submitted to *Geophys. J. Int.*

Mitigating local minima in full-waveform inversion by expanding the search space

Tristan van Leeuwen and Felix J. Herrmann

Dept. of Earth, Ocean and Atmospheric Sciences, University of British Columbia.

Wave-equation based inversions, such as full-waveform inversion, are challenging because of their computational costs, memory requirements, and reliance on accurate initial models. To confront these issues, we propose a novel formulation of full-waveform inversion based on a penalty method. In this formulation, the objective function consists of a data-misfit term and a penalty term which measures how accurately the wavefields satisfy the wave-equation. Because we carry out the inversion over a larger search space, including both the model and synthetic wavefields, our approach suffers less from local minima. Our main contribution is the development of an efficient optimization scheme that avoids having to store and update the wavefields by explicit elimination. Compared to existing optimization strategies for full-waveform inversion, our method differers in two main aspects; *i)* The wavefields are solved from an augmented wave-equation, where the solution is forced to solve the wave-equation *and* fit the observed data, *ii)* no adjoint wavefields are required to update the model, which leads to significant computational savings. We demonstrate the validity of our approach by carefully selected examples and discuss possible extensions and future research.

1 INTRODUCTION

Detailed estimates of subsurface properties can be obtained from seismic data by solving a data-fitting procedure constrained by the wave-equation. The overal optimization problem consists of finding a subsurface model and wavefields (one for each source experiment) that minimize the data-misfit *and* solve the wave-equation for all the sources. Unfortunately, such a *full-space* approach that optimizes over both model and wavefields is not feasible for large-scale seismic problems because it requires storing and updating the wavefields for all sources. Instead, the conventional approach to full-waveform inversion (FWI) is based on a *reduced-space* approach, whose basic workflow consists of four steps (Tarantola & Valette 1982); *i)* solve the wave-equations for all sources to predict the data,

ii) calculate the misfit between observed and predicted data, *iii)* solve the adjoint wave-equations using the data-residual as source-term and *iv)* correlate the two wavefields for each source and stack over the sources in order to obtain a model update. We may view this approach as a special solution technique of the optimization problem where the constraint (the wave-equation) is satisfied at every iteration. Because steps *i)* to *ii)* can be done independently in parallel, this leads to an efficient implementation. However, it also makes the misfit depend on the model parameters in a very non-linear way. It is this non-linearity that causes problems in the form of local minima. Research aimed at mitigating these local minima mainly focusses on continuation from low to high frequencies (Bunks 1995; Sirgue & Pratt 2004), near to far offsets (Virieux & Operto 2009) or small to large Laplace damping parameters (Shin & Cha 2009) and different misfit functionals (Cara & Lévêque 1987; Luo 1991; van Leeuwen & Mulder 2010; Bozda et al. 2011; Moghaddam & Mulder 2012). We refer to Virieux & Operto (2009) for an extensive overview of current approaches to FWI in exploration seismology. Other approaches are based on advanced approaches to migration velocity analysis and typically entail a significant increase in computational costs (de Hoop et al. 2006; Shen & Symes 2008; Weibull et al. 2012). The goal of these approaches is to provide a suitable initial model for FWI that prevents loop-skipping. See Symes (2008) for an extensive overview of state-of-the art approaches to migration velocity analysis and their relation to FWI.

In this paper, we propose an alternative to the conventional FWI formulation that exploits a larger search-space by optimizing over both the wavefields and the model. With this enlarged search space there is more freedom for the optimization to find a solution, which may alleviate problems related to local minima. Our main contribution is an efficient algorithm that explores the full search space without the need of having to store and update the wavefields for all the sources.

We base our method on the observation that the wave equation $A(\mathbf{m})\mathbf{u} = \mathbf{q}$ defines a relation between the medium parameter $\mathbf{m}$, the wavefield $\mathbf{u}$, and the source $\mathbf{q}$. While this relation is typically used to solve for a wavefield given the medium parameters and the source, we could alternatively calculate the medium parameters if we were given the source and the true wavefield everywhere. Of course, we usually do not measure this wavefield everywhere. The question then is, can we reconstruct this wavefield from its measurements at receiver locations and subsequently use this reconstructed wavefield to estimate the medium parameters? If we exploit the fact that we know the observed wavefield at the receiver locations, we can formulate a *constrained* wave-equation that forces the wavefield to obey the wave equation for the current model *and* fit the observed data. Given the solution of the data-constrained wave equation, we update the model parameters and use these updated parameters to repeat this process in an alternating fashion.

It turns out that we can formalize this approach by reformulating the original *full-space* approach,

which defines the optimization problem over both model parameters and wavefields. This novel formulation of full-waveform inversion includes the wave-equation constraint as an additive regularization term. We introduce a trade-off parameter to control the trade-off between fitting the observed data and solving the wave-equation. This makes sense because it is not necessarily useful to solve the wave equation accurately for the wrong model parameters. As such, we arrive at an efficient two-step optimization scheme where *i)* for each source the data-constrained wave equation is solved for a given trade-off parameter, followed by *ii)* the computation of a model update from the resulting wavefields and their corresponding sources.

Compared to traditional methods, our alternating approach has two major advantages. First, there is no need introduce adjoint or reverse-time wavefields. Second, each step is linear or only mildly nonlinear in the unknowns. This latter property, in conjunction with the large search space for the optimization, may lead to a formulation that is less prone to local minima.

The outline of the paper is as follows. First, we review the basic formulation of waveform inversion as a constrained optimization problem (section 2). Next, we introduce the penalty formulation (section 3) and derive the efficient optimization strategy in detail. Finally, we present a few insightful examples (section 4) to illustrate the new approach and conclude the paper with a discussion (section 5) on future work. Throughout the paper we formulate the inversion in the frequency-demain, as this greatly simplifies the derivation. A discussion on how to extend these ideas to a time-domain formulation is given the last section.

2 PDE-CONSTRAINED OPTIMIZATION

Full-waveform inversion in the frequency domain can be formulated as the following PDE-constrained optimization problem:

$$\min_{\mathbf{m},\mathbf{u}} \sum_{i=1}^{M} \tfrac{1}{2}\|P_i\mathbf{u}_i - \mathbf{d}_i\|_2^2 \quad \text{s.t.} \quad A_i(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i, \tag{1}$$

where $\mathbf{m}$ is a vector with gridded medium parameters; i is the experiment index, running over sources and frequencies; $\mathbf{d}_i$ are the observed data; $A_i(\mathbf{m}) = \omega_i^2\mathsf{diag}(\mathbf{m})+L$ is the descretized Helmholtz operator with ω_i the angular frequency of the i^{th} experiment and L the discretized Laplacian; $\mathbf{q}_i$ represents the sources; $\mathbf{u} = [\mathbf{u}_1;\mathbf{u}_2;\ldots]$ are the correpsonding wavefields; and P_i is the detection operator extracting data at the receivers associated with the i^{th} experiment.

An elegant way to solve these constrained optimization problems is via the following Lagrangian (Haber et al. 2000):

$$\min_{\mathbf{m},\mathbf{u},\mathbf{v}} \mathcal{L}(\mathbf{m},\mathbf{u},\mathbf{v}) = \sum_{i=1}^{M} \tfrac{1}{2}\|P_i\mathbf{u}_i - \mathbf{d}_i\|_2^2 + \mathbf{v}_i^* \left(A_i(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i\right), \tag{2}$$

where $\mathbf{v} = [\mathbf{v}_1; \mathbf{v}_2; \ldots]$ are the Lagrange multipliers and the symbol * refers to the adjoint. For FWI, these multipliers can be identified as the adjoint wavefields. Iterative solution techniques of this optimization problem update the model and solve the forward and adjoint wave-equations simultaneously, thus eliminating the need to explicitly solve the wave-equation at each iteration of the optimization. Moreover, the objective $\mathcal{L}(\mathbf{m}, \mathbf{u}, \mathbf{v})$ is linear in any one of its variables if the other two are fixed. The only non-linearity stems from the inner product $\mathbf{v}_i^* (A_i(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i)$, which is fairly benign because it does not involve the solution of the wave equation A_i^{-1}. Because the constrained optimization explores a much larger search space, we argue that derivative-driven optimizations of the constrained formulation are less likely to be affected by local minima. Despite this clear advantage, the constrained formulation is unfortunately not widely adapted because it is unfeasible to store and update all $2M$ wavefields as part of an iterative procedure (Epanomeritakis et al. 2008).

For this reason, one usually eliminates the constraints in Equation (1), yielding the following reduced-space formulation:

$$\min_{\mathbf{m}} \phi_{\text{reduced}}(\mathbf{m}) = \sum_{i=1}^{M} \tfrac{1}{2} \|P_i A_i(\mathbf{m})^{-1} \mathbf{q}_i - \mathbf{d}_i\|_2^2. \tag{3}$$

The gradient of the reduced objective is given by

$$\nabla_{\mathbf{m}} \phi_{\text{reduced}} = \sum_{i=1}^{M} \omega_i^2 \mathsf{diag}(\mathbf{u}_i)^* \mathbf{v}_i. \tag{4}$$

To compute this gradient, we need explicit solves of both the forward $A_i(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$ and adjoint wave equation $A_i(\mathbf{m})^* \mathbf{v}_i = P_i^* (P_i \mathbf{u}_i - \mathbf{d}_i)$ for all M experiments at each iteration of the optimization. To make matters worse, the action of the Hessian requires another three wave-equation solves for each experiment (Pratt et al. 1998). However, compared to the constrained method, the wavefields, gradient, and Hessian can be computed in parallel without the need to store all $2M$ wavefields at any time. While this reduced formulation eliminates the need to store and update the wavefields, its objective is very non-linear as a function of $\mathbf{m}$ because its dependency involves the solution of a wave equation. So the question arises: can we come up with a formulation that combines advantages of both formulations?

3 THE PENALTY METHOD

To arrive at a method that has the best of both worlds, we cast the constrained formulation (cf. Equation (1)) into the following form:

$$\min_{\mathbf{m},\mathbf{u}} \phi_\lambda(\mathbf{m}, \mathbf{u}) = \sum_{i=1}^{M} \tfrac{1}{2} \|P_i \mathbf{u}_i - \mathbf{d}_i\|_2^2 + \tfrac{\lambda^2}{2} \|A_i(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i\|_2^2. \tag{5}$$

The solution of this problem coindices with that of the constrained problem when $\lambda \uparrow \infty$ (Bertsekas 1996). We propose an alternating optimization method that solves this optimization problem while avoiding storage of the wavefields. For this purpose, we first minimize ϕ_λ with respect to $\mathbf{u}$ by solving $\nabla_{\mathbf{u}}\phi_\lambda = 0$. This gives rise to the following overdetermined sparse linear system for each source, which we call the *augmented wave-equation*

$$\begin{pmatrix} \lambda A_i(\mathbf{m}) \\ P_i \end{pmatrix} \mathbf{u}_i = \begin{pmatrix} \lambda \mathbf{q}_i \\ \mathbf{d}_i \end{pmatrix}. \tag{6}$$

This augmented system can be interpreted as a regularized version of the original wave-equation where the solution is forced to not only adhere to the wave equation, but also to fit the observed data, which is related to the solution through the restriction operator. As with the Helmholtz equation, we can solve this system efficiently in 2D via factorization techniques or in 3D using a (preconditioned) iterative solver. Given the wavefields that solve the augmented wave-equation (Equation (6)), we compute the new model by minimizing ϕ_λ with respect to $\mathbf{m}$. This can be done by solving $\nabla_{\mathbf{m}}\phi_\lambda = 0$, where

$$\nabla_{\mathbf{m}}\phi_\lambda(\mathbf{m}, \mathbf{u}) = \sum_{i=1}^{M} \omega_i^2 \mathsf{diag}(\mathbf{u}_i)^* \left(A_i(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i\right). \tag{7}$$

Note that the expression for this gradient is very similar to that of the reduced objective (cf. Equation (4)), except that is does not require an adjoint wavevield. The system of equations $\nabla_{\mathbf{m}}\phi_\lambda = 0$ can be re-written as:

$$\sum_{i=1}^{M} \omega_i^2 \mathsf{diag}(\mathbf{u}_i)^* \mathsf{diag}(\mathbf{u}_i)\mathbf{m} = \sum_{i=1}^{M} \mathsf{diag}(\mathbf{u}_i)^* \left(\mathbf{q}_i - L\mathbf{u}_i\right), \tag{8}$$

By assembling the terms in a running sum, we can form this system of equations without storing all the wavefields. Aside from this important feature, which allows us to compute and accumulate the wavefields independently, the above system is strictly diagonal so it is trivial to calculate the new model from Equation (8). Moreover, the above expression corresponds to a Newton step where the diagonal matrix on the left can be considered as the Hessian of ϕ_λ with respect to $\mathbf{m}$.

In summary, our approach combines advantages of both the Lagrangian and reduced approaches by exploiting the full search space while avoiding storage of all the wavefields. In addition, the method explicitly relaxes the constraint by adding a data-misfit regularization to the wave equation. Next, we consider some illustrative examples.

4 EXAMPLES

These examples serve as a proof of concept of our penalty method and are based on a relatively simple 5-point discretization of the 2D acoustic constant-density Helmholtz equation. To allow people

to evaluate the novel aspects of the proposed method, we have made the code available at `https://github.com/slimgroup/PenaltyMethodGJI` to reproduce the examples below.

4.1 Example 1

To illustrate the information in the solution of the data-augmented wave-equation, we consider a square perturbation embedded in constant background velocity model (2000 m/s), see Figure 1 (a). To mimic a transmission cross-well experiment, we place 51 sources and receivers at either side of the model. The corresponding scattered wavefield (i.e., the difference between the wavefield for the perturbed medium and the background medium) at 10 Hz for a single source located at $x = 10$m, $z = 500$m is plotted in Figure 1 (b). The goal of solving the augmented wave-equation (6) is to reconstruct this wavefield from observations at receivers along $x = 990$m, given the same background velocity model. The resulting wavefield is plotted in figure 2 (a). Because we included the data-constraint in the wave equation, we are able to partly reconstruct the imprint of the perturbation observed in of the original scattered wavefield. We use this information subsequently to calculate a new estimate for the model by solving Equation (8). The result of this exercise is plotted in Figure 2 (b). Repeating this procedure leads to a better reconstruction of the scattered wavefield, which is also observed in the corresponding model estimate as can be seen in Figure 4. For these last examples the region around the receivers was windowed out to suppress artifacts in the model.

4.2 Example 2

To support our claim that the enlarged search space of the penalty method may alleviate problems related to local minima, we conduct an experiment where the unknown 1D velocity profile $v(z) = v_0 + \alpha z$ is parameterized by only two parameters. We plot the objective functions corresponding to the reduced (cf. Equation (3)) and penalty approaches (cf. Equation (5)) as a function of v_0, α for various values of λ. As we can see from Figures 4.2 (a,b), the objective functions for a small λ exhibit no local minima. Moreover, we observe that (i) the global minima for the reduced and the penalty formulation for different values of the control parameters coincide and (ii) that the behavior of the objective function for the penalty method converges to the one of the reduced method as λ increases. Effectiveily, the parameter λ controls the width of the basic of attraction of the penalty objective.

4.3 Example 3

To demonstrate convergence behavior of the penalty method compared to the reduced method, we invert single source, single frequency data for a 1D gridded velocity profile $v(z) = v_0 + \alpha z + \delta v \exp(-\beta^2(z-$

$z_0)^2$) for values $v_0 = 2000$ m/s, $\alpha = 0.7$ 1/s, $\delta v = 200$ m/s ,$z_0 = 2000$ m and $\beta = 10^{-3}$. We use a maximum offset of 3000 m, a frequency of 5 Hz and a gridspacing of 50 m. The initial model is a linear gradient with $v_0 = 2000$ m/s and $\alpha = 0.7$ 1/s. To compare both methods on equal footing, we use a steepest-descent method with a fixed step length. Convergence plots for the value of the objective function and the two-norm of the gradient are included in Figure 5. These results show that the residue of the penalty method decays faster and that the reduced method stalls after about 30 iterations. A comparison of the forward modeled data for the inversion results for both methods is included in Figure 6 (a). This suggests that the reduced method stalled because of a loopskip around receiver 201. Conversely, the penalty method is not loop skipped and explains the data perfectly. While the data-fit may suggest that the reduced method also yields a reasonable result, the corresponding estimate for the velocity is completely wrong as can be seen in Figure 7. These results indicate that the penalty method suffers less from local minima and performs well for an initial model that is too inaccurate for the conventional formulation. Remember, we used a relatively simple optimization method, which explains the large number of iterations. In practice, we would use a more efficient optimization method to speed up the convergence.

4.4 Example 4

To show that the penalty formulation can also be used to image reflected data, we consider a medium with a constant background velocity and three perturbations as depicted in Figure 8. We use this model to generate data for 101 equi-spaced sources and receivers located at the top of the model and frequencies $1, 2, \ldots, 10$ Hz. A so-called reverse-time migration can be obtained by computing the gradient of the reduced objective Lailly (1983), requiring the computation of both forward and adjoint wavefields, followed by a correlation. The resulting image is shown in Figure 8 (b). Likewise, the gradient of the penalty objective (Equation (7)) can be used to obtain an image. However, this does not require the computation of an adjoint wavefield. Moreover, we can easily compenstate for amplitude effects by solving Equation (8) instead of using just the gradient. This requires almost no additional computations as it entails only a scaling with the norm of the wavefield. The resulting image is shown in Figure 8 (c). Aside from requiring only half of the number of wave-equation solves, the image derived from the penalty method is better resolved. The fact that Equation (8) corresponds to a Newton step explains the apparent improvement.

5 CONCLUSIONS AND DISCUSSION

By recasting the constrained formulation of full-waveform inversion into a penalty formation, we arrive at a novel formulation of full-waveform inversion where the misfit consists of two terms. The first term measures the data-misfit term and the second enforces the wave equation. Both terms are balanced by a trade-off parameter. As in the Lagrange formulation, our method explores a larger search space by minimizing over both the model and the wavefields and this helps to mitigate some of the problems with local minima. Moreover, the penalty objective depends on the model parameter through discretized wave-equation itself, instead of through its inverse as is the case for the reduced objective. Our main contribution is the introduction of a two-step approach where an augmented wave-equation is solved, followed by solving a sparse linear system for the model. Because our method solves the augmented wave equations independently, we avoid storage of the wavefields and are able to acculate the gradient and Hessian. As such, our method has the best of both worlds by working in the enlarged search space without the need to update and store the forward and adjoint wavefields for all sources. Initial tests of our formulation confirm the validity of our approach and that it is less sensitive to inaccurate initial models. We also successfully carried out true-amplitude "reverse-time" migration without the need to compute adjoint wavefields.

We envisage several extensions of the proposed method. First, since the problem remains linear in the sources randomized source subsampling techniques, where the inversions are carried out over subsets of sources, still apply. Second, extension to 3D frequency-domain waveform inversion is trivial as long as there exists an efficient scalable method to solve the augmented Helmholtz equation. Third, extension to multi-parameter (elastic) inversion and multi-component measurements is straightforward although the relationship between the wavefields and the model parameters may become nonlinear. Finally, the proposed methodology can be extended to a time-domain formulation by incorporating the data-constraint in a coupled system of ODEs representing a spatial discretization of the wave equation. If an efficient time-stepping algorithm exists for such a system, this approach eliminates the need for checkpointing since it is no longer necessary to compute solutions of the wave equation that propagate backwards in time. As with the frequency-domain approach, the matrices needed to solve for the model updates can be aggregated during the computations.

Future research will be aimed at establishing a formal convergence theory for the proposed method. In particular, it is not clear how the penalty parameter λ should be chosen and what kind of continuation strategy should be used. Although there are initial indications that the method is not overly sensitive to noise more research is needed to establish this.

ACKNOWLEDGMENTS

The authors wish to thank Dan and Rachel Gordon for valuable discussions on the CARP-CG method. This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, BP, Chevron, CGG, ConocoPhillips, ION, Petrobras, PGS, Total SA, WesternGeco and Woodside.

REFERENCES

Bertsekas, D., 1996. Constrained Optimization and Lagrange Multiplier Methods, Athena Scientifics.

Bozda, E., Trampert, J., & Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, *Geophysical Journal International*, **185**(2), 845–870.

Bunks, C., 1995. Multiscale seismic waveform inversion, *Geophysics*, **60**(5), 1457.

Cara, M. & Lévêque, J., 1987. Waveform inversion using secondary observables, *Geophysical Research Letters*, **14**(10), 1046–1049.

de Hoop, M. V., van der Hilst, R. D., & Shen, P., 2006. Wave-equation reflection tomography: annihilators and sensitivity kernels, *Geophysical Journal International*, **167**(3), 1332–1352.

Epanomeritakis, I., Akçelik, V., Ghattas, O., & Bielak, J., 2008. A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion, *Inverse Problems*, **24**(3), 034015.

Haber, E., Ascher, U. M., & Oldenburg, D., 2000. On optimization techniques for solving nonlinear inverse problems, *Inverse Problems*, **16**(5), 1263–1280.

Lailly, P., 1983. The seismic inverse problem as a sequence of before stack migrations, in *Conference on Inverse Scattering–Theory and Application.*, vol. 11, pp. 206–220, Soc. Industr. Appl. Math., Philedelphia.

Luo, Y., 1991. Wave-equation traveltime inversion, *Geophysics*, **56**(5), 645.

Moghaddam, P. & Mulder, W., 2012. The diagonalator: inverse data space full waveform inversion, *SEG Technical Program Expanded Abstracts*.

Pratt, G., Shin, C., & Hicks, G., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophysical Journal International*, **133**(2), 341–362.

Shen, P. & Symes, W. W., 2008. Automatic velocity analysis via shot profile migration, *Geophysics*, **73**(5), VE49–VE59.

Shin, C. & Cha, Y. H., 2009. Waveform inversion in the Laplace-Fourier domain, *Geophysical Journal International*, **177**, 1067–1079.

Sirgue, L. & Pratt, R. G., 2004. Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies, *Geophysics*, **69**(1), 231–248.

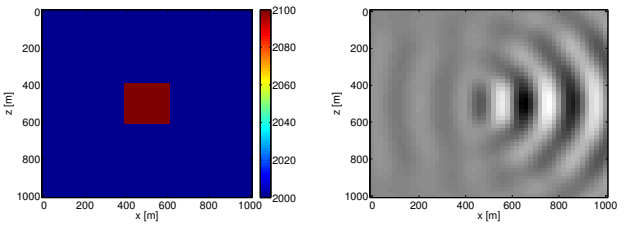
Symes, W. W., 2008. Migration velocity analysis and waveform inversion, *Geophysical Prospecting*, **56**(6), 765–790.

Tarantola, A. & Valette, A., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Reviews of Geophysics and Space Physics*, **20**(2), 129–232.

van Leeuwen, T. & Mulder, W., 2010. A correlation-based misfit criterion for wave-equation traveltime tomography, *Geophysical Journal International*, **182**(3), 1383–1394.

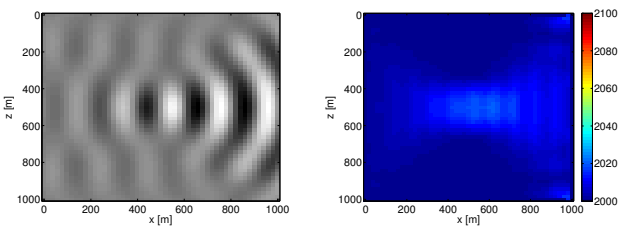
Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26.

Weibull, W., Arntsen, B., & Nilsen, E., 2012. Initial velocity models for Full Waveform Inversion, *SEG Technical Program Expanded Abstracts 2012*, pp. 1–4.



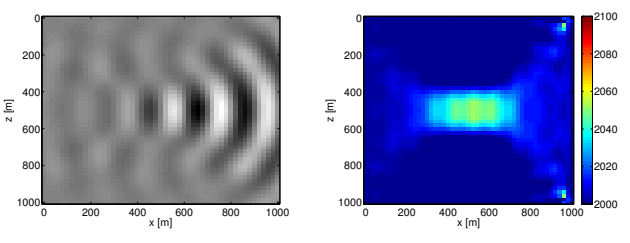
(a) (b)

Figure 1. *(a) Square velocity perturbation embedded in constant background, (b) corresponding scattered wavefield (i.e., the difference between the wavefield for the perturbed model and the wavefield for the constant background model) at 5hz for a point sources at* $z = 500\ m$, $x = 10\ m$.



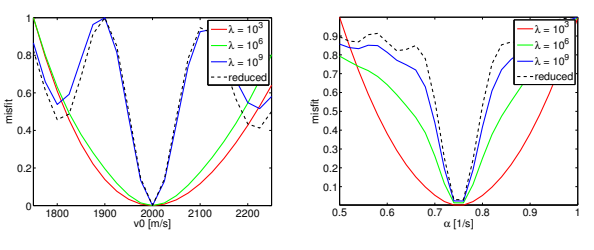
(a) (b)

Figure 2. *(a) Scattered wavefield obtained by solving Equation* (6) *for a constant velocity, (b) corresponding estimate of the model.*



(a) (b)

Figure 3. *(a) Scattered wavefield obtained after 10 iterations, (b) corresponding estimate of the model.*



(a) (b)

Figure 4. *Reduced and penalty objective functions for various values of* λ *as a function of (a)* v_0 *and (b)* α.

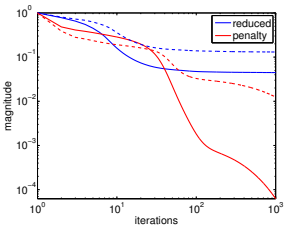


Figure 5. *Convergence histories in terms of the misfit (solid) and two-norm of the gradient (dash).*

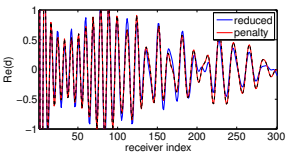


Figure 6. *Data corresponding to the reconstructed velocity profiles. The observed data is plotted with a dashed line.*

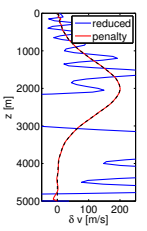
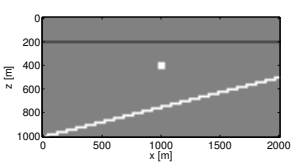
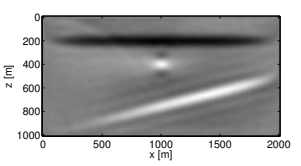


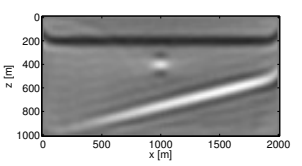
Figure 7. *Reconstructed velocity perturbations (difference between the reconstructed and initial models) after 1000 steepest-descent iterations. The true velocity profile is plotted with a dashed line.*



(a)



(b)



(c)

Figure 8. *(a) True velocity perturbation, (b) traditional reverse-time migration and (c) image obtained with the penalty method by solving Equation* (8)*.*