Introduction

Full waveform inversion (FWI) is a two-way wave-equation based technique aiming at retrieving high-resolution subsurface information from seismograms. FWI is introduced as an inverse problem in which iterative optimization methods are employed to reduce the data misfit in the least-squares sense. Mathematically, the inverse problem is ill-posed with non-unique solutions. Several geophysical methods have been proposed to improve FWI approaches. We generally categorize them into deterministic (Virieux and Operto, 2009) and probabilistic approaches (Tarantola, 2005), which are either based on gradient optimizations or Bayesian inversions respectively. While deterministic FWI methods have been well developed in addressing convergence rate, cycle-skipping and multi-parameter problems, probabilistic approaches in FWI are mostly left behind due to heavy computational burdens. Only few studies have attempted to tackle the problem of uncertainty estimation in FWI (Fichtner and Trampert, 2011; Fichtner and van Leeuwen, 2015; Zhu et al., 2016), where the challenge of uncertainty estimation is closely related to the construction of the Hessian matrix.

In our research, we try to approximate the Hessian information in elastic FWI directly by using a quasi-Newton optimization method named square-root variable metric (SRVM) method (Williamson, 1975; Morf and Kailath, 1975; Hull and Tapley, 1977). SRVM has the same root as L-BFGS from the concept of variable metric (VM). SRVM is an offspring of the Davidson-Fletcher-Powell (DFP) algorithm, whereas L-BFGS is born from the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. DFP is the dual of BFGS. Although the SRVM method has been developed decades ago, it got less attention than L-BFGS. A reason being that SRVM is more complicated and expensive than L-BFGS without obvious benefits. Recently, Tarantola (2005) rediscovered SRVM for FWI and highlighted that the method allows to access the posterior covariance matrix. Here, we slightly modify the SRVM algorithm and express it as a matrix-free vector version to make it memory-affordable for large-scale problems. We incorporate the vector-version SRVM into elastic FWI and run it in a recursive manner. To validate our SRVM based elastic FWI, we test the elastic Marmousi benchmark example as a proxy for exploration setups. After convergence of the inversion, we reconstruct the posterior covariance matrix based on the history of SRVM vectors and scalars and evaluate model posterior distributions through diagonal extraction and random sampling.

Theory and Method

Following the standard Davidon-Fletcher-Powell (DFP) formula, we update the inverse Hessian $\mathbf{B} = \mathbf{H}^{-1}$ as following:

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \frac{\mathbf{B}_k \Delta\mathbf{g}_k \Delta\mathbf{g}_k^T \mathbf{B}_k}{\Delta\mathbf{g}_k^T \mathbf{B}_k \Delta\mathbf{g}_k} + \frac{\Delta\mathbf{m}_k \Delta\mathbf{m}_k^T}{\Delta\mathbf{g}_k^T \Delta\mathbf{m}_k}, \tag{1}$$

where $\Delta\mathbf{g}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$. We repeat the updating procedure in Eq. (1) until the minimization process converges. To ensure that the $\mathbf{B}$ is always positive, Williamson (1975) observed that Eq. (1) could be expressed in a square-root form:

$$\mathbf{S}_{k+1}\mathbf{S}_{k+1}^T = \mathbf{S}_k \left(\mathbf{I} - (1/P_k)\mathbf{S}_k^T \mathbf{y}_k \mathbf{y}_k^T \mathbf{S}_k\right)\mathbf{S}_k^T, \tag{2}$$

where $\mathbf{B}_{k+1} = \mathbf{S}_{k+1}\mathbf{S}_{k+1}^T$, $\mathbf{B}_k = \mathbf{S}_k\mathbf{S}_k^T$, $\mathbf{y}_k = \mu_k \mathbf{g}_k + \Delta\mathbf{g}_k$, $P_k = \mathbf{y}_k^T \mathbf{B}_k \Delta\mathbf{g}_k^T$. , with μ_k being the one-dimension search step at the k th iteration and $\Delta\mathbf{g}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$. For simplification, we also rewrite the term within brackets in a square-root form as

$$\mathbf{A}_k \mathbf{A}_k^T = \mathbf{I} - (1/P_k)\mathbf{S}_k^T \mathbf{y}_k \mathbf{y}_k^T \mathbf{S}_k, \tag{3}$$

where $\mathbf{A}_k = \mathbf{I} - (\upsilon_k / P_k)\mathbf{S}_k^T \mathbf{y}_k \mathbf{y}_k^T \mathbf{S}_k$. The constant υ_k is determined as

$$\upsilon_k = \frac{1 \pm (1 - Q_k/P_k)^{1/2}}{Q_k/P_k}, \tag{4}$$

where $Q_k = \mathbf{y}_k^T \mathbf{S}_k \mathbf{S}_k^T \mathbf{y}_k$, such that the relationship in Eq. (3) holds. In Eq. (4), we need to choose the minus sign to prevent υ_k from being singular at $Q/P = 0$; we also observe that if $Q/P > 1$, υ_k is

imaginary, resulting in a non-real matrix $\mathbf{S}$. Hull and Tapley (1977) suggest using $\upsilon_k = 1$ for $Q/P > 1$, because when $-\infty \le Q/P \le 1$, $0 \le \eta \le 1$. Concerning the workflow of SRVM, we update the square root of the inverse Hessian as following

$$\mathbf{S}_{k+1} = \mathbf{S}_k \left(\mathbf{I} - (\upsilon_k / P_k) \mathbf{S}_k^T \mathbf{y}_k \mathbf{y}_k^T \mathbf{S}_k \right). \tag{5}$$

Note that Eq. (5) is still in a matrix form. Although we can define $\mathbf{S}$ by an analytic Cholesky decomposition, the memory requirement of $\mathbf{S}$ is still at the half of $\mathbf{B}$ in Eq. (2), non-storable as well.

In the following, we rewrite the SRVM algorithm into a vector form to make the memory requirement affordable (Luo, 2012):

$$\begin{aligned}
&\Delta \mathbf{g}_k = \mathbf{g}_{k+1} - \mathbf{g}_k \\
&\mathbf{y}_k = \mu_k \mathbf{g}_k + \Delta \mathbf{g}_k \\
&\mathbf{w}_k = \mathbf{S}_k^T \mathbf{y}_k \\
&\boldsymbol{\beta}_k = \mathbf{S}_k^T \Delta \mathbf{g}_k \\
&P_k = \mathbf{w}_k^T \boldsymbol{\beta}_k \\
&Q_k = \mathbf{w}_k^T \mathbf{w}_k \\
&\upsilon_k = \frac{1 - \left(1 - Q_k / P_k\right)^{1/2}}{Q_k / P_k} \\
&\mathbf{S}_{k+1} = \mathbf{S}_k \left(\mathbf{I} - \frac{\upsilon_k}{P_k} \mathbf{w}_k \mathbf{w}_k^T \right) \\
&\mathbf{B}_{k+1} = \mathbf{S}_{k+1} \mathbf{S}_{k+1}^T = \mathbf{S}_k \left(\mathbf{I} - \frac{\upsilon_k}{P_k} \mathbf{w}_k \mathbf{w}_k^T \right) \left(\mathbf{I} - \frac{\upsilon_k}{P_k} \mathbf{w}_k \mathbf{w}_k^T \right)^T \mathbf{S}_k^T
\end{aligned} \tag{6}$$

in which we only need to store one vector series $\mathbf{w}_k$ and scalar series υ_k / P_k to reconstruct $\mathbf{B}_{k+1}$ recursively. An initial guess about $\mathbf{B}_0$ is the identity matrix $\mathbf{I}$. As stated before, $\mathbf{B}$ is related to the inverse Hessian $\mathbf{H}^{-1}$.

The relationship between the posterior covariance $\mathbf{C}_M$ and the prior covariance $\mathbf{C}_m$ (Tarantola, 2005) is

$$\mathbf{C}_M = \left(\mathbf{H} + \mathbf{C}_m^{-1} \right)^{-1} = \mathbf{C}_m \left(\mathbf{C}_m^{1/2} \mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} \mathbf{C}_m^{1/2} + \mathbf{I} \right)^{-1}, \tag{7}$$

in which $\mathbf{G}\mathbf{C}_m^{1/2}$ means preconditioning $\mathbf{G}$ with $\mathbf{C}_m^{1/2}$ during inversion. We here take the diagonal term $\mathbf{I}$ as a stabilizer and approximately have $\mathbf{H}^{-1} \approx \left(\mathbf{C}_m^{1/2} \mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} \mathbf{C}_m^{1/2} + \mathbf{I} \right)^{-1}$. That is

$$\mathbf{C}_M \approx \mathbf{C}_m \mathbf{H}^{-1}. \tag{8}$$

Based on Eq. (8), we will extract the square-root diagonals of $\mathbf{C}_M$ to draw uncertainty maps, and visualize the prior and posterior distributions with the following equations

$$\begin{aligned}
&\mathbf{m}_{prior} = \mathbf{m}_0 + \mathbf{C}_m^{1/2} \mathbf{n}, \\
&\mathbf{m}_{postorior} = \widetilde{\mathbf{m}} + \mathbf{C}_M^{1/2} \mathbf{n},
\end{aligned} \tag{9}$$

in which $\mathbf{n}$ is 2D random sampler, $\mathbf{m}_0$ and $\widetilde{\mathbf{m}}$ are the prior and posterior models, respectively. We will demonstrate our methods on the 2D elastic Marmousi model, whose size is 3000 m deep and 9200 m wide.

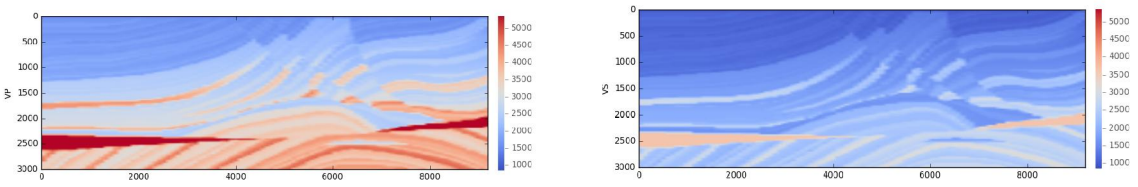

Figure 1 *True Vp (left) and Vs (right) for the elastic Marmousi model.*

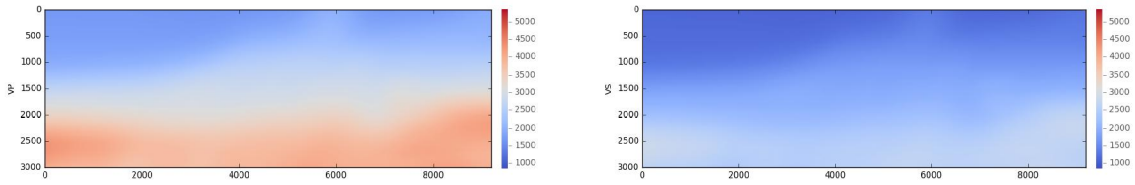

Figure 2 *Initial Vp (left) and Vs (right) in the elastic Marmousi FWI. We obtain the initial models by convolving them with Gaussian smoothers along horizontal and vertical directions.*

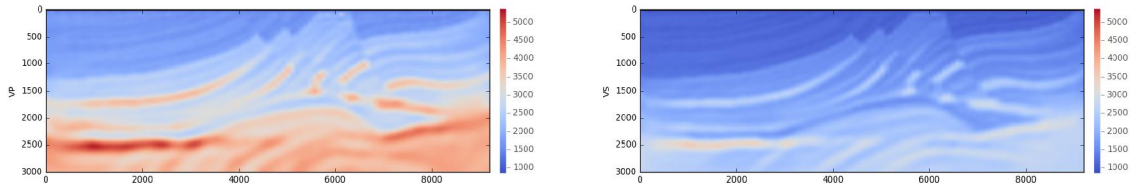

Figure 3 *Inverted Vp (left) and Vs (right) in the elastic Marmousi FWI based on SRVM after 46 iterations. The source used is a 4-Hz Ricker wavelet. The main features of the true models are recovered in the inversion results. We can notice that the shallow part is well recovered while the deep part is not that good. Perhaps we can understand this from the view of data coverage: the shallow part has more data coverages thus it is recovered better.*

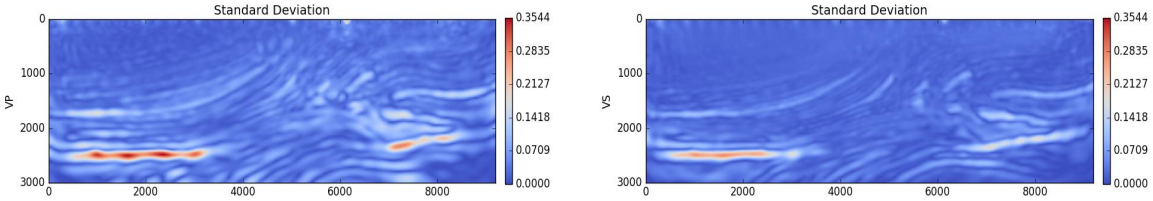

Figure 4 *Variance maps of the standard deviation of Vp (left) and Vs (right) in the elastic Marmousi inversion. Let us explain these maps from the view of data coverage. Because the shallow part has more data coverages, we has more certainties about the inverted result at this part. Note that in areas with no data coverage (for example, the very marginal areas), the uncertainty values drop to zero. This occurs because we use a subspace inversion that will not adjust model spaces where Fréchet derivatives are zero.*

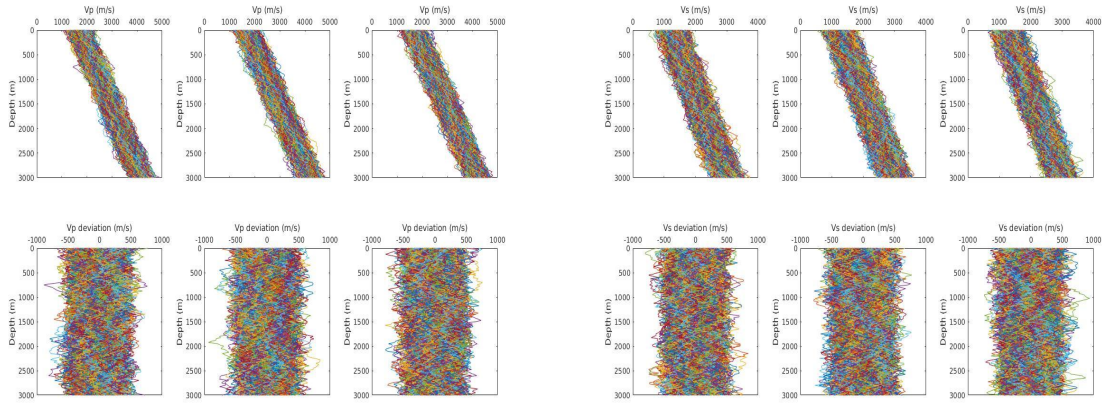

Figure 5 *Prior distributions of Vp (left) and Vs (right) in elastic Marmousi testing. Top row from left to right: Prior distributions against depth of the Vp (left) and Vs (right) models at X=3100m, 4600m, 6900m. Bottom row from left to right: Prior distributions against depth of the Vp (left) and Vs (right) deviations at X=3100m, 4600m, 6900m.*

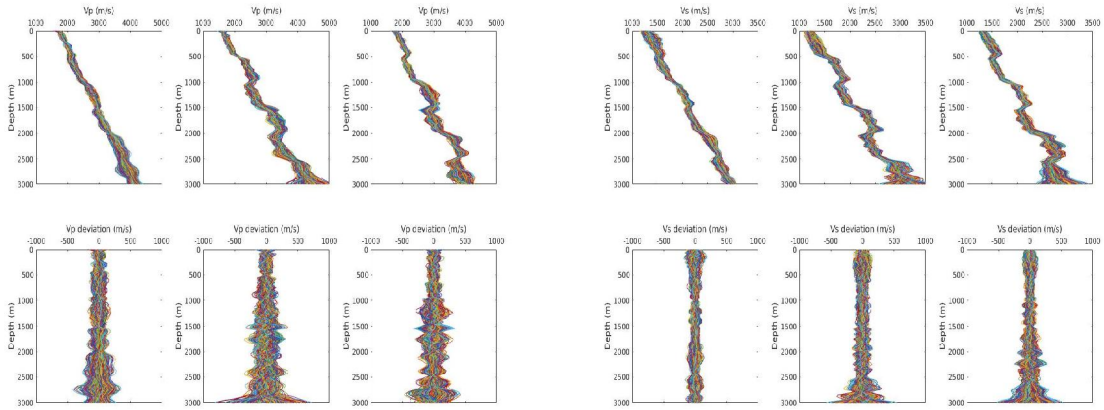

Figure 6 *Posterior distributions of Vp (left) and Vs (right) in elastic Marmousi testing. Top row from left to right: Posterior distributions against depth of the Vp (left) and Vs (right) models at X=3100m, 4600m, 6900m. Bottom row from left to right: Posterior distributions against depth of the Vp (left) and Vs (right) deviations at X=3100m, 4600m, 6900m. From the figures we can see that after inversion, even for one log well, the spatial uncertainty values become inhomogeneous. The areas with "tighter" distributions mean that we can have more certainty on them, and vice versa.*

Conclusions

We find that the vector-version SRVM provides an effective and efficient way to conduct inversions as well as quantify uncertainties in elastic FWI. To that end, variance maps can help to evaluate inversion results in a quantitative manner. Prior and posterior distributions furthermore demonstrate the information gain as well as remaining uncertainties for areas of interest.

Acknowledgements

The research reported in this publication was supported by funding from King Abdullah University of Science and Technology (KAUST). For computer time, this research used the resources of the Information Technology Division (IT) and Extreme Computing Research Center (ECRC) at KAUST.

References

Fichtner, A. and Trampert, J. [2011] Resolution analysis in full waveform inversion. *Geophysical Journal International*, **187**, 1604-1624.

Fichtner, A. and van Leeuwen, T. [2015] Resolution analysis by random probing. *Journal of Geophysical Research*, *Solid Earth*, **120**, 5549-5573.

Hull, D. G. and Tapley, B. D. [1977] Square-root variable-metric methods for minimization. *Journal of optimization theory and applications*, **21**(3), 251-259.

Luo, Y. [2012]. Seismic imaging and inversion based on spectral-element and adjoint methods. *Ph. D. thesis, Princeton*.

Morf, M. and Kailath, T. [1975] Square-root algorithms for least-squares estimation, *IEEE Transactions on Automatic Control*, **20**(4), 487-497.

Tarantola, A. [2005] Inverse problem theory and methods for model parameter estimation. *Society for Industial and Applied Mathematics*, Philadelphia, Pennsylvania, USA.

Virieux, J. and Operto, S. [2009] An overview of full-waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1-WCC26.

Williamson, W. E. [1975] Square-root algorithms for function minimization. *AIAA Journal*, **13**(1), 107-109.

Zhu, H., Li, S., Fomel, S., Stadler, G. and Ghatta, O. [2016] A Bayesian approach to estimate uncertain for full-waveform inversion using a prior information from depth migration. *Geophysics*, **81**(5), R307-R323