DOI:10.1190/image2024-4101084.1

FWI uncertainty analysis with Stein variational gradient descent

Haoyang Cen, Kaihang Guo, Diancheng Wang, Viridien*

Summary

Full waveform inversion (FWI) is becoming an increasingly dominant force in velocity model building and seismic imaging, often providing us with unrivaled focusing and resolution of the subsurface image. The superior velocity models and seismic images achieved through FWI also enable serious and more meaningful discussions of uncertainties on FWI. While the Bayesian inference framework offers a general foundation for FWI uncertainty analysis, the high computational cost associated with classical algorithms, such as Monte Carlo methods, to sample the posterior distribution prohibits it from being applied to industrial-scale problems. The recent development of variational inference (VI) theory presents a promising alternative to traditional sampling algorithms, as it can generate reasonable estimations of the posterior distribution at a more affordable computational cost. In this abstract, we describe an FWI uncertainty analysis method based on a specific type of VI algorithm, the Stein variational gradient descent (SVGD). We demonstrate the efficacy and practicality of this method through 2D synthetic and 3D real data examples.

Introduction

Full waveform inversion (FWI) was proposed four decades ago by pioneers (Lailly, 1983; Tarantola, 1984) with the vision that it could one day become the ultimate elegant solution to the whole seismic imaging problem. Through years of unremitting efforts from the geophysics community, this vision is becoming more and more of a reality today. Especially in the last decade, we have witnessed tremendous breakthroughs in FWI research and practice (Warner and Guasch, 2014; Shen et al., 2017; Wang et al., 2019; Liu et al., 2023). Nowadays, FWI is undoubtedly established as the driving force in seismic velocity model building and imaging, routinely generating high-resolution images of the Earth's subsurface that cannot be matched by any other technique.

These advancements in FWI also enable more serious discussions on FWI uncertainty analysis, as we believe such discussions might be less meaningful if there is not a good velocity model or image to analyze in the first place. Bayesian inference provides a solid framework for us to quantify FWI uncertainties by updating our a priori knowledge with new information from the data and constructing the posterior probability density function (PDF) through Bayes' theorem. Nonetheless, quantifying FWI uncertainty remains a daunting task due to the highly non-linear nature and large dimensions of the FWI problem, as well as the high cost of wave simulations. Various efforts have gone into tackling the challenges posed by FWI uncertainty. One family of such efforts makes a Gaussian approximation to the posterior PDF around the maximum a posteriori (MAP) model and tries to directly probe the structure of the posterior variance by analyzing the Hessian of the FWI misfit (Zhu et al., 2016; Zhai et al., 2022). While the Gaussian assumption may be generally acceptable, evaluating the Hessian of FWI is still computationally prohibitive. Further simplifications are often made to approximate the Hessian, e.g., via randomized singular value decomposition (SVD). These approximations could potentially undermine the reliability of the estimated posterior PDF and the FWI uncertainty. Another family of efforts follows the common practice of evaluating a target distribution by sampling the probability space with Markov chain Monte Carlo (MCMC) (Brooks et al., 2011). Various flavors of MCMC methods have been explored to test their applicability in FWI uncertainty analysis (e.g., Ray et al., 2017; Gebraad et al., 2020). Despite its remarkable success in the broad statistics discipline, MCMC is not immune to the notorious curse of dimensionality, especially when the problem's dimension approaches the industrial FWI level. To obtain any sensible uncertainty results through MCMC, at least thousands, if not millions, of samples are needed, which is plainly impractical.

In recent years, variational inference (VI) (Blei et al., 2017) has emerged as a powerful alternative to MCMC in approximating posterior PDFs. While it does not guarantee convergence to the exact posterior as MCMC does, VI provides an analytical approximation to the posterior, and often at a reduced computational cost, making it especially attractive for FWI uncertainty analysis. Stein variational gradient descent (SVGD) (Liu and Wang, 2016) is a specific type of VI algorithm that uses the particle update approach to iteratively move a group of particles (or samples) towards the target distribution, and it has shown to be effective and practical in large-scale FWI uncertainty analysis (Zhang et al., 2023). As mentioned earlier, one important factor in obtaining more meaningful FWI uncertainty through affordable algorithms such as SVGD is to perform the analysis on top of a good FWI model. Therefore, the FWI results we can achieve today offer us a great opportunity to carry out SVGD uncertainty analysis in field data projects.

In the following sections, we will describe our method and workflow and demonstrate their effectiveness through synthetic and field data examples.

Theory and Method

Bayesian inference is a statistical technique that allows us to update our beliefs about a model based on new evidence or data. It is based on Bayes' theorem, which describes the probability of a model given observed data as

$$\pi(m) \equiv p(m|d) = \frac{p(d|m)*p(m)}{p(d)}, \quad (1)$$

where $\pi(m) \equiv p(m|d)$ is the posterior probability, representing our updated belief about the model m after observing the data d; $p(m)$ is the prior probability, representing our initial knowledge of the model before the data; and $p(d)$ is a normalization factor also commonly referred to as the evidence. $p(d|m)$ is the likelihood, or the probability of observing the data given the model. For FWI, the likelihood function is essentially the forward problem, which is often expressed as

$$p(d|m) \propto \exp\left(-\frac{\|d-F(m)\|^2}{2\sigma^2}\right), \quad (2)$$

where F represents the forward modeling operator, σ^2 is the data variance, and a $L2$ norm misfit is assumed here.

Our objective is to find $\pi(m)$ for the problem of FWI, based on which we can derive uncertainties through basic statistics. However, this is not as simple as it seems, mainly for two reasons. First, the dimension of FWI model is usually very large (in the order of millions), which makes it difficult to approximate the posterior using common Monte Carlo methods, as a very large number of samples are needed. Second, for each sample to be generated, we must evaluate the likelihood function, i.e., the forward modeling. Thus, the price of sample generation is high due to the cost of wave propagation. VI offers a good alternative in this situation as it seeks to approximate the posterior by minimizing the Kullback-Leibler (KL) divergence between a surrogate distribution and the target distribution. Since VI solves the problem through optimization, it can avoid the need of generating a large number of samples, making it more efficient in practice. As a specific type of VI algorithms, SVGD is a non-parametric method that represents the approximating distribution as a set of particles, or samples, that are updated iteratively using a form of functional gradient descent. The key idea behind SVGD is to move the particles in the direction of the gradient of the KL divergence between the true distribution and the approximating distribution. We now lay out the details of our uncertainty analysis workflow based on SVGD.

The first step of SVGD is generating initial particles. Here, we want to reiterate a point we made earlier, that we should perform the analysis around a good reference FWI model, which is the best FWI result that can be obtained with the data. This "around a local optimum" philosophy not only makes our uncertainty result more meaningful but also helps the SVGD convergence and reduces cost. Starting from this reference model, we then add random perturbations to it to construct the initial particles. The random perturbations are generated using Gaussian random fields (GRF) with the Matern covariance kernel (Rasmussen and Williams, 2006):

$$C(x,x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{||x-x'||}{\rho}\right)^{\nu} K_\nu \left(\sqrt{2\nu}\frac{||x-x'||}{\rho}\right), \quad (3)$$

where σ^2 is the variance, ρ is the length scale, ν is a smoothness parameter, Γ is the gamma function, K_ν is the modified Bessel function of the second kind of order ν, and $\|x-x'\|$ is the Euclidean distance between the two input vectors x and x'. Figure 1 shows four example realizations of GRF generated from (3) with $\nu = 1.5$.

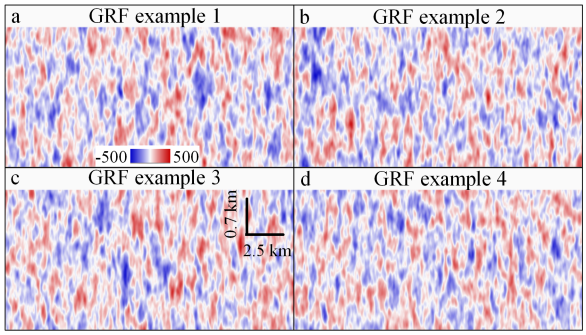

Figure 1: Four examples of GRF perturbations generated according to the Matern covariance kernel.

After we have generated a set of N initial particles $\{m_i: i = 1{\sim}N\}$, we update them iteratively according to the SVGD algorithm:

$$m_i^{l+1} = m_i^l + \epsilon_l * (Part1\ update + Part2\ update), \quad (4)$$

and

$$Part1\ update = \frac{1}{N}\sum_{j=1}^{N} k(m_j^l, m_i^l)\nabla_{m_j^l} \log \pi(m_j^l), \quad (5)$$

$$Part2\ update = \frac{1}{N}\sum_{j=1}^{N} \nabla_{m_j^l} k(m_j^l, m_i^l), \quad (6)$$

where ϵ_l is the step size at the l-th iteration, and $k(m,m')$ is the kernel function from the reproducing kernel Hilbert space (RKHS). Here, we deliberately separate the particle update into two parts, because they each have nice intuitive interpretations that would help our understanding of the dynamics of SVGD, as well as facilitate later discussions. The Part1 update is mainly driven by the gradient of the posterior; therefore, it acts as a *converging force* that steers all particles towards high probability regions. On the other hand, the Part2 update behaves as a *repulsive force* that pushes particles away from each other so that they won't collapse to the same point. A good balance between Part1 and Part2 updates is essential for the particles to eventually converge to a distribution that approximates the target distribution well. Computationally, evaluating the Part1 update for each particle j is almost equivalent to computing one FWI gradient, which is the main cost for the algorithm.

A commonly chosen family of functions to be used as kernels are the radial basis function (RBF) kernels, which satisfy the requirements of being positive definite and continuously differentiable:

$$k(m,m') = \exp\left(-\frac{\|m-m'\|^2}{h^2}\right), \quad (7)$$

where h is an important parameter usually referred to as the bandwidth. The bandwidth h plays a crucial role in properly balancing the strengths of Part1 and Part2 updates and

therefore should be chosen carefully. We follow the heuristic choice of the median trick that is proven effective in various applications:

$$h = Median\{\|m_i - m_j\| : i \neq j\}. \quad (8)$$

Besides being an important parameter in SVGD, h can also serve as a convenient and useful QC parameter telling us how the particles move as a group, as we will see shortly.

Once we run the SVGD algorithm until convergence, we can easily compute certain statistics from the samples, such as the mean and the standard deviation. We use the standard deviation as a quantification of the velocity model uncertainty. In addition to velocity uncertainty, we are also interested in image depth uncertainty in practice. The bridge from the velocity model to the seismic image is any kind of migration procedure. In principle, we could run Kirchhoff or RTM migration on all the final particles to obtain N volumes of image, but that would be costly as well as unnecessary. We might only be focusing on a few target events so a simple map-migration on those target horizons would suffice to generate the depth uncertainties we desire.

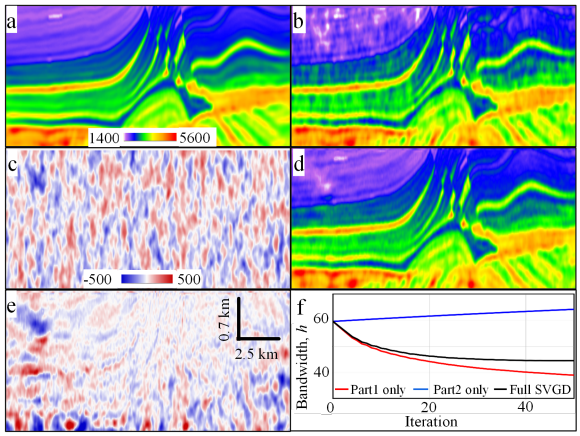

Figure 2: Synthetic test on the Marmousi model. (a) Reference FWI model. (b) An example initial particle. (c) GRF perturbation of the example initial particle. (d) Example particle at convergence. (e) Difference between d and a. (f) h-curve QC; h has arbitrary unit.

Examples

We first validate our method with a 2D synthetic test on the Marmousi model. The synthetic data is generated by a fixed-spread acquisition with a shot spacing of 100 m and a receiver spacing of 50 m. A ricker wavelet peaked at 10 Hz is used. Fifty initial particles are generated based on the reference FWI model (Figure 2a). One such initial particle is shown as an example in Figure 2b, together with its GRF perturbation (Figure 2c). The uncertainty analysis is run at 6 Hz since higher frequency FWI will not have further impact on the kinematics, and the *L2* cost function is used. We run 50 SVGD iterations until the particles appear to converge. The same example particle at convergence is shown in Figure 2d, and its difference against the reference FWI model in Figure 2e. We can see that in the central region where data constraint is good, the particle closely matches the reference model, while in poorly illuminated areas at the edges the difference remains relatively large, as one should expect. Nevertheless, perfect convergence to the reference model is not anticipated even in the good illumination area because the Part2 update as a repulsive force will try to maintain a reasonable diversity among all the particles.

Next, we further examine how the bandwidth parameter h evolves with iterations (Figure 2f) for a better visualization of the collective behavior of the particles. Three sets of tests are run independently: with Part1 update only (red curve), with Part2 update only (blue curve), and with the full SVGD update (black curve). These h-curves clearly illustrate the converging and repulsive forces of the two update parts, as well as the balanced dynamics between them.

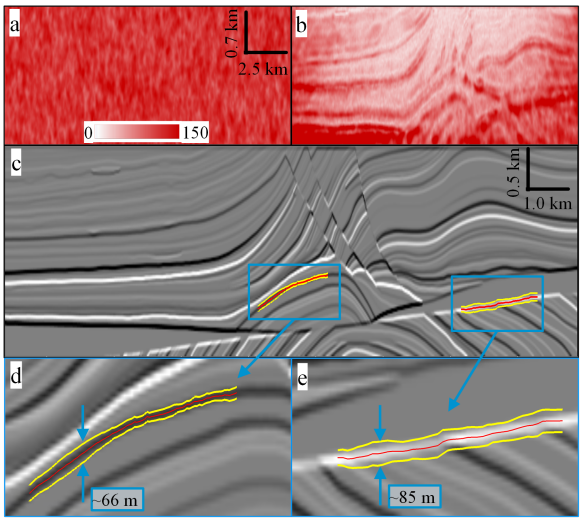

Figure 3: Uncertainty quantification for the Marmousi test. (a) Standard deviation of the initial particles. (b) Standard deviation of the final particles. (c) Target horizons with uncertainty bounds computed from map-migration. (d-e) Zoom-ins of the target horizons.

With the final particles produced by SVGD, we compute the standard deviation of them to quantify the velocity uncertainty and map-migrate target horizons with them to assess depth uncertainties associated with those horizons. Figures 3a-b show the standard deviations for the initial and final particles, from which we can see that the velocity uncertainty is much reduced in well illuminated areas thanks to the new information provided by data. In contrast, our knowledge about the regions with insufficient constraints remains highly uncertain. Figure 3c shows two target horizons together with their uncertainty bounds computed from map-migration. Because we are using only a small number of particles for SVGD, the uncertainty is likely underestimated. To compensate for this underestimation, we choose the 2-standard-deviation (2σ) convention for our

horizons' upper and lower bounds, which represents a 95% confidence level based on the particles used. From the zoom-ins (Figures 3d-e) of the target horizons, we see that the left horizon has relatively smaller uncertainty compared to the right horizon, which agrees with the velocity uncertainty map in Figure 3b.

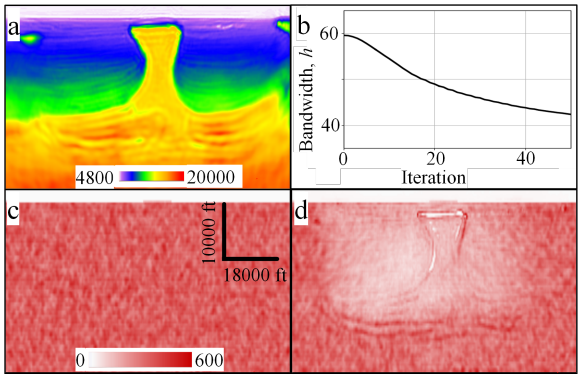

Figure 4: Test on the Herschel OBN data set. (a) Reference FWI model. (b) *h*-curve QC. (c) Standard deviation of the initial particles. (d) Standard deviation of the final particles. The white spaces in (d) are due to model clipping during FWI.

Next, we apply our uncertainty analysis to a field data example: the Herschel OBN data set from the GOM. The geology of the data area features a big salt dome rising from deep (Figure 4a). This data set contains about 900 nodes with a node spacing of 400 × 400 m and a shot spacing of 50 m. A 5-Hz production FWI model is used as the reference model for initial particle generation. We carry out our uncertainty analysis with the FWI engine also at 5 Hz, and the Time-lag cost function (Zhang et al., 2018) is used for this real data study. The SVGD algorithm is executed with 30 particles up to 50 iterations. Figure 4b shows the *h*-curve QC, from which we can see overall convergence is reached. The velocity uncertainties for initial and final particles are shown in Figures 4c-d. Once again, we observe that the uncertainty level is much reduced in the well illuminated regions, thanks to the good constraints from the OBN data.

Lastly, we look at the depth uncertainties on a couple of target horizons (Figure 5). Here, we again follow the 2σ convention in computing the horizon uncertainty bounds. The zoom-in sections of these horizons (Figures 5b-c) tell us that the depth uncertainty of the shallow horizon is significantly smaller than that of the deeper horizon, which is in line with our intuition as the shallow event is better illuminated by data and bears a simpler overburden sediment structure. Such depth uncertainty analysis is straightforward and relatively quick once we have obtained the final particles, making it easily applicable to as many target horizons as we like.

Conclusions

We presented a plausible and practical method for FWI uncertainty analysis based on the Stein variational gradient descent algorithm. The validity and effectiveness of this method were demonstrated by both synthetic and field data examples. This method provides us an affordable pathway to obtain meaningful uncertainty information on today's FWI results. We believe these uncertainty results could play valuable roles in assisting E&P activities such as accessing reservoir levels and de-risking drilling operations.

Acknowledgments

We thank CGG for permission to publish this work.

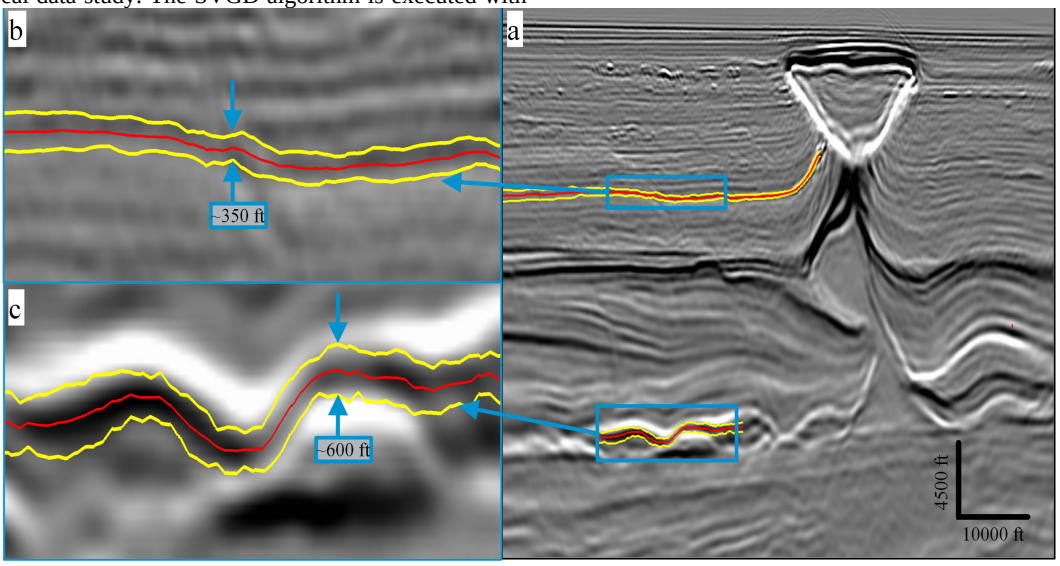

Figure 5: Depth uncertainty quantification for the Herschel OBN data set. (a) Target horizons with uncertainty bounds computed from map-migration. (b-c) Zoom-ins of the target horizons.

REFERENCES

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe, 2017, Variational inference: A review for statisticians: Journal of the American Statistical Association, **112**, no. 518, 859–877.

Brooks, S., A. Gelman, G. Jones, and X.-L. Meng, 2011, Handbook of Markov chain Monte Carlo: Chapman and Hall/CRC.

Gebraad, L., C. Boehm, and A. Fichtner, 2020, Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo: Journal of Geophysical Research: Solid Earth, **125**, no. 3, e2019JB018428.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migration, *in* J. Bednar, eds., Conference on Inverse Scattering: Theory and Application: SIAM, 206–220.

Liu, H., F. Rollins, K. Pratt, E. Da Silva, N. Mootoo, T. Yang, D. Ren, F. Gao, and J. Mei, 2023, Solving Mad Dog subsalt imaging in two decades: From WATS to OBN to elastic FWI: The Leading Edge, **42**, 398–405, doi: https://doi.org/10.1190/tle42060398.1.

Liu, Q., and D. Wang, 2016, Stein variational gradient descent: A general purpose Bayesian inference algorithm: Advances in Neural Information Processing Systems **29**, arxiv:1608.04471.

Rasmussen, C. E., and C. K. I. Williams, 2006, Gaussian Processes for Machine Learning: The MIT Press.

Ray, A., S. Kaplan, J. Washbourne, and U. Albertin, 2017, Low frequency full waveform seismic inversion within a tree based Byesian framework: Geophysical Journal International, **212**, no. 1, 522–542.

Shen, X., I. Ahmed, A. Brenders, J. Dellinger, J. Etgen, and S. Michell, 2017, Salt model building at Atlantis with full- waveform inversion: 87th Annual International Meeting, SEG, Expanded Abstracts, 1507–1511, doi: https://doi.org/10.1190/segam2017-17738630.1.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266, doi: https://doi.org/10.1190/1.1441754.

Wang, P., Z. Zhang, J. Mei, F. Lin, and R. Huang, 2019, Full-waveform inversion for salt: A coming of age: The Leading Edge, **38**, 204–213, doi: https://doi.org/10.1190/tle38030204.1.

Warner, M., and L. Guasch, 2014, Adaptive waveform inversion: Theory: 84th Annual International Meeting, SEG, Expanded Abstracts, 1089–1093, doi: https://doi.org/10.1190/segam2014-0371.1.

Zhu, H., S. Li, S. Fomel, G. Stadler, and O. Ghattas, 2016, A Bayesian approach to estimate uncertainty for full- waveform inversion using a priori information from depth migration: Geophysics, **81**, no. 5, R307–R323, doi: https://doi.org/10.1190/geo2015-0641.1.

Zhai, Y., X. Cheng, R. Bloor, D. Vigh, and Z. Xu, 2022, An efficient L-BFGS-based approach to estimate the uncertainty of full-waveform inversions: 2nd International Meeting for Applied Geoscience & Energy, SEG/AAPG, Expanded Abstracts, 722–726, doi: https://doi.org/10.1190/image2022-3745350.1.

Zhang, X., A. Lomas, M. Zhou, Y. Zheng, A. Curtis, 2023, 3-D Bayesian variational full waveform inversion: Geophysical Journal International, **234**, no. 1, 546–561.

Zhang, Z., J. Mei, F. Lin, R. Huang, and P. Wang, 2018, Correcting for salt misinterpretation with full-waveform inversion: 88th Annual International Meeting, SEG, Expanded Abstracts, 1143–1147, doi: https://doi.org/10.1190/segam2018-2997711.1.