Karlsruhe Institute of Technology

wave phenomena
analysis and numerics

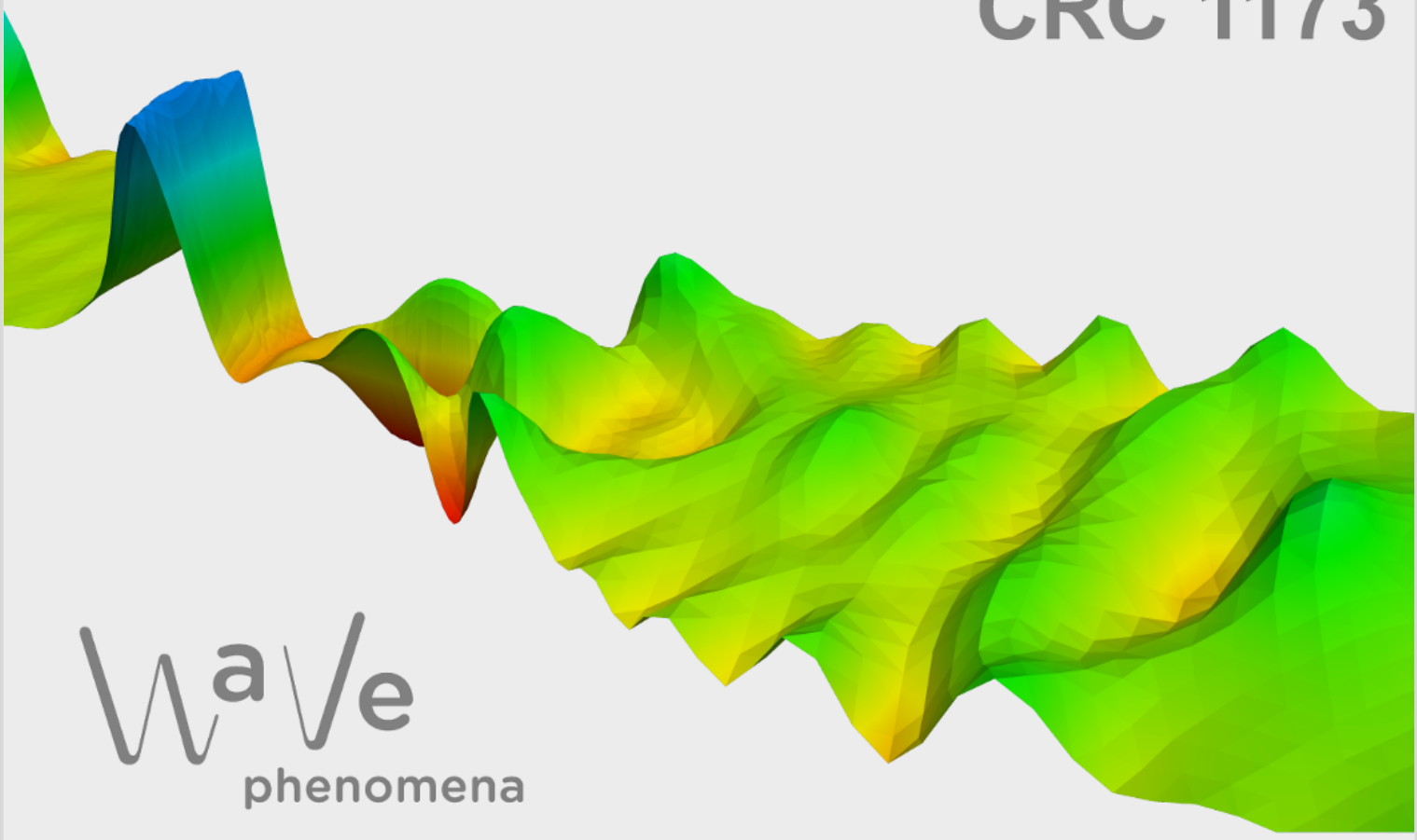
Multiparameter viscoelastic full waveform inversion of shallow seismic surface wave with a preconditioned truncated-Newton method

Lingli Gao, Yudi Pan, Andreas Rieder, Thomas Bohlen

CRC Preprint 2021/15, April 2021

KARLSRUHE INSTITUTE OF TECHNOLOGY

CRC 1173



Participating universities

Universität Stuttgart

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Funded by

ISSN 2365-662X

submitted to *Geophys. J. Int.*

Multiparameter viscoelastic full waveform inversion of shallow seismic surface waves with a preconditioned truncated-Newton method

Lingli Gao[1,2], Yudi Pan[2], Andreas Rieder[1], and Thomas Bohlen[2]

[1] *Institute for Applied and Numerical Mathematics, Karlsruhe Institute of Technology (KIT), Karlsruhe* 76131*, Germany*

[2] *Geophysical Institute, Karlsruhe Institute of Technology (KIT), Karlsruhe* 76187*, Germany*

SUMMARY

2D full waveform inversion (FWI) of shallow seismic Rayleigh waves has become a powerful method for reconstructing viscoelastic multiparameter models of shallow subsurface with high resolution. The multiparameter reconstruction in FWI is challenging due to the potential presence of crosstalk between different parameters and the unbalanced sensitivity of Rayleigh-wave data with respect to different parameter classes. Accounting for the inverse Hessian using truncated Newton methods based on second-order adjoint methods provides as an effective tool to mitigate crosstalk caused by the coupling between different parameters.

In this study, we apply a preconditioned truncated Newton method (PTN) to shallow-seismic FWI to simultaneously invert for multiparameters near-surface models (P- and S-wave velocities, attenuation of P and S waves, and density). We firstly investigate scattered wavefields caused by these parameters to evaluate the coupling between them. Then we investigate the performance of PTN on shallow-seismic FWI of Rayleigh wave for reconstructing all five parameters simultaneously. The application to spatially correlated and uncorrelated models demonstrate that PTN helps to mitigate the crosstalk and improves the resolution of the multiparameter reconstructions, especially for the weak pa-

rameters with small sensitivity such as attenuation and density parameters. The comparison with the classical preconditioned conjugate gradient method highlights the improved performance of PTN and thus the benefit of accounting for the information included in the Hessian.

Key words: Inverse theory; Waveform inversion; Surface wave

1 INTRODUCTION

The reconstruction of shallow-seismic models plays an important role in lithological and geotechnical site investigation to characterise the composition and stability of the sediments. Shallow-seismic wavefields are dominated by Rayleigh waves, which have relatively high signal-to-noise ratio compared to body waves. The inversion of Rayleigh waves is attractive due to their high sensitivity to S-wave velocity. Dispersion-based methods (Xia et al. 1999; Socco et al. 2010; Pan et al. 2016a) are widely used for subsurface imaging. However, they might fail when strong lateral heterogeneities exist (Pan et al. 2019). Secondly, they are limited by the uncertainty in the correct estimation and identification of multi-modal dispersion curves (Gao et al. 2014, 2016).

Full-waveform inversion (FWI) has the potential to estimate high-resolution subsurface models by minimising the differences (or the residuals) between recorded and synthetic seismic seismographs. The implementation are described in the time domain (Tarantola 1986), the frequency domain (Pratt 1999; Brossier et al. 2009), or the Laplace-Fourier domain (Shin & Ho Cha 2009). Successful applications of FWI are reported across different scales (Fichtner et al. 2008; Métivier et al. 2013; Warner & Guasch 2016; Malinowski et al. 2011).

Nowadays, 2D FWI at a near-surface scale has become a novel way to reconstruct shallow subsurface models (Romdhane et al. 2011; Tran et al. 2013; Groos et al. 2014; Pan et al. 2016b, 2019; Köhn et al. 2019; Xing & Mazzotti 2019). However, the previous studies neglect the effects of attenuation or simply implement a passive-viscoelastic FWI approach in which a fixed prior estimation of the attenuation model is used in the forward solver to account for the viscoelastic effects. Pure elastic and passive-viscoelastic FWI approaches are generally valid when the attenuation is weak and the attenuation model is laterally homogeneous. However, strong spatial variation of strong attenuation may exist in shallow subsurface due to its high level of heterogeneity. In this situation, simply neglecting the viscoelastic effect might deteriorate the reconstruction of S-wave velocity (Groos et al. 2014; Gao et al. 2020).

One of the main trends in FWI is to account for more realistic parameters, such as the velocity, attenuation, density and anisotropic parameters. However, multiparameter FWI is complicated and

challenging due to the strong interparameter crosstalk effect in the conventional gradient-based FWI approach (Virieux & Operto 2009; Operto et al. 2013). Suitable parametrization of the subsurface and appropriate hierarchical approach in FWI are helpful to mitigate parameter trade-offs (Köhn et al. 2012; Prieux et al. 2013; Kamei & Pratt 2013; Pan et al. 2018). Another strategy to mitigate the trade-offs between parameters is to account for the Hessian operator (Pratt et al. 1998; Operto et al. 2013). However, the calculation of Hessian and its inverse requires expensive storing memory and is unaffordable in a large-scale problem. Métivier et al. (2013, 2015) apply the truncated Newton method to the 2-D acoustic FWI problem in the frequency domain. It only requires working in matrix-free formalism and computing Hessian-vector products based on second-order adjoint state methods. The Hessian-free Newton method algorithm has two loops: an outer loop for Newton update and an inner loop to search for the Newton direction by solving Newton's equation. Yang et al. (2018) implement the truncated Newton method to visco-acoustic multiparameter FWI. They demonstrate that considering the Hessian into the inversion can improve the multiparameter reconstruction based on a realistic 2D synthetic case.

In this paper, we present a time-domain viscoelastic shallow-seismic FWI by using a preconditioned truncated-Newton (PTN) algorithm for the reconstructions of five parameters (P- and S-wave velocities, attenuation of P- and S-wave and density) simultaneously. The paper is organized as follows. In section 2, we begin with a brief introduction to the forward problem. Then we describe the inverse problem and the time-domain preconditioned truncated Newton algorithm scheme in section 3. In section 4, we analyse the scattered wavefields caused by the different parameters and show the complicated coupling effect among them. Finally, we use spatially uncorrelated and correlated synthetic models to investigate the efficiency of preconditioned truncated Newton FWI approach, and compare the performance of PTN to a preconditioned conjugate gradient method (PCG) in section 5.

2 THE FORWARD PROBLEM

In order to consider the attenuation into time-domain modelling, the generalized standard linear solid (Liu et al. 1976) is widely applied. The viscoelastic wave equation in the velocity-stress formulation is written as (Robertsson et al. 1994; Bohlen 2002):

$$\begin{aligned}
&\rho\partial_t \boldsymbol{v} = \mathrm{div}\boldsymbol{\sigma} + \boldsymbol{f},\\
&\partial_t \boldsymbol{\sigma} = C\left(\left(1+L\tau_s\right)\mu_0, \left(1+L\tau_p\right)\pi_0\right)\boldsymbol{\varepsilon}\left(\boldsymbol{v}\right) + \sum_{l=1}^{L}\boldsymbol{\eta}_l,\\
&-\tau_{\sigma,l}\partial_t \boldsymbol{\eta}_l = C\left(\tau_s\mu_0, \tau_p\pi_0\right)\boldsymbol{\varepsilon}\left(\boldsymbol{v}\right) + \boldsymbol{\eta}_l, \qquad\qquad l = 1, ..., L,
\end{aligned} \tag{1}$$

where $\boldsymbol{v}$ is the velocity vector; $\boldsymbol{\sigma}$ is the stress vector; $\boldsymbol{\eta}_l$ is the memory variable corresponding to the stress tensor $\boldsymbol{\sigma}$; $\tau_{\sigma,l}$ is the stress relaxation time of the lth Maxwell body; L is the total number of Maxwell bodies; τ_s and τ_p are attenuation levels of S and P waves, respectively. Further, $\boldsymbol{f}$ is the external force; μ_0 and π_0 denote the relaxed S- and P- wave moduli which are calculated by $\mu_0 = \frac{\rho V_S^2}{(1+\alpha\tau_s)}$ and $\pi_0 = \frac{\rho V_P^2}{(1+\alpha\tau_p)}$, where V_S and V_P are the S- and P-wave velocities, respectively. The parameter α is used to ensure that the waves travel with the model phase velocity at the reference frequency ω_0 (Bohlen 2002). It can be calculated as $\alpha = \sum_{l=1}^{L} \frac{\omega_0^2\tau_{\sigma,l}^2}{1+\omega_0^2\tau_{\sigma,l}^2}$, where ω_0 is set as the peak frequency of the source wavelet or the observed data.

The linear maps C are defined as $C(\mu_0,\pi_0)\,\boldsymbol{\varepsilon}(\boldsymbol{v}) = 2\mu_0\boldsymbol{\varepsilon}(\boldsymbol{v}) - (\pi_0 - 2\mu_0)\,\mathrm{tr}(\boldsymbol{\varepsilon}(\boldsymbol{v}))\,\boldsymbol{I}$, where $\boldsymbol{\varepsilon}(\boldsymbol{v}) = \frac{1}{2}\left[(\nabla_x \boldsymbol{v})^T + \nabla_x \boldsymbol{v}\right]$. When we consider viscoelasticity in time-domain FWI, the viscoelastic wave equation based on velocity-stress formulation is not self-adjoint. Yang et al. (2016) and Fabien-Ouellet et al. (2017) present different strategies to compute the gradient by the adjoint state methods. Here, we introduce a new transformation proposed by Zeltmann (2019) and Kirsch & Rieder (2019):

$$\begin{pmatrix} \boldsymbol{v} \\ \boldsymbol{\sigma_0} \\ \boldsymbol{\sigma_1} \\ \vdots \\ \boldsymbol{\sigma_L} \end{pmatrix} = \begin{pmatrix} \boldsymbol{v} \\ \boldsymbol{\sigma} + \sum_{l=1}^{L} \tau_{\sigma,l}\boldsymbol{\eta_l} \\ -\tau_{\sigma,l}\boldsymbol{\eta_l} \\ \vdots \\ -\tau_{\sigma,l}\boldsymbol{\eta_L} \end{pmatrix}. \tag{2}$$

Then the forward problen (equation 1) is reformulated as

$$\begin{aligned} \partial_t \boldsymbol{v} &= \frac{1}{\rho}\mathrm{div}\left(\sum_{l=0}^{L}\boldsymbol{\sigma}_l\right) + \frac{1}{\rho}\boldsymbol{f}, \\ \partial_t \boldsymbol{\sigma}_0 &= C(\mu_0,\pi_0)\,\boldsymbol{\varepsilon}(\boldsymbol{v}), \\ \partial_t \boldsymbol{\sigma}_l &= C(\tau_s\mu_0,\tau_p\pi_0)\,\boldsymbol{\varepsilon}(\boldsymbol{v}) - \frac{1}{\tau_{\sigma,l}}\boldsymbol{\sigma}_l. \end{aligned} \tag{3}$$

By considering this equivalent transformation, we can symmetrize the operators and get their self-adjoint operators (Zeltmann 2019; Kirsch & Rieder 2019). The corresponding adjoint equations are given in the appendix.

3 THE INVERSE PROBLEM

FWI estimates the subsurface parameters from the observed seismic data, which is a well-known iterative technique used for minimizing the nonlinear objective function

$$\min_{\boldsymbol{m}} J(\boldsymbol{m}) = \frac{1}{2}\|\boldsymbol{\phi}(\boldsymbol{m}) - \boldsymbol{d_{obs}}\|_2^2. \tag{4}$$

where $\boldsymbol{m}$ is the vector of model parameters (such as velocity, attenuation, density and anisotropy) of the subsurface, which belongs to the model space $\mathcal{M}$ ($\boldsymbol{m} \in \mathcal{M}$). Further, $\boldsymbol{d_{obs}}$ is the vector of observed seismic data, and $\boldsymbol{\phi}(\boldsymbol{m})$ is the vector of modelled data. In the framework of local nonlinear optimization methods, an iterative sequence $\boldsymbol{m_k}$ is built from an initial guess $\boldsymbol{m_0}$ with a descent direction $\Delta \boldsymbol{m_k}$:

$$\boldsymbol{m_k} = \boldsymbol{m_{k-1}} + \alpha_k \Delta \boldsymbol{m_k} \tag{5}$$

where α_k is the step length, which can be estimated via a line search (Nocedal & Wright 2006). Within Newton optimization framework, the search direction $\Delta \boldsymbol{m_k}$ is usually computed by the solution of the Newton linear equation:

$$\boldsymbol{H}(\boldsymbol{m})\, \Delta \boldsymbol{m_k} = -\nabla J. \tag{6}$$

where ∇J is the gradient vector (the first derivative of the misfit function) and $\boldsymbol{H}$ is the Hessian operator (the second-order derivative of the misfit function). In a multiparameter reconstruction framework, we are interested in the reconstruction of several classes of parameters m_i with $i = 1, \ldots, N$, where N is the number of parameter classes to be reconstructed. The gradient of the misfit function with respect to the parameter class m_i is given by

$$\frac{\partial J}{\partial m_i}(\boldsymbol{m}) = \frac{\partial \boldsymbol{\phi}}{\partial m_i}^{\dagger} (\boldsymbol{\phi}(\boldsymbol{m}) - \boldsymbol{d_{obs}}) \tag{7}$$

where the $\frac{\partial \boldsymbol{\phi}}{\partial m_i}$ is the partial Fréchet derivative of the full waveform forward operator $\boldsymbol{\phi}$ with respect to the model parameter m_i. The symbol † denotes the adjoint operator. From equation 7, we see that if two parameter classes have similar scattering responses, the gradient of the misfit function cannot distinguish two parameters classes, which will cause trade-off or crosstalk between them.

We also look at the expression of the Hessian matrix, the ijth block is given by

$$\boldsymbol{H}_{ij} = \nabla^2 J_{ij} = \frac{\partial \boldsymbol{\phi}}{\partial m_i}^{\dagger} \frac{\partial \boldsymbol{\phi}}{\partial m_j} + \frac{\partial^2 \boldsymbol{\phi}}{\partial m_i \partial m_j}^{\dagger} (\boldsymbol{\phi}(\boldsymbol{m}) - \boldsymbol{d_{obs}}) \,. \tag{8}$$

The Hessian matrix can be computed as the zero-lag cross-correlation of the signals scattered by perturbations of parameter classes m_i and m_j (first term on the right-hand side in equation 8) and the cross-correlation between the second-order derivative wavefield and the data residual (second term on the right side in equation 8). The off-diagonal blocks ($i \neq j$) of the Hessian matrix operator reflects the coupling between parameter classes. Here, we give an example to explain the role of Hessian through a simple synthetic example. Two anomalies (V_S and V_P) are overlapped to a homogeneous background model (first row in Figure 1). We compute the gradient (second row in Figure 1) of the misfit with respect to V_S and V_P, respectively. By repeatedly sampling the rows or columns of the Hessian-vector product with an input vector as a Dirac delta function, we can explicitly build the Hessian matrix

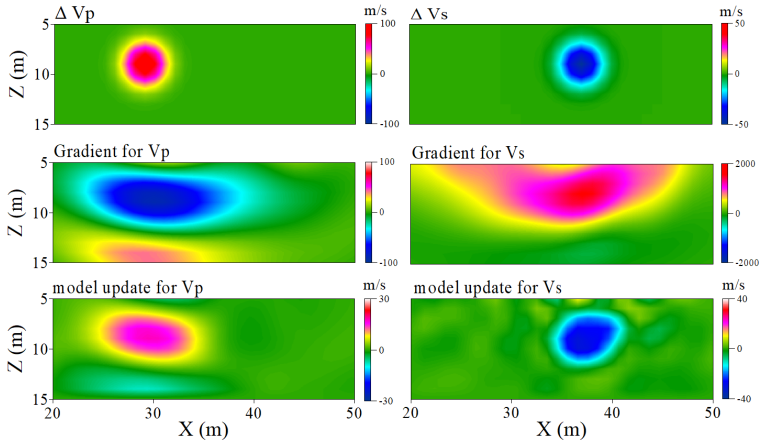


Figure 1. Example for the significant advantage for applying the inverse Hessian to the gradient. The first row represents the perturbations of V_S and V_P anomalies overlapped to a homogeneous background model ($V_P = 800$ m/s and $V_S = 400$ m/s), respectively. We calculated gradients (second row) for V_S and V_P. By applying the inverse Hessian to the gradients ($\Delta m = -H^{-1}\nabla J$), we can get the model updates for V_S and V_P (third row), respectively. The sources are located from $X = 18$ m to 48 m with an interval of 5 m. The receivers are placed on the surface from $X = 20$ m to 50 m with an interval of 1 m.

(left, Figure 2), which has a $2 * 2$ block structure. We can see that the S-wave velocity is a dominant parameter while the P-wave velocity is a secondary parameter that might be contaminated by the crosstalk produced by the S-wave anomaly (right, Figure 2). We can calculate the inverse Hessian by applying a singular value decomposition (SVD) to the Hessian and build an approximate inverse based on the truncation of the operator for a given number of singular values. By multiplying the inverse Hessian to the gradient, we can obtain the model updates which are shown in Figure 1.

As shown in the third row, after applying the inverse Hessian, the final model updates can correctly locate V_S and V_P anomalies at different positions and bring the true orders of the magnitudes. The ambiguity produced by the crosstalk in the gradients of misfit with respect to V_S and V_P is mitigated and the updated results are more focused on the actual location of the anomaly which will lead to improving spatial resolution in the final reconstruction. This indicates that the inverse of the Hessian matrix acts as a refocusing and decoupling operator which can help to mitigate the crosstalk effect. However, the explicit computation of the Hessian or its inverse is unfeasible for large-scale FWI. The truncated Newton method has been proposed to calculate the Hessian-vector product using the 2nd-order adjoint state method. In this paper, we use the adjoint state equation presented in Kirsch & Rieder (2019) to calculate the gradient and Hessian-vector product. The explicit expression of the misfit gradient with respect to parameters (ρ, V_P, τ_p, V_S, τ_s) and the Hessian-vector product are given in the appendix.

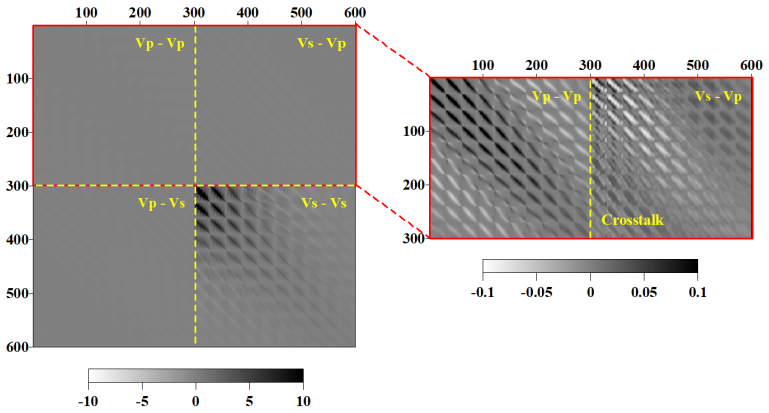


Figure 2. The $2 * 2$ block structure Hessian (left) and a zoom panel for the left and right down blocks of the whole Hessian (red rectangle, right). The diagonal elements of the diagonal blocks account for the geometrical spreading; the off-diagonal elements of the diagonal blocks represent the spatial correlations for the same physical parameter; the diagonal elements of the off-diagonal blocks represent the interparameter coupling at the same position; the off-diagonal elements of the off-diagonal blocks represent both spatial and interparameter trade-offs.

3.1 Preconditioning with parameter scaling

A suitable preconditioning is of importance to improve the convergence rate when solving the Newton equation with the linear CG method (Métivier et al. 2013, 2015; Yang et al. 2018). In multiparameter reconstruction, different parameter classes may be observed with different orders of magnitudes, which will increase the condition number of the Hessian matrix and cause problems when simultaneously updating different parameters. Yang et al. (2018) present an unity-based normalization and an additional user-defined scaling of the preconditioning to promote the weak sensitivity for specific parameters. In our five-parameter case, the scaled pseudo-Hessian preconditioning for the truncated Newton method becomes,

$$P = \begin{bmatrix} s_1 s_1 \mathrm{diag}\widetilde{H}_{11} & s_1 s_2 \mathrm{diag}\widetilde{H}_{12} & s_1 s_3 \mathrm{diag}\widetilde{H}_{13} & s_1 s_4 \mathrm{diag}\widetilde{H}_{14} & s_1 s_5 \mathrm{diag}\widetilde{H}_{15} \\ s_2 s_1 \mathrm{diag}\widetilde{H}_{21} & s_2 s_2 \mathrm{diag}\widetilde{H}_{22} & s_2 s_3 \mathrm{diag}\widetilde{H}_{23} & s_2 s_4 \mathrm{diag}\widetilde{H}_{24} & s_2 s_5 \mathrm{diag}\widetilde{H}_{25} \\ s_3 s_1 \mathrm{diag}\widetilde{H}_{31} & s_3 s_2 \mathrm{diag}\widetilde{H}_{32} & s_3 s_3 \mathrm{diag}\widetilde{H}_{33} & s_3 s_4 \mathrm{diag}\widetilde{H}_{34} & s_3 s_5 \mathrm{diag}\widetilde{H}_{35} \\ s_4 s_1 \mathrm{diag}\widetilde{H}_{41} & s_4 s_2 \mathrm{diag}\widetilde{H}_{42} & s_4 s_3 \mathrm{diag}\widetilde{H}_{43} & s_4 s_4 \mathrm{diag}\widetilde{H}_{44} & s_4 s_5 \mathrm{diag}\widetilde{H}_{45} \\ s_5 s_1 \mathrm{diag}\widetilde{H}_{51} & s_5 s_2 \mathrm{diag}\widetilde{H}_{52} & s_5 s_3 \mathrm{diag}\widetilde{H}_{53} & s_5 s_4 \mathrm{diag}\widetilde{H}_{54} & s_5 s_5 \mathrm{diag}\widetilde{H}_{55} \end{bmatrix}^{-1} . \quad (9)$$

The scaling factors $s_i, i \in \{1, 2, 3, 4, 5\}$, are associated with a priori estimate of the variance of the parameter $m_i \in \mathcal{M}$ (Yang et al. 2018; Kamei & Pratt 2013). The block entries $\widetilde{H}_{ij}$ are constructed by autocorrelation of the sensitivity kernels ($i = j$, diagonal block) or by the correlations between the sensitivity kernels corresponding to different parameters ($i \neq j$, off-diagonal block) at the same

gridpoint (Yang et al. 2018). The general formulation of the time-domain viscoelastic preconditioned truncated Newton method algorithm is summarized in Algorithm 1.

Algorithm 1: Time-domain preconditioned truncated Newton method FWI (PTN)

Input: Given starting point $\boldsymbol{m_0}$, ϵ ;

Output: Solution $\boldsymbol{m}$;

while $\frac{J(\boldsymbol{m_{k+1}})}{J(\boldsymbol{m_k})} > \epsilon$ **do**;

compute $\nabla J(\boldsymbol{m_k})$; %Gradient with equation 10

set $\boldsymbol{r_0} \leftarrow \nabla J(\boldsymbol{m_k})$;

set $\boldsymbol{y_0} \leftarrow P\boldsymbol{r_0}$; %P is calculated by equation 9

$\boldsymbol{p_0} \leftarrow -\boldsymbol{y_0}$, $k \leftarrow 0$;

while $\|H(\boldsymbol{m_k})\boldsymbol{p_k} + \nabla J(\boldsymbol{m_k})\| > \eta\|\nabla J(\boldsymbol{m_k})\|$ or $k < k_{max}$ **do**

compute $H(\boldsymbol{m_k})\boldsymbol{p_k}$; %Hessian-vector product with equation 13

$\beta_1 \leftarrow (H(\boldsymbol{m_k})\boldsymbol{p_k}, \boldsymbol{p_k})$;

if $\beta_1 < 0$ **then**

stop the inner iterations;

else

$\beta_2 \leftarrow (\boldsymbol{r_k}, \boldsymbol{y_k})$;

$\boldsymbol{x_{k+1}} \leftarrow \boldsymbol{x_k} + \frac{\beta_2}{\beta_1}\boldsymbol{p_k}$;

$\boldsymbol{r_{k+1}} \leftarrow \boldsymbol{r_k} + \frac{\beta_2}{\beta_1}H(\boldsymbol{m_k})\boldsymbol{p_k}$;

$\boldsymbol{y_{k+1}} \leftarrow P\boldsymbol{r_{k+1}}$;

$\beta_{k+1} \leftarrow \frac{\boldsymbol{r_{k+1}^T y_{k+1}}}{\boldsymbol{r_k^T y_k}}$

$\boldsymbol{p_{k+1}} \leftarrow -\boldsymbol{y_{k+1}} + \beta_{k+1}\boldsymbol{p_k}$;

end

$k \leftarrow k + 1$;

end

compute α with a line search;

$\boldsymbol{m_{k+1}} \leftarrow \boldsymbol{m_k} + \alpha\boldsymbol{x_k}$;

update the η according to Eisenstat and Walker forcing-term formula (Métivier et al. 2015);

end

4 SCATTERED WAVEFIELDS

The gradient in equation 7 is viewed as the zero-lag cross-correlation between the data residuals and the partial derivatives of wavefields. For two-parameter classes m_i and m_j, if we have $\frac{\partial \boldsymbol{\phi}}{\partial m_i} \approx \frac{\partial \boldsymbol{\phi}}{\partial m_j}$,

the model perturbation provided by gradients cannot decipher numerically between these different parameter classes. This is referred as crosstalk between parameters.

Different from body waves, the seismogram of the near-surface case is dominated by Rayleigh waves, whose particles move elliptically. For a homogeneous half-space model, at the surface and shallow depths, this motion is retrograde. At the greater depths, the particle motion becomes prograde. The transition from retrograde to prograde occurs at the depth where the horizontal displacement becomes zero, which is referred as critical depth. We display the horizontal and vertical displacements for three different frequencies in the fundamental mode of a depth-dependent background model (background model in the synthetic example), which indicates that critical depths are located at around 0.4 m in the model among the frequency range of interest. Figure 4 shows the scattered wavefields of Rayleigh wave which are computed by subtracting the simulated wavefields in the same model with and without a diffracting point perturbations of five different parameters (V_S, V_P, τ_s, τ_p, and ρ) at three different depths ($z = 0$, 0.4 and 0.8 m, red dots in Figure 4) at the location of $X = 60$ m, respectively. These scattered wavefields, also referred as radiation patterns, are good visual proxy to evaluate the term $\frac{\partial \boldsymbol{\phi}}{\partial m_i}$.

For the depth $z = 0$ m, we notice that the scattered wavefields caused by P- and S-wave perturbations are similar. With the depth increase, scattered wavefields caused by V_S and V_P can be decoupled in the backward direction but still coupled in the forward direction. It implies that we are not able to distinguish V_S and V_P anomalies if they are located above the critical depth. While it gives us a hint that we have a higher chance to distinguish them with an appropriate acquisition system if they are located at deeper depth, thanks to their different back-scattered wavefields. The scattered wavefields of the Rayleigh waves have a 180-degree phase change along the backward direction, which is related to the change of Rayleigh-wave particle motion of horizontal displacement at the depth where the critical point exists (where the particle motion changed from retrograde to prograde, Figure 3). The V_S perturbation radiates higher energies compared to the V_P perturbation at a deeper depth. The attenuation parameters are always coupled with the corresponding velocities (τ_s to V_S, and τ_p to V_P) for the same scattering direction but with a phase difference of about 90-degree (Kamei & Pratt 2013; Yang et al. 2018). The scattered wavefield of ρ perturbation appears to be more complicated. We can observe that the density perturbation radiates higher energy at a deeper part and mainly along the 'backward' direction. The scattered wavefields caused by V_S perturbation and ρ perturbation are similar in the backward direction at the deeper depth, while they both can be decoupled in the forward direction. The V_P perturbation is decoupled with ρ perturbation in both directions. Additionally, we observe that forward and backward scattered wavefields are always symmetric in V_P and τ_p, which is not the case

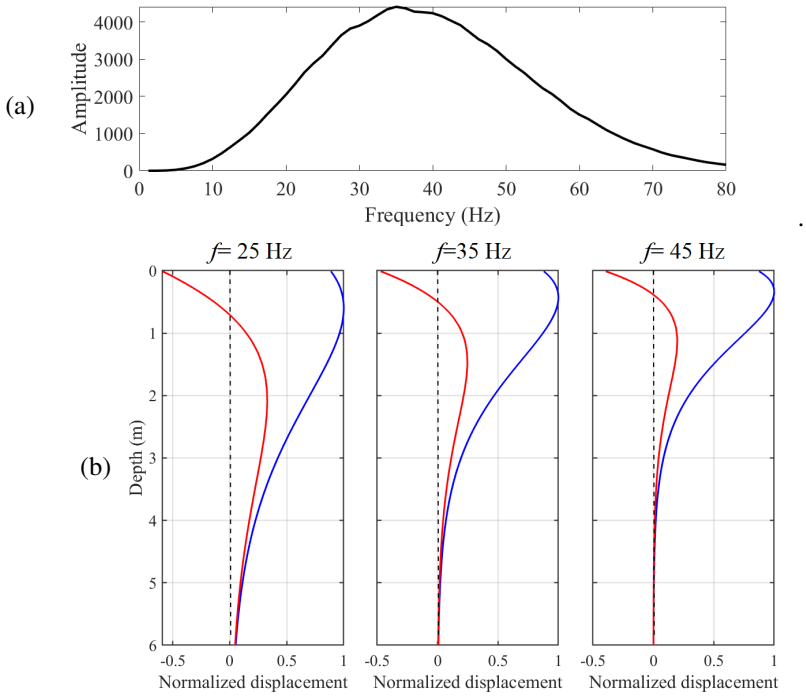


Figure 3. The critical-point depth for three frequencies ($f = 25, 35, 45Hz$). (a) The amplitude of the observed data of the depth-dependent model in the synthetic example. (b) Eigen displacements of vertical (blue line) and horizontal (red line) component calculated from Chen (1993). The vertical and horizontal components are normalized by the maximum of the vertical component, respectively. The critical depths are located around 0.4 m in this case.

in the other parameters. Here we only consider the scattered wavefields of Rayleigh wave due to the dominance of it in the shallow-seismic wavefields.

Overall, we can see that S-wave velocity perturbations can produce relatively strong contaminations into other parameters. Velocities can produce positive contamination into attenuations. And contaminations from S- and P-wave perturbations make density structures highly under- or overestimated. The coupling between different parameter classes is depth-dependent and will introduce interparameter crosstalks to the model updates.

5 SYNTHETIC EXAMPLES

In this section, we perform two synthetic examples to demonstrate the validity of our preconditioned Newton method viscoelastic FWI approach to reconstruct multi parameters of the subsurface. The inverted models using preconditioned conjugate gradient (PCG) approach are also shown for comparison, in which the preconditioner P is calculated by an approximated inverse of the Hessian matrix by using its diagonal term (Plessix & Mulder 2004). We perform a multi-stage inversion with a progres-

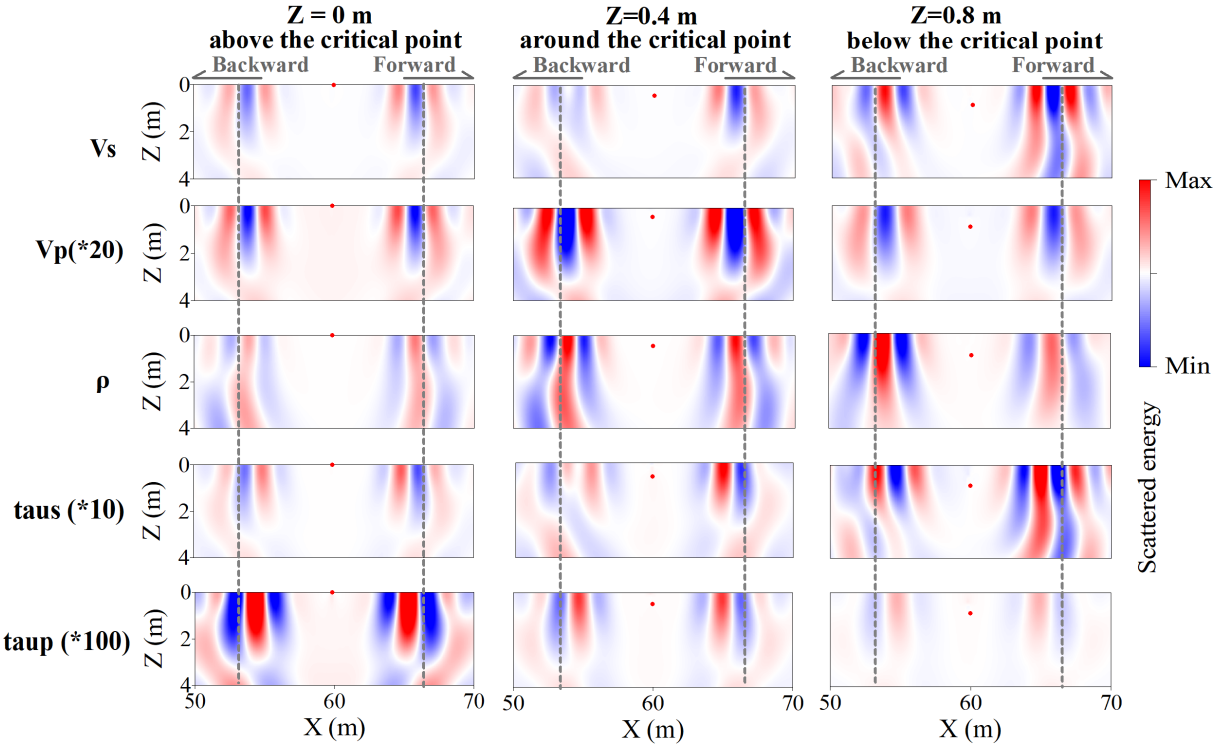


Figure 4. Scattered Rayleigh waves for parameters of V_S,V_P,τ_s,τ_p, and ρ. The perturbation points are located in depths of 0, 0.4, and 0.8 m (red solid dots). The constant number in parantheses denote the amplification factors applied to the scattered Rayleigh waves.

sive frequency band of 0-25, 0-35, 0-45, 0-60, 0-80 Hz. For simplicity, the source time function is set as known information in the synthetic examples.

5.1 A spatially uncorrelated model

We build a true model which consists of a depth-dependent 1-D background model (Figure 5). Two rectangular anomalies are superimposed on each parameter at different depths and positions. Each rectangle is 5 m wide with two different depth (1 m and 4 m, see Figure 5). Eight shots are triggered with a horizontal interval of 10 m, starting from $X = 20$ m to 90 m. The vertical-force source is generated with a delayed Ricker wavelet with a central frequency of 30 Hz. A total of 71 receivers (recording both horizontal and vertical components) are distributed along the surface with an interval of 1 m, starting from $X = 20$ m to 90 m. We use the background models as initial models. We perform a multiparameter viscoelastic FWI on this synthetic data in which five parameters are updated simultaneously during the inversion. A minimum of 10 iterations is performed at every inversion stage. The inversion will move to the next stage when the relative improvement of the misfit becomes less than 1%.

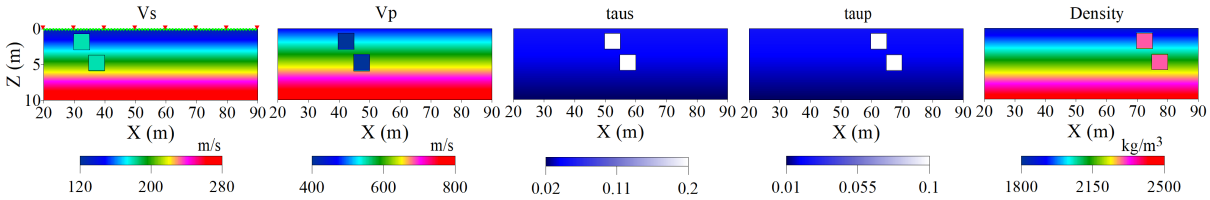


Figure 5. The true model for synthetic example on a spatially uncorrelated model. The red inverse triangles and green dots placed on the surface represent the source and receiver locations, respectively. The black solid boxes overlapped on the results represent the locations of V_S, V_P, τ_s,τ_p, ρ anomalies.

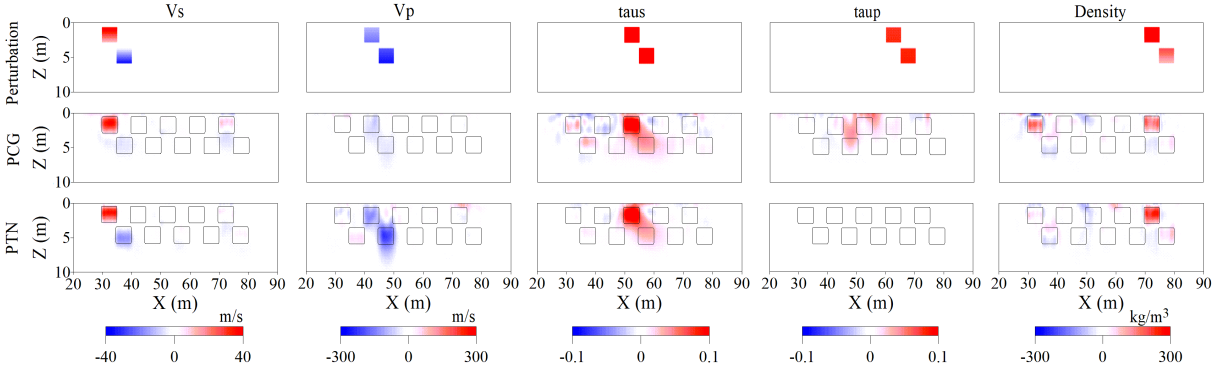


Figure 6. Multiparameter reconstruction example on a spatially uncorrelated model. Three rows represent the true perturbation models, viscoelastic FWI results with preconditioned conjugate gradient method (PCG) and preconditioned truncated Newton method (PTN), respectively. The black solid boxes overlapped on the results represent the locations of V_S, V_P, τ_s,τ_p, ρ anomalies.

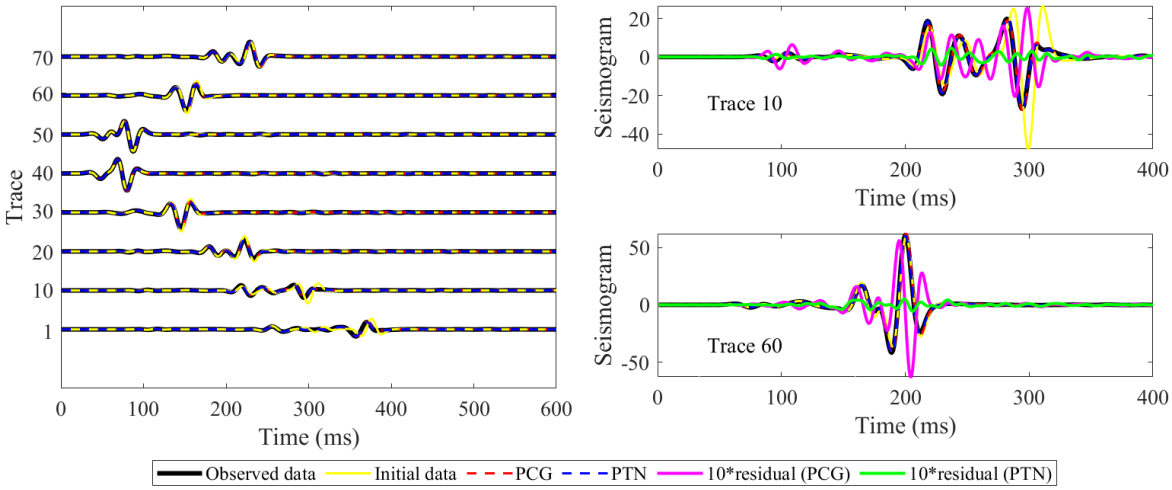


Figure 7. The comparison between the final synthetic data and the observed data. The residuals between the synthetic data and observed data is magnified by 10 times.

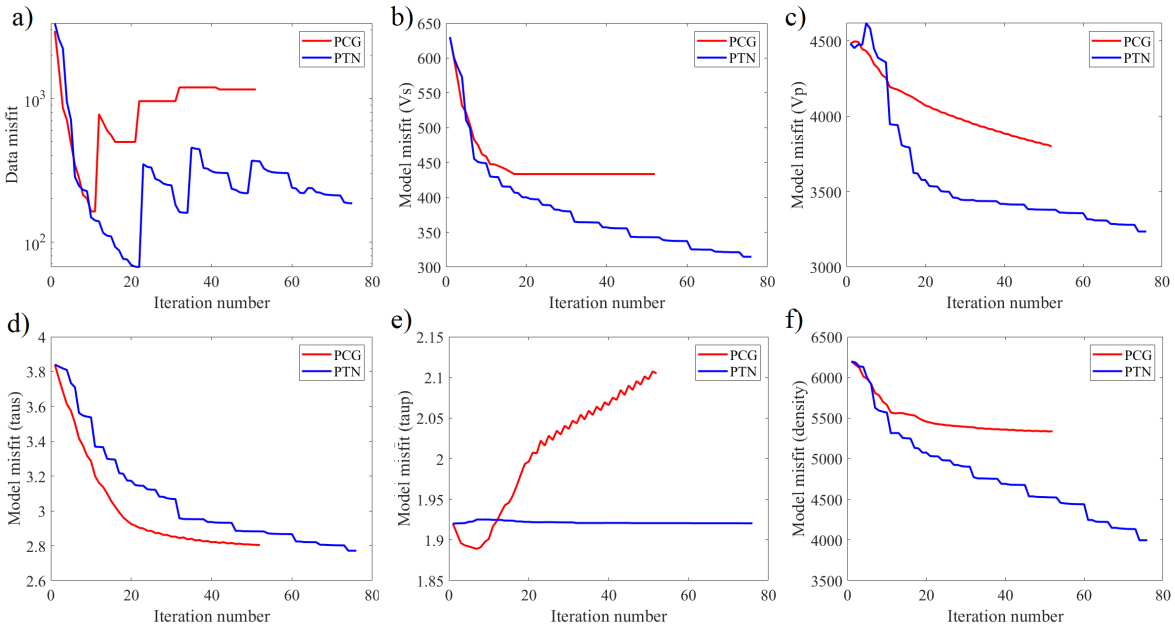


Figure 8. The comparison of the data misfit (a) and model misfits for V_S (b), V_P (c), τ_s (d),τ_p (e), ρ (f), respectively. The red and blue represent the PCG and PTN FWI, respectively.

In the PCG results (second row, Figure 6), the shallow high V_S anomaly can be reconstructed well due to the high sensitivity to S-wave velocity, while the deep low V_S anomaly is hard to be reconstructed due to limited penetrating depth of Rayleigh waves. Moreover, the final V_S model is affected by the crosstalk from the density anomalies. The low V_P perturbation can only be retrieved roughly. The reconstruction of τ_s underlies strong crosstalk from the V_S anomaly and weak crosstalk from secondary parameters like V_P and density. In the inverted τ_p model, the anomalies cannot be identified and the result is contaminated by a strong influence from the V_P anomaly. This observation is mainly caused by the low sensitivity of Rayleigh wave with respect to τ_p. The rectangle density anomaly at shallow depth can be reconstructed in the PCG result, but it suffers strong crosstalk from the V_S anomalies.

We perform multiparameter FWI using the PTN approach with the same setup. The reconstructed multiparameter models (third row, Figure 6) show a nice agreement with the true perturbation models, which is more accurate than the PCG results. It can be seen that the deeper part of V_S is better reconstructed compared to the PCG results and the crosstalk projected from the density anomaly is mitigated. The inverted V_P results delineate clear boundaries and nicely reconstruct the true values of the low-velocity anomalies. Concerning the τ_s results, the reconstructed model contains less crosstalk which are projected from the V_S perturbations. In both PCG and PTN method, the τ_s anomalies in the deeper part are poorly reconstructed. This might be caused by the inaccurate approximation of the Q

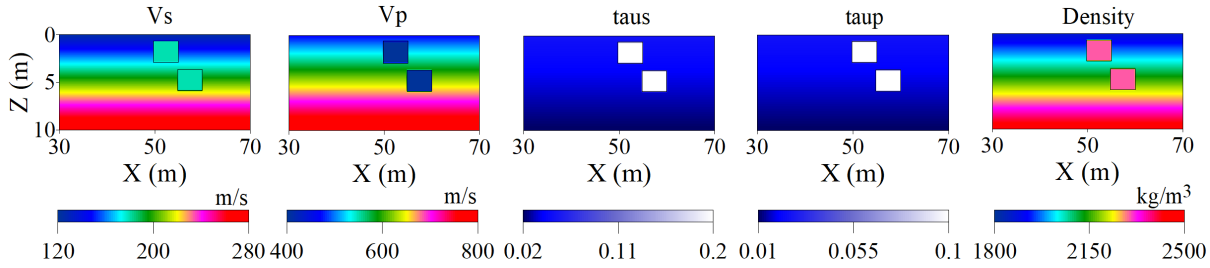


Figure 9. The true model for synthetic example on a spatially correlated model.

in the low-frequency range because we only take one relaxation mechanism ($L = 1$) during the modelling and inversion (Gao et al. 2020). For the inverted density model, the crosstalk effect projected from the V_S anomalies is mitigated and the deeper part of the density anomaly is better reconstructed than by the PCG. Nevertheless, the final synthetic shot gathers in both the PCG and PTN FWIs show a nice agreement with the observed data (left, Figure 7). This can be interpreted either as low sensitivity of the synthetic data to the secondary-order parameters or as the artefacts in the models compensate for the differences between modelled and observed data. However, the magnified waveform residuals (green and pink solid lines) in the PTN FWI results are hardly visible and smaller than the PCG FWI results (right, Figure 7). Plots of data misfit and model misfits also confirm that PTN outperforms PCG in reducing data and model misfits (Figure 8).

Overall, the reconstructed models demonstrate that incorporating the Hessian during the inversion can refocus the deeper part and mitigate the crosstalk between different parameters.

5.2 Spatially correlated models

In the previous synthetic model, we deliberately choose parameters spatially uncorrelated to each other, so that crosstalk between the parameters classes can be easily recognized and analysed. In realistic geological cases, different parameter classes somehow are often spatially correlated or anti-correlated. Here, we conduct two spatially correlated models to investigate the performance of simultaneously inverting five parameters with PCG and PTN FWIs. All parameter perturbations are located at the same positions (Figure 9). The same acquisition geometry, initial models, and inversion strategy as in the first example are used. Similarly, we also perform both viscoelastic PCG and PTN FWIs on the same dataset for comparison.

The shallow V_S and ρ anomalies are well reconstructed in the PCG results, while the deeper part of the anomalies cannot be identified clearly (Figure 10). The τ_s anomalies are also well reconstructed, which is partially supported by the crosstalk from the V_S anomaly. Similarly, the shallow ρ anomaly might also benefit from crosstalk from the V_S anomaly. V_P and τ_p models are roughly reconstructed

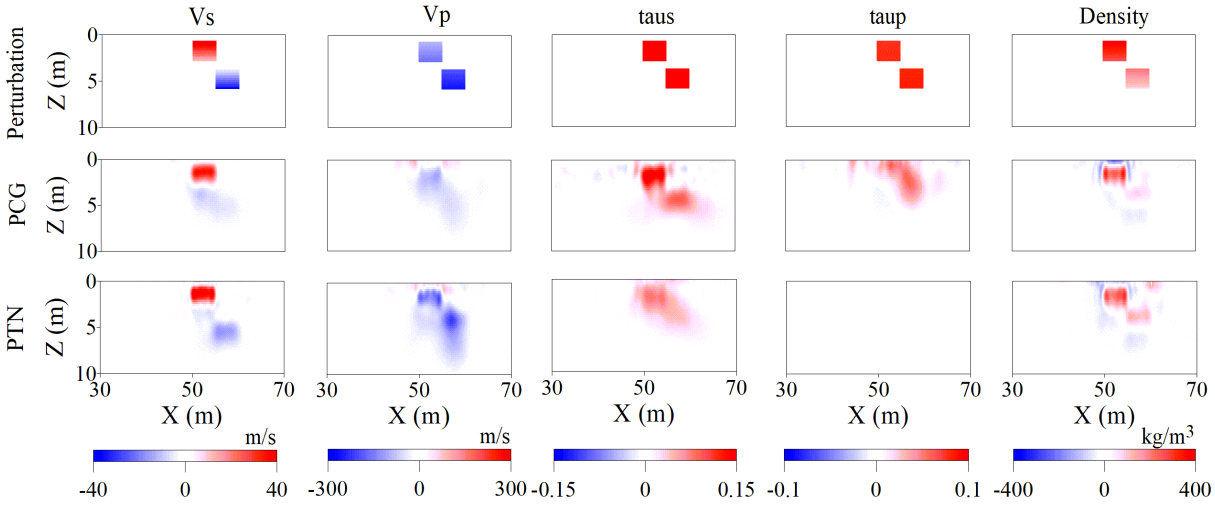


Figure 10. Multiparameter reconstruction example on a spatially correlated model (Figure 9). Three rows represent the true perturbation models, viscoelastic FWI results with preconditioned conjugate gradient method (PCG) and preconditioned truncated Newton method (PTN), respectively.

in the PCG results. PTN reconstructs V_S, V_P and ρ better with a more accurate estimation of values compared to PCG. For brevity, we do not show the comparison for shot gather and misfit. But similar to the previous example, the final synthetic shot gathers in both the PCG and PTN FWIs show a nice agreement with the observed data. The waveform residual in the PTN FWI results is hardly visible and is smaller than the PCG results. The data misfit with the PTN method indicates a better convergence compare to the PCG method.

In the next spatially-correlated synthetic example, we replace the high-density anomalies with low-density anomalies. In this example, the V_S and ρ anomalies are anticorrelated in the shallow part. In the inverted PCG results (Figure 11), the low-value density anomaly can be hardly reconstructed and even produces a wrong high-value anomaly in the shallow part. This may be caused by the crosstalk of the high S-wave velocity anomaly. We notice, however, that the deeper low-density anomaly is better reconstructed, which might be partially caused by the crosstalk between V_S and ρ. The previous synthetic example (Figure 6) showed that the V_S anomaly will cause a correlated footprint on the reconstruction of the density model. In the present experiment, the reconstruction of the shallow low-density anomaly is ruined by the correlated crosstalk from V_S anomaly. All five parameter models are better reconstructed in the PTN results, especially for the deeper V_S and V_P anomalies and the shallow anomaly for density. In both spatially correlated models, attenuation reconstruction from PCG seems to be more accurate compared to the results from the PTN method thanks to the footprint from the velocity model. However, the results may be contaminated when the spatially correlated parameters have anticorrelated values.

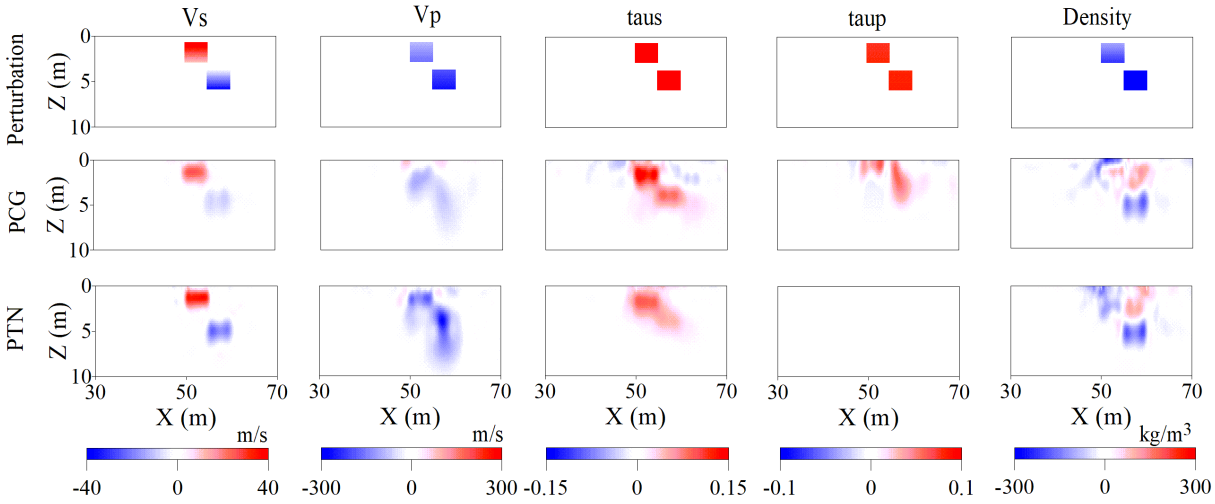


Figure 11. Multiparameter reconstruction example on a spatially correlated model where V_S and ρ are anti-correlated in the shallow part. Three rows represent the true perturbation models, viscoelastic FWI results with preconditioned conjugate gradient method (PCG) and preconditioned truncated Newton method (PTN), respectively.

Overall, the synthetic examples demonstrate that considering the Hessian can significantly improve the multiparameter reconstruction and mitigate the coupling between different parameter classes, which helps to reconstruct subsurface multiparameter models with higher confidence. Furthermore, the Hessian leads to better focusing of deeper anomalies and thus improves spatial resolution.

6 CONCLUSIONS

The truncated Newton method accounts for the inverse Hessian operator and helps to mitigate crosstalk in multiparameter viscoelastic FWI. We presented an efficient implementation of the preconditioned truncated Newton strategy in multiparameter viscoelastic FWI and showed synthetic reconstruction tests of shallow anomalies. The scattering responses of the five parameters show that the scattered wavefields of P- and S-wave velocity perturbations are similar at surface and shallow depth. With the depth increase, the scattered wavefield of Rayleigh wave caused by S-wave velocity perturbation has a 180-degree phase change along the backward direction. Attenuation parameters are always coupled with the corresponding velocity with a phase difference of about 90-degree. The density has partly coupled with S- and P-wave velocity thus is hard to be reconstructed. The complicated coupling effects make viscoelastic FWI difficult to simultaneously reconstruct multiparameter models accurately with a gradient-based optimization algorithm. We performed synthetic examples by using spatially uncorrelated and correlated models. They confirmed that by accounting for the Hessian during the

inversion, the truncated Newton method outperforms the conventional gradient-based optimization algorithm and improves the accuracy of the reconstructed models.

ACKNOWLEDGMENTS

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)-Project-ID 258734477-SFB 1173. The synthetic models and data presented in this study are available upon request to the first author.

7 APPENDICES

For the self-contentedness of this paper, we recall formulas originally derived by Kirsch & Rieder (2019).

APPENDIX A. GRADIENT CALCULATION

Here we give the explicit expression for the misfit gradient calculation.

The adjoint $\phi'(\boldsymbol{m})^\dagger$ at $\boldsymbol{m} = (\rho, V_S, \tau_s, V_P, \tau_p) \in \mathcal{M}$ is given by

$$\phi'(\boldsymbol{m})^\dagger \boldsymbol{g} = \begin{cases} \int_0^T (\partial_t \boldsymbol{v} \cdot \boldsymbol{w} - \frac{1}{\rho}\boldsymbol{\varepsilon}(v) : (\boldsymbol{\psi}_0 + \boldsymbol{\Sigma}))\mathrm{d}t, \\ \frac{2}{V_S}\int_0^t (-\boldsymbol{\varepsilon}(\boldsymbol{v}) : (\boldsymbol{\psi}_0 + \boldsymbol{\Sigma}) + \pi tr(\boldsymbol{\Sigma}^v)\mathrm{div}\boldsymbol{v})\mathrm{d}t, \\ \frac{1}{1+\alpha\tau_s}\int_0^T (\boldsymbol{\varepsilon}(\boldsymbol{v}) : \boldsymbol{\Sigma}^\tau_{s,2} + \pi tr(\boldsymbol{\Sigma}^\tau_{s,1})\mathrm{div}\boldsymbol{v})\mathrm{d}t, \\ -\frac{2\pi}{V_P}\int_0^T tr(\boldsymbol{\Sigma}^v)\mathrm{div}\boldsymbol{v}\mathrm{d}t, \\ \frac{\pi}{1+\alpha V_P}\int_0^T tr(\boldsymbol{\Sigma}^v)\mathrm{div}\boldsymbol{v}\mathrm{d}t, \end{cases} \tag{10}$$

for $\boldsymbol{g} = \Delta\boldsymbol{d} = (\Delta\boldsymbol{v}, \Delta\boldsymbol{\sigma}_0, \ldots, \Delta\boldsymbol{\sigma}_l)$, which is the data residual. Further, $\boldsymbol{v}$ is the first component of the solution of equation 3, $\boldsymbol{w}$ uniquely solves the backward equation

$$\begin{aligned} \partial_t \boldsymbol{w} &= \frac{1}{\rho}\mathrm{div}\left(\sum_{l=0}^{L}\boldsymbol{\sigma}_l\right) + \frac{1}{\rho}\Delta\boldsymbol{v}, \\ \partial_t \boldsymbol{\sigma}_0 &= C\left(\mu_0, \pi_0\right)\left(\boldsymbol{\varepsilon}\left(\boldsymbol{w}\right) + \Delta\boldsymbol{\sigma_0}\right), \\ \partial_t \boldsymbol{\sigma}_l &= C\left(\tau_s\mu_0, \tau_p\pi_0\right)\left(\boldsymbol{\varepsilon}\left(\boldsymbol{w}\right) + \Delta\boldsymbol{\sigma_l}\right) + \frac{1}{\tau_{\sigma,l}}\boldsymbol{\sigma}_l, \end{aligned} \tag{11}$$

with $\boldsymbol{w}(T) = 0$.

And

$$\begin{aligned}\boldsymbol{\Sigma}^v &= \frac{1}{2(\pi-\mu)}\boldsymbol{\psi_0} + \frac{\tau_p}{2(\tau_p\pi-\tau_s\mu)}\boldsymbol{\Sigma},\\ \boldsymbol{\Sigma}^\tau_{s,1} &= -\frac{\alpha}{2(\pi-\mu)}\boldsymbol{\psi_0} + \frac{\tau_p}{2\tau_s(\tau_p\pi-\tau_s\mu)}\boldsymbol{\Sigma},\\ \boldsymbol{\Sigma}^\tau_{p} &= \frac{\alpha}{2(\pi-\mu)}\boldsymbol{\psi_0} - \frac{1}{2(\tau_p\pi-\tau_s\mu)}\boldsymbol{\Sigma},\end{aligned} \tag{12}$$

with $\boldsymbol{\Sigma} = \Sigma_{l=1}^{L}\boldsymbol{\psi}_l$, $\mu = \frac{\mu_0}{\rho}$ and $\pi = \frac{\pi_0}{\rho}$.

Appendix B. Hessian-vector product calculation

Let us define the FWI forward operator $\boldsymbol{\phi} = \Psi \circ F$ which means that $\boldsymbol{\phi}(m) = \Psi(F(m))$, where Ψ is the linear observation operator and F is the parameter-to-solution map $F : (\rho, V_S, \tau_s, V_P, \tau_p) \rightarrow (v, \sigma_0, ..., \sigma_L)$. For any $\hat{m} = (\hat{\rho}, \hat{v}_s, \hat{\tau}_s, \hat{v}_p, \hat{\tau}_p) \in \mathcal{M}$, the Hessian-vector product $\boldsymbol{H(m)\hat{m}}$ can be calculated as

$$\boldsymbol{H(m)\hat{m}} = \underbrace{\boldsymbol{\phi}'(m)^\dagger\boldsymbol{\phi}'(m)\hat{m}}_{P_1} + \underbrace{\Psi''(F(m))F'(m)^\dagger\hat{m}(\phi(m)-d_{obs})}_{P_2} + \underbrace{\Psi'(F(m))F''(m)^\dagger\hat{m}(\boldsymbol{\phi}(m)-d_{obs})}_{P_3} \tag{13}$$

From this equation, we can notice that the Hessian-vector products are given by the sum of three parts.

The calculation of the first term P_1 in equation 13 is explicitly solved with equation 10 with $\boldsymbol{\phi}'(\boldsymbol{m})\hat{\boldsymbol{m}} = \bar{\boldsymbol{u}}$ where $\bar{\boldsymbol{u}} = (\bar{\boldsymbol{v}}, \bar{\boldsymbol{\sigma_0}}, \ldots, \bar{\boldsymbol{\sigma_L}})$ with $\bar{\boldsymbol{u}}(0) = 0$ is the solution of

$$\begin{aligned}\partial_t\bar{\boldsymbol{v}} &= \frac{1}{\rho}\mathrm{div}\left(\sum_{l=0}^{L}\boldsymbol{\sigma}_l\right) - \frac{\hat{\rho}}{\rho}\partial_t\boldsymbol{v},\\ \partial_t\bar{\boldsymbol{\sigma}}_{\mathbf{0}} &= C\left(\mu_0,\pi_0\right)\boldsymbol{\varepsilon}\left(\bar{\boldsymbol{v}}\right) + (\hat{\rho}C(\mu,\pi) + \rho C(\tilde{\mu},\tilde{\mu}))\boldsymbol{\varepsilon}(\boldsymbol{v}),\\ \partial_t\bar{\boldsymbol{\sigma}}_{\boldsymbol{l}} &= C\left(\tau_s\mu_0,\tau_p\pi_0\right)\boldsymbol{\varepsilon}\left(\bar{\boldsymbol{v}}\right) - \frac{1}{\tau_{\sigma,l}}\bar{\boldsymbol{\sigma}}_{\boldsymbol{l}} + (\hat{\rho}C(\tau_s\mu,\tau_p\pi) + \rho C(\hat{\mu},\hat{\mu}))\boldsymbol{\varepsilon}(\boldsymbol{v}),\end{aligned} \tag{14}$$

where $(\boldsymbol{v}, \boldsymbol{\sigma_0}, \ldots, \boldsymbol{\sigma_L})$ is the solution of equation 3.

The second part P_2 is given by

$$\Psi''(F(m))F'(m)\hat{m}^\dagger g = \begin{cases}\int_0^T(\partial_t\bar{\boldsymbol{v}}\cdot\boldsymbol{w} - \frac{1}{\rho}\boldsymbol{\varepsilon}(\bar{\boldsymbol{v}}) : (\boldsymbol{\psi}_0 + \boldsymbol{\Sigma}))\mathrm{d}t,\\ \frac{2}{V_S}\int_0^t(-\boldsymbol{\varepsilon}(\bar{\boldsymbol{v}}) : (\boldsymbol{\psi}_0 + \boldsymbol{\Sigma}) + \pi tr(\boldsymbol{\Sigma}^v)\mathrm{div}\bar{\boldsymbol{v}})\mathrm{d}t,\\ \frac{1}{1+\alpha\tau_s}\int_0^T(\boldsymbol{\varepsilon}(\bar{\boldsymbol{v}}) : \boldsymbol{\Sigma}^\tau_{s,2} + \pi tr(\boldsymbol{\Sigma}^\tau_{s,1})\mathrm{div}\bar{\boldsymbol{v}})\mathrm{d}t,\\ -\frac{2\pi}{V_P}\int_0^T tr(\boldsymbol{\Sigma}^v)\mathrm{div}\bar{\boldsymbol{v}}\mathrm{d}t,\\ \frac{\pi}{1+\alpha V_P}\int_0^T tr(\boldsymbol{\Sigma}^v)\mathrm{div}\bar{\boldsymbol{v}}\mathrm{d}t,\end{cases} \tag{15}$$

where $\bar{\boldsymbol{v}}$ is the solution of equation 14, $\boldsymbol{w} = (\boldsymbol{w}, \boldsymbol{\psi_0}, \ldots, \boldsymbol{\psi_l})$ solves equation 11 with $\boldsymbol{w}(T) = 0$.

The second part P_3 is given by

$$\Psi'(F(m))F''(m)\hat{m}^\dagger g = \begin{cases} \frac{1}{\rho}\int_0^T(\boldsymbol{\varepsilon}(\boldsymbol{v}) : \boldsymbol{\Upsilon}_1^\rho + tr(\boldsymbol{\Upsilon}_2^\rho)\mathrm{div}\boldsymbol{v})\mathrm{d}t, \\ \frac{2}{V_S}\int_0^t(\boldsymbol{\varepsilon}(\boldsymbol{v}) : \boldsymbol{\Upsilon}_{s,1}^v + tr(\boldsymbol{\Upsilon}_{s,2}^v)\mathrm{div}\boldsymbol{v})\mathrm{d}t, \\ \frac{1}{1+\alpha\tau_s}\int_0^T(\boldsymbol{\varepsilon}(\boldsymbol{v}) : \boldsymbol{\Upsilon}_{s,1}^\tau + tr(\boldsymbol{\Upsilon}_{s,2}^\tau)\mathrm{div}\boldsymbol{v})\mathrm{d}t, \\ \frac{2\pi}{V_P}\int_0^T tr(\boldsymbol{\Upsilon}_P^v)\mathrm{div}\boldsymbol{v}\mathrm{d}t, \\ \frac{\pi}{1+\alpha V_P}\int_0^T tr(\boldsymbol{\Upsilon}_P^\tau)\mathrm{div}\boldsymbol{v}\mathrm{d}t, \end{cases} \tag{16}$$

where $\boldsymbol{v}$ is the solution of equation 1 and $\boldsymbol{w} = (\boldsymbol{w}, \boldsymbol{\psi_0}, \ldots, \boldsymbol{\psi_l})$ solves equation 11 with $\boldsymbol{w}(T) = 0$.

In equation 16 the following coefficients are used:

$$\begin{aligned} \boldsymbol{\Upsilon}_1^\rho &= (\frac{\hat{\rho}}{\rho} + \frac{\tilde{\mu}}{\mu})\boldsymbol{\psi_0} + (\frac{\hat{\rho}}{\rho} + \frac{\hat{\mu}}{\tau_s\mu})\boldsymbol{\Sigma}, \\ \boldsymbol{\Upsilon}_2^\rho &= \frac{\tilde{\pi} - \tilde{\mu}\pi}{2\mu(\pi-\mu)}\boldsymbol{\psi_0} + \frac{\hat{\pi}\tau_s\mu - \tilde{\mu}\tau_p\pi}{2\tau_s\mu(\tau_p\pi - \tau_s\mu)}\boldsymbol{\Sigma}, \\ \boldsymbol{\Upsilon}_1^{V_S} &= (\frac{\hat{\rho}}{\rho} + \frac{2\tilde{\mu}}{\mu})\boldsymbol{\psi_0} + (\frac{\hat{\rho}}{\rho} + \frac{2\hat{\mu}}{\tau_s\mu})\boldsymbol{\Sigma}, \\ \boldsymbol{\Upsilon}_2^{V_S} &= K_{S,\psi}\boldsymbol{\psi_0} + K_{S,\Sigma}\boldsymbol{\Sigma}, \\ \boldsymbol{\Upsilon}_1^{\tau_s} &= -\alpha(\frac{\hat{\rho}}{\rho} + \frac{2\tilde{\mu}}{\mu})\boldsymbol{\psi_0} + (\frac{\hat{\rho}}{\tau_s\rho} + \frac{2\hat{\mu}}{\tau_s^2\mu})\boldsymbol{\Sigma}, \\ \boldsymbol{\Upsilon}_2^{\tau_s} &= -\alpha K_{S,\psi}\boldsymbol{\psi_0} + \frac{K_{S,\Sigma}}{\tau_s}\boldsymbol{\Sigma}, \\ \boldsymbol{\Upsilon}_2^{V_P} &= K_{P,\psi}\boldsymbol{\psi_0} + \tau_p K_{P,\Sigma}\boldsymbol{\Sigma}, \\ \boldsymbol{\Upsilon}_2^{V_P} &= -\alpha K_{P,\psi}\boldsymbol{\psi_0} + K_{P,\Sigma}\boldsymbol{\Sigma}, \end{aligned} \tag{17}$$

where

$$\begin{aligned} K_{S,\psi} &= \frac{2\pi\mu\tilde{\mu} - \tilde{\mu}\pi^2 - \tilde{\pi}\mu^2}{\mu(\pi-\mu)^2} - \frac{\hat{\rho}}{\rho}\frac{\pi}{2(\pi-\mu)}, \\ K_{S,\Sigma} &= \frac{2\tau_s\tau_p\pi\mu\hat{\mu} - \hat{\mu}\tau_p^2\pi^2 - \hat{\pi}\tau_s^2\mu^2}{\tau_s\mu(\tau_p\pi - \tau_s\mu)^2} - \frac{\hat{\rho}}{\rho}\frac{\pi}{2(\tau_p\pi - \tau_s\mu)}, \\ K_{P,\psi} &= \frac{\hat{\rho}}{\rho}\frac{1}{2(\pi-\mu)} + \frac{\hat{\pi} - \hat{\mu}}{(\pi-\mu)^2}, \\ K_{P,\Sigma} &= \frac{\hat{\rho}}{\rho}\frac{1}{2(\tau_p\pi - \tau_s\mu)} + \frac{\hat{\pi} - \hat{\mu}}{(\tau_p\pi - \tau_s\mu)^2}. \end{aligned} \tag{18}$$

with

$$\begin{aligned} \tilde{\mu} &= \frac{2V_S}{1+\alpha\tau_s}\hat{v}_s - \frac{\alpha V_S^2}{(1+\alpha\tau_s)^2}\hat{\tau}_s, \\ \hat{\mu} &= \frac{2\tau_s V_S}{1+\alpha\tau_s}\hat{v}_s + \frac{V_S^2}{(1+\alpha\tau_s)^2}\hat{\tau}_s, \\ \tilde{\pi} &= \frac{2V_P}{1+\alpha\tau_p}\hat{v}_p - \frac{\alpha V_P^2}{(1+\alpha\tau_p)^2}\hat{\tau}_p, \\ \hat{\pi} &= \frac{2\tau_p V_P}{1+\alpha\tau_p}\hat{v}_p + \frac{V_P^2}{(1+\alpha\tau_p)^2}\hat{\tau}_p. \end{aligned} \tag{19}$$

REFERENCES

Bohlen, T., 2002. Parallel 3-D viscoelastic finite difference seismic modelling, *Computers & Geosciences*, **28**(8), 887–899.

Brossier, R., Operto, S., & Virieux, J., 2009. Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion, *Geophysics*, **74**(6), WCC105–WCC118.

Chen, X., 1993. A systematic and efficient method of computing normal modes for multilayered half-space, *Geophysical Journal International*, **115**(2), 391–409.

Fabien-Ouellet, G., Gloaguen, E., & Giroux, B., 2017. Time domain viscoelastic full waveform inversion, *Geophysical Journal International*, **209**(3), 1718–1734.

Fichtner, A., Kennett, B. L., Igel, H., & Bunge, H.-P., 2008. Theoretical background for continental- and global-scale full-waveform inversion in the time–frequency domain, *Geophysical Journal International*, **175**(2), 665–685.

Gao, L., Xia, J., & Pan, Y., 2014. Misidentification caused by leaky surface wave in high-frequency surface wave method, *Geophysical Journal International*, **199**(3), 1452–1462.

Gao, L., Xia, J., Pan, Y., & Xu, Y., 2016. Reason and condition for mode kissing in MASW method, *Pure and Applied Geophysics*, **173**(5), 1627–1638.

Gao, L., Pan, Y., & Bohlen, T., 2020. 2-D multiparameter viscoelastic shallow-seismic full-waveform inversion: reconstruction tests and first field-data application, *Geophysical Journal International*, **222**(1), 560–571.

Groos, L., Schäfer, M., Forbriger, T., & Bohlen, T., 2014. The role of attenuation in 2D full-waveform inversion of shallow-seismic body and Rayleigh waves, *Geophysics*, **79**(6), R247–R261.

Kamei, R. & Pratt, R., 2013. Inversion strategies for visco-acoustic waveform inversion, *Geophysical Journal International*, **194**(2), 859–884.

Kirsch, A. & Rieder, A., 2019. Inverse problems for abstract evolution equations II: higher order differentiability for viscoelasticity, *SIAM Journal on Applied Mathematics*, **79**(6), 2639–2662.

Köhn, D., De Nil, D., Kurzmann, A., Przebindowska, A., & Bohlen, T., 2012. On the influence of model parametrization in elastic full waveform tomography, *Geophysical Journal International*, **191**(1), 325–345.

Köhn, D., Wilken, D., De Nil, D., Wunderlich, T., Rabbel, W., Werther, L., Schmidt, J., Zielhofer, C., & Linzen, S., 2019. Comparison of time-domain sh waveform inversion strategies based on sequential low and bandpass filtered data for improved resolution in near-surface prospecting, *Journal of Applied Geophysics*, **160**, 69–83.

Liu, H.-P., Anderson, D. L., & Kanamori, H., 1976. Velocity dispersion due to anelasticity; implications for seismology and mantle composition, *Geophysical Journal International*, **47**(1), 41–58.

Malinowski, M., Operto, S., & Ribodetti, A., 2011. High-resolution seismic attenuation imaging from wide-aperture onshore data by visco-acoustic frequency-domain full-waveform inversion, *Geophysical Journal International*, **186**(3), 1179–1204.

Métivier, L., Brossier, R., Virieux, J., & Operto, S., 2013. Full waveform inversion and the truncated newton method, *SIAM Journal on Scientific Computing*, **35**(2), B401–B437.

Métivier, L., Brossier, R., Operto, S., & Virieux, J., 2015. Acoustic multi-parameter fwi for the reconstruction of p-wave velocity, density and attenuation: Preconditioned truncated newton approach, in *SEG Technical Program Expanded Abstracts 2015*, pp. 1198–1203, Society of Exploration Geophysicists.

Nocedal, J. & Wright, S., 2006. *Numerical optimization*, Springer.

Operto, S., Gholami, Y., Prieux, V., Ribodetti, A., Brossier, R., Metivier, L., & Virieux, J., 2013. A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice, *The leading edge*, **32**(9), 1040–1054.

Pan, W., Geng, Y., & Innanen, K. A., 2018. Interparameter trade-off quantification and reduction in isotropic-elastic full-waveform inversion: Synthetic experiments and hussar land data set application, *Geophysical Journal International*, **213**(2), 1305–1333.

Pan, Y., Xia, J., Xu, Y., & Gao, L., 2016a. Multichannel analysis of Love waves in a 3D seismic acquisition system, *Geophysics*, **81**(5), EN67–EN74.

Pan, Y., Xia, J., Xu, Y., Gao, L., & Xu, Z., 2016b. Love-wave waveform inversion for shallow shear-wave velocity using a conjugate gradient algorithm, *Geophysics*, **81**(1), R1–R14.

Pan, Y., Gao, L., & Bohlen, T., 2019. High-resolution characterization of near-surface structures by surface-wave inversions: From dispersion curve to full waveform, *Surveys in Geophysics*, **40**(2), 167–195.

Plessix, R. & Mulder, W., 2004. Frequency-domain finite-difference amplitude-preserving migration, *Geophysical Journal International*, **157**(1), 957–987.

Pratt, R. G., 1999. Seismic waveform inversion in the frequency domain; part 1, theory and verification in a physical scale model, *Geophysics*, **64**(3), 888–901.

Pratt, R. G., Shin, C., & Hick, G., 1998. Gauss–newton and full newton methods in frequency–space seismic waveform inversion, *Geophysical journal international*, **133**(2), 341–362.

Prieux, V., Brossier, R., Operto, S., & Virieux, J., 2013. Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the valhall field. part 1: Imaging compressional wave speed, density and attenuation, *Geophysical Journal International*, **194**(3), 1640–1664.

Robertsson, J. O., Blanch, J. O., & Symes, W. W., 1994. Viscoelastic finite-difference modeling, *Geophysics*, **59**(9), 1444–1456.

Romdhane, A., Grandjean, G., Brossier, R., Rejiba, F., Operto, S., & Virieux, J., 2011. Shallow-structure characterization by 2D elastic full waveform inversion, *Geophysics*, **76**(3), R81–R93.

Shin, C. & Ho Cha, Y., 2009. Waveform inversion in the laplace—fourier domain, *Geophysical Journal International*, **177**(3), 1067–1079.

Socco, L., Foti, S., & Boiero, D., 2010. Surface wave analysis for building near surface velocity models: established approaches and new perspectives, *Geophysics*, **75**(5), A83–A102.

Tarantola, A., 1986. A strategy for nonlinear elastic inversion of seismic reflection data, *Geophysics*, **51**(10), 1893–1903.

Tran, K., McVay, M., Faraone, M., & Horhota, D., 2013. Sinkhole detection using 2D full seismic waveform tomography, *Geophysics*, **78**(5), R175–R183.

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

Warner, M. & Guasch, L., 2016. Adaptive waveform inversion: Theory, *Geophysics*, **81**(6), R429–R445.

Xia, J., Miller, R., & Park, C., 1999. Estimation of near-surface shear-wave velocity by inversion of Rayleigh wave, *Geophysics*, **64**(4), 691–700.

Xing, Z. & Mazzotti, A., 2019. Two-grid full-waveform rayleigh-wave inversion via a genetic algorithm—part 2: Application to two actual data sets, *Geophysics*, **84**(5), R815–R825.

Yang, P., Brossier, R., Métivier, L., & Virieux, J., 2016. A review on the systematic formulation of 3-d multiparameter full waveform inversion in viscoelastic medium, *Geophysical Journal International*, **207**(1), 129–149.

Yang, P., Brossier, R., Métivier, L., Virieux, J., & Zhou, W., 2018. A time-domain preconditioned truncated newton approach to visco-acoustic multiparameter full waveform inversion, *SIAM Journal on Scientific Computing*, **40**(4), B1101–B1130.

Zeltmann, U. C., 2019. *The Viscoelastic Seismic Model: Existence, Uniqueness and Differentiability with Respect to Parameters*, Ph.D. thesis, Karlsruher Institut für Technologie (KIT), `http://dx.doi.org/10.5445/IR/1000093989`.