Hybrid frequency domain full waveform inversion using ray+Born sensitivity kernels

Ramzi Djebbi[†] and Tariq Alkhalifah^{*}

ABSTRACT

Full waveform inversion (FWI) using the scattering integral (SI) approach is an explicit formulation of the inversion optimization problem. The inversion procedure is straightforward, and the dependence of the data residuals on the model parameters is clear. However, the biggest limitation associated with this approach is the huge computational cost in conventional exploration seismology applications. Modeling from each of the source and receiver locations is required to compute the update at every iteration, and that is prohibitively expensive, especially for 3-D problems. To deal with this issue, we propose a hybrid implementation of frequency-domain FWI, where forward modeling is combined with ray tracing to compute the update. We use the sensitivity kernels computed from dynamic ray-tracing to build the gradient. The data residual is still computed using finite-difference wavefield modeling. With ray theory, the Green's function can be approximated using a coarser grid compared to wave-equation modeling. Therefore, the memory requirements, as well as the computational cost, are largely reduced. Considering that in transmission FWI long-to-intermediate wavelengths are updated during the early iterations, we obtain accurate inverted models. The inversion scheme captured the anomaly embedded in the homogeneous background medium. For more complex models, the hybrid inversion method helps in improving the initial model with little cost compared to conventional SI inversion approaches. The accuracy of the inversion results shows the effectiveness of the hybrid approach for 3-D realistic problems.

INTRODUCTION

In full waveform inversion (FWI) (Lailly, 1983; Tarantola, 1984), the misfits between the observed and modeled data are minimized to update the earth model. However, this misfit, measured through the classic l_2 norm, suffers from many local minima due to cycle skipping between modeled and observed data (Virieux and Operto, 2009). One approach to handle this problem is to consider a frequency-continuation inversion scheme to mitigate sensitivity to the initial model and recover the long-to-intermediate wavelengths. The low frequencies are less sensitive to the initial model inaccuracies and are, thus, inverted first. Then, the inversion result is used as an initial model for higher frequencies (Bunks et al., 1995; Sirgue and Pratt, 2004). Conventionally, FWI is implemented using the adjoint-state method (Tarantola, 1984). It requires an equal number of forward and adjoint modeling steps to compute the gradient. The convergence properties of the least-squares inversion problem can be improved using the Hessian information. Fichtner and Trampert (2011) promote the use of the second-order adjoint-state equations to approximate the Hessian matrix.

Another FWI approach is the scattering integral (SI) method, where the gradient is computed explicitly by multiplying the Fréchet derivatives matrix (or the sensitivity kernels matrix) with the complex conjugate of data residuals (Chen et al., 2007; Liu et al., 2015). The gradient of the misfit function requires more modeling steps compared to the adjoint-state method. The total number of modeling steps needed is proportional to the number of sources and non-repeated receivers. Thus, for seismic exploration applications, the SI method is less relevant than the adjoint-state method. However, when dealing with approximations of the Hessian matrix, the scattering integral approach cost is comparable to the second-order adjoint-state method. Chen et al. (2007) compare the two inversion approaches and described situations where the SI outperforms the adjoint-state method. Liu et al. (2015) propose computing the SI gradient by a matrix decomposition method in which only the wavefields are stored. This method enables the SI method to work with reasonable memory storage. Djebbi and Alkhalifah (2019) present a 2-D anisotropic multi-parameter

Published in Geophysics, Volume 85, Issue 4, July 2020

[†]Formerly King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; presently PGS, Weybridge, UK

^{*}King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

inversion approach that uses a direct Helmholtz solver to reduce the computational cost considerably. For direct wave-equation solvers, the computational cost is mainly in the lower-upper (LU) decomposition of the impedance matrix. Additionally, the impedance matrix is source independent, which enables the same matrix decomposition to be used for all sources and receivers with a reasonable computational cost. For 3-D problems, the computational cost, as well as the memory requirements, increase exponentially. Iterative or hybrid direct-iterative solvers are instead used to solve the frequency-domain wave equation (Virieux and Operto, 2009); thus, solving the wave equation for a large number of sources becomes expensive, and limits the applicability of the SI method.

To deal with the large computational cost associated with 3-D problems, ray theory (Červený, 2001) has been widely used for imaging and inversion. Traveltimes and amplitudes, computed using dynamic ray tracing (Červený and Hron, 1980), has been used for 3-D true amplitude Kirchhoff migration (Bleistein, 1987; Schleicher et al., 1993; Alkhalifah and Bednar, 2000). A different class of methods is ray+Born migration/inversion (Thierry et al., 1999; Lambaré et al., 2003) where a ray tracing approximation of the Green's functions for a smooth background model is used to invert for image or velocity perturbations based on first-order Born approximation. Moreover, ray theory has been used for traveltime tomography (Dziewonski, 1984; Van Der Hilst et al., 1997; Nolet, 2012) and for wave-equation tomography (Hung et al., 2000; Montelli et al., 2004; Tian et al., 2007), where the finite-frequency sensitivity is considered. Liu et al. (2009) propose a combination of wavefield modeling and ray tracing for Fresnel zone tomography. The sensitivity kernels are calculated using frequency-domain modeling, and the traveltime residuals are computed using ray tracing. With this approach, the resolution of the inverted models is improved when compared to traveltime tomography.

In this paper, we propose a hybrid inversion method, for which we combine the efficiency of ray theory (Červený, 2001) and the flexibility of the SI method. As the model is usually updated from low to high wavenumbers, relying on smooth models at the beginning, we use ray tracing as an effective alternative to wavefield modeling to compute Born kernels. We denote the method as hybrid ray+Born inversion. This approach deals with the high computational cost and memory requirements for SI inversion in 3-D cases. Generally, with ray theory, a coarse grid can be used to compute the Green's functions and store it in memory. Then, during the calculation of the sensitivity kernels, traveltimes and amplitudes can be interpolated to the finer inversion grid. An additional benefit is that the availability of ray direction information, which can be used to compute the wavefield derivatives and help in the interpolation to a finer grid, in the framework of anisotropic multi-parameter inversion (Djebbi and Alkhalifah, 2013; Alkhalifah, 2016; Djebbi et al., 2017). The resulting models are limited by the smoothness requirements (for ray tracing) of the background models. Therefore, the inversion result can be used as good initial models for a subsequent FWI, or directly utilized for migration. We briefly describe the SI approach for inversion as well as the proposed hybrid method in the theory section. Then, in the numer-

ical examples section, we start with a transmission experiment to show the accuracy of the proposed approach. In a more realistic Marmousi example, the hybrid approach accurately inverts for the smooth low-frequency model. We also present the gradients computed for a 3-D anisotropic VTI model. The ray-parameter information deals with the angular dependence in the sensitivity kernels. Finally, we analyze the computational cost of the proposed method compared to conventional SI. The hybrid ray+Born inversion method provides an alternative cheaper method to wave equation based SI in building accurate FWI initial models.

THEORY

Full waveform inversion is based on the minimization of the residuals between the modeled data $\mathbf{U}(\mathbf{m})$ and the observed data $\mathbf{d}$. We minimize the l_2 norm misfit function,

$$E(\mathbf{m}) = \frac{1}{2} \Delta \mathbf{d}^T \Delta \mathbf{d}^* \quad (1)$$

where $\Delta \mathbf{d} = \mathbf{U}(\mathbf{m}) - \mathbf{d}$ is the data residuals vector and $\mathbf{m}$ is the model. The superscripts T and * denote the transpose and complex conjugate in the frequency domain, respectively.

The gradient of the misfit function with respect to the model parameters is given as,

$$\mathbf{g}(\mathbf{m}) = \frac{\partial E(\mathbf{m})}{\partial \mathbf{m}} = \Re(\mathbf{F}^T \Delta \mathbf{d}^*) \quad (2)$$

where $\Re$ represents the real part. $\mathbf{F}$ is an $(n \times m)$ matrix, representing the Fréchet derivatives where n and m are the data and model sizes, respectively. The elements of this matrix are given as,

$$k_{ij} = \frac{\partial U_i}{\partial m_j} \quad i = (1, \dots, n); \quad j = (1, \dots, m). \quad (3)$$

Each row of the Fréchet derivatives matrix $\mathbf{F}$ corresponds to a source-receiver pair and relates a perturbation in the recorded data to a perturbation in the earth model. The frequency domain Born approximation of the perturbed wavefield, due to a velocity perturbation of the form: $v = v_0 + \Delta v$, is given by,

$$\begin{aligned} \Delta d(\mathbf{x}_r, \mathbf{x}, \mathbf{x}_s, \omega) &= \int \frac{2\omega^2}{v_0^3} G_0(\mathbf{x}, \mathbf{x}_r, \omega) U_0(\mathbf{x}, \mathbf{x}_s, \omega) \Delta v d\mathbf{x} \\ &= \int \mathbf{K}^{(v)}(\mathbf{x}_r, \mathbf{x}, \mathbf{x}_s, \omega) \Delta v d\mathbf{x} \end{aligned} \quad (4)$$

where v_0 is the background velocity, Δv is the velocity perturbation, $G_0(\mathbf{x}, \mathbf{x}_r, \omega)$ and $U_0(\mathbf{x}, \mathbf{x}_s, \omega)$ are the receiver's Green's function and the source wavefield for the background velocity v_0 , respectively. The velocity Born kernel is denoted as $\mathbf{K}^{(v)}(\mathbf{x}_r, \mathbf{x}, \mathbf{x}_s, \omega)$.

The scattering integral method

The explicit computation of the Fréchet derivatives requires wavefields computed at the shot locations, as well as, the Green's functions from the receiver locations. Liu et al. (2015) propose to compute and save in memory the sources and receivers

Green's functions. Then, a multiplication of the Born kernels with the residuals can be formulated in a matrix-vector form to determine the gradient, as follows:

$$\begin{aligned} \mathbf{g}(\mathbf{m}) &= \begin{pmatrix} k_{11} & \dots & k_{i1} & \dots & k_{n1} \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ k_{1j} & \dots & k_{ij} & \dots & k_{nj} \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ k_{1m} & \dots & k_{im} & \dots & k_{nm} \end{pmatrix} \begin{pmatrix} \Delta d_1^* \\ \vdots \\ \Delta d_j^* \\ \vdots \\ \Delta d_n^* \end{pmatrix} \\ &= \sum_{i=0}^n \begin{pmatrix} k_{i1} \\ \vdots \\ k_{ij} \\ \vdots \\ k_{im} \end{pmatrix} \Delta d_i^* \end{aligned} \quad (5)$$

The gradient update direction $\mathbf{p}_1$ is obtained after multiplication with the step length γ estimated using a second-order approximation of the objective function $E_1(\gamma) = E(\mathbf{m} + \gamma \mathbf{p}_1)$ (Hu et al., 2011). Thus, the step size is given as:

$$\gamma = \frac{\mathbf{p}_1^\dagger \mathbf{p}_1}{\mathbf{p}_1^\dagger \mathbf{H}_a \mathbf{p}_1} = \frac{\mathbf{p}_1^\dagger \mathbf{p}_1}{(\mathbf{Fp})^\dagger (\mathbf{Fp})}. \quad (6)$$

The superscript † denotes the transpose complex conjugate. Depending on the optimization method, $\mathbf{p}$ can be either the gradient direction $\mathbf{p}_1$ or the preconditioned gradient direction $\mathbf{p}_2$. $\mathbf{H}_a = \mathbf{F}^\dagger \mathbf{F}$ is the approximate Hessian.

The truncated Gauss-Newton update direction $\mathbf{p}_2$ can also be estimated using the matrix-vector multiplication approach. In fact, after computing the steepest descent update direction $\mathbf{p}_1$, the system $\mathbf{H}_a \mathbf{p}_2 = \mathbf{p}_1$ is solved using a conjugate gradient method to estimate $\mathbf{p}_2$ (Liu et al., 2015). Therefore, the kernel-based inversion allows the computation of the truncated Gauss-Newton updates without additional modeling steps.

The hybrid inversion approach

Ray theory has been used in a wide range of applications, from Kirchhoff migration to ray+Born inversion. The advantage of using ray theory is the high efficiency it can provide in wavefield construction compared to wavefield modeling (Červený, 2001). For 3-D problems, the main limitation of the SI approach for FWI is the high computational cost. Therefore, we propose to reduce the cost and the memory requirements to store the wavefields by considering a hybrid implementation.

The gradient in equation 2 is composed of two components: the data residuals and the Born sensitivity kernels. In the proposed approach, we solve the Helmholtz wave equation for the source wavefield. Thus, the resulting residuals at the receiver locations contain the wavefield residuals. For the receiver locations, we use ray theory to approximate the Green's functions. We compute traveltime and amplitude maps, which are into memory and used in gradient computation. Ray tracing does not require fine spatial sampling. Therefore, the traveltime and amplitude maps can be computed for a coarse grid

to reduce the cost and memory requirements further (Thierry et al., 1999; Jones, 2010).

The proposed method can be used in early FWI iterations, where the velocity model is often smooth and where low-frequency data are used. Indeed, at this stage, only long-to-intermediate wavelengths of the model are updated, and the update is smoothed, justifying the use of ray theory. The resulting smooth model can be used as a good initial model for FWI, or for imaging. For anisotropy scenario the kernels are angle-dependent as shown by Alkhalifah (2016) and Djebbi et al. (2017). Hence, dynamic ray-tracing can accommodate this angle dependence explicitly using the calculated ray-parameter information.

NUMERICAL RESULTS

The proposed method is applied to a 2-D model containing an anomaly in a transmission geometry and on the 2-D Marmousi model to show the accuracy of the inversion results. Also, we compute the 3-D gradients for the VTI anisotropic model using the hybrid method. The results demonstrate the feasibility of a 3-D anisotropic inversion using the proposed hybrid ray+Born approach.

2-D transmission example

We consider a circular anomaly embedded in a homogeneous background medium (see Figure 1a). The background model velocity is $v_0 = 3500$ m/s. The perturbation is $+200$ m/s with respect to the background velocity. The anomaly is located at the center of the model ($x_a = 2500$ m, $z_a = 2500$ m). Here, a transmission experiment is considered. 21 sources are located on the left and top sides at a distance of 500 m from the edges. The receivers are located on the right and bottom sides with the same distance (500 m) from the edges. We invert eight frequencies from 2 Hz to 16 Hz with five iterations per frequency. In this example, we use truncated Gauss-Newton updates. The inner conjugate gradient loop (to solve for the truncated Gauss-Newton update) is fixed to ten iterations.

Figure 1b shows the conventional SI result using the Green's functions computed with a Helmholtz equation solver. The anomaly is fully recovered, as shown in the vertical profile plot Figure 2a. Figure 1c shows the inversion result using the hybrid implementation. The same inversion parameters as the conventional case are used. For transmission geometry, the phase information is sufficient to constrain the inversion and achieve accurate results. Therefore, only ray-tracing-based traveltimes are stored in memory, which halves the memory requirements compared to conventional SI inversion. Then, the Green's functions are computed on the fly using the stored traveltimes, and a constant amplitude equals 1. Additional cost reductions can be achieved as ray-tracing does not require fine grid sampling. Traveltimes can be computed on a coarse grid then interpolated to a finer grid during the update calculation. The inversion results given by Figures 1c and 2b show high accuracy. We conclude that the hybrid method performs well for transmission geometry with limited memory requirements (only traveltimes maps are saved in memory).

Figure 1: Anomaly inversion using the truncated Gauss-Newton updates: (a) the exact model, (b) the inverted velocity using Helmholtz-based scattering integral (SI) approach, (c) the inverted velocity using the hybrid implementation.

Figure 2: Vertical profiles for the anomaly inversion at $x = 2500$ m. (a) Helmholtz-based SI and (b) Hybrid approach. The red line indicates the exact model and the blue line represents the inverted model.

The 2-D Marmousi model

We invert for the Marmousi model, shown in Figure 3a. The model is modified by including a shallow water layer. This layer is fixed during the inversion to avoid ray-tracing artifacts near the sources. We use 57 shots with a 160 m source interval. The data are recorded on 225 receivers with an interval of 40 m. The initial model is given by a velocity increase with depth, as shown in Figure 3b. We update the model using a conjugate gradient method preconditioned with the diagonal of the approximate Hessian (illumination correction). We also calculate the step length using the Born sensitivity kernels, as shown by equation 6. We invert for ten frequencies from 1 Hz to 10 Hz with ten iterations per frequency. The frequency sampling is variable and is increased at higher inverted frequencies as described by Sirgue and Pratt (2004).

We start by inverting for frequencies from 1.0 Hz to 3.5 Hz. In this stage, the proposed hybrid approach is used to reduce the computational cost. The inversion results are shown in Figures 4a and 4b for the conventional scattering integral and the hybrid approach, respectively. The model recovered using the hybrid approach is smoother compared to the conventional Helmholtz-based SI method. As the ray+Born approach is based on ray theory to approximate the Green's functions,

Figure 3: The Marmousi model (a), and the initial model (b).

it faces similar smoothness limitations. The inverted model smoothness is the result of the ray tracing smooth traveltime and amplitude maps resulting in smoother Green's functions

than the Helmholtz based ones. Despite the smoothness of the inversion results as we see from the vertical profiles shown in Figure 6, the long-wavelength features are well recovered. This long-wavelength information is necessary to avoid local minima observed for higher frequencies in FWI.

Figure 4: Inversion results for the Marmousi model for frequencies between 1.0 Hz and 3.5 Hz. (a) Helmholtz-based inversion and (b) Hybrid inversion.

The models inverted using 1.0 Hz to 3.5 Hz frequencies are used as initial models for conventional SI inversion. Figure 5 shows the inversion results. The inversion results are comparable. We show the vertical profiles at $x = 3000$ m and $x = 7500$ m for both inversions in Figures 6. The inversion accurately recovered most features. The second-stage FWI results (blue curves) are comparable for the two methods. We conclude that using the hybrid approach the long wavelengths are accurately reconstructed. The method is a good tool to provide good initial models for conventional FWI with a reduced computational cost.

The maximum frequency used for the hybrid approach is 3.5 Hz. The choice of the maximum frequency only depends on the robustness of the ray tracing, which relies on the criterion used for the smoothness of the velocity model at each iteration. Therefore, if the model is guaranteed to be smooth, higher frequencies can be used for the hybrid inversion.

For real data, low frequencies are usually not available, which exposes FWI to local minima and hampers convergence. Both conventional SI inversion and the proposed hybrid approach face the same issue and are unable to handle the missing information. Ray+Born gradients are inherently smoother than the conventional SI method, but the missing long wavelengths will not be accurately recovered. To deal with the non-linearity issue and to avoid the need for low frequencies, the ray+Born approach can be implemented using other objective functions that are less sensitive to the FWI non-linearity.

Figure 5: FWI inversion results for the Marmousi model using the models inverted for 1.0 Hz to 3.5 Hz, shown in Figure 4, as initial models. (a) Helmholtz-based inversion and (b) Hybrid inversion.

Acoustic VTI ray+Born inversion

For a transversely isotropic medium with a vertical axis of symmetry (VTI) parameterized using v_{nmo} , η and δ , the Fréchet derivatives matrix is given as:

$$\mathbf{F} = [\mathbf{F}_{v_{\text{nmo}}} \quad \mathbf{F}_{\eta} \quad \mathbf{F}_{\delta}]. \quad (7)$$

The total size of the Fréchet matrix is $3 \times (n \times m)$. Each row of the three sub-matrices, given in equation 7, is the sensitivity kernel for v_{nmo} , η and δ , respectively, for a specific source-receiver pair. The single-frequency Born sensitivity kernels (Djebbi et al., 2017) are given as,

$$\begin{aligned} \mathbf{K}^{(v_{\text{nmo}})}(\mathbf{x}_r, \mathbf{x}, \mathbf{x}_s, \omega) &= \frac{2\omega^2}{v_0^3} U_0(\mathbf{x}, \mathbf{x}_s, \omega) G_0(\mathbf{x}, \mathbf{x}_r, \omega); \\ \mathbf{K}^{(\eta)}(\mathbf{x}_r, \mathbf{x}, \mathbf{x}_s, \omega) &= -\frac{2v_0^2}{\omega^2} \partial_{xx} U_0(\mathbf{x}, \mathbf{x}_s, \omega) \partial_{xx} G_0(\mathbf{x}, \mathbf{x}_r, \omega); \\ \mathbf{K}^{(\delta)}(\mathbf{x}_r, \mathbf{x}, \mathbf{x}_s, \omega) &= -(U_0(\mathbf{x}, \mathbf{x}_s, \omega) \partial_{zz} G_0(\mathbf{x}, \mathbf{x}_r, \omega) \\ &\quad + \partial_{zz} U_0(\mathbf{x}, \mathbf{x}_s, \omega) G_0(\mathbf{x}, \mathbf{x}_r, \omega)). \end{aligned} \quad (8)$$

where $U_0(\mathbf{x}, \mathbf{x}_s, \omega)$ and $G_0(\mathbf{x}, \mathbf{x}_r, \omega)$ are the source wavefield and the receiver Green's function using the background model, respectively.

For anisotropic media, the sensitivity kernels are angle dependent (Djebbi and Alkhalifah, 2013; Alkhalifah, 2016; Djebbi et al., 2017; Djebbi and Alkhalifah, 2019) and the wavefield derivatives are required as shown by equation 8. Dynamic ray-tracing can handle this angle dependence explicitly using the calculated ray-parameter information. The Green's function is approximated asymptotically using ray theory, with multipathing ignored, as $G_0(\mathbf{x}, \mathbf{x}_s, \omega) = A(x, x_s) e^{-i\omega t} = A(x, x_s) e^{-i\frac{\omega}{v_0} \mathbf{p} \cdot (\mathbf{x} - \mathbf{x}_s)}$ where $\mathbf{p}$ is the ray parameter vector. The derivatives, assuming

Figure 6: Vertical profiles for the Marmousi model inversion. (a,b) Helmholtz-based inversion vertical profiles at $x = 3000$ m and $x = 7500$ m, respectively. (c,d) Hybrid inversion vertical profiles at $x = 3000$ m and $x = 7500$ m, respectively. The red line stands for the exact model, the black is the initial model, the purple is the inverted model for 1.0 Hz to 3.5 Hz, and the blue line is the final inverted models using FWI.

smoothly varying background velocity model, are given by,

$$\begin{aligned}\partial_{xx}G_0(\mathbf{x}, \mathbf{x}_s, \omega) &= -p_x^2 \frac{\omega^2}{v_0^2} G_0(\mathbf{x}, \mathbf{x}_s, \omega), \\ \partial_{zz}G_0(\mathbf{x}, \mathbf{x}_s, \omega) &= -p_z^2 \frac{\omega^2}{v_0^2} G_0(\mathbf{x}, \mathbf{x}_s, \omega),\end{aligned}\quad (9)$$

where p_x and p_z represent the ray parameter components. With the ray parameter information, we avoid using finite-difference approximations of the derivatives. Therefore, we can reduce the computational cost with the proposed ray+Born approach.

We show in Figure 7 the gradients (given by the sensitivity kernels weighted by the complex conjugate of the residual value) computed for a smooth version of the 3-D SEG/EAGE salt model, Figure 7a, and for a single source-receiver pair. The background initial model is isotropic, therefore η and δ initial models are zeros and are not shown. The source and the receiver are located at the surface and separated by 11600 m. We notice that the computed 3-D gradients capture the effects of anisotropy and inhomogeneity caused by the high-velocity salt body. Using v_{nmo} , η and δ parameters as our VTI model parameterization, the behaviour of the v_{nmo} , η gradients is similar, whereas the δ parameter gradient is different. The trade-off analysis between the VTI parameters and its implication on multi-parameter FWI is studied by Alkhalifah and Plessix (2014) and Djebbi et al. (2017). The angle dependence of the VTI Born sensitivity kernels is the main reason for the ob-

served differences and will heavily affect the resolution and inversion quality. The proposed ray+Born framework is well adapted for anisotropic 3-D inversion as it grants computing the updates with a minimum cost, especially for low frequencies.

COMPUTATIONAL COST ANALYSIS

In this section, we analyze the computational cost of the proposed hybrid inversion approach. The total number of modeling steps for the SI based inversion is $n_s + n_r$, where n_s and n_r denote the number of sources and receivers, respectively. The cost of calculating the gradient by matrix-vector multiplication is the same for the conventional SI and the proposed approach. It is therefore ignored in this analysis. We assume that the computational cost for one modeling in the frequency domain is given by $t_f = 1$. In general, time-domain modeling costs more than frequency-domain modeling. Therefore, we also assume that the computational cost for the time-domain modeling t_t is given by $t_t = 2 \times t_f$. The cost of one ray tracing is assumed to be much smaller than frequency-domain modeling. In fact, a coarser grid can be used, which further reduces the computational cost as no dispersion issues are encountered in ray tracing. Thus, we assume that the ray-tracing cost conservatively equals $t_r = 0.1 \times t_f$. For instance, computing the

Figure 7: 5.0 Hz peak frequency VTI anisotropy gradients for the SEG/EAGE salt model: (a) the SEG/EAGE salt model, (b) the NMO velocity gradient, (c) the η parameter gradient and (d) the δ parameter gradient. The source is located at $(x_s, y_s, z_s) = (1000, 6760, 0)$ m, the receiver is placed at $(x_r, y_r, z_r) = (12600, 6760, 0)$ m.

wavefield for a single shot, on a 2.7 GHz Intel(R) Xeon processor, using ray tracing costs 9% of the Helmholtz solver cost based on five-points stencil and only 1% the Helmholtz solver cost for a nine-points stencil. This relative cost difference can be much bigger for 3-D problems as t_f increases exponentially. Multiple frequencies are usually used for frequency-domain inversion. The number of frequencies is denoted as n_f .

For 2-D problems, we fix the number of shots to $n_s = 100$ and determine the computational cost for a variable number of frequencies n_f and receivers n_r . Figure 8 shows the gradient computational cost for various modeling approaches. Figure 8a shows the computational cost for frequency-domain modeling. The multi-source feature of the Helmholtz solver is not used here, and LU decomposition is repeated for every source and receiver. This situation is also equivalent to solving the Helmholtz equation using iterative methods. The total cost is given by $C_{2D}^f = n_f \times (n_s + n_r) \times t_f$. Figure 8b shows the computational cost when a single LU decomposition is used to solve for all the sources and receivers. The cost is given as $t_f = t_{LU} + t_{solve}$, and is mainly caused by the LU decomposition. We assume here that solving for a single source costs $t_{solve} = 0.01 \times t_{LU}$. The cost is $C_{2D}^f = n_f \times [t_{LU} + (n_s + n_r) \times t_{solve}]$. For time-domain modeling, Figure 8c, the wavefield is Fourier transformed to the frequency domain. The Fourier transform cost is negligible compared to the modeling cost. The SI using time-domain modeling is independent of the number of frequencies,

and the cost is given by $C_{2D}^t = (n_s + n_r) \times t_t$. The SI inversion using the time-domain modeling cost is reasonable, only, when the number of sources and receivers is small. Finally, using a hybrid ray tracing, shown in Figure 8d, the modeling cost is the sum of the sources modeling cost in the frequency domain, with a single LU decomposition, and the receivers ray-tracing modeling cost. The total cost is proportional to $C_{2D}^h = n_f \times [t_{LU} + n_s \times t_{solve}] + n_r \times t_r$. The proposed approach outperforms all methods especially the iterative solver SI approach and the time-domain solver. The difference compared to the LU based conventional SI method is small, however, when the number of frequencies and receivers are large, the proposed hybrid approach becomes more attractive.

For 3-D problems, the model size and the number of sources and receivers are large. The computational cost for full wavefield modeling methods is expensive. If we consider that the number of shots is $n_s = 1000$, direct methods are not applicable for 3-D problems because of the large memory requirements. Therefore, iterative methods and hybrid direct-iterative methods can be used to solve the Helmholtz equation. We consider the same cost assumptions as for the 2-D case. Figure 9 shows the computational cost for a frequency-domain, a time-domain and a hybrid implementation. The hybrid approach cost is the sum of the 3-D iterative frequency-domain modeling cost for the sources and the ray-tracing cost

Figure 8: Modeling computational cost for a 2-D problem. (a) frequency-domain modeling with repeated LU decomposition, (b) frequency-domain modeling with a single LU decomposition, (c) time-domain modeling (d) Hybrid frequency domain ray+Born method.

for the receivers. The total hybrid method cost in this case is $C_{3D}^h = n_f \times n_s \times t_f + n_r \times t_r$. The time-domain approach is independent of the number of frequencies, however, it is more expensive when the number of receivers is large. The hybrid approach cost changes slightly with increasing the number of receivers and it is more efficient when the number of frequencies is not extensive (<ten frequencies). For instance, using 4000 receivers, and six frequencies, the computational cost is $C_{3D}^h = 5500$ units for the Hybrid approach, compared to $C_{3D}^f = 30000$ for frequency-domain and $C_{3D}^t = 10000$ for time-domain modeling approaches. These improvements in the computational cost are achieved assuming a small number of receivers and that the ray-tracing cost is $t_r = 0.1 \times t_f$, with t_f the frequency-domain modeling cost. As previously mentioned, in real applications, the number of receivers can be much larger and the ray-tracing cost can be much smaller, which results in a better efficiency.

The memory requirements to store the Green's functions in memory is $n_s \times n_r \times m$, where m is the model size. For 3-D problems, we assume that the model size is $\mathcal{O}(N^3)$, where N is the dimension in one direction. Also, assuming $n_s \sim n_r \sim N^2$, the total memory requirement is $\mathcal{O}(N^5)$ (Liu et al., 2015). With the proposed hybrid approach, a coarse grid can be used to save the ray-tracing traveltimes and amplitudes. One order of magnitude can be reduced if ray tracing attributes are

saved every ten grid points. Then, during the kernels calculation, the traveltimes and amplitudes can be interpolated, on the fly, to the original grid (Thierry et al., 1999). The final memory requirements will be $\mathcal{O}(N^4)$. Also, for real applications, the number of sources and receivers is smaller than $\mathcal{O}(N^2)$. Additional cost reduction can also be achieved by saving in memory only the area covered by each seismic shot (Liu et al., 2015). Therefore, for 3-D problems, the memory requirements for the proposed approach can be appropriately handled with a large memory workstation.

CONCLUSIONS

We proposed a hybrid ray+Born implementation of full waveform inversion within the scattering integral framework. The method is based on dynamic ray tracing for the computation of the sensitivity kernels. The residuals are computed using wavefield modeling in the frequency domain. The method has the advantage of reduced memory requirements as well as a lower computational cost. Through numerical examples, we showed that the hybrid inversion approach performs well in a transmission inversion set-up. However, for the case of complex velocity models, it can only be used in the early iterations, as ray tracing cannot handle non-smooth models. The smooth models derived from the hybrid approach can be used for migration or as good initial models for FWI.

Figure 9: Modeling computational cost for a 3-D problem. (a) frequency-domain modeling, (b) time-domain modeling (c) Hybrid frequency domain ray+Born method.

The proposed method is more advantageous for 3-D anisotropic inversion, where wavefield modeling requires large computational resources. We showed through a cost analysis the computational cost reduction achieved using the hybrid approach for 3-D inversion. Also, when inverting for anisotropy parameters, the ray parameter (direction) can be used instead of finite differences to approximate the wavefield derivatives required for anisotropic updates.

ACKNOWLEDGEMENTS

We are grateful to King Abdullah University of Science and Technology (KAUST) for financial support. We thank the members of the seismic wave analysis group (SWAG) for the useful discussions. We also thank Tony Martin from PGS for the useful suggestions and PGS for permission to publish this paper. We finally acknowledge the valuable comments from the reviewers that significantly improved the quality of this paper.

REFERENCES

- Alkhalifah, T., 2016, Research note: Insights into the data dependency on anisotropy: An inversion perspective: *Geophysical Prospecting*, **64**, no. 2, 505–513.
- Alkhalifah, T., and J. B. Bednar, 2000, Building a 3-D anisotropic model: Its implication to traveltime calculation and velocity analysis, *in* SEG Technical Program Expanded Abstracts 2000: Society of Exploration Geophysicists, 965–968.
- Alkhalifah, T., and R.-É. Plessix, 2014, A recipe for practical full-waveform inversion in anisotropic media: An analytical parameter resolution study: *Geophysics*, **79**, no. 3, R91–R101.
- Bleistein, N., 1987, On the imaging of reflectors in the earth: *Geophysics*, **52**, no. 7, 931–942.
- Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: *Geophysics*, **60**, no. 5, 1457–1473.
- Červený, V., and F. Hron, 1980, The ray series method and dynamic ray tracing system for three-dimensional inhomogeneous media: *Bulletin of the Seismological Society of America*, **70**, no. 1, 47–77.
- Červený, V., 2001, *Seismic ray method*: Cambridge University Press.
- Chen, P., T. H. Jordan, and L. Zhao, 2007, Full three-dimensional tomography: a comparison between the scattering-integral and adjoint-wavefield methods: *Geophysical Journal International*, **170**, no. 1, 175–181.
- Djebbi, R., and T. Alkhalifah, 2013, Finite frequency traveltime sensitivity kernels for acoustic anisotropic media: Angle dependent bananas, *in* SEG Technical Program Expanded Abstracts 2013: Society of Exploration Geophysicists, 4858–4863.

- , 2019, Frequency domain multiparameter acoustic inversion for transversely isotropic media with a vertical axis of symmetry: *Geophysics*, **84**, no. 1, C1–C14.
- Djebbi, R., R.-É. Plessix, and T. Alkhalifah, 2017, Analysis of the traveltimes sensitivity kernels for an acoustic transversely isotropic medium with a vertical axis of symmetry: *Geophysical Prospecting*, **65**, no. 1, 22–34.
- Dziewonski, A. M., 1984, Mapping the lower mantle: determination of lateral heterogeneity in P velocity up to degree and order 6: *Journal of Geophysical Research*, **89**, no. B7, 5929–5952.
- Fichtner, A., and J. Trampert, 2011, Hessian kernels of seismic data functionals based upon adjoint techniques: *Geophysical Journal International*, **185**, no. 2, 775–798.
- Hu, W., A. Abubakar, T. Habashy, and J. Liu, 2011, Pre-conditioned non-linear conjugate gradient method for frequency domain full-waveform seismic inversion: *Geophysical Prospecting*, **59**, no. 3, 477–491.
- Hung, S., F. A. Dahlen, and G. Nolet, 2000, Fréchet kernels for finite-frequency traveltimes- II. examples: *Geophysical Journal International*, **141**, no. 1, 175–203.
- Jones, I. F., 2010, Tutorial: Velocity estimation via ray-based tomography: *First Break*, **28**, no. 2, 45–52.
- Lailly, P., 1983, The seismic problem as a sequence of before-stack migrations, in conference on Inverse Scattering: Theory and Applications, ed. Bednar, J.: SIAM.
- Lambaré, G., S. Operto, P. Podvin, and P. Thierry, 2003, 3D ray+ Born migration/inversion? Part 1: Theory: *Geophysics*, **68**, no. 4, 1348–1356.
- Liu, Y., L. Dong, Y. Wang, J. Zhu, and Z. Ma, 2009, Sensitivity kernels for seismic fresnel volume tomography: *Geophysics*, **74**, no. 5, U35–U46.
- Liu, Y., J. Yang, B. Chi, and L. Dong, 2015, An improved scattering-integral approach for frequency-domain full waveform inversion: *Geophysical Journal International*, **202**, no. 3, 1827–1842.
- Montelli, R., G. Nolet, G. Masters, F. Dahlen, and S.-H. Hung, 2004, Global P and PP traveltimes tomography: rays versus waves: *Geophysical Journal International*, **158**, no. 2, 637–654.
- Nolet, G., 2012, *Seismic tomography: with applications in global seismology and exploration geophysics*: Springer Science & Business Media.
- Schleicher, J., M. Tygel, and P. Hubral, 1993, 3-D true-amplitude finite-offset migration: *Geophysics*, **58**, no. 8, 1112–1126.
- Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: *Geophysics*, **69**, no. 1, 231–248.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, no. 8, 1259–1266.
- Thierry, P., S. Operto, and G. Lambaré, 1999, Fast 2-D ray+ Born migration/inversion in complex media: *Geophysics*, **64**, no. 1, 162–181.
- Tian, Y., S.-H. Hung, G. Nolet, R. Montelli, and F. A. Dahlen, 2007, Dynamic ray tracing and traveltimes corrections for global seismic tomography: *Journal of Computational Physics*, **226**, no. 1, 672–687.
- Van Der Hilst, R. D., S. Widiyantoro, and E. R. Engdahl, 1997, Evidence for deep mantle circulation from global tomography: *Nature*, **386**, 578–584.
- Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, no. 6, WCC1–WCC26.