Time-domain wavefield reconstruction inversion for large-scale seismics

Gabrio Rizzuti[1], Mathias Louboutin[1], Rongrong Wang[2], and Felix J. Herrmann[1]

[1] Georgia Institute of Technology,

[2] Michigan State University

Abstract

Wavefield reconstruction inversion is an imaging technique akin to full-waveform inversion, albeit based on a relaxed version of the wave equation. This relaxation aims to beat the multimodality typical of full-waveform inversion. However it prevents the use of time-marching solvers for the augmented equation and, as a consequence, cannot be straightforwardly employed to large 3D problems. In this work, we formulate a dual version of wavefield reconstruction inversion amenable to explicit time-domain solvers, yielding a robust and scalable inversion technique.

Introduction

Wave-equation-based seismic inversion techniques are nowadays mature for 3D large-scale deployment, and are becoming a valuable tool in seismic exploration. State-of-the-art time-stepping solvers ensure feasibility on the hardware available to the oil-and-gas industry (and even to a wider audience thanks to cloud computing). Classical inversion methods, however, suffer from local minimum stagnation, which has limited their scope so far. Recently, model extension methods, such as wavefield reconstruction inversion, have demonstrated remarkable results in mitigating this issue. This class of methods, however, leverages on added complexity, which can seriously curtail their application to large problems. For example, wavefield reconstruction inversion (WRI) requires the solution of augmented version of the wave equation, which cannot be trivially solved in a time-marching fashion, and is, therefore, not readily suited to 3D. Here, we propose a dual (Lagrangian) reformulation of WRI, which naturally lends itself to conventional time-domain solvers and large-scale applications, while retaining the robustness of the original method with respect to local minima.

Theory

Classical inversion

Modern seismic inversion is cast as an optimization problem endowed with partial differential equation (PDE) constraints. Given a suitable discretization of the physical quantities of interest, and the algebraic operators that govern their mutual relationship, the problem is expressed as:

$$\min_{\mathbf{m},\mathbf{u}} \frac{1}{2} \|\mathbf{d} - R\mathbf{u}\|^2 \quad \text{subject to} \quad A(\mathbf{m})\mathbf{u} = \mathbf{q}. \tag{1}$$

Data $\mathbf{d} \in \mathbb{R}^{N_t \times N_r \times N_s}$ is recorded in time at N_r receiver locations for different N_s source positions. When appropriate, after a temporal Fourier transform, $\mathbf{d} \in \mathbb{C}^{N_f \times N_r \times N_s}$ for different N_f frequencies. The variable $\mathbf{u} \in \mathbb{R}^{N \times N_t \times N_s}$ (or $\mathbf{u} \in \mathbb{C}^{N \times N_f \times N_s}$, in frequency domain) represents a wavefield, discretized on an N-dimensional space (N being the number of grid points of the discretized physical domain). The wave equation is encoded by the linear system $A(\mathbf{m})\mathbf{u} = \mathbf{q}$. Note that $A(\mathbf{m})$ is parameterized by the physical quantity $\mathbf{m}$ that will indicate, throughout this paper, the acoustic squared slowness. The right-hand side $\mathbf{q}$ is typically a point source. In Equation (1), data is compared to the interpolated values of $\mathbf{u}$, via the operator R.

For industrial-scale 3D problems, the grid size is $N = O(n^3)$ (in big O notation) with $n \sim 1000$ grid points. With conservative source/receiver acquisition coverage, $N_r = O(n)$, $N_s = O(n)$. Furthermore, when time-domain methods with explicit time-marching schemes are applied, $N_t = O(n)$, due to solver stability conditions. In the frequency domain, for full-bandwidth imaging, we still formally need $N_f = O(n)$, although the complexity constants can be relaxed considerably with respect to the time domain. In this large-scale regime, the need for affordable memory and time complexity poses extraordinary challenges. For instance, the storage for the wavefields $\mathbf{u}$ grows as $O(n^5)$, and cannot be attained for each source/time sample at once. Time-complexity is determined by the computation of the solutions of the wave equation. Frequency domain methods are less favorable to 3D, as this amounts to solving or solutions of a large, sparse, non-definite linear system, and do not scale as efficiently as explicit time-domain approaches, which currently remain the only practical option.

Despite these challenges, many techniques such as full-waveform inversion (FWI) (Virieux and Operto, 2009) are now viable for industrial-size problems, especially when the simulated physics is simple (e.g., acoustics). FWI enforces PDE constraints directly as $\mathbf{u} = A(\mathbf{m})^{-1}\mathbf{q}$:

$$\min_{\mathbf{m}} \frac{1}{2} \|\mathbf{d} - F(\mathbf{m})\mathbf{q}\|^2, \quad F(\mathbf{m}) = RA(\mathbf{m})^{-1}, \tag{2}$$

thus avoiding the need for simultaneously storing wavefields for each source. However, the FWI objective function is multimodal and is not easily tackled by local-search optimization. Traditionally, considerable problem-specific intervention was required to consistently ensure its success.

Wavefield reconstruction inversion (WRI)

Recently, model extension methods (Symes, 2008), based on adding slack unknowns to the original problem, have proven their robustness to local minima, but are usually hampered by the computational

costs associated to working in higher dimension. For example, WRI (van Leeuwen and Herrmann, 2013) aims to the reconstruction of both model parameters $\mathbf{m}$ and wavefields $\mathbf{u}$:

$$\min_{\mathbf{m},\mathbf{u}} \mathscr{J}(\mathbf{m},\mathbf{u}), \quad \mathscr{J}(\mathbf{m},\mathbf{u}) = \frac{1}{2}\|\mathbf{d}-R\mathbf{u}\|^2 + \frac{\lambda^2}{2}\|\mathbf{q}-A(\mathbf{m})\mathbf{u}\|^2. \tag{3}$$

A reduced objective $\bar{\mathscr{J}}(\mathbf{m}) = \min_{\mathbf{u}} \mathscr{J}(\mathbf{m},\mathbf{u}) = \mathscr{J}(\mathbf{m},\bar{\mathbf{u}})$ can be obtained by variable projection, where $\bar{\mathbf{u}}$ is the solution of the augmented wave equation

$$\begin{pmatrix} R \\ \lambda A(\mathbf{m}) \end{pmatrix} \bar{\mathbf{u}} = \begin{pmatrix} \mathbf{d} \\ \lambda\mathbf{q} \end{pmatrix}. \tag{4}$$

Solving for $\bar{\mathbf{u}}$ for each source avoids the need for simultaneous storage (similarly to FWI), but time-marching schemes are not available for Equation (4) (contrary to FWI). Therefore this method cannot be trivially extended to 3D.

A dual formulation for time-domain WRI (TWRI)*

In this section, we formulate our proposal for a scalable time-domain version of WRI, here referred to as TWRI* (where the notation * refers to the following dual approach). The starting point is the denoising reformulation of Equation (4):

$$\min_{\mathbf{m},\mathbf{u}} \frac{1}{2}\|\mathbf{q}-A(\mathbf{m})\mathbf{u}\|^2 \quad \text{s.t.} \quad \|\mathbf{d}-R\mathbf{u}\| \le \varepsilon, \tag{5}$$

for a known noise level ε. The Lagrangian associated to Equation (5) takes the form

$$\mathscr{L}(\mathbf{m},\mathbf{u},\mathbf{y}) = \frac{1}{2}\|\mathbf{q}-A(\mathbf{m})\mathbf{u}\|^2 + \langle\mathbf{y},\mathbf{d}-R\mathbf{u}\rangle - \varepsilon\|\mathbf{y}\|, \tag{6}$$

where the multipliers $\mathbf{y}$ belong to a linear space with the same dimension as data. Similarly to WRI, the wavefield variables $\mathbf{u}$ can be eliminated from Equation (6) by applying variable projection:

$$A(\mathbf{m})\bar{\mathbf{u}} = \mathbf{q} + F(\mathbf{m})^*\mathbf{y}, \tag{7}$$

yielding

$$\mathscr{L}(\mathbf{m},\mathbf{y}) = -\frac{1}{2}\|F(\mathbf{m})^*\mathbf{y}\|^2 + \langle\mathbf{y},\mathbf{r}(\mathbf{m})\rangle - \varepsilon\|\mathbf{y}\|, \tag{8}$$

where the forward operator $F(\mathbf{m})$ was defined in Equation (2), and $\mathbf{r}(\mathbf{m})$ is the data residual for the model $\mathbf{m}$. Furthermore:

$$\nabla_{\mathbf{m}}\mathscr{L} = -J[\mathbf{m},\mathbf{q}+F(\mathbf{m})^*\mathbf{y}]^*\mathbf{y}, \quad \nabla_{\mathbf{y}}\mathscr{L} = \mathbf{d} - F(\mathbf{m})(\mathbf{q}+F(\mathbf{m})^*\mathbf{y}) - \varepsilon\mathbf{y}/\|\mathbf{y}\|. \tag{9}$$

Here, $J[\mathbf{m},\mathbf{f}]$ is the Jacobian of the mapping $\mathbf{m} \mapsto F(\mathbf{m})\mathbf{f}$ (and $J[\mathbf{m},\mathbf{f}]^*$ its adjoint).

The evaluation of the Lagrangian $\mathscr{L}$ and its gradients only requires solutions of a conventional wave equation. The additional multipliers $\mathbf{y}$ can be stored in memory whenever joint $(\mathbf{m},\mathbf{y})$ optimization is carried out.

Note that, for a fixed $\mathbf{m}$, the optimal $\bar{\mathbf{y}}$ must solve the non-linear system:

$$\varepsilon\frac{\bar{\mathbf{y}}}{\|\bar{\mathbf{y}}\|} + F(\mathbf{m})F(\mathbf{m})^*\bar{\mathbf{y}} = \mathbf{r}(\mathbf{m}), \tag{10}$$

An iterative solution of this equation will likely result in an expensive scheme. As a result, we propose—similarly to Wang et al. (2016) — a computationally cheap approximation corresponding to the scaled data residual, $\bar{\mathbf{y}} \approx \tilde{\mathbf{y}} = \alpha\mathbf{r}(\mathbf{m})$, albeit for an unknown scalar α (calculated in closed form by maximizing the Lagrangian in Equation (8)). This substitution leads to a reduced objective $\tilde{\mathscr{L}} = \mathscr{L}(\mathbf{m},\tilde{\mathbf{y}})$ function of only $\mathbf{m}$. A rigorous computation of the gradient with respect to $\mathbf{m}$ yields

$$\nabla_{\mathbf{m}}\tilde{\mathscr{L}} = \nabla_{\mathbf{m}}\mathscr{L}(\mathbf{m},\tilde{\mathbf{y}}) - \alpha J[\mathbf{m},\mathbf{q}]^*\nabla_{\mathbf{y}}\mathscr{L}(\mathbf{m},\tilde{\mathbf{y}}). \tag{11}$$

Compared to FWI, the resulting formulation requires twice as many solutions of the wave equation, per objective evaluation (2 vs 1 solves) or gradient computation (4 vs 2 solves). Despite the increased cost, this still compares favorably to conventional WRI, since we avoid the need to solve for an augmented wave equation. On the other hand, we gained the robustness of conventional WRI against local minima, as demonstrated with numerical examples.

Examples

Low-velocity lens

We consider the Gaussian lens model presented in (Huang et al., 2018), here depicted in Figure (1a). The challenge is to image a low-velocity anomaly of Gaussian shape from transmission data, knowing an homogeneous background model. Synthetic data are generated with a Ricker source wavelet of 10 Hz peak frequency (corresponding to a spatial wavelength of 200 m). The FWI result (1b) clearly represents a spurious local minimum, while TWRI* (1c) is converging to the correct model.
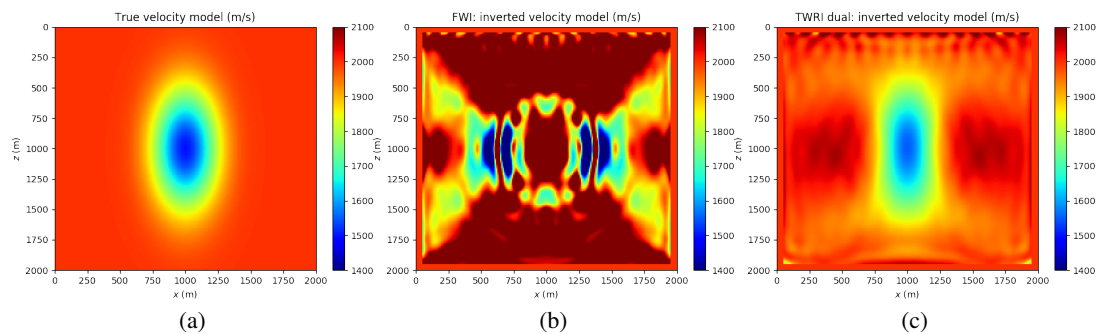

Figure 1: Low-velocity lens problem: (a) true velocity model, (b) FWI inversion result (c) TWRI* inversion result.

BG Compass

For the BG Compass model depicted in Figure (2) (see also Peters et al. (2014)), sources and receivers are placed at the surface, and synthetic data is generated with a 10 Hz Ricker wavelet. Further preprocessing ensures that the frequency components below 5 Hz are removed. In Figure (3), we compare the 5.5 Hz components of the initial gradients of FWI and TWRI*, which illustrates the difficulty of FWI to correctly update the water layer. The final result for TWRI* is shown in Figure (4).
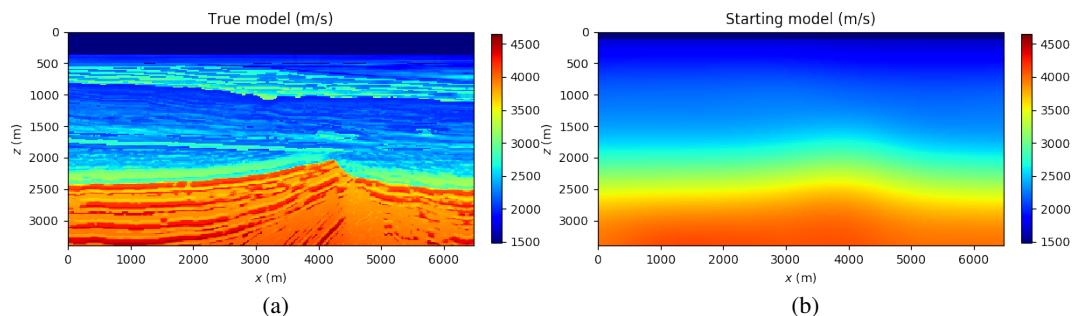

Figure 2: BG Compass: (a) true velocity model, (b) background model.

Conclusions

We introduced a scalable dual reformulation of WRI, TWRI*, based on a model extension by means of dual data multipliers (as opposed to wavefield variables in the original WRI). Since dual variables have the same dimension of data, a joint optimization over model and dual parameters might be devised. In this paper, however, we eliminate these extra unknowns by choosing a filtered version of the data residual, as a cheap estimate. We demonstrated, through numerical experimentation, that the resulting scheme remains relatively robust against local minima. Its objective function can be evaluated and differentiated by using conventional time domain solvers (with a computational cost corresponding to roughly twice what is required for FWI), and is therefore amenable for 3D applications.

Related material

A Julia implementation of the method and examples herein described can be found in

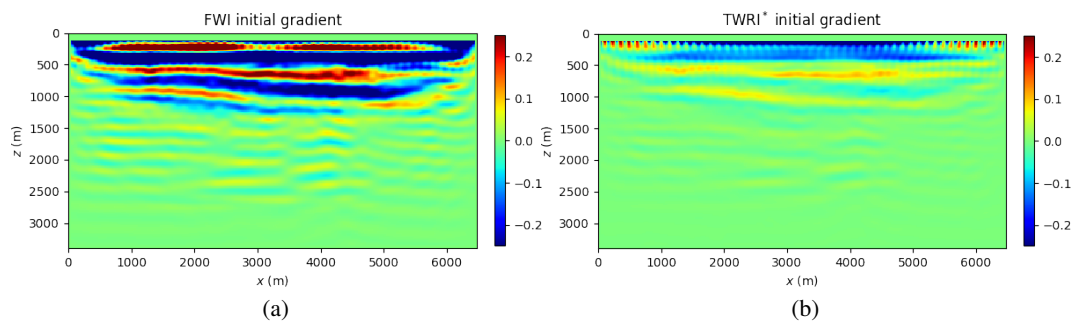https://github.com/slimgroup/Software.rizzuti2020EAGEtwri

(a)

(b)

Figure 3: Initial gradient comparison for a single frequency component 5.5 Hz (with respect to squared slowness parameters): (a) FWI, (b) TWRI*. The FWI gradient points in the wrong direction.
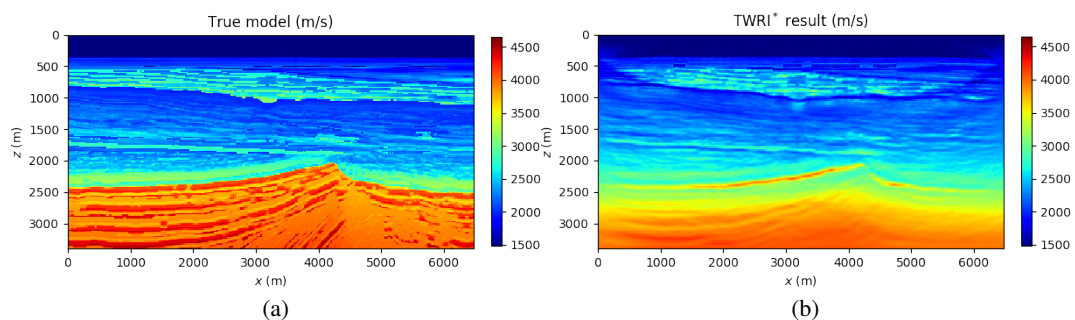

(a)

(b)

Figure 4: BG Compass: (a) true velocity model, (b) TWRI* inversion result.

Acknowledgments

References

Huang, G., Nammour, R. and Symes, W.W. [2018] Volume source-based extended waveform inversion. *Geophysics*, **83**(5), R369–R387.

van Leeuwen, T. and Herrmann, F.J. [2013] Mitigating local minima in full-waveform inversion by expanding the search space. *Geophys. J. Int.*, **195**, 661–667.

Louboutin, M., Lange, M., Luporini, F., Kukreja, N., Witte, P.A., Herrmann, F.J., Velesko, P. and Gorman, G.J. [2019] Devito (v3.1.0): an embedded domain-specific language for finite differences and geophysical exploration. *Geoscientific Model Development*, **12**(3), 1165–1187.

Peters, B., Herrmann, F. and van Leeuwen, T. [2014] Wave-equation Based Inversion with the Penalty Method - Adjoint-state Versus Wavefield-reconstruction Inversion. **2014**(1), 1–5.

Symes, W.W. [2008] Migration velocity analysis and waveform inversion. *Geophysical Prospecting*, **56**, 765–790.

Virieux, J. and Operto, S. [2009] An overview of full-waveform inversion in exploration geophysics. *Geophysics*, **74**(6), 1–26.

Wang, C., Yingst, D., Farmer, P. and Leveille, J. [2016] Full-waveform inversion with the reconstructed wavefield method. In: *SEG Technical Program Expanded Abstracts 2016*, Society of Exploration Geophysicists, 1237–1241.

Witte, P.A., Louboutin, M., Kukreja, N., Luporini, F., Lange, M., Gorman, G.J. and Herrmann, F.J. [2019] A large-scale framework for symbolic implementations of seismic inversion algorithms in Julia. *Geophysics*, **84**(3), F57–F71. (Geophysics).