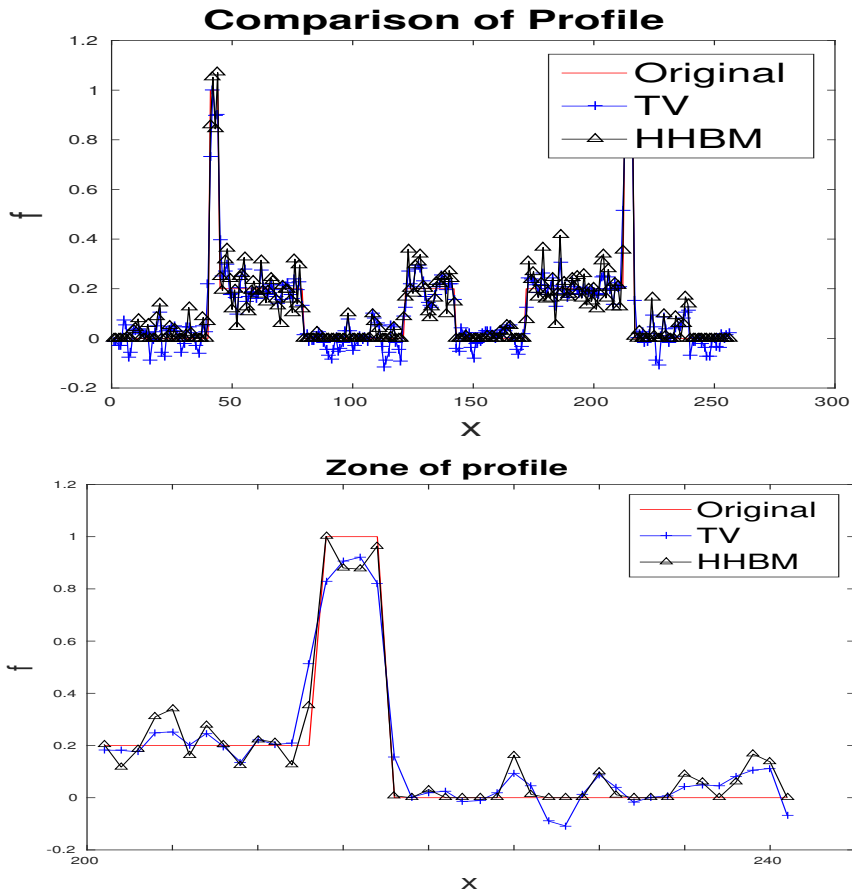


Figure 14: Comparison of profiles of reconstructed Shepp Logan image using different methods with 128 projections and a low SNR=20dB, after 50 iterations. The top image is the profile of original and reconstructed Shepp Logan image, the bottom image is a zone of the top image.

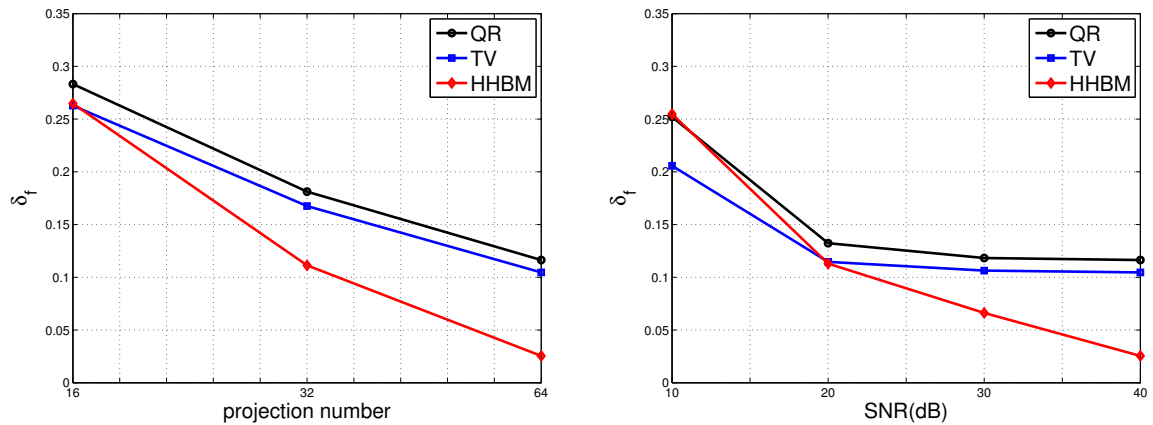


Figure 15: Comparison of the relative error of the reconstruction for the  $64 \times 64$  Shepp Logan image (a) by using different methods with different projection numbers and a high SNR=40dB, and (b) by using different methods with different SNR and 64 projections. The relative error is measured after 50 iterations.

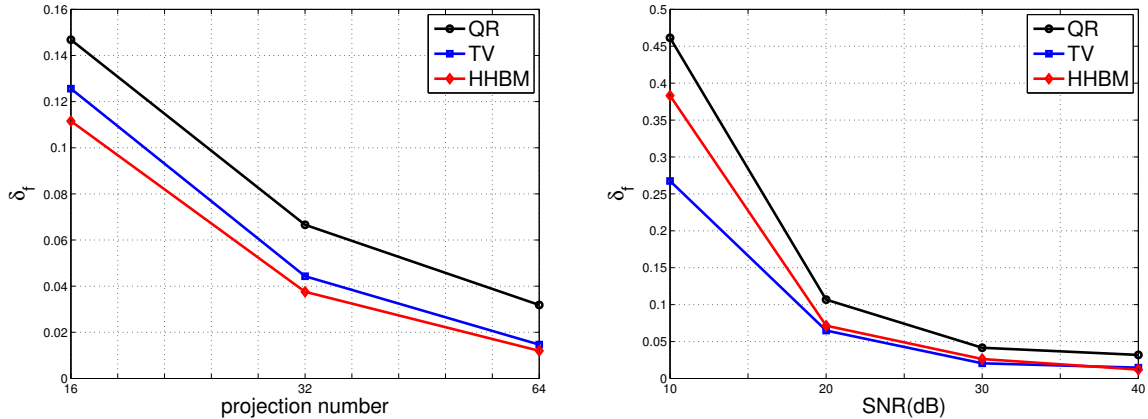


Figure 16: Comparison of the relative error of the reconstruction for the  $256 \times 256$  Shepp Logan image (a) by using different methods with different projection numbers and a high SNR=40dB, and (b) by using different methods with different SNR and 128 projections. The relative error is measured after 50 iterations.

responding to the top-left corner of the coefficient image in Fig.(3)(b), contains most of the information of the homogeneous areas. Meanwhile, the high-frequency components, corresponding to the other areas in Fig.(3)(b), contain the information of the contours of the blocs in the image. While treating the coefficient image, we use the same generalized Student-t model, with different hyper-parameters  $\alpha_{z_0}$  and  $\beta_{z_0}$ , in Eq.(17). The problem is that the information of the low-frequency component is more important than the high-frequency component so that in future work, we are considering applying a multi-resolution estimation for the coefficient  $z$ . The multi-resolution estimation sets out to estimate at the first step the low-frequency component of  $z$ , and then step by step, to estimate the low-frequency component and higher-frequency component.

In the regularization methods, there is always a regularization parameter  $\lambda$  to be fixed. In order to choose a suitable value for this parameter, we compared the relative error of the reconstructed image by using QR and TV methods with  $\lambda = [0.1, 0.2, 0.5, 1, 10, 20, 30, 40, 50, 100, 200]$  respectively. The results are shown in Fig.(17).

We can see that the choice of the value of  $\lambda$  plays an important role in the regularization methods. For example in the low SNR cases, it is better to choose a bigger value for  $\lambda$  than the high SNR cases. The cost of choosing such a value is very expensive. In our simulation we tested for only 10 different values and this step takes already much more time than the proposed Bayesian method, needless to say in the real applications where we know nothing about the noise level and the evaluation of the convergence is not appropriate. Hence, the proposed Bayesian method is more appropriate thanks to the fact that such parameters are all estimated automatically during the simulations.

## 5. Conclusion and Perspectives

The reconstruction of CT image plays an important role in medical and industrial applications. It is becoming important to reconstruct with an increasingly smaller number of projections. In this paper we proposed a Bayesian method with a hierarchical structured prior model. In this model the multilevel Haar transformation is used as a sparse transformation. A generalized Student-t distribution is used to model the transformation coefficients. The simulation results of the Shepp Logan phantom show good performance of the proposed method compared with the conventional quadratic regularization and Total

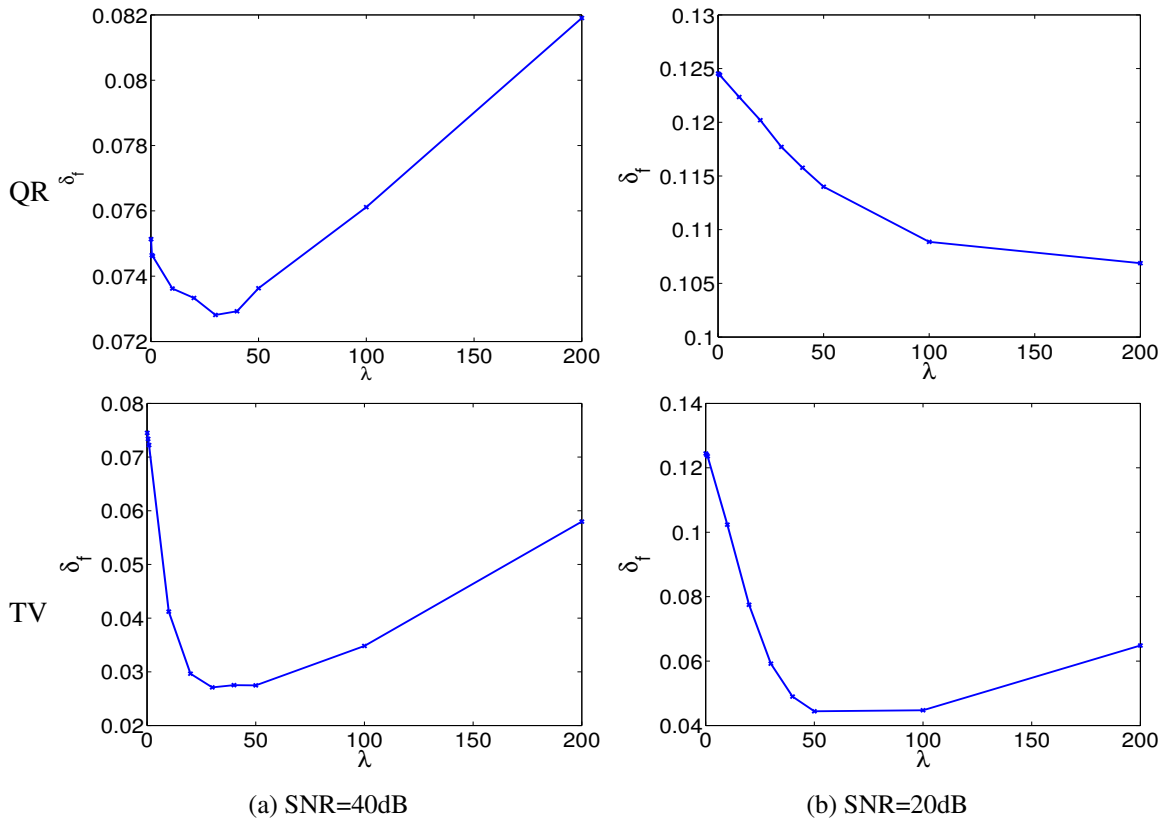


Figure 17: The final relative error by setting different value for  $\lambda$ . The top two figures show the QR method while the bottom two figures show the TV method. The left two figures present a high SNR case and the right two figures present a low SNR case.



Variation methods. The proposed method converges much faster than these state-of-the-art methods. In future work, we will look into different sparse transformations, for example the Dual-Tree CWT transformation [67] which has a better performance than the Haar transformation. We will also consider applying a 'multi-resolution' estimation strategy for the transformation coefficients. The simulation with 3D data is under execution, with the use of the GPU processor and the ASTRA toolbox [68].

## A. Expressions for updating step length $\hat{\gamma}_f$ and $\hat{\gamma}_z$

### A.1. Calculation of parameter $\hat{\gamma}_f^{(k)}$

The gradient descent step length for updating  $\mathbf{f}$  is calculated by minimizing the criterion  $\mathcal{J}_f(\mathbf{f} - \gamma_f d\mathcal{J}_f)$ :

$$\hat{\gamma}_f = \arg \min_{\hat{\gamma}_f} \mathcal{J}_f(\mathbf{f} - \gamma_f d\mathcal{J}_f),$$

where  $\mathcal{J}_f(\mathbf{f} - \gamma_f d\mathcal{J}_f)$  is simplified as:

$$\begin{aligned} & \mathcal{J}_f(\mathbf{f} - \gamma_f d\mathcal{J}_f) \\ &= \frac{1}{2} (\mathbf{g}^T - \mathbf{f}^T \mathbf{H}^T + \gamma_f d\mathcal{J}_f^T \mathbf{H}^T) \mathbf{V}_\epsilon^{-1} (\mathbf{g} - \mathbf{H}\mathbf{f} + \gamma_f \mathbf{H} d\mathcal{J}_f) \\ & \quad + \frac{1}{2} (\mathbf{f}^T - \gamma_f d\mathcal{J}_f^T - \mathbf{z}^T \mathbf{D}^T) \mathbf{V}_\xi^{-1} (\mathbf{f} - \gamma_f d\mathcal{J}_f - \mathbf{D}\mathbf{z}) \\ &= \frac{1}{2} (\gamma_f \mathbf{g}^T \mathbf{V}_\epsilon^{-1} \mathbf{H} d\mathcal{J}_f - \gamma_f \mathbf{f}^T \mathbf{H}^T \mathbf{V}_\epsilon^{-1} \mathbf{H} d\mathcal{J}_f + \gamma_f d\mathcal{J}_f^T \mathbf{H}^T \mathbf{V}_\epsilon^{-1} \mathbf{g} \\ & \quad - \gamma_f d\mathcal{J}_f^T \mathbf{H}^T \mathbf{V}_\epsilon^{-1} \mathbf{H}\mathbf{f} + \gamma_f^2 d\mathcal{J}_f^T \mathbf{H}^T \mathbf{V}_\epsilon^{-1} \mathbf{H} d\mathcal{J}_f) \\ & \quad + \frac{1}{2} \left( -\gamma_f \mathbf{f}^T \mathbf{V}_\xi^{-1} d\mathcal{J}_f - \gamma_f d\mathcal{J}_f^T \mathbf{V}_\xi^{-1} \mathbf{f} + \gamma_f^2 d\mathcal{J}_f^T \mathbf{V}_\xi^{-1} d\mathcal{J}_f + \gamma_f d\mathcal{J}_f^T \mathbf{V}_\xi^{-1} \mathbf{D}\mathbf{z} \right. \\ & \quad \left. + \gamma_f \mathbf{z}^T \mathbf{D}^T \mathbf{V}_\xi^{-1} d\mathcal{J}_f \right) + \text{Cst} \\ &= \gamma_f d\mathcal{J}_f^T \left[ \mathbf{H}^T \mathbf{V}_\epsilon^{-1} \mathbf{g} - \mathbf{H}^T \mathbf{V}_\epsilon^{-1} \mathbf{H}\mathbf{f} - \mathbf{V}_\xi^{-1} \mathbf{f} + \mathbf{V}_\xi^{-1} \mathbf{D}\mathbf{z} \right] \\ & \quad + \frac{1}{2} \gamma_f^2 \left[ \|\mathbf{Y}_\epsilon \mathbf{H} d\mathcal{J}_f\|^2 + \|\mathbf{Y}_\xi d\mathcal{J}_f\|^2 \right] + \text{Cst}. \end{aligned}$$

In Eq.(43) we have:

$$\begin{aligned} \nabla \mathcal{J}(\mathbf{f}) &= -\mathbf{H}^T \mathbf{V}_\epsilon^{-1} (\mathbf{g} - \mathbf{H}\mathbf{f}) + \mathbf{V}_\xi^{-1} (\mathbf{f} - \mathbf{D}\mathbf{z}) \\ &= -\mathbf{H}^T \mathbf{V}_\epsilon^{-1} \mathbf{g} + \mathbf{H}^T \mathbf{V}_\epsilon^{-1} \mathbf{H}\mathbf{f} + \mathbf{V}_\xi^{-1} \mathbf{f} - \mathbf{V}_\xi^{-1} \mathbf{D}\mathbf{z}. \end{aligned}$$

So that:

$$\mathcal{J}_f(\mathbf{f} - \gamma_f d\mathcal{J}_f) = \arg \min_{\hat{\gamma}_f} -\gamma_f \|d\mathcal{J}_f\|^2 + \frac{1}{2} \gamma_f^2 \left[ \|\mathbf{Y}_\epsilon \mathbf{H} d\mathcal{J}_f\|^2 + \|\mathbf{Y}_\xi d\mathcal{J}_f\|^2 \right] + \text{Cst}.$$

The gradient of the criterion is:

$$\frac{\partial \mathcal{J}_f(\mathbf{f} - \gamma_f d\mathcal{J}_f)}{\partial \gamma_f} = -\|d\mathcal{J}_f\|^2 + \gamma_f \left[ \|\mathbf{Y}_\epsilon \mathbf{H} d\mathcal{J}_f\|^2 + \|\mathbf{Y}_\xi d\mathcal{J}_f\|^2 \right].$$

By setting

$$\frac{\partial \mathcal{J}_f(\mathbf{f} - \gamma_f d\mathcal{J}_f)}{\partial \gamma_f} \Big|_{\hat{\gamma}_f} = 0,$$

we get:

$$\hat{\gamma}_f = \frac{\|d\mathcal{J}_f\|^2}{\|\mathbf{Y}_\epsilon \mathbf{H} d\mathcal{J}_f\|^2 + \|\mathbf{Y}_\xi d\mathcal{J}_f\|^2}$$

## A.2. Calculation of parameter $\gamma_z^{(k)}$

The gradient descent step length for updating  $\mathbf{z}$  is calculated by minimizing the criterion  $\mathcal{J}_z(\mathbf{z} - \gamma_z \text{d}\mathcal{J}_z)$ :

$$\hat{\gamma}_z = \arg \min_{\hat{\gamma}_z} \mathcal{J}_z(\mathbf{z} - \gamma_z \text{d}\mathcal{J}_z),$$

where  $\mathcal{J}_z(\mathbf{z} - \gamma_z \text{d}\mathcal{J}_z)$  is simplified as:

$$\begin{aligned} \mathcal{J}_z(\mathbf{z} - \gamma_z \text{d}\mathcal{J}_z) &= \frac{1}{2} (\mathbf{f}^T - \mathbf{z}^T \mathbf{D}^T + \gamma_z \text{d}\mathcal{J}_z^T \mathbf{D}^T) \mathbf{V}_\xi^{-1} (\mathbf{f} - \mathbf{D}\mathbf{z} + \gamma_z \mathbf{D} \text{d}\mathcal{J}_z) \\ &\quad + \frac{1}{2} (\mathbf{z}^T - \gamma_z \text{d}\mathcal{J}_z^T) \mathbf{V}_z^{-1} (\mathbf{z} - \gamma_z \text{d}\mathcal{J}_z) \\ &= \frac{1}{2} \left( \gamma_z \mathbf{f}^T \mathbf{V}_\xi^{-1} \mathbf{D} \text{d}\mathcal{J}_z - \gamma_z \mathbf{z}^T \mathbf{D}^T \mathbf{V}_\xi^{-1} \mathbf{D} \text{d}\mathcal{J}_z + \gamma_z \text{d}\mathcal{J}_z^T \mathbf{D}^T \mathbf{V}_\xi^{-1} \mathbf{f} \right. \\ &\quad \left. - \gamma_z \text{d}\mathcal{J}_z^T \mathbf{D}^T \mathbf{V}_\xi^{-1} \mathbf{D}\mathbf{z} + \gamma - \mathbf{z}^2 \text{d}\mathcal{J}_z^T \mathbf{D}^T \mathbf{V}_\xi^{-1} \mathbf{D} \text{d}\mathcal{J}_z \right) \\ &\quad + \frac{1}{2} \left( -\gamma_z \mathbf{z}^T \mathbf{V}_z^{-1} \text{d}\mathcal{J}_z - \gamma_z \text{d}\mathcal{J}_z^T \mathbf{V}_z^{-1} \mathbf{z} + \gamma_z^2 \text{d}\mathcal{J}_z^T \mathbf{V}_z^{-1} \text{d}\mathcal{J}_z \right) + \text{Cst} \\ &= \gamma_z \text{d}\mathcal{J}_z^T \left[ \mathbf{D}^T \mathbf{V}_\xi^{-1} \mathbf{f} - \mathbf{D}^T \mathbf{V}_\xi^{-1} \mathbf{D}\mathbf{z} - \mathbf{V}_z^{-1} \mathbf{z} \right] + \frac{1}{2} \gamma_z^2 \left[ \|\mathbf{Y}_\xi \mathbf{D} \text{d}\mathcal{J}_z\|^2 + \|\mathbf{Y}_z \text{d}\mathcal{J}_z\|^2 \right] + \text{Cst}. \end{aligned}$$

In Eq.(44) we have:

$$\begin{aligned} \nabla \mathcal{J}(\mathbf{z}) &= -\mathbf{D}^T \mathbf{V}_\xi^{-1} (\mathbf{f} - \mathbf{D}\mathbf{z}) + \mathbf{V}_z^{-1} \mathbf{z} \\ &= -\mathbf{D}^T \mathbf{V}_\xi^{-1} \mathbf{f} + \mathbf{D}^T \mathbf{V}_\xi^{-1} \mathbf{D}\mathbf{z} + \mathbf{V}_z^{-1} \mathbf{z}. \end{aligned}$$

So that:

$$\mathcal{J}_z(\mathbf{z} - \gamma_z \text{d}\mathcal{J}_z) = \arg \min_{\hat{\gamma}_z} -\gamma_z \|\text{d}\mathcal{J}_z\|^2 + \frac{1}{2} \gamma_z^2 \left[ \|\mathbf{Y}_\xi \mathbf{D} \text{d}\mathcal{J}_z\|^2 + \|\mathbf{Y}_z \text{d}\mathcal{J}_z\|^2 \right] + \text{Cst}.$$

The gradient of the criterion is:

$$\frac{\partial \mathcal{J}_z(\mathbf{z} - \gamma_z \text{d}\mathcal{J}_z)}{\partial \gamma_z} = -\|\text{d}\mathcal{J}_z\|^2 + \gamma_z \left[ \|\mathbf{Y}_\xi \mathbf{D} \text{d}\mathcal{J}_z\|^2 + \|\mathbf{Y}_z \text{d}\mathcal{J}_z\|^2 \right].$$

By setting

$$\frac{\partial \mathcal{J}_z(\mathbf{z} - \gamma_z \text{d}\mathcal{J}_z)}{\partial \gamma_z} \Big|_{\hat{\gamma}_z} = 0,$$

we get:

$$\hat{\gamma}_z = \frac{\|\text{d}\mathcal{J}_z\|^2}{\|\mathbf{Y}_\xi \mathbf{D} \text{d}\mathcal{J}_z\|^2 + \|\mathbf{Y}_z \text{d}\mathcal{J}_z\|^2}$$

## References

- [1] Abrams HL, McNeil BJ. Medical implications of computed tomography (CAT scanning). *New England Journal of Medicine*. 1978;298(6):310–318.
- [2] Arridge SR. Optical tomography in medical imaging. *Inverse problems*. 1999;15(2):R41.
- [3] Hanke R, Fuchs T, Uhlmann N. X-ray based methods for non-destructive testing and material characterization. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 2008;591(1):14–18.

- [4] Halmshaw R, Honeycombe R, Hancock P. Non-destructive testing. Arnold; 1991.
- [5] Kalender WA. X-ray Computed Tomography. *Physics in Medicine and Biology*. 2006;51(13):R29.
- [6] Natterer F, Wubbeling F. A propagation-backpropagation method for ultrasound tomography. *Inverse problems*. 1995;11(6):1225.
- [7] Bolomey JC, Pichot C. Microwave tomography: from theory to practical imaging systems. *International Journal of Imaging Systems and Technology*. 1990;2(2):144–156.
- [8] Nelson J, Milner T, Tanenbaum B, Goodman D, Van Gemert M. Infra-red tomography of port-wine-stain blood vessels in human skin. *Lasers in Medical Science*. 1996;11(3):199–204.
- [9] Crowther R, DeRosier D, Klug A. The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. vol. 317. The Royal Society; 1970. p. 319–340.
- [10] Bracewell RN, Riddle A. Inversion of fan-beam scans in radio astronomy. *The Astrophysical Journal*. 1967;150:427.
- [11] Berry M, Gibbs D. The interpretation of optical projections. In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. vol. 314. The Royal Society; 1970. p. 143–152.
- [12] Clackdoyle R, Defrise M. Tomographic reconstruction in the 21st century. *IEEE Signal Processing Magazine*. 2010;27(4):60–80.
- [13] Deans SR. *The Radon transform and some of its applications*. Courier Corporation; 2007.
- [14] Chetih N, Messali Z. Tomographic image reconstruction using Filtered Back Projection (FBP) and Algebraic Reconstruction Technique (ART). In: *Control, Engineering & Information Technology (CEIT), 2015 3rd International Conference on*. IEEE; 2015. p. 1–6.
- [15] Ludwig D. The Radon transform on Euclidean space. *Communications on Pure and Applied Mathematics*. 1966;19(1):49–81.
- [16] Beylkin G. Discrete radon transform. *IEEE transactions on acoustics, speech, and signal processing*. 1987;35(2):162–172.
- [17] Feldkamp L, Davis L, Kress J. Practical cone-beam algorithm. *JOSA A*. 1984;1(6):612–619.
- [18] Alifanov OMI, Artiukhin EA, Rumiantsev SV. *Extreme methods for solving ill-posed problems with applications to inverse heat transfer problems*. Begell house New York; 1995.
- [19] O’Sullivan F. A statistical perspective on ill-posed inverse problems. *Statistical science*. 1986;p. 502–518.
- [20] Gordon R. A tutorial on ART (algebraic reconstruction techniques). *IEEE Transactions on Nuclear Science*. 1974;21(3):78–93.
- [21] Gordon R, Bender R, Herman GT. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of theoretical Biology*. 1970;29(3):471–481.
- [22] Trampert J, Leveque JJ. Simultaneous iterative reconstruction technique: physical interpretation based on the generalized least squares solution. *J geophys Res*. 1990;95(12):553–9.
- [23] Bangliang S, Yiheng Z, Lihui P, Danya Y, Baofen Z. The use of simultaneous iterative reconstruction technique for electrical capacitance tomography. *Chemical Engineering Journal*. 2000;77(1):37–41.
- [24] Ben-Israel A, Greville TN. *Generalized inverses: theory and applications*. vol. 15. Springer Science & Business Media; 2003.
- [25] Rit S, Sarrut D, Desbat L. Comparison of analytic and algebraic methods for motion-compensated cone-beam CT reconstruction of the thorax. *IEEE Transactions on Medical Imaging*. 2009;28(10):1513–1525.
- [26] Chambolle A, Lions PL. Image recovery via total variation minimization and related problems. *Numerische Mathematik*. 1997;76(2):167–188.

- [27] Goldstein T, Osher S. The Split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*. 2009;2(2):323–343.
- [28] Chan TF, Golub GH, Mulet P. A nonlinear primal-dual method for Total Variation-based image restoration. *SIAM Journal on Scientific Computing*. 1999;20(6):1964–1977.
- [29] Wahlberg B, Boyd S, Annergren M, Wang Y. An ADMM algorithm for a class of total variation regularized estimation problems. *IFAC Proceedings Volumes*. 2012;45(16):83–88.
- [30] Rolfs B, Rajaratnam B, Guillot D, Wong I, Maleki A. Iterative thresholding algorithm for sparse inverse covariance estimation. In: *Advances in Neural Information Processing Systems*; 2012. p. 1574–1582.
- [31] Tao S, Boley D, Zhang S. Local Linear Convergence of ISTA and FISTA on the LASSO Problem. *SIAM Journal on Optimization*. 2016;26(1):313–336.
- [32] Mumford D, Shah J. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*. 1989;42(5):577–685.
- [33] Ramani S, Liu Z, Rosen J, Nielsen JF, Fessler JA. Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods. *IEEE Transactions on Image Processing*. 2012;21(8):3659–3672.
- [34] Galatsanos NP, Katsaggelos AK. Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. *IEEE Transactions on Image Processing*. 1992;1(3):322–336.
- [35] Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2005;67(2):301–320.
- [36] Wang G, Schultz L, Qi J. Statistical image reconstruction for muon tomography using a Gaussian scale mixture model. *IEEE Transactions on Nuclear Science*. 2009;56(4):2480–2486.
- [37] Ramani S, Fessler JA. Statistical X-ray CT reconstruction using a splitting-based iterative algorithm with orthonormal wavelets. In: *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE; 2012. p. 1008–1011.
- [38] Redner RA, Walker HF. Mixture densities, maximum likelihood and the EM algorithm. *SIAM review*. 1984;26(2):195–239.
- [39] Moon TK. The expectation-maximization algorithm. *IEEE Signal processing magazine*. 1996;13(6):47–60.
- [40] Shepp LA, Vardi Y. Maximum likelihood reconstruction for emission tomography. *IEEE transactions on medical imaging*. 1982;1(2):113–122.
- [41] Tregouet D, Escolano S, Tiret L, Mallet A, Golmard J. A new algorithm for haplotype-based association analysis: the Stochastic-EM algorithm. *Annals of human genetics*. 2004;68(2):165–177.
- [42] Hudson HM, Larkin RS. Accelerated image reconstruction using ordered subsets of projection data. *IEEE transactions on medical imaging*. 1994;13(4):601–609.
- [43] Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian data analysis*. vol. 2. Chapman & Hall/CRC Boca Raton, FL, USA; 2014.
- [44] Qi J, Leahy RM. Resolution and noise properties of MAP reconstruction for fully 3-D PET. *IEEE transactions on medical imaging*. 2000;19(5):493–506.
- [45] Ericson WA. A note on the posterior mean of a population mean. *Journal of the Royal Statistical Society Series B (Methodological)*. 1969;p. 332–334.
- [46] Fox CW, Roberts SJ. A tutorial on variational Bayesian inference. *Artificial intelligence review*. 2012;38(2):85–95.
- [47] Tzikas DG, Likas AC, Galatsanos NP. The variational approximation for Bayesian inference. *IEEE Signal Processing Magazine*. 2008;25(6):131–146.
- [48] Chen J, Cong J, Vese LA, Villasenor J, Yan M, Zou Y. A hybrid architecture for compressive sensing 3-D CT reconstruction. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*. 2012;2(3):616–625.

- [49] Noël PB, Walczak AM, Xu J, Corso JJ, Hoffmann KR, Schafer S. GPU-based cone beam computed tomography. *Computer methods and programs in biomedicine*. 2010;98(3):271–277.
- [50] Gac N, Mancini S, Desvignes M, Houzet D. High speed 3D tomography on CPU, GPU, and FPGA. *EURASIP Journal on Embedded systems*. 2008;2008:5.
- [51] Wang L, Mohammad-Djafari A, Gac N, Dumitru M. 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. p. 883–887.
- [52] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*. 2006;15(12):3736–3745.
- [53] Chen Y, Nasrabadi NM, Tran TD. Hyperspectral image classification using dictionary-based sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*. 2011;49(10):3973–3985.
- [54] Flury BN, Gautschi W. An algorithm for simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form. *SIAM Journal on Scientific and Statistical Computing*. 1986;7(1):169–184.
- [55] Montgomery D, Zippin L. *Topological transformation groups*. vol. 95. Interscience Publishers New York; 1955.
- [56] Ahmed N, Natarajan T, Rao KR. Discrete cosine transform. *IEEE transactions on Computers*. 1974;100(1):90–93.
- [57] Nadarajah S. A generalized normal distribution. *Journal of Applied Statistics*. 2005;32(7):685–694.
- [58] Rasmussen CE. The infinite Gaussian mixture model. In: *NIPS*. vol. 12; 1999. p. 554–560.
- [59] Aristophanous M, Penney BC, Martel MK, Pelizzari CA. A Gaussian mixture model for definition of lung tumor volumes in positron emission tomography. *Medical physics*. 2007;34(11):4223–4235.
- [60] Klebanov LB. *Heavy tailed distributions*. Matfyzpress; 2003.
- [61] Gomes CP, Selman B, Crato N, Kautz H. Heavy-tailed phenomena in satisfiability and constraint satisfaction problems. *Journal of automated reasoning*. 2000;24(1-2):67–100.
- [62] Lange KL, Little RJ, Taylor JM. Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*. 1989;84(408):881–896.
- [63] Tzikas DG, Likas AC, Galatsanos NP. Variational Bayesian sparse kernel-based blind image deconvolution with Student's-t priors. *IEEE transactions on image processing*. 2009;18(4):753–764.
- [64] Dumitru M. A Bayesian approach for periodic components estimation for chronobiological signals; 2016.
- [65] Boyd S, Vandenberghe L. *Convex optimization*. Cambridge university press; 2004.
- [66] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*. 2004;13(4):600–612.
- [67] Selesnick IW, Baraniuk RG, Kingsbury NC. The dual-tree complex wavelet transform. *IEEE signal processing magazine*. 2005;22(6):123–151.
- [68] Palenstijn WJ, Batenburg KJ, Sijbers J. The ASTRA tomography toolbox. In: 13th International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE. vol. 2013; 2013. .