On the Accuracy of Variational Bayes in Latent Gaussian Spatial Models

Per Sidén, Division of Statistics and Machine Learning, Linköping University, Sweden.

Jointly with Anders Eklund, David Bolin and Mattias Villani. Mail: per.siden@liu.se

LINKÖPING UNIVERSITY

Summary

Bayesian inference in high-dimensional latent Gaussian spatial models is a serious computational challenge. Variational Bayes (VB) is a common and fast option, but suffers from the drawback of producing an approximate posterior that is usually assumed to factorize over the different spatial locations, which is clearly false in applications with spatial dependence. In this work we use modern techniques for efficient high-dimensional Gaussian sampling and develop a fast and practical MCMC algorithm. We compare the resulting posterior with a popular VB approximation in a model for functional magnetic resonance imaging (fMRI) data, with a 3D Gaussian Markov random field (GMRF) spatial smoothing prior on the brain activity, and show that VB can lead to spurious results. We also propose an improved VB method, that is both faster and more accurate. A preprint of this paper is available at `http://arxiv.org/abs/1606.00980`.

GLM with Spatial Prior

Following [1], we define the General Linear Model (GLM) as

$$\underset{T\times N}{\mathbf{Y}} = \underset{T\times K}{\mathbf{X}}\underset{K\times N}{\mathbf{W}} + \underset{T\times N}{\mathbf{E}},$$

where $\mathbf{Y}$ is time-series data with T time-points, measured in N spatial locations, each modeled using the same design matrix $\mathbf{X}$ with K regressors. $\mathbf{E}$ is Gaussian order P AR noise with location-specific precision $\boldsymbol{\lambda}$ and AR parameters $\mathbf{A}$. There is a sparse spatial GMRF prior on the regression coefficients according to

$$\mathbf{W}'_{k,\cdot}|\alpha_k \sim \mathcal{N}\left(0, \alpha_k^{-1}\mathbf{D}_w^{-1}\right),$$

where $\mathbf{D}_w$ is a sparse precision matrix and similar for $\mathbf{A}$. We use gamma priors for α_k and λ_n.

Factorized VB posteriors

We consider two simplifying VB independence assumptions about the posterior

$$(1)\quad q(\mathbf{W},\boldsymbol{\alpha},\boldsymbol{\lambda}|\mathbf{Y}) = q(\mathbf{W}|\mathbf{Y})\,q(\boldsymbol{\alpha}|\mathbf{Y})\,q(\boldsymbol{\lambda}|\mathbf{Y})$$

$$(2)\quad q(\mathbf{W}|\mathbf{Y}) = \prod_{n=1}^{N} q(\mathbf{W}_{\cdot,n}|\mathbf{Y}).$$

Using both assumption results in a spatially factorized posterior, Independent VB (IVB), but we also construct an algorithm that drops the second assumption, Spatial VB (SVB). In addition, we develop a fast MCMC algorithm that drops both assumptions.

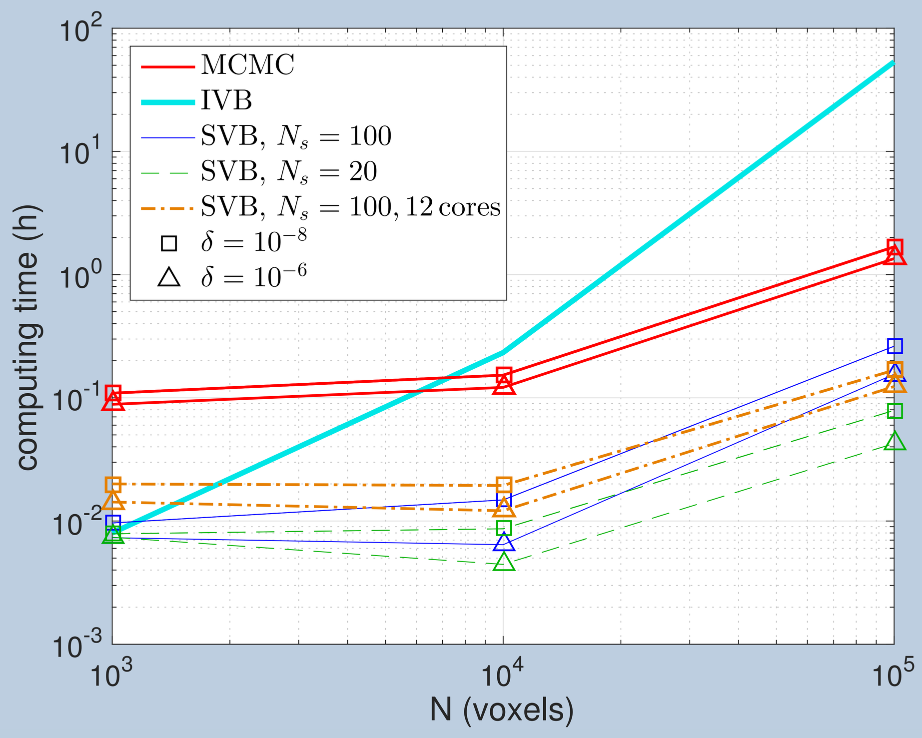
Inference Algorithms

A key to fast inference for both SVB and MCMC is efficient sampling from GMRFs using the preconditioned conjugate gradient (PCG) techniques in [2]. The full conditional posterior for the regression coefficients $\mathbf{w}_r = vec(\mathbf{W}')$ is a GMRF on the form

$$p(\mathbf{w}_r|\mathbf{Y},\boldsymbol{\alpha},\boldsymbol{\lambda}) \propto \exp\left(-\frac{1}{2}\mathbf{w}_r'\tilde{\mathbf{B}}\mathbf{w}_r + \mathbf{b}_w'\mathbf{w}_r\right)$$

$$\mathbf{b}_w = vec\left(diag(\boldsymbol{\lambda})\,\mathbf{Y}'\mathbf{X}\right)$$

$$\tilde{\mathbf{B}} = \mathbf{X}'\mathbf{X}\otimes diag(\boldsymbol{\lambda}) + diag(\boldsymbol{\alpha})\otimes\mathbf{D}_w$$



As the number of voxels increase, PCG sampling can give speed-ups greater than a factor 100, compared to the efficient, exact sparse Cholesky based sampling methods in [3].

We can sample from the full conditionals to perform Gibbs sampling and the SVB update steps are similarly the expectations of the log full conditionals with respect to the VB posteriors of the other parameters, which we approximate using Monte Carlo samples $\mathbf{W}^{(1:N_s)}$ e.g.

$$\log q(\boldsymbol{\alpha}) = E_{q(\mathbf{W}),q(\boldsymbol{\lambda})}\left[\log p(\alpha|\mathbf{Y},\cdot)\right] + C$$

$$\Rightarrow \alpha_k \sim Ga\left(\tilde{\vartheta}_{1k},\tilde{\vartheta}_2\right) \text{ with}$$

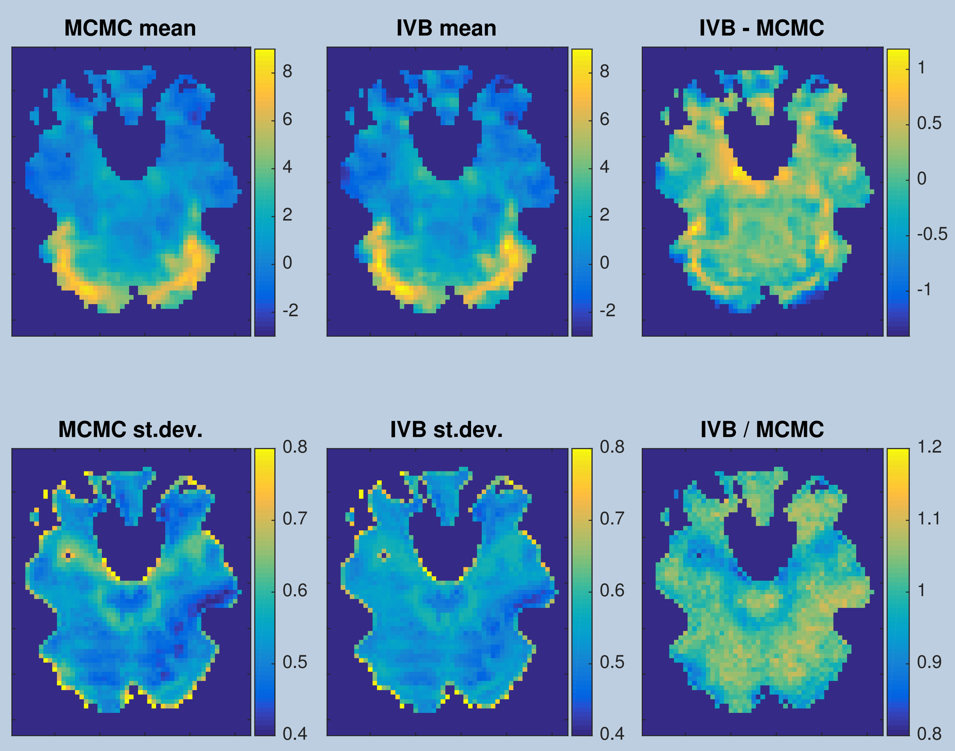
$$\frac{1}{\tilde{\vartheta}_{1k}} = \frac{1}{2}E_{\mathbf{W}}\left[\mathbf{W}_{k,\cdot}\mathbf{D}_w\mathbf{W}_{k,\cdot}'\right] + \frac{1}{\vartheta_1}$$

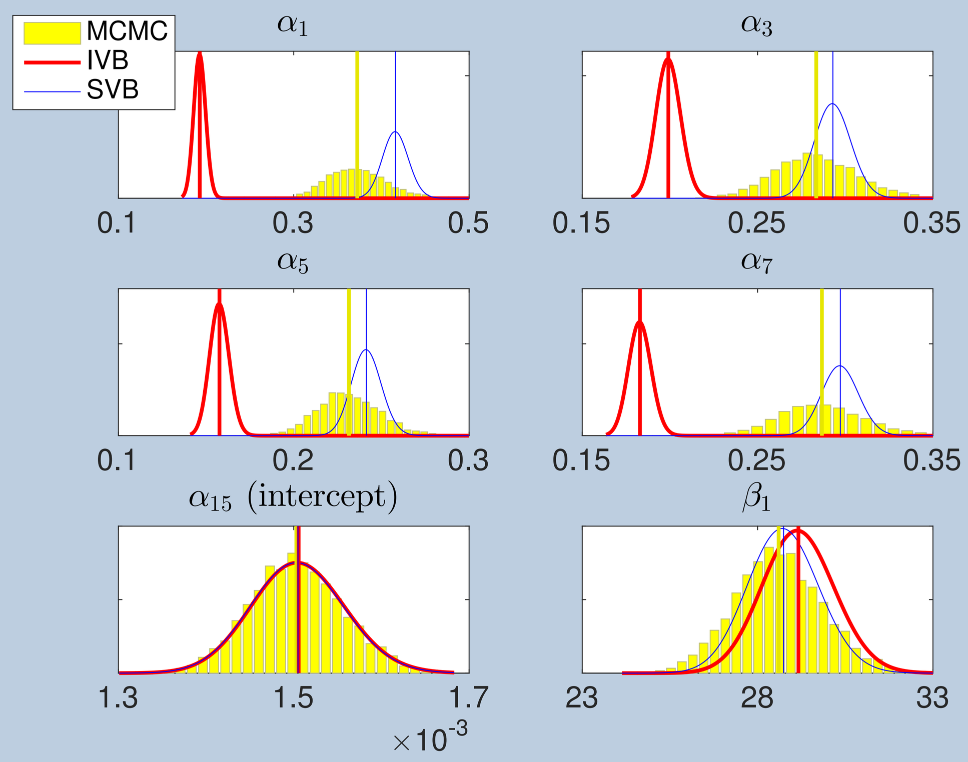
$$\approx \frac{1}{2}\frac{1}{N_s}\sum_{j=1}^{N_s}\mathbf{W}_{k,\cdot}^{(j)}\mathbf{D}_w\mathbf{W}_{k,\cdot}^{(j)\prime} + \frac{1}{\vartheta_1}$$

The resulting MCMC and SVB algorithms scale much better with number of locations than the IVB method implemented in the Statistical Parametric Mapping (SPM) software.

Results

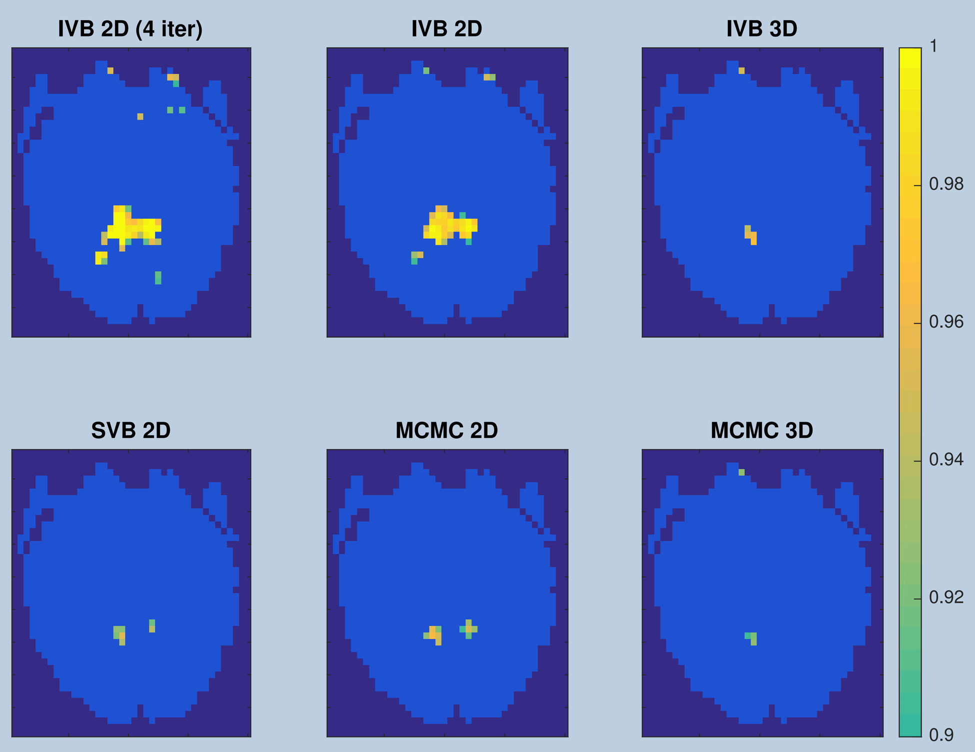
Evaluating the methods on the fMRI data in [1] with a 3D unweighted graph Laplacian prior on brain activity, we observe some differences between the IVB and the exact MCMC posterior mean and standard deviation:





It is well-known that factorized VBs in general tend to underestimate posterior standard deviations [4]. Here, however, much of the differences in the activity posterior can be explained by IVB underestimating also the posterior mean of the hyperparameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ that control the spatial smoothing of activity coefficients $\mathbf{W}$ and AR coefficients $\mathbf{A}$. The error is larger for less informative data.

A different fMRI data set gives evidence than IVB can lead to large errors also for 90% thresholded posterior probability maps of brain activity.



Conclusions and Future Work

We present two fast methods for inference in latent Gaussian spatial models, SVB and MCMC, and show that they outperform IVB in both speed and accuracy. Future work includes adopting model improvements into this framework, e.g. non-stationary spatial priors and spatial measurement noise, and the computation of model selection criteria such as the marginal likelihood.

References

[1] Penny, W. D., Trujillo-Barreto, N. J. and Friston, K. J.: *Bayesian fMRI time series analysis with spatial priors*, NeuroImage (2005)

[2] Papandreou, G. and Yuille, A.: *Gaussian sampling by local perturbations*, Advances in Neural Information Processing Systems 23 (2010)

[3] Rue, H. and Held, L.: *Gaussian Markov Random Fields: Theory and Applications*, CRC Press (2005)

[4] Bishop, C. M.: *Pattern Recognition and Machine Learning*, Springer (2006)

Acknowledgements

This work was funded by Swedish Research Council (Vetenskapsrådet) grant no 2013-5229. Participation in this conference was supported by the "BayesComp @ISBA junior travel support".