Efficient full waveform inversion in the time-ray parameter domain using iteration-dependent sets of ray parameters

Wenyong Pan*, Kristopher A. Innanen, and Gary E. Margrave, CREWES, University of Calgary

SUMMARY

Full waveform inversion (FWI), which has been widely studied in recent years, still cannot be practiced in industry effectively. The reasons for this include computational cost, large memory requirements, slow convergence rate, cycle skipping, etc. In our implementation, a linear source encoding strategy is used for the gradient calculation in time-ray parameter domain. The plane wave encoding approach forms super-gathers by summing densely distributed individual shots, and can reduce the computational burden considerably. We also construct the diagonal part of the pseudo-Hessian using a hybrid source encoding method. The diagonal part of the encoded pseudo-Hessian is a good approximation to the full Hessian matrix, which preconditions the gradient. The preconditioning is equivalent applying a deconvolution imaging condition in prestack reverse time migration. To avoid the local minimum problem, a multi-scale approach in the time domain is employed, by (1) applying a low-pass filtering to the data residuals and (2) increasing the frequency bands step by step. This has been proved to be effective against cycle skipping. We assemble this suite of tools and carry out a numerical experiment with a modified Marmousi model, analyzing the effectiveness of this combination of strategies.

INTRODUCTION

Full waveform inversion (FWI), formulated as a least squares inverse problem, seeks to minimize the difference between the observed data and synthetic data (Virieux and Operto, 2009). It has drawn much recent attention for its potential for the estimation of the subsurface parameters with high resolution. To fulfill its potential, FWI algorithms must cope with significant issues: computational burden, slow convergence rate, cycle skipping being important amongst these.

The efficient gradient calculation strategy proposed by Lailly (1983) and Tarantola (1984), the adjoint state method, allows us to construct the gradient by finding the zero-lag cross-correlation between the forward modelled wavefields from each source, and the back-propagated wavefields from all receivers simultaneously. In this approach, we only need 2Np forward simulations to calculate the gradient, where Np indicates the number of sources. This is a dramatic savings in comparison to global methods; however, for a 3D or a large 2D problem, the computational cost remains very high. The simultaneous source technique has been introduced to address this problem further (Vigh and Starr, 2008; Krebs et al., 2009; Ben-Hadj-Ali et al., 2011), at the cost of the introduction of cross-talk artifacts. These artifacts can in turn be suppressed with a sufficient number of ray parameters, which are controlled by the take-off angle at the surface location (Tao and Sen, 2013). By combining these strategies, the number of forward simulations in one FWI iteration becomes 2Np, where Np is the number of ray parameters. Generally the ray parameter range can be determined by the geological structures of the target area.

In this research, we propose the gradient be calculated using one single ray parameter per FWI iteration, with the ray parameter value varied for different iterations. This means that we only need 2 simulations to calculate the gradient for one FWI iteration. The computational cost is reduced further within this strategy.

FWI suffers from a slow convergence rate. The gradient absent the inverse Hessian does not account for geometric spreading effects (Shin et al., 2001). Gradient-based methods, which replace the Hessian matrix with an identity matrix, suffer from crude scaling, which slows convergence. The un-scaled image can be enhanced considerably with the inverse Hessian (Pratt et al., 1998), but direct inclusion of the inverse Hessian matrix is computationally unfeasible. In this research, we include a preconditioning of the gradient using the diagonal part of the pseudo-Hessian (Shin et al., 2001) which serves as a good approximation to the full Hessian matrix. Preconditioning the gradient using the diagonal part of the pseudo-Hessian is equivalent to a deconvolution imaging condition in reverse time migration.

We adopt a multi-scale approach, moving from lower to higher frequencies as we iterate, to improve the chances that a global rather than local minimum is reached (Vigh and Starr, 2008). The multi-scale approach can be implemented in the frequency domain by selecting specific frequencies. Time domain multi-scale approaches work by through low-pass filtering of the data residual. It has been demonstrated that a multi-scale method is computationally efficient and converges faster than a conventional, single-scale method (Boonyasiriwat et al., 2009).

In this paper, the basic theory for least squares inverse problem is reviewed firstly. We review the plane wave encoded gradient and source encoded pseudo-Hessian approach, and analyze this assembly of FWI tools using a numerical example on a modified Marmousi model. The effectiveness of the proposed suite of strategies is discussed.

THEORY AND METHODS

The least-squares inverse problem

As a least-squares local optimization, full waveform inversion seeks to minimize the difference between the synthetic data and observed data (Lailly, 1983; Tarantola, 1984) and update the model iteratively. The misfit function $\phi$ is given in a least-
Efficient FWI in time-ray parameter domain

squares norm:

\[ \phi = \frac{1}{2} \int d\omega \left( \sum_{r, r_s} \| \delta P(r, r_s, \omega, s_0) \|_2 \right)^2, \]  

(1)

where \( r = (x, y, z) \) is the Euclidean coordinate of a subsurface position, \( r_s \) and \( r_g \) are the positions of the sources and receivers respectively, \( \omega \) is the temporal frequency, \( s_0(r) \) indicate the square of the slowness in the \( nth \) iteration. \( \delta P(r, r_s, \omega, s_0) \) mean the data residuals, the difference between the observed data and synthetic data, \( \| \cdot \|_2 \) indicates the \( \ell_2 \) norm.

The minimum value of the misfit function is sought in the vicinity of the starting model \( s_0(r) \) and the updated model can be written as the sum of the starting model and a model perturbation \( s_0^{(n)}(r) \) (Virieux and Operto, 2009).

\[ s^{(n)}(r) = s_0^{(n)}(r) + \mu^{(n)} \delta s_0^{(n)}(r), \]

(2)

where \( \mu^{(n)} \) is the step length in \( nth \) iteration, which is a scalar constant used to scale the model perturbation and can be obtained through a line search method (Gauthier et al., 1986; Pica et al., 1990).

Applying a second order Taylor-Lagrange development of the misfit function and then taking partial derivative with respect to the model parameters give the model perturbation as:

\[ \delta s_0^{(n)} = - \int d \delta' H^{(n)}(r', \delta') g^{(n)}(r), \]

(3)

where \( g^{(n)}(r') \) is the gradient and \( H^{(n)}(r', r') \) is the Hessian matrix.

For gradient, the first order derivative of the misfit function \( \phi \) with respect to the model parameters can be obtained by a zero-lag correlation between the data residuals and the sensitive matrix (or Jacobian matrix). While it is extremely expensive to calculate the sensitive matrix directly. If we apply a perturbation derivation and truncate the Born series based on the assumption of small model perturbation, the gradient can be written as:

\[ g^{(n)}(r) = - \sum_{r, r_s} \int d\omega \mathfrak{R} \left( \omega^2 \mathcal{F}_s(\omega) G(r, r_s) G(r_g, r) \delta P^* \right), \]

(4)

where \( G(r, r_s) \) and \( G(r_g, r) \) are the source-side and receiverside Green’s functions, respectively and \( \mathcal{F}_s(\omega) \) is the source signature. Then the gradient can be calculated using the adjoint state method by applying a zero-lag convolution between the forward modeling wavefields and back-propagated data residuals, which avoids the direct computation of the partial derivative wavefields.

Plane wave encoded gradient

A linear source encoding method is employed in this research to construct the gradient. Generally, in one FWI iteration, a summation of the gradients by different ray-parameters should be performed to balance the amplitudes. Here, we propose to apply just one simulation for one iteration but vary the ray parameters for different iterations. Thus, the model update can be balanced during a group of iterations with increasing the computational efficiency. And we can formulated the plane wave encoded gradient in one iteration as:

\[ g^{(n)}(r) = - \int d\omega \mathfrak{R} \left\{ \omega^2 \mathcal{F}_s(\omega) G(r, r_s) G(r_g, r) \delta P^* \right\}, \]

(5)

where \( G(r, r_s) = G(r, r_s, e^{i p_s (x_s' - x_s)} \) \( p_s \) is the ray parameter, \( x_s' \) and \( x_s \) are the sources’ coordinates.

Source encoded pseudo-Hessian

The full Hessian matrix is the second order partial derivative of the misfit function with respect to the model parameters. Multiplying the inverse Hessian matrix can compensate the geometrical spreading effects and recover the amplitude especially for the deep reflectors. Unfortunately the explicit calculation of the Hessian matrix is also too expensive to be practiced. The full Hessian matrix can be written as a summation of two terms. The first term, which accounts for the second order multiple scattering, is always neglected. Thus, the second term forms the approximate Hessian used in Gauss-Newton method:

\[ H_a^{(n)} = \sum_{r, r_s} \int d\omega \mathfrak{R} \left\{ \omega^4 G(r, r_s) G(r', r_s) G^*(r_g, r') G^*(r', r_s) \right\} \]

(6)

We can notice that the approximate Hessian consists of two source-side Green’s functions and two receiver-side Green’s functions. And the assumption of the infinite receiver coverage, the receiver-side Green’s functions can be approximated as a constant scalar. This approximation to Hessian matrix is equivalent to the pseudo-Hessian proposed by Shin et al. (2001). The pseudo-Hessian is constructed using the forward modeling wavefields, used as the virtual sources in reverse time migration.

\[ H_p^{(n)} = \sum_{r, r_s} \int d\omega \mathfrak{R} \left\{ \omega^4 G(r, r_s) G^*(r', r_s) \right\} \]

(7)

We can also recognize that the when \( r'' = r' \), the diagonal part of the pseudo-Hessian \( H_p^{(n)} \) is equivalent to the source illumination. So, preconditioning the gradient using the reciprocal of the diagonal pseudo-Hessian is identical to the standard deconvolution imaging condition in reverse time migration.

The hybrid phase encoding strategy used in this research is a combination of linear phase encoding and random phase encoding methods. And it is similar to Perrone and Sava’s dithered phasing encoding strategy (Perrone and Sava, 2012):

\[ H_{pre}^{(n), p} = \sum_{p_r, p_s} \int d\omega \mathfrak{R} \left\{ \omega^4 G(r''_p, r_s) G^*(r', r_s) e^{i \Delta p (x_s' - x_s)} \right\}, \]

(8)

where \( p_r \) are the ray parameters, \( \Delta p = rand \times p_s, rand \in (0, 1) \) is the random coefficient, \( \epsilon \) is the coefficient used to control the amount of dithering. In this research, we call the method with pseudo-Hessian preconditioning as preconditioned method for comparing to the traditional gradient based method.
Efficient FWI in time-ray parameter domain

Multiscale approach
We increase the frequency ranges from lower to higher step by step. The low frequency bands are responsible to catch the low wavenumber component of the velocity model. While the high frequency bands are used to add high wavenumber details to the velocity model. The lowest frequency used in each scale step is held constant from scale step to scale step (Vigh and Starr, 2008). Compared to multiscale method in frequency domain, it doesn’t take advantage of wavenumber redundancy, but it is more robust.

NUMERICAL EXPERIMENTS
The strategies proposed in this research are practiced and applied on a modified Marmousi Model. The Marmousi model is modified by introducing one water layer with a thickness of 70m and a P-wave velocity of 1500m/s. And the model has 180 × 767 grid cells with the same grid interval of 5m in horizontal and vertical. 380 point sources are distributed on the surface with a source interval of 10m from 0 to 3800m and 767 receivers are deployed on the surface with a receiver interval of 5m from 5m to 5750m. The source function is a Ricker wavelet with a dominant frequency of 30Hz. The ray parameter range used for linear phase encoding method is $[-0.3s/km, 0.3s/km]$ with a step of 0.1s/km are tested for comparison. The lowest frequency band used is $[0Hz, 5Hz]$ and the frequency band increases by 5Hz for every 10 iterations. Fig.1a shows the exact P-wave velocity model and Fig.1b shows the smoothed P-wave velocity model used as the reference model in this research. Fig.2a shows the observed shot record when ray parameter $p = 0$ for the true velocity model. Fig.2b is the synthetic shot record when ray parameter $p = 0$ for the initial velocity model. Fig.2c is the corresponding data residual which is the difference of observed data and synthetic data. Fig.2d is the data residual after applying a low-pass filtering with a band of $[0, 5Hz]$.

Figure 1: (a) is true velocity model; (b) is the initial velocity model.

Figure 2: Shot records when ray parameter $p = 0$. (a) is the observed shot record for the true velocity model; (b) is the shot record for the initial velocity model; (c) is the data residual in the first iteration; (d) is the data residual after low-pass filtering with a band of $[0, 5Hz]$.

Figure 3: FWI results after 50 iterations with single ray parameter in one FWI iteration. (a) ray parameter $p = 0s/km$; (b) ray parameter $p = -0.2s/km$; (c) ray parameter $p = 0.2s/km$; (d) ray parameter varies from $-0.3s/km$ to $0.3s/km$ with a step of 0.1s/km.
Efficient FWI in time-ray parameter domain

Fig. 4: Diagonal part of the pseudo-Hessian. (a) is the exact diagonal part of the pseudo-Hessian constructed by 380 simulations; (b) is the diagonal part of the encoded pseudo-Hessian using a hybrid source encoding method by 15 simulations.

Fig. 5: Final inverted results after 200 iterations. (a) is the inverted model by the gradient based method; (b) is the inverted model by the preconditioned method.

Fig. 6: Comparison of the inverted results at 0.5km in distance. The red-solid line is the true velocity model. The black-solid line indicates the initial velocity model. The green-dash-dot line is the inverted model by gradient based method. And the blue-dash line is the inverted model by the preconditioned method.

CONCLUSION

From what we have discussed above, the conclusions can be achieved that the source encoding strategy can reduce the computational burden of FWI drastically and by varying the ray parameters for different iterations as proposed in this research, the computational cost is reduced further and the model can be updated with balance. Preconditioning the gradient using the diagonal part of the encoded pseudo-Hessian compensate the energy loss during wave propagation. Thus, the convergence rate of FWI can be improved. The time domain multiscale ap-

ACKNOWLEDGMENTS

This work was supported by CREWES sponsors. And I wish to thank Danping Cao, Raul Cova and Babatunde Arenrin’s discussions with me.
REFERENCES


