ParCFD'2020
32nd International Conference on Parallel Computational Fluid Dynamics
May 11-13 2020, Nice FRANCE

BAYESIAN VARIATIONAL SEISMIC TOMOGRAPHY

ANDREW CURTIS*, XUEBIN ZHAO* AND XIN ZHANG*

*School of GeoSciences, University of Edinburgh
EH9 3FE, Edinburgh, UNITED KINDOM
E-mail: Andrew.Curtis@ed.ac.uk, Xuebin.Zhao@ed.ac.uk and xzhang15@exseed.ed.ac.uk

Key words: variational inference, computational efficiency, tomography, imaging, inversion

Summary. Variational Bayesian inference (VBI) is an efficient alternative to Monte Carlo methods for solving Bayesian inverse problems. We use VBI to calculate an approximation to the posterior probability distribution that describes the solution to Bayesian tomographic inverse problems. This approximation is found using an optimization framework, yet the method provides fully probabilistic results. We perform Love wave tomography of the British Isles using three different variational methods as well as Monte Carlo sampling. The results show that variational methods can produce similar mean and uncertainty maps to those obtained from Monte Carlo, at significantly decreased computational cost.

1 INTRODUCTION

Seismic travel time tomography is commonly performed to image the Earth's interior structure and to infer subsurface properties. It can be formulated as an inverse problem that estimates parameters of interest (typically underground seismic velocity maps) given observed data[1]. Since this parameter-data relationship is highly non-linear, computationally expensive inverse methods are often deployed in order to obtain fully probabilistic solutions and to understand uncertainties in the tomographic results. Monte Carlo sampling methods estimate the posterior probability distribution which describes the solution to Bayesian tomographic problems, but they are computationally expensive and often intractable for high dimensional model spaces and large data sets due to the curse of dimensionality[2]. We therefore explore variational inference[3] as an alternative to perform Bayesian seismic tomography.

Variational methods have long been recognised as computationally efficient due to their scalability to large datasets. The idea is to approximate the posterior distribution by a simpler distribution q (called the variational distribution) that lies within a predefined variational family Q. We therefore try to find a member in this family that minimizes the difference between the posterior and the variational distributions. Thus, variational inference converts the more usual sampling problem into an efficient optimization, while still providing fully probabilistic results.

2 VARIATIONAL BAYESIAN INFERENCE

We solve non-linear inverse problem using Bayes' rule:

$$p(\mathbf{m}|\mathbf{d}_{obs}) = \frac{p(\mathbf{d}_{obs}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d}_{obs})} \tag{1}$$

where $\mathbf{m}$ is the vector of seismic velocities in discrete cells partitioning the subsurface model and $\mathbf{d}_{obs}$ is the vector of observed travel times of waves crossing those cells. Distribution $p(\mathbf{m})$ is the *prior* probability density function (pdf) that describes information about $\mathbf{m}$ before inversion. The *likelihood* $p(\mathbf{d}_{obs}|\mathbf{m})$ is the conditional probability of observing the data $\mathbf{d}_{obs}$ given that a particular model $\mathbf{m}$ is true. The denominator $p(\mathbf{d}_{obs})$ is called *evidence* and acts as a normalization constant in Bayesian inference. Combining the three terms on the right side of equation 1 gives the so-called *posterior* pdf on the left.

Traditionally, Bayesian problems are solved by Monte Carlo sampling based methods. They provide an ensemble of samples distributed according to the posterior distribution. The computational cost of exploring this posterior increases exponentially with the dimensionality of parameters $\mathbf{m}$, and so may be intractably expensive in high dimensional problems.

Alternatively, variational inference approximates the posterior distribution by a simpler one $q(\mathbf{m})$ (the variational distribution) defined to lie within a variational family or set Q. The method then seeks the best approximation to the true posterior pdf within that family by minimizing the Kullback-Leibler (KL) divergence:

$$\mathrm{KL}[q(\mathbf{m})||p(\mathbf{m}|\mathbf{d}_{obs})] = \mathrm{E}_q\left[\log\left(q(\mathbf{m})/p(\mathbf{m}|\mathbf{d}_{obs})\right)\right] \tag{2}$$

where the expectation is taken with respect to the variational distribution $q(\mathbf{m})$. $\mathrm{KL}[q||p]$ measures the difference (distance) between the two distributions. It has the property $KL[q(\mathbf{m})||p(\mathbf{m}|\mathbf{d}_{obs})] \geq 0$ with equality only when $q(\mathbf{m}) = p(\mathbf{m}|\mathbf{d}_{obs})$. By minimizing KL divergence within the variational family, the resulting optimal distribution $q^*(\mathbf{m})$ is the one closest to the posterior distribution, which serves as an optimal approximation to the left side of equation 1. Combining equations 1 and 2, we have:

$$\log p(\mathbf{d}_{obs}) = \mathrm{ELBO}(q) + \mathrm{KL}[q(\mathbf{m})||p(\mathbf{m}|\mathbf{d}_{obs})] \tag{3}$$

where $\mathrm{ELBO}(q)$ is defined as $\mathrm{ELBO}(q) = \mathrm{E}_q\left[\log p(\mathbf{m},\mathbf{d}_{obs})\right] - \mathrm{E}_q\left[\log q(\mathbf{m})\right]$ and is the evidence lower bound of the logarithmic evidence $log\, p(\mathbf{d}_{obs})$. Minimizing the KL divergence is therefore equivalent to maximizing $\mathrm{ELBO}(q)$, since $log\, p(\mathbf{d}_{obs})$ is fixed for different $q(\mathbf{m})$. Thus an intractable, high dimensional sampling problem is converted into a numerical optimization problem, while still providing fully probabilistic results.

3 VARIATIONAL TOMOGRAPHIC TEST

We perform tomography of the British Isles using the travel time data from 10s period Love waves travelling between pairs of using variational inference shown as triangles in Figure 1. We test three variational methods, namely: automatic differential variational inference (ADVI)[4], Stein variational gradient descent (SVGD)[5] and normalizing flows[6]. We also perform Metropolis-Hastings Markov chain Monte Carlo (MH-McMC) sampling method for comparison.

From left to right, Figure 1 shows the mean and standard deviation maps of the tomographic posterior pdf using the four methods. ADVI and SVGD provide smoother mean and uncertainty maps, whereas MH-McMC and normalizing flows provide more detailed results. The main features of the four results are quite similar to each other, and are consistent with previous studies across this area and the known geology[7]. For example, we observe high velocity regions in the basement of the Scottish Highlands (the Northern mainland), Southern Uplands (at 4°W, 55°N), and East Midlands (at 1°W, 53°N). Low velocity structures are observed along the Eastern coastline of mainland Britain, to the East of Ireland down to Southwest Wales, and in the East Irish Sea. Uncertainty maps display low uncertainty areas of Scotland and southern England, and higher uncertainty in the eastern mainland, as found previously[7]. We expect the uncertainty map from ADVI to be inaccurate as this method applies a log-Gaussian based approximation to the posterior pdf: it finds nearly the same uncertainty level inside the receiver array, except for those regions where the seismometers are densely placed, which deviates significantly from the other methods. Nevertheless, the high consistency between the main features of the four mean models suggests that the group velocity maps may be accurate.

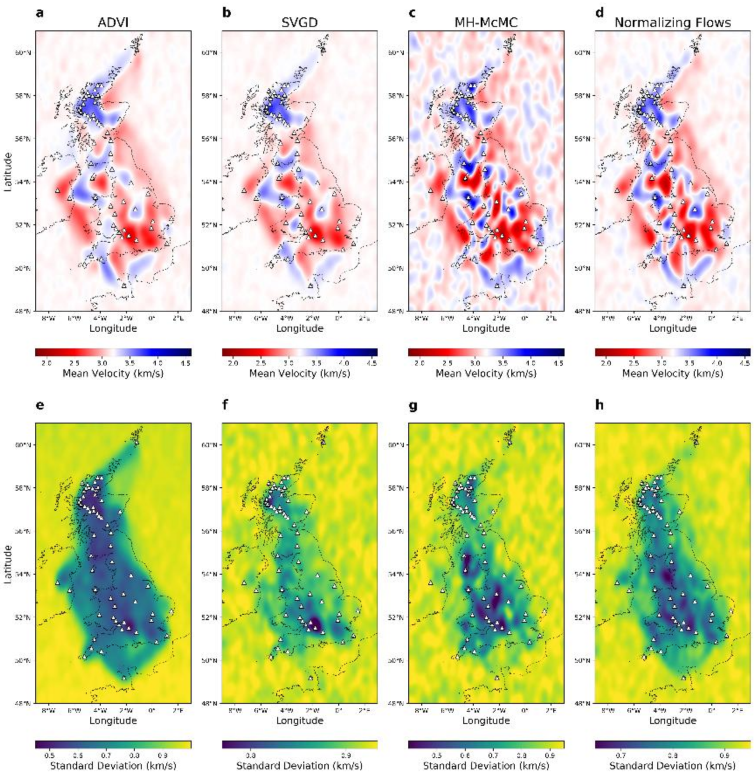

Figure 1: Love wave group velocity maps of the British Isles at 10 s period: mean (top row) and standard deviation (bottom row) of the posterior distributions using different methods. Triangles mark seismometers.

Table 1 lists the computational cost of the four methods. We find ADVI is the cheapest method as it uses fewest forward evaluations, but from this example as well as synthetic tests in previous studies we observe that it fails to provide accurate uncertainty maps due to the above approximation[8]. Normalizing flows and SVGD produce more convincing uncertainty estimates, but the former is far less computationally demanding than the latter. Although we acknowledge some subjectivity in choosing when iterations in each method are halted, all three variational methods have significantly decreased computational cost compared to Monte Carlo which requires millions of forward evaluations for this high dimensional inference problem.

Method	Forward Evaluations
ADVI	10,000
Normalizing Flows	100,000
SVGD	600,000
MH-McMC	15,000,000

Table 1 : Number of forward evaluations for ADVI, normalizing flows, SVGD and MH-McMC to obtain the results shown in Figure 1.

4 CONCLUSIONS

We solve Bayesian travel time tomographic problems under an optimization framework using variational inference. We test three variational methods by performing Love wave tomography to construct seismic group velocity maps of the British Isles, and compare them to Monte Carlo sampling tomography. All four methods give similar and convincing mean velocity maps, and all but ADVI provide similar standard deviation maps, that are consistent with known geology and previous research in this area. Variational methods appear to be the more computationally efficient compared to Monte Carlo in this particular problem.

5 ACKNOWLEDGEMENTS

We thank Edinburgh Imaging Project sponsors (Schlumberger, BP and Total) for supporting this research.

REFERENCES

[1] Curtis, A. and Snieder, R., 2002. Probing the Earth's interior with seismic tomography. *International Geophysics Series*, *81*(A), pp.861-874.

[2] Curtis, A. and Lomax, A., 2001. Prior information, sampling distributions, and the curse of dimensionality. *Geophysics*, 66(2), pp.372-378.

[3] Blei, D.M., Kucukelbir, A. and McAuliffe, J.D., 2017. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518), pp.859-877.

[4] Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A. and Blei, D.M., 2017. Automatic differentiation variational inference. *The Journal of Machine Learning Research*, 18(1), pp.430-474.

[5] Liu, Q. and Wang, D., 2016. Stein variational gradient descent: A general purpose Bayesian inference algorithm. *arXiv preprint* arXiv:1608.04471.

[6] Rezende, D. and Mohamed, S., 2015, June. Variational inference with normalizing flows. *In International Conference on Machine Learning* (pp. 1530-1538). PMLR.

[7] Galetti, E., Curtis, A., Baptie, B., Jenkins, D. and Nicolson, H., 2017. Transdimensional Love-wave tomography of the British Isles and shear-velocity structure of the East Irish Sea Basin from ambient-noise interferometry. *Geophysical Journal International*, 208(1), pp.36-58.

[8] Zhang, X. and Curtis, A., 2020. Seismic tomography using variational inference methods. *Journal of*

Geophysical Research: Solid Earth, 125(4), p.e2019JB018589.