Bayesian Inference with Error Variable Splitting and Sparsity Enforcing Priors for Linear Inverse Problems

Ali Mohammad-Djafari, Mircea Dumitru, Camille Chapdelaine, Nicolas Gac

Bayesian inference with error variable splitting and sparsity enforcing priors for linear inverse problems

Ali Mohammad-Djafari, Mircea Dumitru, Camille Chapdelaine, Nicolas Gac
Laboratoire des signaux et systèmes (L2S),
CNRS-CentraleSupélec-Univ Paris Saclay, 91192 Gif-sur-Yvette, France
Email: djafari@lss.supelec.fr

***Abstract*—Regularization and Bayesian inference based methods have been successfully applied for linear inverse problems. In these methods, often simple Gaussian or Poisson models for the forward model errors have been considered. In this work, we use variable splitting for the errors to model different sources of errors and their possible non-stationarity or impulsive nature using Student-t or other heavy tailed distributions. Also, as a prior model, a sparsity enforcing hierarchical model of Infinite Gaussian Mixture model is introduced. With these prior models, we obtain a complete Bayesian inference framework which can efficiently be implemented for any linear inverse problem. Interestingly, many recent regularization-based algorithms such as Alternating Direction Method of Multipliers (ADMM) as well as more classical Bayesian based methods such as Sparse Bayesian Learning (SBL) are obtained as particular cases. One advantage of the Bayesian approach is the possibility to estimate, jointly with the reconstruction, the hyper-parameters such as the regularization parameter, thus the capability of proposing unsupervised methods. Examples of implementation of the proposed method in different signal and image processing such as deconvolution in mass spectrometry, estimation of periodic components estimation in biological signals and computed tomography are mentioned and referenced.**

keywords: Variable splitting, Bayesian inference; Sparsity enforcing, Inverse problems, Approximate Bayesian computation (ABC)

I. Introduction

The classical single error term forward model for linear inverse problems is:

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}, \tag{1}$$

where all the uncertainties are summarized by $\boldsymbol{\epsilon}$. For this simple model, nowadays, almost everything has been told, starting by deterministic methods: Least Squares (LS), then Quadratic Regularization (QR), L_1 regularization, Total Variation and all associated optimization algorithms such as Augmented Lagrangian (AL), ADMM, ISTA, FISTA, etc. [1], [2], But, also the probabilistic methods and in particular the Bayesian approach with simple Gaussian models for the noise and Gaussian prior model, Double Exponential prior, Student-t prior [3] to much more sophisticated Hierarchical models [4], [5], [6].

However, in many real applications, it is needed to propose forward models which can account for other sources of uncertainties. For example, if we want to distinguish between the measurement noise and model uncertainties, we can propose the following variable splitting model:

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\xi} + \boldsymbol{\epsilon} \tag{2}$$

which can also be written as:

$$\left\{\begin{array}{l} \boldsymbol{g} = \boldsymbol{g}_0 + \boldsymbol{\epsilon}, \\ \boldsymbol{g}_0 = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\xi}, \end{array}\right. \tag{3}$$

where $\boldsymbol{\epsilon}$ represents the measurement noise and $\boldsymbol{\xi}$ represents the modelling errors [7].

In this paper, we consider this noise splitting forward model and propose to model $\boldsymbol{\epsilon}$ by a centered white Gaussian model $\mathcal{N}(\boldsymbol{\epsilon}|\boldsymbol{0}, v_\epsilon \boldsymbol{I})$ with unknown variance v_ϵ and $\boldsymbol{\xi}$ with a heavy tailed Student-t model as we want it to represent a sparse error term. The choice of Student-t is motivated by the equivalent Infinite Gaussian Scaled Mixture (IGSM) model of the Student-t.

For the unknown $\boldsymbol{f}$ also, we propose the following hierarchical model

$$\boldsymbol{f} = \boldsymbol{D}\boldsymbol{z} + \boldsymbol{\zeta} \tag{4}$$

where $\boldsymbol{D}$ is any linear transformation which can be invertible, unitary or represent an overdetermined Dictionary and $\boldsymbol{\zeta}$ an error term for this prior model. We propose to model it by a Gaussian $\mathcal{N}(\boldsymbol{\zeta}|\boldsymbol{0}, v_\zeta \boldsymbol{I})$ with unknown variance v_ζ. Finaly, as $\boldsymbol{z}$ is, in general, sparse, we model it again via the Student-t probability distribution with its equivalent hierarchical IGSM model.

The rest of this paper is organized as follows: In section 2, we present the details of the proposed forward model, expressions of the likelihood, priors and the posterior. In Section 3, the MAP estimator and the details of the optimization algorithm are presented. In Section 4, the details of the VBA algorithm are presented. In section 5, some applications where we have implemented the proposed methods are just mentionned, and finally in section 6, the conclusions are described.

II. Proposed Forward model

The proposed forward model can be summarized in the following three equations:

$$\left\{\begin{array}{l} \boldsymbol{g} = \boldsymbol{g}_0 + \boldsymbol{\epsilon}, \\ \boldsymbol{g}_0 = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\xi}, \\ \boldsymbol{f} = \boldsymbol{D}\boldsymbol{z} + \boldsymbol{\zeta}, \end{array}\right. \tag{5}$$

where $\boldsymbol{g}$ represents the observed data, $\boldsymbol{\epsilon}$ the measurement errors, $\boldsymbol{g}_0$ the ideal no-noise data, $\boldsymbol{H}$ the forward model, $\boldsymbol{\xi}$ the model error, $\boldsymbol{f}$ the interesting real unknowns of the model, $\boldsymbol{D}$ any linear transformation model, $\boldsymbol{z}$ the corresponding representation of the unknowns in this transformed domain which is, in general, sparse and finally, $\boldsymbol{\zeta}$ the errors of this transformation [8].

With this model, a first deterministic regularization based method can be proposed by defining the following criterion

$$\begin{aligned} J(\boldsymbol{f},\boldsymbol{g}_0,\boldsymbol{z}) = \quad & \tfrac{1}{2v_\epsilon}\|\boldsymbol{g}-\boldsymbol{g}_0\|_2^2 + \tfrac{1}{2v_\xi}\|\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}\|_2^2 \\ & + \tfrac{1}{2v_\zeta}\|\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\|_2^2 + \alpha\|\boldsymbol{z}\|_1 \end{aligned} \tag{6}$$

and trying to optimize it. This criterion can be assimilated as the criterion of the Maximum A Posteriori (MAP) estimate in a Bayesian framework

$$(\widehat{\boldsymbol{f}},\widehat{\boldsymbol{g}}_0,\widehat{\boldsymbol{z}}) = \underset{(\boldsymbol{f},\boldsymbol{g}_0,\boldsymbol{z})}{\arg\max}\{p(\boldsymbol{f},\boldsymbol{g}_0,\boldsymbol{z}|\boldsymbol{g})\} \tag{7}$$

where $p(\boldsymbol{f},\boldsymbol{g}_0,\boldsymbol{z}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{g}_0)p(\boldsymbol{g}_0|\boldsymbol{f})p(\boldsymbol{f}|\boldsymbol{z})p(\boldsymbol{z})$ with the following expressions:

$$\left\{\begin{array}{l} p(\boldsymbol{\epsilon}) = \mathcal{N}(\boldsymbol{\epsilon}|\boldsymbol{0},v_\epsilon\boldsymbol{I}) \rightarrow p(\boldsymbol{g}|\boldsymbol{g}_0) = \mathcal{N}(\boldsymbol{\epsilon}|\boldsymbol{g}_0,v_\epsilon\boldsymbol{I}) \\ p(\boldsymbol{\xi}) = \mathcal{N}(\boldsymbol{\xi}|\boldsymbol{0},v_\xi\boldsymbol{I}) \rightarrow p(\boldsymbol{g}_0|\boldsymbol{f}) = \mathcal{N}(\boldsymbol{g}_0|\boldsymbol{f},v_\xi\boldsymbol{I}) \\ p\boldsymbol{\zeta}) = \mathcal{N}(\boldsymbol{\zeta}|\boldsymbol{0},v_\zeta\boldsymbol{I}) \rightarrow p(\boldsymbol{f}|\boldsymbol{z}) = \mathcal{N}(\boldsymbol{f}|\boldsymbol{z},v_\zeta\boldsymbol{I}) \\ p(\boldsymbol{z}) = \mathcal{DE}(\boldsymbol{z}|\alpha) \propto \exp\left[\alpha\|\boldsymbol{z}\|_1\right]. \end{array}\right. \tag{8}$$

The following particular cases are of great interest:

1) Quadratic regularization is obtained with $\boldsymbol{\xi}=\boldsymbol{0}$ (i.e. $\boldsymbol{g}_0=\boldsymbol{H}\boldsymbol{f}$) and $\boldsymbol{D}=\boldsymbol{0}$ (i.e. $\boldsymbol{f}=\boldsymbol{\zeta}$). The criterion becomes the classical quadratic regularization:
$$J(\boldsymbol{f}) = \frac{1}{2}\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|_2^2 + \mu\|\boldsymbol{f}\|_2^2, \quad \text{with} \quad \mu=\frac{v_\epsilon}{2v_\zeta} \tag{9}$$
2) Classical L_1 regularization is obtained with $\boldsymbol{\xi}=\boldsymbol{0}$, $\boldsymbol{D}=\boldsymbol{I}$ and $\boldsymbol{\zeta}=\boldsymbol{0}$ (i.e. $\boldsymbol{f}=\boldsymbol{z}$). The criterion becomes:
$$J(\boldsymbol{f}) = \frac{1}{2}\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|_2^2 + \mu\|\boldsymbol{f}\|_1, \text{ with } \mu=\frac{v_\epsilon}{\alpha} \tag{10}$$
3) Synthesis L_1 regularization is obtained with $\boldsymbol{\xi}=\boldsymbol{0}$ and $\boldsymbol{\zeta}=\boldsymbol{0}$ (i.e. $\boldsymbol{f}=\boldsymbol{D}\boldsymbol{z}$. The criterion becomes:
$$J(\boldsymbol{z}) = \frac{1}{2}\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{D}\boldsymbol{z}\|_2^2 + \mu\|\boldsymbol{z}\|_1, \text{ with } \boldsymbol{f}=\boldsymbol{D}\boldsymbol{z} \tag{11}$$
4) Analysis L_1 regularization is obtained with $\boldsymbol{\xi}=\boldsymbol{0}$, $\boldsymbol{D}$ is an invertible ($\boldsymbol{f}=\boldsymbol{D}\boldsymbol{z}$ and $\boldsymbol{z}=\boldsymbol{D}^{-1}\boldsymbol{f}$) or orthogonal matrix ($\boldsymbol{D}'\boldsymbol{D}=\boldsymbol{I}$, $\boldsymbol{f}=\boldsymbol{D}\boldsymbol{z}$ and $\boldsymbol{z}$=$\boldsymbol{D}$'$\boldsymbol{f}$) and $\boldsymbol{\zeta}=\boldsymbol{0}$. The criterion becomes equivalent to the classical Total Variation (TV):
$$J(\boldsymbol{f}) = \frac{1}{2}\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|_2^2 + \mu\|\boldsymbol{D}'\boldsymbol{f}\|_1, \text{ with } \mu=\frac{v_\epsilon}{\alpha} \tag{12}$$
5) The case $\boldsymbol{\xi}=\boldsymbol{0}$, i.e $\boldsymbol{g}=\boldsymbol{H}\boldsymbol{f}+\boldsymbol{\epsilon}$ and $\boldsymbol{\zeta}=\boldsymbol{0}$ where $\boldsymbol{f}=\boldsymbol{D}\boldsymbol{z}$ and where $\boldsymbol{D}$ is an ivertible or orthogonal operator, i.e. $\boldsymbol{z}=\boldsymbol{D}'\boldsymbol{f}$ is the last interesting particular case because the solution is defined by the following optimization problem:
$$J(\boldsymbol{f},\boldsymbol{z}) = \frac{1}{2v_\xi}\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|_2^2 + \alpha\|\boldsymbol{z}\|_1 \text{ s.t. } \boldsymbol{z}=\boldsymbol{D}'\boldsymbol{f} \tag{13}$$
Using the Augmented Lagrangian (AL) method to this linear constraint optimization problem consists in defining:
$$\begin{aligned} J(\boldsymbol{f},\boldsymbol{z},\lambda) \quad & = \tfrac{1}{2v_\xi}\|\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}\|_2^2 + \alpha\|\boldsymbol{z}\|_1 \\ & \quad + \|\boldsymbol{z}-\boldsymbol{D}'\boldsymbol{f}\|^2 + \lambda'(\boldsymbol{z}-\boldsymbol{D}'\boldsymbol{f}) \end{aligned} \tag{14}$$
and looking for its stationary point via the expressions of its gradients with respect to its arguments or as an alternate optimization algorithm:
$$\left\{\begin{array}{l} \boldsymbol{f} := \underset{\boldsymbol{f}}{\arg\min}\big\{\tfrac{1}{2v_\xi}\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|_2^2 + \|\boldsymbol{z}-\boldsymbol{D}'\boldsymbol{f}\|_2^2 \\ \qquad\qquad +\lambda'(\boldsymbol{z}-\boldsymbol{D}'\boldsymbol{f})\big\}, \\ \boldsymbol{z} := \underset{\boldsymbol{z}}{\arg\min}\left\{\|\boldsymbol{z}-\boldsymbol{D}'\boldsymbol{f}\|_2^2 + \alpha\|\boldsymbol{z}\|_1 + \lambda'(\boldsymbol{z}-\boldsymbol{D}'\boldsymbol{f})\right\}, \\ \lambda := \lambda + (\boldsymbol{z}-\boldsymbol{D}'\boldsymbol{f}). \end{array}\right. \tag{15}$$
where we can compare it with some particular cases of ADMM.
6) In the previous case, considering $\boldsymbol{D}=\boldsymbol{I}$, i.e. $\boldsymbol{f}=\boldsymbol{z}$, we obtain
$$\left\{\begin{array}{l} \boldsymbol{z} := \boldsymbol{z} + \alpha_1^{(k)}\left[\boldsymbol{H}'(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{z}\right] \\ \boldsymbol{f} := \underset{\boldsymbol{f}}{\arg\min}\left\{\|\boldsymbol{f}-\boldsymbol{z}\|_2^2 + \alpha\|\boldsymbol{f}\|_1\right\} = \mathrm{ST}_\alpha(\boldsymbol{z}), \end{array}\right. \tag{16}$$
where $\mathrm{ST}_\alpha(\boldsymbol{z})$ represents the Soft Thresholding (ST) of $\boldsymbol{z}$ and we thus can compare it to the well known Iterative Soft Thresholding Algorithm (ISTA) [9], [10], [11].

More generally, alternate optimization of this criterion with respect to $\boldsymbol{f}$, $\boldsymbol{g}_0$ and $\boldsymbol{z}$ becomes:
$$\left\{\begin{array}{l} \boldsymbol{f}^{(k+1)} = \boldsymbol{f}^{(k)} + \alpha_1^{(k)}\left[\boldsymbol{H}'(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}^{(k)}) - \mu\boldsymbol{f}^{(k)}\right], \\ \boldsymbol{g}_0^{(k+1)} = \boldsymbol{g}_0^{(k)} + \alpha_2^{(k)}\left[(\boldsymbol{g}_0-\boldsymbol{H}\boldsymbol{f}^{(k)})\right], \\ \boldsymbol{z}^{(k+1)} = \arg\min_{\boldsymbol{z}}\left\{\tfrac{1}{2v_\xi}\|\boldsymbol{f}-\boldsymbol{D}\boldsymbol{z}\|_2^2 + \alpha\|\boldsymbol{z}\|_1\right\} \end{array}\right. \tag{17}$$
can be compared to many special purpose optimization algorithms such as Augmented Lagragian (AL) [12], [11] and Bregman Duality (BD) [13], [14], [12], [15], [16], [11], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27].

III. Advantages of Bayesian versus Regularization approaches

The main difficulties in these regularization based methods are twofold:

- How to determine the regularization parameters, and
- How to quantify the remaining uncertainty on the obtained solution.

The Bayesian approach gives these possibilities. To see this, let consider the simplest case where
$$\left\{\begin{array}{l} \boldsymbol{g}=\boldsymbol{H}\boldsymbol{f}+\boldsymbol{\epsilon}, \quad \text{with} \quad p(\boldsymbol{\epsilon})=\mathcal{N}(\boldsymbol{\epsilon}|\boldsymbol{0},v_\epsilon\boldsymbol{I}) \\ \boldsymbol{f}=\boldsymbol{\zeta}, \quad \text{with} \quad p(\boldsymbol{f})=\mathcal{N}(\boldsymbol{f}|\boldsymbol{0},v_f\boldsymbol{I}), \end{array}\right. \tag{18}$$
For this case, we have:
$$\left\{\begin{array}{l} p(\boldsymbol{f}|\boldsymbol{g}) = \mathcal{N}(\boldsymbol{f}|\widehat{\boldsymbol{f}},\widehat{\boldsymbol{\Sigma}}) \quad \text{with} \\ \widehat{\boldsymbol{f}} = \underset{\boldsymbol{f}}{\arg\min}\left\{\|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|^2 + \mu\|\boldsymbol{f}\|^2\right\} \\ \widehat{\boldsymbol{\Sigma}} = v_\epsilon[\boldsymbol{H}'\boldsymbol{H}+\mu\boldsymbol{I}]^{-1}, \quad \mu=\frac{v_\epsilon}{v_f} \end{array}\right. \tag{19}$$

In the supervised case where v_ϵ and v_f are known, the problem is entirely completed. We have access to the posterior law and any question about $\boldsymbol{f}$ can be answered and uncertainties quantified.

In practical situations we may not know these two hyperparameters. In the Bayesian approach, it is possible to include the hyperparameters $\boldsymbol{\theta} = (v_\epsilon, v_f)$ in the estimation process by looking at the joint posterior:

$$p(\boldsymbol{f}, \boldsymbol{\theta}|g) \propto p(\boldsymbol{g}|\boldsymbol{f}, v_\epsilon) p(\boldsymbol{f}|v_f) p(\boldsymbol{\theta}) \tag{20}$$

with $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ [28], [29]. This joint posterior law can be used to infer $\boldsymbol{f}$ and $\boldsymbol{\theta}$ jointly and to quantify the uncertainty on the solution. This can be done, for example by computing the posterior covariance. However, often, the exact computation of the posterior mean and covariance may not be easy or may be too costly. Hopefully, solutions exist. For example, we can use the Variational Bayesian Approximation (VBA) methods to approximate $p(\boldsymbol{f}, \boldsymbol{\theta}|g)$ by $q(\boldsymbol{f}, \boldsymbol{\theta}) = q_1(\boldsymbol{f}) q_2(\boldsymbol{\theta})$ by choosing appropriate families for q_1 and q_2. For example, by choosing the a priori laws:

$$\left\{ \begin{array}{l} p(v_\epsilon) = \mathcal{IG}(v_f|\alpha_{\epsilon_0}, \beta_{\epsilon_0}), \\ p(v_f) = \mathcal{IG}(v_f|\alpha_{f_0}, \beta_{f_0}), \end{array} \right. \tag{21}$$

it is easy to obtain the expression of the joint posterior law:

$$p(\boldsymbol{f}, v_\epsilon, v_f|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f}, v_\epsilon) p(\boldsymbol{f}|v_f) p(v_\epsilon) p(v_f) \tag{22}$$

and by approximating it by

$$q(\boldsymbol{f}, v_\epsilon, v_f) = q_1(\boldsymbol{f}) q_2(v_\epsilon) q_3(v_f) \tag{23}$$

with $q_1(\boldsymbol{f}) = \mathcal{N}(\boldsymbol{f}|\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\Sigma}})$, $q_2(v_\epsilon) = \mathcal{IG}(v_f|\widetilde{\alpha}, \widetilde{\beta}_\epsilon)$ and $q_3(v_f) = \mathcal{IG}(v_f|\widetilde{\alpha}_f, \beta_f)$ we obtain:

$$\left\{ \begin{array}{l} q_1(\boldsymbol{f}) = \mathcal{N}(\boldsymbol{f}|\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\Sigma}}) \\ q_2(v_\epsilon) = \mathcal{IG}(v_\epsilon|\widetilde{\alpha}_\epsilon, \widetilde{\beta}_\epsilon) \\ q_3(v_f) = \mathcal{IG}(v_f|\widetilde{\alpha}_f, \widetilde{\beta}_f) \end{array} \right. \tag{24}$$

where the VBA algorithm becomes an iterative algorithm which updates successively $(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\Sigma}})$, $(\widetilde{\alpha}_\epsilon, \widetilde{\beta}_\epsilon)$ and $(\widetilde{\alpha}_f, \widetilde{\beta}_f)$. At the convergence or at the end of the iterations we get the approximate separable $q(\boldsymbol{f}, v_\epsilon, v_f) = q_1(\boldsymbol{f}) q_2(v_\epsilon) q_3(v_f)$ which we can use to infer on $\boldsymbol{f}$, v_ϵ and v_f. For the expressions of the updating these tilded variables and more details see [30], [4], [31]. Figure 1 shows the generative graphical representations of the supervised and unsupervised models.

IV. Full Bayesian framework for proposed forward models

We keep the general forward model (5) with the same Gaussian prior models for $\boldsymbol{\epsilon}$ and $\boldsymbol{\zeta}$, but letting their variances to be unknown. For $\boldsymbol{\xi}$ we choose a Student-t model to be able to control its sparsity degree. For direct sparsity enforcing of $\boldsymbol{f}$ or for sparsity in the transform domain $\boldsymbol{z}$, we propose also to use Student-t in place of Double Exponential (DE) or Generalized Gaussian (GG) due to its scaled mixture property which gives us to propose a hierarchical Normal-Inverse Gamma model. In the following these two cases are expanded.

A. Direct sparsity case

To summarize, for the direct sparsity of $\boldsymbol{f}$, we have the following relations:

1) $\boldsymbol{g} = \boldsymbol{g}_0 + \boldsymbol{\epsilon}$, $\quad \boldsymbol{\epsilon}$ Gaussian:

$$\left\{ \begin{array}{l} p(\boldsymbol{g}|\boldsymbol{g}_0, v_\epsilon) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{g}_0, v_\epsilon \boldsymbol{I}), \\ p(v_\epsilon) = \mathcal{IG}(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}). \end{array} \right. \tag{25}$$

2) $\boldsymbol{g}_0 = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\xi}$, $\quad \boldsymbol{\xi}$ Student:

$$\left\{ \begin{array}{l} p(\boldsymbol{g}_0|\boldsymbol{f}, \boldsymbol{v}_\xi) = \mathcal{N}(\boldsymbol{g}_0|\boldsymbol{H}\boldsymbol{f}, \boldsymbol{V}_\xi), \quad \boldsymbol{V}_\xi = \text{diag}\left[\boldsymbol{v}_\xi\right], \\ p(\boldsymbol{v}_\xi) = \prod_{i=1}^M p(v_{\xi_i}) = \prod_{i=1}^M \mathcal{IG}(v_{\xi_i}|\alpha_{\xi_0}, \beta_{\xi_0}). \end{array} \right. \tag{26}$$

3) $\boldsymbol{f}$ $\quad$ Sparse:

$$\left\{ \begin{array}{l} p(\boldsymbol{f}|\boldsymbol{v}_f) = \mathcal{N}(\boldsymbol{f}|\boldsymbol{0}, \boldsymbol{V}_f), \quad \boldsymbol{V}_f = \text{diag}\left[\boldsymbol{v}_f\right], \\ p(\boldsymbol{v}_f) = \prod_{j=1}^N p(v_{f_j}) = \prod_{j=1}^N \mathcal{IG}(v_{f_j}|\alpha_{f_0}, \beta_{f_0}) \end{array} \right. \tag{27}$$

which results in:

$$p(\boldsymbol{f}, \boldsymbol{g}_0, v_\epsilon, \boldsymbol{v}_\xi, \boldsymbol{v}_f|\boldsymbol{g}) \propto \exp\left[-J(\boldsymbol{f}, \boldsymbol{g}_0, v_\epsilon, \boldsymbol{v}_\xi, \boldsymbol{v}_f)\right] \tag{28}$$

with

$$\begin{aligned} J(\boldsymbol{f}, \boldsymbol{g}_0, v_\epsilon, \boldsymbol{v}_\xi, \boldsymbol{v}_f) &= \tfrac{1}{2v_\epsilon}\|\boldsymbol{g} - \boldsymbol{g}_0\|_2^2 + (\alpha_{\epsilon_0} + 1)\ln v_\epsilon + \tfrac{\beta_{\epsilon_0}}{v_\epsilon} \\ &+ \tfrac{1}{2}\|\boldsymbol{V}_\xi^{-1/2}(\boldsymbol{g}_0 - \boldsymbol{H}\boldsymbol{f})\|_2^2 + \textstyle\sum_{i=1}^M \left[(\alpha_{\xi_0} + 1)\ln v_{\xi_i} + \tfrac{\beta_{\xi_0}}{v_{\xi_i}}\right] \\ &+ \tfrac{1}{2}\|\boldsymbol{V}_f^{-1/2}\boldsymbol{f}\|_2^2 + \textstyle\sum_{j=1}^N \left[(\alpha_{f_0} + 1)\ln v_{f_j} + \tfrac{\beta_{f_0}}{v_{f_j}}\right]. \end{aligned} \tag{29}$$

Alternate optimization of this criterion results to the optimization of the following criteria:

1) with respect to $\boldsymbol{f}$:
$J(\boldsymbol{f}) = \frac{1}{2}\|\boldsymbol{V}_\xi^{-1/2}(\boldsymbol{g}_0 - \boldsymbol{H}\boldsymbol{f})\|_2^2 + \frac{1}{2v_\zeta}\|\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}\|_2^2$

2) with respect to $\boldsymbol{g}_0$:
$J(\boldsymbol{g}_0) = \frac{1}{2v_\epsilon}\|\boldsymbol{g} - \boldsymbol{g}_0\|_2^2 + \frac{1}{2}\|\boldsymbol{V}_\xi^{-1/2}(\boldsymbol{g}_0 - \boldsymbol{H}\boldsymbol{f})\|_2^2$

3) with respect to v_ϵ:
$J(v_\epsilon) = \frac{1}{2v_\epsilon}\|\boldsymbol{g} - \boldsymbol{g}_0\|_2^2 + (\alpha_{\epsilon_0} + 1)\ln v_\epsilon + \frac{\beta_{\epsilon_0}}{v_\epsilon}$

4) with respect to v_{ξ_i}:
$J(v_{\xi_i}) = \frac{1}{2}\|\boldsymbol{V}_\xi^{-1/2}(\boldsymbol{g}_0 - \boldsymbol{H}\boldsymbol{f})\|_2^2 + \sum_{i=1}^M \left[(\alpha_{\xi_0} + 1)\ln v_{\xi_i} + \frac{\beta_{\xi_0}}{v_{\xi_i}}\right]$

5) with respect to v_{f_j}:
$J(v_{f_j}) = \frac{1}{2}\|\boldsymbol{V}_f^{-1/2}\boldsymbol{f}\|_2^2 + \sum_{j=1}^N \left[(\alpha_{f_0} + 1)\ln v_{f_j} + \frac{\beta_{f_0}}{v_{f_j}}\right]$

which can be compared to ADMM like algorithms with extra advantage which is updating of the hyperparameters which makes the method unsupervized and more robust.

The second main advantage is that we have access to the expressions of all the conditional posterior laws as well as to the joint posterior law upto its normalization constant. All the conditional posterior laws are either Gaussian or Inverse Gamma which can be handled easily and we have access to their means and variances.

Finally, the third advantage is that we can do better than JMAP by using the Variational Bayesian Approximation methods and for example approximate the joint posterior law $p(\boldsymbol{f}, \boldsymbol{g}_0, v_\epsilon, \boldsymbol{v}_\xi, \boldsymbol{v}_f|\boldsymbol{g})$ by the following separable one

$$q(\boldsymbol{f}, \boldsymbol{g}_0, v_\epsilon, \boldsymbol{v}_\xi, \boldsymbol{v}_f) = q_1(\boldsymbol{f})q_2(\boldsymbol{g}_0)q_3(v_\epsilon)q_4(\boldsymbol{v}_\xi)q_5(\boldsymbol{v}_f) \quad (30)$$

using the Kullback-Leibler divergence

$$\mathrm{KL}\left(q : p\right) = \int q \ln \frac{q}{p} \quad (31)$$

which gives rise to the VBA algorithms [32], [33], [4], [34], [35], [36], [37].

The following figure shows the graphical representations of the new forward model.

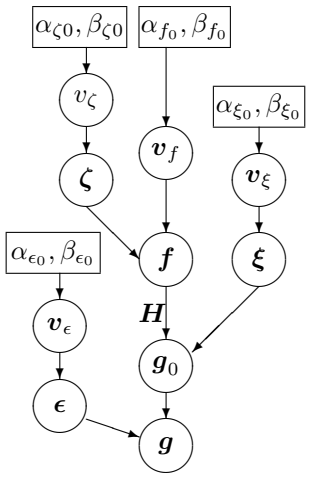

Fig. 1. Graphical representations of the proposed model.

V. Applications

The simple forward model 1 and the more detailed forward model 2 with different prior models for $\boldsymbol{f}$ (without or with another hierarchical level) and the two proposed models have been applied with succes in different application area:

- 1D biological signal processing and in particular periodic components estimation in short time biological time series [36], [5], [38].
- Signal deconvolution in Mass Spectrometry [39], [37].
- Image restoration in RAMAN Mass Spectrometry or in Radio Astronomy [39], [37].
- Detecting and estimating unknown periodic shapes in a short duration signal [39]
- 2D and 3D Industrial Computed Tomography (CT) for Non Destructive Testing (NDT) application [40], [8], [41], [42], [43], [44], [45].
- Low dose and limited angle CT for biological or medical applications [8], [46], [47], [44].

We refer the readers to the appropriate references mentionned for these applications for more details.

VI. Conclusions

In this work, we extended the classical single additive noise, linear forward model to account for different uncertainties by variable splitting techniques.

In the first step, we splitted the error term in two parts to distinguish between observation noise and forward model uncertainty and assigning to them different and appropriate prior probability distributions. Then, we examined in parallel the deterministic regularization and the Bayesian MAP approach focusing more on the second. As a by-product, we could see the links between an alternate optimization algorithm for the MAP estimator and the optimization algorithms such as Alternate Descent Minimization Maximization (ADMM) or ISTA or its fast version FISTA for the particular case of double exponential prior. The main advantage of the Bayesian approach are twofold: i) giving the tools to go further by estimating the hyperparameters and ii) being able to quantify the uncertainties of the solutions of the inversion problems. These two points are often crucial in real applications. However, one may be careful about choosing appropriate hyperparameters, the order of the optimization and the convergency of the algorithms. As the problems are in general very ill-posed, we have to carefully choose the priors to guarantee, at least, the local convergency of the algorithms. But this problem is a very general task for any inverse problem. One main question is still open: finding the reason for better performances of these variable splitting algorithms. This goes in the opposite direction of the regularization idea of restricting the space of the solution to obtain a regularized solution. As a final conclusion, we may try to answer the following question:
Is it better to restrict the space and define a criterion with a global minimum or, in the opposite, increase the dimension of the unknown space, define a criterion which may have many minima and looking for a local minimum of it?

References

[1] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–2434, 2009.

[2] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, pp. 89–97, 2004.

[3] Guobao Wang, Larry Schultz, and Jinyi Qi, "Statistical image reconstruction for muon tomography using a Gaussian scale mixture model," *IEEE Transactions on Nuclear Science*, vol. 56, no. 4, pp. 2480–2486, 2009.

[4] Ali Mohammad-Djafari, "Bayesian inference with hierarchical prior models for inverse problems in imaging systems," in *Int. Workshop on Systems, Signal Processing and Applications (WOSSPA 2013) Proceedings*, 2013, vol. Tutorial.

[5] Mircea Dumitru, Ali Mohammad-Djafari, and Baghai Simona Sain, "Precise periodic components estimation for chronobiological signals through Bayesian inference with sparsity enforcing prior," *EURASIP Journal on Bioinformatics and Systems Biology, Springer*, vol. 3, 2016.

[6] Slimane Arhab, Hacheme Ayasso, Bernard Duchêne, and Ali Mohammad-Djafari, "Optical imaging in a variational Bayesian framework," in *4th International Workshop on New Computational Methods for Inverse Problems*, Cachan, France, May 2014, p. pp 6.

[7] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2345–2356, Sept 2010.

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

[9] Pierre Charbonnier, Laure Blanc-Féraud, Gilles Aubert, and Michel Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Transactions on Image Processing*, vol. 6, no. 2, pp. 298–311, 1997.

[10] Tony F Chan, Gene H Golub, and Pep Mulet, "A nonlinear primal-dual method for total variation-based image restoration," *SIAM journal on scientific computing*, vol. 20, no. 6, pp. 1964–1977, 1999.

[11] C. He, C. Hu, W. Zhang, and B. Shi, "A fast adaptive parameter estimation for total variation image restoration," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 4954–4967, Dec 2014.

[12] D. S. Weller, S. Ramani, and J. A. Fessler, "Augmented lagrangian with variable splitting for faster non-cartesian l_1 -spirit MR image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 351–361, Feb 2014.

[13] J. Duan, Y. Liu, and L. Zhang, "Bregman iteration based efficient algorithm for mr image reconstruction from undersampled k-space data," *IEEE Signal Processing Letters*, vol. 20, no. 8, pp. 831–834, Aug 2013.

[14] Michael Elad, Peyman Milanfar, and Ron Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947, 2007.

[15] N. Zhao, Q. Wei, A. Basarab, N. Dobigeon, D. Kouamé, and J. Y. Tourneret, "Fast single image super-resolution using a new analytical solution for ℓ_2;ℓ_2 problems," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3683–3697, Aug 2016.

[16] I. Y. Chun, B. Adcock, and T. M. Talavage, "Efficient compressed sensing sense pMRI reconstruction with joint sparsity promotion," *IEEE Transactions on Medical Imaging*, vol. 35, no. 1, pp. 354–368, Jan 2016.

[17] S. Chen, H. Liu, Z. Hu, H. Zhang, P. Shi, and Y. Chen, "Simultaneous reconstruction and segmentation of dynamic pet via low-rank and sparse matrix decomposition," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 7, pp. 1784–1795, July 2015.

[18] H. Zhang, C. Wu, J. Zhang, and J. Deng, "Variational mesh denoising using total variation and piecewise constant function space," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 7, pp. 873–886, July 2015.

[19] M. J. Muckley, D. C. Noll, and J. A. Fessler, "Fast parallel MR image reconstruction via B1-Based, adaptive restart, Iterative Soft Thresholding Algorithms (barista)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 2, pp. 578–588, Feb 2015.

[20] M. J. Allison, S. Ramani, and J. A. Fessler, "Accelerated regularized estimation of MR coil sensitivities using Augmented Lagrangian methods," *IEEE Transactions on Medical Imaging*, vol. 32, no. 3, pp. 556–564, March 2013.

[21] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4484–4502, Nov 2012.

[22] S. Xie and S. Rahardja, "Alternating direction method for balanced image restoration," *IEEE Transactions on Image Processing*, vol. 21, no. 11, pp. 4557–4567, Nov 2012.

[23] S. Ramani and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical X-Ray CT," *IEEE Transactions on Medical Imaging*.

[24] S. Ramani and J. A. Fessler, "Parallel MR image reconstruction using Augmented Lagrangian methods," *IEEE Transactions on Medical Imaging*, vol. 30, no. 3, pp. 694–706, March 2011.

[25] D. Chen and L. Cheng, "Alternative minimisation algorithm for non-local total variational image deblurring," *IET Image Processing*, vol. 4, no. 5, pp. 353–364, October 2010.

[26] S. Y. Chun, Y. K. Dewaraja, and J. A. Fessler, "Alternating direction method of multiplier for tomography with nonlocal regularizers," *IEEE Transactions on Medical Imaging*, vol. 33, no. 10, pp. 1960–1968, Oct 2014.

[27] F. Sroubek and P. Milanfar, "Robust multichannel blind deconvolution via fast alternating minimization," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1687–1700, April 2012.

[28] A. Mohammad-Djafari, "On the estimation of hyperparameters in Bayesian approach of solving inverse problems," in *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, April 1993, vol. 5, pp. 495–498 vol.5.

[29] Ali Mohammad-Djafari, "A full Bayesian approach for inverse problems," in *Maximum Entropy and Bayesian Methods*. 1996, pp. 135–143, Kluwer Academic Publisher.

[30] Ali Mohammad-Djafari, "Super-resolution: a short review, a new method based on hidden Markov modeling of HR image and future challenges," *The Computer Journal*, vol. 52, no. 1, pp. 126–141, 2009.

[31] Ning Chu, A. Mohammad-Djafari, N. Gac, and J. Picheral, "A variational Bayesian approximation approach via a sparsity enforcing prior in acoustic imaging," in *Information Optics (WIO), 2014 13th Workshop on*, July 2014, pp. 1–4.

[32] Vàclav Šmídl and Anthony Quinn, *The Variational Bayes Method in Signal Processing*, Signals and Communication Technology. Springer–Verlag Berlin Heidelberg, 1st edition, 2006.

[33] Matthew J. Beal, *Variational Algorithms for Approximate Bayesian Inference*, Ph.D. thesis, University College London, 2003.

[34] S. D. Babacan, S. Nakajima, and M. N. Do, "Bayesian group-sparse modeling and variational inference," *IEEE Trans. Signal Process*, vol. 14, pp. 2906–2921, 2014.

[35] A Mohammad-Djafari, "Bayesian approach with prior models which enforce sparsity in signal and image processing," *EURASIP Journal on Advances in Signal Processing*, vol. Special issue on Sparse Signal Processing, pp. 2012:52, 2012.

[36] A. Mohammad-Djafari, "Variational Bayesian approximation for learning and inference in hierarchical models," in *Seminar at Journe rseau AIGM, Grenoble*, 2015.

[37] Ali Mohammad-Djafari, "Approximate bayesian computation tools for large scale inverse problems and hierarchical models for big data," in *booktitle=37th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Sao Paolo, Bresila, year=2017*.

[38] Ali Mohammad-Djafari, "Direct nonparametric estimation of the period and the shape of a periodic component in short duration signals," in *2017 25th European Signal Processing Conference (EUSIPCO) (EUSIPCO 2017)*, Kos, Greece, Aug. 2017.

[39] Ali Mohammad-Djafari, "Inverse problems in masse spectrometry and bayesian solutions," in *4th World Congress on Mass Spectrometry, June 19-21, 2017, London, UK*. Euro Mass Spectrometry, 2017.

[40] Li Wang, Nicolas Gac, and Ali Mohammad-Djafari, "Bayesian 3d x-ray computed tomography image reconstruction with a scaled gaussian mixture prior model," in *AIP Conference Proceedings*. AIP, 2015, vol. 1641, pp. 556–563.

[41] Camille Chapdelaine, Ali Mohammad-Djafari, Nicolas Gac, and Estelle Parra, "A 3D Bayesian Computed Tomography Reconstruction Algorithm with Gauss-Markov-Potts Prior Model and its Application to Real Data," *Fundamenta Informaticae*, 2017.

[42] Camille Chapdelaine, Ali Mohammad-Djafari, Nicolas Gac, and Estelle Parra-Denis, "Joint Reconstruction and Segmentation of Real 3D Data in Computed Tomography thanks to a Gauss-Markov-Potts prior model," in *Fully 3D*, 2017.

[43] Camille Chapdelaine, Ali Mohamad-Djafari, Nicolas Gac, and Estelle. Parra, "A Joint Segmentation and Reconstruction Algorithm for 3D Bayesian Computed Tomography Using Gauss-Markov-Potts Prior Model," in *ICASSP*, 2017.

[44] Mircea Dumitru, Li Wang, Nicolas Gac, and Ali Mohammad-Djafari, "Computed tomography reconstruction based on a hierarchical model and variational Bayesian method," in *2017 IEEE Int. Conf. on Image, Processing (ICIP)*. IEEE, 2017.

[45] 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.

[46] Mircea Dumitru, Nicolas Gac, Li Wang, and Ali Mohammad-Djafari, "Unsupervised sparsity enforcing iterative algorithms for 3d image reconstruction in x-ray computed tomography," in *The 2017 International Conference on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2017.

[47] Mircea Dumitru, Li Wang, Nicolas Gac, and Ali Mohammad-Djafari, "Comparaison des performances dalgorithmes itératifs bayésiens basés sur trois classes de modeles a priori parcimonieux appliqués à la reconstruction tomographique," in *GRETSI*, 2017.