Optimization driven model-space versus data-space approaches to invert elastic data with the acoustic wave equation

Xiang Li, Anaïs Tamalet*, Tristan van Leeuwen*, Felix J.Herrmann* EOS-UBC*

SUMMARY

Inverting data with elastic phases using an acoustic wave equation can lead to erroneous results, especially when the number of iterations is too high, which may lead to over fitting the data. Several approaches have been proposed to address this issue. Most commonly, people apply "data-independent" filtering operations that are aimed to deemphasize the elastic phases in the data in favor of the acoustic phases. Examples of this approach are nested loops over offset range and Laplace parameters. In this paper, we discuss two complementary optimization-driven methods where the minimization process decides adaptively which of the data or model components are consistent with the objective. Specifically, we compare the Student's t misfit function as the data-space alternative and curvelet-domain sparsity promotion as the model-space alternative. Application of these two methods to a realistic synthetic lead to comparable results that we believe can be improved by combining these two methods.

TIME-HARMONIC ACOUSTIC FWI FORMULATION

Problem formulation

We consider the following frequency-domain formulation of the FWI problem:

$$\min_{\mathbf{m},\mathbf{w}} \phi(\mathbf{m},\mathbf{w}) = \sum_{i=1}^{K} \rho(\mathscr{B}_i(\mathbf{d}_i - w_i \mathscr{F}_i(\mathbf{m}))), \qquad (1)$$

where $\mathbf{d}_i$ is the observed data for one frequency and one source i, $\mathscr{F}_i(\mathbf{m})$ is the corresponding modelling operator, $\mathbf{w}$ are the source weights, $\mathbf{m}$ is the vector with the unknown medium parameters, $\mathscr{B}_i$ is a data-processing operator (to be discussed later), ρ is a penalty function and K is the batch size which is equal to n_s (number of shots) $\times n_f$ (number of frequencies).

For a given model $\mathbf{m}$, we estimate the source wavelet by solving for each source:

$$\underset{w_i}{\text{minimize}}\, \rho(\mathscr{B}_i(\mathbf{d}_i - w_i \mathscr{F}_i(\mathbf{m}))). \qquad (2)$$

This is a scalar optimization problem that can be solved efficiently in a number of ways. The objective function is now effectively a function of the model only: $\overline{\phi}(\mathbf{m}) = \phi(\mathbf{m}, \overline{\mathbf{w}})$ and the gradient of the reduced objective is given by $\nabla \overline{\phi}(\mathbf{m}) = \nabla_{\mathbf{m}} \phi(\mathbf{m}, \overline{\mathbf{w}})$. For more details on this variable projection approach, we refer to (Aravkin et al., 2012; Aravkin and van Leeuwen, 2012)

Inversion of elastic data

In this paper, we use the BG compass model depicted in Figure 3(a) as the true reference and carry out acoustic FWI on elastic data. The S-wave velocity for the elastic data is computed from the P-wave velocity using a fixed Poisson's ratio

equal to 0.25. As we can see in Figure 1(a) and Figure 1(b), the elastic data contain elastic phases that are not present in the acoustic data. These elastic phases are not modelled by the acoustic wave equation and will thus affect the recovery as we can see by comparing Figures 2(a) and 2(b).
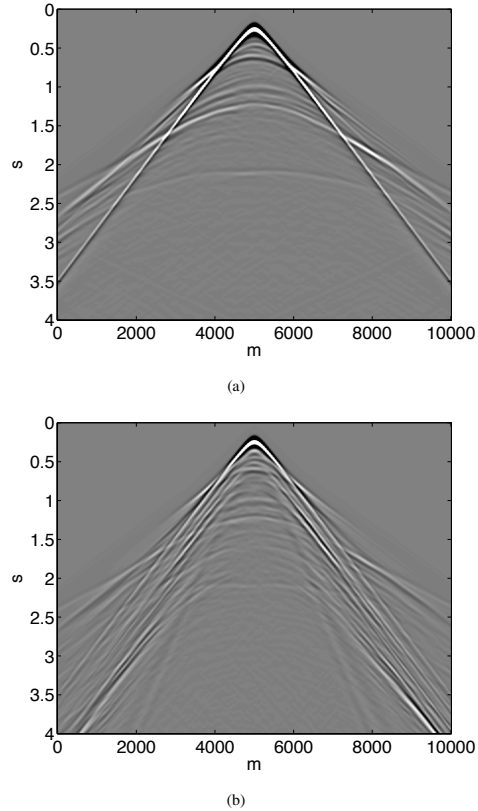

(a)

(b)

Figure 1: Time-domain data. **(a)** Acoustic data. **(b)** Elastic data.

Opposed to commonly used data-space continuation methods, which involve nested loops where Laplace parameters and offset ranges are chosen to first highlight the acoustic and then the elastic phases in the data (Virieux and Operto, 2009), we will consider two complementary optimization-driven approaches to mitigate the affects of ignoring elasticity in the forward model. In the first approach, the missing elastic phases are seen as outliers, whose adverse imprint on the inversion is controlled by using the Student's t misfit function. This type of data-misfit function is known to be insensitive to outliers. These artifacts are known to become more dominant if we run too many iterations, which will lead to an "over fitting" of the elastic phases. During the second approach, we control the affects of the unmodelled elastic phases by promoting sparsity in the curvelet domain. This latter type of regularization is known to penalize spurious artifacts. Before discussing the performance
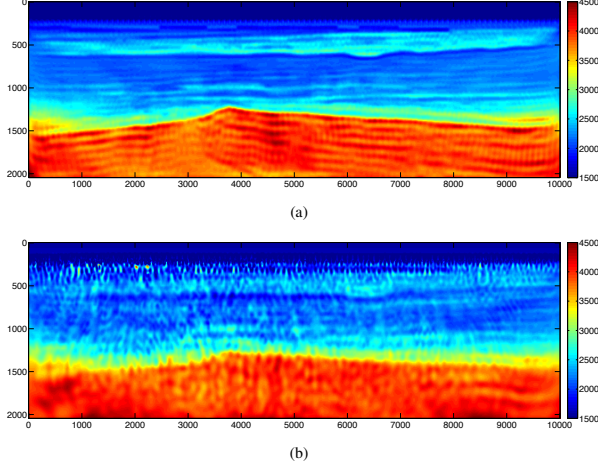
(a)

(b)

Figure 2: Acoustic FWI results **(a)** On acoustic data. **(b)** On elastic data.

of these two approaches, let us first briefly describe these two methods.

Data-space Student's t

As shown in (Aravkin et al., 2011), the Student's t penalty is given by

$$\rho(\mathbf{r}) = \sum_i \log(1 + |r_i|^2/k), \qquad (3)$$

where r_i is the ith datum of the residue vector given by the difference between the observed and modelled data, and k is the degree of freedom of the Student's t distribution.

This misfit function can be used in FWI in order to improve the recovery in the presence of large outliers or unexplained events in the data. Indeed, in contrast to previous robust penalties, the Student's t penalty is non-convex. Thus, large outliers are progressively down-weighted and effectively ignored once they are large enough.

However, for the Student's t penalty to be effective, the outliers must be somehow localized and distinguishable from the good data. Thus, following (van Leeuwen et al., 2013), we first transform the residual into a domain where the outliers are localized via the processing operator $\mathcal{B}_i$ before measuring the Student's t misfit. We can use any transform that would normally be used to filter out the noise (Fourier for periodic noise, Radon for noise with moveout, Curvelets for more complicated coherent events, etc.). Here, we choose to transform the residual in the source-receiver wave-number domain by applying the Fourier transform along both the sources and the receivers. The advantage of this approach is that the optimization with the Student's t misfit carries out the filtering adaptively rather than relying on some user-defined prior filter. So, we let the robust inversion process decide which parts of the data can be fitted and which should be ignored. Thus, the filtering is done implicitly as part of the inversion process.

Model-space sparsity promotion

Aside from controlling the inversion via the Student's t misfit function in the data space, another possible approach is to regularize updates in the model space. For this purpose, we define the data-misfit via the least-squares–i.e., $\rho(\mathbf{r}) = \sum_i |r_i|^2$, where r is the data residual vector as defined before. Contrary to Student's t, we impose ℓ_1-norm constraints on the curvelet representation of the Gauss-Newton updates (Li et al., 2012), i.e., we solve the following modified Gauss-Newton problems:

$$\delta\mathbf{m} = \mathbf{C}^{\mathbf{H}} \arg\min_{\mathbf{x}} \frac{1}{2}\|\underline{\delta\mathbf{D}} - \text{diag}(\underline{\mathbf{w}})\nabla\mathscr{F}[\mathbf{m_0}]\mathbf{C}^{\mathbf{H}}\mathbf{x}\|_F^2$$

$$\text{subject to} \quad \|x\|_1 \le \tau, \qquad (4)$$

for a series of increasing τ's. In this expression, the symbol $\mathbf{C}^H$ stands for the curvelet synthesis given by the adjoint, denoted by ^{H}, of the curvelet transform. The data-residue matrix is given by $\underline{\delta\mathbf{D}} = \underline{\mathbf{D}} - \text{diag}(\underline{\mathbf{w}})\mathscr{F}(\mathbf{m_0})$, with $\underline{\mathbf{D}} = [\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3, \cdots, \mathbf{d}_{K'}]$ the observed data for randomly selected sources and frequencies. Since $K' \ll Ks$, the evaluation of the Jacobian $\text{diag}(\underline{\mathbf{w}})\nabla\mathscr{F}[\mathbf{m_0}]$, scaled by the source weights $\underline{\mathbf{w}}$, becomes cheaper making the solution of Equation 4 computationally feasible (Herrmann et al., 2009, 2008; Li et al., 2012). Following the modified Gauss-Newton approach, we compute the ℓ_1-norm constraints via

$$\tau = \frac{\|\underline{\delta\mathbf{D}}\|_2}{\|\mathbf{C}\nabla\mathscr{F}^T(\mathbf{m_0})\text{diag}(\underline{\mathbf{w}})\underline{\delta\mathbf{D}}\|_\infty} \qquad (5)$$

(Berg and Friedlander, 2008).

EXPERIMENTS

To make a comparison between data-space Student's t and model-space sparsity promotion, we use the BG compass model in Figure 3(a) as the true reference, which is a synthetic P-wave velocity model created constrained by real well-log information. As we mentioned before, we compute the S-wave velocity (Figure 3(b)) using a fixed Poisson's ratio of 0.25. We simulate 101 shot records with a 100m interval using a time-domain elastic finite-difference modeling code (Thorbecke, 2013) with a 9Hz Ricker wavelet. All shots share the same 401 receivers with 25m interval yielding a 10km offset. Because we consider an ocean-bottom node survey with reciprocity, we set the source depth to 200m while receiver depth is set to 10m.

The modeling kernel used in the inversion is based on frequency domain acoustic modeling implemented by solving the Helmholtz system for a 9-point stencil (Jo et al., 1996). To define the starting model for the inversion, we smooth and average laterally the reference velocity model in Figure 3(a). The result is plotted in Figure 3(c).

To make a fair comparison, both inversions are carried out with multiple frequency bands from $3 - 20$Hz. Each frequency band contains 3 frequencies. For data-space Student's t, we compute 5 and 10 l-BFGS updates with all the sources while for model-space sparsity promoting, we solve 5 sparsity-promoting GN subproblems with 50 randomly selected shots. For 5 iterations of l-BFGS, the total number of PDE solves is roughly the same for these two approaches. We varied the number of l-BFGS iterations to show the effect of overfitting (juxtapose Figures 4(a) and 4(b)).

As we have seen from the motivating example in the introduction, carrying out acoustic FWI on data that contain elas-
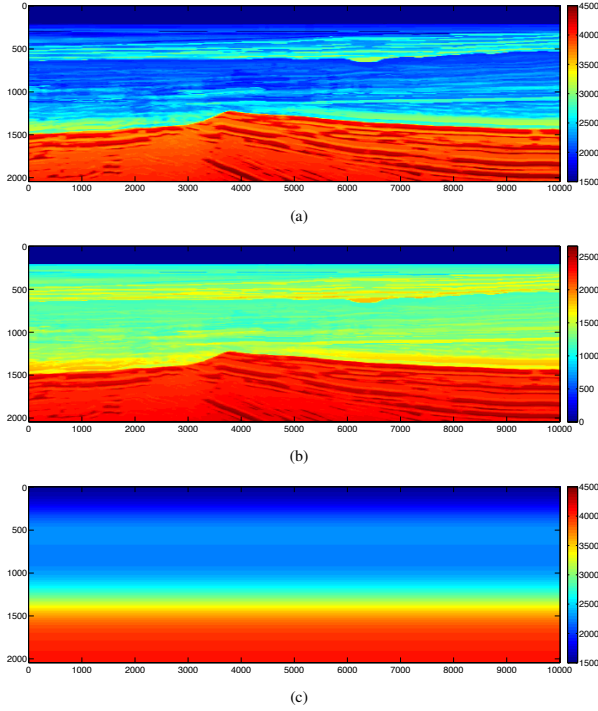
(a)

(b)

(c)

Figure 3: BG compass model. **(a)** True P-wave model. **(b)** True S-wave model. **(c)** Initial model.

(a)

(b)

(c)

Figure 4: l-BFGS inversion results. **(a)** Least-Squares penalty with 5 iterations. **(b)** Least-Squares penalty with 10 iterations. **(c)** Student's t penalty with 10 iterations.

tic phases can be detrimental to the inversion. Using the robust Student's t technique, we hope to mitigate some of the adverse affects of these elastic phases by using the fact that monochromatic frequency slices are relatively sparse in the source-receiver wave-number domain. This means that the elastic events that are not in the range of the acoustic modeling operator tend to be concentrated amongst a limited number of frequencies. The Student's t penalty function is by design relatively insensitive to these outlliers. which should improve the inversion results. As we can see from Figure 4(c), this is indeed the case compared to the inversion result based on the ℓ_2 misfit for 10 iterations included in Figure 4(b). While this result may not be as good as the inversion result for the acoustic only result, our improvement is certainly encouraging. Aside from changing the data-space misfit penalty functional, model-space regularization is another proven tool in inversion problems that suffer from null spaces or from unmodelled noise. Transform-domain sparsity promotion, including one-norm minimization in the curvelet domain, has proven to be an effective tool to remove these unmodelled signal components. For this purpose, we submit the data set with the elastic phases to the above described modified Gauss-Newton method. The results of this exercise are summarized in Figure 5 and were obtained by 5 GN iterations on randomized subsets of 50 shots. From this figure, it is clear that invoking curvelet-domain sparsity promotion on the GN updates also mitigates the adverse affects on the unmodelled elastic phases present in the data. While the Student's t approach works by ignoring the outliers, the curvelet-domain sparsity promotion penalizes complexity on the model that is associated with the unmodelled elastic phases.
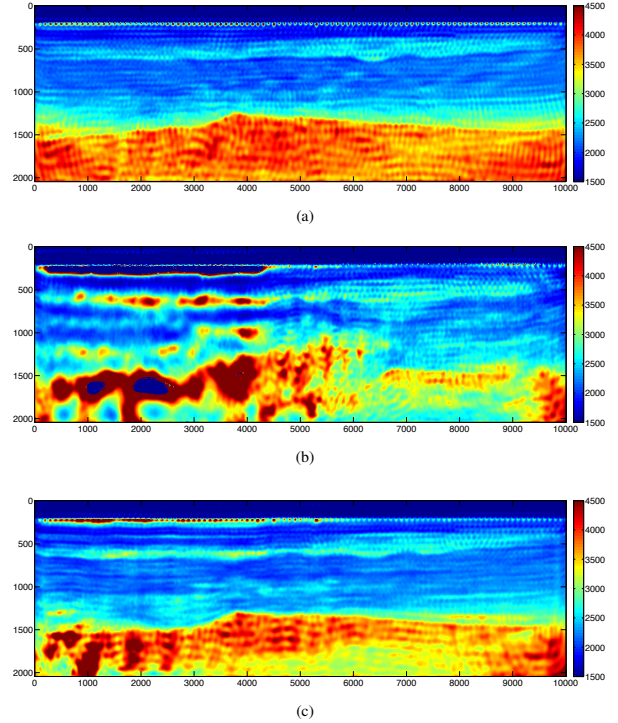
As with the Student's t, the result of this method is encouraging.

DISCUSSION AND CONCLUSIONS

While the results of the two different methods are encouraging, there is still room for improvement. Before outlining a possible road ahead let us first systematically compare the results of the previous two sections. Upon close inspection, the result obtained with the Student's t penalty function has more details compared to the results obtained by GN but also more artifacts related to the presence of elastic phases. When we compare vanilla l-BFGS (Figure 4(b)) without Student's t with the vanilla GN (Figure 5(a)) without curvelet-domain promotion, we observe that the result for GN are apparently less sensitive to unmodelled phases in the data.

From the computational perspective, the Student's t approach is relatively simple since it relies on more-or-less straightforward changes on the definition of the penalty function itself (compared the commonly used ℓ_2 penalty), its gradient, and the estimation of the Student's t parameter by variable projection. The modified Gauss-Newton approach, on the other hand, relies on a somewhat more complicated machinery that includes the curvelet transform itself and a method to choose the ℓ_1-norm constraint for each GN subproblem. The advantage of the GN approach, however, is that it does not rely on computing the gradients for all shots as required by the Student's t method,
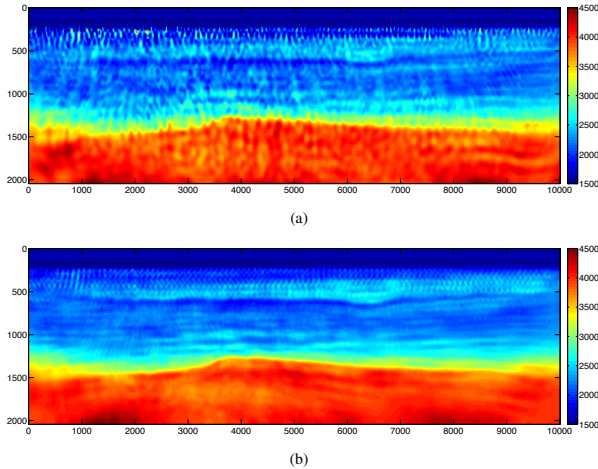
(a)

(b)

Figure 5: Gauss-Newton inversion inversion results. **(a)** GN without sparsity promoting. **(b)** GN with sparsity promoting.

which relies on Fourier transformation along the fully-sampled source and receiver coordinates. Moreover, elastic phases represented as monochromatic frequency slices may not be optimally sparse in the Fourier domain, which jeopardizes our ability to separate the acoustic from the elastic phases using the Student's t. This failure to completely separate the acoustic and elastic phases is properly responsible for the remaining artifacts for the Student's t (cf. Figure 4(c)). The GN is more artifact free but this feature goes at the expense of loss of detail, which is related to the one-norm constraint and perhaps the number of iterations. If we relax the ℓ_1-norm constraint too much, we run the risk of starting to "over fit" the elastic events.

As for the GN method, comparison between the vanilla l-BFGS and GN shows that the latter is apparently less sensitive to unmodelled phases in the data. There may be two possible explanations for this observation. First, the GN updates are regularized because we invert the GN Hessian with a limited number iterations, which corresponds to some kind of model-space regularization. The l-BFGS, on the other hand, is not regularized. Second, l-BFGS attempts to invert the true Hessian via a low-rank approximation while the GN updates are obtained by (approximate) inversion of the GN Hessian formed by the composition of the adjoint of the Jacobian acting on the Jacobian itself. In cases where the modeling operator does not explain all the events in the data, it may not be reasonable to expect that l-BFGS will yield the correct inverse of the Hessian while we can argue that the GN approach will at least get the GN part of the Hessian correct if the velocity model is reasonably accurate.

From our perspective, the results outlined in this abstract are both encouraging and call for a joint formulation based on data-space robust statistics in some transform-domain for the penalty function and on model-space regularization, for instance via curvelet-domain sparsity promotion, on the model updates. The modified GN framework allows for this type of formulation since it offers flexibility regarding the choice of the misfit function and norm on the model. In this way, we anticipate that the

ℓ_1-norm constraint can be relaxed without running the risk of fitting the elastic events, which are, by virtue of the Student's t penalty, discarded.

ACKNOWLEDGMENTS

REFERENCES

Aravkin, A., T. van Leeuwen, and F. Herrmann, 2011, Robust full-waveform inversion using the student's t-distribution: SEG Technical Program Expanded Abstracts, **30**, 2669–2673.

Aravkin, A. Y., and T. van Leeuwen, 2012, Estimating nuisance parameters in inverse problems: Inverse Problems, **28**, 115016.

Aravkin, A. Y., T. van Leeuwen, H. Calandra, and F. J. Herrmann, 2012, Source estimation for frequency-domain fwi with robust penalties: Presented at the EAGE technical program, EAGE, EAGE.

Berg, E. V. D., and M. P. Friedlander, 2008, Probing the Pareto frontier for basis pursuit solutions: Siam J. Sci. Comput., **31**, 890–912.

Herrmann, F. J., C. R. Brown, Y. A. Erlangga, and P. P. Moghaddam, 2009, Curvelet-based migration preconditioning and scaling: Geophysics, **74**, A41.

Herrmann, F. J., P. P. Moghaddam, and C. C. Stolk, 2008, Sparsity- and continuity-promoting seismic imaging with curvelet frames: Journal of Applied and Computational Harmonic Analysis, **24**, 150–173. (doi:10.1016/j.acha.2007.06.007).

Jo, C. H., C. Shin, and J. H. Suh, 1996, An optimal 9-point, finite-difference, frequency-space, 2-D scalar wave extrapolator: Geophysics, **61**, 529–537.

Li, X., A. Y. Aravkin, T. van Leeuwen, and F. J. Herrmann, 2012, Fast randomized full-waveform inversion with compressive sensing: Geophysics, **77**, A13–A17.

Thorbecke, J., 2013, 2d acoustic/visco-elastic finite difference wavefield modeling.

van Leeuwen, T., A. Y. Aravkin, H. Calandra, and F. J. Herrmann, 2013, In which domain should we measure the misfit for robust full waveform inversion?

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.