Viscoacoustic WRI - obstacles for multiparameter cycle-skipping strategies?

Scott Keating and Kris Innanen

ABSTRACT

The obstacle of cycle-skipping in full-waveform inversion has led to the development of many non-L_2 objective functions with better convexity properties. While these are effective in their goal of mitigating cycle skipping, they are typically investigated only with a single-parameter inversion formulation. If simultaneous multi-parameter inversion is to be an effective technology, it is important to understand how these objective functions behave when more than one medium property is considered. Here, we investigate the behaviour of one such strategy, wavefield reconstruction inversion, in the case of simultaneous viscoacoustic inversion. We find that prohibitive cross-talk between inversion variables is recovered when adopting this approach.

INTRODUCTION

Full waveform inversion (FWI) has been successfully used as a tool for intermediate scale velocity-model building in recent years (e.g. Virieux and Operto, 2009). The potential uses of FWI go far beyond velocity-model building however, and FWI has shown promise as a workable inversion strategy for viscoelastic and anisotropic medium properties (e.g. Tarantola, 1986; Hicks and Pratt, 2001; Barnes et al., 2008; Choi et al., 2008; Alkhalifah and Plessix, 2014). At present, there are a few major obstacles limiting the potential of FWI, cycle-skipping and inter-parameter cross-talk being two of the most significant.

Cycle-skipping occurs when the inversion becomes caught in certain local minima of the objective function. The problem arises when the modeled and measured data differ by an integer number of cycles, leading to a partial reduction of the data residuals. Better models can correctly align measured and modeled data, but an incremental step toward such a model will actually cause the objective function to decrease when the model is cycle-skipped. This prevents convergence to a desirable inversion result.

Cycle-skipping can be avoided in FWI through the use of multiscaling strategies (e.g. Bunks et al., 1995), but these suppose access to sufficiently low frequencies. When low frequencies are unavailable, it can be necessary to consider objective functions other than the conventional L_2 norm of data residuals to prevent cycle skipping. Many appropriate objective functions have been proposed. Some re-state the objective function in forms that enlarge the convex region around the minima associated with correct data fitting (Bozdağ et al., 2011; Engquist and Froese, 2014). Others allow for the data to be optimally matched even with poor models, by introducing new inversion parameters designed to enable this matching (van Leeuwen and Herrmann, 2013; Guanghui and Symes, 2015; Warner and Guasch, 2016). These approaches consider both a measure of data residuals and a term penalizing the data-matching variables in the inversion, aiming to eliminate the latter by the end of the inversion. Each of these strategies has shown significant improvement over conventional FWI in preventing cycle-skipping.

A little-studied case in FWI is that of cycle-skipping in simultaneous, multi-parameter inversion. Multi-parameter inversion (e.g., Operto et al., 2013; Plessix et al., 2013; Pan et al., 2016) is important in achieving the potential of FWI because accurate models of seismic wave propagation involve multiple physical properties; P-wave velocity alone is insufficient. Cycle skipping is as significant a problem in multi-parameter inversion as it is in the single-parameter case, and for the same reasons. An additional consideration in multi-parameter inversion, however, is the potential for inter-parameter cross-talk ('cross-talk' hereafter). Cross-talk occurs when data residuals caused by an error in the estimate of one physical property cause the estimate of another property to be changed in the inversion. This behaviour can be seriously detrimental to inversion results as it can greatly slow convergence, and can call into question the accuracy of inversion outputs. Cross-talk is a major concern even when cycle-skipping is not a risk, so the effects of cycle-skipping avoidance techniques on cross-talk are of considerable significance.

In this report, we investigate visco-acoustic inversion using the wavefield reconstruction inversion (WRI) technique (van Leeuwen and Herrmann, 2013). In WRI, the modeled seismic wavefield is considered as an additional inversion parameter, and the objective function penalizes both data misfit and the extent to which the modeled wavefield violates the wave equation. This approach allows for the data to be matched even when the subsurface model is poor, so it has been proposed as an effective way of avoiding cycle-skipping. It also frames the optimization problem driving the inversion in a way that makes the explicit dependence of the objective function on the subsurface model much simpler than in conventional FWI, which might help to mitigate cross-talk. These features make multiparameter WRI an interesting case for studying the relation between cycle-skip prevention and cross-talk.

THEORY

Acoustic wavefield reconstruction inversion

The optimization problem driving full waveform inversion can be expressed as

$$\min_{m,u} \sum_{i,k,\omega} \frac{1}{2} \left(\sum_j R_{ij} u_{jk} - d_{ik} \right)^2, \quad \text{subject to} \quad \sum_{j,k,\omega} A_{ij}(m) u_{jk} = f_{ik}, \tag{1}$$

where m is the subsurface model, u is the pressure wavefield throughout the model, R is a matrix representing receiver sampling, d are the measured data, f is a source term, A is the Helmholtz matrix containing the finite-difference approximation for wave propagation (visco-acoustic in this case), the indices i, j, and k represent receiver number, spatial position and source number, and there is an overall sum over frequencies ω, which we omit for simplicity in later equations. In conventional FWI, we require that the constraint is always satisfied:

$$\min_{m,u} \sum_{i,k} \frac{1}{2} \left(\sum_{j,h} R_{ij} A^{-1}_{jh}(m) f_{hk} - d_{ik} \right)^2, \tag{2}$$

which results in an optimization problem highly nonlinear in m due to the strongly nonlinear dependence of A^{-1} on m. In the WRI approach (van Leeuwen and Herrmann, 2013),

we instead allow violation of the constraint during intermediate steps to achieve a more linear problem, given by

$$\min_{m,u} \sum_{i,k} \frac{1}{2} \left(\sum_j R_{ij} u_{jk} - d_{ik} \right)^2 + \sum_{i,k} v_{ik}^* \sum_j (A_{ij}(m) u_{jk} - f_{ik}), \tag{3}$$

where $v_{i,k}$ are Lagrange multipliers. While this formulation is appealing, it requires calculation and storage of both u and v, which is typically prohibitive. A similar formulation which avoids this obstacle is

$$\min_{m,u} \phi(m,u) = \sum_{i,k} \frac{1}{2} \left(\sum_j R_{ij} u_{jk} - d_{ik} \right)^2 + \frac{\lambda^2}{2} \sum_{i,j,k} \left(A_{ij}(m) u_{jk} - f_{ik} \right)^2, \tag{4}$$

where λ is a penalty term which increases as the inversion proceeds, and ϕ is called the objective function. This objective function can be split into a data misfit penalty term, ϕ_D, and a wave-equation enforcing term, ϕ_u:

$$\phi(m,u) = \phi_D(u) + \lambda^2 \phi_u(m,u), \tag{5}$$

where

$$\phi_D = \sum_{i,k} \frac{1}{2} \left(\sum_j R_{ij} u_{jk} - d_{ik} \right)^2, \tag{6}$$

and

$$\phi_u = \frac{1}{2} \sum_{i,j,k} (A_{ij}(m) u_{jk} - f_{ik})^2. \tag{7}$$

While there is conceptual appeal to inverting for both m and u simultaneously, this is not generally practical due to the necessity for the storage of u. Instead, a sequential minimization of ϕ is considered in WRI, alternating minimization with respect to u and m. In the acoustic case, this can be achieved without storage of u as a whole. The derivative of the objective function with respect to the wavefield for source k, u_k, is given by

$$\frac{d\phi}{du_k} = R^T (R u_k - d_k) + \lambda^2 A^\dagger (A u_k - f_k). \tag{8}$$

When the derivative is zero, we have

$$\begin{bmatrix} \lambda A \\ R \end{bmatrix} u_k = \begin{bmatrix} \lambda f_k \\ d_k \end{bmatrix} \tag{9}$$

When minimizing the objective with respect to u, we set u to the value minimizing the objective, with constant m:

$$u_k = \left(\begin{bmatrix} \lambda A^T & R^T \end{bmatrix} \begin{bmatrix} \lambda A \\ R \end{bmatrix} \right)^{-1} \begin{bmatrix} \lambda A^T & R^T \end{bmatrix} \begin{bmatrix} \lambda f_k \\ d_k \end{bmatrix}. \tag{10}$$

On alternate iterations, we minimize the objective with respect to m, holding u constant. The derivative of the objective with respect to the n^{th} model variable, m_n, is given by

$$\frac{d\phi}{dm_n} = \sum_k \left(\frac{dA}{dm_n} u_k \right)^\dagger (Au_k - f_k) \,. \tag{11}$$

In the original formulation of WRI (van Leeuwen and Herrmann, 2013), acoustic wave-propagation was considered, and the variables were chosen such that m_n represents the squared slowness of the medium at the n^{th} defined location in the medium. For an acoustic medium, A is a matrix approximating the partial differential operator $\frac{\omega^2}{v^2} + \nabla^2$, where v is the vector of wave speeds in the model. The elements $A_{i,j}$ can be defined as

$$A_{i,j} = \delta(i,j)\frac{\omega^2}{v_i^2} + K_{i,j}, \tag{12}$$

where K represents the finite-difference approximation of ∇^2. For this particular choice of problem (acoustic wave propagation and variables representing squared slowness), the derivative in equation 11 can be re-written as

$$\frac{d\phi}{dm_n} = \sum_k \left(\omega^2 u_k \delta(n,l) \right)^\dagger \left(\omega^2 m u_k + [Ku_k - f_k] \right)_l \,. \tag{13}$$

Equivalently,

$$\frac{d\phi}{dm} = \sum_k \omega^4 u_k^\dagger u_k m + \omega^2 u_k^\dagger \left[Ku_k - f_k \right] . \tag{14}$$

The optimal choice of m at fixed u can be found by setting this derivative to zero:

$$m = -\sum_k \frac{u_k^\dagger \left[Ku_k - f_k \right]}{\omega^2 u_k^\dagger u_k}. \tag{15}$$

Because this expression involves a sum over frequencies and sources, only the wavefield for a single source at one frequency u_k is needed at any given step of this calculation. This means that u as a whole does not need to be stored in this approach: it can be calculated on the fly.

Iterations of acoustic wavefield reconstruction then consist of two key steps. First, there is a minimization with respect to u through equation 10. Solving this equation is similar in cost to solving a forward modeling problem ($u = A^{-1}q$) provided the number of rows in R is much less than the number of rows in A, or equivalently the number of model variables is much larger than the number of receivers. Notably, equation 10 holds even if a different model of wave-propagation or choice of variables is used in the inversion. The second step is a minimization with respect to m through equation 15. This equation holds only for acoustic wave propagation and variables describing squared slowness each at one cell in the finite-differences grid. To employ WRI with different wave physics or other variable choices, another solution must be found for this second step.

Viscoacoustic wavefield reconstruction inversion

In this report, we seek to formulate WRI for the viscoacoustic problem. For this type of wave propagation, we can define A by

$$A_{i,j} = \delta(i,j)\frac{\omega^2}{s_i} + K_{i,j}, \tag{16}$$

where

$$s = s_c(1 - s_q\gamma), \tag{17}$$

s_c is the squared slowness of the medium at reference frequency ω_0, s_q is the reciprocal of the quality factor, and

$$\gamma = i + \frac{2}{\pi}\log\left(\frac{\omega}{\omega_0}\right). \tag{18}$$

Attempting to reproduce an expression for the model update in viscoacoustic WRI in analogy to equation 15, we must first determine the derivative of the objective function with respect to each variable. Starting from the general result in equation 11, it can be shown that

$$\frac{d\phi}{ds_{c_n}} = \sum_k \omega^4(u_k^\dagger u_k)\left[s_{c_n}s_{q_n}^2(\gamma^\dagger\gamma) + s_{c_n}s_{q_n}(\gamma^\dagger+\gamma) + s_{c_n}\right] + \left[\omega^2 s_{q_n}\gamma^\dagger u^\dagger + u^\dagger\right](\nabla^2 u_k - q_k), \tag{19}$$

and

$$\frac{d\phi}{ds_{q_n}} = \sum_k \omega^4(u_k^\dagger u_k)\left[s_{c_n}^2\gamma^\dagger + s_{c_n}s_{q_n}(\gamma^\dagger\gamma)\right] + \omega^2 s_{c_n}\gamma^\dagger u^\dagger(\nabla^2 u_k - q_k). \tag{20}$$

Unfortunately, the system of equations defined by equations 19 and 20 is not amenable to direct solution as equation 13 was. This is problematic, as the number of systems of equations which must be solved is the same as the number of velocity and Q variables in the model, which places a major additional computational burden on the inversion. This problem is not specific to the viscoacoustic problem, it will also occur in the elastic case, or even the constant-density acoustic case when variables with spatial extent are considered. In some multi-parameter cases, the associated systems of equations can be quite large, or very nonlinear. Because of this obstacle, we consider an iterative, approximate minimization with respect to m rather than a direct, exact solution, as outlined in van Leeuwen and Herrmann (2015).

To find the m which minimizes the objective function at fixed u, we consider the optimization problem

$$\min_m \phi_m(m) = \sum_{i,j,k}(A_{ij}(m)u_{jk} - q_{ik}). \tag{21}$$

In this report, we solve equation 21 through the L-BFGS optimization strategy (Nocedal and Wright, 2006). The derivatives required within the L-BFGS procedure are given in equations 19 and 20. Iteratively minimizing $\phi_m(m)$ is inefficient in that it requires either storage of u, which is generally considered unfeasible in WRI, or evaluation of the appropriate u at each iteration of the L-BFGS procedure. In this sense, the approach we explore here is not practical, but represents a scenario in which some of the obstacles to

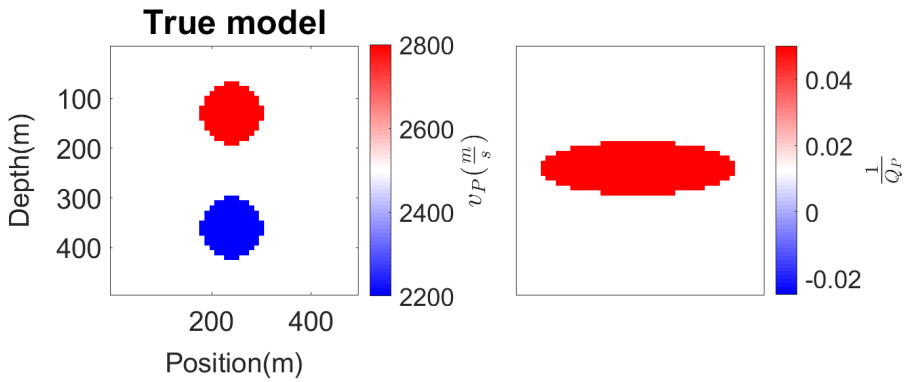

FIG. 1. Model used to generate data in the synthetic test. Left is P-wave velocity at the reference frequency ω_0, right is $\frac{1}{Q}$.

multi-parameter WRI have been removed. The numerical example investigated in the next section highlights that this does not alter our assessment of the approach: the WRI method fails even in this simplified case.

NUMERICAL EXAMPLES

Due to the requirement for storage of u in solving equation 21, we consider a relatively small example, where such storage is feasible. A toy model is used for sythetic generation of the 'measured data' to be inverted, shown in figure 1. The key feature of this model for our purposes are that the v_P and Q_P features do not overlap, making identification of some cross-talk modes easier. For these examples we consider five frequencies, evenly spaced from 1 to 20 Hz. The variables considered in the inversion are $s_c r_i$, parameterizing the squared slowness at each finite-difference grid point used in the wave propagation, and $\hat{s}_q(r_i)$, which defines Q through

$$\frac{1}{Q(r_i)} = \alpha \hat{s}_q(r_i), \tag{22}$$

where α is a fixed scale factor, chosen to make the gradients with respect to $\hat{s}_q$ and s_c similar in amplitude. The initial model used in the examples was a uniform velocity of 2500 m/s, and Q of 80; the background values of the true models.

Before investigating the inversion as a whole, it's valuable to verify that our strategy for solving equation 21 provides a reasonable approximation of the exact solution. For this check, we solve equation 21, using the u corresponding to the true model, but initializing the model with the constant background described above. In this test, and in the inversion later, we consider 20 iterations of L-BFGS optimization in the solution of 21. Figure 2 shows the difference between the model recovered using this approach and the true model. The difference between this result and the true model is extremely small, suggesting that our approach is a good approximation to exact solution.

In the inversion, the variable λ plays an important role. This term controls the trade-off

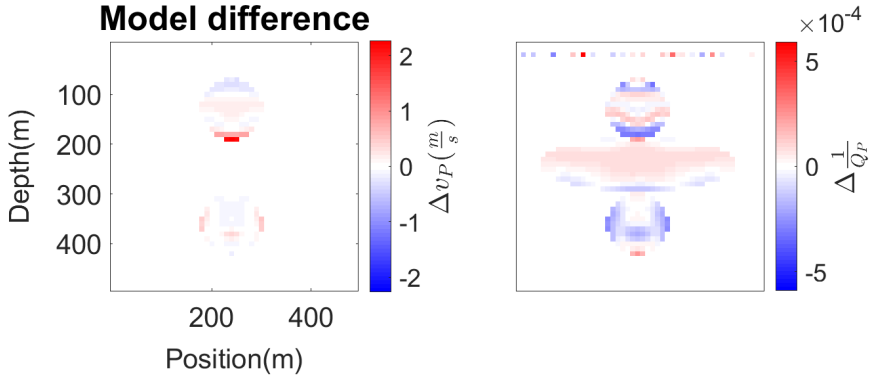

FIG. 2. Difference between true model and recovered model using L-BFGS for solution of equation 21. Left is P-wave velocity at the reference frequency ω_0, right is $\frac{1}{Q}$. Note the change in scale from figure 1

between data-fit and physics violation (equation 4). This term should be very large at the end of the inversion to ensure that the inversion output is linked to the wave equation considered in the inversion. Ideally, however, it should not be so large early in the procedure, because if the wavefields considered never violate the wave equation, the WRI procedure simply reproduces conventional FWI. These requirements suggest that λ should start small, but grow substantially throughout the inversion. We consider a λ that increases by 5% of its previous value at each iteration.

For WRI inversion of the modeled data, we considered all the frequencies of the available data simultaneously. A total of 300 iterations of m updates were used in the inversion. The output of this approach is shown in figure 3. Unfortunately, this result is very poor in several ways. Extremely strong cross-talk seems clearly evident here; the recovered Q model, for instance, is strongly affected by the geometry of the true velocity model, with the upper ball being the main identifiable feature in the output. Additionally, there is negligible recovery of the deeper parts of the model.

The inversion result is dependent in part on the choice of the objective weighting term λ. To verify that a reasonable choice of λ was made throughout the inversion, we examine the objective function history of the inversion for data residual term ϕ_D, and wave equation enforcing term ϕ_u. These are plotted in figure 4. In this case, ϕ_D starts very close to zero - suggesting that the wavefield is chosen to minimize data misfit without regard for the physics constraints, as desired in the WRI procedure. As the number of iterations increases and more weighting is given to ϕ_u, penalizing wave equation violation, ϕ_D begins to increase, while ϕ_u is reduced to negligible values by the end of the inversion - meaning that u and m effectively satisfy the wave equation by the end of the inversion, and the data misfit is not completely reduced in consequence. These starting and ending objective function criteria display exactly the type of behaviour we would hope to see in WRI. In keeping with this observation, it is unlikely that a different choice of λ schedule would result in a substantially better inversion output.

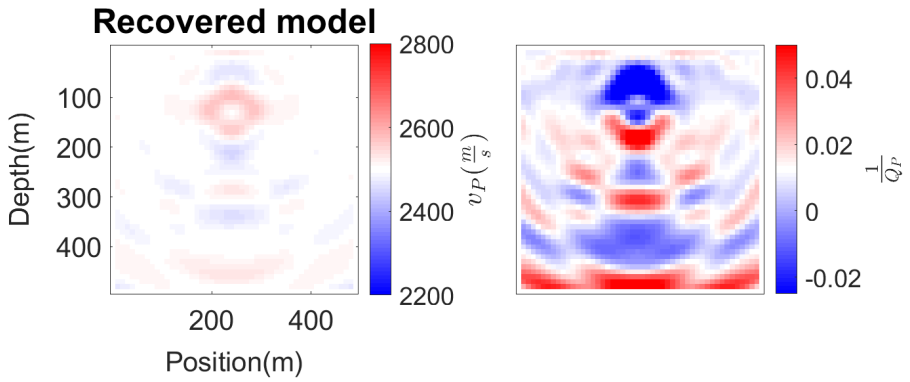

FIG. 3. Model recovered using visco-acoustic WRI approach. Left is P-wave velocity at the reference frequency ω_0, right is $\frac{1}{Q}$.

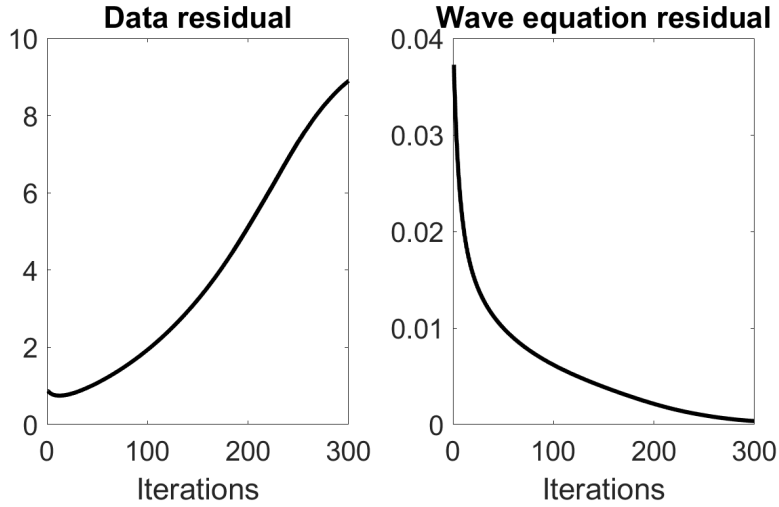

FIG. 4. Data penalty term ϕ_D and wave-equation enforcing term ϕ_u as a function of inversion iteration.

DISCUSSION

How well does this method extend to multi-parameter inversion in general?

One interesting feature of the visco-acoustic WRI problem is the difficulty in reproducing the very simple wavefield-model relation defined in equation 15 for the acoustic case. The visco-acoustic problem is not unique in lacking an analog for this simple relation; most multi-parameter problems will require that nonlinear systems of equations be solved to determine the model exactly corresponding to the wavefield. To avoid solving such a system, we instead considered approximate solutions of the problem defined in equation 21. This approach is generally applicable to a multi-parameter problem, but requires knowledge of u at each step of an iterative optimization. Here we stored u, but this is not practical for problems on the scales of interest. Re-assessment of u at each iteration of the optimization is also impractical in general. In this sense, the strategy used here to implement multi-parameter WRI is impractical. Expressions like equation 15, exact expressions for model parameters in terms of the complex wavefield from a single source, cannot constrain more than two parameters simultaneously. This means that extensions of equation 15 to most multi-parameter problems cannot be formulated. Overall, there are serious concerns about the practicality of extending WRI to multi-parameter inversion, which exist in addition to concerns about the quality of the results.

Do cycle-skip beating methods always become prone to cross-talk?

Methods for cycle-skipping avoidance generally have lower spatial resolution than the conventional L_2 objective. An intuitive explanation for this is that these approaches all seek to make the objective function more sensitive to large shifts in travel times, and less sensitive to local minima associated with small travel-time changes. This results in a strong sensitivity to large wavelengths in the velocity model and weak sensitivity to shorter wavelengths. This focus on long wavelengths makes it easy for inversion variables describing the p-wave velocity of the medium at two nearby points to be confused - strong cross-talk exists between such variables. When considering multi-parameter inversion with these cycle-skipping motivated methods, an important question is whether this variable confusion extends to inter-parameter cross-talk. At first glance it seems unlikely that that this confusion should extend to the inter-parameter case; emphasizing larger travel-time shifts does not have an obvious connection to the relation between different physical properties in the same way that it connects variables at different locations in space. The results presented here, however, strongly suggest the opposite, at least in the case of WRI. Further analysis of the behaviour of cycle-skip preventing multi-parameter FWI algorithms should be done to better establish the effect of cross-talk in these techniques.

CONCLUSIONS

In this report, we developed a technique for extending wavefield reconstruction inversion (WRI) to the constant-density viscoacoustic case. The sequential optimization approach, inverting for first the objective-minimizing wavefield, then the objective-minimizing model, was used, as in the original acoustic algorithm. Direct calculation of the objective-minimizing wavefield was viable in this formulation, but direct solution for the minimizing

model could not be pursued through the original strategy. Instead, an iterative approach was proposed, though this necessitated storage of the calculated wavefield. Synthetic tests using this inversion approach were able to navigate from small data residual, largely wave-equation violating initial models to larger data residual, largely wave-equation enforcing final models, as desired in WRI. These results were strongly contaminated with cross-talk, with the Q model in particular being very poorly recovered. These results raise interesting questions about the viability of FWI approaches designed to mitigate cycle-skipping in the multiparameter case.

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

Alkhalifah, T., and Plessix, R., 2014, A recipe for practical full-waveform inversion in anisotropic media: An analytical parameter resolution study: Geophysics, **79**, R91–R101.

Barnes, C., Charara, M., and Tsuchiya, T., 2008, Feasibility study for an anisotropic full waveform inversion of cross-well data: Geophysical Prospecting, **56**, 897–906.

Bozdağ, E., Trampert, J., and Tromp, J., 2011, Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements: Geophysical Journal International, **185**, 845–870.

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

Choi, Y., Min, D., and Shin, C., 2008, Two-dimensional waveform inversion ofmulti-component data in acoustic-elastic coupled media: Geophysical Prospecting, **56**, 863–881.

Engquist, B., and Froese, B., 2014, Application of the wasserstein metric to seismic signals: Commun. Math. Sci., 12 5 979 988.

Guanghui, H., and Symes, W., 2015, Full waveform inversion via matched source extension: SEG Technical Program Expanded Abstracts, 1320–1325.

Hicks, G., and Pratt, R., 2001, Reflection waveform inversion using local descent methods: Estimating attenuation and velocity over a gas-sand deposit: Geophysics, **66**, 598–612.

Nocedal, J., and Wright, P. S., 2006, Numerical Optimization: Springer, 2nd edn.

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.

Plessix, R.-E., Milcik, P., Rynia, H., Stopin, A., Matson, K., and Abri, S., 2013, Multiparameter full-waveform inversion: marine and land examples: The Leading Edge, **Sep**, 1030–1038.

Tarantola, A., 1986, A strategy for nonlinear inversion of seismic reflection data: Geophysics, **51**, 1893–1903.

van Leeuwen, T., and Herrmann, F. J., 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, 661–667.

van Leeuwen, T., and Herrmann, F. J., 2015, A penalty method for pde-constrained optimization in inverse problems: Inverse Problems, **32**.

Virieux, J., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, No. 6, WCC1.

Warner, M., and Guasch, L., 2016, Adaptive waveform inversion: Theory: Geophysics, **81**, R429–R445.