

# **Multi-resolution Newton optimization in full waveform inversion**

Scott Keating and Kris Innanen

## **ABSTRACT**

Full waveform inversion typically defines many more variables in the inversion than can be expected to be independently recovered. This prevents usage of more powerful optimization techniques, where large numbers of variables become very expensive. Here, FWI is performed with a model resolution defined based on the data frequencies considered at each iteration. This allows for Newton optimization to be employed in the recovery of the low frequency part of the model, considerably reducing cross-talk on long wavelength scales.

## **INTRODUCTION**

Full waveform inversion (FWI) seeks to reconstruct a complete subsurface model which accurately reproduces measured seismic data (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009). This is typically achieved through a numerical optimization procedure. Great advances have been made in, for instance, constructing high resolution velocity models through use of FWI (Virieux and Operto, 2009). Significant obstacles, however, must be confronted to move forward with FWI. Extreme computational cost in particular remains a significant barrier. Closely related is the significant complexity introduced by considering multiple physical parameters in the inversion, where cross-talk, confusion of different parameters in the inversion, is a serious concern (e.g., Operto et al., 2013; Pan et al., 2016). These issues have led to FWI being largely confined to relatively low frequency, velocity-only model recovery.

Many authors have experimented with different ways of improving the efficiency of the FWI algorithm. One approach that has been investigated is to use powerful global optimization techniques to recover a coarse approximation of the subsurface, followed by traditional FWI (Datta and Sen, 2016). This allows for a relatively accurate model at large scales, helping to improve the results of the inversion that builds on this model. While this approach is promising, the extreme cost of performing a global optimization with a large number of parameters severely limits the scales at which it can be employed. Global optimization is far from unique in being too costly to implement for a full resolution FWI. Many techniques intermediate between steepest descent optimization and global optimization are too computationally intensive for the very large number of variables considered. Notably, Newton optimization, which has been shown to be very effective in eliminating cross-talk in multi-parameter FWI, falls into this category. If Newton optimization can be applied at an affordable scale, it may provide useful large-scale information that helps to prevent cross-talk in the FWI result.

While FWI typically considers a fixed number of model variables, this may not be efficient or necessary. In multi-scale FWI early iterations are performed using only the low-frequency content of the data, while later iterations progressively introduce higher frequencies (Bunks et al., 1995). This naturally suggests that the achievable resolution of the recovered model will change as the inversion progresses. In this report we investigate

the idea of applying Newton optimization in FWI with a multi-resolution approach, wherein the inverted model is re-parameterized as the inversion progresses to reflect the frequency content of the data being considered.

## THEORY

Multi-parameter FWI, where several physical properties are recovered in the inversion, is often prone to cross-talk, where data residuals introduced by one parameter mistakenly influences the estimate of another. This problem is not sufficiently treated when employing very simple numerical optimization techniques, such as the steepest descent method. On the other hand, it is well understood to be powerfully mitigated by considering second derivative information in the optimization procedure. A Newton optimization approach, where this second order information is fully employed, has the capacity to greatly reduce cross-talk, an important consideration in multiparameter FWI.

Newton optimization conditions the gradient (first derivatives) with the inverse of the Hessian matrix (second derivatives). This allows for consideration of changes in a derivative with respect to one variable as another variable is changed. Crucially, this prevents one variable from being modified as a result of a data residual that can be entirely explained by a change in another variable. So, for instance, greater resolution can be obtained in a Newton update, as nearby variables are not confused for one another. Similarly, cross-talk between physically distinct properties can be reduced.

Exact Gauss-Newton optimization is typically unachievable in FWI due to the extremely large associated computational costs and memory requirements. The memory requirements consist chiefly of storing the Hessian matrix  $H$ , which for a model of  $N$  elements contains  $N^2$  entries. The computational cost is driven by the solution for the Newton update:

$$\Delta p = -H^{-1}g, \quad (1)$$

where  $\Delta p$  is the descent direction and  $g$  is the gradient. The cost of solving this system is on the order of approximately  $N^3$  operations.

Much of the information present in a Newton update may be useful, but perhaps not crucial to the inversion procedure. For instance the part of the Hessian which allows for better resolution forms a considerable fraction of the size and cost, but does not form the key motivation for using the Hessian. If cross-talk reduction is a major objective, Newton optimization with a coarse set of parameters may be desirable, as it mitigates cross-talk at a large scale, but reduces the number of parameters considered, and so the cost.

Achievable seismic resolution is approximately one quarter to one eighth of the seismic wavelength. Second order finite difference schemes can require ten or more grid points per wavelength to avoid grid dispersion. This means that an inverted model defined on the same grid as the wave simulation may have the capacity to define features much smaller than the achievable limits imposed by the data. Such a model defines many more variables than can actually be accurately recovered independently. The multiscale approach frequently employed in FWI exaggerates the issue even further. In this approach, low frequencies are inverted at early iterations, and successively higher frequencies are gradually introduced as

the inversion proceeds. As the same finite difference grid is usually used at all iterations, this means that an inverted model defined on this grid may have orders of magnitude more variables than can be accurately recovered. As it is the large number of variables that makes powerful optimization techniques prohibitively costly, it is reasonable to consider whether there are advantages to more conservatively parameterized models.

In a multiscale inversion, the achievable level of resolution changes as the inversion progresses. We investigate here the idea of changing in tandem the grid on which the inverted model is defined, in the hopes of achieving efficiency improvements. We refer to this approach as 'multi-resolution FWI'.

While the recovered model may be considerably over-defined in FWI, this does not necessarily imply inefficiency. Some numerical optimization strategies are insensitive to the resolution of the inverted model. Steepest descent optimization, for instance, has a cost dominated by the evaluation of the gradient, the cost of which is closely tied to wave propagation cost. More coarsely defined models offer minimal savings in steepest descent optimization unless the wave propagation cost is also reduced. Other optimization approaches have a more ambiguous relation to the model resolution. The conjugate gradient and BFGS techniques share the property that an  $N$ -dimensional linear problem can be solved in at most  $N$  steps. As the dimensionality drops, this upper cost limit falls as well, supporting the idea that a more coarsely resolved model may be easier to recover. In FWI, however, the problem is nonlinear, and the number of iterations performed is typically much less than the dimensionality of the problem. This makes it difficult to predict the change in the performance of these techniques with changing model resolution. Newton optimization and global optimization have costs closely tied to the model resolution. These are the methods which most benefit by a reduction in the dimensionality of the optimization problem.

## Derivatives in the FWI problem

### *Gradient*

To calculate the Hessian and gradient with respect to coarse model parameters, we employ the adjoint state method. To begin with, we consider the FWI problem as a constrained optimization problem:

$$\min_u \Phi = \frac{1}{2} \|Ru - d\|^2, \quad (2)$$

subject to the constraint that the Helmholtz equation be satisfied, that is

$$S(p)u = f, \quad (3)$$

where  $u$  is the pressure field,  $R$  is a sampling operator representing receiver locations,  $d$  represents the measured data,  $S$  is the Helmholtz operator, and  $f$  is a source term. In the adjoint state method, the Lagrangian of this system is considered:

$$\mathcal{L}(u, p, \lambda) = \frac{1}{2} \|Ru - d\|^2 + \Re(S(p)u - f, \lambda), \quad (4)$$

where  $\lambda$  is a Lagrange multiplier, and  $\Re$  denotes the real part. If we choose a  $\tilde{u}$  satisfying the wave equation, that is

$$S(p)\tilde{u} - f = 0, \quad (5)$$

the Lagrangian reduces to the objective function:

$$\mathcal{L}(\tilde{u}, p, \lambda) = \frac{1}{2} \|Ru - d\|^2 = \Phi. \quad (6)$$

Then, the derivative of the objective function can be given as

$$\frac{d\Phi}{dp} = \frac{d\mathcal{L}(\tilde{u}, p, \lambda)}{dp} = \frac{\partial}{\partial p}(\mathcal{L}(\tilde{u}, p, \lambda)) + \frac{\partial}{\partial \tilde{u}}(\mathcal{L}(\tilde{u}, p, \lambda)) * \frac{d\tilde{u}}{dp}. \quad (7)$$

This expression has an explicit dependence on the derivative  $\frac{d\tilde{u}}{dp}$  which can be eliminated by choosing a Lagrange multiplier  $\tilde{\lambda}$  such that  $\frac{\partial}{\partial \tilde{u}}(\mathcal{L}(\tilde{u}, p, \tilde{\lambda})) = 0$ . This means the Lagrange multiplier satisfies

$$S(p)^\dagger \tilde{\lambda} = R^T(d - R\tilde{u}). \quad (8)$$

After taking this step, the expression for the derivative reduces from equation 7 to

$$\frac{d\Phi}{dp} = \frac{d\mathcal{L}(\tilde{u}, p, \tilde{\lambda})}{dp} = \Re(\partial_p S(p)\tilde{u}, \tilde{\lambda}). \quad (9)$$

The gradient vector can be constructed by calculating this term for each model variable (Metivier et al., 2013).

### Hessian

Calculation of the second derivatives follows a similar route. We now wish to take the derivative of equation 9, subject to the constraints in equations 3 and 8. The associated Lagrangian is given by

$$\mathcal{L}(u, \lambda, \gamma, \delta, p) = \Re(\partial_p S(p)u, \lambda) + \Re(S(p)u - f, \gamma) + \Re(S(p)^\dagger \lambda - R^T(d - Ru), \delta). \quad (10)$$

Once again, choosing  $\tilde{u}$  and  $\tilde{\lambda}$  to satisfy the constraints, only the first term remains, and

$$\mathcal{L}(\tilde{u}, \tilde{\lambda}, \gamma, \delta, p) = \Re(\partial_p S(p)\tilde{u}, \tilde{\lambda}) = \frac{d\Phi}{dp}. \quad (11)$$

The second derivative is then equal to

$$\frac{d^2\Phi}{dpdp'} = \frac{d\mathcal{L}(\tilde{u}, \tilde{\lambda}, \gamma, \delta, p)}{dp'} = \frac{\partial \mathcal{L}}{\partial p'} + \frac{\partial \mathcal{L}}{\partial \tilde{u}} \frac{d\tilde{u}}{dp'} + \frac{\partial \mathcal{L}}{\partial \tilde{\lambda}} \frac{d\tilde{\lambda}}{dp'}. \quad (12)$$

The terms with explicit derivatives can again be eliminated by choosing Lagrange multipliers  $\tilde{\delta}$  and  $\tilde{\gamma}$  satisfying

$$S\tilde{\delta} = -\partial_p S(p)\tilde{u} \quad (13)$$

and

$$S^\dagger \tilde{\gamma} = -(\partial_p S(p)^\dagger \tilde{\lambda} + R^T R\tilde{\delta}) \quad (14)$$

This reduces equation 12 to

$$\frac{d^2\Phi}{dpdp'} = \frac{d\mathcal{L}(\tilde{u}, \tilde{\lambda}, \tilde{\gamma}, \tilde{\delta}, p)}{dp'} = \Re \left\{ (\partial_p \partial_{p'} S \tilde{u}, \tilde{\lambda}) + (\partial_{p'} S \tilde{u}, \tilde{\gamma}) + (\partial_{p'} S^\dagger \tilde{\lambda}, \tilde{\delta}) \right\}. \quad (15)$$

When using the Gauss-Newton approximation, we take the limit as the residual  $d - Ru$  approaches zero, neglecting terms dependent on  $\tilde{\lambda}$ .

### A multiresolution inversion

FWI is an iterative process, wherein model updates, defined on the model grid, and wavefield simulations, defined on a finite-difference grid, are successively repeated. For a model of dimension  $M$ , and a finite difference grid with  $N$  points, we define the  $N$  by  $M$  projection matrix  $P$  such that

$$m_{FD} = Pm \quad (16)$$

and

$$m = P^T m_{FD}, \quad (17)$$

where  $m$  is the model considered in FWI, and  $m_{FD}$  is the corresponding finite difference model. In a multiresolution approach,  $M$  will be different for each range of frequencies considered. The value of  $M$  should be determined by the expected achievable resolution at a given frequency band. Here we choose  $M$ , such that the model is defined on a scale of  $\frac{\lambda}{8}$ , where  $\lambda$  is the wavelength of the highest frequency considered.

The derivatives of the operator  $S$  in the previous section with respect to the model parameters can be easily calculated, given  $P$  and the derivatives with respect to the finite difference points. Let us denote the elements of the finite difference model  $m_i$ , and the variables used in the inversion as  $M_i$ . The relationship between these values is given by

$$M_i = \sum_{j=1}^N P_{j,i} m_j. \quad (18)$$

The derivative of the operator  $S$  is then

$$\frac{\partial S}{\partial M_i} = \sum_{j=1}^N \frac{\partial S}{\partial m_j} \frac{\partial m_j}{\partial M_i} = \sum_{j=1}^N \frac{\partial S}{\partial m_j} \frac{1}{P_{j,i}}. \quad (19)$$

By considering FWI as a multiresolution process, we open the possibility for significant efficiency improvements with a number of different optimization approaches. This report focuses on using multiresolution FWI with exact Gauss-Newton optimization, but efficiency improvements should be attainable using any optimization approach, including truncated Newton, l-BFGS, or conjugate gradients.

### Computational cost and memory considerations for multiresolution Newton FWI

In Newton optimization, there are a number of steps which are demanding from a numerical implementation standpoint. First, the calculation of the Hessian matrix, using

equation 15, requires that equations 13 and 14 be solved once for each variable. This necessitates factoring  $S$  and  $S^\dagger$  only once, as these terms do not change when considering derivatives with respect to different variables  $p$ . For  $N$  variables in both the inversion and the finite difference model, this process requires  $\mathcal{O}(N^3)$  operations, but as  $S$  is typically very sparse, this cost does not usually dominate. Solving equations 13 and 14 once the left hand side matrices are factorized requires a back substitution, necessitating  $\mathcal{O}(N^2)$  operations. As this must be done for  $N$  right hand sides, the cost of this step is  $\mathcal{O}(N^3)$  operations, and is typically much more than the factorization for sparse  $S$ .

The storage of the Hessian matrix can also be very demanding. For a model of  $N$  variables, the Hessian contains  $N^2$  elements. Lastly, the Newton update must be solved for, using equation 1. Solving this system again requires  $\mathcal{O}(N^3)$  calculations. Overall, then, Newton optimization for  $N$  variables in both the inversion and the finite difference model requires  $\mathcal{O}(N^3)$  calculations, and the storage of  $\mathcal{O}(N^2)$  elements.

In a multiresolution approach, the number of finite difference elements,  $N$ , determining the size of  $S$ , may be much larger than the number of variables considered in the optimization,  $M$ . This modifies several costs. The number of operations and storage required for the solution of equation 1 reduce to  $\mathcal{O}(M^3)$  and  $\mathcal{O}(M^2)$  respectively. Obtaining the Hessian matrix reduces to  $\mathcal{O}(N^2M)$  operations, and the cost of factorizing  $S$  remains unchanged. Provided that  $S$  is sufficiently sparse that this cost does not dominate, this leads to a total requirement of  $\mathcal{O}(N^2M)$  operations and storage of  $\mathcal{O}(M^2)$  elements in the multiresolution Newton approach.

## NUMERICAL EXAMPLES

To demonstrate the multiresolution FWI approach, a numerical example is provided in this section. A major benefit of the multiresolution approach is expected to be an increased ability to cope with cross-talk, so a multi-parameter FWI problem is considered. Specifically, the an-acoustic FWI problem is considered, which treats P-wave velocity and quality factor  $Q$ . This problem is very prone to cross-talk, allowing for simple identification of improvements. The an-acoustic wave propagation we consider in this report is given by

$$[\omega^2 s(\mathbf{r}) + \nabla^2] u(\mathbf{r}, \omega) = f(\mathbf{r}, \omega), \quad (20)$$

where the model parameter  $s$  is given by

$$s(\mathbf{r}, \omega) = \frac{1}{c^2(\mathbf{r})} \left\{ 1 + \frac{1}{Q(\mathbf{r})} \left[ i - \frac{2}{\pi} \log \left( \frac{\omega}{\omega_0} \right) \right] \right\}, \quad (21)$$

$c$  is the acoustic wave velocity,  $Q$  is the quality factor,  $\omega_0$  is a reference frequency,  $u$  is the pressure field, and  $f$  is a source term (Innanen, 2015).

### Model

The model considered in this example is two dimensional, 500m by 500m in size and defined on a grid with a spacing of 3.33m in both x and z directions. A simple geometrical model is considered, with uniform background velocity and  $Q$ , and only a few anomalous regions. This model is shown in figure 1. In the inversion the initial model was a uniform velocity and  $Q$ , equal to the background values of the true model.

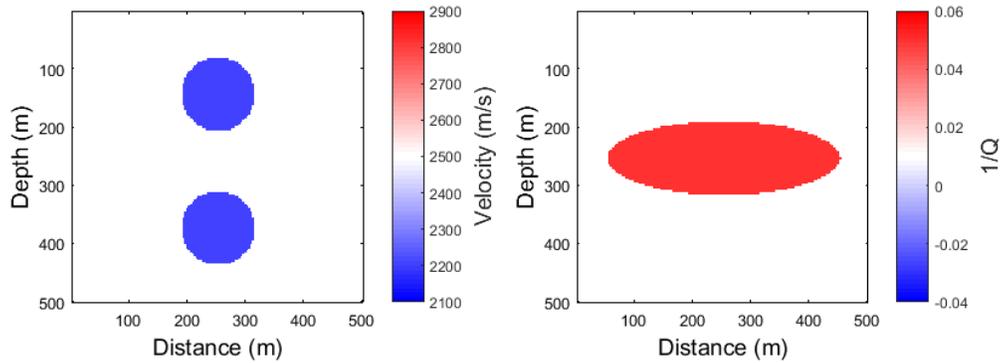


FIG. 1. Velocity (left) and reciprocal Q (right) model used for the numerical examples.

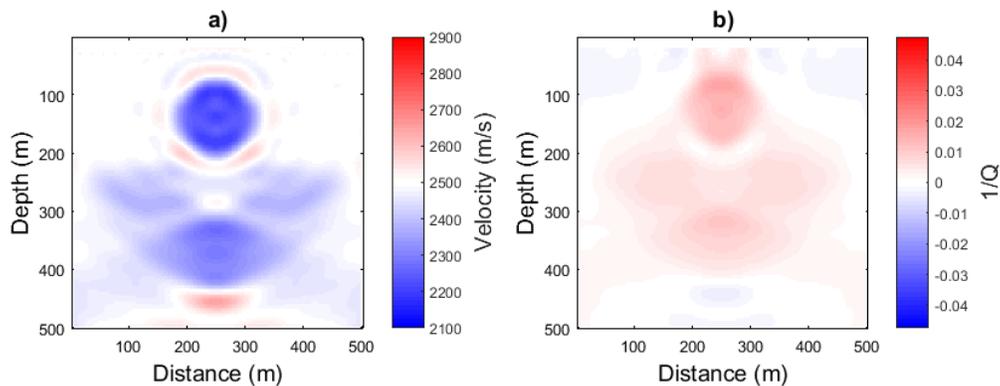


FIG. 2. Inverted velocity (left) and reciprocal Q (right) for a fixed-resolution steepest optimization, with a maximum frequency of 25Hz. Compare with figure 1.

### Fixed-resolution FWI

Traditional FWI does not change the number of variables considered in the inversion based on the frequencies used at each iteration. In this framework, it can be exceedingly difficult to reduce cross-talk in FWI. An example of FWI with steepest-descent optimization is shown in figure 2. Here, 575 iterations were performed, and frequencies from 1-25Hz were considered. The inversion result is severely polluted by cross-talk, which is to be expected when using a method that considers no second derivative information. Newton optimization is extremely expensive in the fixed-resolution approach due to the very large number of variables considered. Approaches which consider approximations to the Newton update are more practical, but can still struggle to significantly reduce cross-talk. One such approach is FWI with truncated Newton optimization, shown in figure 3. This method yields a large improvement over the steepest descent approach, but still leaves significant cross-talk.

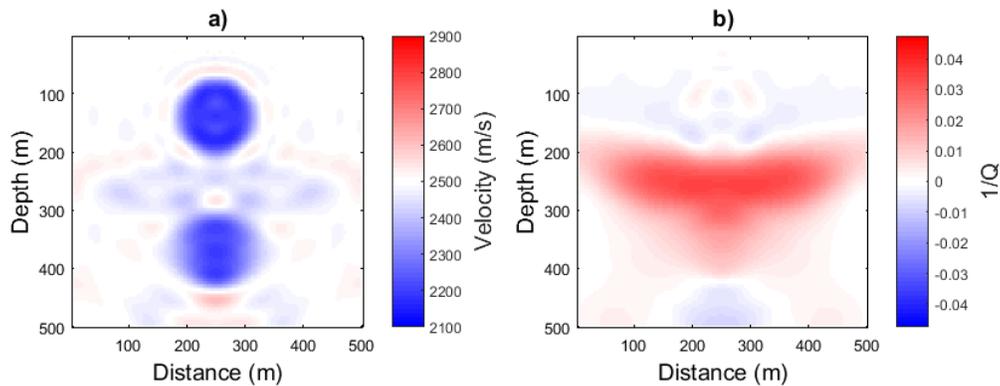


FIG. 3. Inverted velocity (left) and reciprocal Q (right) for a fixed-resolution truncated-Newton optimization, with a maximum frequency of 25Hz. Compare with figures 1, 2.

### Multi-resolution

In the multi-resolution approach implemented in these examples, the grid on which the model update was calculated at each iteration was determined based on the maximum frequency of the data considered at that iteration. More specifically, the model update was calculated on a grid with a spacing equal to  $\frac{\lambda}{8}$ , where  $\lambda$  is the wavelength of a given frequency in the background medium. The value of  $\frac{\lambda}{8}$  was determined by trial and error, larger spatial grids introduced significant errors, likely due to an inability of the model to describe high frequency data. A maximum grid spacing of 100m was also enforced, the computational cost savings of coarser models diminishing rapidly. Higher frequencies were gradually introduced in the inversion in a multi-scale approach (Bunks et al., 1995). In this inversion, five frequencies evenly spaced between a minimum and maximum frequency were inverted at each iteration. At the first iteration, only the 1Hz data was inverted, while at every subsequent iteration the maximum frequency was increased by 1Hz, to a maximum frequency, chosen to be the highest frequency for which this procedure was affordable for this model. The minimum frequency remained fixed at 1Hz.

In these examples, the projection matrix was chosen such that each variable in the coarse parameterization was a square block of the fine parameterization, the size of which was determined by the  $\frac{\lambda}{8}$  resolution criterion. Lengths of non-integer grid points in the fine model were weighted in the projection matrix according to the fraction of the fine element which lay inside the coarse parameter .

While the coarse model descent direction can be projected onto the fine model and applied directly, this was not found to be an optimal approach. Instead, a smoother was applied to the descent direction prior to the line-search and update. In the smoothing process, each point was replaced with the mean of the  $2\frac{N}{M}$  by  $2\frac{N}{M}$  block centered at that point. Examples of a descent direction before and after smoothing are given in figures 4 and 5, respectively.

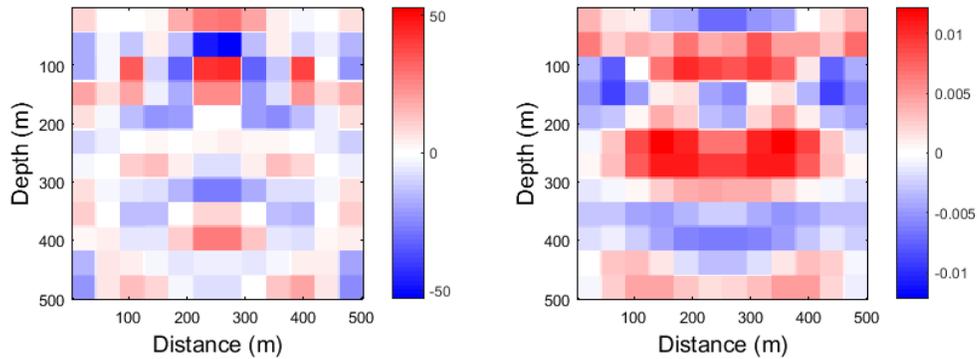


FIG. 4. Calculated coarse update for squared slowness (left) and reciprocal Q (right), projected to full resolution coordinates. Updates are unscaled. This represents a model update for a frequency band of 1-7Hz.

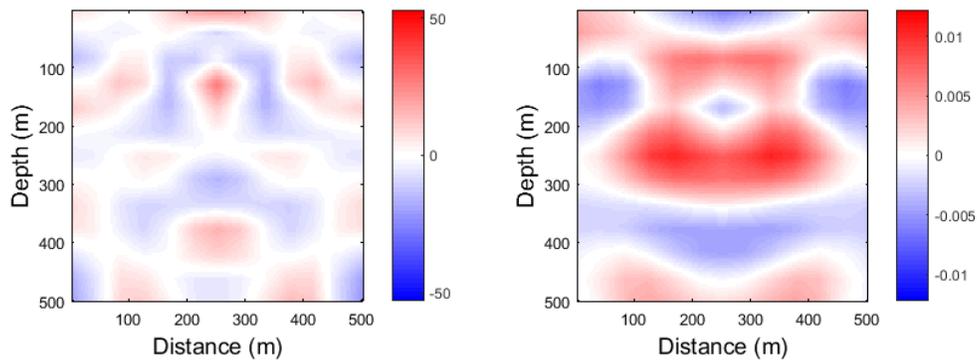


FIG. 5. Smoothed coarse update for squared slowness (left) and reciprocal Q (right). Updates are unscaled. This represents a model update for a frequency band of 1-7Hz. Compare to figure 4.

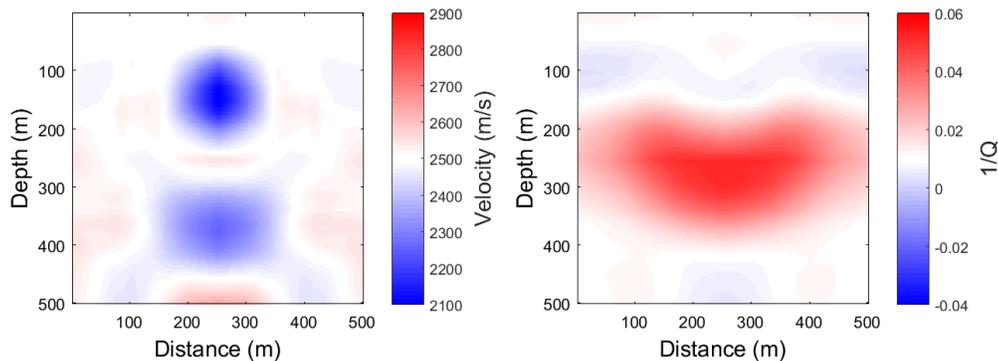


FIG. 6. Inverted velocity (left) and reciprocal  $Q$  (right) for a multi-resolution exact Newton optimization, with a maximum frequency of 7Hz. Compare with figure 1.

### Multi-resolution exact Newton QFWI

The application of an exact Newton update in a multi-resolution FWI is restricted to relatively low frequencies, where there are relatively few variables being inverted for. The true model for these examples (figure 1) consists of  $4.5 \times 10^4$  elements. We consider the performance of the multi-resolution approach for two different cases here, one in which the maximum frequency considered in the exact Newton optimization is 7 Hz and considers only 288 variables in the final step of the inversion, and another in which the maximum frequency is 15Hz and 1152 variables are considered in the final step. These cases represent a conservative and a more computationally demanding implementation of the approach.

The result of the multi-resolution inversion with a maximum frequency of 7Hz is shown in figure 6. As expected, the recovered model has relatively poor resolution, but provides a largely cross-talk free result: the locations of velocity anomalies and  $Q$  anomalies are not confused with one another. This is in sharp contrast to the behaviour of steepest descent, or even truncated Newton optimization.

The result of the multi-resolution inversion with a maximum frequency of 15Hz is shown in figure 7. In this example, the result is very similar to the 7Hz result in figure 6. Like the lower frequency result, cross-talk is largely eliminated here, but higher resolution is achieved.

The large spatial scale of the anomalies allows for relatively coarse model updates to effectively combat cross-talk in this example. Smaller scale anomalies will, of course, require finer scale inversion to effectively remove cross-talk.

## DISCUSSION

While the approach outlined here allows for a significant reduction in the computational cost of a Newton optimization approach, this reduction is not as large as might be expected. The previously limiting  $\mathcal{O}(N^3)$  computational cost and  $\mathcal{O}(N^2)$  memory requirement for solving the full resolution Newton update are reduced to  $\mathcal{O}(M^3)$  and  $\mathcal{O}(M^2)$ , respectively,

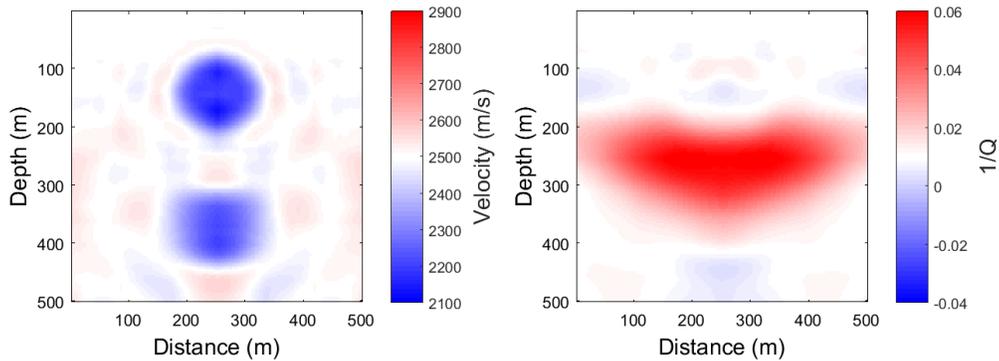


FIG. 7. Inverted velocity (left) and reciprocal  $Q$  (right) for a multi-resolution exact Newton optimization, with a maximum frequency of 7Hz. Compare with figures 1, 6.

but the cost of the algorithm detailed here is dominated instead by the calculation of the elements of the Hessian with a cost of  $\mathcal{O}(N^2M)$ . One alternate approach would be to calculate the Gauss-Newton Hessian through the direct use of Jacobian matrices, using the expression

$$H_{GN} = J^T J, \quad (22)$$

where  $J$  is the  $N$  by  $M$  Jacobian matrix. This approach is less computationally intensive, requiring only  $\mathcal{O}(NM^2)$  calculations, but requires the storage of the Jacobian matrices, with a memory requirement of  $\mathcal{O}(NM)$ . This memory requirement may be unachievable for desirable values of  $M$ . In both these approaches, the size of the finite difference grid  $N$  plays a major role in determining the cost. This suggests yet another possible approach to implementing a multiresolution FWI, in which the finite difference grid also changes as the inversion proceeds. If the inverted model and finite difference grid are both defined on the same  $M$  grid points, this trivially reduces the cost of Newton FWI to  $\mathcal{O}(M^3)$ . This could be helpful for increasing the accuracy of early, low frequency iterations. Complications may arise in pursuing this approach, however, if source and receiver spacing fall below the effective model resolution.

This report has focused on the implementation of Newton optimization on a more coarsely defined FWI problem. While the computational costs and memory requirements of this approach are significantly less than those of a fully resolved Newton optimization, they likely remain sufficiently high that this approach is restricted to considering only the very low resolution component of models. The idea of using Newton optimization for a reduced number of model variables, however, is not restricted to a low resolution approach. This approach could also be used to improve a targeted subsection of a high resolution model.

## CONCLUSIONS

Multiresolution FWI offers the potential for more powerful optimization techniques to be brought to bear on early stages of the FWI procedure. Gauss-Newton optimization in a multiresolution context, offers the potential to reduce large scale cross-talk in multi-

parameter FWI. Using the adjoint state method to obtain the Hessian, this approach cost  $\mathcal{O}(N^2M)$  calculations, and had a memory requirement equal to the greater of  $N$  and  $M^2$ . More efficient implementations based on the same idea may be achievable by applying the multiresolution concept to the finite difference grid as well.

### 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.

### REFERENCES

- Bunks, C., Salek, F., Zaleski, S., and Chavent, G., 1995, Multiscale seismic waveform inversion: *Geophysics*, **60**, 1457–1473.
- Datta, D., and Sen, M. K., 2016, Estimating a starting model for full-waveform inversion using a global optimization method: *Geophysics*, **81**, R211–R223.
- Innanen, K., 2015, Absorption in fwi – some questions and answers: *CREWES Annual Report*, **27**.
- Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: *Conference on Inverse Scattering, Theory and Application, Society for Industrial and Applied Mathematics, Expanded Abstracts*, 206–220.
- 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.
- Operto, S., Gholami, Y., Prieux, V., Ribodetti, A., Brossier, R., Metivier, L., and Virieux, J., 2013, A guided tour of multiparameter full-waveform inversion with multicomponent data: from theory to practice: *The Leading Edge*, **Sep**, 1040–1054.
- Pan, W., Innanen, K. A., Margrave, G. F., Fehler, M., Fang, X., and Li, J., 2016, Estimation of elastic constants for hti media using gauss-newton and full newton multi-parameter full waveform inversion: *Geophysics*, **81**, No. 5, E323–E339.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266.
- Virieux, J., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, No. 6, WCC1.