Bayesian Inference Tools for Inverse Problems

Ali Mohammad-Djafari

Laboratoire des Signaux et Systèmes,
UMR 8506 CNRS-SUPELEC-UNIV PARIS SUD
SUPELEC, Plateau de Moulon, 3 rue Juliot-Curie, 91192 Gif-sur-Yvette, France

Abstract. In this paper, first the basics of Bayesian inference with a parametric model of the data is presented. Then, the needed extensions are given when dealing with inverse problems and in particular the linear models such as Deconvolution or image reconstruction in Computed Tomography (CT). The main point to discuss then is the prior modeling of signals and images. A classification of these priors is presented, first in *separable* and *Markovien* models and then in *simple* or *hierarchical with hidden variables*. For practical applications, we need also to consider the estimation of the hyper parameters. Finally, we see that we have to infer simultaneously on the unknowns, the hidden variables and the hyper parameters.

Very often, the expression of this joint posterior law is too complex to be handled directly. Indeed, rarely we can obtain analytical solutions to any point estimators such the Maximum A posteriori (MAP) or Posterior Mean (PM). Three main tools are then can be used: Laplace approximation (LAP), Markov Chain Monte Carlo (MCMC) and Bayesian Variational Approximations (BVA).

To illustrate all these aspects, we will consider a deconvolution problem where we know that the input signal is *sparse* and propose to use a Student-t prior for that. Then, to handle the Bayesian computations with this model, we use the property of Student-t which is modelling it via an infinite mixture of Gaussians, introducing thus hidden variables which are the variances. Then, the expression of the joint posterior of the input signal samples, the hidden variables (which are here the inverse variances of those samples) and the hyper-parameters of the problem (for example the variance of the noise) is given. From this point, we will present the joint maximization by alternate optimization and the three possible approximation methods. Finally, the proposed methodology is applied in different applications such as mass spectrometry, spectrum estimation of quasi periodic biological signals and X ray computed tomography.

INTRODUCTION

In many generic inverse problems in signal and image processing, the problem is to infer on an unknown signal $f(t)$ or an unknown image $f(r)$ with $r = (x, y)$ through an observed signal $g(t')$ or an observed image $g(r')$ related between them through an operator $\mathcal{H}$ such as convolution $g = h * f$ or any other linear or non linear transformation $g = \mathcal{H}f$. When this relation is linear and we have discretized the problem, we arrive to the relation: $g = Hf + \varepsilon$, where $f = [f_1, \cdots, f_n]'$ represents the unknowns, $g = [g_1, \cdots, g_m]'$ the observed data, $\varepsilon = [\varepsilon_1, \cdots, \varepsilon_m]'$ the errors of modelling and measurement and H the matrix of the system response.
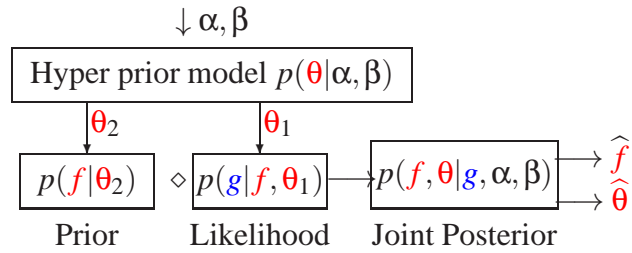
The Bayesian inference approach is based on the posterior law:

$$p(f|g, \theta_1, \theta_2) = \frac{p(g|f, \theta_1)\, p(f|\theta_2)}{p(g|\theta_1, \theta_2)} \propto p(g|f, \theta_1)\, p(f|\theta_2) \tag{1}$$
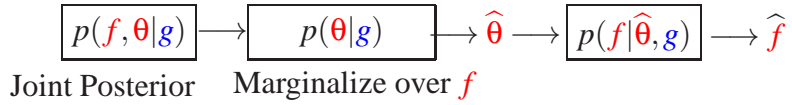
where the sign $\propto$ stands for "proportional to", $p(g|f,\theta_1)$ is the likelihood, $p(f|\theta_2)$ the prior model, $\theta = (\theta_1,\theta_2)$ are their corresponding parameters (often called the hyper-parameters of the problem) and $p(g|\theta_1,\theta_2)$ is called the evidence of the model. When the parameters θ have to be estimated too, a prior $p(\theta|\theta_0)$ with fixed values for θ_0 is assigned to them and the expression of the joint posterior

$$p(f,\theta|g,\theta_0) = \frac{p(g|f,\theta_1)\,p(f|\theta_2)\,p(\theta|\theta_0)}{p(g|\theta_0)} \tag{2}$$

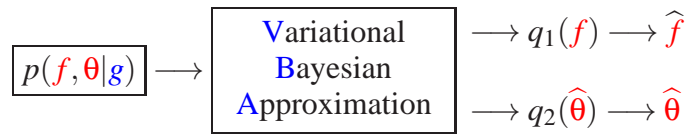is used to infer them jointly. This approach is showed in the following schemes:

$$\downarrow \alpha,\beta$$

$$\boxed{\text{Hyper prior model } p(\theta|\alpha,\beta)}$$

$$\theta_2 \qquad\qquad \theta_1$$

$$\boxed{p(f|\theta_2)} \;\diamond\; \boxed{p(g|f,\theta_1)} \longrightarrow \boxed{p(f,\theta|g,\alpha,\beta)} \begin{array}{l}\longrightarrow \widehat{f} \\ \longrightarrow \widehat{\theta}\end{array}$$

$$\text{Prior} \qquad \text{Likelihood} \qquad \text{Joint Posterior}$$

Full Bayesian Model and Hyperparameter Estimation scheme

$$\boxed{p(f,\theta|g)} \longrightarrow \boxed{p(\theta|g)} \longrightarrow \widehat{\theta} \longrightarrow \boxed{p(f|\widehat{\theta},g)} \longrightarrow \widehat{f}$$

$$\text{Joint Posterior} \qquad \text{Marginalize over } f$$

Marginalization for Hyper-parameter Estimation

Variational Bayesian Approximation (BVA) methods try to approximate $p(f,\theta|g)$ by a separable one $q(f,\theta|g) = q_1(f|\widetilde{\theta},g)\,q_2(\theta|\widetilde{f},g)$ and then using them for estimation [3, 1, 8, 11, 2, 9, 10, 7, 5]. This approach is showed in the following scheme:

$$\boxed{p(f,\theta|g)} \longrightarrow \boxed{\begin{array}{c}\text{Variational}\\ \text{Bayesian}\\ \text{Approximation}\end{array}} \begin{array}{l}\longrightarrow q_1(f) \longrightarrow \widehat{f} \\ \longrightarrow q_2(\widehat{\theta}) \longrightarrow \widehat{\theta}\end{array}$$

For hierarchical prior models with hidden variables z, the problem becomes more complex, because we have to give the expression of the joint posterior law

$$p(f,z,\theta|g) \propto p(g|f,\theta_1)\,p(f|z,\theta_2)\,p(z|\theta_3)\,p(\theta|\theta_0) \tag{3}$$

and then approximate it by a separable one

$$q(f,z,\theta|g) = q_1(f|\widetilde{z},\widetilde{\theta},g)\,q_2(z|\widetilde{f},\widetilde{\theta},g)\,q_3(\theta|\widetilde{z},\widetilde{f},g) \tag{4}$$

and then using them for estimation.

In this paper, first the general VBA method is detailed for the inference on inverse problems with hierarchical prior models. Then, two particular classes of prior models (Student-t and mixture of Gaussians) are considered and the details of BVA algorithms for them are given.

BAYESIAN VARIATIONAL APPROXIMATION WITH HIERARCHICAL PRIOR MODELS

When a hierarchical prior model $p(f|z,\theta)$ is used and when the estimation of the hyperparameters θ has to be considered, the joint posterior law of all the unknowns becomes:

$$p(f,z,\theta|g) \propto p(g|f,\theta_1)\,p(f|z,\theta_2)\,p(z|\theta_3)\,p(\theta) \tag{5}$$

which can also be written as $p(f,z,\theta|g) = p(f|z,\theta;g)\,p(z|\theta;g)\,p(\theta|g)$ where

$$p(f|z,\theta;g) = p(g|f,\theta)\,p(f|z,\theta)/p(g|z,\theta) \text{ with } p(g|z,\theta) = \int p(g|f,\theta)\,p(f|z,\theta)\,\mathrm{d}f \tag{6}$$

and $p(z|\theta;g) = p(g|z,\theta)\,p(z|\theta)/p(g|\theta)$ with $p(g|\theta) = \int p(g|z,\theta)\,p(z|\theta)\,\mathrm{d}z$ and finally

$$p(\theta|g) = p(g|\theta)\,p(\theta)/p(g) \text{ with } p(g) = \int p(g|\theta)\,p(\theta)\,\mathrm{d}\theta \tag{7}$$

We see that the first term

$$p(f|z,\theta,g) \propto p(g|f,\theta)p(f|z,\theta) \tag{8}$$

will be easy to handle because it is the product of two gaussians and so it is a multivariate Gaussian. But the two others are not.

The main idea behind the VBA is to approximate the joint posterior $p(f,z,\theta|g)$ by a separable one, for example

$$q(f,z,\theta|g) = q_1(f|g)\,q_2(z|g)\,q_3(\theta|g) \tag{9}$$

and where the expressions of $q(f,z,\theta|g)$ is obtained by minimizing the Kullback-Leibler divergence

$$\mathrm{KL}(q:p) = \int q\ln\frac{q}{p} = \left\langle \ln\frac{q}{p} \right\rangle_q \tag{10}$$

It is then easy to show that $\mathrm{KL}(q:p) = \ln p(g|\mathcal{M}) - \mathcal{F}(q)$ where $p(g|\mathcal{M})$ is the likelihood of the model

$$p(g|\mathcal{M}) = \int\int\int p(f,z,\theta,g|\mathcal{M})\,\mathrm{d}f\,\mathrm{d}z\,\mathrm{d}\theta \tag{11}$$

with $p(f,z,\theta,g|\mathcal{M}) = p(g|f,\theta)\,p(f|z,\theta)\,p(z|\theta)\,p(\theta)$ and $\mathcal{F}(q)$ is the free energy associated to q defined as

$$\mathcal{F}(q) = \left\langle \ln\frac{p(f,z,\theta,g|\mathcal{M})}{q(f,z,\theta)} \right\rangle_q \tag{12}$$

So, for a given model $\mathcal{M}$, minimizing $\mathrm{KL}(q:p)$ is equivalent to maximizing $\mathcal{F}(q)$ and when optimized, $\mathcal{F}(q^*)$ gives a lower bound for $\ln p(g|\mathcal{M})$.

Without any other constraint than the normalization of q, an alternate optimization of $\mathcal{F}(q)$ with respect to q_1, q_2 and q_3 results in

$$
\begin{cases}
q_1(f) \propto \exp\left\{-\langle \ln p(f,z,\theta,g)\rangle_{q(z)q(\theta)}\right\}, \\
q_2(z) \propto \exp\left\{-\langle \ln p(f,z,\theta,g)\rangle_{q(f)q(\theta)}\right\} \\
q_3(\theta) \propto \exp\left\{-\langle \ln p(f,z,\theta,g)\rangle_{q(f)q(z)}\right\}
\end{cases}
\tag{13}
$$

Note that these relations represent an implicit solution for $q_1(f)$, $q_2(z)$ and $q_3(\theta)$ which need, at each iteration, the expression of the expectations in the right hand of exponentials. If $p(g|f,z,\theta_1)$ is a member of an exponential family and if all the priors $p(f|z,\theta_2)$, $p(z|\theta_3)$, $p(\theta_1)$, $p(\theta_2)$, and $p(\theta_3)$ are conjugate priors, then it is easy to see that these expressions leads to standard distributions for which the required expectations are easily evaluated. In that case, we may note

$$
q(f,z,\theta|g) = q_1(f|\widetilde{z},\widetilde{\theta};g)\, q_2(z|\widetilde{f},\widetilde{\theta};g)\, q_3(\theta|\widetilde{f},\widetilde{z};g)
\tag{14}
$$

where the tilded quantities $\widetilde{z}$, $\widetilde{f}$ and $\widetilde{\theta}$ are, respectively functions of $(\widetilde{f},\widetilde{\theta})$, $(\widetilde{z},\widetilde{\theta})$ and $(\widetilde{f},\widetilde{z})$ and where the alternate optimization results to alternate updating of the parameters $(\widetilde{z},\widetilde{\theta})$ for q_1, the parameters $(\widetilde{f},\widetilde{\theta})$ of q_2 and the parameters $(\widetilde{f},\widetilde{z})$ of q_3.

Finally, we may note that, to monitor the convergence of the algorithm, we may evaluate the free energy

$$
\begin{aligned}
\mathcal{F}(q) &= \langle \ln p(f,z,\theta,g|\mathcal{M})\rangle_q + \langle -\ln q(f,z,\theta)\rangle_q \\
&= \langle \ln p(g|f,z,\theta)\rangle_q + \langle \ln p(f|z,\theta)\rangle_q + \langle \ln p(z|\theta)\rangle_q \\
&\quad + \langle -\ln q(f)\rangle_q + \langle -\ln q(z)\rangle_q + \langle -\ln q(\theta)\rangle_q
\end{aligned}
\tag{15}
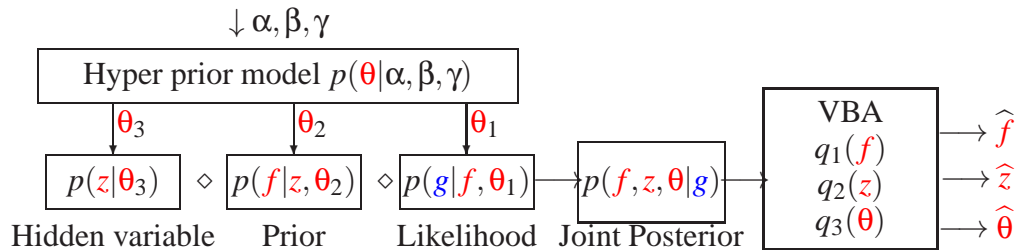$$

where all the expectations are with respect to q.

Other decompositions are also possible:

$$
q(f,z,\theta|g) = \prod_j q_{1j}(f_j|\widetilde{f}_{(-j)},\widetilde{z},\widetilde{\theta};g) \prod_j q_{2j}(z_j|\widetilde{f},\widetilde{z}(-j),\widetilde{\theta};g) \prod_l q_{3l}(\theta_l|\widetilde{f},\widetilde{z},\widetilde{\theta}_{(-l)};g)
\tag{16}
$$

or

$$
q(f,z,\theta|g) = q_1(f|\widetilde{z},\widetilde{\theta};g) \prod_j q_{2j}(z_j|\widetilde{f},\widetilde{z}_{(-j)},\widetilde{\theta};g) \prod_l q_{3l}(\theta_l|\widetilde{f},\widetilde{z},\widetilde{\theta}_{(-l)};g)
\tag{17}
$$

This approach is showed in the following scheme:

Full Bayesian Hierarchical Model and Variational Approximation

In the following section, we consider this case and give some more details with the Hierarchical model of Infinite Mixture model of Student-t.

JMAP AND BAYESIAN VARIATIONAL APPROXIMATION WITH STUDENT-T PRIORS

The Student-t model is:

$$p(f|\nu) = \prod_j \mathcal{S}t(f_j|\nu) \text{ with } \mathcal{S}t(f_j|\nu) = \frac{1}{\sqrt{\pi\nu}} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \left(1 + f_j^2/\nu\right)^{-(\nu+1)/2} \quad (18)$$

Knowing that

$$\mathcal{S}t(f_j|\nu) = \int_0^\infty \mathcal{N}(f_j|0, 1/z_j)\, \mathcal{G}(z_j|\nu/2, \nu/2)\, \mathrm{d}z_j \quad (19)$$

we can write this model via the positive hidden variables z_j:

$$\begin{cases} p(f|z) & = \prod_j p(f_j|z_j) = \prod_j \mathcal{N}(f_j|0, 1/z_j) \propto \exp\left\{-\frac{1}{2}\sum_j z_j f_j^2\right\} \\ p(z_j|\alpha, \beta) & = \mathcal{G}(z_j|\alpha, \beta) \propto z_j^{(\alpha-1)} \exp\left\{-\beta z_j\right\} \text{ with } \alpha = \beta = \nu/2 \end{cases} \quad (20)$$

Cauchy model is obtained when $\nu = 1$:

In this case, let consider the forward model $g = Hf + \varepsilon$ and assign a Gaussian law to the noise ε which which results to $p(g|f, v_\varepsilon) = \mathcal{N}(g|Hf, v_\varepsilon I)$. We also assign a prior $p(\tau_\varepsilon|\alpha_0, \beta_0) = \mathcal{G}(\tau_\varepsilon|\alpha_0, \beta_0)$ to $\tau_\varepsilon = 1/v_\varepsilon$.

Let also note $Z = \mathrm{diag}[z]$, and note $p(f|z) = \prod_j p(f_j|z_j) = \prod_j \mathcal{N}(f_j|0, z_j) = \mathcal{N}(f|0, z)$ and finally, assign $p(z|\alpha_0, \beta_0) = \prod_j \mathcal{G}(z_j|\alpha_0, \beta_0)$.

The following scheme shows the graphical representation of this model.
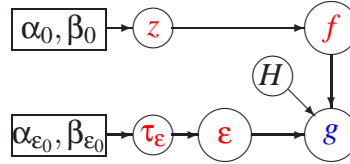
In the following, we summarize all the equations related to this modeling and inference scheme.

• Forward probability laws:

$$\begin{cases} p(g|f, \tau_\varepsilon) = \mathcal{N}(g|Hf, (1/\tau_\varepsilon)I), & p(\tau_\varepsilon|\alpha_{\varepsilon 0}, \beta_{\varepsilon 0}) = \mathcal{G}(\tau_\varepsilon|\alpha_{\varepsilon 0}, \beta_{\varepsilon 0}) \\ p(f|z) = \prod_j \mathcal{N}(f_j|0, 1/z_j), & p(z|\alpha_0, \beta_0) = \prod_j \mathcal{G}(z_j|\alpha_0, \beta_0) \end{cases} \quad (21)$$

• Joint posterior laws:

$$\begin{aligned} p(f, z, \tau_\varepsilon|g, \alpha_0, \beta_0, \alpha_{\varepsilon 0}, \beta_{\varepsilon 0}) & \propto p(g|f, \tau_\varepsilon)\, p(f|z)\, p(z|\alpha_0, \beta_0)\, p(\tau_\varepsilon|\alpha_{\varepsilon 0}, \beta_{\varepsilon 0}) \\ & \propto \tau_\varepsilon^{-M/2} \exp\left\{-\frac{1}{2}\tau_\varepsilon\|g - Hf\|^2\right\} \prod_j z_j^{-1/2} \exp\left\{-\frac{1}{2}z_j f_j^2\right\} \\ & \prod_j z_j^{-\alpha_0+1} \exp\left\{-\beta_0 z_j\right\} \tau_\varepsilon^{-\alpha_{\varepsilon 0}+1} \exp\left\{-\beta_{\varepsilon 0}\tau_\varepsilon\right\} \end{aligned} \quad (22)$$

- Joint MAP alternate maximization algorithm: The objective of the JMAP optimization is:

$$(\widehat{f},\widehat{z},\widehat{\tau_\varepsilon}) = \arg\max_{(f,z,\tau_\varepsilon)} \{p(f,z,\tau_\varepsilon|g,\alpha_0,\beta_0,\alpha_{\varepsilon 0},\beta_{\varepsilon 0})\} \tag{23}$$

The alternate optimization is an iterative optimization, respectively with respect to f, z and τ:

$$\begin{cases} \widehat{f} = \arg\min_f \left\{ \widehat{\tau_\varepsilon}\|g - Hf\|^2 + \sum_j \widehat{z_j} f_j^2 \right\} \\ \widehat{z} = \arg\min_z \left\{ \frac{N+2\alpha_0-2}{2}\ln z_j + \sum_j z_j\left(\frac{1}{2}\widehat{f}_j^2 + \beta_0\right) \right\} \\ \widehat{\tau_\varepsilon} = \arg\min_{\tau_\varepsilon} \left\{ \left(\frac{M}{2}+\alpha_{\varepsilon 0}-1\right)\ln\tau_\varepsilon + \left(\frac{1}{2}\|g-H\widehat{f}\|^2 + \beta_{\varepsilon 0}\right) \right\} \end{cases} \tag{24}$$

The first optimization can be done either analytically or using any gradient based algorithm. The second and the third optimizations have analytical expressions:

$$\begin{cases} \widehat{f} = \widehat{\tau_\varepsilon}\widehat{\Sigma}H'g \text{ with } \widehat{\Sigma} = \left(\widehat{\tau_\varepsilon}H'H + \widehat{Z}\right)^{-1} \text{ where } \widehat{Z} = \text{diag}\left[\widehat{z}^{-1}\right] \\ \widehat{z_j} = \left(\frac{1}{2}\widehat{f}_j^2 + \beta_0\right)/(\frac{M}{2}+\alpha_{\varepsilon 0}-1) \\ \widehat{\tau_\varepsilon} = \left(\frac{1}{2}\|g-H\widehat{f}\|^2 + \beta_{\varepsilon 0}\right)/(\frac{M}{2}+\alpha_{\varepsilon 0}-1) \end{cases} \tag{25}$$

One iteration of this algorithm is shown in the following scheme:
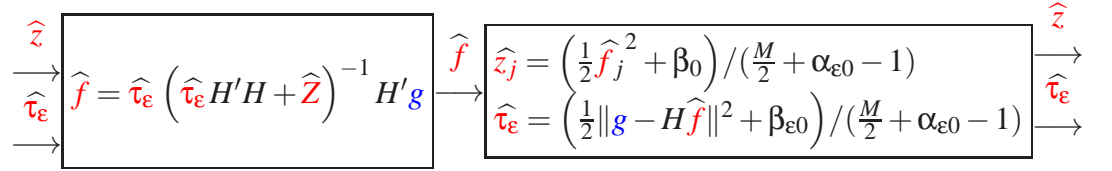
The main drawback of this method is that the uncertainties of the solution at each step is not accounted for for the next step.

- VBA posterior laws:

$$\begin{cases} q_1(f|\widetilde{\mu},\widetilde{\Sigma}) = \mathcal{N}(f|\widetilde{\mu},\widetilde{\Sigma}), \quad \widetilde{\mu} = \widetilde{\tau}\widetilde{\Sigma}H'g, \ \widetilde{\Sigma} = \left(\widetilde{\tau}H'H + \widetilde{Z}\right)^{-1} \text{ with } \widetilde{Z} = \text{diag}\left[\widetilde{z}^{-1}\right] \\ q_{2j}(z_j) = \mathcal{G}(z_j|\widetilde{\alpha}_j,\widetilde{\beta}_j); \quad \widetilde{\alpha}_j = \alpha_0 + \frac{1}{2}, \quad \widetilde{\beta}_j = \beta_0 + <f_j^2>/2 \\ q_3(\tau_\varepsilon) = \mathcal{G}(\tau_\varepsilon|\widetilde{\alpha}_\varepsilon,\widetilde{\beta}_\varepsilon) \\ \widetilde{\alpha}_\varepsilon = \alpha_{\varepsilon 0} + (n+1)/2, \quad \widetilde{\beta}_\varepsilon = \beta_{\varepsilon 0} + \frac{1}{2}[\|g\|^2 - 2\langle f\rangle' H'g + H'\langle ff\rangle H] \end{cases} \tag{26}$$

with

$$<f> = \widetilde{\mu}, \ <ff'> = \widetilde{\Sigma} + \widetilde{\mu}\widetilde{\mu}', \ <f_j^2> = [\widetilde{\Sigma}]_{jj} + \widetilde{\mu}_j^2, \ \widetilde{\tau} = \frac{\widetilde{\alpha}_{\tau_\varepsilon}}{\widetilde{\beta}_{\tau_\varepsilon}}, \ \widetilde{z}_j = \frac{\widetilde{\alpha}_j}{\widetilde{\beta}_j} \tag{27}$$

The expression of the free energie can be obtained as follows:

$$\mathcal{F}(q) = \left\langle \ln \frac{p(f,z,\tau,g|\mathcal{M})}{q(f,z,\tau)} \right\rangle =$$
$$\langle \ln p(g|f,z,\tau)\rangle + \langle \ln p(f|z,\tau)\rangle + \langle \ln p(z|\tau)\rangle + \langle -\ln q(f)\rangle + \langle -\ln q(z)\rangle + \langle -\ln q(\tau)\rangle \tag{28}$$

where

$$\langle \ln p(g|f,\tau_\varepsilon)\rangle = \frac{n}{2}(<\ln\tau_\varepsilon> - \ln(2\pi)) - \frac{1}{2}\{\langle\lambda\rangle g'g - 2 <f>' H'g + H' <ff'> H\}$$

$$\langle -\ln p(f|z)\rangle = -\frac{n+1}{2}\ln(2\pi) - \frac{1}{2}\left\{\sum_j <\ln\alpha_j><\alpha_j><f_j^2>\right\}$$
$$\langle -\ln p(z)\rangle = -(n+1)\alpha_{\varepsilon_0}\ln\beta_{\varepsilon_0} + (\alpha_{\varepsilon_0}-1)\sum_j <\ln\alpha_j> -\beta<\alpha_j> -(n+1)\ln\Gamma(\alpha)$$

$$\langle p(\tau_\varepsilon))\rangle = c\ln d + (c-1)<\ln\tau_\varepsilon> -d\langle\lambda\rangle - \ln\Gamma(c)$$
$$\langle -\ln q(f)\rangle = -\frac{n+1}{2}(1+\ln(2\pi)) - \frac{1}{2}\ln|\Sigma_f|$$

$$\langle -\ln q(z)\rangle = -\sum_j\left[\widetilde{\alpha}_j\ln(\widetilde{\beta}_j) + (\widetilde{\alpha}_j-1)<\ln\widetilde{\alpha}_j> -\widetilde{\beta}_j<\alpha_j> -\ln\Gamma(\widetilde{\alpha}_j)\right]$$

$$\langle q(\tau_\varepsilon)\rangle = \tilde{c}\ln\tilde{d} + (\tilde{c}-1)<\ln\tau> -\tilde{d}\langle\lambda\rangle - \ln\Gamma(\tilde{c})$$

In these equations,

$$(29) \quad \begin{cases} <\ln a_j> = \psi(\tilde{a}_j) - \ln\tilde{b}_j \\ <\ln\tau> = \psi(\tilde{c}) - \ln\tilde{d} \\ \psi(a) = \frac{\partial\ln\Gamma(a)}{\partial a} \end{cases}$$

The three steps of this algorithm is shown in the folowing scheme:

BAYESIAN VARIATIONAL APPROXIMATION WITH MIXTURE OF GAUSSIANS PRIORS

The mixture models are also very commonly used as prior models. In particular the Mixture of two Gaussians (MoG2) model:

$$p(f|\lambda,v_1,v_0) = \prod_j \left(\lambda\mathcal{N}(f_j|0,v_1) + (1-\lambda)\mathcal{N}(f_j|0,v_0)\right) \tag{30}$$

which can also be expressed through the binary valued hidden variables $z_j \in \{0, 1\}$

$$
\begin{cases}
p(f|z) = \prod_j p(f_j|z_j) & = \prod_j \mathcal{N}\left(f_j|0, v_{z_j}\right) \propto \exp\left\{ -\frac{1}{2} \sum_j \frac{f_j^2}{v_{z_j}} \right\} \\
P(z_j = 1) = \lambda, & P(z_j = 0) = 1 - \lambda
\end{cases} \tag{31}
$$

In general $v_1 >> v_0$ and λ measures the sparsity ($0 < \lambda << 1$) [6]. In this case also all the equations are very similarly can be obtained. Here, we do not have enough place to write them.

CONCLUSIONS

In this paper, a VBA method is proposed for doing Bayesian computations for inverse problems where a hierarchical prior is used. In particular, two prior models are considered: the Student-t and the mixture of Gaussian models. In both cases, these priors can be written via hidden variables which gives the model a hierarchical structure which is used to do the factorization. For some applications see for example [12, 4] and two other related papers in this volume.

REFERENCES

1. M. Beal. *Variational Algorithms for Approximate Bayesian Inference*. PhD thesis, Gatsby Computational Neuroscience Unit, University College London, 2003.
2. Sotirios Chatzis and Theodora Varvarigou. Factor analysis latent subspace modeling and robust fuzzy clustering using t-distributionsclassification of binary random patterns. *IEEE Trans. on Fuzzy Systems*, 17:505–517, 2009.
3. R. A. Choudrey. *Variational Methods for Bayesian Independent Component Analysis*. PhD thesis, University of Oxford, 2002.
4. N. Chu, J. Picheral, and A. Mohammad-Djafari. A robust super-resolution approach with sparsity constraint for near-field wideband acoustic imaging. In *IEEE International Symposium on Signal Processing and Information Technology*, pages 286–289, Bilbao, Spain, Dec.14-17,2011.
5. Aurélia Fraysse and Thomas Rodet. A gradient-like variational Bayesian algorithm. In *SSP 2011*, number S17.5, pages 605–608, Nice, France, jun 2011.
6. J. Rao. H. Ishwaran. Spike and Slab variable selection: Frequentist and Bayesian strategies. *Annals of Statistics*, 2005.
7. L. He, H. Chen, and L. Carin. Tree-Structured Compressive Sensing With Variational Bayesian Analysis. *IEEE Signal. Proc. Let.*, 17(3):233–236, 2010.
8. A. C. Likas and N. P. Galatsanos. A variational approach for bayesian blind image deconvolution. *IEEE Transactions on Signal Processing*, 2004.
9. T. Park and G. Casella. The Bayesian Lasso. *Journal of the American Statistical Association*, 2008.
10. M. Tipping. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 2001.
11. John Winn, Christopher M. Bishop, and Tommi Jaakkola. Variational message passing. *Journal of Machine Learning Research*, 6:661–694, 2005.
12. Sha Zhu, Ali Mohammad-Djafari, Hongqiang Wang, Bin Deng, Xiang Li, and Junjie Mao. Parameter estimation for sar micromotion target based on sparse signal representation. *Eurasip Journal of Signal Processing*, special issue "sparse approximations in signal and image processing", 2012.