# Cross-talk and frequency bands in truncated Newton an-acoustic full waveform inversion

*Scott Keating and Kristopher A. Innanen, Department of Geoscience, University of Calgary*

## SUMMARY

Simultaneous use of data within relatively broad frequency bands is essential to discriminating between velocity and $Q$ errors in the construction of an-acoustic full waveform inversion (QFWI) updates. Individual frequencies or narrow bands in isolation cannot provide sufficient information to resolve cross-talk issues in a surface seismic acquisition geometry. Truncated Newton (TN) optimization methods offer the potential for reducing computational cost while incorporating approximate versions of the Newton update to reduce these cross-talk issues, with the trade-off being mediated by the chosen number of inner TN iterations. In fact, in TN-QFWI we are able to choose between two qualitatively distinct "modes" of an-acoustic inversion: one in which the estimation of a velocity model uncorrupted by the influence of $Q$ is the desired outcome, and another in which both a velocity model and a $Q$ model are the desired outcomes. Both can in principle be accomplished in the context of TN-QFWI, with the former at significantly reduced computational expense.

## INTRODUCTION

Full waveform inversion (FWI) is a technique which attempts to recover the true subsurface parameters by iteratively minimizing the difference between measured data and modeled data generated from the current estimated subsurface parameters (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009). While multiparameter versions of FWI have been formulated and studied, the majority of research on FWI is focused on a single parameter problem, specifically that in which acoustic wave propagation is assumed and density is treated as constant. In this problem, only P-wave velocity varies in the model.

However, in the effort to make FWI effective in the determination of larger numbers of smaller scale (e.g., reservoir) properties, the multiparameter FWI problem must be brought to bear. In multiparameter FWI (e.g., Operto et al., 2013; Plessix et al., 2013; Pan et al., 2016), allowance is made in the gradient/Hessian quantities for simultaneous and independent variations of several parameters, either to support velocity model building, or to push towards elastic characterization of the subsurface (Tarantola, 1986; Choi et al., 2008).

Attenuation and dispersion play important roles in both of these applications of multiparameter FWI. It can be a powerful nuisance to acoustic and elastic FWI, strongly influencing the amplitude and phase of the waveforms we would like to interrogate for acoustic/elastic information, but it can also be a rich source of information by which fluids and viscosities can be determined or discriminated. So, we can also distinguish between whether we wish to specifically determine $Q$ in FWI, or merely "protect" the recovery of other parameters from its influence. Either motivation requires that the physics of attenuation be included in an FWI scheme. An-acoustic FWI (QFWI for short), in which attenuation and dispersion parameters are determined simultaneously alongside their elastic counterparts, has been carefully investigated (e.g., Hak and Mulder, 2011; Hicks and Pratt, 2001; Malinowski et al., 2011; Kamei and Pratt, 2013; Métivier et al., 2015). In much of this existing research, however, incorporating attenuation is treated as a small addition to the classical acoustic/elastic FWI problem, with relatively little focus on how the nature of the problem changes. Parameter cross-talk, in which one parameter is mistakenly updated to account for data residuals caused by another, affects an-acoustic FWI significantly and in a unique manner requiring special study.

Simultaneous variations in acoustic and/or elastic properties can be separately estimated in FWI primarily because of differences in the angle-dependence of scattering from one parameter to another. With P-wave velocity and $Q$ it is, in contrast, the differences in the frequency-dependence of scattering amplitudes which permits them to be distinguished (Innanen and Weglein, 2007; Hak and Mulder, 2011). This fact ties together, in an unusually close manner, issues of multiscale FWI, in which iterations or groups of iterations involve different frequency bands, parameter cross-talk, and the degree of approximation with which off-diagonal elements of the inverse Hessian are incorporated through TN iterations. In this paper we analyze this relationship in the context of synthetic an-acoustic frequency domain FWI. Because the exact manner in which dispersion is modelled determines the character of the cross-talk, the attenuation model type, which must be selected prior to formulating a detailed FWI algorithm, plays a key role. This issue is discussed in a companion paper (Keating and Innanen, 2017).

## THEORY

### Cross-talk

The FWI problem considered here has an objective function given by

$$\phi(\mathbf{m}) = \frac{1}{2}||\mathbf{d}_{obs} - \mathbf{d}_{mod}||_2^2, \tag{1}$$

where $\phi$, a function of the subsurface model $\mathbf{m}$, measures the discrepancy between the measured data $\mathbf{d}_{obs}$ and the modelled data $\mathbf{d}_{mod}$. To recover the true properties of the subsurface, this objective is minimized in FWI through gradient-based, or Newton type updates.

Cross-talk is the phenomenon where data residuals introduced by an error in one model parameter are attributed to errors in the estimate of another parameter. For example, cross-talk is present if an estimate of density is modified due to data residuals introduced by errors in a velocity estimate. Cross-talk is a major concern in FWI, as it can severely harm the accuracy of the recovered model and the convergence of the scheme (e.g., Plessix et al., 2013; Innanen, 2014; Pan et al., 2016). Gradient updates are particularly vulnerable to cross-talk. This is due to the fact that the gradient considers only the derivative of the objective function with respect to each variable parameterizing the model. If changes in several different variables can reduce the same part of the data residual, all will be changed in a gradient update.

Newton optimization employs both the first order (gradient) and second order (Hessian) derivatives of the objective function. In Newton optimization, the update $\mathbf{p}$ is given by

$$\mathbf{p} = -\mathbf{H}^{-1}\mathbf{g} \quad , \tag{2}$$

where $\mathbf{g}$ is the gradient of the objective function, and $\mathbf{H}$ is the Hessian matrix. The Hessian provides information about how the derivative with respect to one variable will change as another variable changes. This helps to prevent several variables from being used in reducing the data residual introduced by an error in one, mitigating cross-talk. Unfortunately, in realistic FWI applications, Newton optimization tends not to be a viable approach, because of the excessive cost for the storage and inversion of the Hessian.

### Optimization

Two approaches which attempt to approximate exact Newton optimization but at reduced cost are quasi-Newton methods and truncated Newton methods. Quasi-Newton methods obtain an exact solution to an equation approximating equation 2, whereas truncated Newton

(TN) methods are those which obtain an approximate solution to equation 2. Both attempt to provide an efficient alternative to exact Newton optimization while still retaining important information about the Hessian, which helps to mitigate cross-talk. In this report we focus on the TN method.

TN optimization is similar to exact Newton optimization, but rather than directly solving 2, an approximate solution is obtained by iteratively minimizing (Nocedal and Wright, 2006)

$$\theta(\mathbf{p}) = \frac{1}{2}\mathbf{p}^T\mathbf{H}\mathbf{p} + \mathbf{g}^T\mathbf{p}. \quad (3)$$

At a minimum of this objective function, the gradient of $\theta$ is zero, so

$$\mathbf{H}\mathbf{p} + \mathbf{g} = 0, \quad (4)$$

satisfying equation 2. In this research, the truncated Gauss-Newton method is used, where $\mathbf{H}$ is replaced with $\mathbf{H}_{GN}$, the residual independent part of the Hessian. Following Metivier et al. (2013), FWI updates are iteratively constructed in what will be called the *outer* loop, and the minimization of equation 3, which occurs once for each FWI update, but is itself iterative, involves what will be called the *inner* loop. Provided a suitable optimization approach is employed in the inner loop, this method does not require the storage or inversion of the Hessian matrix $\mathbf{H}_{GN}$, only the product of the Hessian with an arbitrary $\mathbf{p}$. This Hessian-vector product can be efficiently calculated using the adjoint state method, as described in Metivier et al. (2013).

We implement the inner loop of the TN FWI algorithm with a BFGS inner solver, wherein $\mathbf{p}$ is determined by iteratively solving

$$\mathbf{p}_k = \mathbf{p}_{k-1} + \alpha_k\Delta\mathbf{p}, \quad \text{where} \quad \Delta\mathbf{p} = -\mathbf{Q}\nabla\mathbf{p}_{k-1}, \quad (5)$$

$\mathbf{Q}$ is the BFGS approximation of the inverse Hessian of $\theta$ (which is the same as the inverse Hessian of $\phi$), and

$$\alpha_k = -(\partial\theta/\partial\mathbf{p})^T\Delta\mathbf{p} \Big/ \left(\Delta\mathbf{p}^T\mathbf{Q}\Delta\mathbf{p}\right). \quad (6)$$

The computational cost of the truncated Newton method is determined largely by the number of inner iterations used in minimizing (3) in each FWI iteration. This cost is controlled by specifying stopping conditions. The stopping conditions used here are satisfied when a maximum number of iterations are reached, or the condition

$$||\mathbf{H}_{GN}\mathbf{p} + \mathbf{g}|| \leq ||\eta\mathbf{g}|| \quad (7)$$

holds, where $\eta$ is a chosen forcing term. The smaller this forcing term, the greater the cost and lower the cross-talk; the larger the forcing term, the less the computational cost and the greater the cross talk.

**Wave equations**

In order to study in isolation new aspects of cross talk (etc.) in multi-parameter FWI which are introduced by attenuation and dispersion, we consider waves whose propagation is governed by

$$\left[\omega^2 s(\mathrm{r},\omega) + \nabla^2\right] u(\mathrm{r},\omega) = f(\mathrm{r},\omega), \quad (8)$$

where $u$ is the pressure field, $f$ is a source term, and the model parameter $s$ includes a dispersive velocity and an attenuation. No single attenuative-dispersive model is likely always to be entirely correct, so many exist, meaning that several anacoustic models could be considered. This variation, which has its own set of issues for FWI (Keating and Innanen, 2017), is reflected in different specific forms of $s$. Here, the constant $Q$ Kolsky-Futterman attenuation model is considered:

$$s(\mathrm{r},\omega) = \frac{1}{c^2(\mathrm{r})}\left\{1 + \frac{1}{Q(\mathrm{r})}\left[i - \frac{2}{\pi}\log\left(\frac{\omega}{\omega_0}\right)\right]\right\}, \quad (9)$$

where $c$ is the acoustic wave velocity at the reference frequency $\omega_0$, and $Q$ is the quality factor. For a chosen $\omega_0$, the an-acoustic FWI problem is to determine the unknown spatial distributions of two parameters, $c$ and $Q$. Inspection of equation 9 identifies a specific challenge that the QFWI problem faces. The size of the frequency dependent term in $s$, which models dispersion, is determined by $Q$. In effect, both $c$ and $Q$ co-determine the wave velocity at a given frequency. This opens the possibility of considerable cross-talk, and is suggestive that variations from one frequency to another will be instrumental in mitigating it.

**Predicting cross-talk with an-acoustic scattering potentials**

The radiation patterns, or scattering potentials, of point perturbations in active FWI parameters, plotted as functions of experimental variables (e.g., angle between incoming and outgoing rays, frequency, etc.), are often used to determine the degree of expected parameter cross talk in multi-parameter FWI. Parameters which generate potentials with proportional amplitude variations over a given range of these experimental variables are easily confused with one another. The scattering potential $\mathscr{V}$ for position $\mathbf{x}$ and frequency $\omega$ associated with our chosen an-acoustic wave equation is

$$\mathscr{V}(\mathbf{x},\omega) \approx -\frac{\omega^2}{c_0(\mathbf{x})^2}\left[V_Q(\mathbf{x},\omega) + V_c(\mathbf{x},\omega)\right], \quad (10)$$

where

$$V_Q(\mathbf{x},\omega) = \frac{F(\omega)}{Q_0(\mathbf{x})}\Delta Q(\mathbf{x}), \quad V_c(\mathbf{x},\omega) = \left(1 + \frac{F(\omega)}{Q_0(\mathbf{x})}\right)\Delta c(\mathbf{x}) \quad (11)$$

and where $F(\omega) = i - (2/\pi)\log(\omega/\omega_0)$; the $\Delta\cdot$ quantities,

$$\Delta Q(\mathbf{x}) = 1 - \frac{Q_0(\mathbf{x})}{Q(\mathbf{x})}, \quad \Delta c(\mathbf{x}) = 1 - \frac{c_0(\mathbf{x})^2}{c(\mathbf{x})^2}, \quad (12)$$

represent localized jumps in their corresponding model parameters $Q$ and $c$.

Notably, the $Q$ and $c$ components do not vary independently of one another with scattering angle. This means that angle-based considerations in the discrimination of different model parameters, which are crucial in elastic and anisotropic FWI, do not apply here. The components do, however, undergo relative variation as frequency changes; our conclusion is that only through simultaneous inversion of a range of frequencies can a QFWI update distinguish between the influence of $c$ and that of $Q$. This is illustrated in Figure 1. By inspection of this plot, we can furthermore predict that if only a small range of frequencies are considered, over which the two scattering potentials vary roughly in proportion, $c$ and $Q$ will be extremely difficult to distinguish. Over broader ranges of frequencies, however, significant differences between the two scattering potentials become prevalent, which should enable a QFWI iteration to create meaningful updates in both. This introduces a new feature to multi-scale FWI workflows, in which demands already exist on the frequencies considered.

**Gradient for QFWI**

Gradients for $c$ and $Q$, consistent with equations (1), (8), and (9) can be written

$$g_c(\mathbf{r}) = \sum_{\mathbf{r}_g,\mathbf{r}_s,\omega}\omega^2\left[1 + \beta(\omega)s_{q_0}(\mathbf{r})\right]G_0(\mathbf{r}_g,\mathbf{r})G_0(\mathbf{r},\mathbf{r}_s)\ \delta d^* \quad (13)$$

and

$$g_q(\mathrm{r}) = \sum_{\mathrm{r}_g,\mathrm{r}_s,\omega}\omega^2\left[\beta(\omega)s_{c_0}(\mathrm{r})\right]G_0(\mathrm{r}_g,\mathrm{r})G_0(\mathrm{r},\mathrm{r}_s)\delta d^*, \quad (14)$$

where $\delta d = \delta d(\mathbf{r}_g,\mathbf{r}_s)$ are the residulas, $s_{c_0} = c_0^{-2}$, $s_{q_0} = Q_0^{-1}$, are the current model iterates, $G_0(\mathrm{r},\mathrm{r}')$ is the Green's function describing propagation from $\mathrm{r}'$ to $\mathrm{r}$ in the current medium iterate, and

$$\beta = i - \frac{2}{\pi}\log\left(\omega/\omega_0\right). \quad (15)$$

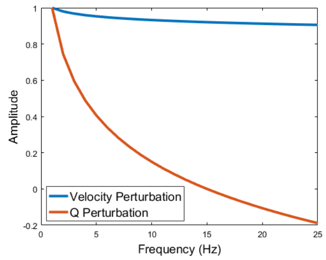


Figure 1: Amplitude of the scattering potential as a function of frequency for velocity perturbation (blue) and $Q$ perturbation (red). Amplitudes have been normalized to 1 at 1Hz. Background $Q_0 = 20$, and the reference frequency was 15Hz for this example.

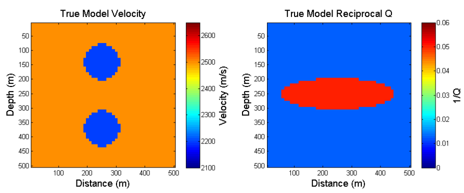


Figure 2: Benchmark model, velocity (left) and $Q$ (right). The velocity values correspond with reference frequency $\omega_0/2\pi = 30$Hz.

## NUMERICAL EXAMPLES

We use the QFWI framework in the previous section to examine the effects of (1) frequency groups used in individual iterations, and (2) optimization method on parameter cross-talk. 2D frequency-domain finite difference modelling is used for simulations; models used are defined on a 2D 50×50 grid with 10m grid cells. 24 sources at 30m depth are spaced 20m apart from 10-470m; 48 receivers at 20m depth are spaced 10m apart from 10-480m. Frequencies from 1Hz to 25Hz are assumed to be available, and the source function is considered to have a uniform amplitude spectrum over this range. First order Engquist boundary conditions are implemented at every boundary. The reference frequency considered was 30 Hz. Figure 2 illustrates the benchmark model used for all of the examples; the initial model consists of homogeneous $c_0$ and $Q_0$ values equal to the background values in Figure 2. Velocity perturbations placed above and below a $Q$ anomaly were chosen to highlight cross-talk issues as they arise due to optimization strategies and frequency bandwidth.

### Frequencies in multiscale QFWI

In the examples shown in this section, the effect of frequency information on cross-talk in QFWI is investigated. Full Gauss-Newton optimization is employed to ensure that cross-talk is not being introduced by numerical optimization.

Figure 3 illustrates the results of QFWI when one frequency is inverted at each iteration. 6 iterations were performed at each frequency, beginning at the lowest frequency, 1Hz, then increasing in 1Hz increments up to 25Hz. The problems in Figure 3 highlight important features of QFWI. Cross-talk impairs the $Q$ estimate far more strongly than the $c$ estimate, with the $Q$ anomaly not meaningfully recovered. The $c$ estimate is also strongly impacted by cross-talk, with the lower anomaly (to illuminate which raypaths have had to traverse the $Q$-anomaly) being much more poorly reconstructed. Because this cross-talk occurs despite the use of full Gauss-Newton optimization, and comprehensive simulated acquisition, it is reliably traceable to the use of a single frequency, which is similar to attempting to solve for several elastic parameters with a single angle of data.

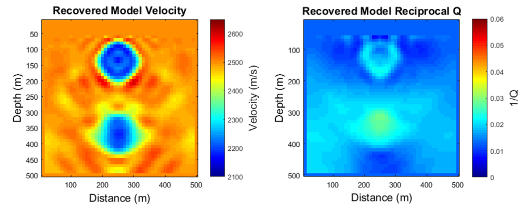


Figure 3: Gauss-Newton QFWI, inverting only one frequency at each iteration.

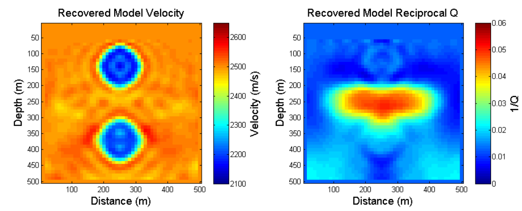


Figure 4: Gauss-Newton QFWI, inverting a 1Hz band of frequencies at each iteration. Compare with Figure 3.

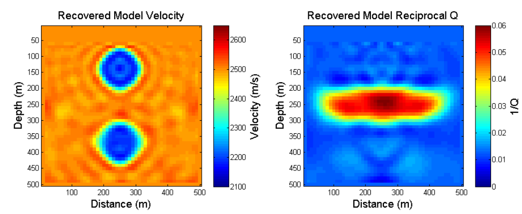


Figure 5: Gauss-Newton QFWI, inverting a broad band of frequencies at each iteration. Compare with Figures 3 and 4.

Figure 4 illustrates the results of QFWI in which a narrow band of frequencies (6 evenly-spaced frequencies in a 1Hz range) was inverted at each iteration. One iteration was carried out per band, beginning with a band centered at 1.5Hz, and increasing the center frequency by 1Hz at each iteration, up to 24.5Hz. The improved recovery of the deeper velocity anomaly and the $Q$ anomaly are notable (compare with Figure 3). This is further evidence supporting the prediction from scattering potential analysis that groups of frequencies at each iteration offer the only tangible means of discriminating between velocity and $Q$, and mitigating cross-talk.

Figure 5 illustrates the results of QFWI in which a broad band of frequencies (6 evenly-spaced frequencies over a growing range) was inverted at each iteration. The 6 frequencies were distributed from 1Hz to a maximum frequency which began at 2Hz, and increased by 1Hz per iteration to a maximum of 25Hz. This approach produces the best recovery of the three strategies. All else having been held fixed in these experiments, we conclude that iterations involving a large, varied range of frequencies are optimal for suppressing cross-talk. This is in keeping with the general principle that, to separate any two parameters, data must span experimental variables across which the two have different characteristic scattering signatures.

### Optimization strategy

Based on our conclusions above, we will from this point on use the broad frequency-band multiscale approach (i.e., the scheme by which Figure 5 was generated). An outstanding weakness in those results,

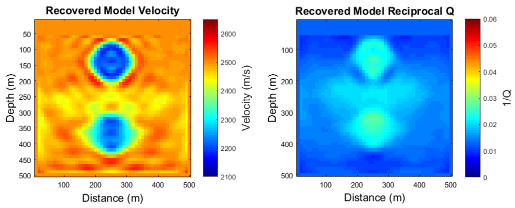


Figure 6: Result of steepest descent FWI, inverting a large band of frequencies at each iteration. Severe cross-talk is evident, despite using the same frequencies for inversion as in figure 5.

however promising, is that to achieve them we have used an exact Gauss-Newton optimization strategy, which, while viable for small models, is computationally very expensive. To improve efficiency, a steepest descent scheme can be employed, in which the Hessian is approximated with the unit operator. Figure 6 illustrates the results of doing so. For this example, 23 iterations were carried out at each frequency band. Cross-talk is dominant in the result, despite the very large number of iterations used. So, although a broad frequency band brings into the inversion sufficient information to suppress cross-talk, only with the weights provided by the Hessian is that information correctly used. In this example, the gradient was evaluated 575 times, which, in addition to the 2356 objective function evaluations required in the 575 necessary line searches, brought the total cost to 3506 wavefield simulations.

In common with other multiparameter FWI problems, something lying between the full Gauss-Newton and steepest descent schemes is needed to efficiently suppress cross-talk. The truncated Newton (TN) optimization approach plays this role. Figure 7 illustrates the result of applying truncated Gauss-Newton QFWI (hereafter TN-QFWI) with a forcing term of $10^{-3}$. One iteration was carried out at each frequency band. This result shows little evidence of cross-talk. The small forcing term makes this a close approximation to exact Gauss-Newton optimization. This forcing term magnitude also meant that a total of 1789 Hessian-vector products were evaluated, each with the same computational cost as a gradient calculation. In addition to the 25 gradient calculations, and the 100 objective function evaluations required in the 25 line searches performed, this brought the total cost to 3728 wavefield propagation problems solved.

We next consider to what degree cross-talk suppression like this can be obtained with larger forcing terms, i.e., greater efficiency. In Figure 8 the result with a forcing term of $10^{-2}$, with all else unchanged, is illustrated. The recovered model contains significant errors, mostly in the recovered Q. This approach required a total of 803 Hessian-vector product calculations, in addition to 25 gradient calculations and 100 objective function evaluations required in 25 line searches, for a total cost of 1756 wavefield propagation problems. This is suggestive that two modes of TN-QFWI be considered: a relatively low-cost mode involving a high forcing term, producing a velocity model that is very successfully "protected" from the influence of $Q$, but a $Q$ model which is itself of limited use. A higher-cost mode involving a low forcing term, produces high fidelity reconstructions of both parameters.

An interesting case to consider is that in which a less accurate TN approximation is used, but with a greater number of outer iterations performed. Figure 9 illustrates the results of this experiment: $\eta = 10^{-2}$ is used, with 2 iterations per frequency band. Very little cross-talk is observed in the recovered model for this example. The computational cost is high, however, a total of 2475 Hessian-vector products are evaluated, 683 on first iterations of a frequency band, and 1792 on the second. Combined with 50 gradient calculations and 219 objective function evaluations required in 50 line searches, a total of 5269 wavefield propagation problems were solved. There appears therefore to be a close relationship between outer iterations performed, and inner iterations required; a simple compromise between these is not easily achieved.

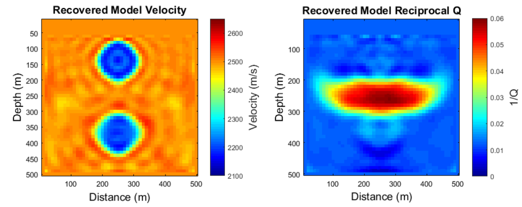


Figure 7: TN-QFWI, inverting a large band of frequencies at each iteration with a forcing term of $10^{-3}$.

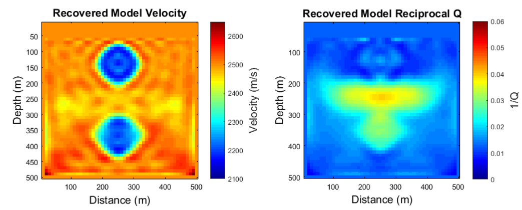


Figure 8: TN-QFWI, inverting a large band of frequencies at each iteration with a forcing term of $10^{-2}$.

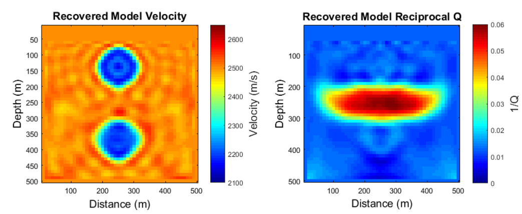


Figure 9: TN-QFWI, inverting a large band of frequencies at each iteration with a forcing term of $10^{-2}$, 2 iterations per frequency band.

## CONCLUSIONS

Cross-talk is a serious concern in QFWI, and has a particularly strong impact on the recovered Q model. Frequency dependent effects play a major role in eliminating this cross-talk. Single frequency updates produce results dominated by cross-talk, even when exact Gauss-Newton optimization is used. Inverting even a small band of frequencies per iteration offers a notable improvement, and the best results were achieved inverting the largest range of frequencies possible. Adequate consideration of Hessian information, for instance through truncated Newton methods, is crucial to using these frequencies to suppress cross-talk. A greater number of FWI iterations can compensate for a less precise estimate of the Hessian, but the cost of approximating the Hessian is not constant, and the cost of performing additional iterations at a chosen level of accuracy may not be easy to predict. We observe that TN-QFWI mode in "efficient" mode, with a high forcing term, may provide efficient and robust velocity estimates, protected from attenuation, at the cost of a well-resolved $Q$ model estimate.

## ACKNOWLEDGMENTS

We 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.