Velocity model building by slope tomography

Bernard Law and Daniel Trad

ABSTRACT

Slope tomography method uses slopes and traveltimes of locally coherent reflected events to estimate the macro-velocity model from reflection data for depth imaging and full waveform inversion (FWI). Without the requirement of picking traveltimes on continuous reflection events, slope tomography is operationally more efficient than traditional reflection tomography. It is computationally more efficient than migration velocity (MVA) because it estimates the global velocity model simultaneously without layer stripping and expensive depth migration iterations. We review the slope tomography methods including CDR tomography, stereotomography, and adjoint stereotomography.

INTRODUCTION

An accurate starting model is very important to depth migration and full waveform inversion (FWI). Depth migration can update the velocity model iteratively using migration velocity analysis (MVA) methods; however, each iteration requires an update to the velocity model and depth migration process. The goal of FWI is to converge to the global minimum of the objective function and to arrive at the correct model. However, FWI is an ill-posed problem, its solution often represents only a local minimum. Therefore, an accurate initial model can improve the efficiency and accuracy of depth migration and FWI. Traditional reflection tomography methods (Bishop et al. 1985) inverse traveltime picks to a velocity model and requires difficult interpretive traveltime picking of continuous reflection events and computationally complex two-point ray tracing. Sword (1987) uses the concept of controlled directional reception (CDR) method (Rieber 1936; Riabinkin 1957) and ray parameters of the waves transmitted from a shot to a receiver to invert for the velocity model. The ray parameters of a locally coherent reflection event can be picked interactively or automatically on localized slant stacks of shot and geophone gathers. Stereotomography (Billette and Lambaré 1998) is a generalized slope tomography method that extended the CDR tomography to include a larger data and model space to make the inversion less sensitive to the input slope picks and rapid velocity changes. However, this generalization also results in a large Fréchet derivative matrix that can make this method infeasible for large seismic surveys. Adjoint stereotomography (Tavakoli et al. 2017) uses the adjoint-state method to avoid the computation and memory cost of building the Fréchet derivative matrix. We review these slope tomography methods and evaluate their effectiveness as a velocity model building tool.

CDR method

The principles of controlled directional sensitivity (CDS) method was first introduced by Rieber in 1936. In standard seismic recording, signals from multiple geophones are summed to suppress surface waves (Fig.1a). Rieber recognized this method could also suppress reflection events from dipping reflectors (Fig.1b) and proposed to align the geophone summing in the direction normal to reflection arrival (Fig.1c). Instead of summing multiple geophones during recording, the signal from individual geophones was recorded on film in variable density. An optical apparent dip analyzer was used to align the reflection arrival along the direction of the apparent dip (Fig.1d).

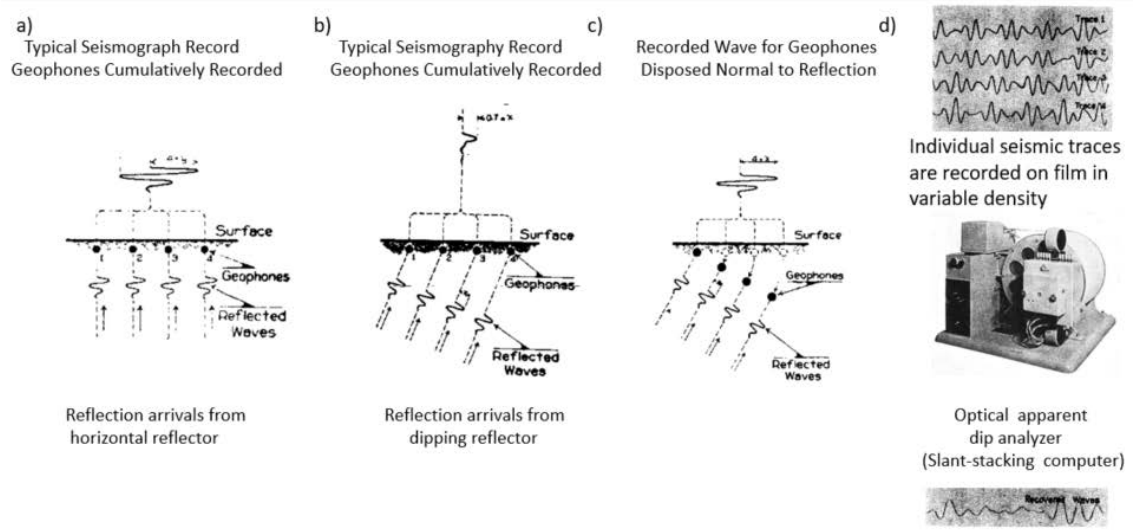


Figure 1. (a) Reflection arrivals from horizontal reflector are enhanced by summing geophones during recording, (b) reflection arrivals from dipping reflector are suppressed, (c) aligning the geophone summing can preserve reflection arrivals from dipping reflector, (d) optical apparent dip analyzer. (Rieber 1936)

Riabinkin (1957) described the development of the CDR method in Russia. Using a similar concept as Rieber (1936), the seismic waves were summed according to the apparent velocity and recorded on film as a summofilm. This process is equivalent to slant stacking and is used to separate seismic waves according to their apparent velocity or localized slopes. When slant stacking is performed on traces within a shot record x_s, it identifies the slopes of localized events p_g at the central geophone location x_g. Similarly, slant stacking performed on traces within a geophone gather x_g identifies the slopes of localized events p_s at the central shot location x_s (Fig. 2a). These parameters are referred to as the reciprocal parameters. Assuming straight ray paths (Fig. 2b) the location of the reflection point (X_R, Z_R) and the dip angle ϕ can be computed (Billette et. al, 2003). Therefore, each triplet of P_s, P_g and T_{SR} pick is equivalent to a localized dipping event represented by X_R, Z_R and ϕ. By picking locally coherent events on shot and geophone slant stacks, dip bars can be constructed for structural interpretation. Figure 3 displays the dip bars computed from the reciprocal parameters picks of a synthetic dataset created from the Marmousi model. The dip bars show reasonable correlations with velocity model.

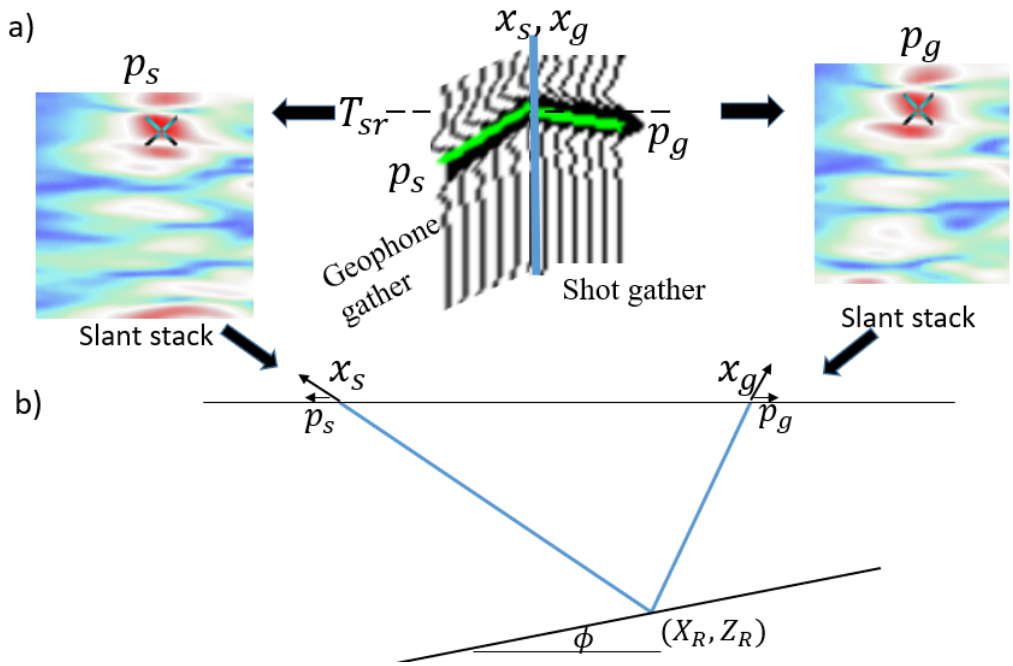


Figure 2. (a) Shot and geophone slopes (p_s, p_g) and reflection arrival time T_{sr} of localized coherent event are picked from slant stacks created from geophone gather x_g and shot record x_s, (b) Reflector position (X_R, Z_R) and reflector dip ϕ associated with the reciprocal parameters p_s, p_g and T_{sr} can be determined assuming straight ray paths and constant velocity.

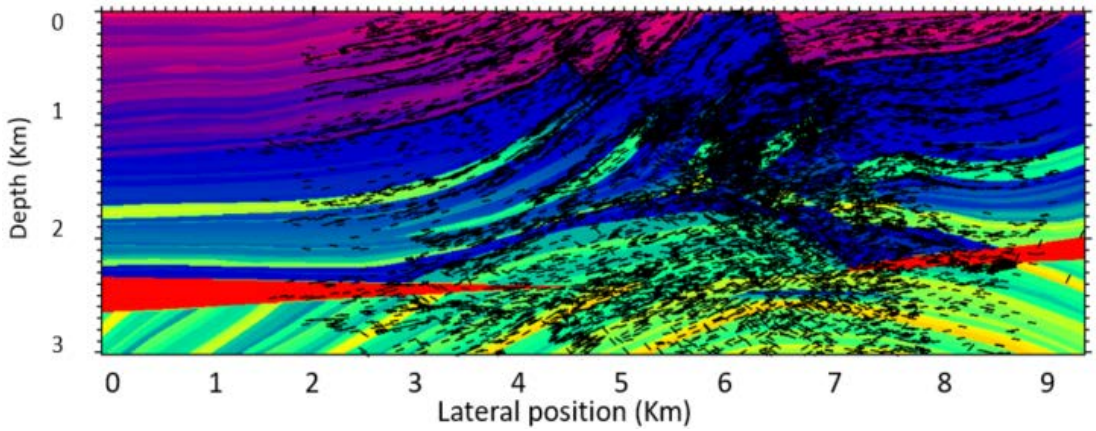


Figure 3. Marmousi model overlaid with dip bars computed from reciprocal parameter picks.

CDR tomography

Sword (1987) proposed to use the reciprocal parameters to estimate the velocity. If we trace rays from x_s and x_g using ray take off angles associated with ray parameters p_s and p_r, the rays will meet when the sum of the source and receiver ray path traveltimes T_s, T_g equals the reflection arrival time T_{sr} (Fig. 4). However, if an incorrect velocity model is used in the ray tracing, the rays will not meet (Fig. 5) and the distance error X_{err} will be non-zero:

$$Xerr = x_g - x_s \tag{1}$$

X_{err} is used in the cost function $\boldsymbol{J}(\boldsymbol{v})$ for CDR tomographic inversion:

$$\boldsymbol{J}(\boldsymbol{v}) = \| \boldsymbol{X_{err}} \|^2 \tag{2}$$

Damping factors λ_x and λ_z are added to avoid rapid changes in velocity:

$$\boldsymbol{J}(\boldsymbol{v}) = \| \boldsymbol{X_{err}} \|^2 + \lambda_x^2 \| \frac{\boldsymbol{\partial v^2}}{\boldsymbol{\partial x}} \| + \lambda_z^2 \| \frac{\boldsymbol{\partial v^2}}{\boldsymbol{\partial z}} \| \tag{3}$$

The tomographic inversion problem is solved by finding the value of v that minimize the cost function $J(v)$. This is done by solving the following least squares system:

$$\boldsymbol{A}^{(k)} \boldsymbol{\Delta v} = -\boldsymbol{X}_{\boldsymbol{err}}^{(k)} , \tag{4}$$

where $X_{err}^{(k)} = X_{err}(v^{(k)})$ and is computed by ray tracing the current velocity model and takeoff angles computed from p_s and p_g, Δv is the value to update the velocity model with and $A^{(k)}$ is the Fréchet derivative matrix:

$$A_{mn}^{k} = \left(\frac{\partial X_{err}(n)}{\partial v_m}\right)_{v=v^k} , \tag{5}$$

where k is the iteration count, i is the model index and j is the input index. Equation (4) in expanded form using equation (5) is:

$$\begin{bmatrix} \frac{\partial X_{err\,1}}{\partial v_1} & \frac{\partial X_{err\,1}}{\partial v_2} & .. & \frac{\partial X_{err\,1}}{\partial v_I} \\ \frac{\partial X_{err\,2}}{\partial v_1} & \frac{\partial X_{err\,2}}{\partial v_2} & .. & \frac{\partial X_{err\,2}}{\partial v_I} \\ .. & .. & .. & .. \\ \frac{\partial X_{err\,J}}{\partial v_1} & \frac{\partial X_{err\,J}}{\partial v_2} & .. & \frac{\partial X_{err\,J}}{\partial v_I} \end{bmatrix} \begin{bmatrix} \Delta v_1 \\ \Delta v_2 \\ .. \\ \Delta v_I \end{bmatrix} = - \begin{bmatrix} X_{err\,1} \\ X_{err\,2} \\ .. \\ X_{err\,J} \end{bmatrix} \tag{6}$$

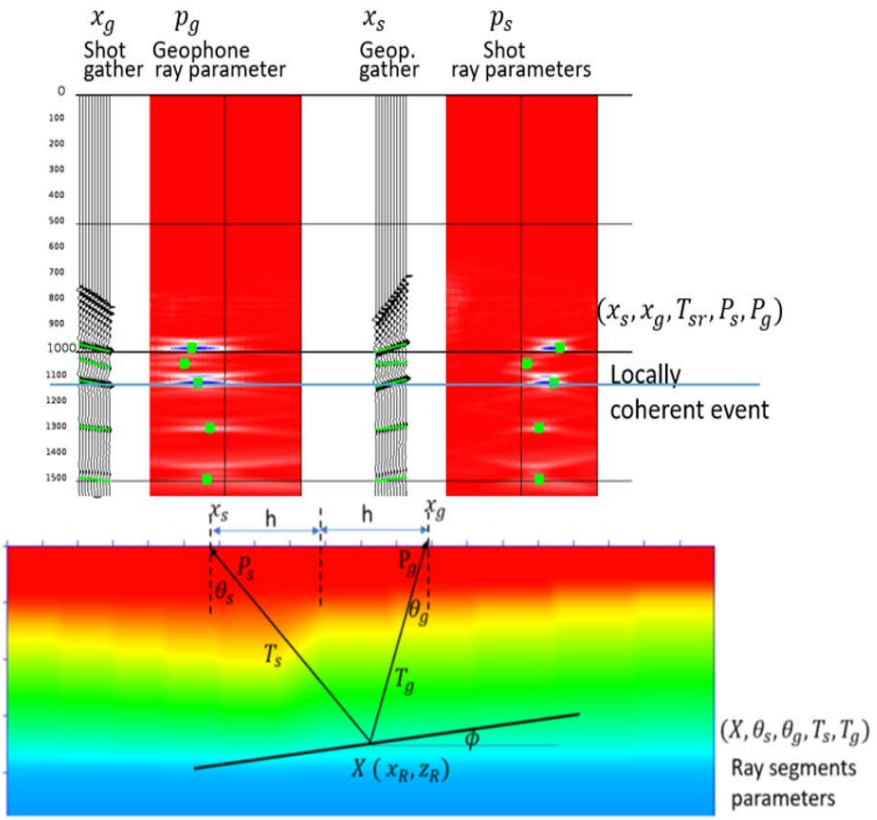


Figure 4. A locally coherent event can be picked on the localized shot and receiver slant stacks. The event is characterized by the traveltime T_{sr} and the ray parameters p_s and p_r and is associated with a ray segment pair in the velocity model. Ray segment parameters including the scatter point location X, ray shooting angles θ_s and θ_g can be computed from the half offset h, the ray parameters and two-way travel time T_{sr}.

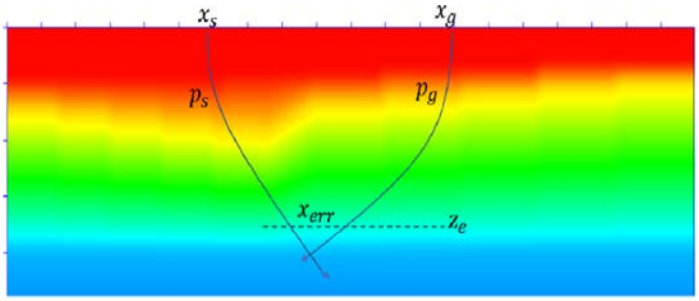


Figure 5. Forward modelling is done by shooting rays from the source location X_s and receiver location X_g using picked ray parameters P_s and P_g. If errors exist in the velocity model, the source and receiver ray paths will not meet at the depth where the sum of ray traced times $T_s + T_g$ equals the measured time T_{sr}

CDR tomography algorithm

1: Input starting velocity model and slope and travel time picks
2: Inversion:

- Loop on number of iterations
  - Loop on number of picks
    Compute X_{err} , equation (1)
    Compute Fréchet derivatives A_{mn}^{k}, equation (5)
  - Build equations $\boldsymbol{A}^{(k)}\ \boldsymbol{\Delta v} = -\boldsymbol{X}_{\boldsymbol{err}}^{(k)}$
  - Solve for $\boldsymbol{\Delta v}$
  - Update model parameters
- Next inversion iteration

Stereotomography

The CDR tomography method uses the picked ray parameters P_s and P_g to compute the shooting angles for the forward modelling. This can result in the solution being sensitive to picking errors and complex structures in the velocity model. Billette and Lambaré (1997) presented the stereotomography method as a generalized slope tomography method. The name "Stereotomography" is chosen because each event consists of both source and receiver ray paths. Stereotomography uses the same data picking concept as CDR tomography; however, it shoots rays from a scatter point toward the source and receiver using the initial ray segment parameters $(X, \theta_s, \theta_g, T_s$ and $T_g)$ computed from the input data (Fig. 6) . If the velocity model and initial ray segment parameters are incorrect, the computed X_s, X_g, P_s, P_g and T_{sr} will be different than the input measurements. Using all these differences can reduce the sensitivity of the inverse problem to the ray parameters picks and make stereotomography more stable than CDR tomography.

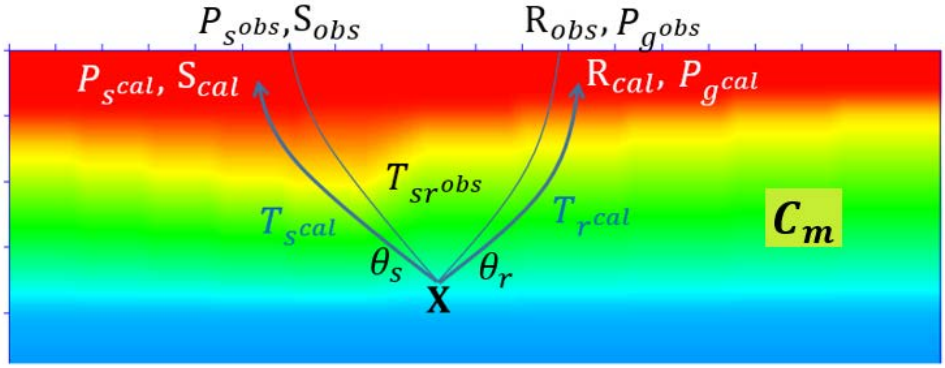


Figure 6: Model space of stereotomography consists of the scatter position X, ray shooting anlges θ_s, θ_r, traveltimes T_s, T_r and velocity C_m. The data space consists of source and receiver positions X_s, X_g, source and receiver ray parameters P_s, P_g and arrival time T_{sr}.

For a dataset with J picked events, the data space d is represented as:

$$d = \left[\left(X_s, X_g, P_s, P_g, T_{sr}\right)_j\right]_{j=1}^{J}. \quad (7)$$

The model space is also expanded to include reflection/diffraction point X, ray scattering angles θ_s, θ_r , and one-way traveltime T_s and T_r (Fig. 6):

$$m = \left[\left[(X, \theta_s, \theta_r, T_s, T_r)_j\right]_{j=1}^{J}, [C_i]_{i=1}^{I}\right]. \quad (8)$$

The initial values for X, θ_s and θ_r are computed using straight ray paths and constant velocity assumption. The $l_2\ norm$ cost function for the inversion of the model parameter is

$$\boldsymbol{J(m)} = ||\ \boldsymbol{\Delta d(m)}\ ||^2, \quad (9)$$

where $\boldsymbol{\Delta d(m)} = \boldsymbol{d_{obs} - d_{calc}(m)}$ (10)

The model update $\boldsymbol{\Delta m}$ can be found by solving the linear system of equations

$$\boldsymbol{J(m)\Delta m} = \boldsymbol{\Delta d(m)}, \quad (11)$$

where $\boldsymbol{J(m)}$ is the Fréchet derivative matrix:

$$\boldsymbol{J}(m) = \partial\left(\left(\boldsymbol{X_s}, \boldsymbol{P_s}, \boldsymbol{X_g}, \boldsymbol{P_g}, T_{sr}\right)_j\right)_{j=1}^{J} \Big/ \partial\left(\left(\left(\boldsymbol{X}, \theta_s, \theta_g, T_s, T_g\right)_j\right)_{j=1}^{J}, (C_i)_{i=1}^{I}\right) \quad (12)$$

Equation (11) in expanded form is:

$$\begin{bmatrix} \frac{\partial X_{s_1}}{\partial X_1} & .. & \frac{\partial X_{s_1}}{\partial X_I} & .. & \frac{\partial X_{s_1}}{\partial \theta_{s_1}} & .. & \frac{\partial X_{s_1}}{\partial \theta_{s_I}} & .. & .. & \frac{\partial T_{sr_1}}{\partial C_1} & .. & \frac{\partial T_{sr_1}}{\partial C_J} \\ \frac{\partial X_{s_2}}{\partial X_1} & .. & \frac{\partial X_{s_2}}{\partial X_I} & .. & \frac{\partial X_{s_2}}{\partial \theta_{s_1}} & .. & \frac{\partial X_{s_2}}{\partial \theta_{s_I}} & .. & .. & \frac{\partial T_{sr_2}}{\partial C_1} & .. & \frac{\partial T_{sr_2}}{\partial C_J} \\ .. & .. & .. & .. & .. & .. & .. & .. & .. & .. & .. & .. \\ .. & .. & .. & .. & .. & .. & .. & .. & .. & .. & .. & .. \\ \frac{\partial X_{s_I}}{\partial X_1} & .. & \frac{\partial X_{s_I}}{\partial X_I} & .. & \frac{\partial X_{s_I}}{\partial \theta_{s_1}} & .. & \frac{\partial X_{s_I}}{\partial \theta_{s_I}} & .. & .. & \frac{\partial T_{sr_I}}{\partial C_1} & .. & \frac{\partial T_{sr_I}}{\partial C_J} \end{bmatrix} \begin{bmatrix} \Delta X_1 \\ \Delta X_I \\ .. \\ \Delta \theta_{s_1} \\ .. \\ \Delta \theta_{s_I} \\ .. \\ .. \\ \Delta C_1 \\ .. \\ \Delta C_J \end{bmatrix} = \begin{bmatrix} \Delta X_{s_1} \\ \Delta X_{s_I} \\ .. \\ \Delta X_{g_1} \\ .. \\ \Delta X_{g_I} \\ .. \\ .. \\ \Delta T_{sr_1} \\ .. \\ \Delta T_{sr_J} \end{bmatrix}, \quad (13)$$

The row dimension of $J(m)$ matrix is the number of picks J . The column dimension of $J(m)$ is 6 x J plus the number of velocity cells I.

The stereotomography algorithm involves the initialization, localization and the inversion steps (Billette and Lambare, 1998; Billette et.al,2003) . The initialization step uses ray parameters and two-way traveltime picks and straight rays and constant velocity assumption to calculate the initial estimates of the ray segment parameters $(X_R, Z_R, \theta_s, \theta_r, T_s, T_r)$. The localization step refines the ray segment parameters using a fixed velocity model by minimizing the difference between observed and modelled data $(X_s, Z_s, P_s, X_g, Z_g, P_g, T_{sr})$. The inversion step iteratively builds the Fréchet deriviative matrix $\boldsymbol{J}(\boldsymbol{m})$ and solves for the model updates vector $\boldsymbol{\Delta m}$.

Stereotomogaphy algorithm

1: Input starting velocity model and slope and traveltime picks

2: Multiscale loop
- Start with large grid spacing, progress to smaller grid spacing
- Use starting velocity model or velocity model from previous multiscale iteration and resample to new smaller grid
- Input velocity model, slope and traveltime data

3: Initialization:
- Compute initial estimates of ray segment parameters $(X_R, Z_R, \theta_s, \theta_r, T_s, T_r)$ using straight ray and constant velocity assumption

4: Localization:
- Loop on number of input data
  - Refine the ray segment parameters using a fixed starting velocity model by minimizing the difference between the observed data and modeled data.
    As the $X_R and\ Z_R$ are refined, new velocity will be retrieved from the model and new angles and ray path times will be recomputed.

5: Inversion:
- Loop on number of iterations
  - Compute data residuals and update cost function
  - Compute Fréchet derivatives
  - Build equations $\boldsymbol{J}(\boldsymbol{m})\boldsymbol{\Delta m} = \boldsymbol{\Delta d}(\boldsymbol{m})$
  - Solve for $\boldsymbol{\Delta m}$
  - Update model parameters
- Next inversion iteration

-Next multiscale iteration

Adjoint-state method

Stereotomography requires both the derivatives of the traveltimes with respect to the velocity parameters as well as the derivatives of the traveltimes with respect to the ray path. The former can be computed efficently using the ray segment length along the ray path; while the latter requires the solutions to the paraxial ray tracing equations and the propagator associated with these equations. Furthermore, for large-scale data, direct

computation, and handling of this Fréchet derivative matrix can be computationaly inefficient.

The adjoint-state method is a constrained optimzation technique that can provide a more efficient framework to compute the gradient of the misfit function with respect to its model parameters without the computation and memory requirements of building the Fréchet matrix (Plessix 2006). It has been used in other geophsyical application including least-squares migration and full waveform inversion. In the adjoint-state method, the misfit function $J(m)$ is subjected to the constraint of the realization of the state variable in the state equation $F(u,m)$, where m is the model parameter and u is the state variable. When J depends on $u(m), J$ is defined with $h(u(m),m)$.

$$J(m) = h(u(m),m), \tag{14}$$

$$L(u,\lambda,m) = h(u,m) - <\lambda, F(u,m)>, \tag{15}$$

where $L(u,\lambda,m)$ is the Lagrangian or the augmented misfit function, u is the state variable, m is the model parameter, $F(u,m)$ is the state equation of the forward modeling equation that map m to u, and λ is the Lagragian multiplier or the adjoint-state variable. λ, m and u are independent quantities but are linked through the state and adjoint-state equations.

Since u is a physical realization, $F(u,m)$ is zero. Therefore,

$$L(u,\lambda,m) = h(u,m) = J(m). \tag{16}$$

At the point of optimization, the gradient of L with respect to u, λ and m should be zero:

$$\frac{\partial L(u,\lambda,m)}{\partial u} = 0, \tag{17}$$

$$\frac{\partial L(u,\lambda,m)}{\partial \lambda} = 0, \tag{18}$$

$$\frac{\partial J}{\partial m} = \frac{\partial L(u,\lambda,m)}{\partial m} = \frac{\partial h(u,m)}{\partial m} - <\lambda, \frac{\partial F(u,m)}{\partial m}> = 0 \tag{19}$$

Zeroing the gradient of L with respect to u leads to the adjoint-state equation, which leads to the solution of λ:

$\frac{\partial L(u,\lambda,m)}{\partial u} = \frac{\partial h(u,m)}{\partial u} - <\lambda, \frac{\partial F(u,m)}{\partial \mathrm{u}}> = 0,$

$\frac{\partial h(u,m)}{\partial u}$ can be expanded as:

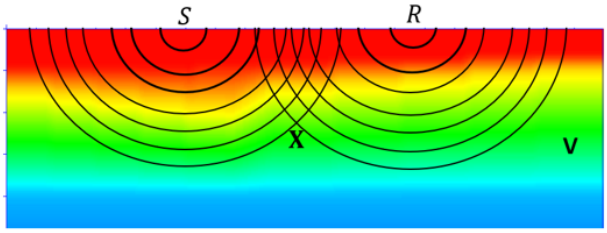
$\frac{\partial h(u,m)}{\partial u} = \frac{\partial}{\partial u}\left(\frac{1}{2} \| u^{cal} - u^{obs} \|^2\right) = \left(u^{cal} - u^{obs}\right)\frac{\partial}{\partial u}\left(u^{cal} - u^{obs}\right) = u^{cal} - u^{obs}$, therefore,

$$\left(\frac{\partial F(u,m)}{\partial u}\right)^* \lambda = u^{cal} - u^{obs}. \tag{20}$$

Zeroing the gradient of L with respect to λ leads to the state equation. Finally, zeroing the gradient of L with respect to m, leads to the gradient of the misfit function $\frac{\partial J}{\partial m}$. For wavefield progagation, the state equation $F(u,m)$ represents the wave equation, m is the model parameter or parameters and state variable $u(x,t)$ is the modeled wavefield displacement. The adjoint-state equation (26) discribes the back propagation of the residuals, $u^{cal} - u^{obs}$.

Adjoint stereotomography

Tavakoli et.al (2016) presented a new slope tomography method using the adjoint state method and a reduced data and model space. By computing traveltimes through eikonal solvers from known source and receiver positions, X_s and X_g are eliminated from the data space and $\theta_s, \theta_r, T_s, T_r$ are eliminated from the model space (Fig.7). Following the



Model space : $[X_{j=1,N}],[V]_{i=1,M}$]

Data space : $\left[T_{sr}, P_s, P_g \right]_{j=1,N}$

Inversion : $\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \frac{\partial J}{\partial \mathbf{m}}$

Figure 7: The model space of adjoint stereotomography consists of the scatter position X and velocity model V. The data space consists of reciprocal parameter picks, T_{sr}, P_s, P_g. Adjoint stereotomography updates the model space **m** by cocomputing the gradients $\partial J/\partial m$ and the step length α.

adjoint-state method procedure, the state variables are $T_{sr}, P_s, P_g, t_s \; and \; t_r$. The corresponding adjoint state variables are $\mu_{s,r}, \xi_s, \xi_g, \lambda_s \; and \; \lambda_r$. The gradient of the velocity and scatter position can be computed using equation (21) and (22):

$$\frac{\partial J}{\partial V_m} = -\sum_{s=1}^{N_s} \frac{\lambda_s(x)}{V_m^3} - \sum_{r=1}^{N_r} \frac{\lambda_r(x)}{V_m^3}, \tag{21}$$

$$\frac{\partial J}{\partial X_{n_{s,r}}} = \mu_{s,r,n_{s,r}} \left(\frac{\partial Q_{n_{s,r}}}{\partial X_{n_{s,r}}}(t_s + t_r)\right) + \frac{\xi_{s,n_{s,r}}}{2\Delta s}\left(\frac{\partial Q_{n_{s,r}}}{\partial X_{n_{s,r}}}(t_{s+1} - t_{s-1})\right) + \frac{\xi_{r,n_{s,r}}}{2\Delta r}\left(\frac{\partial Q_{n_{s,r}}}{\partial X_{n_{s,r}}}(t_{r+1} - t_{r-1})\right), \tag{22}$$

where m is the index for the velocity model, $\lambda_s(x)$ and $\lambda_r(x)$ are the adjoint state variables for source ray path traveltime t_s, receiver ray path traveltime t_r; $\left(\frac{\partial Q_{n_{s,r}}}{\partial X_{n_{s,r}}}(t_s + t_r)\right)$ is the windowed and weighted slowness vector and is normal to the wavefront and $\left(\frac{\partial Q_{n_{s,r}}}{\partial X_{n_{s,r}}}(t_{s+1} - t_{s-1})\right)$ and $\left(\frac{\partial Q_{n_{s,r}}}{\partial X_{n_{s,r}}}(t_{r+1} - t_{r-1})\right)$ are gradient vectors normal to rays connecting source and receivers to scatters.

$\lambda_s(x)$ and $\lambda_r(x)$ are solutions to the adjoint state equations (23) and (24):

$$\frac{\partial}{\partial x}\frac{\partial T_s}{\partial x}\,\lambda_s + \frac{\partial}{\partial z}\frac{\partial T_s}{\partial z}\,\lambda_s = -\sum_{r=1}^{N_r^s}\sum_{n_{s,r=1}}^{N_n^{s,r}}\mu_{s,r,n_{s,r}} + \sum_{r=1}^{N_r^{s+1}}\sum_{n_{s+1,r=1}}^{N_n^{s+1,r}}\frac{\xi_{s+1}}{2\Delta s} - \sum_{r=1}^{N_r^{s-1}}\sum_{n_{s-1,r=1}}^{N_n^{s-1,r}}\frac{\xi_{s-1}}{2\Delta s} \quad , \quad (23)$$

$$\frac{\partial}{\partial x}\frac{\partial T_s}{\partial x}\,\lambda_s + \frac{\partial}{\partial z}\frac{\partial T_s}{\partial z}\,\lambda_s = -\sum_{r=1}^{N_r^{s}}\sum_{n_{s,r=1}}^{N_n^{s,r}}\mu_{s,r,n_{s,r}} + \sum_{r=1}^{N_r^{s+1}}\sum_{n_{s+1,r=1}}^{N_n^{s+1,r}}\frac{\xi_{s+1}}{2\Delta s} - \sum_{r=1}^{N_r^{s-1}}\sum_{n_{s-1,r=1}}^{N_n^{s-1,r}}\frac{\xi_{s-1}}{2\Delta s} \quad , \quad (24)$$

where $\mu_{s,r,n_{s,r}}$, ξ_s and ξ_r and the adjoint state variable for traveltime, t_{sr} and source and geophone slope p_s and p_g, and can be computed directly using equation (25) to (27):

$$\mu_{s,r,n_{s,r}} = \frac{1}{2\sigma_{T_{sr}}^2}\left(T_{s,r,n_{s,r}} - T^*_{s,r,n_{s,r}}\right), \quad (25)$$

$$\xi_{s,n_{s,r}} = \frac{1}{2\sigma_{P_s}^2}\left(P_{s,n_{s,r}} - \quad P^*_{s,n_{s,r}}\right) \quad , \quad (26)$$

$$\xi_{r,n_{s,r}} = \frac{1}{2\sigma_{P_r}^2}\left(P_{r,n_{s,r}} - \quad P^*_{r,n_{s,r}}\right), \quad (27)$$

where $n_{s,r}$ denotes scatter point positions, * denotes observed parameters, $\sigma_{T_{sr}}, \sigma_{p_s}$ and σ_{p_r} are data covarance for the input data.

The right-hand sides of equation (23) and (24) are the residual errors in traveltimes and slopes; therefore, equations (23) and (24) back-propagates the residual errors in traveltimes and slopes from scatter positions to the source and receiver locations respectively. As shown in equation (21), the summation of $\lambda_s(x)$ and $\lambda_r(x)$ forms the kernel of the velocity gradient $\partial J/\partial v$. With the gradients The velocity and scatter positions V_m and X_m can be updated using Newton-based local optimization schemes:

$$V_{m_{k+1}} = V_{m_k} + \alpha_{V_k}\left(\frac{\partial^2 J}{\partial V_{m_k}^2}\right)^{-1}\left(\frac{\partial J}{\partial V_{m_k}}\right) \quad (28)$$

$$X_{n_{k+1}} = X_{n_k} + \alpha_{X_k}\left(\frac{\partial^2 J}{\partial X_{n_k}^2}\right)^{-1}\left(\frac{\partial J}{\partial X_{n_k}}\right) \quad (29)$$

where α_{V_k} and α_{X_k} are the step lengths for the velocity and scatter position updates at iteration k.

Implementaion procedure

Similar to the classical stereotomography (Billette and Lambare, 1998; Billette et.al,2003), velocity estimation includes the initialization, localization and the inversion steps.

Adjoint stereotomogaphy algorithm

1: Multiscale loop
- Start with large grid spacing, progress to smaller grid spacing
- Use velocity model from previous multiscale iteration and resample to new smaller grid spacing

2: Input initial velocity model, measured state variables, source and receiver positions, slopes and traveltime data.

3: Initialization:

- Compute initial estimates of ray segment parameters (X_R, Z_R, T_s, T_r) using straight ray and constant velocity assumptions.

4: Localization:

- Loop on number of input data
  - Refine X_R and Z_R by minimizing the difference between measured and modelled P_s, P_g and T_{sr}.

5: Inversion:

- Loop on number of iterations
  - Compute modeled state variables T_s, T_r, P_s and P_g
  - Compute data residual and update cost function $J(m)$

  Loop on number of shot points
  - Compute adjoint state variables μ, ξ_s, ξ_r ,equation (25) to (27) and update $\frac{\partial J}{\partial X_n}$, equation (22)
  - Compute adjoint state variables λ_s , eqs (23)
  - Sum to $\frac{\partial J}{\partial V_m}$, equation (21)

  Loop on number of receivers
  - Compute λ_r , equation (24)
  - Sum to $\frac{\partial J}{\partial V_m}$, equation (22)

  Compute step lengths α_k, equation (28) and (29)
  Update model parameter m_{k+1}, equation (28) and (29)
- Next inversion iteration

-Next multi-scale iteration

Comparison of the slope tomography methods

CDR tomography, stereotomography, and adjoint stereotomography use slopes and traveltimes of locally coherent events to reconstruct the subsurface velocity. They differ on how the model space, data space are chosen and the inversion method used. Figure 8 summaries the differences among the three methods. CDR tomography reconstructs the source and receiver ray paths using source and receiver ray parameters picked from locally coherent events on unmigrated prestack seismic data. The model space is the velocity and the data space is the position errors at the endpoints of source and receiver ray pairs. Linear inversion using Fréchet derivative matrix is used to update the velocity. The inherent problem with this approach is that ray path errors can be caused by errors in the initial velocity model as well as the shooting angles computed from the ray parameters. Therefore, the CDR tomography is sensitive to picking errors in

the ray parameters and the solution can be unstable. Stereotomography remedies the above limitation in CDR tomography by shooting the shot and receiver ray pairs from a scatter point toward the source and receiver. Before the optimal solution for the velocity model and the ray segment parameters is achieved, the source ray parameter, geophone ray parameter, arrival time and end points of the source and receiver ray paths can be different from the values measured from the seismic data. Therefore, the model space for stereotomography includes the velocity model and the ray segment parameters. The ray segment parameters include the scatter point position, shooting angle and traveltime for the source and receiver ray paths. The data space of stereotomography includes the difference between the modeled and measured values of source and receiver coordinates, source and receiver ray parameters and the arrival time. Since both ray parameters and arrival time picked from seismic data are minimized in the optimization process, steretomography is less sensitive to picking errors. However, because of the expanded model and data space, stereotomography is computationally expensive. Adjoint stereotomography uses an eikonal solver to perform forward modeling; therefore, source and receiver coordinates, shooting angles and traveltimes are not required in the model space. Adjoint stereotomography retains the benefit of stereotomography by minimizing the difference between the modeled and measured values of the source and receiver ray parameters and arrival time. To address the computation efficiency, adjoint stereotomography uses the matrix free approach by using the adjoint-state method to compute the gradients for the model updates. With the adjoint-state method, the memory requirement is linearly proportional to the dimension of the velocity model size and the number of input picks.

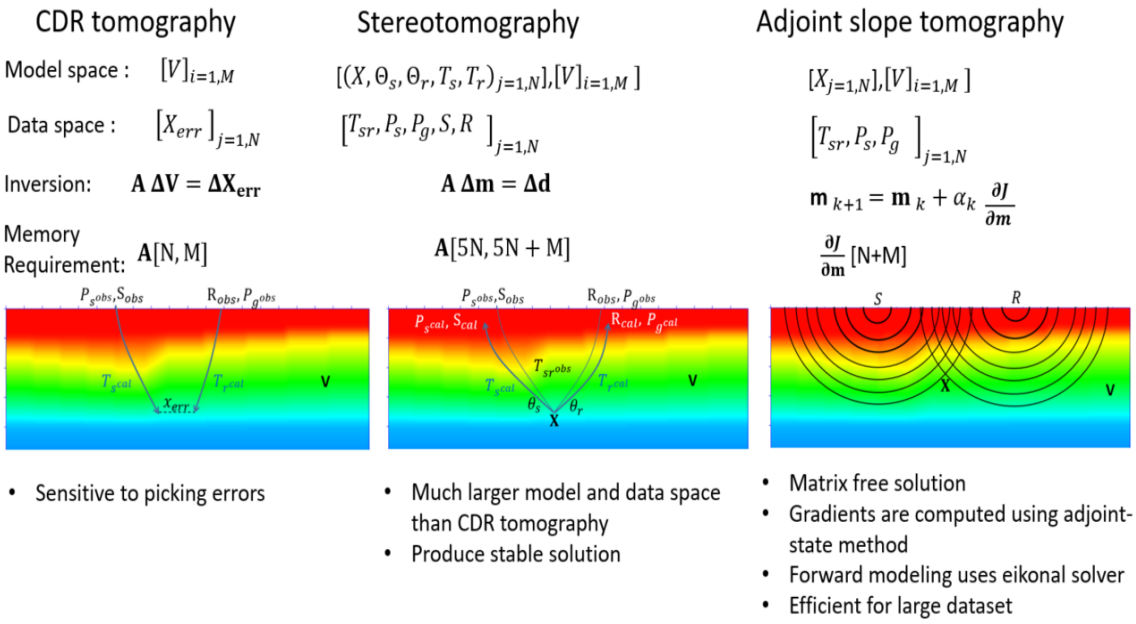


Figure 8: Comparison of CDR tomography, stereotomography and adjoint stereotomography

Numerical examples

Circular anomaly

To validate the adjoint stereotomography method, a velocity model with a linear vertical gradient and a circular anomaly (Fig. 9a and 9b) is created. Scatter points are placed around the velocity anomaly. 37 shots with shot spacing of 200m and maximum of 4000 m are used to compute the reciprocal parameters from these scatters (Fig. 9c). 1305 reciprocal parameter picks were created and used to compute the velocity gradients. Fig. 10 shows the solution of adjoint state variables $\lambda_s \; and \; \lambda_r$ for a single scatter. This demonstrates that the solutions of λ_s using equations (22) back propagates the source related residuals from the scatter point to the source location. Similarly, the solution of λ_r using equations (23) back propagates the receiver related residuals from the scatter point to the receiver location. A velocity model update is computed by summing $\lambda_s \; and \; \lambda_r \; from \; all \; picks$ and scaling the result by $\frac{1}{v(x)^3}$. The model update is done

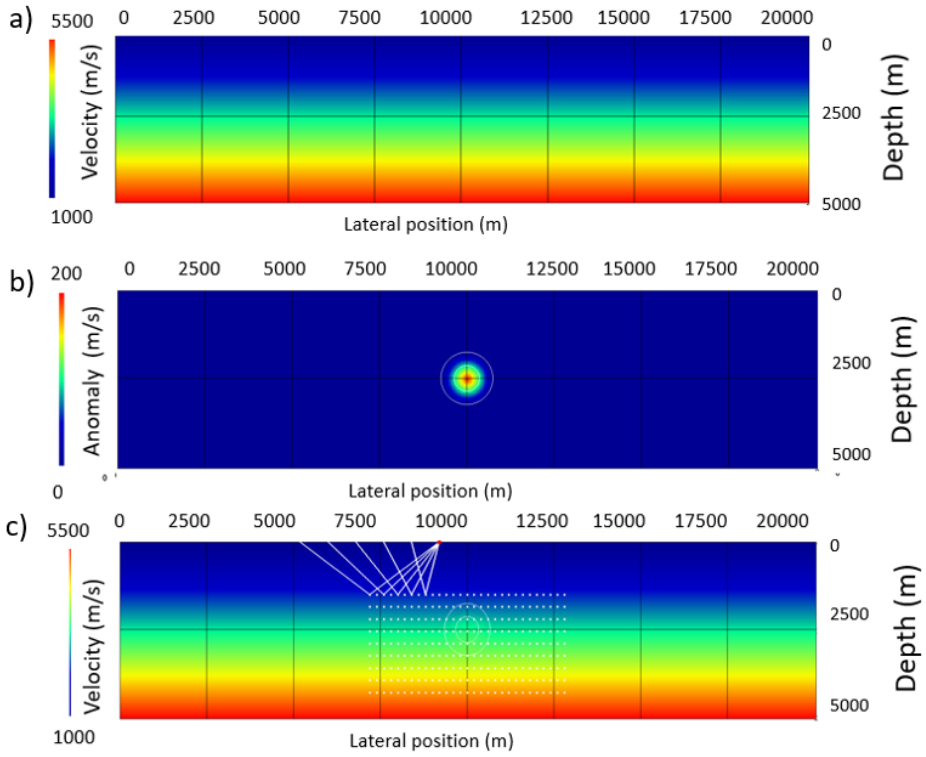


Figure 9. (a) Linear velocity gradient, (b) circular velocity anomaly, (c) combined velocity model, scatters and receiver spread for one shot.

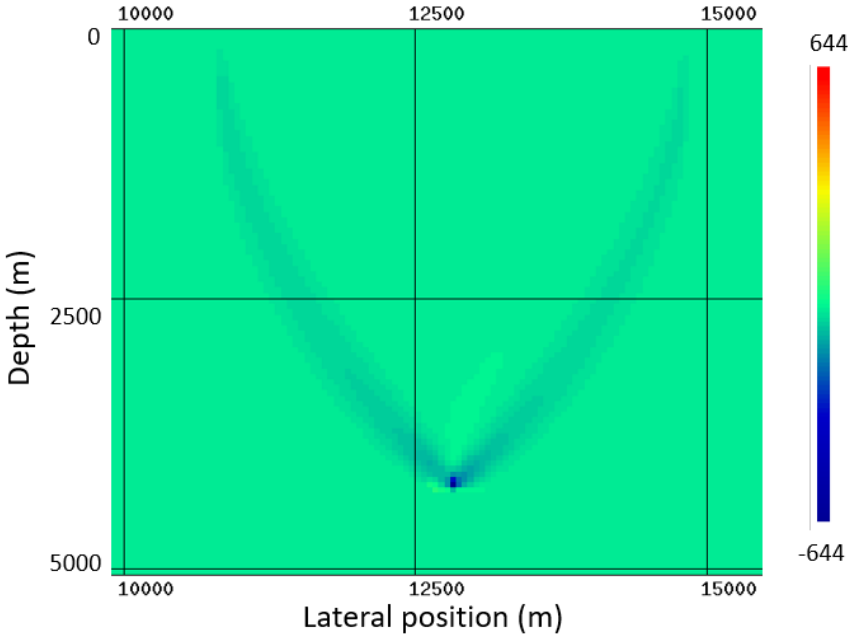


Figure 10. A solution of adjoint state variables λ_s and λ_r for a single scatter with a source located at 14800m. and a receiver located at 10800m.

iteratively by first updating the scatter positions and then the velocity. Figure 11a shows the initial velocity and the initial estimates of the scatter position using straight ray and constant velocity assumption. Figure 11b shows the updated scatter positions and velocity update after 40 iterations. This test result shows that the adjoint stereotomography algorithm implemented in this study can restore the scatter positions, the location and the shape of the velocity anomaly. However, there are artifacts below the main anormaly that suggest potential cross-talk from the scatter position. Another test is performed using the actual scatter positions to compute the velocity update (Fig. 12a). The velocity update (Fig. 12b) from this test does not show any artifact. Figure 13 compares the velocity updates from these two tests to the actual velocity anomaly.

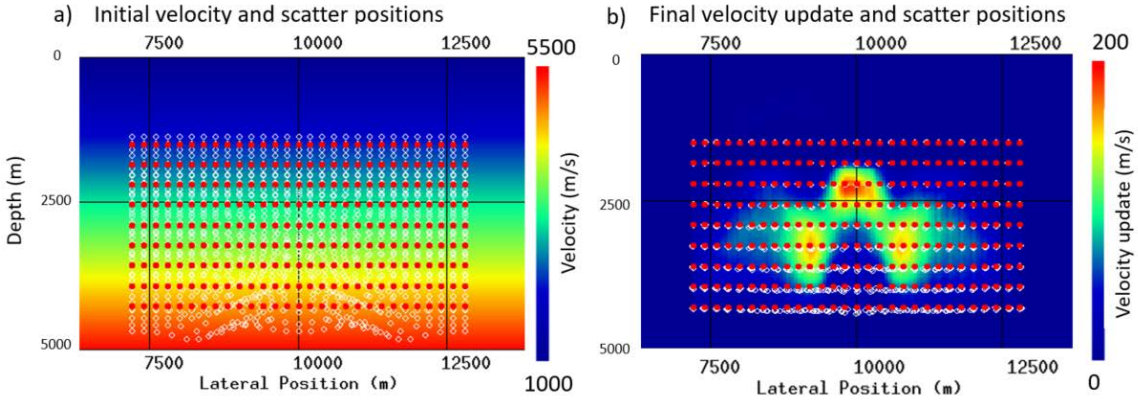


Figure 11. (a) True scatter positions (red dots) and initial scatter positions (white circles) computed from the CDR method, (b) Final velocity update and scatter positions (white circles) after 40 iterations.

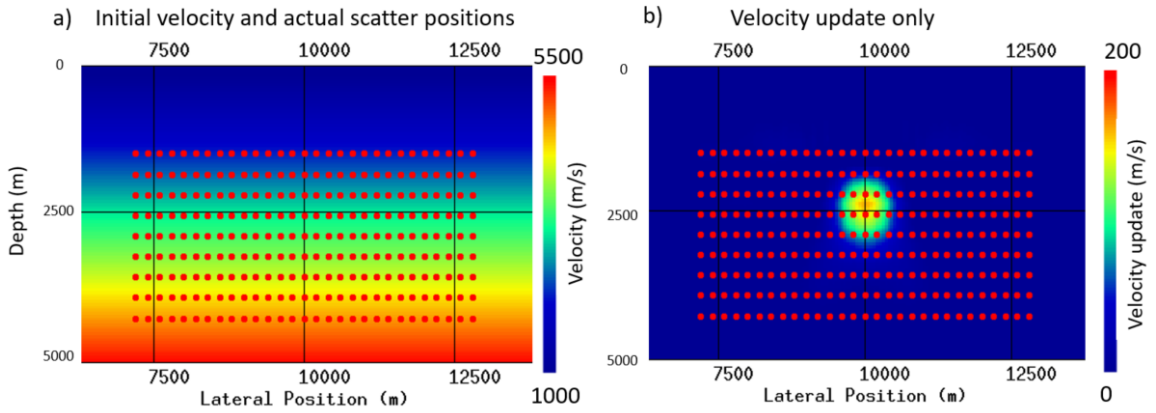


Figure 12. (a) True scatter positions (red dots) are used to compute the velocity updates, (b) Final velocity update after 40 iterations.

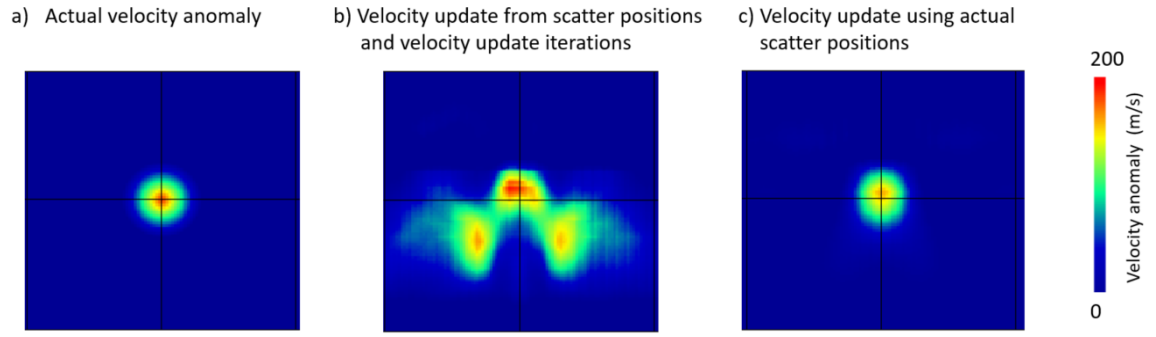


Figure 13. (a) Actual velocity anomaly, (b) Velocity update from scatter positions and velocity update iterations, (c) Velocity update using actual scatter positons.

Marmousi model

To further validate the implementation of the adjoint stereotomography algorithm, horizons are picked on the Marmousi model (Fig. 13a). Scatter points are created every 100m along the horizons. Shots with a shot spacing of 100m, geophone spacing of 40m and a maximum offset of 3400 m are used to compute the reciprocal parameters from these scatters. 1700 reciprocal parameter picks were created and used to compute the velocity updates. The relocation of the scatter positions computed from CDR equations failed to converge; therefore the actual scatter positions were used in this test. The velocty model after 5 iterations is shown in Fig. 13b. The solution from adjoint stereotomography matches the long wavelength trend of the actual velocity model.

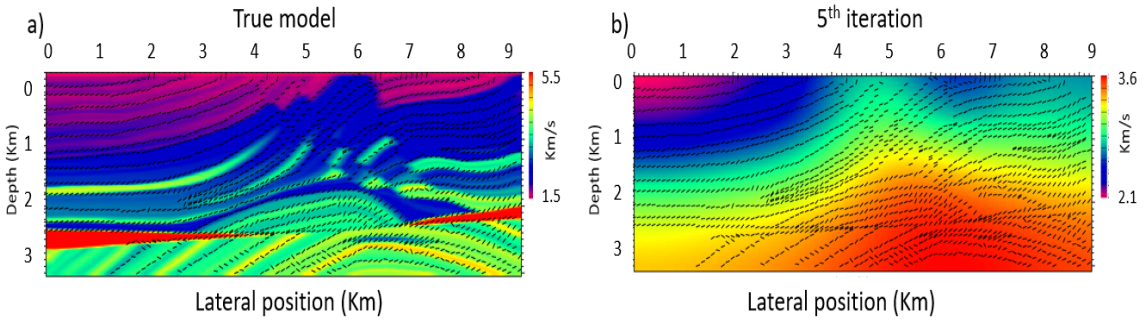


Figure 13. (a) True velocity model and dip bars computed from CDR picks and CDR equations, (b) Final velocity after 5 iterations using actual scatter positions.

CONCLUSION

We reviewed the CDR tomography, stereotomography, and adjoint stereotomography methods. We find the adjoint stereotomogaphy method can retain the stability of the stereotomography while improve the computation efficiency using the adjoint-state method. We used a simple velocity model with constant vertical gradient and circular velocity anomaly as well as the Marmousi model to verify the ability of adjoint stereotomography to recover velocity information using ray parameters and traveltime picks from unmigrated pre-stack seismic data. The test results show that the adjoint stereotomography implemented in this study can recover the long wavelength trend of the simple velocity model with a circular anormaly and the more complex Marmousi model. However, it fails to recover the scatter positions for the Marmousi model. Furthermore, velocity artifiacts caused by the scatter position errors are observed in the simple circular velocity anomaly model. Future work includes further inverstigation of more robust estimation of the scatter positions and simultaneous estimation of the step lengths for scatter positon and velocity updates using Newton-based method.

ACKNOWLEDGEMENTS

This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 46119-13.

REFERENCES

Aki, K., and P. Richards, 2001, Quantitative Seismology: University Science Books, Sausalito, California.

Bishop, T. N., K. P. Bube, R. T. Langan, P.L. Love, J. R. Resnick, T. T. Shuey, and D. A. Spinder, 1985, Tomographic determination of velocity and depth in laterally varying media: Geophysics, **50**, 903-923.

Billette, F., and G. Lambaré, 1998, Velocity macro-model estimation from seismic reflection data by stereotomography: Geophys. J. Int., **135**, 671-690.

Billette, F., S. Le B*é*gat, P. Podvin, and G. Lambar*é*, 2003, Practical aspects and applications of 2D stereotomography: Geophysics, **68**, 1008-1021.
Rieber, F, 1936, A new reflection system with controlled directional sensitivity: Geophysics, **1**, 97-106.
Riabinkin, L. A., 1957. Fundamentals of resolving power of controlled directional reception (CDR) of seismic waves, in Slant stack processing: 1991, SEG, Geophysics Reprint Series No.**14**, 36-60 (translated and paraphrased from Prikladnaya, **16**, 3-36).
Sword C. H., 1987, Tomographic determination of interval velocities from reflection seismic data: The method of controlled directional reception. PhD thesis, Stanford University.
Tavakoli, B. F., S. Operto, A. Ribodeti, J. Virieux, 2017, Slope tomography based on eikonal solvers and the adjoint-state method: Geophy, J. Int., **209**, 1629-1647.

APPENDIX A

Fréchet Derivatives for CDR Tomography

The major step in solving equation (10) is building the Fréchet derivatives $\frac{\partial X_{err}(j)}{\partial v_i}$, which can be rewritten as:

$$\frac{\partial X_{err}(j)}{\partial v_i} = \frac{\partial X_{err}(j)}{\partial \Delta X_i}\frac{\partial \Delta X_i}{\partial v_i} + \frac{\partial X_{err}(j)}{\partial \Delta t_i}\frac{\partial \Delta t_i}{\partial v_i} \quad . \tag{A1}$$

The first term in equation (A1) can be defined by considering the ray path geometry in figure 2 and definition of $Xerr$ in equation (7),

$$\frac{\partial X_{err}(j)}{\partial \Delta X_i} = -1 \quad \text{for source ray path,}$$
$$\frac{\partial X_{err}(j)}{\partial \Delta X_i} = \quad 1 \quad \text{for receive ray path.}$$

To derive the third term $\frac{\partial X_{err}(j)}{\partial \Delta t_i}$, we rewrite it as:

$$\frac{\partial X_{err}(j)}{\partial \Delta t_i} = \frac{dX_{err}(j)}{dZe}\frac{d\mathrm{Ze}}{d\Delta t_i} \quad , \tag{A2}$$

where Ze is the depth of the end points of the rays as show in figure 2, and

$$\frac{dX_{err}(j)}{dZe} = \frac{dX_{ge}}{dZe} - \frac{dX_{se}}{dZe} \quad \text{, and} \tag{A3}$$
$$\frac{dX_{err}(j)}{dZe} = \tan\theta_{ge} - \tan\theta_{se} \tag{A4}$$

By rewriting $\frac{d\mathrm{Ze}}{d\Delta t_i}$ as

$$\frac{d\mathrm{Ze}}{d\Delta t_i} = \frac{1}{dt/dZe} \quad \text{,and} \tag{A5}$$
$$\frac{dt}{dZe} = \frac{1}{v_{se}}\frac{1}{cos\theta_{se}} + \frac{1}{v_{ge}}\frac{1}{cos\theta_{ge}} \tag{A6}$$

Substitute equation (A4) and equation (A6) into equation (A2):

$$\frac{\partial X_{err}(j)}{\partial \Delta t_i} = -v_{se}v_{ge}\frac{\sin\theta_{ge}cos\theta_{se} - \sin\theta_{se}cos\theta_{ge}}{v_{se}cos\theta_{se} + v_{ge}cos\theta_{ge}} \tag{A7}$$

The second and fourth terms in equation (12) can be computed by perturbing the velocity model and compute the differences in ray path position and traveltime.

APPENDIX B

Fréchet Derivatives for Stereotomography

Equation (18) can be expanded to matrix form:

$$\boldsymbol{J}(\boldsymbol{m}) = \begin{bmatrix} \frac{\partial(\boldsymbol{X_s},\boldsymbol{P_s})}{\partial \boldsymbol{X}} & \frac{\partial(\boldsymbol{X_s},\boldsymbol{P_s})}{\partial \theta_s} & \frac{\partial(\boldsymbol{X_s},\boldsymbol{P_s})}{\partial \theta_g} & \frac{\partial(\boldsymbol{X_s},\boldsymbol{P_s})}{\partial T_s} & \frac{\partial(\boldsymbol{X_s},\boldsymbol{P_s})}{\partial T_g} & \frac{\partial(\boldsymbol{X_s},\boldsymbol{P_s})}{\partial C_i} \\ \frac{\partial(\boldsymbol{X_g},\boldsymbol{P_g})}{\partial \boldsymbol{X}} & \frac{\partial(\boldsymbol{X_g},\boldsymbol{P_g})}{\partial \theta_s} & \frac{\partial(\boldsymbol{X_g},\boldsymbol{P_g})}{\partial \theta_g} & \frac{\partial(\boldsymbol{X_g},\boldsymbol{P_g})}{\partial T_s} & \frac{\partial(\boldsymbol{X_g},\boldsymbol{P_g})}{\partial T_g} & \frac{\partial(\boldsymbol{X_g},\boldsymbol{P_g})}{\partial C_i} \\ \frac{\partial T_{sr}}{\partial \boldsymbol{X}} & \frac{\partial T_{sr}}{\partial \theta_s} & \frac{\partial T_{sr}}{\partial \theta_g} & \frac{\partial T_{sr}}{\partial T_s} & \frac{\partial T_{sr}}{\partial T_g} & \frac{\partial T_{sr}}{\partial C_i} \end{bmatrix}, \quad \text{(B1)}$$

and reduced to:

$$\boldsymbol{J}(\boldsymbol{m}) = \begin{bmatrix} \frac{\partial(\boldsymbol{X_s},\boldsymbol{P_s})}{\partial \boldsymbol{X}} & \frac{\partial(\boldsymbol{X_s},\boldsymbol{P_s})}{\partial \theta_s} & 0 & \frac{\partial(\boldsymbol{X_s},\boldsymbol{P_s})}{\partial T_s} & 0 & \frac{\partial(\boldsymbol{X_s},\boldsymbol{P_s})}{\partial C_i} \\ \frac{\partial(\boldsymbol{X_g},\boldsymbol{P_g})}{\partial \boldsymbol{X}} & 0 & \frac{\partial(\boldsymbol{X_g},\boldsymbol{P_g})}{\partial \theta_g} & 0 & \frac{\partial(\boldsymbol{X_g},\boldsymbol{P_g})}{\partial T_g} & \frac{\partial(\boldsymbol{X_g},\boldsymbol{P_g})}{\partial C_i} \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}. \quad \text{(B2)}$$

These derivatives can be derived from the ray equations and from paraxial ray tracing.

A ray Hamiltonian can be written as:

$$H = V^2\mathbf{p}^2 - 1 = 0, \quad \text{and} \quad \text{(B3)}$$

$$\mathbf{p} = \frac{\partial T}{\partial \mathbf{x}}, \quad \text{(B4)}$$

$$V = \frac{1}{|\mathbf{p}|} \quad \text{(B5)}$$

where $\mathbf{p}$ is the slowness vector, T is the traveltime, $\mathbf{x}$ is the position vector and V is the phase velocity. Defining $y = \begin{pmatrix} \mathbf{x} \\ \mathbf{p} \end{pmatrix}$, the Hamiltonian ray equations for can be written as:

$$\frac{\partial y}{\partial t} = \frac{\partial}{\partial t}\begin{pmatrix} x \\ \mathrm{p} \end{pmatrix} = \begin{pmatrix} \nabla_p H \\ -\nabla_x H \end{pmatrix}, \quad \text{(B6)}$$

Fréchet derivatives can be derived from the paraxial approximation (Farra & Madariaga 1987):

$$\delta y(t) = P(t,t_0)\delta y_0 + \int_{t_0}^{t} P(t,t')B\left(\delta C\left(x(t')\right)\right)dt' + \begin{pmatrix} \nabla_p H \\ -\nabla_x H \end{pmatrix} \quad \text{(B7)}$$

The third term is the ray equations and represents $\frac{\partial(\boldsymbol{X_s},\boldsymbol{P_s})}{\partial T_s}$, and can be used to evaluate the perturbation due to the perturbation in time:

$$\frac{\partial(\boldsymbol{X_s},\boldsymbol{P_s})}{\partial T_s} = \begin{pmatrix} \nabla_p H \\ -\nabla_x H \end{pmatrix} = \begin{pmatrix} \frac{\partial(V^2\boldsymbol{P_s^2})}{\partial \boldsymbol{P_s}} \\ -\frac{\partial(V^2\boldsymbol{P_s^2})}{\partial \boldsymbol{X_s}} \end{pmatrix} = \begin{pmatrix} 2V^2\boldsymbol{P_s} \\ -\frac{2}{\mathrm{v}}\frac{\partial V}{\partial \boldsymbol{X_s}} \end{pmatrix}, \tag{B8}$$

The first term is the initial perturbation in ray parameters $\delta\boldsymbol{y}(t_0)$ and can be decomposed into $\delta\theta$ and $\delta\boldsymbol{X}$:

$$\delta\boldsymbol{y}(t_0) = \begin{pmatrix} 0 \\ \hat{p} \end{pmatrix}\delta\theta + \begin{pmatrix} \mathrm{I}_2 \\ -\frac{\partial(V^2\boldsymbol{P_s^2})}{\partial \boldsymbol{X_s}} \end{pmatrix} = \begin{pmatrix} 2V^2\boldsymbol{P_s} \\ -\frac{2}{\mathrm{v}}\frac{\partial V}{\partial \boldsymbol{X_s}} \end{pmatrix}, \tag{B9}$$

From $\boldsymbol{P_s} = [sin\theta_s/V(x,z), \quad -cos\theta_s/V(x,z)]$, $\frac{\partial \boldsymbol{P_s}}{\partial \boldsymbol{X}}$, $\frac{\partial \boldsymbol{P_s}}{\partial \theta_s}$ and $\frac{\partial \boldsymbol{P_s}}{\partial C_i}$ in the following can be evaluated directly:

$$\frac{\partial(\boldsymbol{X_s},\boldsymbol{P_s})}{\partial \boldsymbol{X}} = [I, \quad \frac{\partial \boldsymbol{P_s}}{\partial \boldsymbol{X}}], \tag{B10}$$

$$\frac{\partial(\boldsymbol{X_s},\boldsymbol{P_s})}{\partial \theta_s} = [0, \quad \frac{\partial \boldsymbol{P_s}}{\partial \theta_s}], \tag{B11}$$

$$\frac{\partial(\boldsymbol{X_s},\boldsymbol{P_s})}{\partial C_i} = [0, \quad \frac{\partial \boldsymbol{P_s}}{\partial C_i}]. \tag{B12}$$

The similar can be derived for the geophone terms.

APPENDIX C

Adjoint stereotomography

Following the adjoint-state method procedure, the state variables, model parameters, state equations, adjoint-state variable, misfit function and augmented misfit function of the new methods are defined in equation (C1) to (C11):

State variables:

$$\boldsymbol{d} = (T_{sr}, P_s, P_r, t_s, t_r, t_s(x_s), t_r(x_r)). \tag{C1}$$

Model parameters:

$$\boldsymbol{m} = (\,(v_m)_{m=1}^{M}, (X_n)_{n=1}^{N}\,). \tag{C2}$$

State equations:

$$T_{sr} - Q_{n_{s,r}}(t_s + t_r) = 0, \tag{C3}$$

$$P_s - \frac{Q_{n_{s,r}} T_{s+1} - Q_{n_{s_r}} T_{s-1}}{2\Delta S} = 0\,, \tag{C4}$$

$$P_r - \frac{Q_{n_{s,r}} T_{r+1} - Q_{n_{s_r}} T_{r-1}}{2\Delta r} = 0\,, \tag{C5}$$

$$|\nabla t_s(x)|^2 - \frac{1}{v(x)^2} = 0, \tag{C6}$$

$$|\nabla t_r(x)|^2 - \frac{1}{v(x)^2} = 0, \tag{C7}$$

$$t_s(x_s) = 0, \tag{C8}$$

$$t_r(x_r) = 0. \tag{C9}$$

Misfit function:

$$\begin{aligned} J(m) = {} & \frac{1}{2\sigma_{T_{sr}}^2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r^s} \sum_{n_{s,r=1}}^{N_n^{s,r}} \left(T_{s,r,n_{s,r}}(m) - T^*_{s,r,n_{s,r}}\right)^2 + \\ & \frac{1}{2\sigma_{P_s}^2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r^s} \sum_{n_{s,r=1}}^{N_n^{s,r}} \left(P_{s,n_{s,r}}(m) - P^*_{s,n_{s,r}}\right)^2 + \\ & \frac{1}{2\sigma_{P_r}^2} \sum_{r=1}^{r} \sum_{s=1}^{N_s^r} \sum_{n_{s,r=1}}^{N_n^{s,r}} \left(P_{r,n_{s,r}}(m) - P^*_{r,n_{s,r}}\right)^2, \end{aligned} \tag{C10}$$

Lagrangian misfit function:

$$\begin{aligned} L = {} & h(u.m) - \sum_{s=1}^{N_s} \sum_{r=1}^{N_r^s} \sum_{n_{s,r=1}}^{N_n^{s,r}} \mu_{s,r,n_{s,r}} \left(T_{s,r,n_{s,r}} - Q_{n_{s,r}} t_s - Q_{n_{s,r}} t_r\right) - \\ & \sum_{s=1}^{N_s} \sum_{r=1}^{N_r^s} \sum_{n_{s,r=1}}^{N_n^{s,r}} \xi_{s,n_{s,r}} \left(P_{s,n_{s,r}} - \frac{Q_{n_{s,r}} T_{s+1} - Q_{n_{s,r}} T_{s-1}}{2\Delta S}\right) - \\ & \sum_{r=1}^{N_r} \sum_{s=1}^{N_s^r} \sum_{n_{s,r=1}}^{N_n^{s,r}} \xi_{r,n_{s,r}} \left(P_{r,n_{s,r}} - \frac{Q_{n_{s,r}} T_{r+1} - Q_{n_{s,r}} T_{r-1}}{2\Delta r}\right) - \\ & \frac{1}{2} < \lambda_s(x), |\nabla T_s(x)|^2 - \frac{1}{v(x)^2} > \Omega - \\ & \frac{1}{2} < \lambda_r(x), |\nabla T_r(x)|^2 - \frac{1}{v(x)^2} > \Omega - \sum_{s=1}^{N_s} \psi_s t_s(x_s) - \sum_{r=1}^{r} \psi_r t_r(x_r) - \\ & \sum_{s=1}^{N_s} \psi_s t_s(x_s) - \sum_{r=1}^{r} \psi_r t_r(x_r)\,, \end{aligned} \tag{C11}$$

where $\frac{1}{\sigma_{T_{s,r}}}, \frac{1}{\sigma_{P_s}}$, and $\frac{1}{\sigma_{P_r}}$ are the standard deriations of $T_{s,r}, P_s. P_r.$ T_{sr}, P_s, P_r, t_s and t_r are the state variables, $\mu, \xi_s\,, \xi_r, \lambda_s, \lambda_r, \psi_s \; and \; \psi_r$ are corresponding adjoint-sate variables. Ω denotes subsruface domain. $Q_{n_{s,r}}$ is windowing function mapping computed traveltimes from discreted grid locations to scatter positions.

Zeroing the gradient of the Lagrangian:

Zeroing the gradient of the Lagrangian with respect to the model parameters $V\ and\ X$ results in the respective gradient equations:

$$\frac{\partial J}{\partial V_m} = \frac{\partial L}{\partial V_m} = -\frac{1}{2}\frac{\partial}{\partial V_m}\Big(< \lambda_s(x), |\nabla T_s(x)|^2 - \frac{1}{v(x)^2} >_\Omega \Big) - \frac{1}{2}\frac{\partial}{\partial V_m}\Big(< \lambda_r(x), |\nabla T_r(x)|^2 - \frac{1}{v(x)^2} >_\Omega \Big),$$

$$\frac{\partial J}{\partial V_m} = -\sum_{s=1}^{N_s} \frac{\lambda_s(x)}{V_m^3} - \sum_{r=1}^{N_r} \frac{\lambda_r(x)}{V_m^3} \quad \text{(C12)}$$

$$\frac{\partial J}{\partial X_{n_{s,r}}} = \frac{\partial L}{\partial X_{n_{s,r}}} = \frac{\partial}{\partial X_{n_{s,r}}}\Big(-\mu_{s,r,n_{s,r}}\Big(T_{s,r,n_{s,r}} - Q_{n_{s,r}}(t_s + t_r)\Big)\Big) - \frac{\partial}{\partial X_n}\Big(\xi_{s,n_{s,r}}\Big(P_{s,n_{s,r}} - \frac{Q_{n_{s,r}}T_{s+1} - Q_{n_{s,r}}T_{s-1}}{2\Delta S}\Big)\Big) - \frac{\partial}{\partial X_n}\Big(\xi_{r,n_{s,r}}\Big(P_{r,n_{s,r}} - \frac{Q_{n_{s,r}}T_{r+1} - Q_{n_{s,r}}T_{r-1}}{2\Delta r}\Big)\Big),$$

$$\frac{\partial J}{\partial X_{n_{S,r}}} = \mu_{s,r,n_{s,r}}\left(\frac{\partial Q_{n_{s,r}}}{\partial X_{n_{s,r}}}(t_s + t_r)\right) + \frac{\xi_{s,n_{s,r}}}{2\Delta s}\left(\frac{\partial Q_{n_{s,r}}}{\partial X_{n_{s,r}}}(t_{s+1} - t_{s-1})\right) + \frac{\xi_{r,n_{s,r}}}{2\Delta r}\left(\frac{\partial Q_{n_{s,r}}}{\partial X_{n_{s,r}}}(t_{r+1} - t_{r-1})\right), \quad \text{(C13)}$$

where $\left(\frac{\partial Q_{n_{s,r}}}{\partial X_{n_{s,r}}}(t_s + t_r)\right)$ is the windowed and weighted slowness vector and is normal to the wavefront and $\left(\frac{\partial Q_{n_{s,r}}}{\partial X_{n_{s,r}}}(t_{s+1} - t_{s-1})\right)\ and\ \left(\frac{\partial Q_{n_{s,r}}}{\partial X_{n_{s,r}}}(t_{r+1} - t_{r-1})\right)$ are gradient vectors normal to rays connecting source and receivers to scatters.

The adjoint state variables $\lambda_s\ and\ \lambda_r$ are needed to compute $\frac{\partial J}{\partial v_m}$; while the adjoint state variables $\mu_{s,r,n_{s,r}}, \xi_{s,n_{s,r}}\ and\ \xi_{r,n_{s,r}}$ are needed to compute $\frac{\partial J}{\partial X_n}$.

$\boldsymbol{V_m}$ and $\boldsymbol{X_n}$ can be updated using Newton-based local optimization schemes:

$$V_{m_{k+1}} = V_{m_k} + \alpha_{V_k}\left(\frac{\partial^2 J}{\partial V_{m_k}^2}\right)^{-1}\left(\frac{\partial J}{\partial V_{m_k}}\right) \quad \text{(C14)}$$

$$X_{n_{k+1}} = X_{n_k} + \alpha_{X_k}\left(\frac{\partial^2 J}{\partial X_{n_k}^2}\right)^{-1}\left(\frac{\partial J}{\partial X_{n_k}}\right) \quad \text{(C15)}$$

Zeroing the gradient of the Lagrangian with respect to the state variables results in the following adjoint state equations:

$$\frac{1}{2\sigma_{T_{sr}}^2}\left(T_{s,r,n_{s,r}} - T^*_{s,r,n_{s,r}}\right) - \mu_{s,r,n_{s,r}} = 0\ , \quad \text{(C16)}$$

$$\frac{1}{2\sigma_{P_s}^2}\left(P_{s,n_{s,r}} - P^*_{s,n_{s,r}}\right) - \xi_{s,n_{s,r}} = 0\ , \quad \text{(C17)}$$

$$\frac{1}{2\sigma_{P_r}^2}\left(P_{r,n_{s,r}} - P^*_{r,n_{s,r}}\right) - \xi_{r,n_{s,r}} = 0\ , \quad \text{(C18)}$$

$$\psi_s = 0\ , \quad \text{(C19)}$$

$$\psi_r = 0\ , \quad \text{(C20)}$$

$$\sum_{r=1}^{N_r^s}\ \sum_{n_{s,r=1}}^{N_n^{s,r}} Q_{n_{s,r}}^{t}\ \mu_{s,r,n_{s,r}} - \frac{1}{2\Delta s}\sum_{r=1}^{N_r^{s+1}}\ \sum_{n_{s+1,r=1}}^{N_n^{s+1,r}} Q_{n_{s+1,r}}^{t}\xi_{s+1,n_{s+1,r}} -$$
$$\frac{1}{2\Delta s}\sum_{r=1}^{N_r^{s-1}}\ \sum_{n_{s-1,r=1}}^{N_n^{s-1,r}} Q_{n_{s-1,r}}^{t}\xi_{s-1,n_{s-1,r}} -$$
$$\frac{\partial}{\partial T_s}\left(\frac{1}{2}\sum_{s=1}^{N_s} < \lambda_r(x), |\nabla T_r(x)|^2 - \frac{1}{v(x)^2} >_{\Omega}\right) = 0, \qquad \text{(C21)}$$

$$\sum_{s=1}^{N_s^r}\ \sum_{n_{s,r=1}}^{N_n^{s,r}} Q_{n_{s,r}}^{t}\ \mu_{s,r,n_{s,r}} - \frac{1}{2\Delta r}\sum_{s=1}^{N_s^{r+1}}\ \sum_{n_{s,r+1=1}}^{N_n^{s,r+1}} Q_{n_{s,r+1}}^{t}\xi_{r+1,n_{s,r+1}} -$$
$$\frac{1}{2\Delta r}\sum_{s=1}^{N_s^{r-1}}\ \sum_{n_{s,r-1=1}}^{N_n^{s,r-1}} Q_{n_{s,r-1}}^{t}\xi_{r-1,n_{s,r-1}} -$$
$$\frac{\partial}{\partial T_r}\left(\frac{1}{2}\sum_{r=1}^{N_r} < \lambda_r(x), |\nabla T_r(x)|^2 - \frac{1}{v(x)^2} >_{\Omega}\right) = 0\,. \qquad \text{(C22)}$$

The last term in equation (C21) and (C22) can be expanded using multivariate integration by parts and divergence therorem and results in:

$$\sum_{r=1}^{N_r^s}\ \sum_{n_{s,r=1}}^{N_n^{s,r}} Q_{n_{s,r}}^{t}\ \mu_{s,r,n_{s,r}} - \frac{1}{2\Delta s}\sum_{r=1}^{N_r^{s+1}}\ \sum_{n_{s+1,r=1}}^{N_n^{s+1,r}} Q_{n_{s+1,r}}^{t}\xi_{s+1,n_{s+1,r}} -$$
$$\frac{1}{2\Delta s}\sum_{r=1}^{N_r^{s-1}}\ \sum_{n_{s-1,r=1}}^{N_n^{s-1,r}} Q_{n_{s-1,r}}^{t}\xi_{s-1,n_{s-1,r}} -$$
$$\sum_{s=1}^{N_s}\left(\left(\nabla\cdot\left(\lambda_s(x)\nabla T_s(x)\right)\right)_{\Omega} - \left(\lambda_s(x)\nabla T_s(x)\cdot n\right)_{\Gamma}\right) = 0, \qquad \text{(C23)}$$

$$\sum_{s=1}^{N_s^r}\ \sum_{n_{s,r=1}}^{N_n^{s,r}} Q_{n_{s,r}}^{t}\ \mu_{s,r,n_{s,r}} - \frac{1}{2\Delta r}\sum_{s=1}^{N_s^{r+1}}\ \sum_{n_{s,r+1=1}}^{N_n^{s,r+1}} Q_{n_{s,r+1}}^{t}\xi_{r+1,n_{s,r+1}} -$$
$$\frac{1}{2\Delta r}\sum_{s=1}^{N_s^{r-1}}\ \sum_{n_{s,r-1=1}}^{N_n^{s,r-1}} Q_{n_{s,r-1}}^{t}\xi_{r-1,n_{s,r-1}} -$$
$$\sum_{r=1}^{N_r}\left(\left(\nabla\cdot\left(\lambda_r(x)\nabla T_r(x)\right)\right)_{\Omega} - \left(\lambda_r(x)\nabla T_r(x)\cdot n\right)_{\Gamma}\right) = 0, \qquad \text{(C24)}$$

where Γ denotes the boundary of Ω and n is the outward normal vector of Γ. With λ_s and λ_r being computed over the domain of Ω, the sum of n of Γ is zero; therefore the last term in equation (C23) and (C24) are zero and the adjoint-state equations for $\lambda_s\ and\ \lambda_r$ become:

$$\left(\nabla\cdot\left(\lambda_s(x)\nabla T_s(x)\right)\right)_{\Omega} = -\sum_{r=1}^{N_r^s}\ \sum_{n_{s,r=1}}^{N_n^{s,r}}\ \mu_{s,r,n_{s,r}} + \sum_{r=1}^{N_r^{s+1}}\ \sum_{n_{s+1,r=1}}^{N_n^{s+1,r}}\ \frac{\xi_{s+1}}{2\Delta s} - \sum_{r=1}^{N_r^{s-1}}\ \sum_{n_{s-1,r=1}}^{N_n^{s-1,r}}\ \frac{\xi_{s-1}}{2\Delta s} \quad ,\text{(C25)}$$

$$\left(\nabla\cdot\left(\lambda_r(x)\nabla T_r(x)\right)\right)_{\Omega} = -\sum_{s=1}^{N_s^r}\ \sum_{n_{s,r=1}}^{N_n^{s,r}}\ \mu_{s,r,n_{s,r}} + \sum_{s=1}^{N_s^{r+1}}\ \sum_{n_{s,r+1=1}}^{N_n^{s,r+1}}\ \frac{\xi_{r+1}}{2\Delta r} - \sum_{s=1}^{N_s^{r-1}}\ \sum_{n_{s,r-1=1}}^{N_n^{s,r-1}}\ \frac{\xi_{r-1}}{2\Delta r} \quad ,\text{(C26)}$$

The right-hand side of equation (C25) acts as the source terms at the scatter positions for all the receivers of the shot . The left-hand side of equation (C25) can be expanded as:

$\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial z}\right)\cdot\left(\lambda_s\left(\frac{\partial T_s}{\partial x}, \frac{\partial T_s}{\partial z}\right)\right) = \frac{\partial}{\partial x}\frac{\partial T_s}{\partial x}\lambda_s + \frac{\partial}{\partial z}\frac{\partial T_s}{\partial z}\lambda_s$, and equation (C25) becomes:

$$\frac{\partial}{\partial x}\frac{\partial T_s}{\partial x}\lambda_s + \frac{\partial}{\partial z}\frac{\partial T_s}{\partial z}\lambda_s = -\sum_{r=1}^{N_r^s}\ \sum_{n_{s,r=1}}^{N_n^{s,r}}\ \mu_{s,r,n_{s,r}} + \sum_{r=1}^{N_r^{s+1}}\ \sum_{n_{s+1,r=1}}^{N_n^{s+1,r}}\ \frac{\xi_{s+1}}{2\Delta s} - \sum_{r=1}^{N_r^{s-1}}\ \sum_{n_{s-1,r=1}}^{N_n^{s-1,r}}\ \frac{\xi_{s-1}}{2\Delta s} \quad , \qquad \text{(C27)}$$

Equation (C26) can be expanded in a similar way.

Following the fast marching method of Leung and Qian (2006), equation (C27) can be solved using fast marching method with the consideration of the characteristics of the propagation of the source terms from the scatter positions to the source.