Scientific Cooperations International Workshops on Electrical and Computer Engineering Subfields
22-23 August 2014, Koc University, ISTANBUL/TURKEY

BAYESIAN BLIND DECONVOLUTION USING A STUDENT-T PRIOR MODEL AND VARIATIONAL BAYESIAN APPROXIMATION

A. Mohammad-Djafari

Laboratoire des signaux et systèmes (L2S)
UMR 8506 CNRS-SUPELEC-UNIV PARIS SUD
plateau de Moulon, 3 rue Joliot-Curie, 91192 GIF-SUR-YVETTE Cedex, France

ABSTRACT

Deconvolution consists in estimating the input of a linear and invariant system from its output knowing its Impulse Response Function (IRF). When the IRF of the system is unknown, we are face to Blind Deconvolution. This inverse problem is ill-posed and needs prior information to obtain a satisfactory solution. Regularization theory, well known for simple deconvolution, is no more enough to obtain a satisfactory solution. Bayesian inference approach with appropriate priors on the unknown input as well as on the IRF has been used successfully, in particular with a Gaussian prior on the IRF and a sparsity enforcing prior on the input. Joint Maximum A posteriori (JMAP), Expectation-Maximization (EM) algorithm for marginalized MAP and Variational Bayesian Approximation (VBA) are the methods which have been considered recently with some advantages for the last one. In this paper, first we review these methods and give some original insights by comparing them, in particular for their respective properties, advantages and drawbacks and their computational complexity. Then, we propose to use a Student-t prior law for the unknown input which has the property of sparsity enforcing and which gives the possibility to give a hierarchical graphical structure for the generating model of the observations. Finally, we present detailed algorithms of JMAP, EM and VBA for the joint estimation of the input, the IRF and the hidden variables of the infinite Gaussian mixture model of the Student-t probability law.

Keywords
Blind Deconvolution, Bayesian JMAP, Expectation Maximization (EM), Variational Bayesian Approximation (VBA), Gaussian, Mixture of Gaussians (MoG) and Student-t prior models

I. INTRODUCTION

In a Linear and Invariant System (LIS), the output $g(t)$ can be modeled as the convolution of the input $f(t)$ with the impulse response function (IRF) $h(t)$:

$$g(t) = f(t) * h(t) + \epsilon(t), \tag{1}$$

where * represents the convolution operation and $\epsilon(t)$ the errors. The inverse problem of the deconvolution consists in estimating $f(t)$ from the output $g(t)$ when the IRF $h(t)$ of the blurring system is known a priori. This inverse problem is ill-posed and needs prior information on the input signal $f(t)$. Regularization theory and the Bayesian inversion have been successfully for this task [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19].

Blind Deconvolution consists in estimating both the input $f(t)$ and the IRF $h(t)$ from the output $g(t)$. This inverse problem is still more ill-posed and need strong prior information to obtain a satisfactory solution. Regularization theory and simple Bayesian inversion, well known, for simple deconvolution are no more enough. Bayesian inference approach with appropriate priors on the input as well as on the IRF has been used successfully, in particular with a Gaussian prior on the IRF and a sparsity enforcing prior on the input[5], [20], [21], [22].

Joint estimation of the input $f(t)$ and the IRF $h(t)$ can be done by Joint Maximum A posteriori (JMAP), Bayesian Expectation-Maximization (BEM) or the Variational Bayesian Approximation (VBA) which are three main methods which have been considered recently with some advantages for the last one [23], [24], [25], [26], [27], [28].

In this paper, first we review these methods in general and give some original insights by comparing them for the Gaussian priors model. Then, we propose to keep the Gaussian model for the IRF $h(t)$ but to use a Student-t prior model for the input $f(t)$. The Student-t model has the advantage of sparsity enforcing property and its Infinite Gaussian Mixture property gives the possibility of proposing a hierarchical structure generative graphical model for the output data. Finally, we give details of the three estimation methods of JMAP, BEM and VBA for this prior model and discuss more in detail their properties as well as their computational costs and complexities.

Even if we have applied this prior model and some of these algorithms in different 1D signal deconvolution and 2D image restoration, in this paper no particular result is shown, but the focus will be more on the comparison of these algorithms and in particular on their computational costs.

II. BASICS OF THE BAYESIAN APPROACH FOR BLIND DECONVOLUTION

Assuming a forward convolution model $g(t) = f(t) * h(t) + \epsilon(t)$ with additive noise, and discretized model, we have:

$$\boldsymbol{g} = \boldsymbol{h} * \boldsymbol{f} + \boldsymbol{\epsilon} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon} = \boldsymbol{F}\boldsymbol{h} + \boldsymbol{\epsilon}, \tag{2}$$

where $\boldsymbol{f}$ is the vector of the unknown input samples, $\boldsymbol{h}$ is the vector of the unknown IRF samples, $\boldsymbol{\epsilon}$ is the vector of errors, $\boldsymbol{H}$ is a Toeplitz convolution matrix obtained from the IRF $\boldsymbol{h}$ and $\boldsymbol{F}$ is a Toeplitz convolution matrix obtained from the samples of the input $\boldsymbol{f}$.

Using this forward model and assigning the forward $p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{h})$ and the prior laws $p(\boldsymbol{f})$ and $p(\boldsymbol{h})$, the Bayesian approach starts with the expression of the joint posterior law

$$p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g}) = \frac{p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{h})\, p(\boldsymbol{f})\, p(\boldsymbol{h})}{p(\boldsymbol{g})}. \tag{3}$$

From here, as we will see in the next sections, basically three approaches have been proposed to estimate both $\boldsymbol{f}$ and $\boldsymbol{h}$. The first one is Joint Maximum A Posteriori (JMAP):

$$(\widehat{\boldsymbol{f}},\widehat{\boldsymbol{h}}) = \arg\max_{(\boldsymbol{f},\boldsymbol{h})} \left\{p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g})\right\}. \tag{4}$$

The second one is based first on the computation of the Marginal likelihood:

$$\widehat{\boldsymbol{h}} = \arg\max_{\boldsymbol{h}} \left\{p(\boldsymbol{h}|\boldsymbol{g})\right\} \tag{5}$$

followed by the MAP estimate

$$\widehat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \left\{p(\boldsymbol{f}|\widehat{\boldsymbol{h}},\boldsymbol{g})\right\}. \tag{6}$$

But as we will see the first step needs a marginalization which is hard to do. However, the Bayesian Expectation-Maximization (BEM) method tries to find a solution to this method through an iterative algorithm which converges to a local maximum of the marginal likelihood. The third one is based on the approximation of this joint posterior in such a way that its use can be done more easily. These methods are summarized in the next three subsections.

II-A. Joint MAP

The first one is easily understood and linked to the classical regularization, if we note that:

$$\begin{aligned}(\widehat{\boldsymbol{f}},\widehat{\boldsymbol{h}}) &= \arg\max_{(\boldsymbol{f},\boldsymbol{h})} \left\{p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g})\right\} \\ &= \arg\min_{(\boldsymbol{f},\boldsymbol{h})} \left\{J_{\mathrm{MAP}}(\boldsymbol{f},\boldsymbol{h})\right\}\end{aligned} \tag{7}$$

with

$$J_{\mathrm{MAP}}(\boldsymbol{f},\boldsymbol{h}) = -\ln p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{h}) - \ln p(\boldsymbol{f}) - \ln p(\boldsymbol{h}) \tag{8}$$

which, with the following Gaussian priors :

$$\left\{\begin{array}{l} p(\boldsymbol{\epsilon}) = \mathcal{N}(\boldsymbol{\epsilon}|0, v_\epsilon \boldsymbol{I}), \\ p(\boldsymbol{f}) = \mathcal{N}(\boldsymbol{f}|0, v_f \boldsymbol{I}), \\ p(\boldsymbol{h}) = \mathcal{N}(\boldsymbol{h}|0, v_h (\boldsymbol{C}_h'\boldsymbol{C}_h)^{-1}) \end{array}\right. \tag{9}$$

becomes:

$$\begin{aligned} &J_{\mathrm{MAP}}(\boldsymbol{f},\boldsymbol{h}) \\ &\quad = \tfrac{1}{v_\epsilon}\|\boldsymbol{g} - \boldsymbol{h} * \boldsymbol{f}\|_2^2) + \tfrac{1}{v_f}\|\boldsymbol{f}\|_2^2 + \tfrac{1}{v_h}\|\boldsymbol{C}_h\boldsymbol{h}\|_2^2 \\ &\quad = \tfrac{1}{v_\epsilon}\left[\|\boldsymbol{g} - \boldsymbol{h} * \boldsymbol{f}\|_2^2) + \lambda_f\|\boldsymbol{f}\|_2^2 + \lambda_h\|\boldsymbol{C}_h\boldsymbol{h}\|_2^2\right]. \end{aligned} \tag{10}$$

where $\lambda_f = \frac{v_f}{v_\epsilon}$ and $\lambda_h = \frac{v_h}{v_\epsilon}$.

Noting that:

$$\|\boldsymbol{g} - \boldsymbol{h} * \boldsymbol{f}\|_2^2 = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|_2^2 = \|\boldsymbol{g} - \boldsymbol{F}\boldsymbol{h}\|_2^2,$$

its alternate optimization with respect to $\boldsymbol{f}$ (with fixed $\boldsymbol{h}$) and $\boldsymbol{h}$ (with fixed $\boldsymbol{f}$) result to:

$$\left\{\begin{array}{l} \textbf{Algorithm JMAP:} \\ \textbf{Initialization:} \\ \boldsymbol{h}^{(0)} = \boldsymbol{h}_0, \quad \boldsymbol{H} = \mathtt{Convmtx}(\boldsymbol{h}^{(0)}) \\ \textbf{Iterations:} \\ \boldsymbol{f}^{(k)} = (\boldsymbol{H}'\boldsymbol{H} + \lambda_f \boldsymbol{I})^{-1}\boldsymbol{H}'\boldsymbol{g} \\ \boldsymbol{F} = \mathtt{Convmtx}(\boldsymbol{f}^{(k-1)}) \\ \boldsymbol{h}^{(k)} = (\boldsymbol{F}'\boldsymbol{F} + \lambda_h \boldsymbol{C}_h'\boldsymbol{C}_h)^{-1}\boldsymbol{F}'\boldsymbol{g} \\ \boldsymbol{H} = \mathtt{Convmtx}(\boldsymbol{h}^{(k-1)}) \end{array}\right. \tag{11}$$

We may note that the computation of $\boldsymbol{f}^{(k)}$ and $\boldsymbol{h}^{(k)}$ can be done via gradient based optimization algorithms:

$$\left\{\begin{array}{l} \boldsymbol{f}^{(k)} = \arg\min_{\boldsymbol{f}} \left\{J_{\mathrm{MAP}}(\boldsymbol{f},\boldsymbol{h})\right\} \\ \boldsymbol{h}^{(k)} = \arg\min_{\boldsymbol{h}} \left\{J_{\mathrm{MAP}}(\boldsymbol{f},\boldsymbol{h})\right\} \end{array}\right. \tag{12}$$

II-B. Algorithm BEM

The second method, needs first the integration (marginalization):

$$p(\boldsymbol{h}|\boldsymbol{g}) = \int p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g})\, \mathrm{d}\boldsymbol{f} \tag{13}$$

which can not often be done analytically and needs approximation methods to obtain the solution. The Expectation-Maximization (EM) and its Bayesian version try to find this solution by alternate maximizing of some lower bound $p^*(\boldsymbol{h}|\boldsymbol{g})$ to it. In summary, the BEM algorithm can be written as a two step iterative algorithm:

- E step: Compute the Expected value:

$$Q(\boldsymbol{h},\boldsymbol{h}^{(k-1)}) = \langle \ln p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g})\rangle_{p(\boldsymbol{f}|\boldsymbol{h}^{(k-1)},\boldsymbol{g})} \tag{14}$$

  which is now considered as a function of $\boldsymbol{h}$.

- M step (Maximization):

$$\boldsymbol{h}^{(k)} = \arg\max_{\boldsymbol{h}} \left\{Q(\boldsymbol{h},\boldsymbol{h}^{(k-1)})\right\} \tag{15}$$

It is shown that, subject to some mild conditions, this algorithm converges to a local maximum of the marginal likelihood.

For the Gaussian case, noting that

$$\begin{aligned} -\ln p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g}) &= c + \frac{1}{2v_\epsilon} J_{\text{MAP}}(\boldsymbol{f},\boldsymbol{h}) \\ &= c + \frac{1}{2v_\epsilon}\left[\|\boldsymbol{g}-\boldsymbol{h}*\boldsymbol{f}\|_2^2) + \lambda_f\|\boldsymbol{f}\|_2^2 + \lambda_h\|\boldsymbol{C}_h\boldsymbol{h}\|_2^2\right] \end{aligned} \tag{16}$$

where c is a constant which will be eliminated since after. Now, looking at the expression of $\langle -\ln p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g})\rangle_{p(\boldsymbol{f}|\boldsymbol{h}^{(k-1)},\boldsymbol{g})}$ and keeping only the terms depending on $\boldsymbol{h}$ we obtain:

$$\begin{aligned} &< -\ln p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g}) > \\ &=< \|\boldsymbol{g}-\boldsymbol{h}*\boldsymbol{f}\|_2^2) > +\lambda_f < \|\boldsymbol{f}\|_2^2) > +\lambda_h\|\boldsymbol{C}_h\boldsymbol{h}\|_2^2 \\ &= \left[\|\boldsymbol{g}\|^2 - 2\boldsymbol{g}'\boldsymbol{H}* < \boldsymbol{f} > + < \|\boldsymbol{H}\boldsymbol{f}\|_2^2 >\right] \\ &+\lambda_f < \|\boldsymbol{f}\|_2^2) > +\lambda_h\|\boldsymbol{C}_h\boldsymbol{h}\|_2^2 \\ &= \left[\|\boldsymbol{g}\|^2 - 2\boldsymbol{g}'\boldsymbol{H} < \boldsymbol{f} > +\text{Tr}\left\{\boldsymbol{H} < \boldsymbol{f}\boldsymbol{f}' > \boldsymbol{H}'\right\}\right] \\ &+\lambda_h\|\boldsymbol{C}_h\boldsymbol{h}\|_2^2 \\ &= \Big[\|\boldsymbol{g}\|^2 - 2\boldsymbol{g}'\boldsymbol{H} < \boldsymbol{f} > \\ &+\text{Tr}\left\{\boldsymbol{H}(\text{Cov}\,[f] + < \boldsymbol{f} >< \boldsymbol{f}' >)\boldsymbol{H}'\right\}\Big] + \lambda_h\|\boldsymbol{C}_h\boldsymbol{h}\|_2^2 \\ &= \frac{1}{v_\epsilon}\Big[\|\boldsymbol{g}\|^2 - 2\boldsymbol{g}' < \boldsymbol{F} > \boldsymbol{h} + \| < \boldsymbol{F} > \boldsymbol{h}\|^2 + \\ &\text{Tr}\left\{\boldsymbol{H}\text{Cov}\,[f]\,\boldsymbol{H}'\right\}\Big] + \lambda_h\|\boldsymbol{C}_h\boldsymbol{h}\|_2^2 \\ &= \left[\|\boldsymbol{g}- < \boldsymbol{F} > \boldsymbol{h}\|^2 + \|\boldsymbol{D}_f\boldsymbol{h}\|^2\right] + \lambda_h\|\boldsymbol{C}_h\boldsymbol{h}\|_2^2 \end{aligned} \tag{17}$$

where we assumed that $\text{Tr}\left\{\boldsymbol{H}\text{Cov}\,[f]\,\boldsymbol{H}'\right\}$ can be written as $\|\boldsymbol{D}_f\boldsymbol{h}\|^2$ which is possible. Then, with this relation, it is easy to write down the Bayesian EM algorithm as follows:

$$\left\{\begin{array}{l} \textbf{Algorithm BEM:} \\ \textbf{Initialization:} \\ \boldsymbol{h}^{(0)} = \boldsymbol{h}_0, \quad \boldsymbol{H} = \texttt{Convmtx}(\boldsymbol{h}^{(0)}) \\ \textbf{Iterations:} \\ \boldsymbol{\Sigma}_f = v_\epsilon(\boldsymbol{H}'\boldsymbol{H} + \lambda_f\boldsymbol{I})^{-1} \\ \boldsymbol{f}^{(k)} = (\boldsymbol{H}'\boldsymbol{H} + \lambda_f\boldsymbol{I})^{-1}\boldsymbol{H}'\boldsymbol{g} \\ \boldsymbol{F} = \texttt{Convmtx}(\boldsymbol{f}^{(k-1)}) \\ \text{Tr}\left\{\boldsymbol{H}\boldsymbol{\Sigma}_f\boldsymbol{H}'\right\} = \|\boldsymbol{D}_h'\boldsymbol{h}\|_2^2 \\ \boldsymbol{h}^{(k)} = (\boldsymbol{F}'\boldsymbol{F} + \lambda_h\boldsymbol{C}_h'\boldsymbol{C}_h + \boldsymbol{D}_h'\boldsymbol{D}_h)^{-1}\boldsymbol{F}'\boldsymbol{g} \\ \boldsymbol{H} = \texttt{Convmtx}(\boldsymbol{h}^{(k-1)}) \end{array}\right. \tag{18}$$

II-C. Algorithm VBA

The third approach which, in some way, generalizes BEM, is the VBA method which consists in approximating the joint posterior law $p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g})$ by a separable one $q(\boldsymbol{f},\boldsymbol{h}) = q_1(\boldsymbol{f})\,q_2(\boldsymbol{h})$ by minimizing the Kullback-Leibler $\text{KL}(q:p)$. It is easily shown that the alternate optimization of this criterion results to the following iterative algorithm:

- E step: Compute the expected values $\langle\ln p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g})\rangle_{q_1}$ and $\langle\ln p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g})\rangle_{q_2}$ and deduce:

$$\begin{cases} q_1(\boldsymbol{f}|\boldsymbol{h}^{(k)}) \propto \exp\left\{\langle\ln p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g})\rangle_{q_2(\boldsymbol{h}|\boldsymbol{f}^{(k-1)})}\right\} \\ q_2(\boldsymbol{h}|\boldsymbol{f}^{(k)}) \propto \exp\left\{\langle\ln p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g})\rangle_{q_1(\boldsymbol{f}|\boldsymbol{h}^{(k-1)})}\right\} \end{cases} \tag{19}$$

- M step:

$$\begin{cases} \boldsymbol{f}^{(k+1)} = \arg\max_{\boldsymbol{f}}\left\{q_2(\boldsymbol{h}|\boldsymbol{h}^{(k)})\right\} \\ \boldsymbol{h}^{(k+1)} = \arg\max_{\boldsymbol{h}}\left\{q_2(\boldsymbol{h}|\boldsymbol{f}^{(k)})\right\} \end{cases} \tag{20}$$

Here too, it can be shown that with the Gaussian priors, we obtain the following algorithm:

$$\left\{\begin{array}{l} \textbf{Algorithm VBA:} \\ \textbf{Initialization:} \\ \boldsymbol{h}^{(0)} = \boldsymbol{h}_0; \quad \boldsymbol{H} = \texttt{Convmtx}(\boldsymbol{h}^{(0)}) \\ \boldsymbol{\Sigma}_f = v_\epsilon(\boldsymbol{H}'\boldsymbol{H} + \lambda_f\boldsymbol{I})^{-1} \\ \boldsymbol{f} = (\boldsymbol{H}'\boldsymbol{H} + \lambda_f\boldsymbol{I})^{-1}\boldsymbol{H}'\boldsymbol{g} \\ \boldsymbol{F} = \texttt{Convmtx}(\boldsymbol{f}) \\ \text{Tr}\left\{\boldsymbol{H}\boldsymbol{\Sigma}_f\boldsymbol{H}'\right\} = \|\boldsymbol{D}_h'\boldsymbol{h}\|_2^2 \\ \textbf{Iterations:} \\ \boldsymbol{\Sigma}_h = v_\epsilon(\boldsymbol{F}'\boldsymbol{F} + \lambda_h\boldsymbol{C}_h'\boldsymbol{C}_h + \boldsymbol{D}_h'\boldsymbol{D}_h)^{-1} \\ \boldsymbol{h}^{(k)} = (\boldsymbol{F}'\boldsymbol{F} + \lambda_h\boldsymbol{C}_h'\boldsymbol{C}_h + v_\epsilon\boldsymbol{D}_h'\boldsymbol{D}_h)^{-1}\boldsymbol{F}'\boldsymbol{g} \\ \boldsymbol{H} = \texttt{Convmtx}(\boldsymbol{h}^{(k-1)}) \\ \text{Tr}\left\{\boldsymbol{F}\boldsymbol{\Sigma}_h\boldsymbol{F}'\right\} = \|\boldsymbol{D}_f'\boldsymbol{f}\|_2^2 \\ \boldsymbol{\Sigma}_f = v_\epsilon(\boldsymbol{H}'\boldsymbol{H} + \lambda_f\boldsymbol{I} + v_\epsilon\boldsymbol{D}_h'\boldsymbol{D}_h)^{-1} = \|\boldsymbol{D}_f'\boldsymbol{f}_h\|_2^2 \\ \boldsymbol{f}^{(k)} = (\boldsymbol{H}'\boldsymbol{H} + \lambda_f\boldsymbol{I} + v_\epsilon\boldsymbol{D}_f'\boldsymbol{D}_f)^{-1}\boldsymbol{H}'\boldsymbol{g} \\ \boldsymbol{F} = \texttt{Convmtx}(\boldsymbol{f}^{(k-1)}) \\ \text{Tr}\left\{\boldsymbol{H}\boldsymbol{\Sigma}_f\boldsymbol{H}'\right\} = \|\boldsymbol{D}_h'\boldsymbol{f}\|_2^2 \end{array}\right. \tag{21}$$

II-D. Comparison between JMAP, BEM and VBA

Comparing the three algorithms JMAP (11), BEM (18) and VBA (21), we can make the following remarks:

- In Joint MAP, there is no need to matrix inversion. At each step, we can find $\boldsymbol{f}^{(k)}$ and $\boldsymbol{h}^{(k)}$ using an optimization algorithm.
- In BEM, at each step, wee need to compute $\boldsymbol{\Sigma}_f$ and do the matrix decomposition $\text{Tr}\left\{\boldsymbol{H}\boldsymbol{\Sigma}_f\boldsymbol{H}'\right\} = \|\boldsymbol{D}_h'\boldsymbol{h}_h\|_2^2$. This is a very costly operation due to the size of the matrix $\boldsymbol{H}'$ and the matrix $\boldsymbol{\Sigma}_f$.
- In VBA, at each step, wee need to compute $\boldsymbol{\Sigma}_f$ and do the matrix decomposition $\text{Tr}\left\{\boldsymbol{H}\boldsymbol{\Sigma}_f\boldsymbol{H}'\right\} = \|\boldsymbol{D}_h'\boldsymbol{h}_h\|_2^2$ and also to compute $\boldsymbol{\Sigma}_f$ and do the

matrix decomposition $\mathrm{Tr}\left\{\boldsymbol{F}\boldsymbol{\Sigma}_h\boldsymbol{F}'\right\} = \|\boldsymbol{D}_f'\boldsymbol{f}_h\|_2^2$. There are two very costly operations.

For practical applications, we have to write specialized algorithm taking account of the particular structures of the matrix operators $\boldsymbol{H}$ and $\boldsymbol{F}$. In particular, in Blind deconvolution, these matrices are Toeplitz (or Block-Toeplitz) and we can approximate them with appropriate circulant (or Bloc-circulant) matrices and use the Fast Fourier Transform (FFT) to write appropriate algorithms.

III. JMAP, BEM AND VBA WITH A STUDENT-T PRIOR

As we are, in general, looking for a sharp input signal, a Gaussian prior is not very appropriate. We may use any sparsity enforcing priors. Between those prior law, one is very interesting, the Student-t prior with its Infinite Gaussian Mixture (IGM) property:

$$\begin{array}{l}\mathcal{T}(f_j|\nu,\mu_j,v_f) = \\ \quad \displaystyle\int_0^\infty \mathcal{N}(f_j|\mu_j,z_j^{-1}v_f)\,\mathcal{G}(z_j|\nu/2,\nu/2)\,\mathrm{d}z_j\end{array} \tag{22}$$

where

$$\begin{array}{l}\mathcal{N}(f_j|\mu_j,z_j^{-1}v_f) = \\ \quad (\frac{2\pi v_f}{z_j})^{-1/2}\exp\left\{-\frac{1}{2v_f}z_j(f_j-\mu_j)^2\right\}\end{array} \tag{23}$$

and

$$\mathcal{G}(z_j|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} z_j^{\alpha-1}\exp\left\{-\beta z_j\right\}. \tag{24}$$

This property can be used to propose a hierarchical prior structure which can be used to propose a hierarchical graphical generating model for the observed signal $\boldsymbol{g}$ which can be summarized as follows:

$$\left\{\begin{array}{l}p(\boldsymbol{\epsilon}|v_\epsilon) = \mathcal{N}(\boldsymbol{\epsilon}|0,v_\epsilon\boldsymbol{I}) \\ p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{h},v_\epsilon) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{h}*\boldsymbol{f},v_\epsilon\boldsymbol{I}), \\ p(\boldsymbol{h}|v_h) = \mathcal{N}(\boldsymbol{h}|0,v_h(\boldsymbol{C}_h'\boldsymbol{C}_h)^{-1}) \\ p(\boldsymbol{f}|\boldsymbol{z},v_f) = \mathcal{N}(\boldsymbol{f}|0,v_f\boldsymbol{Z}^{-1}) \\ \quad \text{with } \boldsymbol{Z} = \mathrm{Diag}[z_1,\cdots,z_N] \\ p(\boldsymbol{z}|\alpha,\beta) = \prod_{j=1}^N \mathcal{G}(z_j|\alpha,\beta)\end{array}\right. \tag{25}$$

Then, from the expression of the joint posterior:

$$\begin{array}{l}p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{h}|\boldsymbol{g}) \\ \propto p(\boldsymbol{g}|\boldsymbol{h},\boldsymbol{f})\,p(\boldsymbol{h}|v_h)\,p(\boldsymbol{f}|\boldsymbol{z},v_f)\,p(\boldsymbol{z}|\alpha,\beta) \\ \propto \mathcal{N}(\boldsymbol{g}|\boldsymbol{h}*\boldsymbol{f},v_\epsilon\boldsymbol{I})\,\mathcal{N}(\boldsymbol{h}|0,v_h(\boldsymbol{C}_h'\boldsymbol{C}_h)^{-1}) \\ \quad \mathcal{N}(\boldsymbol{f}|0,v_f\boldsymbol{Z}^{-1})\prod_{j=1}^N\mathcal{G}(z_j|\alpha,\beta) \\ \propto \exp\left\{-\frac{1}{v_\epsilon}J_{\mathrm{MAP}}(\boldsymbol{f},\boldsymbol{z},\boldsymbol{h})\right\}\end{array} \tag{26}$$

we can deduce the JMAP criterion:

$$\begin{array}{l}J_{\mathrm{MAP}}(\boldsymbol{f},\boldsymbol{z},\boldsymbol{h}) = \|\boldsymbol{g}-\boldsymbol{h}*\boldsymbol{f}\|^2 + \lambda_h\|\boldsymbol{C}_h\boldsymbol{h}\|^2 \\ \quad +\lambda_f\|\boldsymbol{Z}^{1/2}\boldsymbol{f}\|^2 + 2v_\epsilon\left[\sum_{j=1}^N(\alpha-1)\ln z_j + \beta z_j\right]\end{array} \tag{27}$$

Using this expression, we can obtain easily the necessary developments to describe the the algorithms JMAP, EM and VBA for which, we added the appendix 2 to distinguish them from the Gaussian models of the last section.

$$\left\{\begin{array}{l}\textbf{Algorithm JMAP2:} \\ \textbf{Initialization:} \\ \boldsymbol{h}^{(0)} = \boldsymbol{h}_0, \quad \boldsymbol{H} = \mathtt{Convmtx}(\boldsymbol{h}^{(0)}), \quad \boldsymbol{z}^{(0)} = 1 \\ \textbf{Iterations:} \\ \boldsymbol{f}^{(k)} = (\boldsymbol{H}'\boldsymbol{H}+\lambda_f\boldsymbol{Z})^{-1}\boldsymbol{H}'\boldsymbol{g} \\ \widehat{z}_j^{(k)} = \frac{\widehat{\beta}_j}{\widehat{\alpha}_j} \\ \quad \text{with} \quad \widehat{\alpha}_j = \frac{1}{2}+\alpha \quad \text{and} \\ \qquad\qquad \widehat{\beta}_j = \beta+\frac{1}{2}+\|\boldsymbol{g}-\boldsymbol{h}*\boldsymbol{f}\|^2 \\ \boldsymbol{F} = \mathtt{Convmtx}(\boldsymbol{f}^{(k-1)}) \\ \boldsymbol{h}^{(k)} = (\boldsymbol{F}'\boldsymbol{F}+\lambda_h\boldsymbol{C}_h'\boldsymbol{C}_h)^{-1}\boldsymbol{F}'\boldsymbol{g} \\ \boldsymbol{H} = \mathtt{Convmtx}(\boldsymbol{h}^{(k-1)})\end{array}\right. \tag{28}$$

where $\lambda_f = \frac{v_f}{v_\epsilon}$ and $\lambda_h = \frac{v_h}{v_\epsilon}$.

Following the same approach and finding the expressions of $\langle -\ln p(\boldsymbol{f},\boldsymbol{h}|\boldsymbol{g})\rangle$ with respect to the marginals $p(\boldsymbol{f}|\boldsymbol{h},\boldsymbol{g})$ and $p(\boldsymbol{h}|\boldsymbol{f},\boldsymbol{g})$, we obtain the necessary relations for BEM and VBA with the proposed hierarchical IGM prior. These two algorithms are summarized in the following:

$$\left\{\begin{array}{l}\textbf{Algorithm BEM2:} \\ \textbf{Initialization:} \\ \boldsymbol{h}^{(0)} = \boldsymbol{h}_0, \quad \boldsymbol{H} = \mathtt{Convmtx}(\boldsymbol{h}^{(0)}), \quad \boldsymbol{z}^{(0)} = 1 \\ \textbf{Iterations:} \\ \boldsymbol{\Sigma}_f = v_\epsilon(\boldsymbol{H}'\boldsymbol{H}+\lambda_f\boldsymbol{Z})^{-1} \\ \boldsymbol{f}^{(k)} = (\boldsymbol{H}'\boldsymbol{H}+\lambda_f\boldsymbol{Z})^{-1}\boldsymbol{H}'\boldsymbol{g} \\ \boldsymbol{F} = \mathtt{Convmtx}(\boldsymbol{f}^{(k-1)}) \\ \mathrm{Tr}\left\{\boldsymbol{H}\boldsymbol{\Sigma}_f\boldsymbol{H}'\right\} = \|\boldsymbol{D}_h'\boldsymbol{h}_h\|_2^2 \\ \widehat{z}_j^{(k)} = \frac{\widehat{\beta}_j}{\widehat{\alpha}_j} \\ \quad \text{with} \quad \widehat{\alpha}_j = \frac{1}{2}+\alpha \quad \text{and} \\ \qquad\qquad \widehat{\beta}_j = \beta+\frac{1}{2}+\boldsymbol{g}-\boldsymbol{h}*\boldsymbol{f}\|^2 \\ \boldsymbol{h}^{(k)} = (\boldsymbol{F}'\boldsymbol{F}+\lambda_h\boldsymbol{C}_h'\boldsymbol{C}_h+\boldsymbol{D}_h'\boldsymbol{D}_h)^{-1}\boldsymbol{F}'\boldsymbol{g} \\ \boldsymbol{H} = \mathtt{Convmtx}(\boldsymbol{h}^{(k-1)})\end{array}\right. \tag{29}$$

$$
\left\{
\begin{array}{l}
\textbf{Initialization:} \\
\boldsymbol{h}^{(0)} = \boldsymbol{h}_0; \quad \boldsymbol{H} = \texttt{Convmtx}(\boldsymbol{h}^{(0)}), \quad \boldsymbol{z}^{(0)} = 1; \\
\boldsymbol{\Sigma}_f = v_\epsilon (\boldsymbol{H}'\boldsymbol{H} + \lambda_f \boldsymbol{I})^{-1} \\
\boldsymbol{f} = (\boldsymbol{H}'\boldsymbol{H} + \lambda_f \boldsymbol{I})^{-1} \boldsymbol{H}'\boldsymbol{g} \\
\boldsymbol{F} = \texttt{Convmtx}(\boldsymbol{f}) \\
\text{Tr}\left\{\boldsymbol{H}\boldsymbol{\Sigma}_f\boldsymbol{H}'\right\} = \|\boldsymbol{D}_h'\boldsymbol{h}_h\|_2^2 \\
\textbf{Iterations:} \\
\boldsymbol{\Sigma}_h = v_\epsilon (\boldsymbol{F}'\boldsymbol{F} + \lambda_h \boldsymbol{C}_h'\boldsymbol{C}_h + \boldsymbol{D}_h'\boldsymbol{D}_h)^{-1} \\
\boldsymbol{h}^{(k)} = (\boldsymbol{F}'\boldsymbol{F} + \lambda_h \boldsymbol{C}_h'\boldsymbol{C}_h + v_\epsilon \boldsymbol{D}_h'\boldsymbol{D}_h)^{-1} \boldsymbol{F}'\boldsymbol{g} \\
\boldsymbol{H} = \texttt{Convmtx}(\boldsymbol{h}^{(k-1)}) \\
\text{Tr}\left\{\boldsymbol{F}\boldsymbol{\Sigma}_h\boldsymbol{F}'\right\} = \|\boldsymbol{D}_f'\boldsymbol{f}_h\|_2^2 \\
\boldsymbol{\Sigma}_f = v_\epsilon (\boldsymbol{H}'\boldsymbol{H} + \lambda_f \boldsymbol{I} + v_\epsilon \boldsymbol{D}_h'\boldsymbol{D}_h)^{-1} = \|\boldsymbol{D}_f'\boldsymbol{f}_h\|_2^2 \\
\boldsymbol{f}^{(k)} = (\boldsymbol{H}'\boldsymbol{H} + \lambda_f \boldsymbol{I} + v_\epsilon \boldsymbol{D}_f'\boldsymbol{D}_f)^{-1} \boldsymbol{H}'\boldsymbol{g} \\
\boldsymbol{F} = \texttt{Convmtx}(\boldsymbol{f}^{(k-1)}) \\
\text{Tr}\left\{\boldsymbol{H}\boldsymbol{\Sigma}_f\boldsymbol{H}'\right\} = \|\boldsymbol{D}_h'\boldsymbol{h}_h\|_2^2 \\
\widehat{z}_j^{(k)} = \frac{\widehat{\beta}_j}{\widehat{\alpha}_j} \\
\quad \text{with} \quad \widehat{\alpha}_j = \frac{1}{2} + \alpha \quad \text{and} \\
\qquad\qquad \widehat{\beta}_j = \beta + \frac{1}{2} + \|\boldsymbol{g} - \boldsymbol{h} * \boldsymbol{f}\|^2
\end{array}
\right. \tag{30}
$$

As we can see the complexity of these algorithms with the Student-t prior compared to the equivalent Gaussian cases are not too much different. However, in real applications, still we need to do simplification. One way to go ahead is to choose a full separable approximation.

IV. CONCLUSIONS

In this paper, we considered the blind deconvolution problem in a Bayesian framework. Then, first, using Gaussian priors, we compared three main algorithms based on JMAP, BEM and VBA giving some insight and comparison between these methods using a Gaussian prior for both IRF $\boldsymbol{h}$ and the input signal $\boldsymbol{f}$. Then, using a Gaussian prior for the IRF but a Student-t prior for the images to enhance or to account for the sparsity structure of the input signal, again, we compared those three methods and their corresponding algorithms. The main conclusion can be summarized as follows:

- In JMAP, at each iteration, only the value of the previous estimate of h and f are transferred to the next step. So, in one hand, the computational cost of this approach is low because there is no need for matrix inversion. At the other hand, we do not know a lot about the convergence and the properties of the obtained solution.

- In EM, the value of the IRF h is transferred, but for $\boldsymbol{f}$, its expected value and its uncertainty (covariance matrix) are transferred for the next iteration computation of $\boldsymbol{h}$. So, in one hand, the computational cost of this approach is higher than JMAP because here we need the computation of $\boldsymbol{\Sigma}_f$ which needs a huge matrix inversion. At the other hand, we know a little more about the convergence (to local maximum of the marginal likelihood) and the properties of the obtained solution.

- In VBA, at each step, not only the values of the estimates, but also theirs uncertainties (in fact the whole approximated marginal laws) are transferred. So, in one hand, the computational cost of this approach is still higher than BEM because here we need the computation of $\boldsymbol{\Sigma}_f$ and $\boldsymbol{\Sigma}_h$ which needs two huge dimensional matrix inversion. At the other hand, not only we get the estimates of $\boldsymbol{f}$ and $\boldsymbol{h}$ but also their approximated marginals $q_1(\boldsymbol{f})$ and $q_2(\boldsymbol{h})$, from which, we can compute any statistical properties of these estimates.

V. REFERENCES

[1] T.J. Deeming, "Deconvolution and reflection coefficient estimation using a generalized minimum entropy principle," *Proc. 51st Ann. Meeting of the Soc. of Exploration Geophysicists*, p. 1, 1981.

[2] D.L. Donoho, "On minimum entropy deconvolution," *Applied Time Series Analysis II*, p. 1, 1981.

[3] G. Wahba, "Constrained regularization for ill-posed linear operator equations with applications in meteorology and medicine," *in Statistical Decision Theory and Related Topics III New-York: Academic*, pp. 383–417, 1982.

[4] G. Demoment and R. Reynaud, "Fast minimum-variance deconvolution," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. ASSP-33, pp. 1324–1326, 1985.

[5] T. F. Chan and C. K. Wong, "Total variation blind deconvolution.," *IEEE Trans Image Process*, vol. 7, no. 3, pp. 370–375, 1998.

[6] T. J. Holmes, "Blind deconvolution of quantum-limited incoherent imagery: maximum-likelihood approach.," *J Opt Soc Am A*, vol. 9, no. 7, pp. 1052–1061, Jul 1992.

[7] S. U. Pillai and B. Liang, "Blind image deconvolution using a robust gcd approach.," *IEEE Trans Image Process*, vol. 8, no. 2, pp. 295–301, 1999.

[8] M. K. Ng, R. J. Plemmons, and S. Qiao, "Regularization of rif blind image deconvolution.," *IEEE Trans Image Process*, vol. 9, no. 6, pp. 1130–1134, 2000.

[9] Stuart Jefferies, Kathy Schulze, Charles Matson, Kurt Stoltenberg, and E. Keith Hege, "Blind deconvolution in optical diffusion tomography.," *Opt Express*, vol. 10, no. 1, pp. 46–53, Jan 2002.

[10] Simone Fiori, "Fast fixed-point neural blind-deconvolution algorithm.," *IEEE Trans Neural Netw*, vol. 15, no. 2, pp. 455–459, Mar 2004.

[11] Liang Wei, Lei Hua-ming, and Que Pei-wen, "Sparsity enhancement for blind deconvolution of ultrasonic signals in nondestructive testing application.," *Rev Sci Instrum*, vol. 79, no. 1, pp. 014901, Jan 2008.

[12] Anat Levin, Yair Weiss, Fredo Durand, and William T Freeman, "Understanding blind deconvolution algo-

rithms.," *IEEE Trans Pattern Anal Mach Intell*, Jul 2011.

[13] Haiyong Liao and Michael K Ng, "Blind deconvolution using generalized cross-validation approach to regularization parameter estimation.," *IEEE Trans Image Process*, vol. 20, no. 3, pp. 670–680, Mar 2011.

[14] Andrs G Marrugo, Michal Sorel, Filip Sroubek, and Mara S Milln, "Retinal image restoration by means of blind deconvolution.," *J Biomed Opt*, vol. 16, no. 11, pp. 116016, Nov 2011.

[15] Mohammad Rostami, Oleg Michailovich, and Zhou Wang, "Image deblurring using derivative compressed sensing for optical imaging application.," *IEEE Trans Image Process*, vol. 21, no. 7, pp. 3139–3149, Jul 2012.

[16] Filip Sroubek, Gabriel Cristbal, and Jan Flusser, "A unified approach to superresolution and multichannel blind deconvolution.," *IEEE Trans Image Process*, vol. 16, no. 9, pp. 2322–2332, Sep 2007.

[17] Filip Sroubek and Peyman Milanfar, "Robust multichannel blind deconvolution via fast alternating minimization.," *IEEE Trans Image Process*, vol. 21, no. 4, pp. 1687–1700, Apr 2012.

[18] Luxin Yan, Houzhang Fang, and Sheng Zhong, "Blind image deconvolution with spatially adaptive total variation regularization.," *Opt Lett*, vol. 37, no. 14, pp. 2778–2780, Jul 2012.

[19] Xiang Zhu and Peyman Milanfar, "Removing atmospheric turbulence via space-invariant deconvolution.," *IEEE Trans Pattern Anal Mach Intell*, vol. 35, no. 1, pp. 157–170, Jan 2013.

[20] J. Idier and Y. Goussard, "Markov modeling for bayesian multi-channel deconvolution," *Proceedings of IEEE ICASSP*, p. 2, 1990.

[21] Sevket Babacan, Jingnan Wang, Rafael Molina, and Aggelos Katsaggelos, "Bayesian blind deconvolution from differently exposed image pairs.," *IEEE Trans Image Process*, vol. 19, no. 11, Nov 2010.

[22] 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.

[23] S. Derin Babacan, Rafael Molina, and Aggelos K Katsaggelos, "Variational bayesian blind deconvolution using a total variation prior.," *IEEE Trans Image Process*, vol. 18, no. 1, pp. 12–26, Jan 2009.

[24] Michael M Bronstein, Alexander M Bronstein, Michael Zibulevsky, and Yehoshua Y Zeevi, "Blind deconvolution of images using optimal sparse representations.," *IEEE Trans Image Process*, vol. 14, no. 6, pp. 726–736, Jun 2005.

[25] Rafael Molina, Javier Mateos, and Aggelos K Katsaggelos, "Blind deconvolution using a variational approach to parameter, image, and blur estimation.," *IEEE Trans Image Process*, vol. 15, no. 12, pp. 3715–3727, Dec 2006.

[26] Se Un Park, Nicolas Dobigeon, and Alfred O Hero, "Semi-blind sparse image reconstruction with application to mrfm.," *IEEE Trans Image Process*, vol. 21, no. 9, pp. 3838–3849, Sep 2012.

[27] Zhimin Xu and Edmund Y Lam, "Maximum a posteriori blind image deconvolution with huber-markov random-field regularization.," *Opt Lett*, vol. 34, no. 9, pp. 1453–1455, May 2009.

[28] L. Blanco and L. M. Mugnier, "Marginal blind deconvolution of adaptive optics retinal images.," *Opt Express*, vol. 19, no. 23, pp. 23227–23239, Nov 2011.