EAGE

HPC01

Keynote Presentation: Design of Seismic Modeling Engines and Optimization Algorithms for FWI Applications

J.M. Virieux* (Universite Joseph Fourier)

SUMMARY

Different challenges of the full waveform inversion (FWI) will be encountered as we need to move away from acoustic approximation to improved elastic modeling. Quantitative high-resolution imaging could be achieved through efficient gradient estimation of the cost function but also through Hessian influence estimation which is crucial for multiparameter reconstruction. We propose an analysis of the different steps required to achieve the computation of these ingredients: their complexities, the algorithm organization bringing the attention to issues for performing the medium reconstruction.

EAGE

Title

Design of seismic modeling engines and optimization algorithms for FWI applications

Authors

R. Brossier (ISTerre Univ. Grenoble Alpes), L. Combe (Géoazur, CNRS, Univ. Nice Sophia Antipolis L. Métivier (LJK & ISTerre, Univ. Grenoble Alpes), S. Operto (Géoazur, CNRS, Univ. Nice Sophia Antipolis), B. Pajot (ISTerre Univ. Grenoble Alpes) & J. Virieux* (ISTerre, Univ. Grenoble Alpes)

Summary

Different challenges of the full waveform inversion (FWI) will be encountered as we need to move away from acoustic approximation to improved elastic modeling. Quantitative high-resolution imaging could be achieved through efficient gradient estimation of the cost function but also through Hessian influence estimation which is crucial for multiparameter reconstruction. We propose an analysis of the different steps required to achieve the computation of these ingredients: their complexities, the algorithm organization bringing the attention to issues for performing the medium reconstruction.

Introduction

Full waveform inversion (FWI) is a data-fitting inverse problem which allows for the estimation of subsurface parameters from the recordings of seismic waves near the surface with a theoretical spatial resolution of half a wavelength (Virieux and Operto (2009) for a recent review). Since its first introduction in seismics (Lailly, 1983; Tarantola, 1984), this imaging tool has shown to be quite successful and inversions of 3D data have been possible thanks to the access to increased computational resources (Plessix, 2009; Plessix and Perkins, 2010; Sirgue et al., 2010), essentially under the acoustic approximation for single parameter reconstruction. However, FWI remains a computationally intensive imaging tool and resources become a limitation, specially as more 3D data becomes available, higher frequencies are injected in the inversion, or, more complex wave physics (visco-elastic) is accounted for during seismic modeling.

Quantitative reconstruction of multiple parameters will require efficient estimations of the gradient operator of the misfit function and will also require the contribution of the Hessian operator. These estimations are obtained by solving seismic simulations aside other data manipulations we must achieve in the framework of the FWI.

After the short presentation of the theory, we shall analyze the algorithm structures and the computational complexities before showing a typical computational example related to gradient estimations. We shall address also the computational burden when considering multi-parameter reconstruction as they require an estimation of the Hessian influence.

Theory

We first start from the time-domain wave equation written in the generic form

$$\partial_t \mathbf{u}(\mathbf{x},t) - N(\mathbf{m}(\mathbf{x}))H(\nabla)\mathbf{u}(\mathbf{x},t) = \mathbf{s}(\mathbf{x},t), \tag{1}$$

where $\mathbf{u}(\mathbf{x},t)$ is the wavefield unknown formed by particle velocity and stress components. The symmetrical operator $N(\mathbf{m})$ depends on physical properties only and the symmetrical operator $H(\nabla)$ is related to spatial derivatives. This expression, coming from Burridge (1996), allows to write the wave equation as a first-order (implicit) symmetrical hyperbolic system even for general anisotropic media, which is quite appealing to derive adjoint formulation as the forward operator, under this expression

$$N^{-1}(\mathbf{m}(\mathbf{x}))\partial_t \mathbf{u}(\mathbf{x},t) - H(\nabla)\mathbf{u}(\mathbf{x},t) = N^{-1}(\mathbf{m}(\mathbf{x}))\mathbf{s}(\mathbf{x},t), \tag{2}$$

is self-adjoint. The solution can be still obtained from the first equation through an explicit time-marching algorithm using numerical discretization methods while not possible with the second equation.

We can then naturally derive the time-domain FWI gradient expression as

$$\mathscr{G}(\mathbf{x}) = \int_T dt \lambda^T(\mathbf{x},t) \frac{\partial N^{-1}(\mathbf{m}(\mathbf{x}))}{\partial \mathbf{m}} \partial_t \mathbf{u}(\mathbf{x},t), \tag{3}$$

where $\lambda(\mathbf{x},t)$ is the adjoint wavefield which satisfies a similar equation than 1 with a specific source term which is given by residues at receivers weighted by the N operator.
We may also consider the frequency-domain elastodynamic equation from either explicit system (1) or implicit system (2) and we could spatially discretize it into a linear system

$$\mathbf{B}(\mathbf{x},\omega)\mathbf{u}(\mathbf{x},\omega) \quad = \quad \mathbf{s}(\mathbf{x},\omega), \tag{4}$$

where the matrix **B** is the so-called impedance matrix (Marfurt, 1984). Even if we reduce the number of unknowns through parsimonious elimination (Luo and Schuster, 1990) when possible, the linear system (4) could be quite large, sparse and indefinite: it can be solved either with a direct or with an iterative solver.
Direct solver: although the LU decomposition of the impedance matrix is expensive (Amestoy et al., 2000), Operto et al. (2007) have shown the feasibility of direct solvers at low frequencies for the acoustic case. Recent developments that take advantage of some low-rank properties of the impedance matrix have further improved the efficiency of 3D direct solvers and allow to consider anisotropy as well (Wang et al., 2012; Weisbecker et al., 2013). Figure 1 displays the memory improvement as it is the bottle-neck for this solving approach very appealing when considering many sources.
Iterative solver: iterative methods are other possible modeling engines for FWI applications in the acoustic cases (Plessix, 2007, 2009) although fast convergence of the iterative solve might be an issue. Recently, Gordon and Gordon (2010) have proposed a parallel iterative method CARG-CG based on Kaczmarz approach. This method has been proven efficient for the resolution of the 2D and 3D Helmholtz equation in heterogeneous media, even at high frequencies (van Leeuwen et al., 2012; Gordon and Gordon, 2013). Extension to the elastodynamic equations is currently investigated by Li et al. (2014).
Steady state: finally, frequency solutions could be obtained from the steady state of the time-domain solutions of both $\mathbf{u}(\mathbf{x},t)$ and $\lambda(\mathbf{x},t)$ by computing their Fourier transform in the entire domain.
Considering these frequency-domain solutions, we can rely on a frequency-domain FWI gradient expressed as

$$\mathscr{G}(\mathbf{x}) = \sum_{\omega} -i\omega\lambda^T(\mathbf{x},\omega) \frac{\partial N^{-1}(\mathbf{m}(\mathbf{x}))}{\partial \mathbf{m}} \mathbf{u}(\mathbf{x},\omega). \tag{5}$$

Therefore, we may consider four strategies for computing FWI gradients: a time-domain forward modeling and a time-domain gradient computation (TD/TD), a time-domain forward modeling and a frequency-domain gradient estimation (TD/FD), a direct-solver frequency-domain modeling and gradient estimation (FD/FD_D) and an iterative-solver frequency-domain modeling and gradient estimation (FD/FD_I).
For both time-domain and frequency-domain FWI formulations, additional tasks should be performed we shall consider in our complexity analysis.

Computational complexities

We consider a 3D grid of size N^3 with N_{rec} receivers, nt time steps for time integration, nf frequencies, ni iterations for solving the frequency linear system. We need to extract solutions at receivers through projections which are the sinc approximation of Hicks (2002) for finite-difference grids as well as adjoint source injections. We may assume complexity behaviours such that $N_{rec} \approx N^2$, $nt \approx N$, $nf \approx 1$ and $N^3_{hicks} \approx N$. By judicious preconditioning of the acoustic equation (Plessix, 2007; Riyanti et al., 2007; Erlangga and Nabben, 2008), the quantity ni should be independent of N (some strategies are still sensitive to the frequency) while the CARG-CG which could be applied to the acoustic or elastic equations independent of the frequency still requires a significant number of iterations. We consider now the

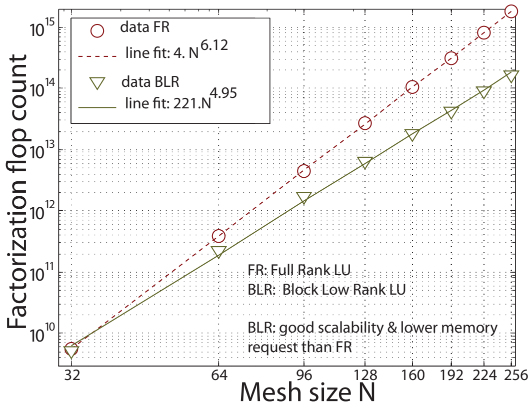

Figure 1 *Flop scalability of the BLR multifrontal factorization of the Laplacian operator discretized with a 3D 7-point stencil, revealing a $O(N^{4.95})$ time complexity for $\varepsilon = 10^{-14}$. Results are obtained with an experimental version of MUMPS embedding Block Low-Rank (BLR) technologies and the standard Full-Rank (FR). The numerical parameter Îţ defines how much approximation we accept to put in the factorization. A small value leads to a good accuracy but potentially less complexity improvements, while a large value leads to a loosely accurate solution but achieved through a substancially cheaper process. Very large values often lead to unacceptable solutions (after Weisbecker).*

complexity behaviour per source.
• The required steps to compute the TD/TD gradient include the forward modeling, the extraction of the solution at receivers, the misfit evaluation and residual computations. These residuals are used as sources for adjoint modeling while stored fields at boundaries and final stage will allow computing again the incident field backward in time for getting the gradient. In practice, only half of the elapsed time is used for modeling purposes with complexity $\mathscr{O}(N^4)$ while other tasks as projections of complexity $\mathscr{O}(N^4)$ require significant resources. The gradient computational complexity is of the same order.
• The TD/FD gradient implementation stores the incident field avoiding repeating the incident modeling but the additional Fourier extraction of similar complexity as the modeling and the gradient $\mathscr{O}(N^4)$ leads to one third of the elapsed time for modeling purposes.
• The FD/FD_D gradient implementation avoids the Fourier extraction which leads to a gradient computational complexity of $\mathscr{O}(N^3)$ while the modeling has a complexity in $\mathscr{O}(N^6)$ for the factorization and in $\mathscr{O}(N^4)$ for the solving step. Projections stays at the complexity $\mathscr{O}(N^4)$. In spite of its time complexity and its memory-demanding feature, this approach may be appealing when considering many sources.
• The FD/FD_I gradient implementation should have a modeling computational complexity of $\mathscr{O}(N^3)$ with a multiplicative coefficient which turns out to behave as N for current iterative solvers with potential improvements.
We may also investigate other ways to reduce the complexity as source encoding, compressive sensing and random fluctuations strategies where cross-talks should be mitigated (Krebs et al., 2009; Herrmann et al., 2009; Ben Hadj Ali et al., 2011; van Leeuwen et al., 2011)

Improving FWI efficiency will require to pay attention not only to modeling issues but also to other operations as projections on acquisition deployment or Fourier transforms if needed. In the TD/FD FWI code *GeoInv3D* and the TD/TD code *TOYxDAC_TIME* developed in the SEISCOPE group with two-level or three-level parallelisms (one MPI parallelism over seismic sources, and one OpenMP parallelism for each source and eventually a domain-decomposition MPI parallelism), we have observed the significant elapsed time of Fourier transforms. Of course, the estimation of the Hessian influence will still put modeling kernels as the main task of the FWI.

From monoparameter to multi-parameter reconstruction

Multi-parameter FWI remains however a very challenging task. Promotion of local updating of various parameters requires the Hessian information as this operator partially corrects for dimensionalities of

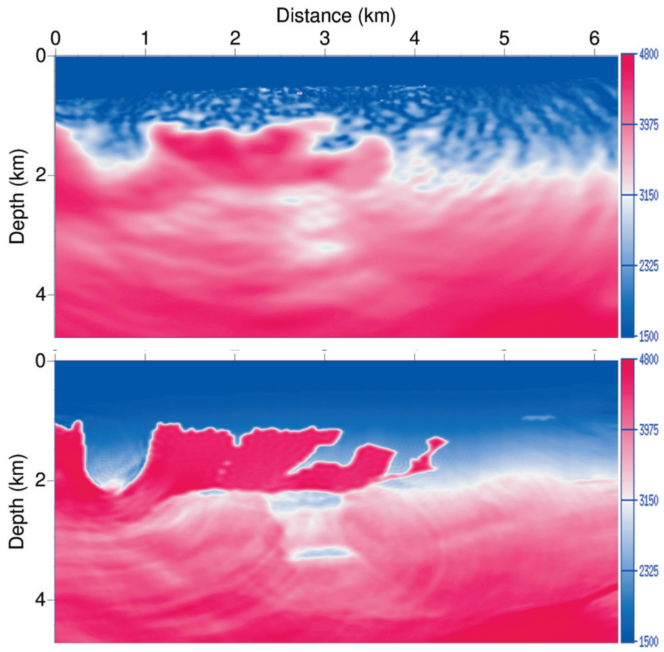

Figure 2 *A zoom on the BP model. The top panel is obtained with l-BFGS-B approach while the bottom panel nearby the real model is obtained by the truncated Newton method (after Métivier)*

parameters, for parameter crosstalk, for acquisition imprint and limited source bandwidth, for poor illumination and second-order scattering effects (Fichtner and Trampert, 2011; Métivier et al., 2013), relaxing requirements on initial model and frequency content for specific applications. This will require additional forward modeling steps and could increase significantly the computational cost of the FWI. A review of these issues is proposed in Operto et al. (2013).

Different behaviours between l-BFGS-B based only on gradient estimations and storages and truncated Newton methods based on second-order adjoint formulation leading to additional forward and adjoint modeling with different source terms have been illustrated by Métivier et al. (2014). One can see that these two ways of considering the Hessian impact do provide different results when considering the same framework for the inversion (Figure 2). These very different behaviours will have an impact on multi-parameter reconstruction.

Conclusions

FWI challenges could be of different nature related to efficient algorithms for forward and adjoint modellings, for gradient and Hessian evaluations but also to data manipulations, such as Fourier transforms or Hicks interpolation or shape functions interpolation, which could be a noticeable part of the FWI workflow. Multi-parameter reconstruction, a necessary step for better quantification of the subsurface, will be a challenging step for FWI as trade-off between parameters may exist. After the different investigations on the gradient estimation performed by many groups, we should be now more concerned by the Hessian operator and efficient methods for estimating its effects. Handling these issues will be strongly related to specific implementation of FWI on modern HPC platforms.

Acknowledgements

This study was funded by the SEISCOPE consortium (*seiscope2.osug.fr*), sponsored by BP, CGG, CHEVRON, EXXON-MOBIL, JGI, PETROBRAS, SAUDI ARAMCO, SCHLUMBERGER, SHELL, SINOPEC, STATOIL, TOTAL, WOODSIDE. This study was granted access to the HPC facilities of CIMENT (Université Joseph Fourier Grenoble) and of GENCI-CINES under Grant 046091.