DOI:10.1190/segam2021-3582659.1

Reflection FWI with an augmented wave equation and quasi-Newton adaptive gradient scheme

James McLeman, Tim Burgess, Mrinal Sinha, Gary Hampson, Troy Thompson, DownUnder GeoSolutions*

Summary

Conventional Full Waveform Inversion (FWI) produces high-resolution model updates from diving wave energy in recorded seismic data. However, the penetration depth of refraction energy is limited by the maximum recorded offset and the Earth's velocity field, so we rely on reflected energy to update the deeper parts of the model. The inclusion of reflections in FWI requires special care due to the relatively weak amplitude of the "rabbit ear" tomographic term. We propose a novel reflection FWI (R-FWI) simultaneous inversion for velocity and derivative quantities which may include properties related to amplitude variation with angle of incidence (AVA). Such multi-parameter inversions face difficulties with parameter scaling, Hessian estimation and slow convergence rates. Therefore, we also propose an efficient adaptive gradient extension to a conventional quasi-Newton optimization scheme to complement our multi-parameter approach. We demonstrate these new approaches on data from the North-West shelf of Australia.

Introduction

FWI has become a common method for determining a range of different subsurface parameters from seismic data, e.g., velocity, anisotropy and Q-absorption. Its ability to produce high-resolution model updates and fill the resolution gap between traditional tomographic inversion and the migration reflectivity makes it an essential tool in the model building workflow. Conventionally, FWI relies on the difference between the recorded and modelled diving waves to drive a velocity update, which yields the classic "banana" shaped gradient. However, the penetration depth of diving waves is typically only 1/3 to 1/5 of the maximum offset (Zhou et al., 2015), which is inadequate for deeper targets of interest. One solution is to acquire longer offsets; however, this comes with attendant cost and operational drawbacks. R-FWI offers a solution to this penetration problem, but requires special care during implementation (Yao et al., 2020).

Modelling data with a smooth model produces only transmitted waves. If the input data to FWI contains only reflections, the initial gradient will contain only the higher-wavenumber 'events' (the "migration" term) due to the cross-correlation imaging condition. These events will have amplitude proportional to reflection coefficient, *R*. Subsequent iterations will scatter the wavefield from these events and additionally produce the lower wavenumber tomographic updates (the "rabbit ears" term) responsible for the kinematics with amplitude of order R^2. The inversion will therefore preferentially invert for the stronger migration term. Meaningful tomographic updates can be achieved via simultaneous or cascaded multi-parameter inversion such as velocity and a perturbed parameter (Xu et al., 2012), velocity-reflectivity (Yao et al., 2014) or velocity-density (Wang et al., 2015). Their gradients are decomposed so that the migration term is placed in reflectivity or density and the tomographic term into velocity. These parameterizations, however, face difficulties. The reflectivity or perturbation approach typically involves the Born approximation (linearization) of the scattering problem; thus, the input data should be free of multiples and ghosts. This data is not likely to be available at the beginning of a project. Alternatively, a velocity-density formulation where only the migration term is kept in density means it acts only to scatter the wavefield rather than contain useful lower frequency information.

It is well known that there is crosstalk between parameters in multi-parameter inversion if the quantities are coupled (Operto et al., 2013). There is also slow convergence of weaker parameters when parameters have very different scales. These issues cannot be overcome using steepest descent, even with enhancements from machine learning, such as "momentum" or adaptive gradient (AdaGrad) (Duchi et al., 2011) schemes. The absence of curvature information, that higher order methods could provide, renders these approaches undesirable for complex non-linear problems. Higher order quasi-Newton methods which have realistic memory footprints, such as L-BFGS, are viable because they use diagonal plus low-rank approximations to the Hessian. Although this improves convergence compared to steepest descent, it still fails to adequately decouple model parameters or address parameter scaling issues.

We demonstrate a new multi-scattering reflection FWI that simultaneously solves for velocity and vector derivative quantities which can have AVA-like properties. The convergence rate of this multi-parameter inversion is accelerated and parameter scaling issues addressed using a quasi-Newton adaptive gradient scheme that is equally well suited to single parameter problems.

Reflection FWI with an augmented wave equation

We begin with the variable density acoustic wave equation, which is given by,

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial^2 t} - \rho\nabla\cdot\left(\frac{1}{\rho}\nabla u\right) = \rho s\delta(\boldsymbol{x} - \boldsymbol{x_0}) \qquad (1)$$

where c is the P-wave velocity, ρ is density, s is the source term, δ acts at the source position and u is the forward

DOI:10.1190/segam2021-3582659.1

pressure wavefield. Equation (1) can be re-written equivalently as,

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial^2 t} + \nabla \log \rho \cdot \nabla u - \nabla^2 u = \rho s \delta(\boldsymbol{x} - \boldsymbol{x_0}). \quad (2)$$

In this form, computation of nested spatial derivatives at each timestep is avoided. Instead, $\nabla \log \rho$ can be precomputed and stored in memory. Equation (2) provides insight into the role of density in the wave equation. It indicates that, away from *s*, only contrasts in $\log \rho$ matter and there is no zero-wavenumber information about density available from the wavefield. This is consistent with formulations of reflectivity, where only ratios involving density contribute (Li and Peng, 2017). Note, the similarity between the density contribution above and the intercept in the AVA equation,

$$R(\theta = 0) = \frac{1}{2}\left(\frac{\Delta c}{c} + \frac{\Delta \rho}{\rho}\right) \cong \frac{1}{2}(\nabla(\log c + \log \rho)). \quad (3)$$

We can generalize equation (2) to represent the dot product of a spatially varying vector parameter ∇r_i and the gradient of the scalar wavefield ∇u,

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial^2 t} + \nabla r_i \cdot \nabla u - \nabla^2 u = \rho s \delta(\boldsymbol{x} - \boldsymbol{x_0}). \quad (4)$$

The ∇r_i term can be the intercept-reflectivity given by equation (3). It follows that we could modify the wave equation by adding additional scattering terms to serve other purposes such as including the AVA gradient term,

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial^2 t} + \nabla r_i \cdot \nabla u + \nabla r_g sin^2\theta \cdot \nabla u - \nabla^2 u = \rho s \delta(\boldsymbol{x} - \boldsymbol{x_0}) \quad (5)$$

In general, any number of scattering, ∇r_j, and directivity, d_j, terms can be included (e.g., AVA curvature term),

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial^2 t} + \sum_{j=0}^{m} \nabla r_j \cdot d_j \nabla u - \nabla^2 u = \rho s \delta(\boldsymbol{x} - \boldsymbol{x_0}) \quad (6)$$

The ∇r_j terms provide the scattering needed to generate the reflection tomographic term (rabbit ears).

Given an objective function (*F*) of the form,

$$\mathrm{F} = \frac{1}{2}\|\boldsymbol{d} - \boldsymbol{L}(\boldsymbol{m})\|_2^2 \quad (7)$$

where $\boldsymbol{d}$ is the observed data, $\boldsymbol{L}$ is the wave equation operator and $\boldsymbol{m}$ are models of interest, we can compute gradients with respect to velocity and ∇r_j via the adjoint-state method.

The separation of migration and tomographic terms are achieved using non-stationary filters discriminating by scattering angle, which is applied as a wavenumber domain preconditioner. Further constraints can be included such as a structurally oriented smoothing approach applied to velocity (Fehmers & Höcker, 2003).

Note that the observed data $\boldsymbol{d}$ contains ghosts, multiples and the source wavelet but the inverted quantity ∇r_j is a migrated image with these attenuated. This inversion performs deghosting, demultiple and designature. Running R-FWI to high frequency provides reflectivity image(s) for both structural and quantitative interpretation, without the need for a conventional processing and imaging workflow.

Although this approach applies easily to any anisotropic symmetry, for simplicity, we have only presented the isotropic case. We also note that it is easily extended to least squares reverse-time migration.

Because the foregoing has increased the number of parameters to be inverted for, it is now imperative that we deal with the problems associated with multi-parameter inversion as discussed in the introduction.

Quasi-Newton adaptive gradient scheme

L-BFGS is a second order quasi-Newton optimizer which approximates a diagonal plus low-rank inverse Hessian matrix. A step in this algorithm estimates the diagonal component of the inverse Hessian from a scaled identity matrix based on the curvature along the most recent search direction (Nocedal and Wright, 2006),

$$\alpha = \frac{\boldsymbol{s}^T\boldsymbol{y}}{\boldsymbol{y}^T\boldsymbol{y}} \quad (6)$$

where $\boldsymbol{y}$ and $\boldsymbol{s}$ are the change in gradient and change in model, respectively, since the previous iteration. Note that L-BFGS' positive definite requirement necessitates $\alpha > 0$. Since the Wolfe conditions mean that the curvature condition $\mathbf{s}^T\mathbf{y} > 0$, and so must $\mathbf{y}^T\mathbf{y} > 0$, which guarantees that $\alpha > 0$. Equation (6) results in a single scalar to modify the Hessian so that a unit step length is likely to be accepted.

In a multiparameter inversion where the inverted parameters are of different scales, a single scalar, α, is inappropriate as it will be biased towards the stronger parameter(s). Additionally, in both single and multi-parameter inversion, the inverse Hessian controls the compensation for variations in illumination. It is typically the case that the eigenvalues of the Hessian matrix contain a wide range of different values, causing slow convergence in regions of the inverted model and between parameters. Typically, image domain preconditioners are used in conjunction with L-BFGS to

10.1190/segam2021-3582659.1

DOI:10.1190/segam2021-3582659.1

improve convergence and help with illumination compensation because the quasi-Newton optimization alone can take many iterations to resolve acquisition sampling issues. As a result, there will be poor convergence in model locations where the preconditioning is insufficient. Unfortunately, this is likely to correspond with complex deeper structures where the R-FWI update is most needed.

An improvement is to choose $q = pm$ scalars where p is the number of inverted parameters and m is the number of model points per parameter. This is achieved using an AdaGrad-type scheme to improve the L-BFGS inverse Hessian estimate. We propose to replace $\alpha\mathbf{I}$ with a new quantity $\beta\mathbf{M}$,

$$\beta = \frac{\boldsymbol{y}^T\boldsymbol{s}}{\boldsymbol{y}^T\boldsymbol{M}\boldsymbol{y}} \tag{7}$$

$$\mathbf{M} = \left[\mathrm{diag}\left(\sum_{i=0}^{n}\mathbf{g_i}\mathbf{g_i^T}\right) + \varepsilon\right]^{-\frac{1}{2}} \tag{8}$$

where n is the iteration number, $\boldsymbol{g}$ is the gradient and ε is some small number to stabilize the inverse. Like α, β is guaranteed to be greater than 0 and ensures the positive definite property required by L-BFGS. This general approach is not limited to only improving the diagonal estimate of the Hessian: off-diagonal terms and matching filters can be included to help reduce parameter crosstalk.

Using this formulation, initially each parameter class, indeed, every element in the model, will receive an equal weight in the inversion. With each subsequent iteration every element in the model will receive updated weights that modify the sensitivity of the objective function to each element of the model so that the relative sensitivities and coupling of the various parameters are compensated for. This overcomes parameter bias and avoids the need for illumination compensation preconditioners. Regularization of the inverse Hessian is incorporated to avoid biasing the inversion towards noise.

Real data example

We applied our new approach to a dual-source / 8-streamer dataset from the Australian NW continental shelf. The initial velocity model was tomographically derived from residual moveout picks. Anisotropy was initially parameterized using wells, with epsilon further refined using tomography. Several rounds of diving wave FWI were then performed increasing from 4 Hz up to 12 Hz. We then applied reflection FWI to update the deeper parts of the model.

In Figure 1, using our new R-FWI formulation, we compare the convergence rates of the AdaGrad + L-BFGS scheme to a standard L-BFGS optimizer. The cost function achieved using the standard L-BFGS approach after 15 iterations is achieved using AdaGrad + L-BFGS in ~7 iterations.

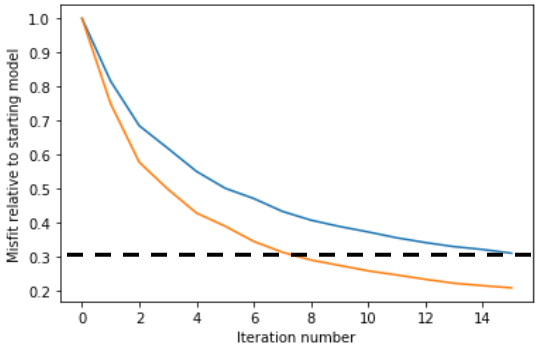

Figure 1: Convergence of conventional L-BFGS (blue) versus AdaGrad + L-BFGS (red).

Figures 2a & b show Kirchhoff migrated common image gathers with the initial and updated velocity models. The new reflection FWI model demonstrates a clear improvement in gather flatness.

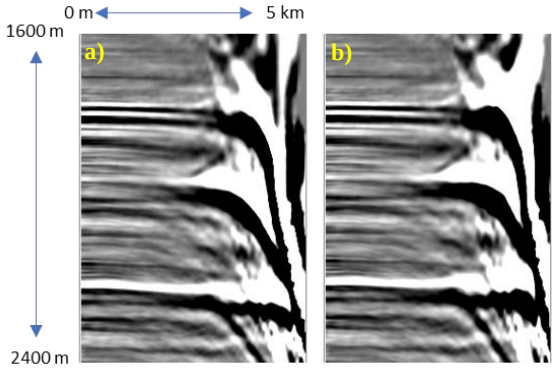

Figure 2: a), b) Kirchhoff migrated common image gathers with the initial and updated velocity model, respectively.

The initial and updated models overlain on their Kirchhoff preSDM images are shown in Figures 3a & b. The updated velocity model demonstrates good geological conformance, as shown by the black arrows. Consequently, their stacks in Figures 4a & b are better focused and deeper events appear structurally simplified, with the flat spot becoming flatter. The green arrows show areas of notable improvement, the red arrows show the same location with the initial model.

Figures 5a & b show a depth slice at 1340 m (~130 m below seabed) between a Kirchhoff preSDM stack with the updated velocity and the inverted intercept-reflectivity. The Kirchhoff stack was filtered to the intercept-reflectivity bandwidth and is using processed data as input. We observe a good correlation between them but better event delineation in the latter despite only using 1/3 of the unprocessed shots.

10.1190/segam2021-3582659.1

Conclusions

We have described a novel solution to the reflection FWI problem. This approach involves a simultaneous inversion for velocity and a vector derivative quantity with intercept-like scattering properties. This can be extended to include any number of scattering terms in general. An application of this method involving velocity and intercept-reflectivity was demonstrated on real data where the velocity update showed a clear improvement in imaging. Despite the input being raw recorded data, the intercept-reflectivity output has attenuated multiples, ghosts and the source wavelet and compares well with the band-passed Kirchhoff migration which has processed gathers as input. This solution was implemented with an efficient generic multi-parameter inversion scheme that accelerates convergence by resolving parameter scale differences and providing a better estimate of the Hessian matrix. It is equally applicable to single parameter class inversions as it can better compensate for illumination and does not rely on image domain amplitude compensation preconditioners.

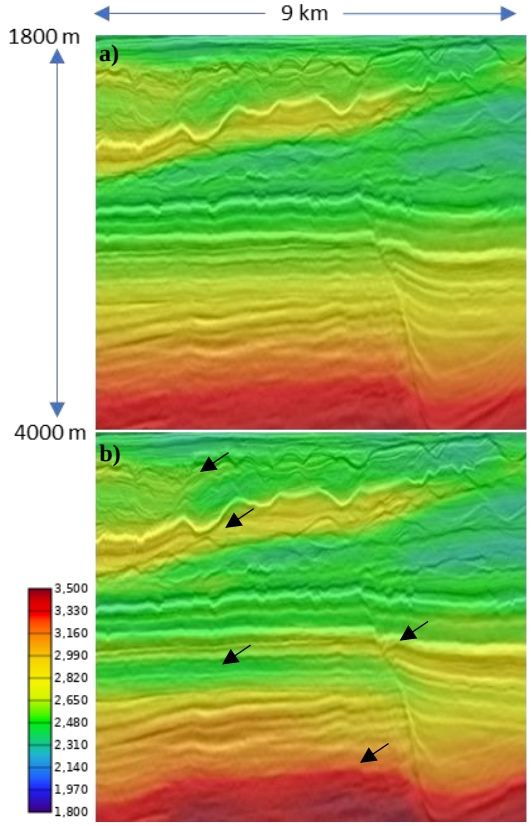

Figure 3: a), b) Kirchhoff preSDM stacks using initial and updated velocity, respectively, overlain with their velocity model.

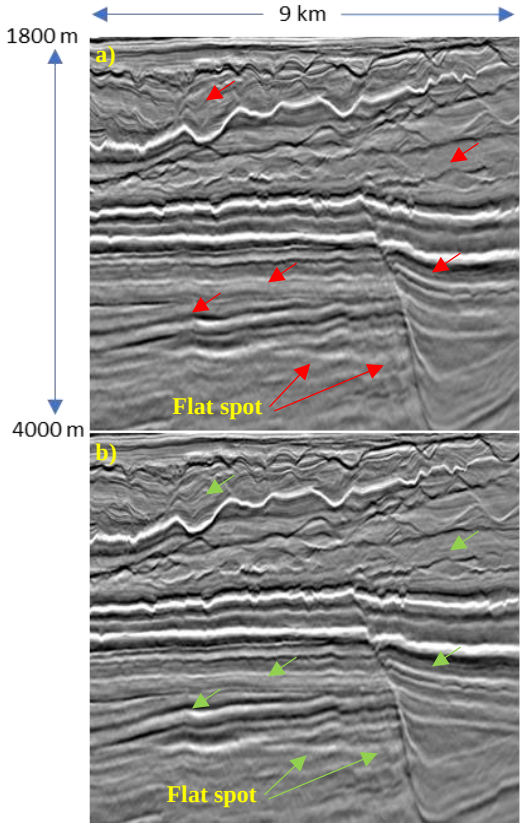

Figure 4: a), b) Kirchhoff preSDM stacks using initial and reflection FWI updated velocity, respectively.

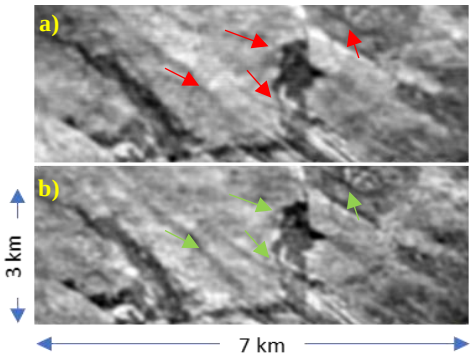

Figure 5: A depth slice at 1340 m of a) the Kirchhoff preSDM using the updated velocity model, and b) the intercept-reflectivity computed using the proposed inversion method.

Acknowledgements

We would like to thank Multi-Client Resources (MCR) for permission to use the BEX MC3D real data example and DownUnder GeoSolutions (DUG) for allowing us to present this work.

REFERENCES

Duchi, J., E. Hazan, and Y. Singer, 2011, Adaptive subgradient methods for online learning and stochastic optimization: Journal of Machine Learning Research, **12**, 2121–2159.

Fehmers, G. C., and C. F. W. Höcker, 2003, Fast structural interpretations with structure-oriented filtering: Geophysics, **68**, 1286–1293, doi: https://doi.org/10.1190/1.1598121.

Li, S., and Z. Peng, 2017, Seismic acoustic impedance inversion with multi-parameter regularization: Journal of Geophysics and Engineering, **14**, 520–532, doi: https://doi.org/10.1088/1742-2140/aa5e67.

Nocedal, J., and S. J. Wright, 2006, Numerical optimization: Springer, 178.

Operto, S., Y. Cholami, V. Prieux, A. Ribodetti, R. Brossier, L. Metivier, and J. Virieux, 2013, A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice: The Leading Edge, **32**, 1040–1054, doi: https://doi.org/10.1190/tle32091040.1.

Wang, H., S. Singh, F. Audebert, and H. Calandra, 2015, Inversion of seismic refraction and reflection data for building long-wavelengths velocity models: Geophysics, **80**, no. 2, R81–R93, doi: https://doi.org/10.1190/geo2014-0174.1.

Xu, S., D. Wang, F. Chen, G. Lambaré, and Y. Zhang, 2012, Inversion on reflected seismic wave: 72nd Annual International Meeting, SEG, Expanded Abstracts, 1–7, doi: https://doi.org/10.1190/segam2012-1473.1.

Yao, G., M. Warner, and A. Silverton, 2014, Reflection FWI for both reflectivity and background velocity: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, 1–5, doi: https://doi.org/10.3997/2214-4609.20141089.

Yao, G., D. Wu, and S. X. Wang, 2020, A review on reflection-waveform inversion: Petroleum Science, **17**, 334–351, doi: https://doi.org/10.1007/s12182-020-00431-3.

Zhou, W., R. Brossier, S. Operto, and J. Virieux, 2015, Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation: Geophysical Journal International, **202**, 1535–1554, doi: https://doi.org/10.1093/gji/ggv228.