A subspace-based variational Bayesian method

Yuling Zheng[1], Aurélia Fraysse[1], Thomas Rodet[1]

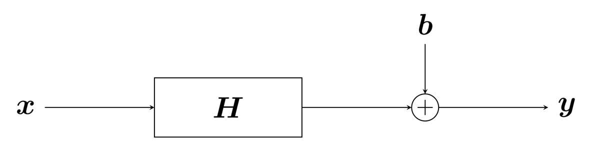
[1]**L2S**, 3 rue Joliot Curie, 91192 Gif Sur Yvette, France

The problem

- Ill-posed inverse problems
- Bayesian methodology :
  - Large dimensional.
  - Non Gaussian prior.
- Approximation of the posterior distribution by simpler laws.

I Sparse linear inverse problems

Model



$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{b}$ where $\boldsymbol{H} \in \mathcal{M}_{M\times N}$ is a given matrix and $\boldsymbol{b} \in \mathbb{R}^M$ is a Gaussian iid noise.

Problem :

- $\boldsymbol{H}$ or $\boldsymbol{H}^T\boldsymbol{H}$ not easily invertible
- ill posed problem.
- large dimension.

Prior information :

Given by a GSM (Gaussian Scale Mixture) model, see [1] :

$$\forall i = 1,\dots,N, \;\; p(x_i) = \int p(x_i/z_i)\phi(z_i)dz_i$$
$$= \frac{b^a}{\Gamma(a)} \int_{\mathbb{R}} \frac{\sqrt{z_i}}{\sqrt{2\pi}|\sigma_1^2|^{1/2}} e^{-\frac{z_i x_i^2}{2\sigma_1^2}} z_i^{a-1} e^{-bz_i} dz_i.$$

where $z_i \sim Gamma(a,b)$ is an hidden variable to be estimated.

Posterior distribution :

$$p(\mathbf{x},\mathbf{z}|\mathbf{y}) \propto \exp\left[-\frac{\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2}{2\sigma_b^2}\right] \prod_i \frac{\sqrt{z_i}}{\sigma_1} \exp\left[-\frac{z_i x_i^2}{2\sigma_1^2}\right] \frac{b^a z_i^{a-1} e^{-bz_i}}{\Gamma(a)}.$$

Difficulties :

– Link between $\boldsymbol{X}$ and $\boldsymbol{Z}$.
– When the dimension of $\boldsymbol{X}$ increases, the correlation matrix cannot be inversed efficiently.

II Variational Bayesian methodology

Introduced by [4]. Let us denote $\mathbf{W} = (\mathbf{X},\mathbf{Z})$ in the following.

- **Principle :** Approximation of the posterior by separable distributions $q = \prod_i q_i$
- Minimization of the Kullback-Leibler divergence.

The convex optimization problem given by variational Bayesian method is

$$q^{opt} = \underset{q \text{ p.d.f. separable}}{\arg\min} \mathcal{KL}[q(\mathbf{X},\mathbf{Z})||p(\mathbf{X},\mathbf{Z}|\mathbf{Y}))]. \quad (1)$$

- Minimizing the Kullback-Leibler divergence is equivalent to the maximization of the negative free energy.

The equivalent concave optimization problem :

$$q^{opt} = \underset{q \text{ p.d.f. separable}}{\arg\max} F(q), \quad (2)$$

where the negative free energy is

$$F(q) = \int \ln(p(\boldsymbol{y},\boldsymbol{w}))q(\boldsymbol{w})d\boldsymbol{w} - \int \ln(q(\boldsymbol{w}))q(\boldsymbol{w})d\boldsymbol{w} \quad (3)$$

The solution of Problem 2 is given by $q(\mathbf{W}) = \prod_i q_i(w_i)$ such that

$$q_i(w_i) = \frac{1}{K_i} \exp\left(\langle \log p(\mathbf{Y},\mathbf{W})\rangle_{\prod_{j\neq i} q_j(w_j)}\right). \quad (4)$$

– **Advantage :**
  – q is a probability density function.
  – Approach faster than MCMC.
– **Inconvenient :**
  – Needs a conjugate prior.
  – Implicit form.

- Aternate minimization algorithms

$$q^{(k+1)}(w_i) = \frac{1}{Z_i} \exp\left[\langle \log(p(\mathbf{Y},\mathbf{W}))\rangle_{\prod_{j\neq i} q^{(k)}(w_j)}\right]$$

Large dimensional problems. ⇒ Too important computation time!

Exponentiated Gradient like algorithm, [3] :

- Update of the density q^{k+1} :

$$h(\mathbf{W}) = \exp(\alpha\nabla F(q(\mathbf{W}))) = \left(\prod_i \frac{1}{K_i} \frac{\exp\left(\langle \log p(\mathbf{Y},\mathbf{W})\rangle_{\prod_{j\neq i} q_j^k(w_j)}\right)}{q_i^k}\right)^{\alpha}.$$

where ∇F stands for the Gateaux derivatives of F whereas $\alpha > 0$ is the algorithm step-size.

$$q^{k+1}(\mathbf{W}) = q^k(\mathbf{W}) \left(\prod_i \frac{1}{K_i} \frac{\exp\left(\langle \log p(\mathbf{Y},\mathbf{W})\rangle_{\prod_{j\neq i} q_j^k(w_j)}\right)}{q_i^k}\right)^{\alpha} \quad (5)$$

How to improve this method ?

III Subspace based Optimization

Subspace algorithms in a Hilbert vector space :

- Iterative algorithms given by

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{D}^k\mathbf{s}^k, \quad (6)$$

where $\mathbf{D}^k$ spans a I dimensional space, [2].

Different constructions of the subspace :
– Consider gradient directions.
– Consider previous directions.

Subspace algorithms for Variational Bayesian methodology :

Define :

$$q_i^r = \exp\left(\langle \log p(\mathbf{y},\mathbf{w})\rangle_{\prod_{j\neq i} q_j^k}\right).$$

Then

$$q^{k+1} = \exp(\mathbf{D}^k\mathbf{s}^k)q^k \quad (7)$$

With $\mathbf{D}^k$ spanning a 2 dimensional space
$\mathbf{D}_1^k = \mathrm{d}f(q^k)$ is the Gateaux differential of F at q^k
and $\mathbf{D}_2^k = \ln\left(\frac{q^k}{q^{k-1}}\right)$ is the previous direction.

$$q^{\mathbf{s}}(\mathbf{w}) = K^k \prod_i q_i^k \left(\frac{q_i^r}{q_i^k}\right)^{s_1} \left(\frac{q_i^k}{q_i^{k-1}}\right)^{s_2}. \quad (8)$$

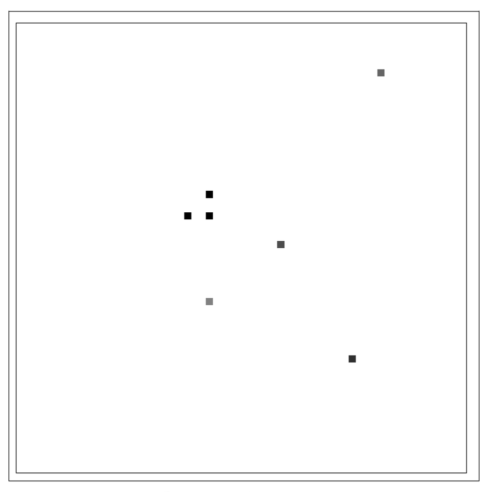
- s_k is the optimal stepsize in the two dimensional space.

IV Proposed algorithm

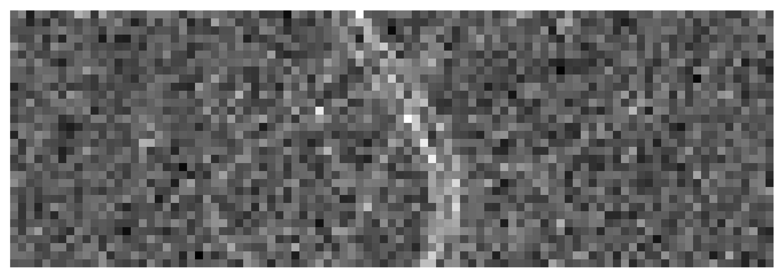
1. Initialize $(q_i^0)_{i=1,\dots,N}$ and $(\tilde{q}_j^0)_{j=1,\dots,N}$
2. Update the parameters of Gamma distributions $(\tilde{q}_j^{k+1})_{j=1,\dots,N}$ with classical VBA
3. Compute auxiliary function $h = \exp(\mathbf{D}^k\mathbf{s}^k)$
4. Determine the *subspace* : $\left[\frac{q^r}{q^k}, \frac{q^k}{q^{k-1}}\right]$
5. Compute the step-size
6. Update means and variances of $(q_i^{k+1})_{i=1,\dots,N}$
7. Go back to 2 until convergence

V Simulation

- Data : 64×64 sparse phantom, composed of 7 peaks with a magnitude between 0.5 and 1.
- Tomography parameter : Data in parallel beam geometry : 32 angles, uniformly spaced over [0, 180[, and 95 detector cells.
- White noise with a 0.3 standard deviation.
- hyper-parameters : $\sigma_b^2 = 1$, $\sigma_1^2 = 0.05$ and $a = b = \frac{\nu}{2}$ where $\nu = 0.1$.
- Comparison with FBP, MCMC, Classical Bayesian Variational method.

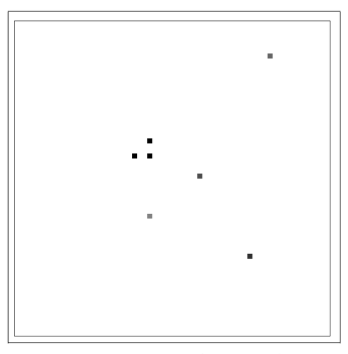


Original image

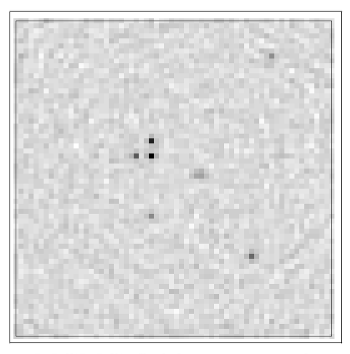


Data collected

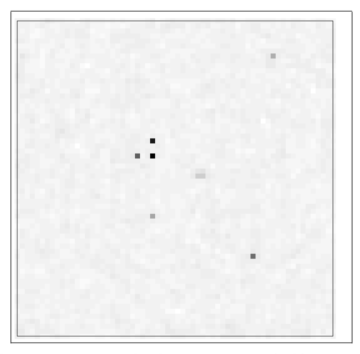
VI Results



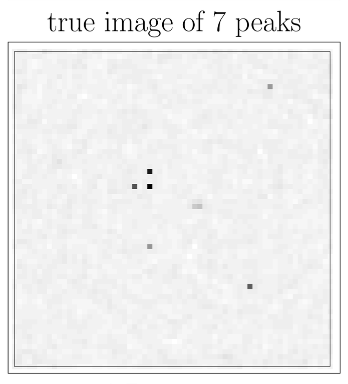
true image of 7 peaks



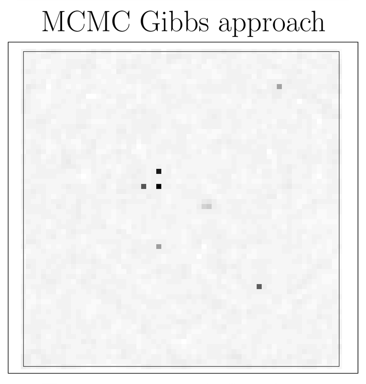
MCMC Gibbs approach



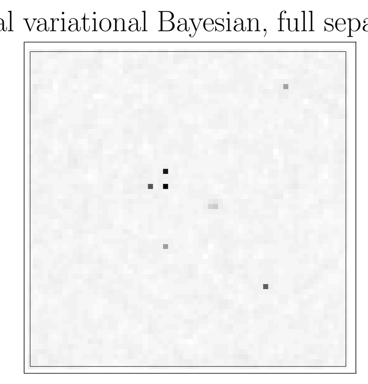
classical variational Bayesian, full separability



Classical variational Bayesian, partial separability



Exponentiated Gradient



Our method

Computation time (s) :

Method	MCMC	VBFS	VBPS	VBGrad	Proposed
PSNR(dB)	28.8	35.1	35.2	35.9	35.9
Time(s)	69313.8	723.5	327.6	23.4	3.2

Conclusions

- Definition of a different iterative algorithm based on the subspace based optimization in the context of variational Bayesian methods.
- Converges faster than the classical Bayesian methods and the exponentiated gradient like algorithm.
- Application on a small tomographic example. Better performances than previous methods.

Références

[1] G. Chantas, N. Galatsanos, A. Likas, and M. Saunders. Variational Bayesian image restoration based on a product of t-distributions image prior. *IEEE Transactions on Image Processing*, 17(10) :1795–1805, October 2008.

[2] E. Chouzenoux, J. Idier, and S. Moussaoui. A majorize-minimize strategy for subspace optimization applied to image restoration. *IEEE Transactions on Image Processing*, 20(18) :1517–1528, 2011.

[3] A. Fraysse and T. Rodet. A gradient-like variational bayesian algorithm. In *Statistical Signal Processing Workshop (SSP)*, pages 605–608. IEEE, 2011.

[4] D. J. C. MacKay. Ensemble learning and evidence maximization. http ://citeseerx.ist.psu.edu/viewdoc/summary ?doi=10.1.1.54.4083, 1995.