Multi-task Unscented Kalman Inversion for joint inversion of receiver function and surface wave dispersion
Wang Longlong*, Liu Youshan, Chen Yun, State Key Laboratory of Lithospheric Evolution, Institute of Geology and Geophysics, Chinese Academy of Sciences, Nanqiao Du, Department of Physics, University of Toronto

SUMMARY
Based on the recently developed theory of Unscented Kalman Inversion in computational mathematics, we proposed a Bayesian joint inversion framework, i.e., Multi-task Unscented Kalman Inversion (MTUKI), and apply it in the joint inversion of receiver function (RF) and surface wave dispersion (SWD). This method can share information between different observations in a derivative-free way and provide an efficient Gaussian approximation to the posterior distribution of model parameters (thickness and S-wave velocity in each layer of media). The theory and experiments show that our proposed framework demonstrates superior performance in terms of robustness, accuracy, high efficiency.

METHOD
For the joint inversion of RF and SWD, we use the Thomson-Haskell matrix method to compute the RF and SWD data [Herrmann, 2013]. To get a better understanding of this joint inversion framework, we follow the work of Daniel Z. Huang and introduce the UKI from Gaussian Approximation Algorithm (GAA). Then we show how to employ unscented transformation to well approximate Gaussian expectations and covariance using the UKI algorithm. Next, we introduce the Multi-task Unscented Kalman method, which is a new joint inversion framework for geophysical data.

Gaussian Approximation Algorithm
Inverse problems can be formulated as following:

\[ y = G(\theta) + \eta, \quad (1) \]

where \( G : \mathbb{R}^N \rightarrow \mathbb{R}^N \) is a nonlinear function (Specially, Forward model with RF or SWD), and we want to recover \( \theta \in \mathbb{R}^N \) from \( y \in \mathbb{R}^N \), with the observational noise \( \eta \) assumed with a Gaussian distribution \( \mathcal{N}(0, \Sigma_\eta) \).

From bayesian viewpoint, the \( \theta \) and \( y \) are treated as random vector, and the inverse process aims to maximum a posterior (MAP) estimation for \( \theta \), which can be formulated as \( p(\theta) \) and be set as following:

\[ p(\theta|Y) \propto \exp(-\Phi(\theta))p_0(\theta), \quad (2) \]

\[ \Phi(\theta) = \frac{1}{2} \| \Sigma^{-\frac{1}{2}} (y - G(\theta)) \|^2. \quad (3) \]

Generally, The posterior distribution \( p(\theta|Y) \), where \( Y := y_1, y_2, ..., y_n \), is intractable to evaluate. So we assume it drawn from Gaussian distribution.

Gaussian Approximation Algorithm, originally in geostatistics, also known as Gaussian process regression (GPR), is a method of interpolation that maps Gaussian to Gaussian for a Gaussian process and gives insight into the Kalman methodology [Huang et al., 2022b]. At the beginning of the algorithm, we need to assume that the posterior distribution of model parameter \( p_n \sim \mathcal{N}(m_n, \Sigma_n) \) with mean \( m_n \) and covariance \( \Sigma_n \), then the algorithm proceeds with two steps. In the first step (also known as prediction step in view of Kalman Filter), \( p_{n+1} = \mathcal{N} \left( \hat{m}_{n+1}, \hat{C}_{n+1} \right) \) is also Gaussian and satisfies:

\[ \hat{m}_{n+1} = \mathbb{E}[\theta_{n+1}|Y_n] = m_n, \quad (4) \]

\[ \hat{C}_{n+1} = \text{Cov}[\theta_{n+1}|Y_n] = C_n + \Sigma_w, \quad (5) \]

where \( \Sigma_w \) are evolution error, and \( \Sigma \) denotes the updated parameters.

Then in the second step (analysis step) we can get the joint distribution of \( \{\theta_{n+1}, y_{n+1}\} | Y_n \).

\[ \mathcal{N} \left( \begin{bmatrix} \hat{m}_{n+1} \\ \hat{y}_{n+1} \end{bmatrix}, \begin{bmatrix} \hat{C}_{n+1} & \hat{C}_{np} \\ \hat{C}_{np}^T & \hat{C}_{pp} \end{bmatrix} \right), \quad (6) \]

where

\[ \hat{y}_{n+1} = \mathbb{E}[y_{n+1}|Y_n] = \mathbb{E}[G(\theta_{n+1})|Y_n], \quad (7) \]

\[ \hat{C}_{np} = \text{Cov}[\theta_{n+1}, y_{n+1}|Y_n] = \text{Cov}[\theta_{n+1}, G(\theta_{n+1})|Y_n], \quad (8) \]

\[ \hat{C}_{pp} = \text{Cov}[y_{n+1}|Y_n] = \text{Cov}[G(\theta_{n+1})|Y_n] + \Sigma_v. \quad (9) \]

Computing the conditional distribution of joint distribution [Bishop, 2006] to find \( \theta_{n+1}|Y_{n+1} : \)

\[ m_{n+1} = \hat{m}_{n+1} + \hat{C}_{np}(\hat{C}_{pp})^{-1}(y_{n+1} - \hat{y}_{n+1}), \quad (10) \]

\[ C_{n+1} = \hat{C}_{n+1} - \hat{C}_{np}(\hat{C}_{pp})^{-1}\hat{C}_{np}^T. \quad (11) \]

We can see that GAA not only provide the expected value \( m_{n+1} \) for the given observation \( y_n \) but also offer the uncertainty through the variance \( C_{n+1} \). When assuming all observations \( y_n \) are identical to \( y \) and cycle from Equations (4) to Equations (11), the Kalman inversion will be established.

UKI
The vital operation performed in the Kalman method is how to reflect the statistical properties of the system states, i.e., approximates the integrals in Equations (11). The UKI method approximates the integrals by deterministic quadrature rules, which are called Unscented Transform [Julier, 1996]. To evaluate the process, \( 2N_\theta + 1 \) sigma points have been selected deterministic according to the posterior distribution \( \mathcal{N}(m, C) : \)

\[ \theta^0 = m, \theta^j = m + c_j|\sqrt{C}|, \theta^j+N_\theta = m - c_j|\sqrt{C}|, 1 \leq j \leq N_\theta, \quad (12) \]
Multi-task Unscented Kalman Inversion for joint inversion of receiver function and surface wave dispersion

where $\sqrt{C}_{ij}$ is the jth column of the Cholesky factor of C. The quadrature rule approximates the mean and covariance of the transformed variable $\hat{\eta}(\theta)$ as follows,

$$E[G_i(\theta)] \approx G_i(\theta^0),$$

$$\text{Cov}[G_i(\theta), G_j(\theta)] \approx 2n_w \sum_{j=1}^{2n_w} W_j \left( G_i(\theta^j) - E[G_i(\theta)] \right) \left( G_j(\theta^j) - E[G_j(\theta)] \right)^T$$

The parameters are chosen as $\eta$ and $\eta_{mod}$. The generalized $y$ is set as:

$$y = [y_{RF}, \ldots, y_{RF}, \alpha * y_{SWD}, \ldots, \alpha * y_{SWD}^T]^T,$$

where $\alpha$ is a weight factor selected according to the relatively data quality.

We also set thickness and S-wave velocity value in each layer are both variables so that their posterior distributions can be adjusted with Equation $\text{10}$ and $\text{11}$. In the joint inversion of RF and SWD, the initial posterior mean of model parameter is set as $\theta = [\theta_{\text{RF}}, \theta_{\text{thickness}}]^T$.

Although the Kalman Inversion does not depend on a objective function, we use the squared error cost function to depict the convergence of the model parameter vector $\theta$:

$$\Phi(\theta) = \frac{1}{2} \| \Sigma^{-\frac{1}{2}} (y - G(\theta)) \|^2.$$

where $\eta = [\alpha * y_{RF}, \eta_{mod}]^T$. The objective function is defined by the linear combination of misfits of the weighted receiver functions $\Phi_{RF}$ and the $\Phi_{SWD}$, thus takes the form:

$$\Phi(\theta) = \alpha \Phi_{RF} + \Phi_{SWD}.$$

EXAMPLE

In this section, we present several tests for the MTUKI with synthetic and real data of RF and SWD. Throughout all applications, we focus mainly on the MTUKI, some comparisons with the Reversible Jump Markov Chain Monte Carlo (RJMCMC) and gradient method are also presented.

Synthetic data

We use a known velocity model to verify our algorithm, the model consists of 8 horizontal layers with a low S-wave velocity layer in the crust and a strong velocity increase at the Moho (Bodin et al., 2012). To initialize MTUKI, we assume that the initial model has 25 layers, and the prior thickness mean of the first three layers are all 1.5 km, those of the fourth- to seventh-layer are 2 km and those of the other layers are all 3 km, so the thickness of layer $m_{thk}$ can be set as $m_{thk} = [1.5, 1.5, 1.5, 2, 2, 2, 2, 3, 3]$. After many experiments, we find that the inversion is almost independent on the parameter $m_{sw}$, therefore we set it as $m_{sw} = [4, \ldots, 4]$ and the initial mean $m_0 = [m_{sw}, m_{thk}]^T$. Given the covariance of parameter can be adapted in the inversion process (Equation 10), we set covariance as $C_n = 0.072I$ and initialize at $\theta_0 \sim N(m_0, C_n)$. Figures 1.2 show the inversion process of this 8-layers model. In Figure 1, it can be seen that the algorithm can adaptively adjust the model parameter (i.e., velocity and thickness) and fits the observations well (Figure 1a,e). The convergence curve is shown in Figure 2, while Figure 2b shows the inverted Vs profile. Compared with the RJMCMC (Figure 2c), it can be demonstrated that our method can better recover the model in much less time.

Field case

Our algorithm is further tested using a real data set from a KIGAM station at Seoul National University. This data set is available from the web page http://www.eas.slu.edu/eqc/eqccps/TUTORIAL/STRUCT/index.html. In this data set, the dispersion measurements and receiver functions at station KIGAM have been inverted and performed well (Herrmann 2013). These RFs all have a Gaussian factor=2.5 and SWD are the same with periods from 5-90 s. The initial model thickness parameters are set to follow Figure 3a except for more layers adding in the sedimentary layer.

The joint inversion results for seismic station KIGAM are summarized in Figure [3]. Figure [3] shows that the inversion iterates only 10 times and it can simultaneously adjust the distribution of thickness and S-wave velocity at each iteration. The results obtained by MTUKI (red dashed line) and Computer Program in Seismology (CPS), which are plotted by green line, are shown in Figure 3b. The probability density of the posterior distribution is plotted as the base map. As the figure shows, the two methods have similar results. Compared with panels (c) and (d), we can find that our method can simultaneously recover the shallow sedimentary structure and deep structure (up to 140 km).

CONCLUSION

The purpose of the current study was to introduce MTUKI as an effective inversion framework that can perform joint inversion of RF and SWD. In synthetic and real dataset tests, this inversion framework demonstrates the outstanding ability for at least three reasons: a) It provide good flexibility when modeling multimodal geophysical data by using a derivative-free way; b) It can effectively obtain uncertainty of the solution by Gaussian approximation; c) It generally requires 10 iterations to obtain an “optimal” result.
Multi-task Unscented Kalman Inversion for joint inversion of receiver function and surface wave dispersion

Figure 1: The inversion of the Multi-task Unscented Kalman Inversion in an 8-layers model. (a) The Posterior probability distribution for the S-wave velocity at each layer, the grey dotted line denotes the initial model, the blue dashed line denotes the real model, while the red dashed line denotes the inverted model. (b) Simulated RF data with the Gaussian random noise and the inverted result. (c) Simulated SWD data with Gaussian random noise and the inverted result.

Figure 2: Comparison of inverted results obtained by the MTUKI and the RJMCMC methods in an 8-layers model. (a) Convergence curve for the MTUKI. (b) The inverted model was obtained by MTUKI with a CPU time of 6.01s. (c) The inverted model solved by the RJMCMC (Bayhunter) arise from 196591 models from 7 chains with a CPU time of 6866.91s.
Multi-task Unscented Kalman Inversion for joint inversion of receiver function and surface wave dispersion

Figure 3: Comparison results obtained by the MTUKI and CPS in a real data at station KIGAM. (a) Evolution of the inverted models, the final model is denoted by the black line. (b) The posterior probability for S-wave velocity. (c) Receiver functions of the real data (grey dashed line) and predicted data obtained by the MTUKI (red line) and CPS (green line), respectively. (d) Fundamental mode Rayleigh wave phase velocity dispersion of real data (grey line) and predicted data obtained by the MTUKI (red line) and CPS (green line), respectively.
Multi-task Unscented Kalman Inversion for joint inversion of receiver function and surface wave dispersion

REFERENCES

Beardsmore, G., H. Durrant-Whyte, L. McCalman, S. O’Callaghan, and A. Reid, 2016, A Bayesian inference tool for geophysical joint inversions: ASEG Extended Abstracts, 2016, 1–10.

Bishop, C. M., 2006, Pattern recognition and machine learning: Springer. Information science and statistics.

Bodin, T., M. Sambridge, H. Tkalčič, P. Arroucau, K. Gallagher, and N. Rawlinson, 2012, Transdimensional inversion of receiver functions and surface wave dispersion: TRANSDIMENSIONAL INVERSION OF RF AND SWD: Journal of Geophysical Research: Solid Earth, 117.

Bonilla, E. V., K. Chai, and C. Williams, 2007, Multi-task Gaussian Process Prediction: Presented at the Advances in Neural Information Processing Systems.

Cao, S., and D. Z. Huang, 2021, Bayesian Calibration for Large-Scale Fluid Structure Interaction Problems Under Embedded/Immersed Boundary Framework: arXiv print arXiv:2105.09497.

Herrmann, R., 2013, Computer programs in seismology: An evolving tool for instruction and research.

Huang, D., J. Huang, S. Reich, and A. Stuart, 2022a, Efficient Derivative-free Bayesian Inference for Large-Scale Inverse Problems.

Huang, D. Z., and J. Huang, 2021, Unscented Kalman Inversion: Efficient Gaussian Approximation to the Posterior Distribution: arXiv print arXiv:2103.00277.

Huang, D. Z., T. Schneider, and A. M. Stuart, 2022b, Iterated Kalman Methodology For Inverse Problems: arXiv print arXiv:2102.01580.

Julier, S., 1996, A General Method for Approximating Nonlinear Transformations of Probability Distributions: 27.