FREQUENCY-DOMAIN FULL WAVEFORM INVERSION WITH PLANE-WAVE DATA

Yi Tao[1] , Mrinal K. Sen[2]

[1]*Formerly The University of Texas at Austin; Presently ConocoPhillips, Houston, Texas*
[2]*The University of Texas at Austin and National Geophysical Research Institute, India*

ABSTRACT

We propose an efficient frequency-domain full waveform inversion (FWI) method using plane-wave encoded shot records. The forward modeling involves application of position dependent linear time shifts at all source locations. This is followed by propagation of wavefields into the medium from all shot points simultaneously. The gradient of the cost function needed in the FWI, is calculated first by transforming the densely sampled seismic data into frequency-ray parameter domain and then back-propagating the residual wavefield using an adjoint-state approach. We use a Gauss-Newton framework for model updating. The approximate Hessian matrix is formed with a plane-wave encoding strategy, which requires a summation over source and receiver ray parameters of the Green's functions. Plane-wave encoding considerably reduces the computational burden and cross-talk artifacts are effectively suppressed by stacking over different ray parameters. It also has the advantage of directional illumination of the selected targets. Numerical examples show the accuracy and efficiency of our method.

INTRODUCTION

During the past several decades, full waveform inversion (FWI) has become increasingly popular for velocity analysis in seismic exploration applications (e.g., Tarantola, 1984; Pratt et al., 1998; Pratt, 1999; Shin and Cha, 2008). This is because FWI utilizes both amplitude and travel time information from band-limited seismic wavefields. The forward problem is solved by the two-way wave equation, which can model full wave propagation effects including internal multiples. The FWI overcomes several limitations of some of the common imaging techniques such as those based on ray theory and one-way wave equation and can provide high-resolution estimates of the medium parameters.

Mathematically, FWI can be formulated as a nonlinear optimization problem that iteratively minimizes a misfit functional between observed seismic data and synthetic data. The optimization for FWI is under a gradient-type inversion framework. In the frequency domain, the misfit is a summation of residuals over all sources, receivers and frequencies. Thus, although

gradient-based methods generally provide fast convergence with a good starting model, they are still computationally expensive mainly because a large number of sources need to be simulated at each iteration. Modern acquisitions with high-fold, short source and receiver interval and wide azimuthal coverage make the size of the dataset very large and therefore, the cost of computation is increased significantly.

The simultaneous source encoding approach has been proposed to reduce the computational burden for the shot profile simulations required for time-domain prestack depth migration and FWI (Romero et al., 2000; Capdeville et al., 2005; Krebs et al., 2009; Boonyasiriwat et al., 2009; Guitton and Díaz, 2011). This approach is also implemented in the frequency-domain FWI (Herrmann et al., 2009; Ben-Hadj-Ali et al., 2011). The key step of this approach is forming super-shots by summing densely distributed individual shots. In the frequency domain, a random phase delay is applied to each of the shot records before summation. This, however, results in strong incoherent cross-talk artifacts. The artifacts can be suppressed with stacking the super-shots. However, the performance of this suppression is dependent on the choice of the random shots. This approach is also more sensitive to random noise than conventional FWI method, with the degree of sensitivity dependent on the number of shots assembled (Ben-Hadj-Ali et al., 2011). This method also cannot effectively deal with coherent noise, which can be improved with a structure oriented smoothing operator (Ma and Hale, 2012). The noise in the data can also be mitigated by using the array information with a double beamforming approach (Brossier and Roux, 2011).

An alternative approach to reduce the computational cost is to decompose the densely acquired data into plane-wave domain and to select a limited number of ray parameters for modeling and inversion (Whitmore, 1995; Zhang et al., 2005; Liu et al., 2006; Stoffa et al., 2006). In this approach, a linear phase delay is applied to each shot record to form the super-shot. The process of transforming the raw data into plane-wave gather is also known as slant-stacking since it sums amplitudes along a line of constant slope in the time-space domain (e.g., Stoffa 1989). Cross-talk artifacts are suppressed reasonably well with a sufficient number of ray parameters. The ray parameter is directly related to the take-off angle at the surface location and it can be selected based on an illumination analysis of the targets to be imaged, resulting in further saving of computational time. Plane-wave transformation also has the advantage of easily incorporating anisotropic analysis into the imaging process (Sen and Mukherjee 2003; Sil and Sen, 2009). Better imaging results can be obtained with split-spread receiver gathers, which are generated by the reciprocity principle (Liu et al., 2004).

Vigh and Starr (2008) applied a time-domain plane-wave FWI with applications to synthetic data sets and a real 3D marine data. Compared with the standard FWI, computational time is reduced significantly with comparable quality of the inverted velocity model. When seismic data

do not have sufficient spatial sampling and aperture, plane-wave transform can introduce artifacts in the data space. However, modern seismic acquisitions typically satisfy this requirement and plane-wave approaches can have a wide range of applications. Zhang and Wang (2009) proposed a different propagation scheme with a time-domain wave-equation approach applied to a similar plane-wave dataset. This approach has similar computation cost per iteration as the approach proposed by Vigh and Starr (2008).

In this paper, we focus on the frequency domain methods and investigate some practical issues. A frequency-domain implementation of FWI remains an attractive approach since the coefficients of the allocated matrix can remain the same for different shots. This domain also favors a natural choice of different groups of frequencies to avoid cycle skipping and to obtain multi-scale solutions (Virieux and Operto, 2009).

Preconditioning FWI with the Hessian matrix can provide considerable improvement of convergence by scaling the perturbed model (e.g., Pratt et al., 1998). Inversion with the use of the Hessian matrix is known as a Gauss-Newton optimization procedure, which generally takes fewer iterations to attain quadratic convergence than pure gradient methods. Constructing the full Hessian matrix is prohibitively expensive for very large models. An approximate Hessian is normally used for practical applications (Shin et al., 2001). Further reduction in the computational cost can be attained by using the diagonal part of the Hessian only. However, even the diagonal Hessian requires $Ns + Nr$ forward modeling steps, with Ns equals the number of shots and Nr equals the number of receivers. A phase encoding scheme (Tang, 2009; Tang and Lee, 2010) has been proposed to save computational time and disk storage for Hessian calculation. Another alternative to Gauss-Newton procedure is to use a quasi-Newton approach, which does not compute the Hessian matrix explicitly, but instead iteratively update it (Nocedal, 1980; Guitton and Díaz, 2011; Ma and Hale, 2011).

To improve on the efficiency and accuracy of this full-wave inverse problem, we propose performing FWI with frequency domain plane-wave data. With limited selected ray parameters of plane-wave data, computation time is saved both in forward modeling and back-propagation of residuals for gradient calculation. This approach simplifies the inversion procedure with an effective suppression of cross-talk artifacts. The space-domain Hessian is formed using a plane-wave phase encoding approach, allowing efficient calculation with comparable results as those obtained with the standard method. Unlike random phase-encoding approach, this method is not sensitive to random noise. This approach can also be easily applied to marine geometry, which requires special treatment with a random phase encoding method (Krebs et al., 2009).

This paper is organized as follows. I first review the conventional FWI approach and introduce our plane-wave encoding scheme for gradient and Hessian calculations. Next, I show some numerical examples to demonstrate the validity and efficiency of this method.

METHODOLOGY

Three primary steps comprise our method. The first involves slant stacking and temporal Fourier transform, which provides a frequency-domain plane-wave gather from recorded seismic data. This step uses data from different sources and compresses them into a regularized super-shot. Second, the synthetic data are simulated with a distance dependent linear time shift at all receiver locations. This delay is determined by the ray parameter in consideration, which can be determined by the take-off angle at the surface location. Similar approach is adopted for back propagation of the residuals. Third, the diagonal Hessian is formed with a plane-wave encoding approach. These are then used in model update in which the optimal step length is determined by a line search. The above process is repeated for multiple groups of frequencies until convergence is reached.

Frequency-domain plane-wave data

For a recorded time-domain shot gather, frequency-domain plane-wave data is obtained first through a plane-wave transformation, also known as $\tau - p$ transform (e.g., Stoffa 1989; Foster and Mosher, 1992). This transform can be implemented either in the time-domain or in the frequency-domain. In the frequency-domain, it involves a phase shift to each trace that is dependent on the offset and ray-parameter. This is followed by a stacking operator as follows

$$\tilde{u}(\mathbf{p} \mid \mathbf{x}_s, \omega) = \int u(\mathbf{x}_s \mid \mathbf{x}, \omega) \exp(i\omega \mathbf{p} \cdot (\mathbf{x} - \mathbf{x}_s)) d\mathbf{x}, \tag{1}$$

where $u(\mathbf{x}_s \mid \mathbf{x}, \omega)$ is the frequency-domain recorded data at position $\mathbf{x}$ for frequency ω and shot $\mathbf{x}_s$, $\tilde{u}(\mathbf{p} \mid \mathbf{x}_s, \omega)$ is the transformed plane-wave gather. The vectors $\mathbf{p}$ and $\mathbf{x}$ are used to include 3D scenarios. The multiplication $\mathbf{p} \cdot (\mathbf{x} - \mathbf{x}_s)$ is then an inner product. For 2D cases, $\mathbf{p}$ and $\mathbf{x}$ are scalars.

Note that in the time-domain, this transform needs to be modified by applying a time-shift to transform the data from $\tau - p$ domain to "physical" time-ray parameter ($t - p$) domain to match the modeling strategy discussed next (Whitmore, 1995). Also, application of the reciprocity principle can produce a split-spread receiver gather, which helps to overcome the aliasing problem in plane-wave composition (Liu et al., 2006).

In a conventional frequency-domain FWI framework, the forward modeling step is a simulation of the wavefield propagated from a surface location $\mathbf{x_s}$. For a constant density acoustic medium, is governed by the following equation

$$\nabla^2 u(\mathbf{x},\omega)+\frac{\omega^2}{v^2(\mathbf{x})}u(\mathbf{x},\omega)=f(\omega)\delta(\mathbf{x}-\mathbf{x_s}), \tag{2}$$

where $u(\mathbf{x},\omega)$ represents the scalar pressure wavefield in the frequency domain, $v(\mathbf{x})$ is the acoustic velocity, and $f(\omega)$ is the frequency domain source signature.

After discretizing the model with a finite difference or finite element scheme, equation 2 can be rewritten in a matrix form as follows (e.g., Marfurt 1984)

$$\mathbf{A}(\mathbf{x},\omega)u(\mathbf{x},\omega)=f(\omega)\delta(\mathbf{x}-\mathbf{x_s}), \tag{3}$$

where $\mathbf{A}(\mathbf{x},\omega)$ is the complex impedance matrix. This matrix is typically sparse and symmetric, though it can be non-symmetric for many boundary conditions (Hustedt et al., 2004). Because of the pattern of the matrix, equation 3 can be solved with a scientific linear solver package such as the Multifrontal massively parallel solver (MUMPS) (Amestory et al., 2006, Operto et al., 2007) and Portable Extendable Toolkit for Scientific Computation (PETSc) (Balay et al., 2001). We adopt the direct solver approach with MUMPS to speed up the computation.

The modeled plane-wave gather is obtained with a linear source, which is obtained by using all the sources simultaneously each with a phase shift along the surface seismic line (Whitmore, 1995). Correspondingly, the forward problem can be formulated as

$$\mathbf{A}(\mathbf{x},\omega)u(\mathbf{x},\mathbf{p},\omega)=\sum_{\mathbf{x_s}}\varphi(\mathbf{x},\mathbf{x_s},\mathbf{p},\omega)f(\omega)\delta(\mathbf{x}-\mathbf{x_s}), \tag{4}$$

with the *i*th component of $\varphi(\mathbf{x},\mathbf{x_s},\mathbf{p},\omega)$ is given by

$$\varphi(x_i,x_{s_i},p_i,\omega)=\begin{cases}\exp\left(i\omega p_i(x_{s_i}-x_{0_i}\right) & p_i\geq 0\\ \exp\left(i\omega p_i(x_{s_i}-x_{\max_i})\right) & p_i<0\end{cases}, \tag{5}$$

where x_{0_i} and $x_{\max_i}$ are the *i*th component of the start and end receiver locations, respectively.

Mathematically, equation 4 is the same as the time-domain approach. However, a frequency-domain implementation does not need temporal interpolation and has the flexibility to include frequency scaling and phase rotation, which is useful when the source wavelet is not well

estimated at first and when frequency-dependent processing is needed to obtain optimal results (Whitmore, 1995).

Gauss-Newton inversion strategy

Like other inverse problems, FWI is formulated as an optimization problem searching for the minimum of a suitably defined misfit functional that measures the difference between the observed and calculated data. A regularization term is necessary to make the inversion better posed (Menke, 1984). In the least-square sense, where L-2 norm is used for the misfit functional, this problem is stated as

$$\min_{\mathbf{m}} E(\mathbf{m}) \text{ such that } F(\mathbf{m}) = 0 , \tag{6}$$

with

$$E(\mathbf{m}) = \frac{1}{2}\sum_{\omega}\sum_{\mathbf{x_s}}\sum_{\mathbf{x_r}}\left[\Delta\mathbf{d}^T\mathbf{W}_d\Delta\mathbf{d}^* + \lambda\left(\mathbf{m}-\mathbf{m}_{prior}\right)^T\mathbf{W}_m\left(\mathbf{m}-\mathbf{m}_{prior}\right)^*\right], \tag{7}$$

and

$$\Delta\mathbf{d} = \mathbf{d}_{obs} - \mathbf{d}_{cal} , \tag{8}$$

where $\mathbf{m}$ is the model, $\Delta\mathbf{d}$ is data residual between the observe data $\mathbf{d}_{obs}$ and calculated data $\mathbf{d}_{cal}$, $\Delta\mathbf{d}^T$ and $\Delta\mathbf{d}^*$ are the transpose and complex conjugate of the data residual, respectively; $\mathbf{W}_d$ and $\mathbf{W}_m$ are the inverses of the data and model covariance matrices, respectively; $\mathbf{m}_{prior}$ is the prior information about the model known from travel time tomography, for example; λ is a damping factor to weigh the data space and model space errors.

The above nonlinear problem is linearized by expanding the misfit function in a generalized Taylor's series about a reference model $\mathbf{m}_0$. Assuming that the Hessian (second partial derivative of the data with respect to model parameters) can be constructed easily and ignore the model gradient, the model perturbation can be obtained by the following Gauss-Newton framework formula (Pratt et al., 1998)

$$\begin{aligned}\Delta\mathbf{m} &= -\left(\left(\frac{\partial^2 E}{\partial\mathbf{m}^2}\right)+\lambda\mathbf{I}\right)^{-1}\mathbf{W}_m^{-1}\frac{\partial E}{\partial\mathbf{m}} \\ &= -\mathbf{W}_m^{-1}\left(\mathrm{Re}\,\mathbf{H}+\lambda\mathbf{I}\right)^{-1}\mathrm{Re}\left[\mathbf{J}^T\mathbf{W}_d\Delta\mathbf{d}^*\right]\end{aligned} , \tag{9}$$

where $\mathbf{J}$ is the Fréchet derivative of the data with respect to model parameters, $\mathbf{H}$ is the Hessian matrix and $\mathbf{H} = \mathbf{J}^T \mathbf{W}_d \mathbf{J}^*$, $\mathbf{I}$ is the identity matrix (Re means the real part), and $\mathbf{W}_m^{-1}$ acts as a smoothing operator with a correct regularization approach (Virieux and Operto, 2009).

Under Born approximation, the gradient is obtained by taking the partial derivative of the perturbed model with respect to the perturbed misfit functional. It can be explicitly expressed as (Sirgue and Pratt, 2004, Tromp et al., 2005)

$$g(\mathbf{x}) = \frac{\partial E}{\partial \mathbf{m}} = \sum_{\omega} \sum_{\mathbf{x_s}} \sum_{\mathbf{x_r}} \mathrm{Re}\left\{ \frac{-2\omega^2 f(\omega)}{v^3(\mathbf{x})} G(\mathbf{x}, \mathbf{x_s}, \omega) G(\mathbf{x}, \mathbf{x_r}, \omega) W_d(\mathbf{x}, \mathbf{x_r}) \Delta d^*(\mathbf{x_r}, \mathbf{x_s}, \omega) \right\}, \quad (10)$$

where $G(\mathbf{x}, \mathbf{x_s}, \omega)$ is the Green's function at the space location $\mathbf{x}$ with a source term at $\mathbf{x_s}$, $v(\mathbf{x})$ is the velocity, $W_d(\mathbf{x}, \mathbf{x_r})$ and $\Delta d^*(\mathbf{x_r}, \mathbf{x_s}, \omega)$ are the elements of the matrix $\mathbf{W}_d$ and $\Delta \mathbf{d}^*$, $\mathbf{W}_d \Delta \mathbf{d}^*$ is the weighted data residual and '*' is complex conjugation.

Efficient calculation of the gradient is based on the well-known adjoint state approach, where the forward-propagated source wavefield is crosscorrelated with the back-propagated data residual. In practice, the source wavefield is usually simulated for every source location and the back-propagated residual wavefield is usually simulated simultaneously for all the receiver locations. Under such a framework, the gradient can be written as (e.g., Plessix and Mulder, 2004)

$$g(\mathbf{x}) = \frac{\partial E}{\partial \mathbf{m}} = \sum_{\omega} \sum_{\mathbf{x_s}} \mathrm{Re}\left\{ \frac{-2\omega^2 f(\omega)}{v^3(\mathbf{x})} S(\mathbf{x}, \mathbf{x_s}, \omega) \Delta R^*(\mathbf{x}, \mathbf{x_s}, \omega) \right\}, \quad (11)$$

where $S(\mathbf{x}, \mathbf{x_s}, \omega)$ is the forward simulated source wavefield at a particular shot location and $\Delta R(\mathbf{x}, \mathbf{x_s}, \omega)$ is the reverse-time propagated data residual for all the receiver locations. The gradient can be expressed as the crosscorrelation between the forward simulated wavefield and back propagated data residual at the zero lag time.

The full Hessian matrix is never formed explicitly for practical implementations because of the prohibitive computation cost. Instead, the second-order term, which accounts for multiple scattering effects, is often neglected. The first-order term (approximate Hessian), which is due to single scattering, can be explicitly written as (Plessix and Mulder 2004; Tang, 2009)

$$H_{\alpha}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{\omega} \sum_{\mathbf{x_s}} \sum_{\mathbf{x_r}} \mathrm{Re}\left\{ \frac{4\omega^4 \left| f(\omega) \right|^2}{v^3(\mathbf{x}_i) v^3(\mathbf{x}_j)} G^*(\mathbf{x}_i, \mathbf{x_s}, \omega) G(\mathbf{x}_j, \mathbf{x_s}, \omega) G^*(\mathbf{x}_i, \mathbf{x_r}, \omega) G(\mathbf{x}_j, \mathbf{x_r}, \omega) \right\}, \quad (12)$$

where $H_\alpha(\mathbf{x}_i, \mathbf{x}_j)$ is the approximate Hessian. The rows of the Hessian can be interpreted as source and receiver illumination energy. Preconditioning the gradient with the approximate Hessian can remove the geometrical amplitude decay of the Green's functions. Therefore, it can scale the deep perturbations constrained by far-offset data and shallow perturbations constrained by near-offset data (Pratt et al., 1998). The diagonal part of the Hessian can be interpreted as the resolution of individual model parameters. The off-diagonal part of the Hessian can be viewed as the trade-off between different parameters. Large off-diagonal elements indicate large uncertainties for parameter estimation since the misfit functional can remain unchanged when changes of some parameters are compensated with changes of other parameters.

Computing this Hessian is still expensive for present real data applications since it requires $Ns \times Nr$ forward simulations to obtain the source-side and receiver-side Green's functions, with Ns and Nr representing the number of shots and number of receivers, respectively. To further reduce the computational burden, the diagonal part of the Hessian ($\mathbf{x}_i = \mathbf{x}_j$) can be used. It is given by

$$H_0(\mathbf{x}) = \sum_{\omega} \sum_{\mathbf{x_s}} \sum_{\mathbf{x_r}} \mathrm{Re}\left\{ \frac{4\omega^4 \left| f(\omega) \right|^2}{v^6(\mathbf{x})} \left| G(\mathbf{x}, \mathbf{x_s}, \omega) \right|^2 \left| G(\mathbf{x}, \mathbf{x_r}, \omega) \right|^2 \right\}, \tag{13}$$

where $H_0(\mathbf{x})$ is used to represent the diagonal Hessian, the L-2 norm of the complex scalar vector $\left| f(\omega) \right|^2 = f^*(\omega) f(\omega)$, the same relation holds for $G(\mathbf{x}, \mathbf{x_s}, \omega)$ and $G(\mathbf{x}, \mathbf{x_r}, \omega)$. The diagonal Hessian is the autocorrelation of the partial derivative of the source-side and receiver-side Green's functions. This can still provide a good approximation to scale the gradient when the off-diagonal elements are relatively small. The computation cost of the diagonal Hessian requires $Ns + Nr$ forward simulations, which is equivalent to the calculation of the gradient, which requires $2Ns$ forward simulations.

Plane-wave encoding strategy

The computational cost of FWI can be reduced significantly by incorporating the plane-wave forward modeling strategy because of reduction in the number of wavefield simulations. The source wavefield, back-propagated residual wavefield and the source and receiver Green's functions are calculated using equations 4 and 5. After numerical simulations, plane-wave encoded gradient can be written as

$$\tilde{g}(\mathbf{x}) = \sum_{\omega} \sum_{\mathbf{p_s}} \mathrm{Re}\left\{ \frac{-2\omega^2 |\omega|}{v^3(\mathbf{x})} f(\omega) \tilde{S}(\mathbf{x}, \mathbf{p_s}, \omega) \Delta \tilde{R}^*(\mathbf{x}, \mathbf{p_s}, \omega) \right\}. \tag{14}$$

where $\mathbf{p_s}$ is the source ray parameter. It corresponds to the take-off angle of the receiver gather. Equation 14 is equivalent to shot-profile gradient (Appendix A). However, the summation in plane-wave gradient is over ray parameters instead of shot locations, which can considerably reduce the number of simulations when a large number of sources are simulated in a conventional FWI.

Similar to the gradient calculation, the diagonal Hessian can also exploit a plane-wave phase encoding strategy to reduce computational burden. Using an algorithm similar to wave-equation shot-profile migration, we can obtain the receiver-side phase-encoded Hessian as follows

$$\tilde{H}_0(\mathbf{x}) = \sum_{\omega}\sum_{\mathbf{x_s}}\sum_{\mathbf{p_r}} \mathrm{Re}\left\{ \frac{4\omega^4 \left|\omega\right| \left| f(\omega)\right|^2}{v^6(\mathbf{x})} \left|G(\mathbf{x},\mathbf{x_s},\omega)\right|^2 \left|\tilde{G}(\mathbf{x},\mathbf{p_r},\omega)\right|^2 \right\}. \qquad (15)$$

We further consider a simultaneous encoding of source and receiver-side Green's functions with plane-wave simulations. This computing process can be easily implemented. It requires only two forward wavefield simulations: one for source plane-wave wavefiled and the other for receiver plane-wave wavefield. Mathematically, it can be written as

$$\tilde{\tilde{H}}_0(\mathbf{x}) = \sum_{\omega}\sum_{\mathbf{p_s}}\sum_{\mathbf{p_r}} \mathrm{Re}\left\{ \frac{4\omega^4 \left|\omega f(\omega)\right|^2}{v^6(\mathbf{x})} \left|\tilde{G}(\mathbf{x},\mathbf{p_s},\omega)\right|^2 \left|\tilde{G}(\mathbf{x},\mathbf{p_r},\omega)\right|^2 \right\}. \qquad (16)$$

Details of the derivation of equations 15 and equation 16 are given in Appendix B and Appendix C respectively. Equations 13, 15 and 16 have similar forms, but they differ from each other in the propagation scheme and in the number of forward simulations required. Assuming that a survey has Ns shots and Nr receivers, for one step in line search, direct computation of the diagonal Hessian requires $Ns + Nr$ simulations while simultaneous plane-wave encoding requires $Np_s + Np_r$ simulations. Unlike random phase encoding approach, where the number of simulations varies for marine geometry and for land-acquisition or Ocean Bottom Seismometer (OBS) geometry, plane-wave encoding approach just needs a similar number of ray parameters. The number of ray-parameters can be determined by the imaging targets and by the maximum dips present in the shot and receiver gathers.

Any kind of inversion approach in a Gauss-Newton framework can be applied for optimization. For simplicity, we use a steepest decent approach to update the velocity model. This method exploits a parabolic or cubic search algorithm with a modification of the fixed length model perturbation of equation 9. The model update can be written as

$$\mathbf{m}_k(\mathbf{x}) = \mathbf{m}_{k-1}(\mathbf{x}) + \alpha_k \left(H_{0_{k-1}}(\mathbf{x}) + \lambda \mathbf{I} \right)^{-1} \mathbf{W}_m^{-1} g_{k-1}(\mathbf{x}) \,. \qquad (17)$$

where α_k is a scalar for the k th iteration. It is defined by line searching along the preconditioned gradient direction. In this paper, we use a parabolic fitting scheme for line search.

A multi-scale solution is obtained through the inversion of discrete frequencies. Groups of frequencies are selected and two successive loops are formed. The first loop is over a group of frequencies, and the second loop is over the frequencies of the frequency group (Brossier et al., 2009). Inverted model of the previous frequency loop is used as the starting model for the new loop. The stopping criterion for each iteration is given by

$$\varepsilon = \frac{\left\| \mathbf{d}_{obs}(\mathbf{x},\omega) - \mathbf{d}_{cal}(\mathbf{x},\omega) \right\|}{\left\| \mathbf{d}_{cal}(\mathbf{x},\omega) \right\|} \quad . \tag{18}$$

A large ε value means we can stop when the error is large, therefore we can save computational time. On the other hand, a small ε value means a more accurate data fitting is required. In this paper, we use $\varepsilon = 0.001$ and this value is fixed for all iterations and for each frequency group.

NUMERICAL EXAMPLES

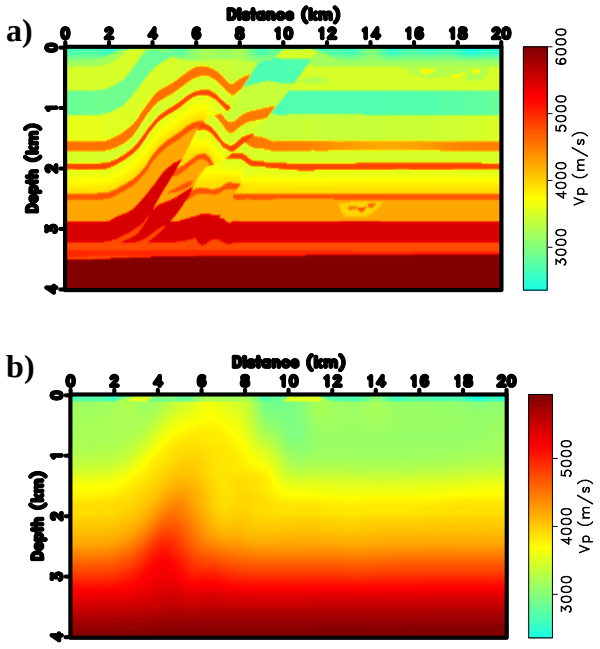

Figure 1. A 2-D profile of the Overthrust model. (a) P-wave velocity of the true model. (b) Initial model used to generate the gradient and for inversion in this paper.

In this section, we will demonstrate the numerical applications of our approach with synthetic examples. The synthetic examples are based on a 2D profile of the SEG/EAGE Overthrust model

(Figure 1). Analysis of effects of different ray parameters on the gradient is presented to demonstrate the dip-selective advantage of our method. In other words, the gradient calculation can be benefited from plane-wave decomposition by focusing the image on selected targets. Calculation of the Hessian with plane-wave encoding is more powerful since only approximated Hessian with a few ray parameters are necessary to ensure the illumination compensation of the gradient. We analyze the effectiveness of the encoding schemes with selective ray parameters. We finally show the inversion results from our examples with different strategies on the selection of the ray parameters.

Dip-selective gradient

The P-wave velocities of the true model and initial model (a smoothed version of the true model) are shown in Figure 1. This model has 187×801 grids with 25m of horizontal and vertical grid intervals. While a 2-D example is presented for simplicity, extension of all the following analysis to a bigger model and to 3-D is trivial. Similar to Ben-Hadj-Ali et al. (2011) we use seven different frequencies ranging from 3 Hz to 20 Hz. The range of ray parameters can be defined by the imaging targets and the maximum dip of the recorded data. The range for the ray parameters used in this paper is $[-0.4 \text{ s/km}, 0.4 \text{ s/km}]$.

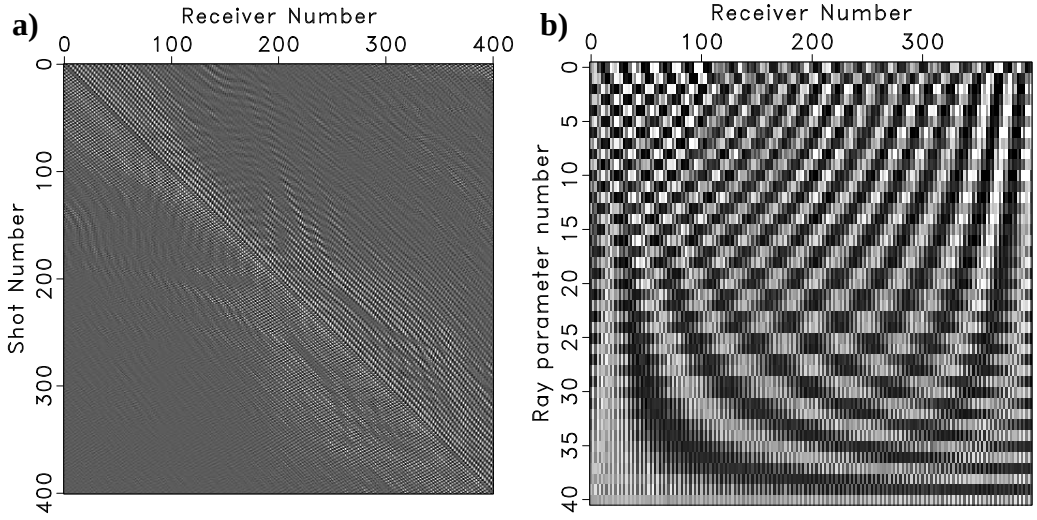

Figure 2. Monochromatic frequency domain (f=20 Hz) seismic data (a) in the source-receiver domain. (b) ray parameter-receiver domain.

The data for a single frequency component with shot-domain modeling and plane-wave domain are shown in Figure 2. The sparsity in the shot-domain does not exist in the plane-wave domain. The effect of a single ray parameter to the gradient is shown in Figure 3. The gradient

shown here is the first iteration with the initial model in Figure 1b. Though a sequential calculation for each frequency or each frequency band is typically required for the inversion process, the gradient is summed over all the frequencies for a better display. Each ray parameter is related to the take-off angle at the surface location, it contributes to the gradient with directional illumination. More energy is focused on the left side of the model for a negative ray parameter (Figure 3b) and more illumination on the right side of the model for a positive ray parameter (Figure 3d). A zero ray parameter (Figure 3c) tends to provide balanced illumination on both sides of the model. For reference, the gradient with 81 ray parameters is shown in Figure 3a.

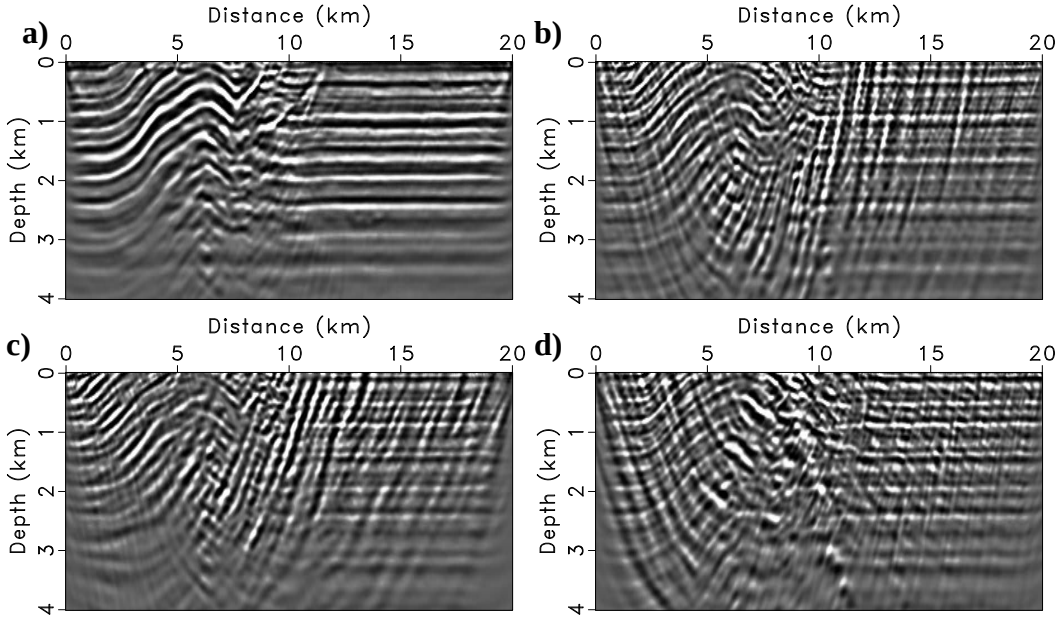

Figure 3. Plane-wave gradient at the first iteration with different choice of parameters. (a) 81 source and receiver ray parameters; (b) single source ray parameter $p_s = -0.3$ s/km**; (c)** $p_s = 0$ s/km**; (d)** $p_s = 0.3$ s/km**.**

Figure 4 shows how the number of ray parameters affects the gradient calculation. For only zero ray parameter (zero offset section), there are plenty of artifacts that are generated by the footprints of plane-wave encoding (Figure 4a). For 9 ray parameters (Figure 4b), the artifacts reduce substantially. For 41 ray parameters (Figure 4c), the artifacts become unnoticeable compared with 81 ray parameters (Figure 4d). This provides a qualitative way to define the number of ray-parameters to be used in the inversion. For this problem, we assess that 41 ray parameters are adequate for inversion, provided that the 41 ray parameters are enough for Hessian calculation. Another quantitative requirement of the number of ray parameters can be written as (Zhang et al., 2005)

$$N_p \geq \frac{Lf_{\max}(\sin\alpha_{\max} - \sin\alpha_{\min})}{v}, \tag{19}$$

where N_p is the number of ray parameters. L is the distance of the receiver gather, $f_{\max}$ is the maximum frequency for inversion, $\alpha_{\min}$ and $\alpha_{\max}$ are the minimum and maximum take-off angle, respectively, v is the velocity at the surface location.

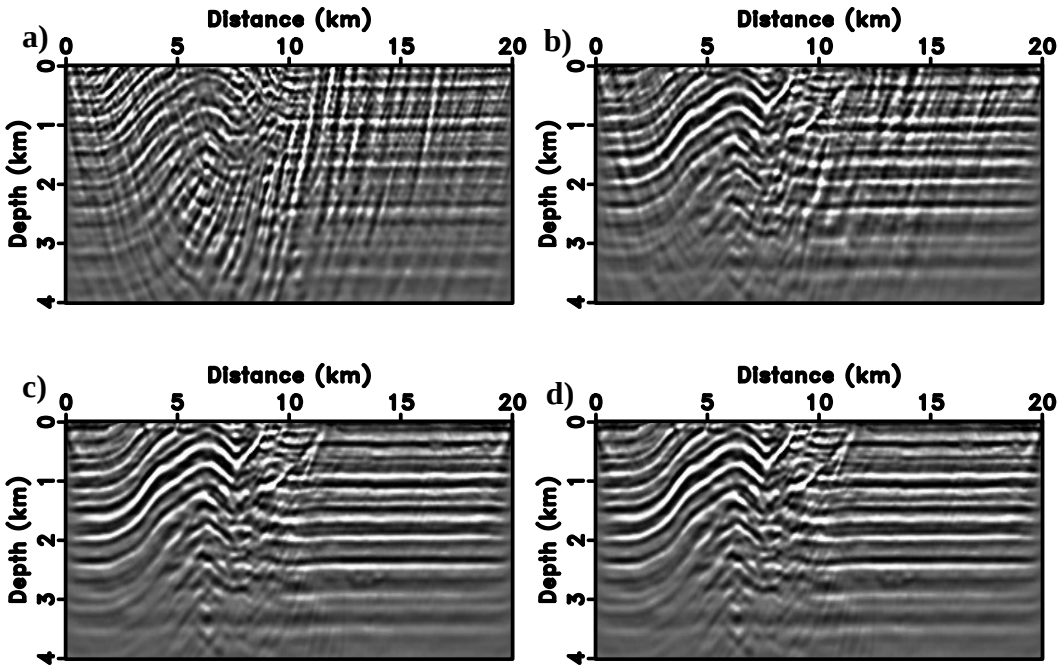

Figure 4. The effect of the ranges of ray parameters on the plane-wave gradient. (a) 1 source ray parameter ($p_s = 0$ s/km); (b) 9 source ray parameters equally distributed from -0.4 s/km **to** 0.4 s/km **; (c) 41 source ray parameters; (d) 81 source ray parameters.**

Plane-wave encoded Hessian

To test the effect of directional illumination for the source and receiver plane-wave encoded Hessian, we first apply this strategy to a medium with a constant velocity gradient. The background velocity is described by *v(z)* = 2000+ 0.4*z*. The diagonal Hessian for a single frequency at 3 Hz is shown in Figure 5. In addition, the Hessian for the true Overthrust model is shown in Figure 6. Unlike the gradient calculation where only a single ray parameter is used, the Hessian is calculated with 1 source ray parameter and all receiver ray parameters. Similar to the gradient, single ray parameter contributes to directional illumination of the model. This decomposition process can be used for directional choice of the ray parameters for the inversion.

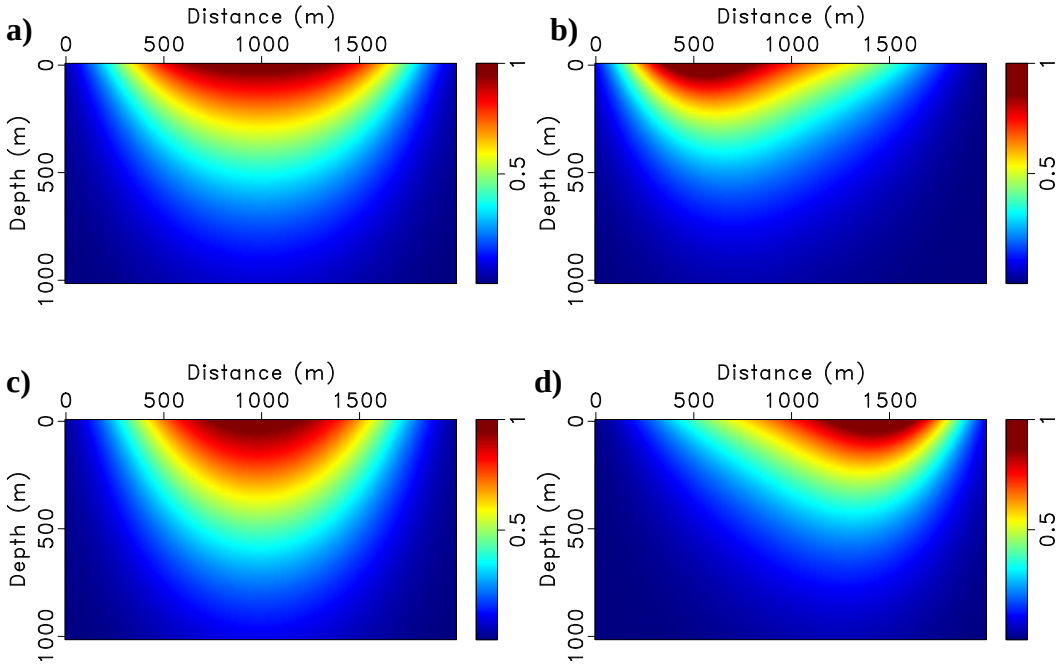

Figure 5. Diagonal Hessian for a constant velocity gradient model with different selection of parameters. (a) 101 source and receiver ray parameters; (b) single source ray parameter $p_s = -0.3$ s/km and 101 receiver ray parameters; (c) $p_s = 0$ s/km; (d) $p_s = 0.3$ s/km.

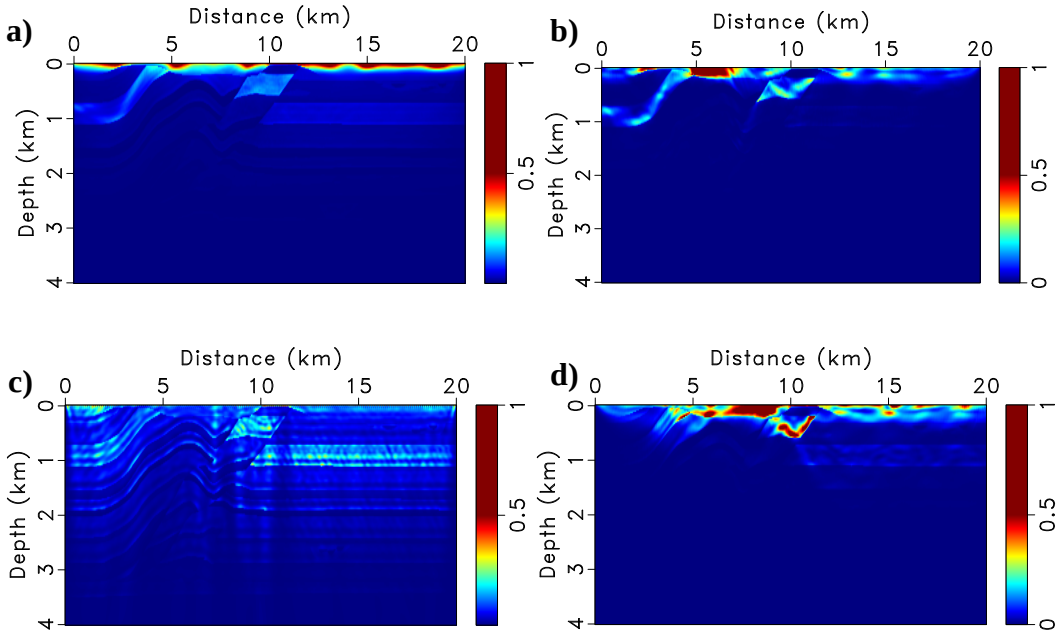

Figure 6. Diagonal Hessian for the true model with different selection of parameters. (a) 81 source and receiver ray parameters; (b) single source ray parameter $p_s = -0.3$ s/km and 81 receiver ray parameters; (c) $p_s = 0$ s/km; (d) $p_s = 0.3$ s/km.

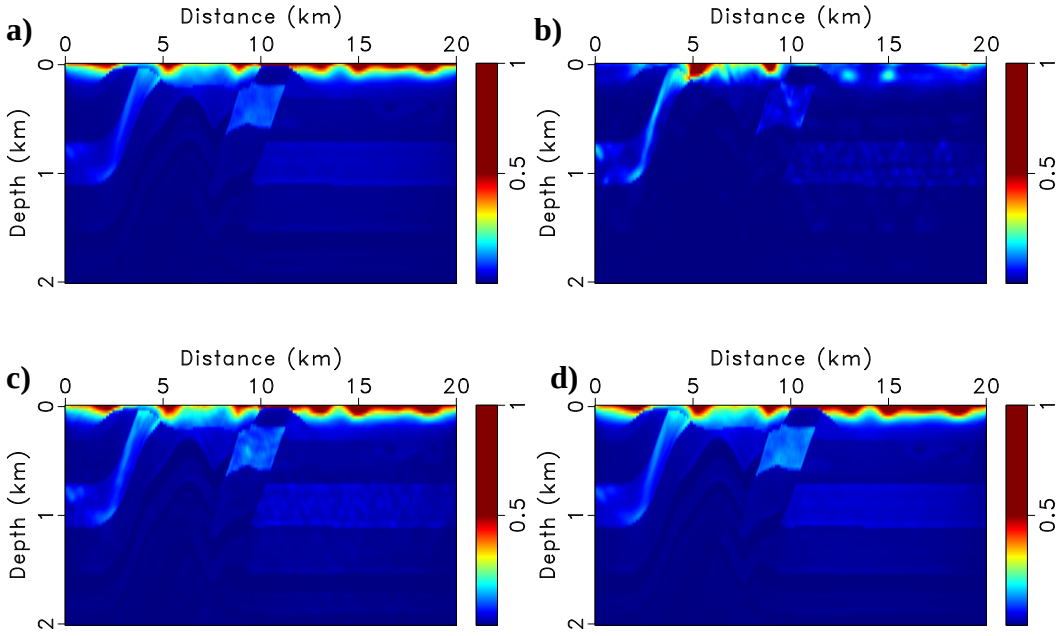

Figure 7. Comparison of the diagonal Hessian for the true model with different parameters. (a) true Hessian with 401 sources and 401 receivers; (b) plane-wave Hessian with 9 source and receiver ray parameters; (c) plane-wave Hessian with 41 source and receiver ray parameters; (d) plane-wave Hessian with 81 source and receiver ray parameters.

Plane-wave encoding provides a good approximation of the Hessian. For the plane-wave 9 source and receiver ray parameters (Figure7b), this approximation has notable artifacts. For 41source and receiver ray parameters, theses artifacts are very small (Figure7c). 81 source and receiver ray parameters (Figure 7d) produce visually identical results as the true Hessian calculated with 401 sources and 401 receivers (Figure 7a). This also confirms that 41 ray parameters may be enough for this problem. Note that if the off-diagonal parts of the Hessian are needed for the accuracy requirement in a Gauss-Newton framework, plane-wave encoding is more appealing because significant amount of computational time can be saved.

Inversion of the Overthrust model

We employ a sequential inversion approach for each frequency component. Inverted result from the previous step is used as the starting model for the next step and so on. Figure 8b shows the inverted model with 9 source and receiver ray parameters. The gradient only requires decomposing the receiver gather into source ray parameters. The receiver ray parameters are used for Hessian calculation. For each iteration, only 9 wavefield simulations are needed for the gradient and also 9 simulations for the Hessian when the reciprocity is considered. From the inverted result, we can see that with only 9 ray parameters, promising result can be obtained.

However, it still has some errors (more on the shallow part) because the gradient is not preconditioned very well. With 41 ray parameters (Figure 8c), the inverted result is almost comparable with the result using 401 sources and 401 receivers (Figure 8a). Ben-Hadj-Ali et al. (2011) used 199 sources and 200 receivers and showed similar results as Figure 8a.

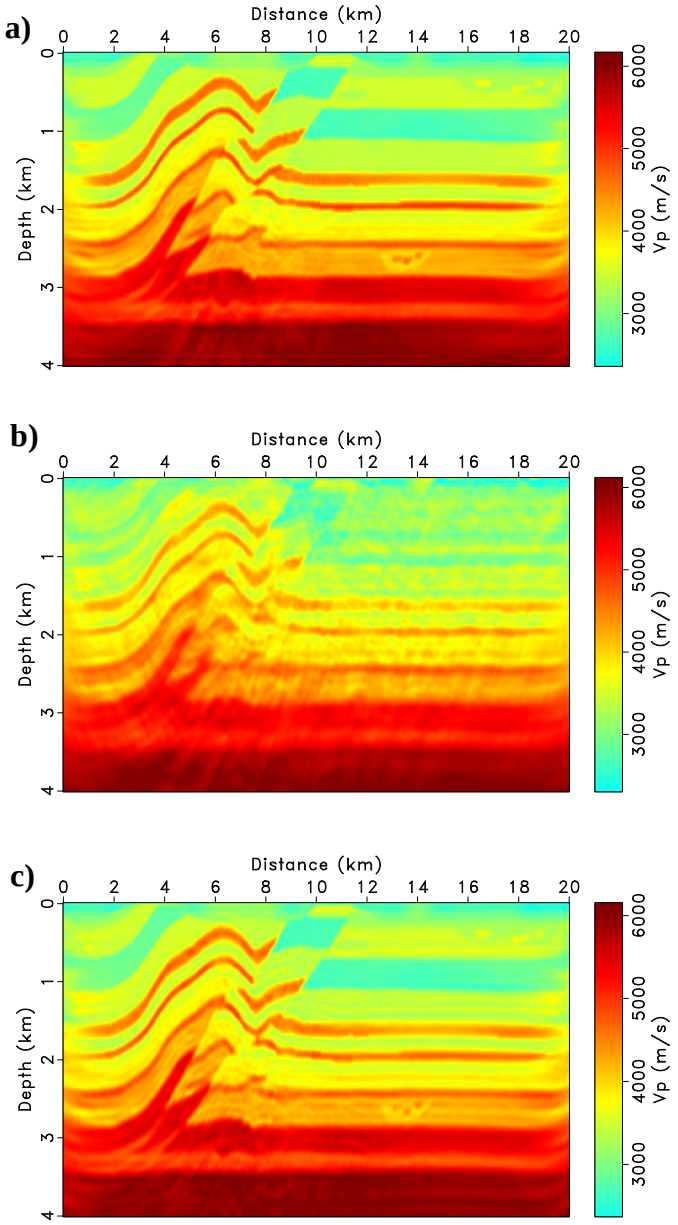

Figure 8. Inverted P-wave velocity model with (a) shot domain approach with 401 sources and 401 receivers; (b) plane-wave domain with 9 source and receiver ray parameters; (c) plane-wave Hessian with 41 source and receiver ray parameters.

Figure 9 plots the normalized data residuals as a function of the iteration number. For a lower frequency, the error converges faster and using more iterations does not reduce the error much.

For a higher frequency, since the starting model is already close to the true model, a slow search process is shown in this plot. Figure 10 shows the time domain plane-wave gather for $p_s = 0$ s/km and $p_s = 0.3$ s/km, respectively. This is obtained by an inverse Fourier Transform of the frequency domain data. While little structure can be observed from the starting model, the inverted model shows features that are similar to those of the true model.

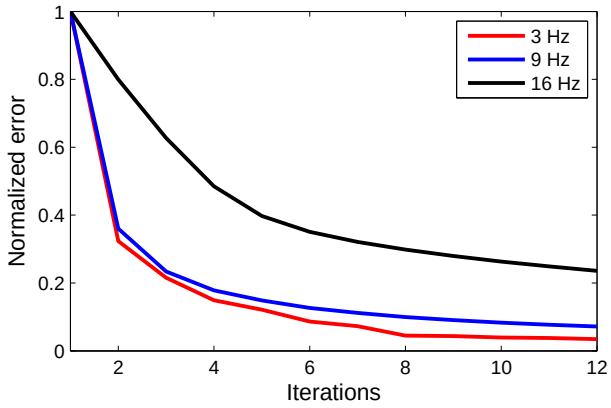

Figure 9. Normalized error versus iterations for three different frequencies.

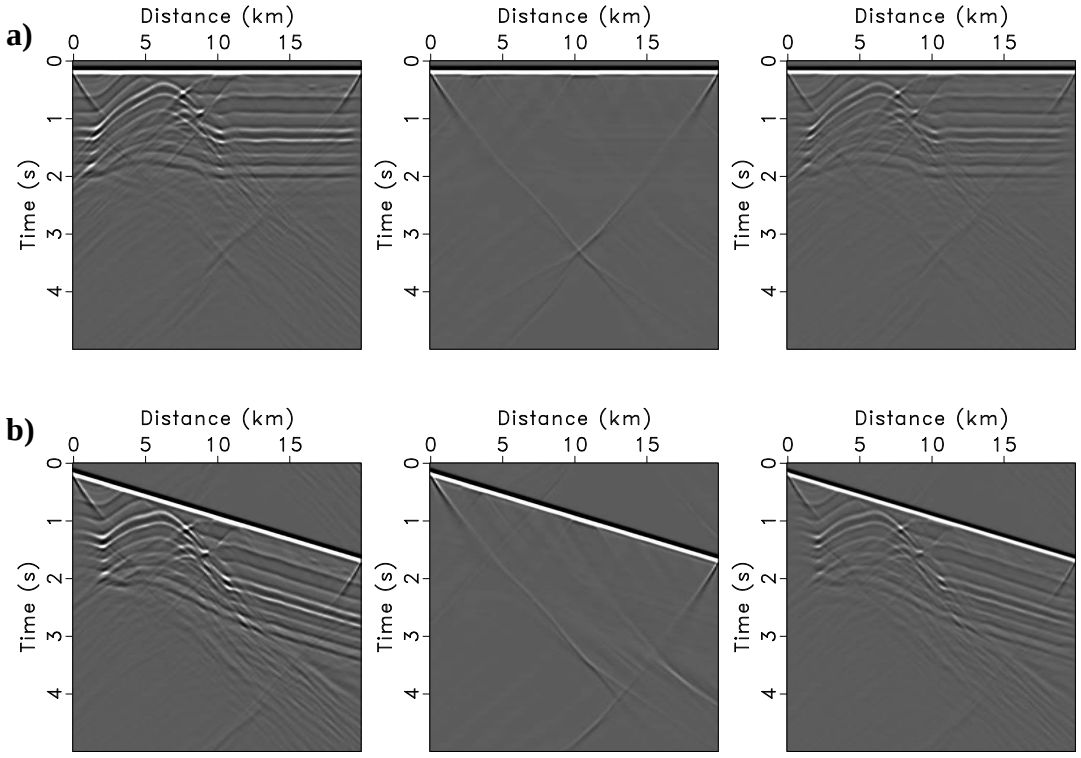

Figure 10. Time domain plane-wave gather for (a) $p_s = 0$ s/km and (b) $p_s = 0.3$ s/km. Left: true model. Middle: initial model. Right: inverted model.

CONCLUSIONS

In this study, we have presented a plane-wave encoding strategy for efficient frequency-domain full waveform inversion (FWI). This approach employs a simultaneous-shot approach for gradient calculation and a source and receiver encoding scheme for Hessian calculation. Compared to random phase encoding approach, it does not introduce strong cross-talk artifacts. Efficiency of this method is achieved by the use of a limited range of ray parameters for wavefield simulations. This encoding is robust with respect to noise. Directional illumination can be obtained with the selection of ray parameters, which may result in more savings of the computational time. Compared to time domain methods, frequency domain forward modeling can model multiple frequencies efficiently. Thus, this approach provides an efficient and robust tool for full waveform inversion.

ACKNOWLEDGMENTS

Yi Tao is supported by a ConocoPhillips fellowship and by the funding from Exploration Development Geophysics Education and Research (EDGER) Forum, the University of Texas at Austin. The Overthrust model is provided by SEISCOPE consortium http://seiscope.oca.edu. We thank Rui Zhang and Mohammed Alhussian for useful discussions.

REFERENCES

Amestoy, P. R., A. Guermouche, J. Y. L'Excellent, and S. Pralet, 2006, Hybrid scheduling for the parallel solution of linear systems: Parallel computing, 32, 136–156.

Balay, S., K. Buschelman, W. D. Gropp, D. Kaushik, L. C. McInnes, and B. F. Smith, PETSc home page. http://www.mcs.anl.gov/petsc, 2001.

Boonyasiriwat, C., P. Valasek, P. Routh, W. Cao, G. T. Schuster, and B. Macy, 2009, An efficient multiscale method for time-domain waveform tomography: Geophysics, 74, WCC59-WCC68.

Ben-Hadj-Ali, H., S. Operto and J. Virieux, 2011, An efficient frequency-domain full-waveform inversion method using simultaneous encoded sources: Geophysics, 76, R109–R124.

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

Brossier, R. and P. Roux, 2011, Hierarchical waveform inversion with double beamforming. 81st Annual International Meeting, SEG, Expanded Abstracts, 2389-2394.

Capdeville, Y., Y. Gung, and B. Romanowicz, 2005, Towards global earth tomography using the spectral elementmethod:Atechnique based on source stacking: Geophysical Journal International, 162, 541–554.

Foster, D. J., and C. C. Mosher, 1992, Suppression of multiple reflections using the Radon transform: Geophysics, 57, 386–395.

Guitton, A. and E. Díaz, 2011, Attenuating crosstalk noise with simultaneous source full waveform inversion: Geophysical Prospecting, 60, 759-768.

Herrmann, F. J., Y. Erlangga, and T. Y. Lin, 2009, Compressive sensing applied to full-wave form inversion: 71st Annual International Conference & Exhibition, EAGE, Extended Abstracts, S016.

Hustedt, B., S. Operto, and J. Virieux, 2004, Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling: Geophysical Journal International, 157, 1269–1296.

Krebs, J., J. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M. D. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources: Geophysics, 74, WCC177–WCC188.

Liu, F., D. W. Hanson, N. D. Whitmore, R. S. Day, and R. H. Stolt, 2006, Toward a unified analysis for source plane-wave migration: Geophysics, 71, S129–S139.

Liu, F., N. D. Whitmore, D. W. Hanson, R. S. Day, and C. C. Mosher, 2004, The impact of reciprocity on prestack source plane wave migration: 74th Annual International Meeting, SEG, Expanded Abstracts, 1045–1048.

Ma, Y. and D. Hale, 2012, Quasi-Newton full-waveform inversion with a projected Hessian matrix: Geophysics, 77, R207-R216.

Marfurt, K., 1984, Accuracy of finite-difference and finite-elements modeling of the scalar and elastic wave equation: Geophysics, 49, 533–549.

Menke, W., 1984, Geophysical data analysis: Discrete inverse theory: Academic Press, Inc.

Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, 95, 339–353.

Operto, S., J. Virieux, P. Amestoy, J. Y. L'Excellent, L. Giraud, and H. Ben- Hadj-Ali, 2007, 3D finite-difference frequency-domain modeling of viscoacoustic wave propagation using a massively parallel direct solver: A feasibility study: Geophysics, 72, SM195–SM211.

Plessix, R.-E., and W. A. Mulder, 2004, Frequency-domain finite-difference amplitude-preserving migration: Geophysical Journal International, 157, 975–987.

Pratt, R. G., C. Shin, and G. J. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, 133, 341–362.

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model: Geophysics, 64, 888–901.

Romero, L. A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 2000, Phase encoding of shot records in prestack migration: Geophysics, 65, 426–436.

Sen. M. K., and A. Mukherjee, 2003, t-p analysis in transversely isotropic media: Geophysical Journal International, 154, 647-658.

Shin, C., S. Jang, D. J., Min, 2001, Improved amplitude preservation for prestack depth migration by inverse scattering theory: Geophysical Prospecting, 49, 592-606.

Shin, C., and Y. H. Cha, 2008, Waveform inversion in the Laplace domain: Geophysical Journal International, 173, 922–931.

Sil, S. and M. K. Sen, 2009, Seismic critical-angle anisotropy analysis in the tau-p domain: Geophysics, 74, A53–A57.

Sirgue L. and R. G. Pratt, 2004, Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies: Geophysics, 69, 231–248.

Stoffa, P. L., ed., 1989, Tau-p: A plane wave approach to the analysis of seismic data: Kluwer Academic Publishers.

Stoffa, P. L., M. K. Sen, R. Seifoullaev, R. Pestana, and J. Fokemma, 2006, Plane-wave Depth migration: Geophysics, 71, S261-S272.

Tang, Y., 2009, Target-oriented wave-equation least-squares migration/inversion with phase-encoded Hessian: Geophysics, 74, WCA95–WCA107.

Tang Y. and S. Lee, 2010, Preconditioning full waveform inversion with phase-encoded Hessian, 80th Annual International Meeting, SEG, Expanded Abstracts, 1034–1038.

Tarantola A. 1984. Inversion of seismic reflection data in the acoustic approximation: Geophysics, 49, 1259–1266.

Vigh, D., and E. Starr, 2008, 3D prestack plane-wave, full waveform inversion: Geophysics, 73, VE135–VE144.

Virieux J. and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: Geophysics, 74, WCC127–WCC152.

Whitmore, N., 1995, An imaging hierarchy for common angle plane wave seismograms: PhD thesis, University of Tulsa.

Zhang, Y., J. Sun, C. Notfors, S. H. Gray, L. Chernis, and J. Young, 2005, Delayed-shot 3d depth migration: Geophysics, 70, E21–E28.

Zhang, Y. and D. Wang, 2009, Traveltime information-based wave-equation inversion: Geophysics, 74, WCC27–WCC36.

APPENDIX A

Proof of the equivalence of shot-profile and plane-wave gradient

Incorporating the plane-wave encoding strategy, the forward simulation of the source wavefield is the solution to equation 4 and 5. It can be written as

$$\tilde{S}(\mathbf{x},\mathbf{p},\omega)=\begin{cases}\sum_{\mathbf{x_s}} S(\mathbf{x},\mathbf{x_s},\omega)\exp(i\omega\mathbf{p}(\mathbf{x_s}-\mathbf{x}_0)) & p_i\geq 0\\ \sum_{\mathbf{x_s}} S(\mathbf{x},\mathbf{x_s},\omega)\exp(i\omega\mathbf{p}(\mathbf{x_s}-\mathbf{x}_{\max})) & p_i<0\end{cases}, \tag{A-1}$$

where $S(\mathbf{x},\mathbf{x_s},\omega)$ is the source wavefield for the shot location at ($\mathbf{x_s}$,0) simulated with the shot-profile wave equation; $\tilde{S}(\mathbf{x},\mathbf{p},\omega)$ is the source wavefield with the plane-wave simulation.

Similarly, the back propagated data residual can be expressed as

$$\Delta\tilde{R}(\mathbf{x},\mathbf{p},\omega)=\begin{cases}\sum_{\mathbf{x_s}} \Delta R(\mathbf{x},\mathbf{x_s},\omega)\exp(i\omega\mathbf{p}(\mathbf{x_s}-\mathbf{x}_0)) & p_i\geq 0\\ \sum_{\mathbf{x_s}} \Delta R(\mathbf{x},\mathbf{x_s},\omega)\exp(i\omega\mathbf{p}(\mathbf{x_s}-\mathbf{x}_{\max})) & p_i<0\end{cases}. \tag{A-2}$$

After crosscorrelating the source wavefield and residual receiver wavefield, the gradient, with the ray parameter being $\mathbf{p}$, can be written as follows

$$\begin{aligned}\tilde{g}(\mathbf{x},\mathbf{p_s}) &= \sum_{\omega}\mathrm{Re}\left\{\omega^2|\omega| f(\omega)\tilde{S}(\mathbf{x},\mathbf{p_s},\omega)\Delta\tilde{R}^*(\mathbf{x},\mathbf{p_s},\omega)\right\}\\ &= \sum_{\omega}\sum_{\mathbf{x_1}}\sum_{\mathbf{x_2}}\frac{-2\omega^2|\omega| f(\omega)}{v^3(\mathbf{x})}S(\mathbf{x},\mathbf{x}_1,\omega)\Delta R^*(\mathbf{x},\mathbf{x}_2,\omega)\exp(i\omega\mathbf{p_s}(\mathbf{x}_1-\mathbf{x}_2))\end{aligned}, \tag{A-3}$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ represent different shot locations for the source wavefiled and residual receiver wavefield.

Note that

$$\sum_{p=-\infty}^{\infty}\exp(i\omega p(\mathbf{x}_1-\mathbf{x}_2))=\frac{1}{|\omega|}\delta(\mathbf{x}_1-\mathbf{x}_2). \tag{A-4}$$

Stacking equation A-3 over all ray parameters results in the plane-wave domain gradient

$$\begin{aligned}\tilde{g}(\mathbf{x}) &= \sum_{\omega}\sum_{\mathbf{x_1}}\sum_{\mathbf{x}_2}\sum_{|\mathbf{p_s}|=-\infty}^{\infty}\frac{-2\omega^2|\omega| f(\omega)}{v^3(\mathbf{x})}S(\mathbf{x},\mathbf{x}_1,\omega)\Delta R^*(\mathbf{x},\mathbf{x}_2,\omega)\exp(i\omega\mathbf{p_s}(\mathbf{x}_1-\mathbf{x}_2)) \\ &= \sum_{\omega}\sum_{\mathbf{x_1}}\sum_{\mathbf{x}_2}\frac{-2\omega^2 f(\omega)}{v^3(\mathbf{x})}S(\mathbf{x},\mathbf{x}_1,\omega)\Delta R^*(\mathbf{x},\mathbf{x}_2,\omega)\delta(\mathbf{x}_1-\mathbf{x}_2) \\ &= \sum_{\omega}\sum_{\mathbf{x_s}}\frac{-2\omega^2 f(\omega)}{v^3(\mathbf{x})}S(\mathbf{x},\mathbf{x}_\mathbf{s},\omega)\Delta R^*(\mathbf{x},\mathbf{x}_\mathbf{s},\omega)\end{aligned} \quad . \text{(A-5)}$$

Thus, we have proved that

$$\tilde{g}(\mathbf{x}) = g(\mathbf{x}) \, . \quad \text{(A-6)}$$

However, real application of the gradient calculation is to simultaneously back-propagate the residual wavefield at all the receiver locations. This will result in some different filtering effects between plane-wave encoded gradient and shot-profile gradient.

APPENDIX B

Proof of the equivalence of shot-profile plane-wave encoded diagonal Hessian

For writing simplicity, the proof shown below is based on the diagonal Hessian. However, it does not involve any modification to extend to the full Hessian matrix.

Similar to the relationship between the shot-profile and plane-wave extrapolated wavefield in the gradient calculation, the Green's function satisfies

$$\tilde{G}(\mathbf{x},\mathbf{p},\omega) = \begin{cases}\sum_{\mathbf{x_s}} G(\mathbf{x},\mathbf{x}_\mathbf{s},\omega)\exp(i\omega\mathbf{p}(\mathbf{x}_\mathbf{s}-\mathbf{x}_0)) & p_i \geq 0 \\ \sum_{\mathbf{x_s}} G(\mathbf{x},\mathbf{x}_\mathbf{s},\omega)\exp(i\omega\mathbf{p}(\mathbf{x}_\mathbf{s}-\mathbf{x}_{\max})) & p_i < 0\end{cases} . \quad \text{(B-1)}$$

The receiver-side plane-wave encoded Hessian is expressed as

$$\begin{aligned}\tilde{H}_0(\mathbf{x},\mathbf{p}_\mathbf{r}) &= \sum_{\omega}\sum_{\mathbf{x_s}}\mathrm{Re}\left\{\frac{4\omega^4|\omega||f(\omega)|^2}{v^6(\mathbf{x})}\left|G(\mathbf{x},\mathbf{x}_\mathbf{s},\omega)\right|^2\left|\tilde{G}(\mathbf{x},\mathbf{p}_\mathbf{r},\omega)\right|^2\right\} \\ &= \sum_{\omega}\sum_{\mathbf{x_s}}\mathrm{Re}\left\{\frac{4\omega^4|\omega||f(\omega)|^2}{v^6(\mathbf{x})}\left|G(\mathbf{x},\mathbf{x}_\mathbf{s},\omega)\right|^2\tilde{K}_\mathbf{r}(\mathbf{x},\mathbf{p}_\mathbf{r},\omega)\right\}\end{aligned} \, , \quad \text{(B-2)}$$

With

$$\tilde{K}_{\mathbf{r}}(\mathbf{x},\mathbf{p}_{\mathbf{r}},\omega)=\sum_{\mathbf{x}_1}\sum_{\mathbf{x}_2}G(\mathbf{x},\mathbf{x}_1,\omega)G^{*}(\mathbf{x},\mathbf{x}_2,\omega)\exp(i\omega\mathbf{p}_{\mathbf{r}}(\mathbf{x}_1-\mathbf{x}_2))\,. \qquad \text{(B-3)}$$

Integrating the equation B-3 over ray parameters, we get

$$\begin{aligned}\tilde{K}_{\mathbf{r}}(\mathbf{x},\omega)&=\sum_{\mathbf{x}_1}\sum_{\mathbf{x}_2}\sum_{\mathbf{p}_{\mathbf{r}}}G(\mathbf{x},\mathbf{x}_1,\omega)G^{*}(\mathbf{x},\mathbf{x}_2,\omega)\exp(i\omega\mathbf{p}_{\mathbf{r}}(\mathbf{x}_1-\mathbf{x}_2))\\&=\sum_{\mathbf{x}_1}\sum_{\mathbf{x}_2}G(\mathbf{x},\mathbf{x}_1,\omega)G^{*}(\mathbf{x},\mathbf{x}_2,\omega)\frac{1}{|\omega|}\delta(\mathbf{x}_1-\mathbf{x}_2)\\&=\sum_{\mathbf{x}_{\mathbf{r}}}\frac{1}{|\omega|}\left|G(\mathbf{x},\mathbf{x}_{\mathbf{r}},\omega)\right|^2\end{aligned}\quad , \qquad \text{(B-4)}$$

$$\begin{aligned}\tilde{H}_0(\mathbf{x})&=\sum_{\omega}\sum_{\mathbf{x}_{\mathbf{s}}}\sum_{\mathbf{x}_{\mathbf{r}}}\mathrm{Re}\left\{\frac{4\omega^4\left|f(\omega)\right|^2}{v^6(\mathbf{x})}\left|G(\mathbf{x},\mathbf{x}_{\mathbf{s}},\omega)\right|^2\left|\tilde{G}(\mathbf{x},\mathbf{x}_{\mathbf{r}},\omega)\right|^2\right\}.\\&=H_0(\mathbf{x})\end{aligned} \qquad \text{(B-5)}$$

Equation B-5 shows the equivalence of the receiver-side plane-wave encoded Hessian and shot-profile Hessian.

A simultaneous plane-wave encoding uses both source and receiver ray parameters. It can be formulated as

$$\begin{aligned}\tilde{\tilde{H}}_0(\mathbf{x},\mathbf{p}_{\mathbf{s}},\mathbf{p}_{\mathbf{r}})&=\sum_{\omega}\mathrm{Re}\left\{\frac{4\omega^4\left|\omega f(\omega)\right|^2}{v^6(\mathbf{x})}\left|\tilde{G}(\mathbf{x},\mathbf{p}_{\mathbf{s}},\omega)\right|^2\left|\tilde{G}(\mathbf{x},\mathbf{p}_{\mathbf{r}},\omega)\right|^2\right\}\\&=\sum_{\omega}\mathrm{Re}\left\{\frac{4\omega^4\left|\omega f(\omega)\right|^2}{v^6(\mathbf{x})}\left|\tilde{K}_{\mathbf{s}}(\mathbf{x},\mathbf{p}_{\mathbf{s}},\omega)\right|^2\left|\tilde{K}_{\mathbf{r}}(\mathbf{x},\mathbf{p}_{\mathbf{r}},\omega)\right|^2\right\}\end{aligned}\quad , \qquad \text{(B-6)}$$

With $\tilde{K}_{\mathbf{r}}(\mathbf{x},\mathbf{p}_{\mathbf{r}},\omega)$ given in equation B-3 and $\tilde{K}_{\mathbf{s}}(\mathbf{x},\mathbf{p}_{\mathbf{s}},\omega)$ given as follows

$$\tilde{K}_{\mathbf{s}}(\mathbf{x},\mathbf{p}_{\mathbf{s}},\omega)=\sum_{\mathbf{x}_3}\sum_{\mathbf{x}_4}G(\mathbf{x},\mathbf{x}_3,\omega)G^{*}(\mathbf{x},\mathbf{x}_4,\omega)\exp(i\omega\mathbf{p}_{\mathbf{s}}(\mathbf{x}_3-\mathbf{x}_4))\,, \qquad \text{(B-7)}$$

The integration relationship in B-3 also holds for $\tilde{K}_{\mathbf{s}}(\mathbf{x},\mathbf{p}_{\mathbf{s}},\omega)$, i.e.

$$\tilde{K}_{\mathbf{s}}(\mathbf{x},\omega)=\sum_{\mathbf{x}_{\mathbf{s}}}\frac{1}{|\omega|}\left|G(\mathbf{x},\mathbf{x}_{\mathbf{s}},\omega)\right|^2\,. \qquad \text{(B-8)}$$

Thus

$$\begin{aligned}\tilde{\tilde{H}}_0(\mathbf{x}) &= \sum_{\omega}\sum_{\mathbf{x_s}}\sum_{\mathbf{x_r}} \mathrm{Re}\left\{\frac{4\omega^4\left|f(\omega)\right|^2}{v^6(\mathbf{x})}\left|G(\mathbf{x},\mathbf{x_s},\omega)\right|^2\left|\tilde{G}(\mathbf{x},\mathbf{x_r},\omega)\right|^2\right\}. \\ &= H_0(\mathbf{x})\end{aligned} \qquad \text{(B-9)}$$

Equation B-9 proves the equivalence of the source and receiver plane-wave encoded Hessian and shot-profile Hessian.