Penalty terms in the truncated Newton inner-loop for full waveform inversion

Scott Keating and Kris Innanen

ABSTRACT

Inter-parameter cross-talk, where physical properties are confused with one another, is a major concern in multi-parameter full waveform inversion. Cross-talk is difficult to prevent, not least because it is difficult to identify in a recovered model. Certain modes of cross-talk, however, may be more evident in the model updates used in the inversion. We propose penalizing evidence of cross-talk in the model updates within a truncated Newton optimization. We test this approach to penalize cross-talk between v_P and Q_P in a viscoelastic inversion. We find that an update penalty term can be effectively included in a truncated Newton strategy, but that the penalty terms we investigate are poor measures of cross-talk.

INTRODUCTION

Full waveform inversion (FWI) is a powerful tool for recovering sub-surface seismic properties, driven by a numerical optimization procedure (Tarantola, 1984). Several optimization strategies are used in FWI, including conjugate-gradient, L-BFGS, and truncated Newton optimization. In truncated Newton optimization, solving for each update is itself an iterative process, minimizing a linear objective function (e.g. Métivier et al., 2013). This is usually referred to as 'the inner problem'. The fact that each update is solved for by minimizing the inner objective, and not simply by direct calculation, has significant implications. It means that any knowledge about the observable features of a good or bad update can be included in the update calculation by adding appropriate regularization terms to the inner objective function. So, insofar as we have an understanding of what constitutes a good update independently of how it relates to the current model, we should be able to better guide the inversion. One such case may arise in viscoelastic inversion.

Viscoelastic inversion, and multiparameter inversion in general, suffer from the obstacle of cross-talk, where data residuals caused by one model parameter are mistakenly attributed to another in the inversion. This effect can be very difficult to identify in an inversion result because the true subsurface is unknown. Cross-talk in a given update, however, is often relatively simple to identify. In the case of squared P slowness v_P^{-2} and inverse of P quality factor Q_P, for instance, a phase shift relation between these parameters is often strongly indicative of cross-talk. This is because the gradients of the objective function with respect to these variables themselves differ chiefly by a phase shift. The close relation of these variables means that the unadjusted gradient is often very contaminated with cross-talk - any proposed v_P change is accompanied by a phase-shifted Q_P change. In general, there is little to reason to expect that both parts of the v_P Q_P phase-shifted pair represent real model residuals; generally we would expect that the similarity is introduced by cross-talk. Consequently, an update which adequately suppresses cross-talk can be expected to reduce these pairs to only the part corresponding to a real model residual. For this reason, we may expect that an update which does not obey the relation existing between gradients with respect to different model variables may generally be a better update than one which does. This is the concept we explore in this report.

THEORY

A conventional FWI objective function is given by

$$\phi = \frac{1}{2}||Ru - d||_2^2, \tag{1}$$

where R is a matrix applying the receiver sampling, d are the measured data, u is a forward modelled wavefield, satisfying

$$Su = f, \tag{2}$$

S is a Helmholtz matrix, representing a finite-difference approximation of the frequency domain wave equation, and f is a source term. The gradient of the objective function ϕ with respect to a given variable m can be expressed as

$$\frac{\partial \phi}{\partial m} =< \frac{\partial S}{\partial m}u, \lambda >, \tag{3}$$

where λ are the back-propagated data residuals, satisfying

$$S^{\dagger}\lambda = R^T \left(Ru - d\right). \tag{4}$$

The visco-elastic expressions used for S in this report are taken from Pratt (1990), and are elaborated on in Keating et al. (2018). All dependence of S on P-wave velocity v_P and P-quality factor Q_P is contained in its dependence on the complex Lamé parameter

$$\tilde{\lambda} = \left\{ v_P \left[1 + \frac{1}{Q_P} \left(\frac{1}{\pi} \log \frac{\omega}{\omega_0} + \frac{i}{2} \right) \right] \right\}^2 \rho - 2\tilde{\mu}, \tag{5}$$

where ω is angular frequency, ω_0 is a reference frequency, ρ is density, and $\tilde{\mu}$ is the other complex Lamé parameter.

Notably, the only term in equation 3 dependent on the choice of variable m is $\frac{\partial S}{\partial m}$; the wavefields in the inner product are independent of the variable chosen. This means that changing m simply re-weights the product of the two wavefields. The consequence of this fact is that the derivatives of the objective function with respect to two different variables are related in a relatively simple way, defined by the derivatives $\frac{\partial S}{\partial m}$ for these variables.

The relationships between these derivatives can generally be important in describing cross-talk. If we consider variables which describe different physical properties (e.g. v_P and ρ) at the same point in space, the derivative of the finite difference star in S will often correspond to gradients at the same location in space. While this can be a result of cross-talk (for instance, where there is really only a v_P or only a ρ model residual), it may also be correctly describing a change in both variables at this location. Given that a change in lithology is likely to involve differences in multiple physical properties, it is difficult to rule out either case: it is unclear whether the update contains cross-talk as both variables at that location may be changing. In this way, even though we know that cross-talk will

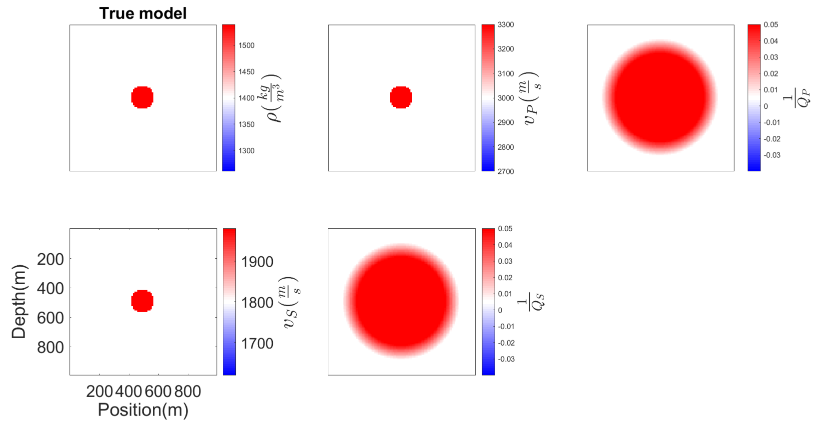

FIG. 1. True model used in synthetic tests.

often manifest as a same-location update in different physical properties, the existence of same-location updates in the inversion is not necessarily indicative of cross-talk.

The behaviour of cross-talk manifesting as a same-location change primarily holds only in the case where the $\frac{\partial S}{\partial m}$ terms are real, or have only small imaginary components. This is not the case when comparing, for instance, derivatives with respect to v_P and Q_P. Instead, these terms differ by a substantial imaginary part, due to the imaginary term in equation 5. This imaginary term introduces a phase component in the inner product in equation 3. Consequently, the gradients with respect to these variables tend to be spatially shifted from one another. Unlike in the previously discussed case, there is little reason to suspect that this relation between updates should exist based on geological concerns, so model updates which observe this relation can generally be considered as suggestive of cross-talk. Unfortunately, while this relation can be a strong indicator of cross-talk in model updates, it doesn't necessarily correspond to a simple regularization term in the inversion. This is due to the influence of iteration and multi-scaling: model updates in v_P and Q_P may be spatially shifted copies, but the net change in these parameters may not be. Instead of trying to design a cross-talk penalty term in the inversion objective function, we here explore using such a penalty term in our calculation of the model update.

NUMERICAL EXAMPLES

Penalty term design

To illustrate the effects of different penalty terms, we consider an inversion using data from a simple visco-elastic model, shown in figure 1. The inversion variables considered are squared wave slownesses, quality factors Q_P and Q_S, and density, defined at each cell in the finite difference grid used for wave propagation. The initial model is the constant background value for each parameter. The gradient of the objective function with respect to the v_P and Q_P dependent variables for the conventional L_2 objective function is shown in figure 2. In the example shown, sources and receivers are located at the surface, and five frequencies evenly spaced from 1 Hz to 10 Hz are considered.

There are strong reasons to suspect that the gradient shown in figure 2 is substantially contaminated by cross-talk, independent of substantial prior knowledge of the true model.

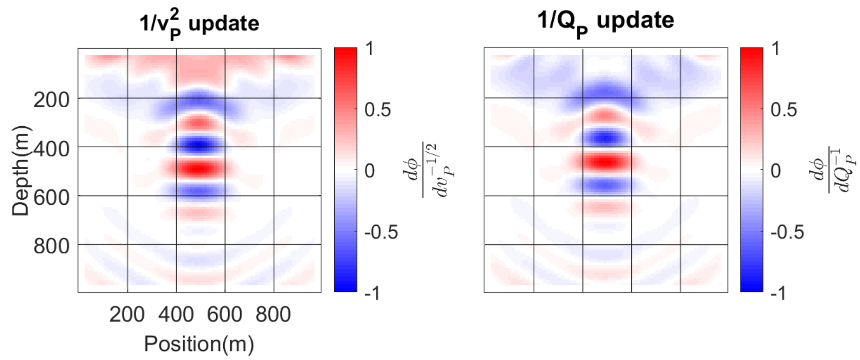

FIG. 2. Gradient of objective function with respect to v_P^{-2} (left) and Q_P^{-1} (right). The scale here is relative to the maximum amplitude of each. These gradients are strongly suggestive of cross-talk.

The gradient with respect to each variable appears as a spatial shift of the other. If we allow that there is negligible likelihood of geologic trends duplicating themselves in these two parameters at small spatial removes, we can confidently infer that cross-talk is the source of this connection. In fact, the phase term in the gradient calculation is causing the same data residuals to suggest these similar, but geologically inconsistent updates for these two parameters. The spatial shift associated with this phase term is determined by the frequency considered, but as we consider several frequencies the net shift is some combination of those for each frequency, and is not easy to predict. There is, however, a spatial shift which can bring these gradients into close agreement.

A key strategy for cross-talk mitigation is to employ second-derivative information in the inversion's optimization procedure. In this report, we focus on the use of truncated Gauss-Newton (TGN) optimization to include such information in the inversion (e.g. Métivier et al., 2013). This style of optimization is meant to approximate Gauss-Newton optimization, wherein a model update at each iteration is given by

$$\Delta m = -H_{GN}^{-1} g, \tag{6}$$

where $H_G N$ is the residual-independent part of the second derivative matrix (the Hessian), and g is the gradient. Exact Gauss-Newton optimization is typically not employed in FWI due to the prohibitive cost of calculating, storing, and inverting H_{GN}. In TGN optimization, the update used is an approximation to the Gauss-Newton update, calculated at each iteration of the inversion. This approximation is usually calculated by finding an approximate minimum of an 'inner' objective function (the FWI objective being the 'outer' objective) with respect to Δm, given by

$$\psi = \Delta m^T H_{GN} \Delta m + \Delta m^T g. \tag{7}$$

This function and its derivative can calculated without explicit calculation of H_{GN}, only the product $H_{GN} \Delta m$ is needed, and this can be efficiently calculated. More details on

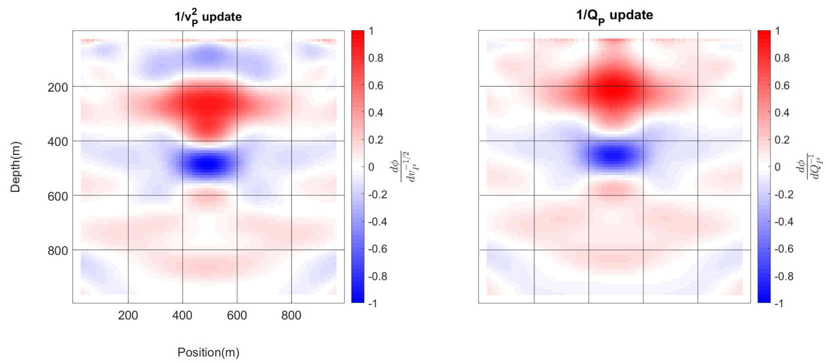

FIG. 3. TGN model update calculated for v_P^{-2} (left) and Q_P^{-1} (right). The scale here is relative to the maximum amplitude of each. Though different from the gradient, these updates are strongly suggestive of cross-talk.

the calculation of H_{GN} and approaches for minimizing ψ can be found in (Keating and Innanen, 2016). In general, ψ is minimized iteratively, with larger numbers of iterations corresponding to a more accurate approximation of the Gauss-Newton update.

While Gauss-Newton optimization has powerful cross-talk reduction capabilities (Pratt et al., 1998), an approximation to this update which sufficiently reduces cross-talk can be expensive to calculate. Furthermore, much of the information in the Hessian is unrelated to cross-talk (e.g. Innanen, 2014; Keating et al., 2018), and the inner objective ψ in equation 7 is unable to prioritize the cross-talk related information. Figure 3 shows a TGN model update after 20 iterations of inner-loop optimization. Clearly, the update has been modified substantially from the gradient in the TGN inner loop, but the cross-talk problem is little changed: both updates are still primarily shifted versions of one another. Using the same criteria as in the gradient case above, we can identify this update as being undesirable from a cross-talk perspective. Knowledge like this motivates an alteration to the inner loop in equation 7. Provided we can define some metric $\xi(\Delta m)$ which is large when an update contains cross-talk, and small when it does not, we can change the inner objective of TGN to better achieve our inversion goals:

$$\psi = \Delta m^T H_{GN} \Delta m + \Delta m^T g + \xi(\Delta m). \tag{8}$$

With an appropriate choice of ξ, this approach may allow for more efficient cross-talk avoidance in FWI.

Cross-correlation penalty term

Unfortunately, an appropriate choice of ξ may be difficult to make. When penalizing cross-talk between Q_P and v_P, one intuitive choice is a cross-correlation based term. We define a cross-correlation version of the inner objective as

$$\xi_{cc} = \sum_i \frac{1}{\lambda} \Delta m_{Q_P}(r_i) \Delta m^*_{v_P}(r_i), \tag{9}$$

where r_i represents the i^{th} position in the model, Δm_{Q_P} is the current $\frac{1}{Q_P}$ update, $\Delta m^*_{v_P}$ is the current s_P update, with a spatial shift applied, and λ is a normalizing term, de-

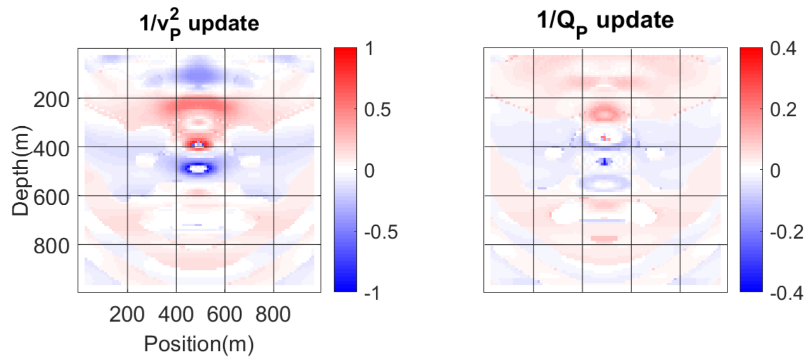

FIG. 4. TGN model update calculated for v_P^{-2} (left) and Q_P^{-1} (right) using ξ_{cc} as a penalty term. The scale here is relative to the maximum amplitude of each. Severe high-frequency artifacts are present.

emphasizing the magnitude of Δm_{Q_P} and $\Delta m^*_{v_P}$. If this shift is chosen to be the same as the one which matches the gradients with respect to these variables, this metric should be large for some cross-talk heavy updates.

A major obstacle to the effective use of penalty terms in this way is the high degree of accuracy needed in the cross-talk penalizing term ξ. If ξ can be reduced without removing cross-talk, it is very possible that its inclusion will have unwelcome effects on the inversion result. Figure 4 shows a TGN update calculated in the same way as figure 3, but with the inclusion of a penalty term like ξ_{cc}. There are still strong similarities between these models suggestive of cross-talk, but high frequency artifacts are also present, and these manage to reduce the cross-correlation between Δm_{Q_P} and $\Delta m^*_{v_P}$. In this sense, ξ_{cc} is insufficiently linked to cross-talk to be able to improve TGN updates.

The artifacts present in figure 4 are problematic, but also highly distinctive. Changes on scales much smaller than the seismic wavelengths considered are introduced to lower the cross-correlation, while longer scale features remain highly correlated. This suggests that ξ_{cc} may not be fatally flawed as a measure of cross-talk, only incomplete. To prevent the type of artifacts associated with ξ_{cc}, we can add an additional term ξ_ω which penalizes wavenumbers in the update which aren't present in the gradient. Figure 5 shows a TGN update calculated with both the ξ_{cc} and ξ_ω penalty terms. Here, the artifacts associated with ξ_{cc} alone have been eliminated, but the more significant problems with the cross-correlation penalty term have been uncovered. Cross-talk is still clearly evident in that each update is very nearly a scaled version of the other. The almost uniform amplitudes in this update suggest that the normalized cross-correlation has been reduced by minimizing the variation in amplitude. As this update is both highly effective in reducing ξ_{cc} and heavily cross-talked, it is suggestive of serious flaws in the design of ξ_{cc} as a penalty term. Rather than further modifying this penalty term, we consider other possible penalties.

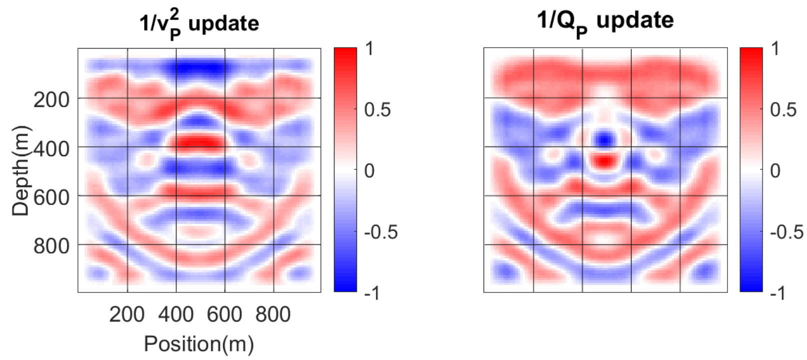

FIG. 5. TGN model update calculated for v_P^{-2} (left) and Q_P^{-1} (right) using ξ_{cc} and ξ_ω as a penalty terms. The scale here is relative to the maximum amplitude of each. The regularization term results in near-uniform amplitudes.

Subtraction penalty term

The errors associated with the cross-correlation-type penalty term arise because the cross-correlation between the shifted gradients is not a good measure of the cross-talk. In particular, identical structures could be reduced to negligible cross-correlation by equalizing the amplitudes throughout. A better metric for undesirable behaviour may improve results. We next consider the metric defined by

$$\xi_{-} = -\frac{\sum_i \frac{1}{\lambda}\left(\Delta m^*_{v_P}(r_i) - \alpha \Delta m_{Q_P}(r_i)\right)^2}{\sum_i \Delta m^*_{v_P}(r_i)^2}, \tag{10}$$

where

$$\alpha = \frac{\sum_i \left(\Delta m^*_{v_P}(r_i)\Delta m_{Q_P}(r_i)\right)^2}{\sum_i \Delta m_{Q_P}(r_i)^2}. \tag{11}$$

This metric measures the relative amplitude of the difference between the shifted v_P update and a scaled version of the Q_P update. The scale is chosen to give these terms similar amplitudes. There are several advantages of this metric. Like ξ_{cc}, it is large when the updates in different model variables are shifted versions of one another. It also achieves lower values when these updates have different geometry from one another. Unlike ξ_{cc}, however, an overall equalization in amplitudes is unable to reduce this metric. For this reason, it may be expected that ξ_{-} can work as a better penalty term.

Figure 6 shows a TGN update calculated using the ξ_{-} and ξ_ω penalty terms. There are several improvements visible here as compared to figures 4 and 5. Significantly, no major artifacts are evident in the update, the main features are those that were present in the original gradient. Additionally, some progress has been made toward reducing the cross-talk feature we identified: the updates in these variables are no longer near-identical under a spatial shift. On the other hand, substantial overlap between these updates is still in evi-

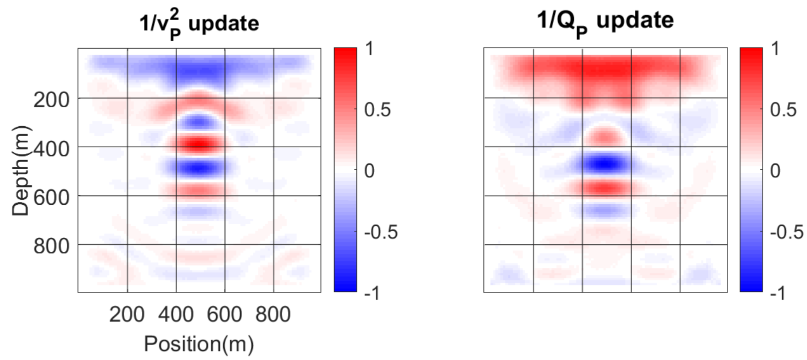

FIG. 6. TGN model update calculated for v_P^{-2} (left) and Q_P^{-1} (right) using ξ_- and ξ_ω as a penalty terms. The scale here is relative to the maximum amplitude of each. Evidence of cross-talk persists here.

dence, though less in extent, and the updates are not very different from the original gradient. Overall, the TGN update in figure 6 shows both promising and potentially problematic features, we investigate a full FWI procedure using this approach to better understand its performance.

To test the effectiveness of ξ_-, we attempted an inversion of synthetic data generated using the model in figure 1. The initial model for the inversion was the same one used for the model update calculations above, a constant background. Data from 49 explosive sources and 98 multi-component receivers evenly-spaced along the top of the model were assumed to be available. Ten frequency bands of five evenly-spaced frequencies were considered in the inversion, starting with 1-2 Hz, and ending with 1-20 Hz. At each frequency band, one iteration of truncated-Gauss-Newton (TGN) optimization was performed, with a maximum of 20 inner loop iterations per FWI iteration.

The result of this inversion approach is shown for v_P and Q_P in figure 7. This result suggests the penalty term we used was unsatisfactory in a few ways. First, substantial cross-talk is evident based on comparison with figure 1. The Q_P changes in particular strongly represent the structure of the true v_P model. This makes clear that the penalty terms did not mitigate cross-talk as they were designed to. Another disappointing feature of the result is the small amplitude of changes from the background. A conventional TGN viscoelastic inversion with similar cost is capable of recovering features on the scale of the true anomalies, as demonstrated by Keating et al. (2018). The fact that this could not be reproduced here suggests that the penalty terms actually hampered the effectiveness of the TGN algorithm in this case.

DISCUSSION

The numerical tests investigated here seem to establish that neither of the proposed metrics, ξ_{cc} or ξ_{GN}, are effective in reducing the v_P - Q_P cross-talk mode we studied. In

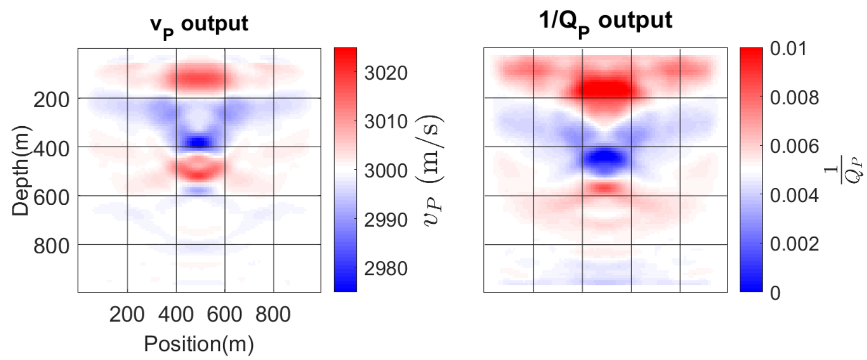

FIG. 7. Inversion output for v_P (left) and Q_P (right) using ξ_- and ξ_ω as a penalty terms. Comparison with figure 1 suggests severe cross-talk remains.

both cases, however, the TGN optimization strategy was effective in calculating an update which reduced the metric we defined. The question which will determine the effectiveness of this penalty approach is then whether any metric can be defined that is effective in quantifying cross-talk.

CONCLUSIONS

In this report, we attempted to prevent cross-talk in full-waveform inversion through a penalty term in the inner loop of the truncated Gauss-Newton optimization. We suggest that a penalty term which correctly identifies cross-talk in updates may allow for suppression of cross-talk. This offers a potential advantage over penalty terms on the FWI objective function, which will be restricted to penalizing model features, rather than features of the model update. We proposed several penalty terms to avoid v_P - Q_P cross-talk, but none of these were found to be effective. The possibility remains that a more appropriately chosen penalty term may be effective in reducing cross-talk.

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

Innanen, K., 2014, Seismic avo and the inverse hessian in precritical reflection full waveform inversion: Geophys. J. Int., **199**, 717–734.

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

Keating, S., Li, J., and Innanen, K. A., 2018, Viscoelastic fwi: solving for q_p, q_s, v_p, v_s, and density:

CREWES Annual Report, **30**.

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

Pratt, R., 1990, Frequency-domain elastic wave modeling by finite differences: A tool for crosshole seismic imaging: Geophys. J. Int., **133**, 341–362.

Pratt, R., C[]Shin, and G[]J[]Hicks, 1998, Gauss-newton and full newton methods in frequency-space seismic waveform inversion: Geophys. J. Int., **133**, 341–362.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.