Using multi-resolution truncated Newton optimization for cross-talk reduction in FWI

Scott Keating and Kris Innanen

ABSTRACT

Cross-talk, where data signatures of different physical properties are confused, is a major concern in multi-parameter FWI. This can be mitigated by using a good estimate of the Newton update in the inversion procedure, but such an approach is typically too computationally intensive to be pursued. In this report, the cost of approximating the Newton update is reduced by considering a multi-resolution approach, in which the grid defining the model is varied with frequency. This approach allows for a smaller computational burden at low frequencies, and effectively mitigates the cost of approximating the Newton update. This allows for cross-talk to be more effectively prevented.

INTRODUCTION

When full waveform inversion (FWI) is used to recover multiple physical properties of the subsurface, there is a risk of confusing the data signatures of different properties. This can result in cross-talk, where different variables are confused in the inversion result. The prevention of cross-talk is a major priority in multi-parameter FWI, and is the focus of substantial research. Consideration of the second derivative of the objective function in the model update is well understood to play an important role in the prevention of cross-talk. Unfortunately, making use of this second derivative information is very computatationally intensive.

Keating and Innanen (2017) discussed the multigrid approach, a strategy for reducing the computational expense of minimizing cross-talk at the long wavelength scales in the model. In this approach, the grid on which the model was defined was changed during the inversion based on the frequencies being considered at each stage of the inversion. By considering a smaller inversion problem at the low frequencies more powerful optimization techniques were made feasible, and this led to a reduction in cross-talk on the corresponding model scales. The main focus of Keating and Innanen (2017) was on the use of exact Gauss-Newton optimization. While this approach is very powerful, the frequencies needed for practical application may be extremely low. Application of this approach can be applied to less demanding optimization approaches. This report builds on the previous work by investigating numerical simulations of the multigrid approach when using the truncated Gauss-Newton (TGN) method for numerical optimization.

THEORY

The multi-resolution approach

In conventional FWI using two dimensional, finite-difference wave modeling, the variables considered in the inversion are coefficients. These coefficients describe the values of physical properties in regions of size dz by dx in the subsurface, where dz and dx are the position and depth increments used in the forward modeling. In effect, the inversion is for the variables used in the finite-difference wave propagation calculations. When the problem is extended to three dimensions, or finite element modeling is used, the variables used are still just the coefficients used in the forward modeling engine. The multi-resolution approach is a strategy based on the work of Bunks et al. (1995) that was proposed by Keating and Innanen (2017) which uses a different parameterization for the inversion. In this approach, the variables used in the inversion are still coefficients for a region of the model, but this region is allowed to have a length of $\frac{\lambda}{8}$, where λ is the wavelength of the highest frequency considered at a given iteration in the background medium. This is different from the conventional approach because the frequencies considered change at each iteration of the FWI problem, and so at low frequencies much larger regions are defined in the multi-resolution inversion. The conventional model a is related to the multi-resolution model m at each iteration through the relation

$$a = Pm, \tag{1}$$

where P is a matrix defining the multi-resolution variables. Because the variables defined in P are spatially large at low frequencies, m has far fewer elements than a at these frequencies. Keating and Innanen (2017) showed that the smaller problem size allows for more effective optimization techniques to be used, allowing for multiparameter FWI problems like cross-talk to be more effectively prevented.

Truncated Newton optimization

Full waveform inversion is typically posed as an optimization problem, in which a scalar objective function quantifying the current level of mismatch between predicted data and those measured is minimized. The minimization of this objective function requires the use of an optimization strategy. In Newton optimization descent directions are generated by solving the system

$$H\Delta m = -g, \tag{2}$$

where g is the gradient of the objective function, H is the Hessian (second derivatives), and Δm is the model update being calculated. Solving this system is very computationally intensive for a large, non-sparse H. The main purpose of Keating and Innanen (2017) was to develop an optimization strategy in which Newton optimization could be used in the multiparameter FWI problem for long length scales. The multi-grid approach described there defines a model parameterization which changes with frequency. At low frequencies the model can be characterized by relatively few variables describing large spatial regions of the model. This allows for the use of Newton optimization at very low frequencies, where the number of required variables is quite small. In this way the approach is similar to the global optimization strategy of Debens et al. (2015). The multi-resolution approach is not specific to Newton optimization however, and offers efficiency improvements for other optimization strategies, including the truncated Gauss-Newton method.

In truncated Newton optimization the solution of equation 2 is approximated instead of being solved directly. The truncated Gauss-Newton (TGN) method approximates the solution of the same system, but with H replaced by H_{GN}, the residual independent part of the Hessian. The solution to this system is approximated through iterative minimization of the function

$$\phi(\Delta m) = \frac{1}{2}\Delta m^T H_{GN} \Delta m + \Delta m^T g \tag{3}$$

with respect to the update Δm. This problem is equivalent to solving equation 2 because at a minimum $\frac{\partial \phi}{\partial \Delta m} = 0$, and

$$\frac{\partial \phi}{\partial \Delta m} = H_{GN} \Delta m + g. \tag{4}$$

This problem is referred to as the 'inner loop', because its solution is iteratively solved at each step of the iterative FWI problem. Minimizing ϕ is a problem of the same dimension as the FWI problem, but because it is a linear problem it is generally easier to solve. The cost of solving this problem is also highly dependent on the dimensionality of the problem. Conjugate gradient and BFGS methods are effective approaches for solving a linear problem. In the absence of rounding errors, these methods have guaranteed convergence in a number of iterations less than or equal to the number of variables in the problem. Consequently, these methods can obtain a much more accurate approximation of the Newton step in a given number of iterations is the number of variables is decreased. This will make the TGN method more effective. Because of this problem size dependence, TGN optimization should perform considerably better in a multi-resolution framework.

Gradient and Hessian-vector product calculation

To perform the inner loop optimization problem, it is necessary to calculate the gradient of ϕ (equation 4). This means that g and the product of H with a given vector need to be calculated. These terms are derived in a general way in, for example, Metivier et al. (2013). This derivation is repeated here, with reference to how these terms with respect to m can be related to the traditional terms with respect to a and the matrix P.

We assume here an objective function of the form

$$\theta = \frac{1}{2}||Ru - D||_2^2, \tag{5}$$

where u is the modeled wavefield, D is the measured data, and R is a matrix applying receiver sampling. The minimization problem in FWI then becomes

$$min_m \theta \text{ subject to } S(m)u = f, \tag{6}$$

where S is a Helmholtz matrix enforcing a model of wave propagation, and f is a source term. This objective penalizes the L_2 norm of difference between measured data and modeled data with the current model m. The Lagrangian of this problem is

$$L = ||Ru - D||_2^2 + \langle S(m)u - f, \lambda \rangle, \tag{7}$$

where λ is an as-yet unconstrained Lagrange multiplier, and $\langle , \rangle$ represents an inner product. If $u = \bar{u}$, where $\bar{u}$ satisfies $S(m)\bar{u} = f$, then $L = \theta$. The derivative of θ with respect to m is then equal to the derivative of L with respect to m at $u = \bar{u}$. This derivative is

$$\frac{dL}{dm} = \frac{\partial L}{\partial \bar{u}} \frac{\partial \bar{u}}{\partial m} + \frac{\partial L}{\partial m}. \tag{8}$$

Because the term $\frac{\partial \bar{u}}{\partial m}$ is prohibitively expensive to calculate, we choose to eliminate the contribution of this term by choosing $\lambda = \bar{\lambda}$ such that the term $\frac{\partial L}{\partial \bar{u}}$ becomes zero. This requirement becomes

$$\frac{\partial L}{\partial \bar{u}} = R^T(Ru - D) + S^\dagger \bar{\lambda} = 0. \tag{9}$$

This means that $\bar{\lambda}$ is calculated by back-propagating the data residuals. This choice reduces equation 8 to

$$\frac{dL}{dm} = \frac{\partial L}{\partial m} = \langle \frac{\partial S}{\partial m}\bar{u}, \bar{\lambda}\rangle. \tag{10}$$

Because $\frac{\partial S}{\partial m}$ is typically very sparse, this can be quickly calculated for many variables m, with the main cost being the calculation of $\bar{u}$ and $\bar{\lambda}$. The derivative with respect to the variables m is related to the derivative with respect to traditional variables a through the relation

$$\frac{\partial L}{\partial m} = \frac{\partial L}{\partial a}\frac{\partial a}{\partial m} = P^T \frac{\partial L}{\partial a}. \tag{11}$$

The Hessian-vector product term is calculated in a similar way. Again, we follow the derivation of Metivier et al. (2013). The Gauss-Newton Hessian is given by the relation

$$J^\dagger R^T R J, \tag{12}$$

where J is the Jacobian matrix $\frac{du}{dm}$. The matrix J is too costly to directly calculate, so the adjoint state method is used to prevent direct calculation of this term. We note that the derivative of the function

$$p = \langle u(m), w\rangle, \tag{13}$$

where w is an arbitrary vector, with respect to m is

$$\nabla g = J^\dagger w. \tag{14}$$

Consequently, if w is chosen to be $R^T RJv$, then the Hessian-vector product $H_{GN}v$ is equal to the derivative of p. This derivative can be calculated in exactly the same way as the gradient was by considering the Lagrangian

$$L = \langle u(m), w\rangle + \langle S(m)u - f, \lambda\rangle \tag{15}$$

instead of equation 7. The same procedure follows, but instead of requiring Lagrange multiplier $\bar{\lambda}$ to satisfy equation 9, the removal of $\frac{\partial \bar{u}}{\partial m}$ requires that the Lagrange multiplier for this problem, ξ, satisfies

$$S^\dagger \xi = -w. \tag{16}$$

As before, if the Lagrange multiplier satisfies this condition, then

$$\frac{dL}{dm} = \frac{\partial L}{\partial m} = \langle \frac{\partial S}{\partial m}\bar{u}, \bar{\xi}\rangle. \tag{17}$$

The calculation of ξ still requires that w is known, and we cannot directly calculate J. We calculate the product of J with the vector v through consideration of the derivative of the forward problem with respect to variables m_i multiplied by vector elements v_i. Using the relation $Su - f = 0$, we find

$$\frac{\partial(Su-f)v_i}{\partial m_i} = S(\frac{\partial u}{\partial m_i}v_i) = -u(\frac{\partial S}{\partial m_i}v_i). \tag{18}$$

A sum over i is equal to the product $S(Jv)$ for the left hand side, so

$$S(Jv) = -u\sum(\frac{\partial S}{\partial m_i}v_i) \tag{19}$$

This equation can be solved for the product Jv. Using this term to define w as $R^T RJv$, we can solve equation 16 for ξ. With this choice for ξ, the derivative $\frac{dL}{dm}$ from equation 17 becomes the Gauss-Newton Hessian vector product $H_{GN}v$.

Once again, this calculation can be related back to the equivalent expression for the conventional FWI variables. In terms of the Jacobian matrix with respect to variables a, J_a, the Gauss-Newton Hessian becomes

$$P^T J_a^\dagger R^T RJ_a P \tag{20}$$

in comparison to equation 12. This can be calculated by replacing J with $P^T J_a$. The steps which must be modified are equation 17, which becomes

$$H_{GN} = P^T\frac{dL}{da} = P^T\langle\frac{\partial S}{\partial a}\bar{u}, \bar{\xi}\rangle, \tag{21}$$

and equation 19, which becomes

$$S(Jv) = -u\sum\left(P_i^T\frac{\partial S}{\partial a}\right)v_i, \tag{22}$$

where P_i is the ith column of P.

NUMERICAL EXAMPLES

To investigate the effectiveness of this approach, we perform numerical tests of a viscoacoustic inversion, where P-wave velocity c and quality factor Q are the only variables considered in the inversion. This problem is prone to cross-talk (e.g. Keating and Innanen, 2016), which the second derivative information considered in the TGN method should be able to help prevent. The viscoacoustic wave propagation we consider in this report is given by

$$\left[\omega^2 s(\mathrm{r}) + \nabla^2\right]u(\mathrm{r},\omega) = f(\mathrm{r},\omega), \tag{23}$$

where the model parameter s is given by

$$s(\mathrm{r},\omega) = \frac{1}{c^2(\mathrm{r})}\left\{1 + \frac{1}{Q(\mathrm{r})}\left[i - \frac{2}{\pi}\log\left(\frac{\omega}{\omega_0}\right)\right]\right\}, \tag{24}$$

and ω_0 is a reference frequency (Innanen, 2015).

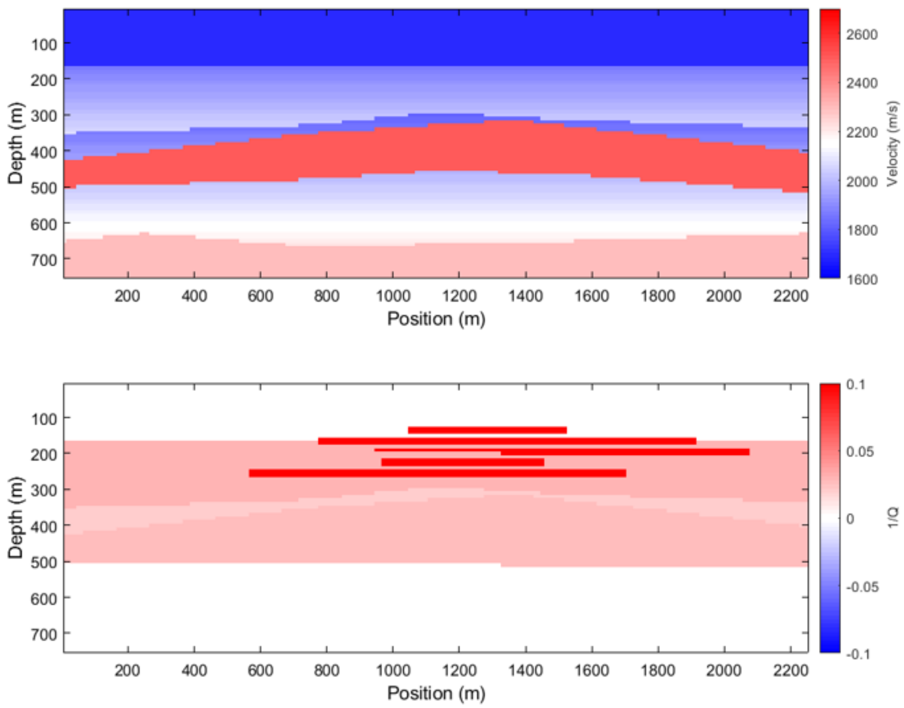

FIG. 1. Velocity (left) and reciprocal Q (right) model used for the numerical examples.

The model to be inverted is shown in figure 1. Finite difference modeling was used with a grid spacing of 10 m. A surface acquisition was simulated, with 225 receivers placed at the surface, and 112 sources placed at 10 m depth. The starting models used were constant values, infinite Q and 1700 m/s velocity. Frequencies from 1 Hz to 25 Hz were used in the inversion. At each FWI iteration, five frequencies were inverted. These frequencies changed every two iterations, and were equally spaced from 1 Hz to a maximum frequency which increased throughout the inversion. This frequency began at 2 Hz, and increased to 25 Hz at the final frequency band. At each FWI iteration, 30 inner loop iterations were performed in the TGN approach.

In the first test, the TGN method was used in the inversion with a fixed grid spacing of 10 m. This is the conventional approach, in which the variables a are inverted for. The result of this inversion is shown in figure 2. Cross-talk is evident in this result, especially away from the center of the Q model between 300 m and 500 m depth, where the $\frac{1}{Q}$ model has decreased to negative values instead of increasing. The velocity model is poorly recovered below 500 m depth, likely a consequence of inadequate Q recovery. This cross-talk occurs despite the use of TGN optimization, implying that insufficient second derivative information is being considered in the inversion.

In the second test, the same inversion is attempted with a multi-resolution strategy. The results of this approach are shown in figure 3. While the negative Q artifacts of the previous example are not completely eliminated, they are substantially reduced. Both the Q and velocity models recovered are closer to the true models in this inversion, despite using the computational cost as the previous inversion. This suggests that the multi-resolution

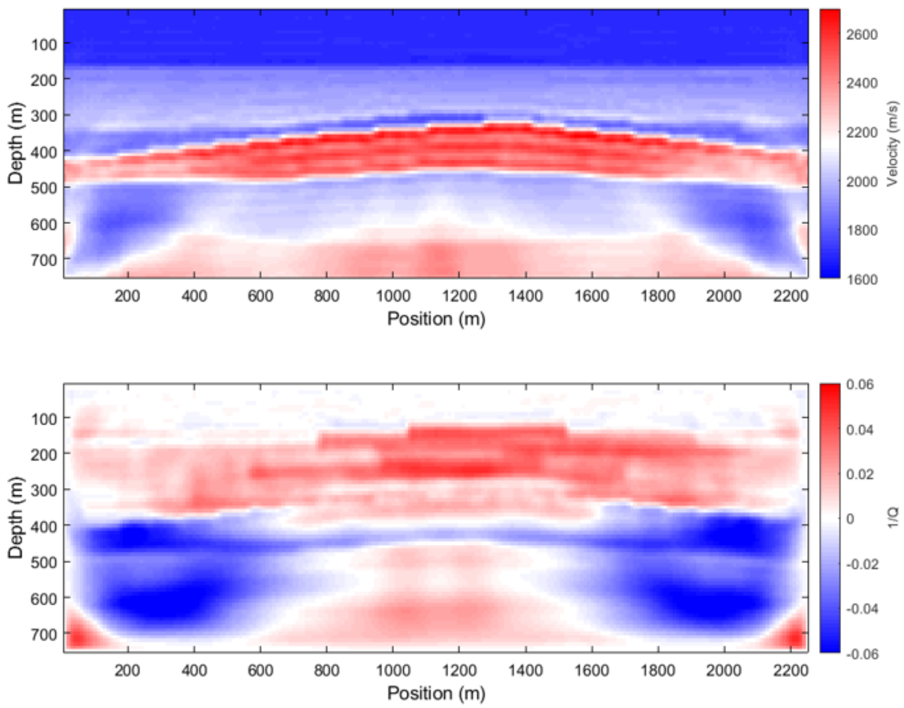

FIG. 2. Inverted velocity (left) and reciprocal Q (right) for a fixed resolution inversion.

strategy has successfully improved the capacity of the inversion to prevent cross-talk.

DISCUSSION

The TGN method has a computational cost that can be easily changed by altering the number of inner iterations performed. This allows it to be applied on problems much larger than those which Newton optimization would be appropriate for. This should allow for more efficient cross-talk reduction in most multi-parameter FWI problems.

CONCLUSIONS

The multiparameter FWI problem is very computationally difficult to solve, and is heavily affected by cross-talk effects. By reducing the dimensionality of the problem in a multi-grid approach, significantly better step directions can be calculated to hellp prevent cross-talk. The truncated Gauss-Newton optimization step can be considerably improved at low frequencies by using a multi-resolution approach. Unlike exact Gauss-Newton optimization, this method is applicable for very large problem sizes, and should result in improved model updates even at relatively high frequencies.

ACKNOWLEDGEMENTS

The authors thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 461179-13. Scott Keating was also supported by the Earl D. and Reba C. Griffin Memorial Scholarship.

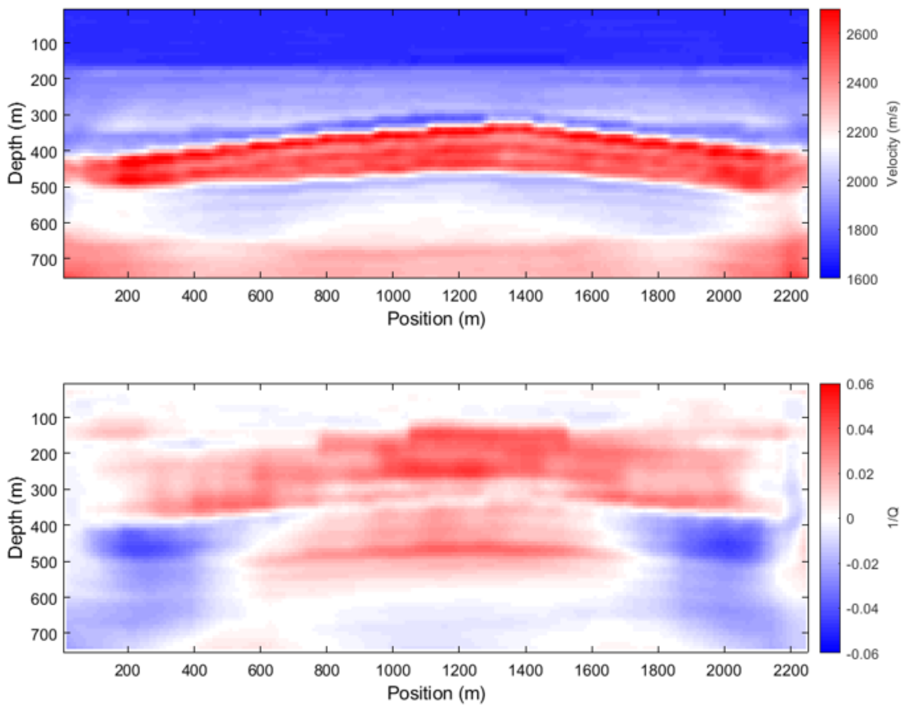

FIG. 3. Inverted velocity (left) and reciprocal Q (right) for a multi-resolution inversion.

REFERENCES

Bunks, C., Salek, F., Zaleski, S., and Chavent, G., 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Debens, H., Warner, M., Umpleby, A., and da Silva, N., 2015, Global anisotropic 3d fwi: SEG Expanded Abstracts.

Innanen, K., 2015, Absorption in fwi – some questions and answers: CREWES Annual Report, **27**.

Keating, S., and Innanen, K., 2016, Challenges in anacoustic full waveform inversion: CREWES Annual Report, **28**.

Keating, S., and Innanen, K., 2017, Multi-resolution newton optimization in full waveform inversion: CREWES Annual Report, **29**.

Metivier, L., Brossier, R., Virieux, J., and Operto, S., 2013, Full waveform inversion and the truncated newton method: Siam J. Sci. Comput., **35**, B401–B437.