Regularization tunneling for full waveform inverstion

Scott Keating and Kris Innanen

ABSTRACT

Prior information can be a powerful tool in seismic inversion that can substantially improve on the results that seismic data alone can provide. Knowledge about clustering of rock physics properties may be especially significant, but this type of information is problematic in full waveform inversion due to the local optimization methods which are typically used. Here, we propose a regularization tunneling strategy for full waveform inversion, in which global regularization information is partially accounted for. By introducing the potential for elements of the subsurface model to tunnel between clusters, this approach is able to overcome obstacles associated with local minima in regularization terms from a priori data. We test a tunneling inversion approach on a simple synthetic problem, and find that when information about rock physics clustering is available, the proposed technique has the potential to better use this information than a conventional inversion strategy.

INTRODUCTION

Seismic inversion techniques, like full waveform inversion (FWI), are typically driven by objective function minimization (Tarantola, 1984). In these methods, an objective function is defined such that good models have small objective function values, while poor models have large objective function values. The problem of minimizing the objective function is then equivalent to finding the best model (by the metric of the chosen objective function). Generally, an objective function consists of two key parts. One part of the objective function rewards models which accurately reproduce the measured data. This term drives the inversion to use the measured data to improve the model. The other part of the objective function is the regularization term. This term rewards models which are consistent with our prior information about the subsurface. If the data-fit term were neglected in inversion, the inversion would learn nothing new about the subsurface, while if the regularization term were neglected, the inversion may produce a result consistent with the data considered, but not consistent with our other sources of information. Both terms play major roles in ensuring that a good inversion result is obtained (Tarantola, 2005).

While some inversion approaches are able to use global optimization techniques, the minimization of the objective function in FWI is usually restricted to local optimization due to the computational cost of the problem (Virieux and Operto, 2009). With this type of approach, the local behaviour of the objective function becomes very important. Because local optimization techniques exclusively consider update directions which locally reduce the objective function, certain types of model-space steps are discouraged. The data-fit term of objective functions will discourage model-space steps which locally decrease the model fit. Similarly, regularization terms will discourage steps which locally move away from regions of model space which are a priori likely. While these behaviours are highly desirable in the vicinity of the global minimum, they can also cause problematic convergence to local minima. In the cycle-skipping phenomenon, for instance, an inversion can be trapped at a local minimum because short steps towards the global minimum actually

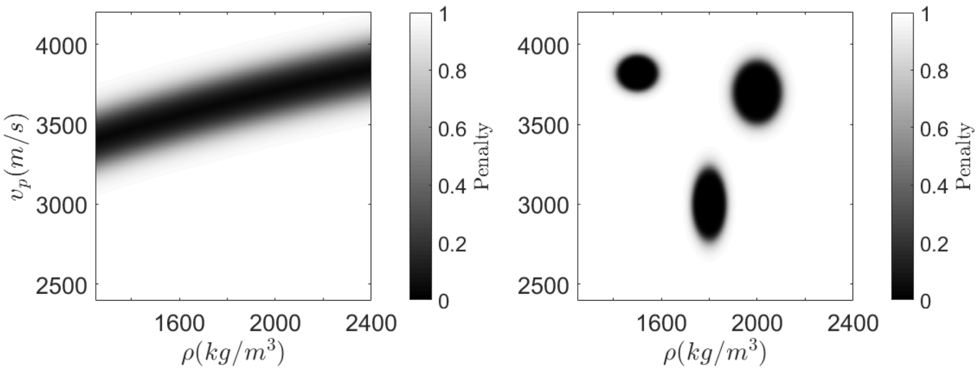

FIG. 1. Penalty term well suited to local optimization (left) and penalty term representing clustering (right). Note that several local minima exist in the clustering case.

decrease the data fit. These problems can be very challenging to solve in FWI.

One important source of a priori information in many geophysics problems is lithology. Based on geologic knowledge of a region, we may know that only certain types of rocks are likely to be found and, from study of those rock types, we may know which combinations of physical properties (density, bulk modulus, etc.) are feasible. If the feasible relations of physical properties form a single cluster, then the corresponding regularization term will usually be well suited for local minimization approaches. If, on the other hand, the possible physical properties form multiple, separated clusters, there will be several local minima in the regularization problem, which could seriously hamper the inversion. These two cases are illustrated in figure 1. The challenge in the multi-cluster case is that an element of the model that is initially classed in the wrong cluster will be strongly discouraged from changing clusters by the regularization term due to the local optimization techniques used.

To allow for a change of clusters in FWI, we propose a strategy for including non-local regularization information in the inversion. This means that instead of considering only the derivatives of the regularization term at a given point in model space, we will also consider the value of the regularization term at other points in model space. While this type of approach would be infeasible for the data-fit term of the objective function due to the computational cost of wavefield modeling, the regularization term will typically require no such modeling, and so can be evaluated at many models at negligible cost. Specifically, we propose here a regularization tunneling approach, wherein model elements which push strongly against the regularization terms in their local clusters are moved to other clusters, based on our global knowledge of the regularization function.

THEORY

The principle of the approach we suggest is relatively simple. First, we suggest that model elements which push only slightly against the regularization function likely do so because of minor disagreements between the local character of the data-fit function and the regularization term, and that these elements need no non-local correction. Model elements which push strongly against the regularization function, however, are likely elements that the data-fit is pushing towards another cluster of the regularization term. If this is the case, these elements should be allowed to “tunnel” directly from their current location in model space to an acceptable region in the other cluster.

For the tunneling approach we propose, we assume that the regularization function (or at least the key part of it for our tunneling approach) is defined as a sum over elements of the finite-difference model. This function should take all of the physical properties estimated at a given element of the model, and assign a regularization penalty term to the element based on how well or poorly it corresponds to expected rock physics relations. With this type of regularization, the algorithm we propose is as follows. At each iteration of the inversion, a local optimization strategy (conjugate gradient, L-BFGS, truncated Newton, etc.) is used to calculate a descent direction, and the model is updated by choosing an appropriate step length in this direction. In a conventional local optimization, the next iteration would begin after the model was updated. In our approach, we instead follow each model update with a tunneling step. In the tunneling step, we calculate for each element a *potential*, which quantifies how strongly the element pushes against the regularization, and a *momentum*, which estimates the model-space direction the element would move in if there were no regularization. The calculated potential and momentum are then fed into a *tunneling probability function*, which assigns a probability that the element is tunneled by having its physical properties changed to match those of an acceptable region in another cluster. The tunneling probability function should be large if the potential is high and the momentum points near another acceptable region of model space, and should be small if only one or neither of these conditions are met. After the decision whether to tunnel for each element is made based on the tunneling probability function, the appropriate elements are tunneled, and the next iteration of the inversion begins.

Possible definitions of key functions

While the preceding section outlines the algorithm in principle, we have explained only the intent of the potential, momentum and tunneling probability function terms. The specific choices made for these terms may play a major role in determining how a regularization tunneling approach performs in practice. Accordingly, we provide some motivations for specific choices of these terms in this section.

Potential

Perhaps the most natural choice for the potential term is simply the regularization penalty incurred by each model element. As this penalty term is designed to push the model toward acceptable regions of model space, the larger this term is, the more strongly the data must be pushing toward other solutions. While this penalty term may be problematic in very specific cases, we expect that it will generally work as an acceptable potential.

Momentum direction

This term should be designed to estimate the direction that the element would have been moved in by the inversion if no regularization were considered. A very simple approach to calculating this term is to re-compute the descent direction, omitting the contribution of the regularization. This should result in a viable momentum, providing useful information about the model-space direction the data suggest. This approach does, however, have some potential drawbacks. While this recalculation step should incur little additional computational or memory cost in some optimization approaches (e.g. conjugate gradient or L-BFGS), these are approaches in which Hessian information is either neglected or inferred from previous model iterates. The large, abrupt steps introduced by tunneling could greatly interfere with accurate Hessian estimation in these approaches. Other methods, like truncated Newton optimization, do not rely on previous iterates to provide Hessian information, but do incur a considerable computational cost in estimating the effect of the Hessian, which would have to be reduplicated in order to calculate a regularization-free descent direction. An alternate approach to determining a momentum would be to define it as the gradient of the regularization term with respect to the element's parameters. This calculation is simple, and naturally identifies the direction the data most push against the regularization. This direction may not be ideal, however, as this will generally be the direction the data would push if no regularization existed. In this report, we will consider a regularization-free truncated Gauss Newton step direction as the momentum. While this incurs extra computational cost, it should give us a good idea of how the approach performs in an ideal case.

Probability function

The tunneling probability function plays a major role in the proposed algorithm, using the potential and momentum to determine both whether a given element tunnels, as well as where it tunnels to. There are many different ways to formulate such a function, here we outline the one we consider in this report. First, we define a base probability of tunneling:

$$p_b = \gamma\psi, \tag{1}$$

where ψ is the potential, and γ is a constant scale factor chosen for the inversion. The base probability of tunneling, p_b is the probability that an element will tunnel if there is an acceptable region in another cluster in the direction of the momentum. While it would be possible to use p_b as the complete tunneling probability, this may be overly restrictive, as it may often be the case that the momentum does not point directly at another cluster. To help address this issue, we define the tunneling probability as

$$p(\hat{r}_\theta) = p_b f_\theta(\hat{r}_\theta, \hat{r}_m) g(\hat{r}_\theta), \tag{2}$$

where $\hat{r}_\theta$ is a unit vector, $\hat{r}_m$ is the unit vector for the momentum, f_θ is a function that is equal to 1 when $\hat{r}_\theta = \hat{r}_m$ and grows smaller as the angle between these directions grows, and g is a function that returns 1 if there an acceptable region of model space in the direction $\hat{r}_\theta$ and returns 0 otherwise. This function has the advantages of not requiring that the momentum point directly at another cluster, and providing greater probability for clusters that have a larger cross-sectional area from the element considered. After assembling $p(\hat{r}_\theta)$ for a representative sample of directions with $f_\theta(\hat{r}_\theta, \hat{r}_m) \neq 0$, we can compare it with a random number. If the number is greater than the sum of p, no tunneling occurs, otherwise, the element is changed to have the properties of a random, acceptable point in the the appropriate direction $\hat{r}_\theta$.

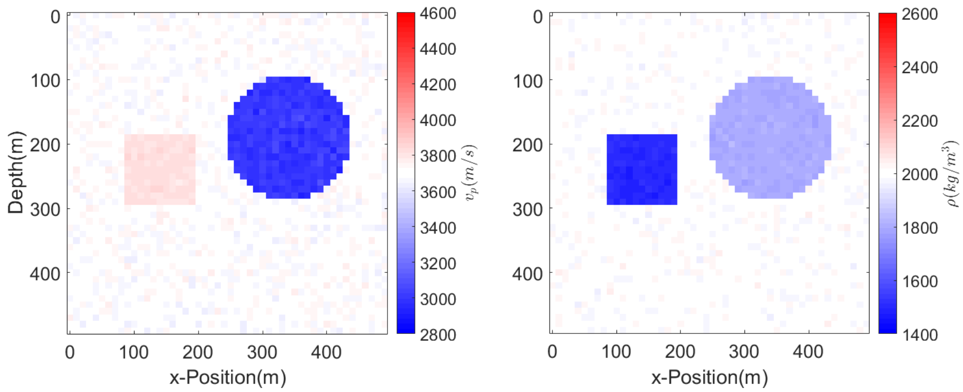

FIG. 2. True v_P and ρ used in the synthetic test.

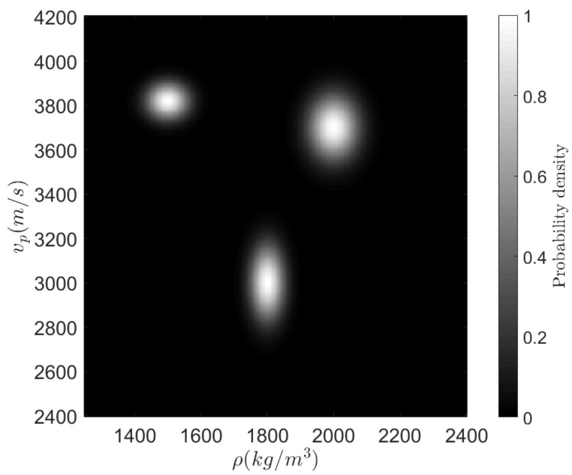

FIG. 3. Probability density as function of element v_P and ρ.

NUMERICAL EXAMPLE

In this section, we consider a simple synthetic example to illustrate the basic features of the approach. The true subsurface properties, used to generate the data, are shown in figure 2. This is a small two dimensional model, 500m by 500m in size, with a grid spacing of 10 m in both directions. Three different rock-types are present in this model, each with a different mean and standard deviation for P-wave velocity, v_P, and density, ρ. The probability distribution for v_P and ρ is shown in figure 3. While such specific information would be expected to greatly aid the inversion process, conventional FWI will struggle to cope with the clustered nature of the a priori information. We define a simple, sigmoidal regularization term, such that the half-maximum amplitude is reached where the probability density is 10% the maximum probability density. The resulting regularization, as a function of v_P and ρ, is shown in figure 4.

The starting model used for the inversion was a constant medium, corresponding to one of the rock physics clusters, with $v_P = 3700m/s$ and $\rho = 2000kg/m^3$. We consider

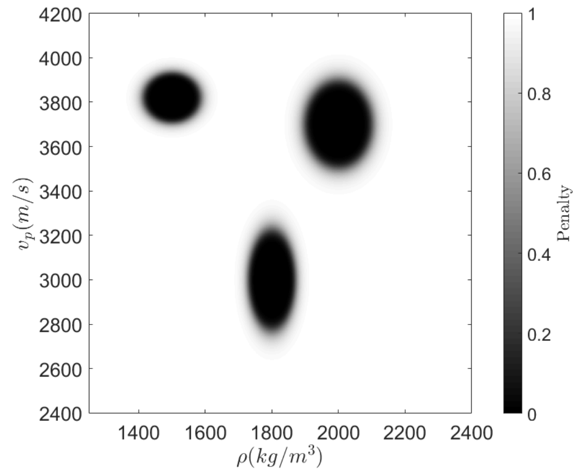

FIG. 4. Regularization penalty term as function of element v_P and ρ.

two possible acquisition geometries for the synthetic tests: a surface acquisition in which one line of sources and one line of receivers are located near the top of the model, and a more comprehensive acquisition, in which the surface acquisition is augmented by a line of receivers at the bottom of the model. In the inversion, we consider 20 frequency bands of four frequencies each. The frequencies in each band are linearly spaced from a low frequency of 1 Hz to a high frequency that begins at 2 Hz and increases with each FWI iteration to a maximum of 20 Hz. At each frequency band, one iteration of truncated Gauss Newton (TGN) optimization is used, with ten inner iterations per FWI iteration. In the tunneling approach, the inner loop is repeated at each iteration in order to calculate the momentum term.

The inversion results for the surface and comprehensive acquisition geometries are shown in figures 5 and 6, respectively. In both of these results, there is some FD cell-level noise; isolated cells have assumed very different values than the bulk of their neighbors. As this is an undesirable feature of the inversion result linked to the probabilistic nature of our tunneling approach, we apply an edge-preserving filter before further analyzing the results. The filter we use examines the seven by seven FD cell region around each element of the output model. If the element is in the same cluster as a chosen fraction of the elements in this region (we use a 40% threshold), then it is unchanged, otherwise it is moved to a random set of acceptable values in the majority cluster for this region if a majority exists, and to the background model otherwise. The filtered results are shown in figures 7 and 8 for the surface and comprehensive acquisitions, respectively. In the comprehensive acquisition case, the inversion result matches the true model very closely, recovering both anomalies effectively, and making minimal updates elsewhere in the model. In the surface acquisition case, the circular anomaly on the right side of the model is still recovered accurately, and the interior portion of the left anomaly is also correctly identified. In this case there are more spurious minor changes in the model, but overall the inversion result is a good estimate of the true model.

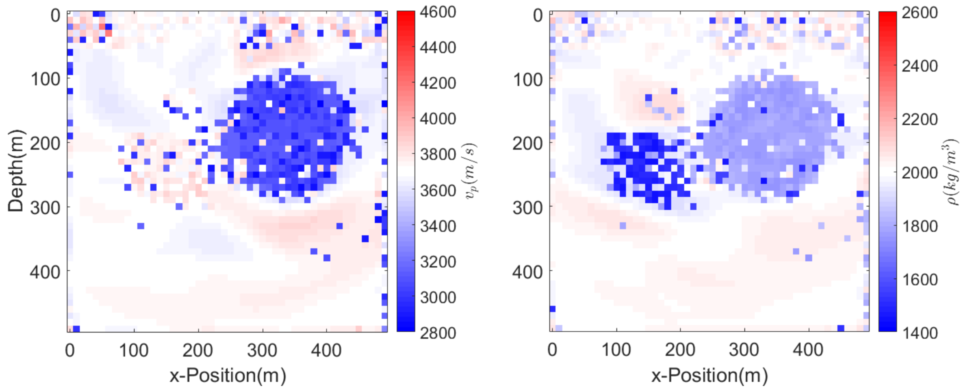

FIG. 5. Estimated v_P and ρ after inversion of surface-only data.

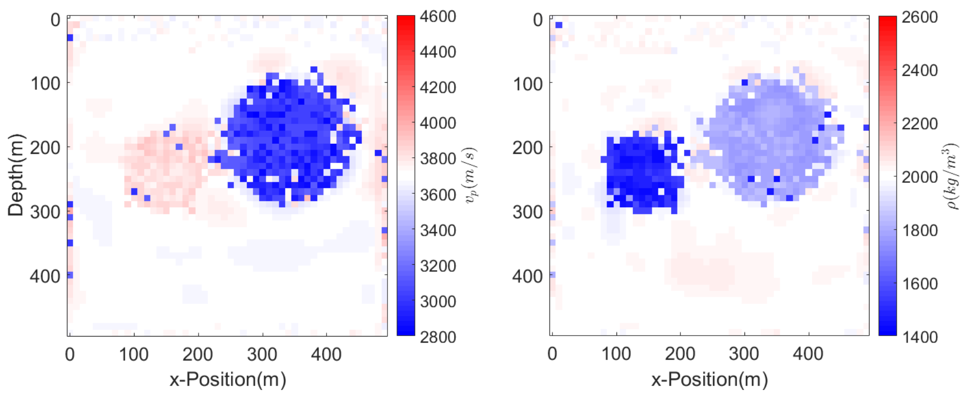

FIG. 6. Estimated v_P and ρ after inversion of comprehensive acquisition geometry data.

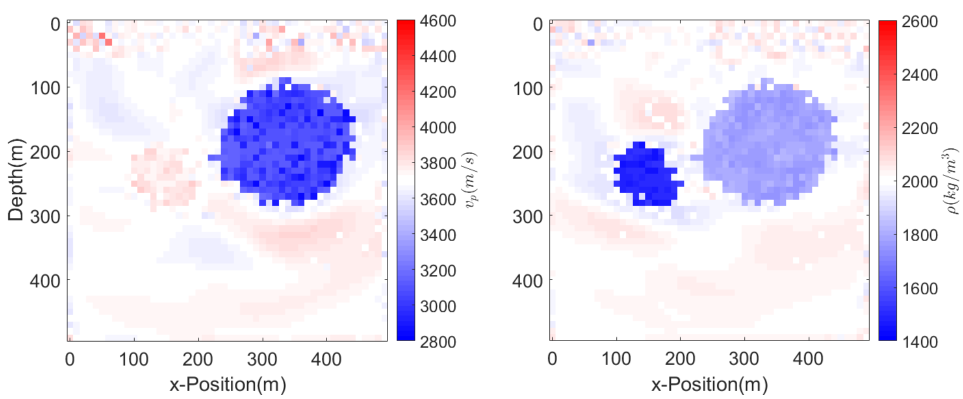

FIG. 7. Filtered surface-only inversion result. Compare with figures 5 and 2.

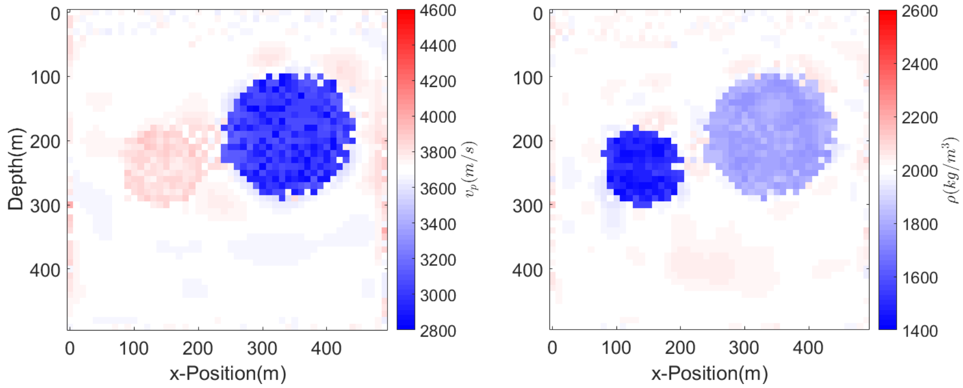

FIG. 8. Filtered comprehensive acquisition geometry inversion result. Compare with figures 6 and 2.

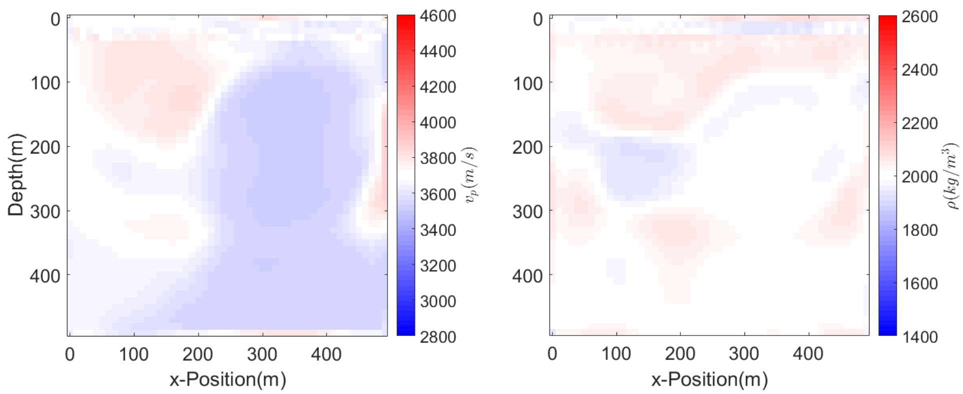

FIG. 9. Estimated v_P and ρ after inversion of comprehensive acquisition geometry data without tunneling. Compare with figures 8 and 2.

For comparison, we have also tested the same inversion problem with two conventional FWI strategies, using the comprehensive acquisition with no tunneling component. In the first approach, the regularization term is preserved. The result of this inversion is shown in figure 9. In this case, the inversion has moved both anomalies toward their true values, but has been prevented from large updates away from the starting cluster by the regularization term. Consequently, the inversion is held back by the regularization term in this example, and the inversion result is very poor. In a second approach, the regularization term is omitted in order to prevent the convergence problems. The result of this approach is shown in figure 10. In this case, the data are more free to improve the model, so there are significant improvements over figure 9. By completely discarding a priori information, however, this approach has failed to reproduce the true model as accurately as the tunneling approach, even in the case where the tunneling approach uses only surface data (figure 7).

DISCUSSION

While the filtered results shown in the previous section accurately estimate the true model, the need for filtering suggests that the current implementation of this approach could be improved. The cause of the noisy elements in these examples is likely the probabilistic approach to tunneling used here, which introduces a nonzero chance that an element

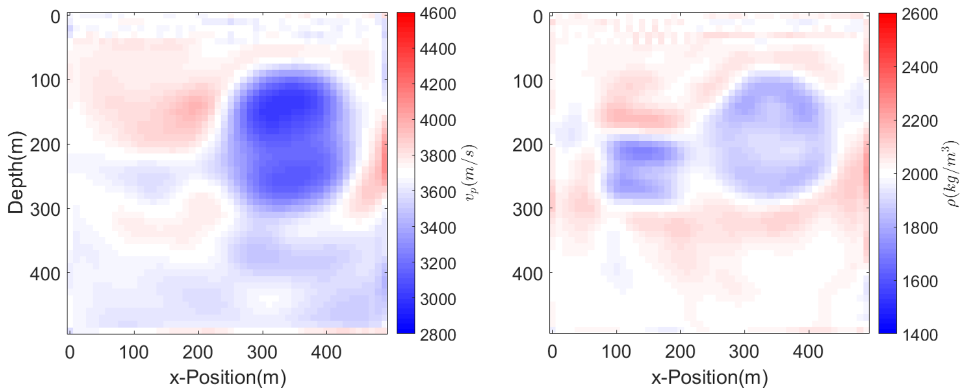

FIG. 10. Estimated v_P and ρ after inversion of comprehensive acquisition geometry data without tunneling or regularization. Compare with figures 8 and 2.

will tunnel to another cluster at any iteration. If one element jumps at a different iteration or to a different cluster than its neighbours, it may introduce the noisiness seen in these results. In principle, an element that tunnels to the wrong cluster should be capable of self-correction at later iterations, as the momentum term will continue to point at or near the correct cluster. In this case, however, the elements are defined at the smallest possible scale, while the momentum is calculated based a descent direction, and is fundamentally limited in wavelength by the data considered. This means that an element in the wrong cluster may not naturally self-correct, because the momentum calculated at the element is chiefly determined by the surrounding neighborhood as a whole. Prevention of the noisiness caused by this effect could be achieved by either moving to a more deterministic tunneling approach, or by defining model elements to be large enough to individually control the momentum in their region.

CONCLUSIONS

Prior information has the capability to dramatically improve seismic inversion results through regularization. Conventional FWI, using local optimization strategies, struggles to cope even with very simple regularization terms if they have multiple local minima. Information about different clusters of plausible rock physics relations is particularly problematic. In this report, we propose a regularization tunneling strategy, which introduces a global component to FWI based on knowledge of the regularization term. On simple synthetic tests, this tunneling approach allows for the inversion to make use of prior information which would not be usable in a conventional approach.

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 and CRDPJ 543578-19.

REFERENCES

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

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation, vol. 89: siam.

Virieux, J., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, No. 6, no. 6, WCC1–WCC26.