Wave-equation based inversion with the penalty method: adjoint-state versus wavefield-reconstruction inversion

B. Peters[1], F.J. Herrmann[1] and T. van Leeuwen[2]

[1]University of British Columbia, dept. of Earth, Ocean and Atmospheric Sciences, Vancouver, Canada
[2]Centrum Wiskunde & Informatica, Amsterdam, The Netherlands

January 17, 2014

Abstract

In this paper we make a comparison between wave-equation based inversions based on the adjoint-state and penalty methods. While the adjoint-state method involves the minimization of a data-misfit and exact solutions of the wave-equation for the current velocity model, the penalty-method aims to first find a wavefield that jointly fits the data and honours the physics, in a least-squares sense. Given this reconstructed wavefield, which is a proxy for the true wavefield in the true model, we calculate updates for the velocity model. Aside from being less nonlinear–the acoustic wave equation is linear in the wavefield and model parameters but not in both–the inversion is carried out over a solution space that includes both the model and the wavefield. This larger search space allows the algortihm to circumnavigate local minima, very much in the same way as recently proposed model extentions try to acomplish. We include examples for low frequencies, where we compare full-waveform inversion results for both methods, for good and bad starting models, and for high frequencies where we compare reverse-time migration with linearized imaging based on wavefield-reconstruction inversion. The examples confirm the expected benefits of the proposed method.

Introduction

Nonlinear frequency domain seismic waveform inversion is a partial-differential-equation (PDE) constrained optimization problem (shown here using a least-squares misfit and without regularization)

$$\min_{\mathbf{m},\mathbf{u}} \frac{1}{2}\|P\mathbf{u}-\mathbf{d}\|_2^2 \quad \text{s.t. } A(\mathbf{m})\mathbf{u}=\mathbf{q}, \tag{1}$$

where P is a linear operator restricting the predicted wavefield $\mathbf{u}$ to the receiver locations, A is the discretized Helmholtz equation (+ boundary conditions) system matrix, source term $\mathbf{q}$, observed data $\mathbf{d}$ and the current model estimate is $\mathbf{m}$ (inverse slowness squarred).
The PDE-constrained problem can be recast as a Lagrangian where the PDE-constraint is supplemented using a Lagrangian multiplier. 'All-at-once' methods update the wavefield, model and Lagrangian multiplier (adjoint wavefield) simultaneously, Haber et al. (2000). Therefore this requires storage of all wavefields for each frequency and source, which is a major disadvantage. In seismic waveform inversion it is often preferred to avoid this storage requirement. This can be achieved by eliminating the 'forward' and 'adjoint' variables from the Lagrangian. This is then a reduced Lagrangian problem, also known as the adjoint-state method, see for example Plessix (2006). A major disadvantage of this formulation is the extra nonlinearity introduced by eliminating the PDE constraint. This results in a very nonlinear objective and the associated local minima are a major problem in waveform inversion. Much effort is spent to prevent getting stuck in a local minimum, see for example Shah et al. (2012).
Recently, another approach was introduced by van Leeuwen and Herrmann (2013b,a). Their method does not augment the PDE-constraint to a Lagrangian, but augments it as a least-squares penalty. We refer to this method as 'Wavefield Reconstruction Inversion' (WRI). In this abstract we present an example guided explanation when and why WRI can result in a better model estimation than the reduced Lagrangian method. Specific focus is on the behavior of the gradient of the objective w.r.t. the medium parameters in the low-frequency range (full-waveform inversion) and the high-frequency range (linearized two-way wave-equation imaging). The relevant equations are summarized below:

	Reduced Lagrangian	Wavefield Reconstruction Inversion
Objective	$\phi_{\text{red}}(\mathbf{m}) = \frac{1}{2}\|P\mathbf{u}-\mathbf{d}\|_2^2$	$\bar{\phi}_\lambda(\mathbf{m}) = \frac{1}{2}\|P\bar{\mathbf{u}}-\mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|A(\mathbf{m})\bar{\mathbf{u}}-\mathbf{q}\|_2^2$
Field	$\mathbf{u} = A^{-1}\mathbf{q}$	$\bar{\mathbf{u}} = \arg\min_{\bar{\mathbf{u}}} \left\| \begin{pmatrix} \lambda A(\mathbf{m}) \\ P \end{pmatrix} \bar{\mathbf{u}} - \begin{pmatrix} \lambda\mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2$
Adjoint	$A^*\mathbf{v} = -P^*(P\mathbf{u}-\mathbf{d})$	none
Gradient	$\nabla_{\mathbf{m}}\phi_{\text{red}}(\mathbf{m}) = G(\mathbf{m},\mathbf{u})^*\mathbf{v}$	$\nabla_{\mathbf{m}}\bar{\phi}_\lambda = \lambda^2 G(\mathbf{m},\bar{\mathbf{u}})^*\left(A(\mathbf{m})\bar{\mathbf{u}}-\mathbf{q}\right)$
Partial derivative	$G(\mathbf{m},\mathbf{u}) = \partial A(\mathbf{m})\mathbf{u}/\partial\mathbf{m}$	$G(\mathbf{m},\bar{\mathbf{u}}) = \partial A(\mathbf{m})\bar{\mathbf{u}}/\partial\mathbf{m}$
Gauss-Newton Hessian	$G(\mathbf{m},\mathbf{u})^*A^{-*}P^*PA^{-1}G(\mathbf{m},\mathbf{u})$	$\lambda^2 G(\mathbf{m},\bar{\mathbf{u}})^*G(\mathbf{m},\bar{\mathbf{u}})$ (for small λ)

The WRI objective $\bar{\phi}_\lambda(\mathbf{m})$ can be described as the sum of data and PDE misfits, balanced by a tradeoff parameter (scalar) λ. Solving for $\bar{\mathbf{u}}$ is a variational projection as described in Aravkin and van Leeuwen (2012). This projection is used in every iteration of the nonlinear optimization algorithm used to minimize $\bar{\phi}_\lambda(\mathbf{m})$. The vector $\bar{\mathbf{u}}$ is not a wavefield in the conventional sense, but a solution satisfying both the Helmholtz equation (the physics) and observed data in a least-squares sense. The two gradients w.r.t. the medium parameters have similar structure, but the components are very different. The reduced Lagrangian gradient $\nabla_{\mathbf{m}}\phi_{\text{red}}(\mathbf{m})$ can be thought of as a correlation between the wavefield and back-propagated data residue in the current model estimate. The WRI gradient $\nabla_{\mathbf{m}}\bar{\phi}_\lambda$ on the other hand, is a correlation between the reconstructed wavefield and the PDE-residual in the current model estimate. There is no 'adjoint' field present in the WRI formulation, it only requires one least-squares problem to be solved, compared to two PDE-solves for the reduced Lagrangian method. Because the Gauss-Newton Hessian for the reduced Lagrangian contains the term A^{-1}, it is dense and matrix-vector products with it take extra PDE-solves. The WRI Gauss-Newton Hessian is, for sufficiently small λ, sparse and diagonal (van Leeuwen and Herrmann (2013b)). This means the Gauss-Newton search direction can be computed

at the cost of computing a gradient. A companion abstract explains the relation of WRI with the physics and properties of the solution of the data-augmented wave-equation $\bar{\mathbf{u}}$.

Comparing different formulations of waveform inversion by example

The following examples are generated using a finite-difference discretization of the Helmholtz equation on a regular grid with PML-boundaries on all sides and we invert for the inverse slowness squared $\mathbf{m}$. All examples use synthetic data. There are 64 sources and 64 receivers located near the surface and equally distributed over the horizontal coordinate, maximum offset is 3800 m. The inverse problem is approached with a multi-stage frequency continuation approach where the final model of the first frequency batch is used as initial guess for the second batch and so on. To minimize the reduced Lagrangian and WRI objectives, we use the quasi-Newton algorithm l-BFGS with a Wolfe line-search. Every frequency batch was allowed to use 10 l-BFGS iterations. More iterations per batch did not bring down the objective significantly. The source function is a Ricker waveform with a peak frequency of 30 Hz.

Example 1: Favourable setting for waveform inversion The first example is meant to illustrate that both methods will converge if the problem is relatively easy. The available data has frequencies (Hertz) $\{2,3,4,\ldots,20\}$, is noise free and the frequency batches are designed to be $\{2,3\},\{3,4\},\{4,5\},\ldots,\{19,20\}$. So each frequency batch contains just two frequencies, where one of the frequencies overlaps with the previous batch and the other overlaps with the next batch in line. The true model is a piece from the BG Compass model. The initial model in this example is a quite accurate smoothed version of the true model. True and initial models are shown in fig.1.

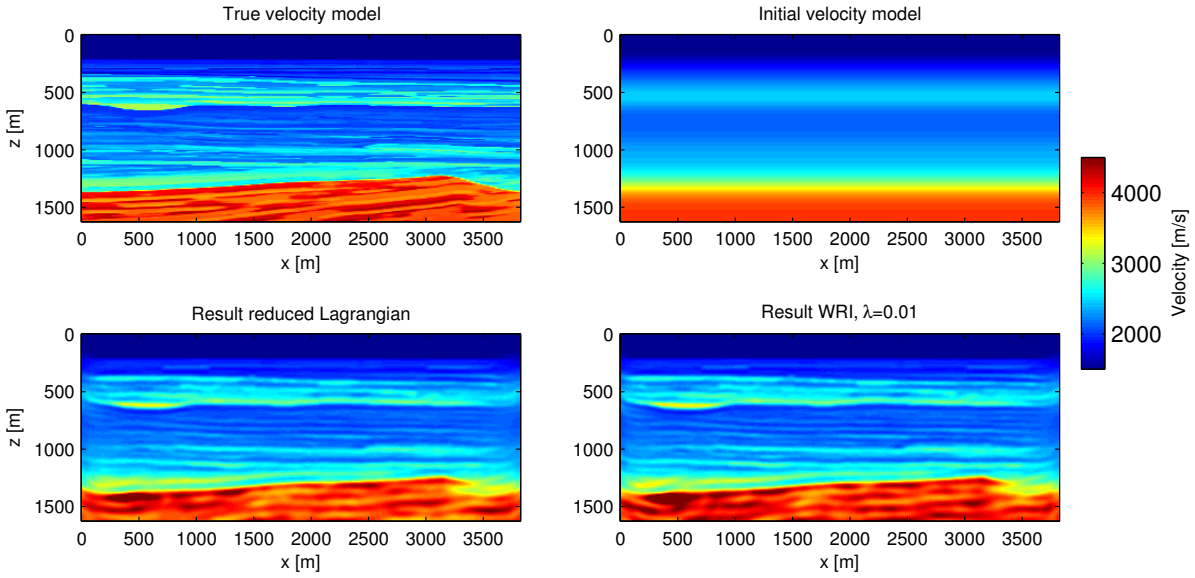

Figure 1 *True model, initial model and final waveform inversion results (velocity) for example 1.*

The resulting models from the WRI method and reduced Lagrangian method are shown in fig.1. Both methods result in a very similar model estimation. The results are good, considering the limited total frequency band and the limited number of frequencies in each batch. The edges are recovered poorly, because all sources and receivers are within the domain shown in fig.1. The WRI result was computed in less than half the time the reduced Lagrangian method took. For grid sizes of a few hundred nodes squared to a thousand squared, direct solvers tend to solve one least-squares problem faster than two PDE-systems. We also observed that WRI requires less l-BFGS line search steps than the reduced Lagrangian method. This also reduces the number of linear systems to be solved. Why the number of line search steps is smaller requires further investigation.

Example 2: No data below 7 Hz. and a poor start model The poor quality start model (fig.2) contains large parts with deviations from the true model up to 1500 m/s. The data also contains random and zero mean noise, $\|\text{noise}\|_2/\|\text{signal}\|_2 = 0.05$. All frequencies up to 7 Hz are missing. The frequency batches are designed to be $\{7\},\{7,8\},\{8,9\},\ldots,\{19,20\}$. Final results are shown in fig.2. This shows there is a clear difference between the reduced Lagrangian and WRI methods. The WRI method results in a much better reconstruction in this example; all main features and interfaces are present at the correct locations. The reconstruction is not perfect, because there is noise in the data and there is no data below 7 Hz. There are also no sources or receivers outside the computational domain.

To explain why the two methods result in different models, we can look at the first gradient at the first

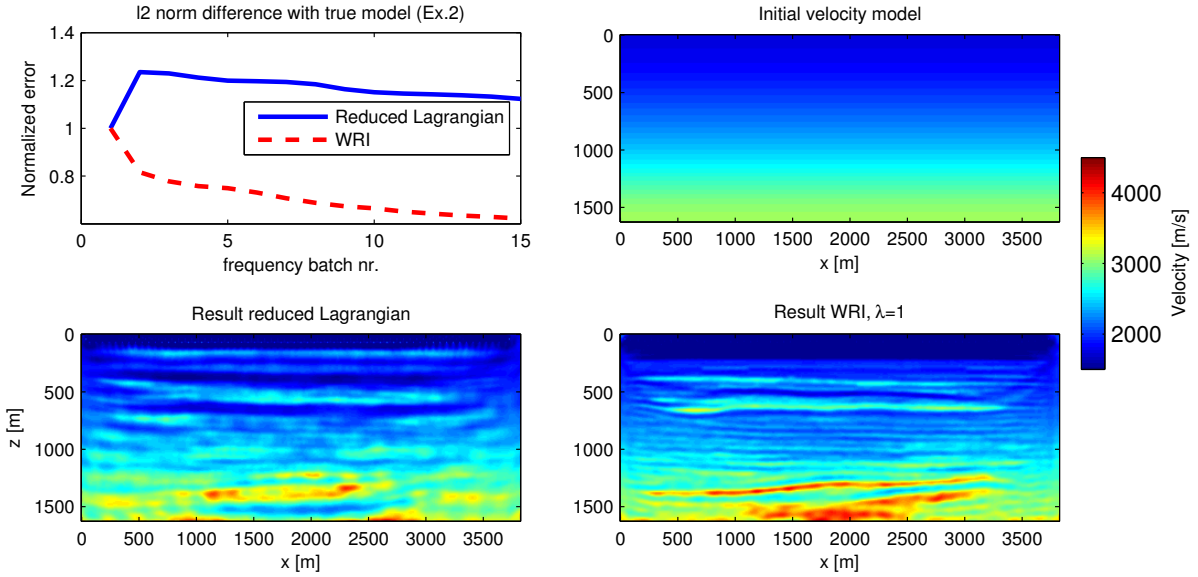

Figure 2 *Initial and final models for ex.2. Top left panel shows the ℓ_2 difference with the true model.*

iteration. These are shown in fig.3. The figure shows a striking difference. The WRI method starts by updating the area near the receivers, while the reduced Lagrangian method starts updating the entire domain. This initial gradient is quite oscillatory, because the first iteration in the first frequency batch contains only one frequency (7Hz.). The reduced Lagrangian gradient also starts updating in the wrong direction in some parts of the model, note the blue area around (x=2000, z=1500) in fig.3 (left panel) and fig.2 (left bottom panel). Later iterations were not able to change that area into the right direction. The ℓ_2 norm difference between the true model and estimated model at the end of every frequency batch is shown in figure 2 (top left), supporting the observation that WRI outperforms the reduced Lagrangian method. When we look at the estimated model at every iteration (not shown here), we observe that the reduced Lagrangian roughly attains its final structure after a dozen iterations, followed by minor changes. The WRI method keeps on updating the model estimate throughout the iterations, including the deeper parts of the model. This illustrates the concept that the reduced Lagrangian method gets stuck near a local minimum quite soon after the start, while the WRI method does not.

Example 3: Imaging This example is intended to investigate if the reduced Lagrangian and the WRI methods can result in very different images, just like we saw in the waveform inversion example. The images in this section are formed as the gradient $\nabla_{\mathbf{m}}\phi_{\text{red}}(\mathbf{m})$ of the reduced Lagrangian objective the Gauss-Newton search direction of the WRI objective ($[\lambda^2 G(\mathbf{m},\bar{\mathbf{u}})^* G(\mathbf{m},\bar{\mathbf{u}})]^{-1}\nabla_{\mathbf{m}}\bar{\phi}_\lambda$). This image can be formed for sufficiently small λ and comes at the same cost as the computation of a single gradient. The estimated models used to form the images are taken as the waveform inversion results from fig.2. There are 150 sources and 150 receivers with 3800 m offset. The source function is a Ricker wavelet with 30Hz. peak frequency. Available data are 30 equispaced frequencies between 4 and 50 Hz. The results are shown in fig.4. RTM imaging using the reduced Lagrangian waveform inversion final model

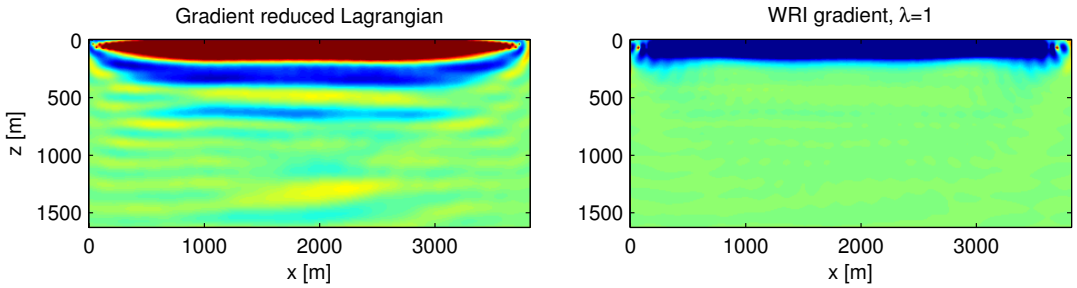

Figure 3 *Gradients at the first iteration for both methods.*

results in an image that shows hardly any structure. The WRI image based on the WRI waveform inversion result shows all main reflectors, except in areas where the model was not very well estimated by waveform inversion. Computation time for WRI was less than half of the RTM computation time.

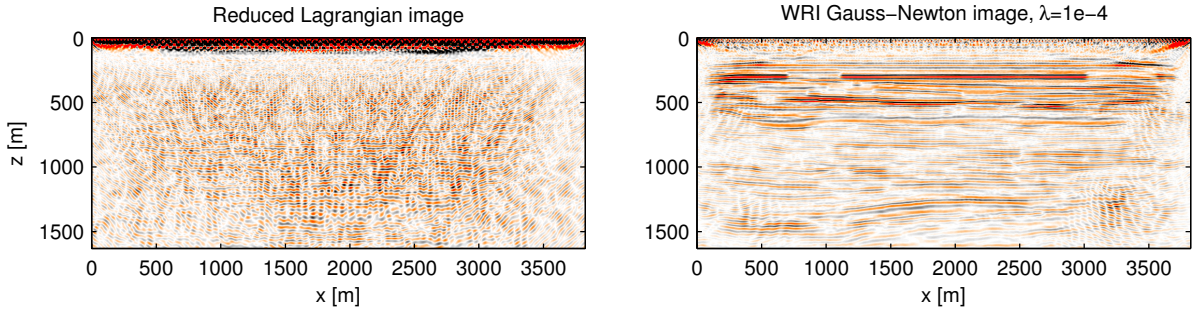

Figure 4 *Left: Reduced Lagrangian imaging (RTM) using the reduced Lagrangian waveform inversion result (fig.2 bottom left). Right: WRI Gauss-Newton image using the WRI result (fig.2 bottom right).*

Conclusions

In this abstract we have tested the performance of the recently introduced Wavefield Reconstruction Inversion (WRI) formulation for full-waveform inversion and compared it to the widely used adjoint-state method. WRI first reconstructs a 'wavefield' that satisfies the data and the Helmholtz equation in a least-squares sense and then estimates medium parameters from it. This leads to a less nonlinear problem and optimization over a larger space of the medium parameters and wavefield. We have shown that this methodology is able to overcome, at least partially, problems related to a poor start model in combination with missing low-frequency data. Subsequent imaging based on the WRI gradient and Gauss-Newton Hessian also indicate the added value of WRI for imaging. WRI can be extended straightforwardly to the simultaneous source scenarios and offers potential benefits for multi-parameter inversion.

References

Aravkin, A.Y. and van Leeuwen, T. [2012] Estimating nuisance parameters in inverse problems. *Inverse Problems*, **28**(11), 115016, ISSN 0266-5611.

Haber, E., Ascher, U.M. and Oldenburg, D. [2000] On optimization techniques for solving nonlinear inverse problems. *Inverse Problems*, **16**(5), 1263–1280, ISSN 0266-5611, doi:10.1088/0266-5611/16/5/309.

Plessix, R.E. [2006] A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, **167**(2), 495–503, ISSN 1365-246X, doi:10.1111/j.1365-246X.2006.02978.x.

Shah, N. et al. [2012] *Quality assured full-waveform inversion: Ensuring starting model adequacy*, chap. 499. 1–5, doi:10.1190/segam2012-1228.1.

van Leeuwen, T. and Herrmann, F.J. [2013a] Mitigating local minima in full-waveform inversion by expanding the search space. *Geophysical Journal International*, **195**, 661–667, doi:10.1093/gji/ggt258.

van Leeuwen, T. and Herrmann, F.J. [2013b] A penalty method for pde-constrained optimization.