Variational Bayesian Approximation methods for inverse problems

Ali Mohammad-Djafari

Laboratoire des signaux et systèmes (L2S),
UMR 8506 CNRS-SUPELEC-UNIV PARIS SUD,
SUPELEC, Plateau de Moulon, 91192 Gif-sur-Yvette, France

E-mail: `djafari@lss.supelec.fr`

Abstract. Variational Bayesian Approximation (VBA) methods are recent tools for effective Bayesian computations. In this paper, these tools are used for inverse problems where the prior models include hidden variables and where where the estimation of the hyper parameters has also to be addressed. In particular two specific prior models (Student-t and mixture of Gaussian models) are considered and details of the algorithms are given.

1. 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 $f(\boldsymbol{r})$ with $\boldsymbol{r}=(x,y)$ through an observed signal $g(t')$ or an observed image $g(\boldsymbol{r}')$ 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: $\boldsymbol{g}=\boldsymbol{H}\boldsymbol{f}+\boldsymbol{\epsilon}$, where $\boldsymbol{f}=[f_1,\cdots,f_n]'$ represents the unknowns, $\boldsymbol{g}=[g_1,\cdots,g_m]'$ the observed data, $\boldsymbol{\epsilon}=[\epsilon_1,\cdots,\epsilon_m]'$ the errors of modelling and measurement and $\boldsymbol{H}$ the matrix of the system response.

The Bayesian inference approach is based on the posterior law:

$$p(\boldsymbol{f}|\boldsymbol{g},\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)=\frac{p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)\,p(\boldsymbol{f}|\boldsymbol{\theta}_2)}{p(\boldsymbol{g}|\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}\propto p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)\,p(\boldsymbol{f}|\boldsymbol{\theta}_2) \tag{1}$$

where the sign $\propto$ stands for "proportional to", $p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)$ is the likelihood, $p(\boldsymbol{f}|\boldsymbol{\theta}_2)$ the prior model, $\boldsymbol{\theta}=(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ are their corresponding parameters (often called the hyper-parameters of the problem) and $p(\boldsymbol{g}|\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ is called the evidence of the model. When the parameters $\boldsymbol{\theta}$ have to be estimated too, a prior $p(\boldsymbol{\theta}|\boldsymbol{\theta}_0)$ with fixed values for $\boldsymbol{\theta}_0$ is assigned to them and the expression of the joint posterior

$$p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g},\boldsymbol{\theta}_0)=\frac{p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{\theta}_1)\,p(\boldsymbol{f}|\boldsymbol{\theta}_2)\,p(\boldsymbol{\theta}|\boldsymbol{\theta}_0)}{p(\boldsymbol{g}|\boldsymbol{\theta}_0)} \tag{2}$$

is used to infer them jointly.

Variational Bayesian Approximation (BVA) methods try to approximate $p(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g})$ by a separable one $q(\boldsymbol{f},\boldsymbol{\theta}|\boldsymbol{g})=q_1(\boldsymbol{f}|\widetilde{\boldsymbol{\theta}},\boldsymbol{g})\;q_2(\boldsymbol{\theta}|\widetilde{\boldsymbol{f}},\boldsymbol{g})$ and then using them for estimation [1, 2, 3, 4, 5, 6, 7, 8, 9].

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

$$p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)\, p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_2)\, p(\boldsymbol{z}|\boldsymbol{\theta}_3)\, p(\boldsymbol{\theta}|\boldsymbol{\theta}_0) \tag{3}$$

and then approximate it by a separable one

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

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.

2. Bayesian Variational Approximation with hierarchical prior models

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

$$p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)\, p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_2)\, p(\boldsymbol{z}|\boldsymbol{\theta}_3)\, p(\boldsymbol{\theta}) \tag{5}$$

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

$$q(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f}|\boldsymbol{g})\, q_2(\boldsymbol{z}|\boldsymbol{g})\, q_3(\boldsymbol{\theta}|\boldsymbol{g}) \tag{6}$$

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

$$\mathrm{KL}(q:p) = \int q \ln \frac{q}{p} = \left\langle \ln \frac{q}{p} \right\rangle_q \tag{7}$$

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

$$p(\boldsymbol{g}|\mathcal{M}) = \int\!\!\int\!\!\int p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g}|\mathcal{M})\; \mathrm{d}\boldsymbol{f}\; \mathrm{d}\boldsymbol{z}\; \mathrm{d}\boldsymbol{\theta} \tag{8}$$

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

$$\mathcal{F}(q) = \left\langle \ln \frac{p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g}|\mathcal{M})}{q(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta})} \right\rangle_q \tag{9}$$

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 $\ln p(\boldsymbol{g}|\mathcal{M})$.

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

$$\left\{\begin{array}{l} q_1(\boldsymbol{f}) \propto \exp\left\{-\left\langle \ln p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g})\right\rangle_{q(\boldsymbol{z})q(\boldsymbol{\theta})}\right\}, \\ q_2(\boldsymbol{z}) \propto \exp\left\{-\left\langle \ln p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g})\right\rangle_{q(\boldsymbol{f})q(\boldsymbol{\theta})}\right\} \\ q_3(\boldsymbol{\theta}) \propto \exp\left\{-\left\langle \ln p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g})\right\rangle_{q(\boldsymbol{f})q(\boldsymbol{z})}\right\} \end{array}\right. \tag{10}$$

Note that these relations represent an implicit solution for $q_1(\boldsymbol{f})$, $q_2(\boldsymbol{z})$ and $q_3(\boldsymbol{\theta})$ which need, at each iteration, the expression of the expectations in the right hand of exponentials. If $p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}_1)$ is a member of an exponential family and if all the priors $p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_2)$, $p(\boldsymbol{z}|\boldsymbol{\theta}_3)$, $p(\boldsymbol{\theta}_1)$, $p(\boldsymbol{\theta}_2)$, and $p(\boldsymbol{\theta}_3)$ 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

$$q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f}|\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{\theta}};\boldsymbol{g})\, q_2(\boldsymbol{z}|\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{\theta}};\boldsymbol{g})\, q_3(\boldsymbol{\theta}|\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{z}};\boldsymbol{g}) \tag{11}$$

where the tilded quantities $\widetilde{\boldsymbol{z}}$, $\widetilde{\boldsymbol{f}}$ and $\widetilde{\boldsymbol{\theta}}$ are, respectively functions of $(\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{\theta}})$, $(\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{\theta}})$ and $(\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{z}})$ and where the alternate optimization results to alternate updating of the parameters $(\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{\theta}})$ for q_1, the parameters $(\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{\theta}})$ of q_2 and the parameters $(\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{z}})$ of q_3. Finally, we may note that, to monitor the convergence of the algorithm, we may evaluate the free energy

$$\begin{aligned}\mathcal{F}(q) = \left\langle \ln p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{g}|\mathcal{M})\right\rangle_q + \left\langle -\ln q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta})\right\rangle_q &= \left\langle \ln p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta})\right\rangle_q + \left\langle \ln p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta})\right\rangle_q + \left\langle \ln p(\boldsymbol{z}|\boldsymbol{\theta})\right\rangle_q \\ &+ \left\langle -\ln q(\boldsymbol{f})\right\rangle_q + \left\langle -\ln q(\boldsymbol{z})\right\rangle_q + \left\langle -\ln q(\boldsymbol{\theta})\right\rangle_q\end{aligned} \tag{12}$$

where all the expectations are with respect to q. Other decompositions are also possible:

$$q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g}) = \prod_j q_{1j}(f_j|\widetilde{\boldsymbol{f}}_{(-j)},\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{\theta}};\boldsymbol{g})\; \prod_j q_{2j}(z_j|\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{z}}(-j),\widetilde{\boldsymbol{\theta}};\boldsymbol{g})\; \prod_l q_{3l}(\theta_l|\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{\theta}}_{(-l)};\boldsymbol{g}) \tag{13}$$

or

$$q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f}|\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{\theta}};\boldsymbol{g}) \prod_j q_{2j}(z_j|\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{z}}_{(-j)},\widetilde{\boldsymbol{\theta}};\boldsymbol{g}) \prod_l q_{3l}(\theta_l|\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{\theta}}_{(-l)};\boldsymbol{g}) \tag{14}$$

Here, we consider this case and give some more details on it.

3. Bayesian Variational Approximation with Student-t priors

The Student-t model is:

$$p(\boldsymbol{f}|\nu) = \prod_j \mathcal{S}t(f_j|\nu) \text{ with } \mathcal{S}t(f_j|\nu) = \frac{1}{\sqrt{\pi\nu}} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \left(1 + f_j^2/\nu\right)^{-(\nu+1)/2} \tag{15}$$

Knowing that

$$\mathcal{S}t(f_j|\nu) = \int_0^\infty \mathcal{N}(f_j|0,1/\tau_j)\, \mathcal{G}(\tau_j|\nu/2,\nu/2)\; \mathrm{d}\tau_j \tag{16}$$

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

$$\begin{array}{ll} p(\boldsymbol{f}|\boldsymbol{\tau}) & = \prod_j p(f_j|\tau_j) = \prod_j \mathcal{N}(f_j|0,1/\tau_j) \propto \exp\left\{-\frac{1}{2}\sum_j \tau_j f_j^2\right\} \\ p(\tau_j|\alpha,\beta) & = \mathcal{G}(\tau_j|\alpha,\beta) \propto \tau_j^{(\alpha-1)} \exp\left\{-\beta\tau_j\right\} \text{ with } \alpha=\beta=\nu/2 \end{array} \tag{17}$$

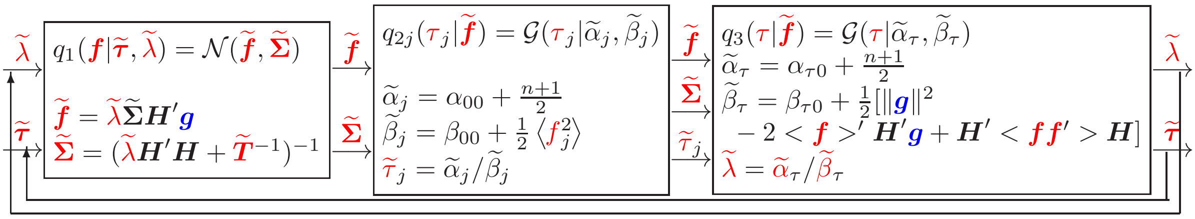
Cauchy model is obtained when $\nu = 1$:

In this case, let consider the forward model $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$ and assign a Gaussian law to the noise $\boldsymbol{\epsilon}$ which which results to $p(\boldsymbol{g}|\boldsymbol{f},v_\epsilon) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f},v_\epsilon\boldsymbol{I})$. We also assign a prior $p(\tau_\epsilon|\alpha_{\tau0},\beta_{\tau0}) = \mathcal{G}(\tau_\epsilon|\alpha_{\tau0},\beta_{\tau0})$ to $\tau_\epsilon = 1/v_\epsilon$. Let also note $\boldsymbol{\tau} = [\tau_1,\cdots,\tau_N]$, $\boldsymbol{T} = \operatorname{diag}[\boldsymbol{\tau}]$, $z_j = 1/\tau_j$, $\boldsymbol{Z} = \operatorname{diag}[\boldsymbol{z}] = \boldsymbol{T}^{-1}$ and note $p(\boldsymbol{f}|\boldsymbol{\tau}) = \prod_j p(f_j|\tau_j) = \prod_j \mathcal{N}(f_j|0,\tau_j) = \mathcal{N}(\boldsymbol{f}|0,\boldsymbol{T})$ and finally, assign $p(\boldsymbol{\tau}|\alpha_0,\beta_0) = \prod_j \mathcal{G}(\tau_j|\alpha_0,\beta_0)$.

Then, we obtain the following expressions for the VBA:

$$\left\{\begin{array}{l} p(\boldsymbol{g}|\boldsymbol{f},\tau_\epsilon) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f},(1/\tau_\epsilon)\boldsymbol{I}) \\ p(\tau_\epsilon|\alpha_{\tau0},\beta_{\tau0}) = \mathcal{G}(\tau_\epsilon|\alpha_{\tau0},\beta_{\tau0}) \\ p(\boldsymbol{f}|\boldsymbol{\tau}) = \prod_j \mathcal{N}(f_j|0,1/\tau_j) \\ p(\boldsymbol{\tau}|\alpha_0,\beta_0) = \prod_j \mathcal{G}(\tau_j|\alpha_0,\beta_0) \\ \\ q_1(\boldsymbol{f}|\widetilde{\boldsymbol{\mu}},\widetilde{\boldsymbol{\Sigma}}) = \mathcal{N}(\boldsymbol{f}|\widetilde{\boldsymbol{\mu}},\widetilde{\boldsymbol{\Sigma}}) \\ \widetilde{\boldsymbol{\mu}} = \langle\lambda\rangle\, \widetilde{\boldsymbol{\Sigma}}\boldsymbol{H}'\boldsymbol{g} \\ \widetilde{\boldsymbol{\Sigma}} = (\langle\lambda\rangle\, \boldsymbol{H}'\boldsymbol{H} + \widetilde{\boldsymbol{Z}})^{-1}, \\ \text{with } \; \widetilde{\boldsymbol{Z}} = \widetilde{\boldsymbol{T}}^{-1} = \operatorname{diag}[\widetilde{\boldsymbol{\tau}}] \end{array}\right. \quad \left\{\begin{array}{l} q_{2j}(\tau_j) = \mathcal{G}(\tau_j|\widetilde{\alpha}_j,\widetilde{\beta}_j) \\ \widetilde{\alpha}_j = \alpha_{00} + 1/2 \\ \widetilde{\beta}_j = \beta_{00} + < f_j^2 > /2 \\ \\ q_3(\tau_\epsilon) = \mathcal{G}(\tau_\epsilon|\widetilde{\alpha}_{\tau_\epsilon},\widetilde{\beta}_{\tau_\epsilon}), \\ \widetilde{\alpha}_{\tau_\epsilon} = \alpha_{\tau0} + (n+1)/2 \\ \widetilde{\beta}_{\tau_\epsilon} = \beta_{\tau0} + 1/2[\|\boldsymbol{g}\|^2 \\ -2\,\langle\boldsymbol{f}\rangle'\,\boldsymbol{H}'\boldsymbol{g} + \boldsymbol{H}'\,\langle\boldsymbol{f}\boldsymbol{f}'\rangle\,\boldsymbol{H}] \end{array}\right. \quad \left\{\begin{array}{l} < \boldsymbol{f} > = \widetilde{\boldsymbol{\mu}} \\ < \boldsymbol{f}\boldsymbol{f}' > = \widetilde{\boldsymbol{\Sigma}} + \widetilde{\boldsymbol{\mu}}\widetilde{\boldsymbol{\mu}}' \\ < f_j^2 > = [\widetilde{\boldsymbol{\Sigma}}]_{jj} + \widetilde{\mu}_j^2 \\ \widetilde{\lambda} = \widetilde{\alpha}_\tau/\widetilde{\beta}_\tau \\ \widetilde{\tau}_j = \widetilde{\alpha}_j/\widetilde{\beta}_j \end{array}\right. \tag{18}$$

We can also express the free energy expression $\mathcal{F}(q)$ which can be used as a stopping criterion for the algorithm. Its expression is given in the appendix. The resulting algorithm can be summarized as follows


$\widetilde{\lambda}$
$\widetilde{\boldsymbol{\tau}}$
$q_1(\boldsymbol{f}|\widetilde{\boldsymbol{\tau}},\widetilde{\lambda}) = \mathcal{N}(\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{\Sigma}})$
$\widetilde{\boldsymbol{f}} = \widetilde{\lambda}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{H}'\boldsymbol{g}$
$\widetilde{\boldsymbol{\Sigma}} = (\widetilde{\lambda}\boldsymbol{H}'\boldsymbol{H} + \widetilde{\boldsymbol{T}}^{-1})^{-1}$
$\widetilde{\boldsymbol{f}}$
$\widetilde{\boldsymbol{\Sigma}}$
$q_{2j}(\tau_j|\widetilde{\boldsymbol{f}}) = \mathcal{G}(\tau_j|\widetilde{\alpha}_j,\widetilde{\beta}_j)$
$\widetilde{\alpha}_j = \alpha_{00} + \frac{n+1}{2}$
$\widetilde{\beta}_j = \beta_{00} + \frac{1}{2}\left\langle f_j^2 \right\rangle$
$\widetilde{\tau}_j = \widetilde{\alpha}_j/\widetilde{\beta}_j$
$\widetilde{\boldsymbol{f}}$
$\widetilde{\boldsymbol{\Sigma}}$
$\widetilde{\tau}_j$
$q_3(\tau|\widetilde{\boldsymbol{f}}) = \mathcal{G}(\tau|\widetilde{\alpha}_\tau,\widetilde{\beta}_\tau)$
$\widetilde{\alpha}_\tau = \alpha_{\tau 0} + \frac{n+1}{2}$
$\widetilde{\beta}_\tau = \beta_{\tau 0} + \frac{1}{2}[\|\boldsymbol{g}\|^2$
$- 2 < \boldsymbol{f} >' \boldsymbol{H}'\boldsymbol{g} + \boldsymbol{H}' < \boldsymbol{f}\boldsymbol{f}' > \boldsymbol{H}]$
$\widetilde{\lambda} = \widetilde{\alpha}_\tau/\widetilde{\beta}_\tau$
$\widetilde{\lambda}$
$\widetilde{\boldsymbol{\tau}}$

4. 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:

$$p(\boldsymbol{f}|\lambda, v_1, v_0) = \prod_j \left(\lambda\mathcal{N}(f_j|0, v_1) + (1-\lambda)\mathcal{N}(f_j|0, v_0)\right) \tag{19}$$

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

$$\left\{ \begin{array}{ll} p(\boldsymbol{f}|\boldsymbol{z}) = \prod_j p(f_j|z_j) & = \prod_j \mathcal{N}\left(f_j|0, v_{z_j}\right) \propto \exp\left\{-\frac{1}{2}\sum_j \frac{f_j^2}{v_{z_j}}\right\} \\ P(z_j = 1) = \lambda, & P(z_j = 0) = 1 - \lambda \end{array} \right. \tag{20}$$

In general $v_1 >> v_0$ and λ measures the sparsity $(0 < \lambda << 1)$ [10]. In this case also all the equations are very similarly can be obtained. Here, we do not have enough place to write them. They will be provided in the final version as an appendix.

5. 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. All the details are given in the appendix. For some applications see for example [11, 12].

6. References

[1] Choudrey R A 2002 *Variational Methods for Bayesian Independent Component Analysis* Ph.D. thesis University of Oxford

[2] Beal M 2003 *Variational Algorithms for Approximate Bayesian Inference* Ph.D. thesis Gatsby Computational Neuroscience Unit, University College London

[3] Likas A C and Galatsanos N P 2004 *IEEE Transactions on Signal Processing*

[4] Winn J, Bishop C M and Jaakkola T 2005 *Journal of Machine Learning Research* **6** 661–694

[5] Chatzis S and Varvarigou T 2009 *IEEE Trans. on Fuzzy Systems* **17** 505–517

[6] Park T and Casella G 2008 *Journal of the American Statistical Association*

[7] Tipping M 2001 *Journal of Machine Learning Research*

[8] He L, Chen H and Carin L 2010 *IEEE Signal. Proc. Let.* **17** 233–236

[9] Fraysse A and Rodet T 2011 *SSP 2011* S17.5 (Nice, France) pp 605–608

[10] H Ishwaran J R 2005 *Annals of Statistics*

[11] Zhu S, Mohammad-Djafari A, Wang H, Deng B, Li X and Mao J 2012 *Eurasip Journal of Signal Processing* **special issue "sparse approximations in signal and image processing"**

[12] Chu N, Picheral J and Mohammad-Djafari A Bilbao, Spain, Dec14-17,2011 *IEEE International Symposium on Signal Processing and Information Technology* pp 286–289