TOMOGPI - A MATLAB PLUGIN FOR ASTRA TOOLBOX

BAYESIAN ALGORITHMS FOR TOMOGRAPHY

MIRCEA DUMITRU*, LI WANG*, NICOLAS GAC*, ALI MOHAMMAD-DJAFARI*

April 10, 2018

CONTENTS

INTRODUCTION

This is a report containing the details about the algorithms and programs developed in *Groupe Problèmes Inverses (GPI) – Laboratoire des signaux et systèmes* for Computed Tomography. Some specific programs, using some specific algorithms are to be integrated in the **ASTRA Toolbox**, as a library providing Bayesian algorithms. The report provides some basic details about the algorithms, indicating the sources for more details. It contains also the implementation details of the algorithms, using the *ASTRA Toolbox*. Most of the algorithms discussed in this report can be found in [1]. First the Bayesian framework is presented then the HHBM algorithm is provided in the *TomoGPI* plugin

* *Laboratoire des signaux et systèmes, CNRS – CentraleSupélec – Université Paris-Sud, Université Paris-Saclay, France*

1 BAYESIAN FRAMEWORK

This report contains the technical details of some specific algorithms designed for Computed Tomography (CT) that are to be integrated in the ASTRA Toolbox as a library. The algorithms are designed in the Bayesian framework and their implementations is done in Matlab, using the ASTRA Toolbox and implicitly all the conventions for geometry, phantom, etc.

It contains a brief description of how the CT problem is tackled in the Bayesian framework, introducing the forward model (considered throughout this report), the Bayesian hierarchical model considered for modeling the *a priori* knowledge, the estimators used and the corresponding iterative algorithms. Depending on the particular *a priori* distribution used to model the image (or the representation of the image, via a transformation, which will be the case, inducing a *sparse* representation of the image to be reconstructed), on the particular estimation method, on the hypothesis of the noise structure (stationary or non-stationary), on the choice of the transformation used to obtain a sparse representation of the image (Harr, wavelet, *etc...*) the resulting algorithm are different. We will present, briefly, the particularities of each corresponding algorithm. The general framework can be resumed in the following steps:

- **The linear model:** Considers a mathematical model to describe the process of Computed Tomography. Typically, under the hypothesis of linearity this is often the forward model, (see (1)). It describes in matricial form the recorded data as a sum between an operator (typically the Radon transformation) applied on the unknown image plus a term accounting for the noise and the uncertainties of the model (for instance the hypothesis of linearity). This report considers a linear model that accounts for uncertainties also between the image to be reconstructed and its sparse representation, (see (2)).

- **Build an hierarchical (Bayesian) prior model:** Consider *a priori* distributions modeling the noise(s), the image (or, in our case, for its sparse corresponding representation) and the other parameters (in our case, the variances). Consider the *a posteriori* corresponding distribution, obtained via Bayes rule.

- **Estimate the unknowns (volume and parameters)** From the *a posteriori* distribution, estimate the unknowns of the model. It is worth mentioning that, along with the image to be reconstructed (or it sparse representation), all the other parameters involved in the hierarchical model (notably the variances) can be estimated. This is important, since this framework allows the uncertainty quantification of the estimation, via the estimated variances. The estimation leads typically to iterative algorithms.

1.1 Forward model and sparse representation of the volume

The methods considered in this report are developed using the classical forward model used in inverse problems,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{x}$ represents the image to be reconstructed, $\mathbf{y}$ represents the recorded data (the sinograme) used for reconstructing the image and $\mathbf{H}$ the measurement matrix, and considers a supplementary linear equation for the *sparse representation* of the image $\mathbf{x}$ via a *sparse transformation operator* $\mathbf{D}$ (Haar, gradient, DT-CWT) and introducing an uncertainty $\boldsymbol{\xi}$,

$$\begin{cases} \mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\epsilon} \\ \mathbf{x} = \mathbf{D}\mathbf{z} + \boldsymbol{\xi} \end{cases}, \tag{2}$$

Throughout this report, $\mathbf{z}$ is the main sparse-unknown of the model from which $\mathbf{f}$ is easily obtained using $\mathbf{D}$. Also, $\mathbf{D}$ is a sparse transformation operation, in most cases corresponding to the Haar transform.

1.2 Bayesian framework

All the algorithms discussed in this report are issued from a Bayesian framework, where *a priori* probability density functions are considered for modeling our knowledge of the parameters. Three (zero mean) infinite Gaussian scale mixtures (iGSM) will form the prior model, but each one appearing for different reason. As mentioned, in (2), z is the sparse representation of the image to be reconstructed. Therefore, for modeling it, heavy tailed distribution are to be considered, since this family is known in the literature to promote sparsity. In particular, we consider the class of such distributions that are expressed as iGSM, where the mixing $\mathcal{M}$ distribution is conjugate prior. The model can be expressed by

$$z \mid \theta_{v_z} \sim \mathrm{iGSM}\left(z \mid \theta_{v_z}\right) \leftrightarrow \begin{cases} z \mid v_z \sim \mathcal{N}\left(z \mid \mathbf{0}, V_z\right) \\ v_z \sim \prod_{j=1}^{M} \mathcal{M}\left(v_{z_j} \mid \theta_{v_z}\right) \end{cases}, \text{with iGSM being} \begin{cases} \mathcal{S}t \\ \mathcal{VG} \\ \mathcal{NIG} \end{cases} \text{for } \mathcal{M} \text{ being} \begin{cases} \mathcal{IG} \\ \mathcal{G} \\ \mathcal{GIG} \end{cases} \tag{3}$$

where $\mathcal{N}$ denotes the Normal distribution, $\mathcal{M}$ denotes a conjugate mixing distribution, θ_{v_z} denotes its parameter(s) (note that θ_{v_z} does not depend on j, i.e. we assume the same mixing distribution parameters for each variance vector element) and $V_z = \mathrm{diag}\left[v_z\right]$. The conjugate mixing distributions that will be considered are the Inverse Gamma ($\mathcal{IG}$), Gamma ($\mathcal{G}$) and Generalized Inverse Gaussian ($\mathcal{GIG}$). While (3) is considered in order to model the sparse structure of z via a heavy tailed distribution, the same model as in (3) is used for modeling noise ϵ and ϵ. This time, the reason for introducing it is different and is justified by the supposition of non-stationary noise: the noise vector is supposed to be zero-mean (multivariate) Gaussian distributed, with each corresponding variance following the same (mixing) $\mathcal{M}$ distribution, with different corresponding parameters. The graphical representation of the hierarchical Bayesian model corresponding to the forward model (2) is presented in Figure 1.

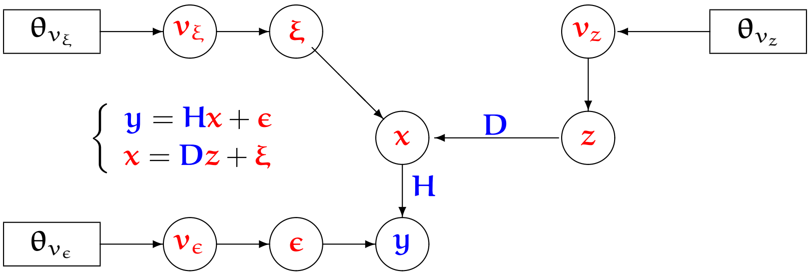


Figure 1: Hierarchical Bayesian Model

1.3 Estimation

From the posterior distribution, different estimation techniques can be considered in order to infer on the parameters. Mainly, the algorithms presented in this report correspond to the *Maximum A Posteriori* (MAP) point estimator. In particular all the MAP algorithms presented here correspond to a simple alternate minimization. The next section presents the Haar Hierarchical Bayesian Model HHMM algorithms.

2 HAAR HIERARCHICAL BAYESIAN MODEL (HHBM) – JMAP ESTIMATION

This section presents the Algorithm provided in the TomoGPI plugin:

- the sparse representation of the image is obtained via the Haar transform
- the noise ϵ is considered non-stationary
- the estimation is done using Joint MAP (by alternate optimization)

The sparse modeling of the sparse structure will be done via heavy-tailed distributions, in particular expressed as iGSM. We refer to this model throughout the report as the HHBM. In (2), D will denote the Haar transform. The sparse representation image z and its corresponding variance v_z are modeled by (3), with a conjugate mixing distribution. Some possible choices are presented in Table 1. For example, for an Inverse Gamma ($\mathcal{IG}$) distribution, the corresponding prior is a Student-t marginal distribution, modeling the sparse image representation. In this case, the parameters θ_{v_z} are the shape and scale parameters of the

Normal ($\mathcal{N}$) ($z_j \mid v_{z_j}$)	Mixing ($\mathcal{M}$) ($v_{z_j} \mid \theta_z$)	Prior distribution ($z_j \mid \theta_z$)
$\mathcal{N}\left(z_j \mid v_{z_j}\right)$	Inverse Gamma ($\mathcal{IG}$)	Student-t ($\mathcal{S}t$)
$\mathcal{N}\left(z_j \mid v_{z_j}\right)$	Gamma ($\mathcal{G}$)	Variance-Gamma ($\mathcal{VG}$)
$\mathcal{N}\left(z_j \mid v_{z_j}\right)$	particular form of Generalized Inverse Gaussian ($\mathcal{GIG}$)	Normal Inverse Gaussian ($\mathcal{NIG}$)

Table 1: iGSM models: Priors ($z_j \mid \theta_z$) and their corresponding *conjugate* mixing distribution ($v_{z_j} \mid \theta_z$)

$\mathcal{IG}$ distribution. We will see that in the hierarchical modeling, those particular parameters represent the highest level and therefore, an *a priori* choice has to be done for their values. In the Bayesian framework this is referred to as the *parameter selection model* and when the Bayesian framework is interpreted as a regularization approach, the choice of those parameters (often referred to as the hyperparameters) is associated with the choice for the regularization parameters in the regularization approach. As mentioned, the same iGSM is used for modeling the noise ϵ and its corresponding variance v_ϵ, respectively the noise ξ and its corresponding variance v_ξ.

The iGSM model for the noise is justified by the non-stationary hypothesis of the noise. In this case this model is considered not because of we are interested in the marginal modeling (only) the noise but we rather consider a Gaussian distribution for the noise, conditioned by the its corresponding variance and then, for each variance element we assume a distribution in the same family in order to describe the non-stationary behaviour. The likelihood is obtained from the ϵ noise model, and first equation of the forward model (2).

There is no *a priori* reason to assume the same mixing distribution for the three iGSM. While, for example, a Student-t is considered as the *sparsity enforcing prior* modeling the sparse representation of the image, corresponding to a $\mathcal{IG}$ mixing distribution, we can image an hierarchical model where the non-stationary noise is modeled by any other mixing distribution. However, for simplicity, we present during this report the algorithms that corresponds to prior models that are symmetric with respect to the mixing distribution, i.e. we use the same mixing distributions for all three iGSM.

In a straightforward formulation, the hierarchical Bayesian model described above, resumes to a posterior distribution proportional to the product of **a)** three (multivariate) Normal distributions, modeling $z \mid v_z$, $\epsilon \mid v_\epsilon$ and $\xi \mid v_\xi$ and **b)** three products of Inverse Gamma distributions (or other conjugate mixing distribution).

The JMAP estimator can be directly applied, by minimizing the corresponding log-posterior distribution via an alternate optimization. The detailed mathematical development of the model is not the goal of this report. This development is presented however in Appendices. Here we present the corresponding algorithm. Another possible estimator is the Posterior Mean (PM).

3 LIBRARY: INTEGRATION IN ASTRA TOOLBOX

This section presents the implementation details corresponding to the Bayesian algorithms, Sec. 2 in Matlab using the ASTRA Toolbox. In order the present the behaviour of the algorithms, some Bayesian algorithms are available in a user friendly program, in order to be tested for different dimensions, levels of noise or different parameters. First, in Subsection 3.1, program is presented, with generic details about the corresponding parameters used for each algorithm. Then, in Subsection 3.2, the implementation details are presented for the Bayesian algorithms.

3.1 Demo

The program is designed to be user friendly and can be navigated using menus. The first menu available to the user is presented in Fig. 2.

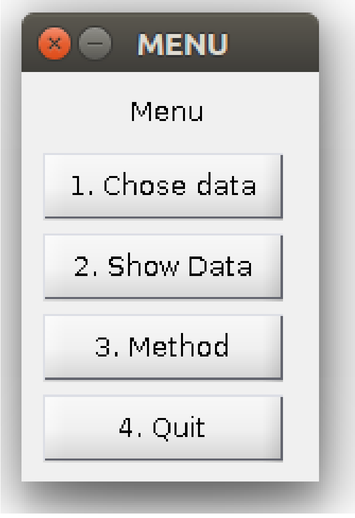


Figure 2: Menu

The options are:

- Chose data - user sets the synthetic data used as input for the algorithms: dimension, number of projections and noise level;
- Show data - user can see the data, i.e. the original Sheep Logan phantom and its projection;
- Method - user can perform the reconstruction using different methods;
- Quit - quits program;

The first option, *Chose data* selects the data dimension (i.e. the cube dimension), the projection number and the level of noise (SNR), Fig. 3. If it is not accessed, the default values are used: 64^3, 64 projection and SNR=30dB.

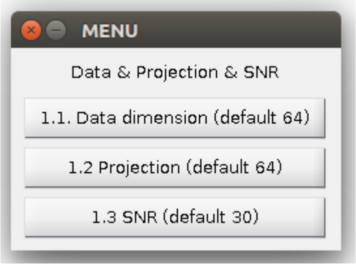
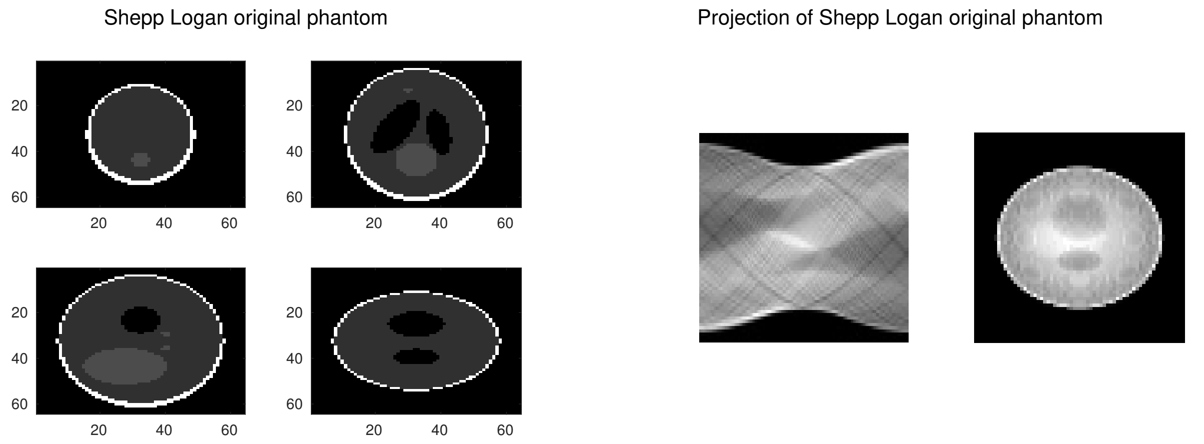


Figure 3: *Show data* Menu

The second option, *Show data* is showing the original Sheep Logan phantom and its projection, Fig. (4)



(a) Original Sheep Logan phantom slices

(b) original Sheep Logan phantom slice and its projection

Figure 4: Data

The third option, *Method* selects the reconstruction method, Fig. 5. The Bayesian methods, using three different prior laws, (see Tab 1), using the JMAP as estimation, constructed as presented in Subsection 1.1 and detailed in Appendix A are *JMAP – St-t*, *JMAP – NIG* and *JMAP – VG*. In order to compare the results, in term of reconstruction error or time, two other reconstruction methods are available: the classical *Filtered Back Projection* method and the Total Variation (TV) method.

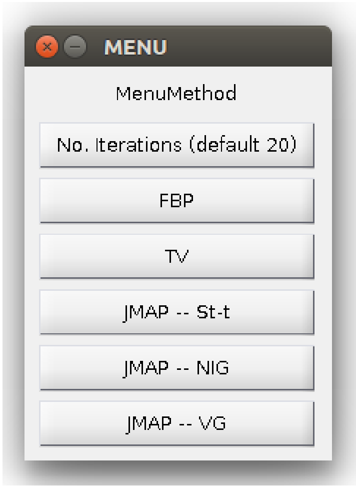


Figure 5: *Method* Menu

Each method is producing the corresponding reconstruction volume and plots four slices for comparison, Fig. 6 with the original volume, Fig. 4.

3.2 Implementation details

The program is designed such that it can be integrated in the ASTRA Toolbox. It is therefore using the ASTRA Toolbox *mex* functions. The 3D phantom, representing the unknown $\mathbf{f}$ of the linear forward model (1) is generated using the MATLAB function *phantom3D*,

```
f = phantom3d(N);
```

The measurement matrix $\mathbf{H}$ of the linear forward model (1) is generated using the *opTomo* and the ASTRA functions for the volume and projection geometry

```
vol_geom = astra_create_vol_geom(N, N, N);
proj_geom = astra_create_proj_geom('parallel3d', 1, 1, N, N, linspace2(0,pi,nproj));
H_base = opTomo('cuda', proj_geom, vol_geom);
```

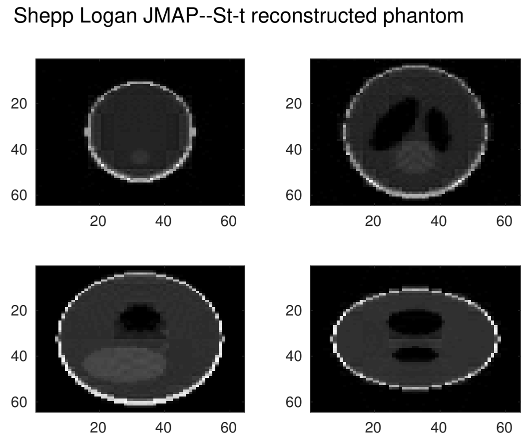


Figure 6: *Method* Reconstructed volume, via Bayesian algorithm, Student-t model

A HHBM - STUDENT-T - JMAP: MODEL, DEVELOPMENT AND ALGORITHM

This section presents the development of the hierarchical prior model corresponding to the Student-t prior. The equations corresponding to the prior model are presented and then the algorithms corresponding to the JMAP and PM estimation.

- the hierarchical model is using as a *a priori* model the **Student-t** distribution;
- the Student-t prior distribution is expressed via iGSM (3), considering the variance vector $\boldsymbol{v}_z$ as unknown and each element Inverse Gamma distributed;
- the likelihood is derived from the distribution proposed for modeling the uncertainties vector $\boldsymbol{\epsilon}$;
- for the noise and uncertainties vectors $\boldsymbol{\epsilon}$ and $\boldsymbol{\xi}$, non-stationary Student-t models are proposed, i.e. a (multivariate) Student-t distribution expressed via iGSM (3), again with the Inverse Gamma distribution modeling each element of the corresponding variances. Equation (4) presents the Bayesian hierarchical prior model: heavy tailed distribution in order to model the sparse structure of $\boldsymbol{z}$, in particular the Student-t prior, expressed as an iGSM and non-stationary models for the noise and uncertainties, both expressed using the same iGSM model.

$$
\text{Student-t HHBM}\quad
\begin{cases}
\begin{cases}
p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{v}_\epsilon) = \mathcal{N}(\boldsymbol{y}|\boldsymbol{H}\boldsymbol{x},\boldsymbol{V}_\epsilon) \propto \prod_{i=1}^{N} v_{\epsilon_i}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\|\boldsymbol{V}_\epsilon^{-\frac{1}{2}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})\|^2\right\},\\
p(\boldsymbol{v}_\epsilon|\alpha_\epsilon,\beta_\epsilon) = \prod_{i=1}^{N}\mathcal{IG}(v_{\epsilon_i}|\alpha_\epsilon,\beta_\epsilon) \propto \prod_{i=1}^{N} v_{\epsilon_i}^{-\alpha_\epsilon-1}\exp\left\{-\sum_{i=1}^{N}\frac{\beta_\epsilon}{v_{\epsilon_i}}\right\},\\
\qquad \boldsymbol{V}_\epsilon = \operatorname{diag}[\boldsymbol{v}_\epsilon],\ \boldsymbol{v}_\epsilon = [\ldots, v_{\epsilon_i},\ldots],
\end{cases}\\
\begin{cases}
p(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{v}_\xi) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{D}\boldsymbol{z},\ \boldsymbol{V}_\xi) \propto \det(\boldsymbol{V}_\xi)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\|\boldsymbol{V}_\xi^{-\frac{1}{2}}(\boldsymbol{x}-\boldsymbol{D}\boldsymbol{z})\|^2\right\},\\
p(\boldsymbol{v}_\xi|\alpha_f,\beta_f) = \prod_{j=1}^{M}\mathcal{IG}(v_{\xi_j}|\alpha_\xi,\beta_\xi) \propto \prod_{j=1}^{M} v_{\xi_j}^{-\alpha_\xi-1}\exp\left\{-\sum_{j=1}^{M}\frac{\beta_\xi}{v_{\xi_j}}\right\},\\
\qquad \boldsymbol{V}_\xi = \operatorname{diag}[\boldsymbol{v}_\xi],\ \boldsymbol{v}_\xi = [\ldots, v_{\xi_j},\ldots],
\end{cases}\\
\begin{cases}
p(\boldsymbol{z}|0,\boldsymbol{v}_z) = \mathcal{N}(\boldsymbol{z}|0,\ \boldsymbol{V}_z) \propto \det(\boldsymbol{V}_z)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\|\boldsymbol{V}_z^{-\frac{1}{2}}\boldsymbol{z}\|^2\right\},\\
p(\boldsymbol{v}_z|\alpha_z,\beta_z) = \prod_{j=1}^{M}\mathcal{IG}(v_{z_j}|\alpha_z,\beta_z) \propto \prod_{j=1}^{M} v_{z_j}^{-\alpha_z-1}\exp\left\{-\sum_{j=1}^{M}\frac{\beta_z}{v_{z_j}}\right\},\\
\qquad \boldsymbol{V}_z = \operatorname{diag}[\boldsymbol{v}_z],\ \boldsymbol{v}_z = [\ldots, v_{z_j},\ldots].
\end{cases}
\end{cases}
\tag{4}
$$

The hierarchical model showed in Figure 1, which is build over the linear forward model, Equation (2), using as a prior for z_j a Student-t distribution, expressed via the iGSM (3) and modelling the noise and uncertainties of the forward model $\boldsymbol{\epsilon}$ and $\boldsymbol{\xi}$, using a non-stationary Student-t model, is presented in Equation (4). The posterior distribution

$$
\textbf{Posterior} = p(\boldsymbol{x},\boldsymbol{z},\boldsymbol{v}_\epsilon,\boldsymbol{v}_\xi,\boldsymbol{v}_z|\boldsymbol{y},\alpha_\epsilon,\beta_\epsilon,\alpha_\xi,\beta_\xi,\alpha_z,\beta_z) \tag{5}
$$

is obtained via the Bayes rule, Equation (6):

$$
\begin{aligned}
\textbf{Posterior} &\propto p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{v}_\epsilon)\ p(\boldsymbol{v}_\epsilon|\alpha_\epsilon\beta_\epsilon)\ p(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{v}_\xi)\ p(\boldsymbol{v}_\xi|\alpha_f,\beta_f)\ p(\boldsymbol{z}|0,\boldsymbol{v}_z)\ p(\boldsymbol{v}_z|\alpha_z,\beta_z)\\
&\propto \prod_{i=1}^{N} v_{\epsilon_i}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\|\boldsymbol{V}_\epsilon^{-\frac{1}{2}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})\|^2\right\} \prod_{i=1}^{N} v_{\epsilon_i}^{-\alpha_\epsilon-1}\exp\left\{-\sum_{i=1}^{N}\frac{\beta_\epsilon}{v_{\epsilon_i}}\right\}\\
&\quad \det(\boldsymbol{V}_\xi)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\|\boldsymbol{V}_\xi^{-\frac{1}{2}}(\boldsymbol{x}-\boldsymbol{D}\boldsymbol{z})\|^2\right\} \prod_{j=1}^{M} v_{\xi_j}^{-\alpha_\xi-1}\exp\left\{-\sum_{j=1}^{M}\frac{\beta_\xi}{v_{\xi_j}}\right\}\\
&\quad \det(\boldsymbol{V}_z)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\|\boldsymbol{V}_z^{-\frac{1}{2}}\boldsymbol{z}\|^2\right\} \prod_{j=1}^{M} v_{z_j}^{-\alpha_z-1}\exp\left\{-\sum_{j=1}^{M}\frac{\beta_z}{v_{z_j}}\right\}
\end{aligned}
\tag{6}
$$

The goal is to jointly estimate the unknowns of the hierarchical model, namely x and z, the main unknowns of the linear forward model (2), and the three variances appearing in the hierarchical model, the variance corresponding to the sparse structure z, namely v_z and the two variances corresponding to the uncertainties of model ϵ, and ξ, namely v_ϵ and v_ξ.

A.1 Joint MAP estimation

The unknowns are estimated on the basis of the available data y, by maximizing the posterior distribution:

$$\begin{aligned}(\hat{x}, \hat{z}, \hat{v}_\xi, \hat{v}_\epsilon, \hat{v}_z) &= \underset{(x, z, v_\xi, v_\epsilon, v_z)}{\arg\max} \quad p(x, z, v_\xi, v_\epsilon, v_z|y, \alpha_\xi, \beta_\xi, \alpha_\epsilon, \beta_\epsilon, \alpha_z, \beta_z) \\ &= \underset{(x, z, v_\xi, v_\epsilon, v_z)}{\arg\min} \quad \mathcal{L}(x, z, v_\xi, v_\epsilon, v_z),\end{aligned} \tag{7}$$

where for the second equality the criterion $\mathcal{L}(x, z, v_\xi, v_\epsilon\ v_z)$ is defined as:

$$\mathcal{L}(x, z, v_\xi, v_\epsilon, v_z) = -\ln p(x, z, v_\xi, v_\epsilon, v_z|y, \alpha_\xi, \beta_\xi, \alpha_\epsilon, \beta_\epsilon, \alpha_z, \beta_z) \tag{8}$$

The MAP estimation corresponds to the solution minimizing the criterion $\mathcal{L}(x, z, v_\xi, v_\epsilon, v_z)$. From the analytical expression of the posterior distribution, Equation (6) and the definition of the criterion $\mathcal{L}$, Equation (8), we obtain:

$$\begin{aligned}\mathcal{L}(x, z, v_\xi, v_\epsilon, v_z) &= -\ln p(x, z, v_\xi, v_\epsilon, v_z|y, \alpha_\xi, \beta_\xi, \alpha_\epsilon, \beta_\epsilon, \alpha_z, \beta_z) \\ &+ \frac{1}{2}\|V_\epsilon^{-\frac{1}{2}}(y - Hx)\|^2 + \left(\alpha_\epsilon + \frac{3}{2}\right)\sum_{i=1}^N \ln v_{\epsilon_i} + \sum_{i=1}^N \frac{\beta_\epsilon}{v_{\epsilon_i}} \\ &+ \frac{1}{2}\|V_\xi^{-\frac{1}{2}}(x - Dz)\|^2 + \left(\alpha_\xi + \frac{3}{2}\right)\sum_{i=1}^N \ln v_{\xi_i} + \sum_{i=1}^N \frac{\beta_\xi}{v_{\xi_i}} \\ &+ \frac{1}{2}\|V_z^{-\frac{1}{2}} z\|^2 + \left(\alpha_z + \frac{3}{2}\right)\sum_{j=1}^M \ln v_{z_j} + \sum_{j=1}^M \frac{\beta_f}{v_{z_j}}\end{aligned} \tag{9}$$

One of the simplest optimisation algorithm that can be used is an alternate optimization of the criterion $\mathcal{L}(x, z, v_\xi, v_\epsilon, v_z)$ with respect to the each unknown:

- With respect to x:

$$\begin{aligned}\frac{\partial \mathcal{L}(x, z, v_\xi, v_\epsilon, v_z)}{\partial x} = 0 &\Leftrightarrow \frac{\partial}{\partial x}\left(\|V_\epsilon^{-\frac{1}{2}}(y - Hx)\|^2 + \|V_\xi^{-\frac{1}{2}}(x - Dz)\|^2\right) = 0 \\ &\Leftrightarrow -H^T V_\epsilon^{-1}(y - Hx) + V_\xi^{-1}(x - Dz) = 0 \\ &\Leftrightarrow \left(H^T V_\epsilon^{-1} H + V_\xi^{-1}\right) x = H^T V_\epsilon^{-1} y + V_\xi^{-1} D z \\ &\Rightarrow \hat{x} = \left(H^T V_\epsilon^{-1} H + V_\xi^{-1}\right)^{-1}\left(H^T V_\epsilon^{-1} y + V_\xi^{-1} D z\right)\end{aligned}$$

- With respect to z:

$$\begin{aligned}\frac{\partial \mathcal{L}(x, z, v_\xi, v_\epsilon, v_z)}{\partial z} = 0 &\Leftrightarrow \frac{\partial}{\partial x}\left(\|V_\xi^{-\frac{1}{2}}(x - Dz)\|^2 + \|V_z^{-\frac{1}{2}} z\|^2\right) = 0 \\ &\Leftrightarrow -D^T V_\xi^{-1}(x - Dz) + V_z^{-1} z = 0 \\ &\Leftrightarrow \left(D^T V_\xi^{-1} D + V_z^{-1}\right) z = D^T V_\xi^{-1} x \\ &\Rightarrow \hat{z} = \left(D^T V_\xi^{-1} D + V_z^{-1}\right)^{-1} D^T V_\xi^{-1} x\end{aligned}$$

- With respect to v_{ϵ_i}, $i \in \{1, 2, \ldots, N\}$:
  First, we develop the norm $\|V_\epsilon^{-\frac{1}{2}}(y - Hx)\|^2$:

$$\begin{aligned}\|V_\epsilon^{-\frac{1}{2}}(y - Hx)\|^2 &= y^T V_\epsilon^{-1} y - 2 y^T V_\epsilon^{-1} H x + H^T x^T V_\epsilon^{-1} H x \\ &= \sum_{i=1}^N v_{\epsilon_i}^{-1} g_i^2 - 2\sum_{i=1}^N v_{\epsilon_i}^{-1} g_i H_i x + \sum_{i=1}^N v_{\epsilon_i}^{-1} x^T H_i^T H_i x,\end{aligned}$$

where $\mathbf{H}_i$ denotes the line i of the matrix $\mathbf{H}$, $i \in \{1, 2, \ldots, N\}$, i.e. $\mathbf{H}_i = [h_{i1}, h_{i1}, \ldots, h_{iM}]$.

$$\frac{\partial \mathcal{L}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{v}_\xi, \boldsymbol{v}_\epsilon, \boldsymbol{v}_z)}{\partial v_{\epsilon_i}} = 0 \Leftrightarrow \frac{\partial}{\partial v_{\epsilon_i}} \left[\left(\alpha_\epsilon + \frac{3}{2} \right) \ln v_{\epsilon_i} + \left(\beta_\epsilon + \frac{1}{2} \left(g_i^2 - 2 g_i \mathbf{H}_i \boldsymbol{x} + \boldsymbol{x}^T \mathbf{H}_i{}^T \mathbf{H}_i \boldsymbol{x} \right) \right) v_{\epsilon_i}^{-1} \right] = 0$$
$$\Leftrightarrow \left(\alpha_\epsilon + \frac{3}{2} \right) v_{\epsilon_i} - \left(\beta_\epsilon + \frac{1}{2} \left(g_i - \mathbf{H}_i \boldsymbol{x} \right)^2 \right) = 0$$
$$\Rightarrow \widehat{v}_{\epsilon_i} = \frac{\beta_\epsilon + \frac{1}{2} \left(g_i - \mathbf{H}_i \boldsymbol{x} \right)^2}{\alpha_\epsilon + \frac{3}{2}}$$

- With respect to v_{ξ_j}, $j \in \{1, 2, \ldots, M\}$:
  First, we develop the norm $\| \mathbf{V}_\xi{}^{-\frac{1}{2}} (\boldsymbol{x} - \mathbf{D}\boldsymbol{z}) \|^2$:

$$\| \mathbf{V}_\xi{}^{-\frac{1}{2}} (\boldsymbol{x} - \mathbf{D}\boldsymbol{z}) \|^2 = \boldsymbol{x}^T \mathbf{V}_\xi{}^{-1} \boldsymbol{x} - 2 \boldsymbol{x}^T \mathbf{V}_\xi{}^{-1} \mathbf{D}\boldsymbol{z} + \mathbf{D}^T \boldsymbol{z}^T \mathbf{V}_\xi{}^{-1} \mathbf{D}\boldsymbol{z}$$
$$= \sum_{j=1}^{M} v_{\xi_j}^{-1} f_j^2 - 2 \sum_{j=1}^{M} v_{\xi_j}^{-1} f_j D_j \boldsymbol{z} + \sum_{j=1}^{M} v_{\xi_j}^{-1} \boldsymbol{z}^T D_j{}^T D_j \boldsymbol{z},$$

where $\mathbf{D}_j$ denotes the line j of the matrix $\mathbf{D}$, $j \in \{1, 2, \ldots, M\}$, i.e. $\mathbf{D}_j = \left[d_{j1}, d_{j1}, \ldots, d_{jM} \right]$.

$$\frac{\partial \mathcal{L}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{v}_\xi, \boldsymbol{v}_\epsilon, \boldsymbol{v}_z)}{\partial v_{\xi_j}} = 0 \Leftrightarrow \frac{\partial}{\partial v_{\xi_j}} \left[\left(\alpha_\xi + \frac{3}{2} \right) \ln v_{\xi_j} + \left(\beta_\xi + \frac{1}{2} \left(f_j^2 - 2 f_j \mathbf{D}_j \boldsymbol{z} + \boldsymbol{z}^T \mathbf{D}_j{}^T \mathbf{H}_j \boldsymbol{z} \right) \right) v_{\xi_j}^{-1} \right] = 0$$
$$\Leftrightarrow \left(\alpha_\xi + \frac{3}{2} \right) v_{\xi_j} - \left(\beta_\xi + \frac{1}{2} \left(f_j - \mathbf{D}_i \boldsymbol{z} \right)^2 \right) = 0$$
$$\Rightarrow \widehat{v}_{\xi_j} = \frac{\beta_\xi + \frac{1}{2} \left(f_j - \mathbf{D}_j \boldsymbol{z} \right)^2}{\alpha_\xi + \frac{3}{2}}$$

- With respect to v_{zj}, $j \in \{1, 2, \ldots, M\}$:

$$\frac{\partial \mathcal{L}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{v}_\xi, \boldsymbol{v}_\epsilon, \boldsymbol{v}_z)}{\partial v_{z_j}} = 0 \Leftrightarrow \frac{\partial}{\partial v_{z_j}} \left[\left(\alpha_z + \frac{3}{2} \right) \ln v_{z_j} + \left(\beta_z + \frac{1}{2} z_j{}^2 \right) v_{z_j}^{-1} \right] = 0$$
$$\Leftrightarrow \left(\alpha_z + \frac{3}{2} \right) v_{z_j} - \left(\beta_z + \frac{1}{2} z_j{}^2 \right) = 0$$
$$\Rightarrow \widehat{v}_{z_j} = \frac{\beta_z + \frac{1}{2} z_j{}^2}{\alpha_z + \frac{3}{2}}$$

Algorithm 1 Joint MAP - Student-t hierarchical model, non-stationary Student-t uncertainties model

Ensure: INITIALIZATION $\mathbf{x}^{(0)}, \mathbf{z}^{(0)}$

function JMAP($\alpha_\xi, \beta_\xi, \alpha_\epsilon, \beta_\epsilon, \alpha_z, \beta_z, \mathbf{y}, \mathbf{H}, \mathbf{D}, \mathbf{x}^{(0)}, \mathbf{z}^{(0)}, M, N, NoIter$)

  for $n = 0$ to $NoIter$ **do**

    for $j = 1$ to M **do**

      $\widehat{v}_{\xi_j}^{(n)} = \frac{\beta_\xi + \frac{1}{2}\left(\widehat{f}_j^{(n)} - \mathbf{D}_j \widehat{\mathbf{z}}^{(n)}\right)^2}{\alpha_\xi + \frac{3}{2}}$

    end for

    for $i = 1$ to N **do**

      $\widehat{v}_{\epsilon_i}^{(n)} = \frac{\beta_\epsilon + \frac{1}{2}\left(g_i - \mathbf{H}_i \widehat{\mathbf{x}}^{(n)}\right)^2}{\alpha_\epsilon + \frac{3}{2}}$

    end for

    for $j = 1$ to M **do**

      $\widehat{v}_{z_j}^{(n)} = \frac{\beta_z + \frac{1}{2}\left(z_j^{(n)}\right)^2}{\alpha_z + \frac{3}{2}}$

    end for

    $\widehat{\mathbf{x}}^{(n+1)} = \left(\mathbf{H}^T \operatorname{diag}\left[\left(\widehat{v}_{\epsilon_i}^{(n)}\right)^{-1}\right] \mathbf{H} + \operatorname{diag}\left[\left(\widehat{v}_{\xi_j}^{(n)}\right)^{-1}\right]\right)^{-1} \left(\mathbf{H}^T \operatorname{diag}\left[\left(\widehat{v}_{\epsilon_i}^{(n)}\right)^{-1}\right] \mathbf{y} + \operatorname{diag}\left[\left(\widehat{v}_{\xi_j}^{(n)}\right)^{-1}\right] \mathbf{D}\widehat{\mathbf{z}}\right)$

    $\widehat{\mathbf{z}}^{(n+1)} = \left(\mathbf{D}^T \operatorname{diag}\left[\left(\widehat{v}_{\xi_j}^{(n)}\right)^{-1}\right] \mathbf{D} + \operatorname{diag}\left[\left(\widehat{v}_{z_j}^{(n)}\right)^{-1}\right]\right)^{-1} \mathbf{D}^T \operatorname{diag}\left[\left(\widehat{v}_{\xi_j}^{(n)}\right)^{-1}\right] \widehat{\mathbf{x}}^{(n+1)}$

  end for

  return $\left(\widehat{\mathbf{x}}^{(NoIter+1)}, \widehat{\mathbf{z}}^{(NoIter+1)}, \widehat{v}_{\xi_j}^{(NoIter)}, \widehat{v}_{\epsilon_i}^{(NoIter)}, \widehat{v}_{z_j}^{(NoIter)}\right)$

end function

B ALGORITHMS CONTEXT AND BIBLIOGRAPHY

Ref.	Method	Sparse coefficient $\mathbf{D}$	Noise model ϵ	Estimation method	Additional uncertainty ξ	Data
[2]	HHBM	Haar	Stationary	JMAP & VBA	Non	2D
[3] [4] A.1	HHBM	Haar	Non-stationary	JMAP	Yes	2D
[5] [6]	HHBM	Haar	Non-stationary	JMAP	Yes	3D
IEEE	HHBM	Haar	Non-stationary Hyperparam. analysis	JMAP	Yes	3D

Table 2: Algorithms, context, bibliography

REFERENCES

[1] Li Wang. *Fast and accurate 3D X-ray image reconstruction for Non Destructive Testing industrial applications*. PhD thesis, Université Paris–Saclay, 2017.

[2] Li Wang, Ali Mohammad-Djafari, Nicolas Gac, and Mircea Dumitru. Computed tomography reconstruction based on a hierarchical model and variational Bayesian method. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 883–887. IEEE, 2016.

[3] Li Wang, Ali Mohammad-Djafari, and Nicolas Gac. Bayesian X-ray computed tomography using a three-level hierarchical prior model. In *AIP Conference Proceedings*, volume 1853, page 060003. AIP Publishing, 2017.

[4] Li Wang, Ali Mohammad-Djafari, and Nicolas Gac. X-ray Computed Tomography using a sparsity enforcing prior model based on haar transformation in a Bayesian framework. *Fundamenta Informaticae XX*, 2017.

[5] Li Wang, Ali Mohammad-Djafari, Nicolas Gac, and Mircea Dumitru. 3D X-ray Computed Tomography reconstruction using sparsity enforcing hierarchical model based on Haar transformation. In *The 2017 International Conference on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2017.

[6] Li Wang, Nicolas Gac, and Ali Mohammad-Djafari. Reconstruction 3D en tomographie à rayons X à l'aide d'un modèle a priori hiérarchique utilisant la transformation de Haar. In *Colloque GRETSI 2017*, 2017.