Bayesian AVA Elastic Seismic Inversion using Stein Variational Gradient Descent (SVGD)

R. Biswas[1], M. Walker[1], J. Zhang[1], P. Paramo[1], K. Wolf[1], S. Gerth[1], J. Winterbourne[1], A. Roy[1], P. Morris[1], C. Decalf[1], Y. Zheng[2], R. Warnick[3]

[1] bp; [2] Formerly bp now Google; [3] Formerly NAG now Microsoft Security Research

Summary

We present a new scheme to solve seismic Amplitude Vs. Angle (AVA) inversion problem in a probabilistic framework. We have used a recently developed Stein Variational Gradient Descent (SVGD) method in our proposed Bayesian Integrated Reservoir Characterization (BIRCh) workflow. BIRCh is based on a Bayesian framework inference algorithm that seeks a set of points (or particles) to approximate the target distribution using iterative gradient-based updates. It surpasses several of the disadvantages of traditional Bayesian techniques pose, like a huge number of sampling required in the case of Markov Chain Monte Carlo (MCMC) and Variational Inference (VI) struggle to represent a complex posterior distribution. We demonstrate the BIRCh workflow on a seismic inversion problem.

Bayesian AVA Elastic Seismic Inversion using Stein Variational Gradient Descent (SVGD)

Introduction

For many inverse problems, no unique solution exists due to uncertainty on the observed data and/or fundamental non-uniqueness in the inverse relationship between data and model parameters. Such problems are often solved using a Bayesian approach where model parameters are cast as random variables, which allows the resultant uncertainty over the parameters to be described using the so-called posterior probability distribution. This distribution is calculated using Bayes' rule as the normalized product of the likelihood and prior distributions; the former embodies information derived from the data, and the latter embodies information derived from additional, independent prior sources. The posterior may be written in a convenient parameterized form if certain parametric distributions are deemed appropriate for the likelihood and prior. However, for most problems, such assumptions cannot be made, and the posterior cannot be written in closed form. Furthermore, suppose the dimensionality of the model space is high. In that case, numerical evaluation is practically impossible due to a large number of evaluations of likelihood and prior, which would be required. In this case, approximate inference methods must be used to estimate the posterior distribution, with which we can do further practical analysis such as point estimation, uncertainty characterization, and, ultimately, decision-making.

The most common approximate approaches are sampling methods which attempt to draw samples from the posterior. They are designed such that the distribution of samples is guaranteed to converge to the posterior as the number of iterations approaches infinity. However, for complex posteriors (e.g., those with multimodality) on high-dimension model spaces, they often struggle to converge in a finite number of iterations and are thus often impractical to run in practice. An alternative class of methods is Variational Inference (VI) (David et al., 2017) methods which attempt to approximate the posterior using another known distribution by minimizing some metric of difference between it and the posterior target distribution. Such methods can be very efficient, but there is no guarantee of convergence to the posterior even after an infinite number of iterations (the algorithm may converge to a local minimum in the space of approximating distributions). Furthermore, they are highly dependent on the suitability to the problem of the parametric approximation choice, thus requiring careful case-by-case selection.

Stein Variational Gradient Descent (SVGD) is a recently devised algorithm (Liu and Wang, 2016) that overcomes some of the drawbacks of VI by seeking a set of points (known as "particles") in the model space which optimally approximates the target distribution. Unlike sampling algorithms, it relies solely on iterative gradient-based updates to the particles to obtain these, rather than random sampling, so it can be much more efficient. Also, unlike typical VI algorithms, there is no requirement to choose an appropriate candidate parametric distribution for each problem. Furthermore, it can be easily combined with various state-of-the-art techniques responsible for the success of gradient optimization, including stochastic gradient, adaptive learning rates, and momentum. This makes it uniquely flexible and scalable for solving inverse problems where no closed form exists for the posterior.

Inference of elastic parameters from amplitude versus angle (AVA) data is a highly non-unique inverse problem; AVA data suffers from both measurement and epistemic error, and the inverse relationship between this data and the parameters is fundamentally non-unique (due to band limitation of the data and its dominant sensitivity to contrasts in the elastic parameters). The dimensionality of the problem is also often huge, with subsurface models with thousands or millions of grid cells being the norm. Thus, AVA inversion is often solved in the Bayesian framework. Simplifying assumptions can be made about the AVA-physics, which allow the likelihood to take on a parameterized form, and, similarly, a simple parameterized distribution can be assumed for the prior. Such simplifications allow a closed form to be found for the posterior but lead to the loss of both prior and data information. If such simplifications are to be avoided, we must typically use sampling or VI algorithms to do the inversion, but they often suffer from the general drawbacks of those approaches described above.

Thus, in this contribution, we demonstrate how SVGD can be used to do efficient AVA inversion without making simplifying assumptions about AVA physics and the prior. We first describe the details of Bayesian AVA inversion and how SVGD can be used to solve it. We then show an example of SVGD being used to solve the elastic inversion problem without simplifying the AVA physics. Finally, we discuss how SVGD has recently been used to great effect to solve the elastic inversion part of the Bayesian Integrated Reservoir Characterization (BIRCh) of the two-step inversion scheme (Paramo et al., 2023), in which a complex prior distribution (i.e., with no closed form) has been successfully incorporated into the SVGD inversion (Walker et al., 2023).

Theory

The posterior is related to the observed data and prior information via Bayes' rule (Sen & Stoffa, 2013),

$$p(\boldsymbol{m}|\boldsymbol{d}) = \frac{p(\boldsymbol{d}|\boldsymbol{m})p(\boldsymbol{m})}{p(\boldsymbol{d})}, \tag{1}$$

where $p(\boldsymbol{d}|\boldsymbol{m})$ is the likelihood distribution which describes the probability of the observed data $\boldsymbol{d}$ given an instance of the elastic parameter model $\boldsymbol{m}$. It is a function of a forward model relating the parameters to the data and some models of data noise. The AVA data is angle-domain post-migration data (either gathers or partial stacks). The forward model comprises 1D convolution of an angle-dependent wavelet and reflectivity series calculated using the Zoeppritz equations (or some approximation of them). $p(\boldsymbol{d})$ is a constant given the data has been observed, and is hence fixed, and can be thought of as a normalization constant ensuring the integral (or sum) of the posterior is equal to one; it may be calculated simply in theory, but in practice, it can be hard to estimate accurately. $p(\boldsymbol{m})$ is the prior distribution which describes prior knowledge we have about the elastic parameters.

For completeness, we will describe the SVGD algorithm in brief. The objective is to approximate a positive density function $\boldsymbol{p}(\boldsymbol{x})$, where $\boldsymbol{x} \subseteq \mathbb{R}^{\boldsymbol{d}}$, using a set of particles $\{\boldsymbol{x_i}\}_{\boldsymbol{i=1}}^{\boldsymbol{n}}$. In our case, $\boldsymbol{p}(\boldsymbol{x})$ is the posterior distribution $\boldsymbol{p}(\boldsymbol{m}|\boldsymbol{d})$ described in eq. 1. The particles are first initialized with some known simpler distribution $\boldsymbol{q_0}$ and then iteratively updated using the equation $\boldsymbol{x_i} \leftarrow \boldsymbol{x_i} + \boldsymbol{\epsilon\phi}(\boldsymbol{x_i}), \forall_{\boldsymbol{i}} = \boldsymbol{1}, \dots, \boldsymbol{n}$. Where $\boldsymbol{\epsilon}$ is the small step size and $\boldsymbol{\phi}(\boldsymbol{x})$ is the perturbation direction or velocity field. The perturbation direction is chosen such that the decrease in the KL divergence between the distribution of the updated particle and the target distribution is maximum. The perturbation direction $\boldsymbol{\phi}(\boldsymbol{x})$ can be described in equation as:

$$\begin{aligned}\boldsymbol{\phi} &= \mathbf{arg}\ \max_{\boldsymbol{\phi}\in\mathcal{F}}\{\underbrace{\boldsymbol{KL}(\boldsymbol{q}||\boldsymbol{p})}_{\textit{old particles}} - \underbrace{\boldsymbol{KL}(\boldsymbol{q}_{[\epsilon\phi]}||\boldsymbol{p})}_{\textit{updated particles}}\} \\ &= \mathbf{arg}\ \max_{\boldsymbol{\phi}\in\mathcal{F}}\{-\frac{\boldsymbol{d}}{\boldsymbol{d\epsilon}}\boldsymbol{KL}(\boldsymbol{q}_{[\epsilon\phi]}||\boldsymbol{p})|_{\epsilon=0}\}\end{aligned} \tag{2}$$

Where $\boldsymbol{q}$ represents the density of the old particles $\boldsymbol{x}$ and $\boldsymbol{q}_{[\epsilon\phi]}$ is the density of the updated particles. And $\mathcal{F}$ is the set of the directions for perturbations where optimization is performed. The objective function in the eqn. 2 has a known simple linear functional form represented by the Stein's operator $\mathbf{T}$ as:

$$-\frac{\boldsymbol{d}}{\boldsymbol{d\epsilon}}\boldsymbol{KL}(\boldsymbol{q}_{[\epsilon\phi]}||\boldsymbol{p})\Big|_{\epsilon=0} = \boldsymbol{E}_{x\sim q}[\mathrm{T}_p\boldsymbol{\phi}(\boldsymbol{x})], \mathbf{with}\ \mathrm{T}_p\boldsymbol{\phi}(\boldsymbol{x}) = \nabla_{\mathrm{x}} \log \boldsymbol{p}(\boldsymbol{x})^{\mathrm{T}} + \dot{\nabla}_{\mathrm{x}}\boldsymbol{\phi}(\mathrm{x}) \tag{3}$$

Now, the optimization can be rewritten in the Stein Discrepancy form as:

$$\boldsymbol{D}(\boldsymbol{q}||\boldsymbol{p}) = \max_{\boldsymbol{\phi}\in\mathcal{F}}\{\boldsymbol{E}_q[\mathrm{T}_p\boldsymbol{\phi}]\} \tag{4}$$

Where the Stein discrepancy $\boldsymbol{D}(\boldsymbol{q}||\boldsymbol{p}) = \boldsymbol{0}$ iff $\boldsymbol{q} = \boldsymbol{p}$ and $\mathcal{F}$ is large enough. To simplify the perturbation directions $\mathcal{F}$is taken as the unit ball of a vector valued Reproducible Kernel Hilbert Space (RKHS) $\mathcal{H} = \mathcal{H}_{\boldsymbol{o}} \times \dots \times \mathcal{H}_{\boldsymbol{o}}$. $\mathcal{H}_{\boldsymbol{o}}$ is the scalar valued RKHS calculated using a scalar valued positive definite kernel $\boldsymbol{k}(\boldsymbol{x}, \boldsymbol{x}')$. Thus, the Kernelized Stein Discrepancy can be written as:

$$\boldsymbol{D}\left(\boldsymbol{q}\middle||\boldsymbol{p}\right) = \max_{\boldsymbol{\phi}\in\mathcal{H}}\left\{\boldsymbol{E}_q[\mathrm{T}_p\boldsymbol{\phi}]\ \ \boldsymbol{s.t.}\,||\boldsymbol{\phi}||_{\mathcal{H}} \leq \mathbf{1}\right\}. \tag{5}$$

Now, there exist a closed-form solution to the above objective function as

$$\boldsymbol{\phi}^*(\boldsymbol{x}) \propto \boldsymbol{E}_{x\sim q}[\mathrm{T}_p\boldsymbol{k}(\boldsymbol{x}, \boldsymbol{x}')] \tag{6}$$

$$= E_{x\sim q}[\nabla_x \log p(x)\ k(x,x') + \nabla k(x,x')]$$

Thus, the final SVGD update for the particles related to each other using a scalar RKHS $\boldsymbol{k}(\boldsymbol{x},.)$ Is

$$\boldsymbol{x_i} \leftarrow \boldsymbol{x_i} + \epsilon \widehat{\mathbb{E}}_{x\sim\{x_i\}_{i=1}^{n}} \left[\underbrace{\nabla_x \log p(x)\, k(x,x_i)}_{Attraction\ Force} + \underbrace{\nabla_x k(x,x_i)}_{Repulsive\ Force} \right], \quad \forall i = 1, \dots, n. \quad (7)$$

For the inversion, first we calculate the angle dependant scalar to scale the input seismic data to that of the synthetic data. We used Zoeppritz equation to model the P-P reflection at the interface and convolved it with the wavelet to calculate the synthetic seismogram. Second, step is to generate the initial particles for the SVGD algorithm. For the AVA problem, a particle is a full time-series pseudo-log containing V_P, V_S, and density values for each layer. The initial particles can be generated using some distribution.

We then run the SVGD inversion, where the particles are updated iteratively in the direction dictated by eqn. 6. The attraction force in the eqn. 7, is guided by the gradient of the posterior distribution $p(x)$, and it forces the particles to move in the nearest higher probability region. And the second term, repulsive term, which is the gradient of the kernel, creates a repulsive force between each of the particles and prevent them to collapse at the same location and spread them apart in the model space. The kernel dictates how a particle interact with each individual particle in the model space. In our case, we used a radial basis function as the kernel defined as $k(x,x') = \exp(-h||x-x'||_2^2)$. Overall, this particle update creates a fascinating "momentum" effect in which the particles move cooperatively to avoid the local optimum and converge to a variety of positions that closely resemble the target distribution. However, if we use a single particle in SVGD, the algorithm works similar to a gradient descent algorithm as the gradient of the kernel becomes zero, and the update direction is only determined from the attraction force which will lead it to the nearby minima.

Synthetic Example

To demonstrate the methodology, we used a real well-log from a gas field. We used Zoeppritz equation to generate the seismic data and added some noise to it to represent our observed seismic data. Figure 1 shows the plot of the synthetic seismic data generated, which is treated as the observed seismic data and is inverted for the elastic properties. We generated total of nine angles starting from $\mathbf{0^o}$ to $\mathbf{50^o}$ at an interval of $\mathbf{5^o}$.

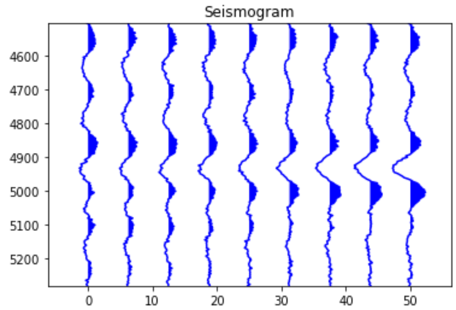

Figure 1. *Synthetic seismic data added with noise is treated as an observed seismic angle gathers.*

Scalars for the synthetic example is unity for all angles. Using the well-log, we calculated the correlation coefficients for the V_P, V_S, and density. SVGD algorithm requires multiple initial particles, we generated 1000 particles randomly using a correlated Gaussian noise on top of the low-frequency model. Figure 2a shows the cloud of 1000 initial particles in purple, low-frequency model in black and the true V_P, V_S, and density in red.

After running the SVGD algorithm, Figure 2b represent the final updated 1000 particles representing the posterior distribution at the end of 2000 iterations. The blue cloud is representing the true model in red remarkably.

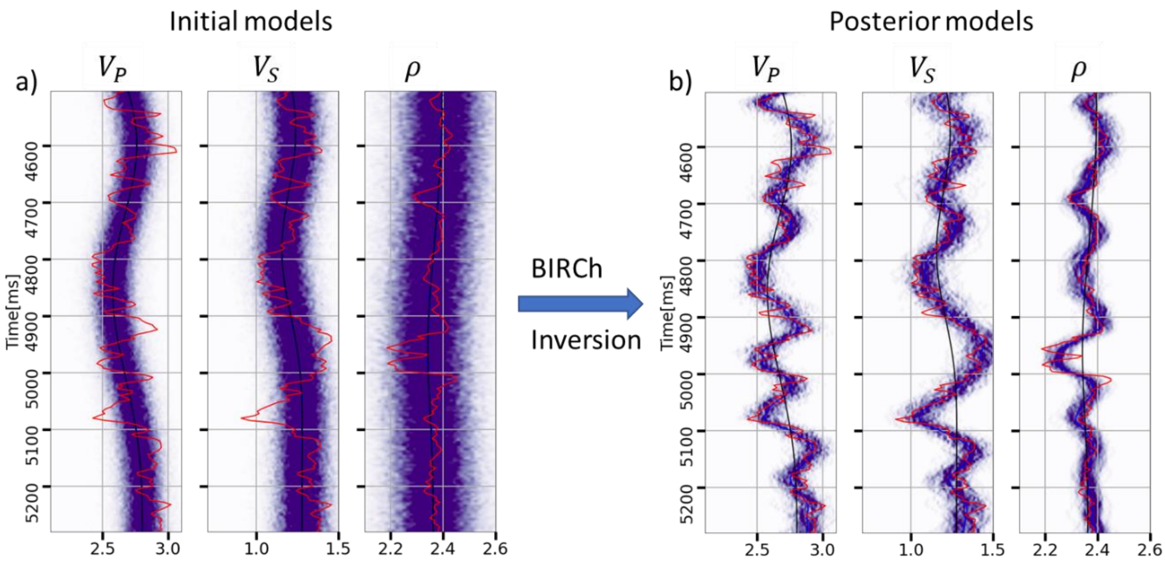

***Figure 2** shows the a) 1000 Initial models generated from a prior distribution, and b) shows the 1000 final posterior distribution as a result of the BIRCh inversion workflow. The purple cloud represents the representative distribution from the 1000 particles. True model in red for V_P, V_S, and density, the blue line represents the mean model of the 1000 particles, and the black line is the low-frequency model used.*

Conclusion

A new workflow, namely BIRCh, was devised which presents an alternative technique to solve the seismic AVA inversion in probabilistic manner utilizing Bayesian framework and estimate the posterior distribution of the model parameters V_P, V_S, and density. The method utilizes a particle-based method, where it solves a local optimization, but the particles can interact with each and thus collectively can overcome the local minima and reach a global minimum and represent the uncertainty. We demonstrated the method using a synthetic example. However, we also present some of the real data case studies in Walker et. al. (2023), and Paramo et. al. (2023).

Acknowledgments

We wish to thank BP management for permission to publish.

References

David M. Blei, Alp Kucukelbir & Jon D. McAuliffe (2017) Variational Inference: A Review for Statisticians, Journal of the American Statistical Association, 112:518, 859-877, DOI: 10.1080/01621459.2017.1285773

Liu, Qiang, and Dilin Wang. "Stein variational gradient descent: A general purpose Bayesian inference algorithm." Advances in neural information processing systems 29 (2016).

Paramo, P., Wolf, K., Zhang, J., Biswas, R., Walker, M., Winterbourne, J., Roy, A., Morris, P., Decalf, C., Cook, P., Verhaeghe, J., Brown, R., Gerth, S., Packer, K. & Zheng, Y. [2023]. Bayesian Integrated Reservoir Characterization (BIRCh) and its application to the Raven field. Submitted to 84th EAGE Conference & Exhibition, Extended Abstracts.

Walker, M., Paramo, P., Wolf, K., Zhang, J., Biswas, R., Gerth, S., Winterbourne, J., Morris, P., Decalf, C., Cook, P., Verhaeghe, J., Brown, R., Packer, K., Zheng, Y., Warnick, R. [2023]. Gaussian Mixture Models as priors in a gradient-based variational algorithm for Bayesian elastic inversion. Submitted to 84th EAGE Conference & Exhibition, Extended Abstracts.

Zhang, J., Biswas, R., Kumar, D., Walker, M., Paramo, P., Wolf, K., Yusifov, A., Gerth, S., Winterbourne, J., Liu, Q. & Liu, X. [2022]. Got noisy impedance? Try this. In Second International Meeting for Applied Geoscience & Energy, pp. 160-164. SEG and AAPG, 2022.

Sen, Mrinal K., and Paul L. Stoffa. Global optimization methods in geophysical inversion. Cambridge University Press, 2013.