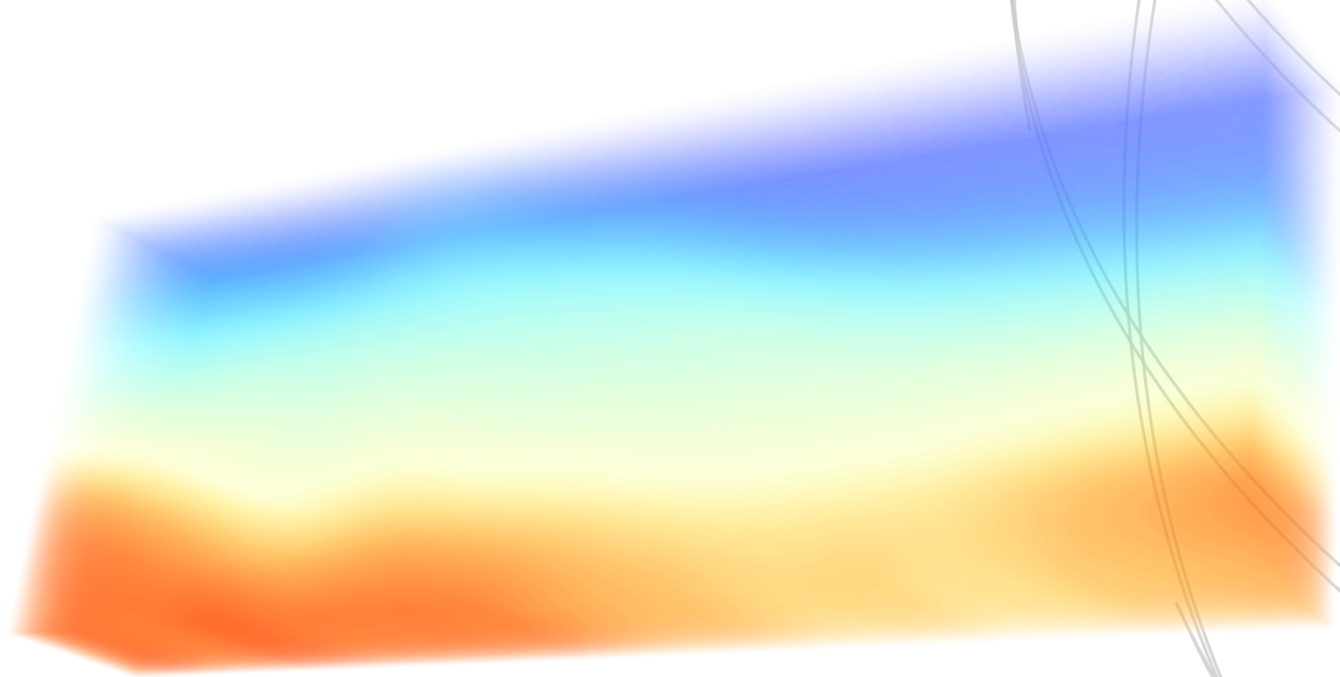
Square Root Variable Metric based null-space shuttle: a characterisation of the non-uniqueness in elastic full-waveform inversion

QIANCHENG LIU
KING ABDULLAH UNIVERSITY OF SCIENCE AND TECHNOLOGY (KAUST), SAUDI ARABIA
qiancheng.liu@kaust.edu.sa

DANIEL PETER
KING ABDULLAH UNIVERSITY OF SCIENCE AND TECHNOLOGY (KAUST), SAUDI ARABIA
daniel.peter@kaust.edu.sa



Abstract

Full-waveform inversion is an **ill-posed inverse problem**, with **non-unique solutions**. We examine its non-uniqueness by exploring the **null-space shuttle**, which can **generate an ensemble of data-fitting solutions efficiently**.

We construct this shuttle based on a quasi-Newton method, the **square-root variable-metric (SRVM)** method. This method enables a retrieval of the inverse data-misfit Hessian after the SRVM-based elastic FWI converges. Combining **SRVM** with **randomised singular value decomposition (SVD)**, we obtain the **eigenvector subspaces** of the inverse data-misfit Hessian. **The first one** among them is considered to determine the **null space** of the elastic FWI result.

Using the **SRVM-based null-space shuttle** we can modify the inverted result **a posteriori**in a highly efficient manner without corrupting data misfit. Also, because the SRVM method is embedded through elastic FWI, our method can be extended to **multi-parameter** problems. We confirm and highlight our methods with the elastic Marmousi example.

SRVM-based inverse Hessian construction

In Newton's method, the relationship between the model perturbation $\Delta\mathbf{m}$ and the gradient $\mathbf{g}$ is

$$\Delta\mathbf{m} \approx -\mu\mathbf{H}^{-1}\mathbf{g}^T = -\mu\mathbf{B}\mathbf{g}^T, \tag{1}$$

where μ is the one-dimension search step, $\mathbf{H}$ the Hessian, and $\mathbf{B}$ the inverse Hessian. The standard Davidon-Fletcher-Powell (**DFP**) method gives an iterative approach to the inverse Hessian

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \frac{\mathbf{B}_k\Delta\mathbf{g}_k\Delta\mathbf{g}_k^T\mathbf{B}_k}{\Delta\mathbf{g}_k^T\mathbf{B}_k\Delta\mathbf{g}_k} + \frac{\Delta\mathbf{m}_k\Delta\mathbf{m}_k^T}{\Delta\mathbf{g}_k^T\Delta\mathbf{m}_k}, \tag{2}$$

with $\Delta\mathbf{g}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$ and $\Delta\mathbf{m}_k = \mathbf{m}_{k+1} - \mathbf{m}_k$. **SRVM is a modified DFP algorithm** which ensures the **positive definiteness** of B.

In a **vector-version SRVM algorithm**, we only need to keep one vector $\mathbf{w}_k$ and one scalar $\frac{\upsilon_k}{P_k}$ additionally per iteration, which are **memory-affordable** even for large-scale problems. The initial guess of $\mathbf{S}_0$ is an identity matrix $\mathbf{I}$ such that all the operations are in vectors except for those regarding $\mathbf{S}_{k+1}$ and $\mathbf{B}_{k+1}$. Note that in the SRVM-based elastic FWI, we determine the search direction with the gradient via a matrix-vector product operation $\mathbf{S}_{k+1}\mathbf{S}_{k+1}^T\mathbf{g}_{k+1}$, which is finally in a vector version.

Once the **SRVM-based FWI** converges after n iterations, we have $\mathbf{B}_{n+1}$, a positive-definite approximation to retrieve the inverse Hessian over all the iterations, as follows

$$\mathbf{B}_{n+1} = \mathbf{B}_0 + \sum_{k=0}^{n}\left(\frac{\Delta\mathbf{m}_k\Delta\mathbf{m}_k^T}{\Delta\mathbf{g}_k^T\Delta\mathbf{m}_k} - \frac{\mathbf{B}_k\Delta\mathbf{g}_k\Delta\mathbf{g}_k^T\mathbf{B}_k}{\Delta\mathbf{g}_k^T\mathbf{B}_k\Delta\mathbf{g}_k}\right), \tag{3}$$

in which $\mathbf{B}_0$ is usually chosen as an identity matrix $\mathbf{I}$, an initial guess about the inverse Hessian. Also, $\mathbf{B}_0 = \mathbf{I}$ acts as a stabilizer to ensure a sufficiently stable $\mathbf{B}_k$ during iterative inversion. As the iterations proceed, $\mathbf{B}_{k+1}$ tends to the inverse Hessian. Eq. (3) indicates that the history information about the gradients and the model updates only contributes to $\mathbf{B}_{n+1} - \mathbf{B}_0$, so we retrieve the **approximated inverse Hessian** by

$$\mathbf{H}^{-1} = \mathbf{B}_{n+1} - \mathbf{I}. \tag{4}$$

In Eq. (4) $\mathbf{H}^{-1}$ is the reconstructed inverse data-misfit Hessian when elastic FWI completes.

Randomised SVD into SRVM

The SRVM method has already been a low-rank approach to reconstruct $\mathbf{H}^{-1}$. To furthermore **facilitate samplings** from $\mathbf{H}^{-1}$, we attempt to employ the **randomised SVD** to factorise $\mathbf{H}^{-1}$ into **eigenvalues and eigenvectors**. Given one symmetric matrix $\mathbf{Z} = \mathbf{H}^{-1}$ of $M \times M$ and a set of random vectors $\mathbf{X}$ of $M \times NR$, where NR is the estimated rank order of $\mathbf{Z}$.

We run such kind of random samplings on $\mathbf{Z}$ with a collection of independent $\mathbf{X}$ for $niter$ times, with $niter$ being the iteration number at convergence, because $\mathbf{Z}$ has already been in a low-rank form. Then, we merge the resulting collection into a matrix $\mathbf{Y}$. After putting $\mathbf{Y}$ into the single-channel randomised SVD, we can represent the **inverse Hessian** in a **SVD** form as

$$\mathbf{H}^{-1} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T, \tag{5}$$

in which the **eigenvectors** of $\mathbf{H}^{-1}$ are ordered within $\mathbf{V}$ and the **eigenvalues** within $\mathbf{\Lambda}$. In the next subsection, we will introduce the SRVM-based null-space shuttle from the subspace spanned from $\mathbf{V}$.

SRVM-based null-space shuttle

We consider the construction of the **conservative models** a posteriori from the sampling of the **posterior distributions**:

$$\mathbf{m}_{post} = \widetilde{\mathbf{m}} + \mathbf{C}_M^{1/2}\mathbf{n}, \tag{6}$$

with $\widetilde{\mathbf{m}}$ being the inverted model (also known as the maximum a posterior model), $\mathbf{C}_M$ being the posterior covariance matrix, $\mathbf{n}$ of zero mean and identity covariance being a 2D/3D Gaussian random sampler in the 2D/3D case. The classic **posterior covariance** can be calculated by

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

with $\mathbf{C}_d$ being the data covariance matrix, $\mathbf{C}_m$ being the prior model covariance matrix, and ε being an additional tuning factor. We can rewrite Eq. (7) as

$$\mathbf{C}_M = \left(\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G} + \varepsilon\mathbf{C}_m^{-1}\right)^{-1} = \left(\mathbf{C}_m^{1/2}\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G}\mathbf{C}_m^{1/2} + \varepsilon\mathbf{I}\right)^{-1}\mathbf{C}_m. \tag{8}$$

Incorporating $\mathbf{C}_m^{1/2}$ into $\mathbf{G}\mathbf{C}_m^{1/2}$ and $\mathbf{C}_m^{1/2}\mathbf{G}^T$ means that the prior covariance takes part in the linearized inversion of elastic FWI, yielding an approximated inverse data-misfit Hessian as $\mathbf{H}^{-1} \approx \left(\mathbf{C}_m^{1/2}\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G}\mathbf{C}_m^{1/2} + \varepsilon\mathbf{I}\right)^{-1}$. Then, we have

$$\mathbf{C}_M \approx \mathbf{H}^{-1}\mathbf{C}_m, \tag{9}$$

in which the data-misfit Hessian acts as a "big filter" upon $\mathbf{C}_m$ for $\mathbf{C}_M$. We can thus rewrite Eq. (6) as

$$\mathbf{m}_{post} = \widetilde{\mathbf{m}} + \mathbf{C}_m^{1/2}\mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{V}^T\mathbf{n}. \tag{10}$$

By taking the first largest eigenvalue λ_1 from $\mathbf{\Lambda}$ and its corresponding eigenvector $\mathbf{v}_1$ from $\mathbf{V}$, we construct the **SRVM-based conservative filtered models a posteriori** as

$$\mathbf{m}_{cons} = \widetilde{\mathbf{m}} + \mathbf{C}_m^{1/2}\mathbf{v}_1\lambda_1^{1/2}\mathbf{v}_1^T\mathbf{n}, \tag{11}$$

in which the term $\mathbf{C}_m^{1/2}\mathbf{v}_1\lambda_1^{1/2}\mathbf{v}_1^T\mathbf{n}$ is our **SRVM-based null-space shuttle**. This shuttle allows for a more significant change on $\widetilde{\mathbf{m}}$ but only with a slight change on the data misfit of $\mathbf{m}_{cons}$.

If we look further into **the term** $\mathbf{C}_m^{1/2}\mathbf{v}_1\lambda_1^{1/2}\mathbf{v}_1^T\mathbf{n}$, we will figure out that $\mathbf{v}_1^T\mathbf{n}$ is a **scalar** . We can, therefore, simplify Eq. (11) into

$$\mathbf{m}_{cons} = \widetilde{\mathbf{m}} + \mathbf{C}_m^{1/2}\mathbf{v}_1\eta\left(\mathbf{n}\right), \tag{12}$$

in which $\eta\left(\mathbf{n}\right) = \lambda_1^{1/2}\mathbf{v}_1^T\mathbf{n}$, a scalar function regarding $\mathbf{n}$, is a scaling factor to $\mathbf{C}_m^{1/2}\mathbf{v}_1$, and $\mathbf{C}_m^{1/2}\mathbf{v}_1\eta\left(\mathbf{n}\right)$ is the **null-space shuttle**. Eq. (12) suggests that we can add in an arbitrary constant times the null-space model vector but not affect the data misfit.

SRVM-based elastic FWI results

To validate our method we take the elastic Marmousi benchmark. For the optimisation method in elastic FWI, we take the SRVM method, with the L-BFGS one as a reference.

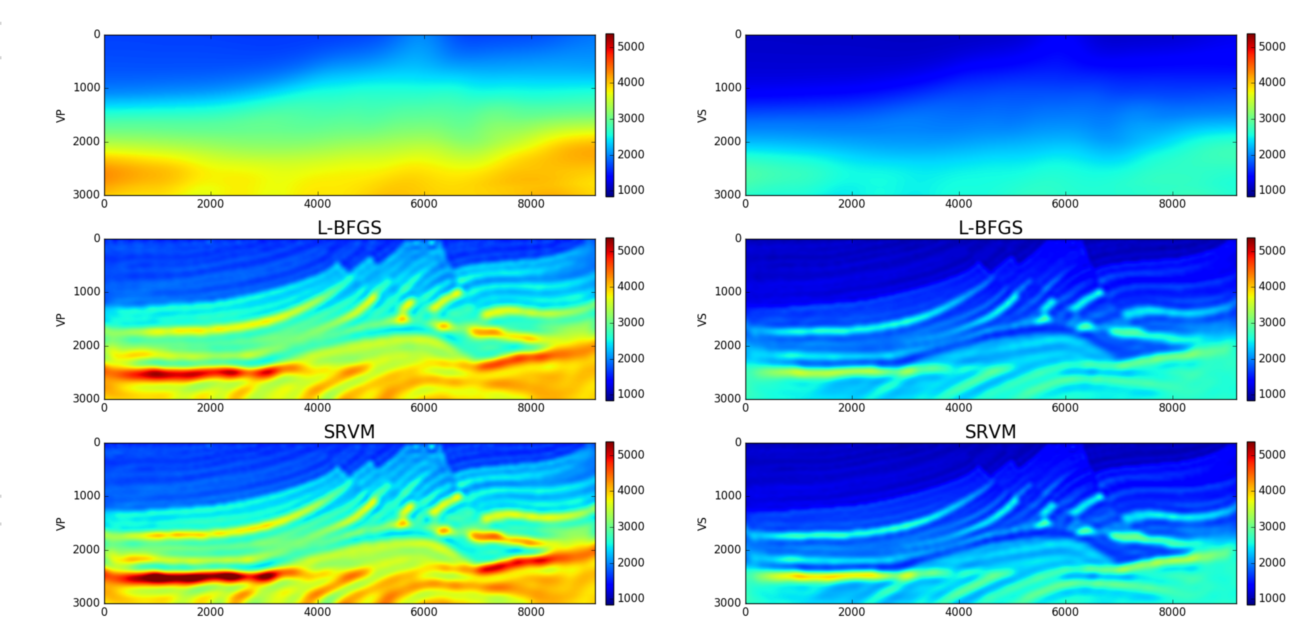


Figure 1: Top: the initial model for elastic FWI; Middle: inverted results by **L-BFGS-based elastic FWI**; Bottom: inverted results by **SRVM-based elastic FWI**.

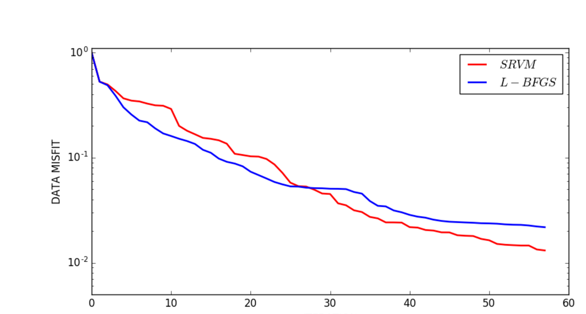


Figure 2: The convergence comparison of L-BFGS and SRVM-based elastic FWIs.

SRVM-based null-space shuttle in elastic FWI

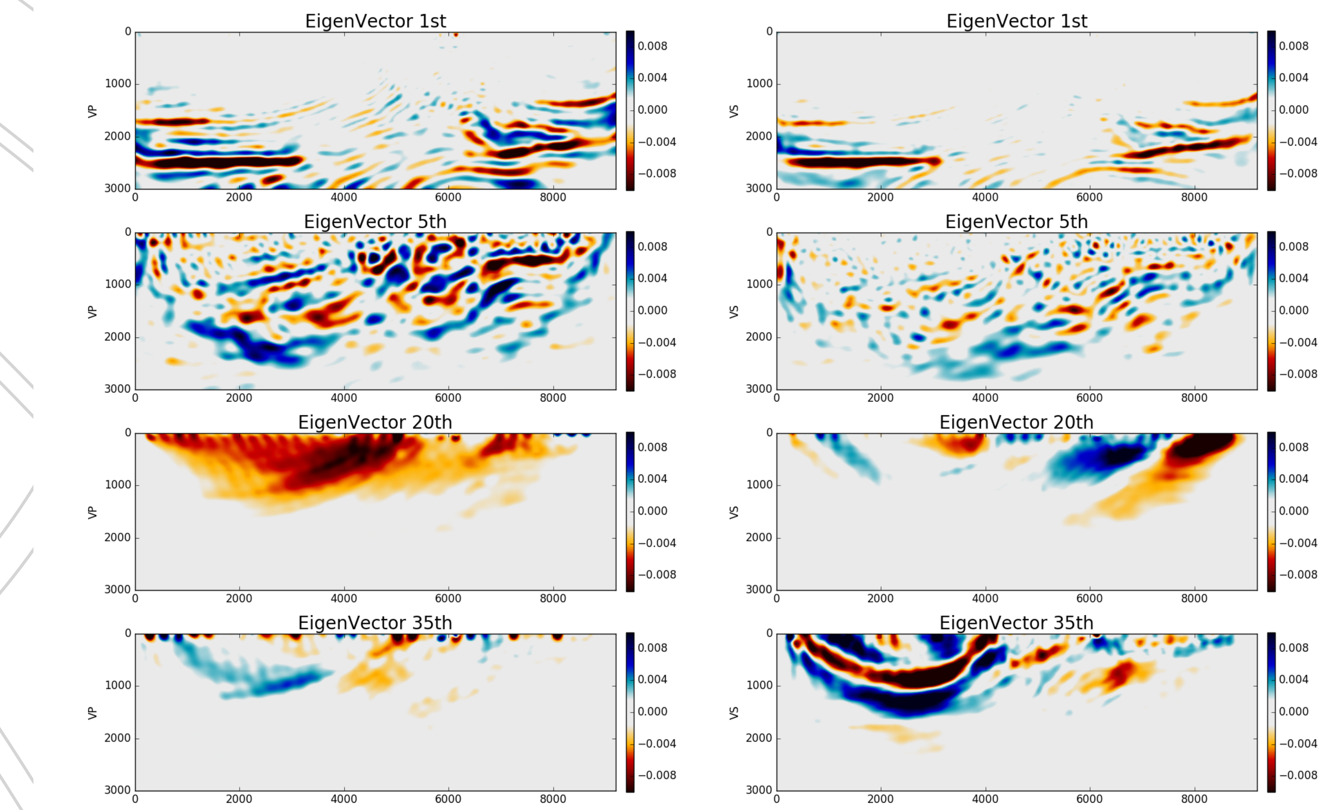


Figure 3: The $1st$, $5th$, $20th$, $35th$, eigenvectors, respectively, of the SRVM-based inverse Hessian factorised by randomised SVD. We can see as the eigen-order increases the enery distributions tend to move upwards. **The 1st eigenvector will be chosen as the null-space shuttle vector**.

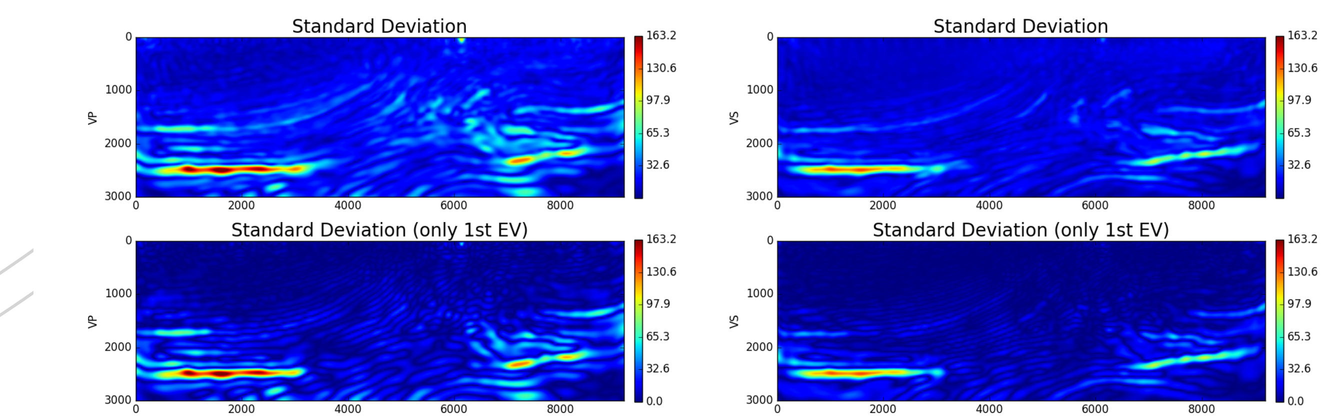


Figure 4: Top: the full uncertainty map from the posterior covariance; Bottom: the contribution to the uncertainty map from **the 1st eigenvector of the inverse Hessian**, which **dominates the uncertainty map and will be taken as the null-space shuttle vector**.

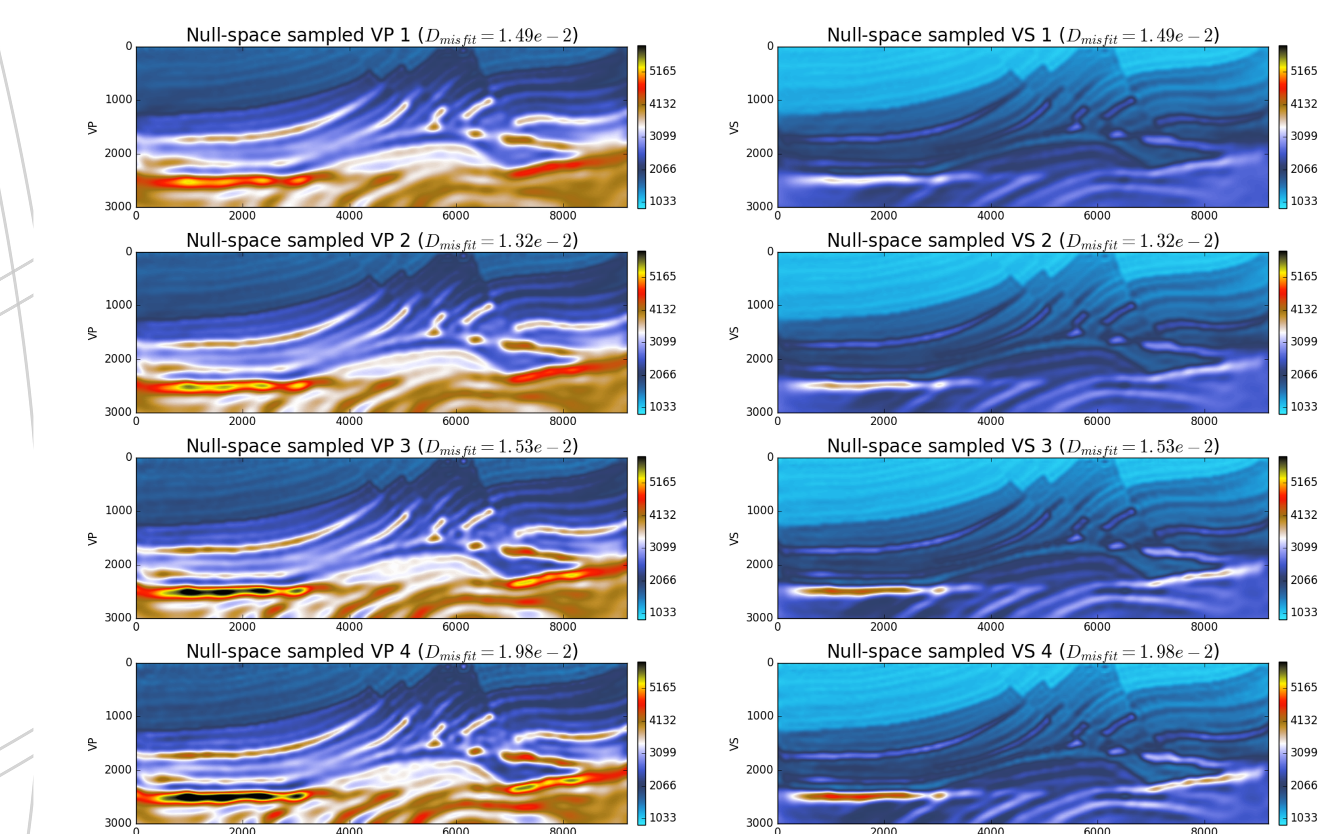


Figure 5: Conservative models by SRVM-based null-space shuttle. **The structures within the model null-space perturbing between** $[-600m/s, 600m/s]$ **lead to slightly distorted normalized data-misfits:** 1.49%, 1.32%, 1.53%, 1.98%, **with the optimal value** 1.31% **for reference**.

Conclusions & Acknowlegment

We explore the null-space shuttle in elastic FWI by using **SRVM** together with **randomised SVD**. The inverted structures associated with null-space shuttle are with the least data coverage from the data-misfit Hessian point of view. Therefore, even if we perturb these structures significantly, only slight data-misfit distorts arise. **The SRVM-based null-space shuttle runs a posterior at a highly efficient manner, generating thousands of posterior models at a cheap overhead.** The numerical applications highlight that our method is a characterisation of the non-uniqueness in elastic FWI.

This work is supported by the King Abdullah University of Science & Technology (KAUST) under award UAPN#2605-CRG4. Computational resources were provided by KAUST's Extreme Computing Research Center (ECRC). For our implementation, we used the **SPECFEM2D** package, which is hosted by the Computational Infrastructure for Geodynamics (CIG), and the **Seisflows** package by Modrak & Tromp (2016).