Amsterdam '14

We G102 05

Time and Frequency-domain FWI Implementations Based on Time Solver - Analysis of Computational Complexities

R. Brossier* (ISTerre, Universite Grenoble Alpes), B. Pajot (ISTerre, Universite Grenoble Alpes), L. Combe (Geoazur, Universite Nice-Sophia Antipolis), S. Operto (Geoazur, CNRS, Universite Nice-Sophia Antipolis), L. Metivier (LJK, CNRS, Universite Grenoble Alpes) & J. Virieux (ISTerre, Universite Grenoble Alpes)

SUMMARY

Full Waveform Inversion is an appealing technique to derive Earth subsurface models. With the development of modern HPC architectures, FWI implementations should benefit from the available computing power. In this study, after a review of time and frequency-domain FWI formulations based on time-domain solver for the forward problem, we discuss the theoretical computational complexities of major steps required to build the FWI gradients. It appears that some significant steps are common to both inversion domain, while other steps remain specific. Some experimental values coming from a real 3D FWI application confirm this theoretical analysis, and show that efficient FWI implementation should focus on all these significant steps for an overall speed up.

Introduction

Full Waveform Inversion (FWI) is an appealing technique to retrieve quantitative high-resolution models of the subsurface (see Virieux and Operto, 2009, for a review). FWI relies on a local optimization problem that tries to minimize differences between observed and simulated seismic data. The misfit function can be formulated either in the time-domain or in the frequency-domain. A necessary ingredient to perform optimization is the gradient of the misfit function, estimated efficiently by an adjoint technique (Plessix, 2006). This gradient involves the solution of two forward seismic simulations: one for the incident wavefield and one for the back-propagated adjoint wavefield.

When considering time-domain forward problem, we can still consider the two domains to perform inversion : the natural time-domain or the frequency-domain after transforming the time-domain fields to the frequency-domain as proposed by Nihei and Li (2007); Sirgue et al. (2007). On the one hand, time-domain FWI takes benefit of all the time-domain processing that can be applied to observed and computed data prior inversion, such as surgical windowing and/or weighting to favor some arrivals. On the other hand, when acquisition is sufficiently dense, frequency-domain FWI allows to decrease significantly the complexity of processed data by considering only few frequencies required to sample the missing wavenumber spectrum of the model.

In this study, we describe time-domain and frequency-domain FWI based on time-domain forward solver, compare and discuss their algorithmic complexity and bottlenecks for efficient implementation on large scale HPC platforms. We will finally show performance on a real 3D FWI application on the Valhall field.

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 operator $N(\mathbf{m})$ depends on physical properties only and $H(\nabla)$ is related to spatial derivatives. This expression, coming from Burridge (1996), allows to write the wave-equation as a first-order (implicit) symmetric hyperbolic system even for general anisotropic media, which is quite appealing to derive adjoint formulation as the forward operator is self-adjoint :

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

Discretization of equation (1) is achieved by an explicit time-marching algorithm. We consider here the VTI anisotropic acoustic wave-equation with a staggered-grid finite-difference (FD) formulation, where the time-derivative is discretized by a second-order leap-frog scheme while the spatial derivatives are discretized by a staggered-grid 4^{th} or 8^{th} order FD schemes (Fornberg, 1988). CPML (Komatitsch and Martin, 2007) absorbing layers are implemented on edges of the simulation box while a flat free surface condition is considered on top.

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

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

where $\boldsymbol{\lambda}(\mathbf{x},t)$ is the adjoint wavefield which satisfies a similar equation than 1 with a specific source term.

Assuming the simulation of both $\mathbf{u}(\mathbf{x},t)$ and $\boldsymbol{\lambda}(\mathbf{x},t)$ have reached the steady state, we can compute their Fourier transform, leading to the frequency-domain solution $\mathbf{u}(\mathbf{x},\omega) = \mathscr{F}_\omega \mathbf{u}(\mathbf{x},t)$ and $\boldsymbol{\lambda}(\mathbf{x},\omega) = \mathscr{F}_\omega \boldsymbol{\lambda}(\mathbf{x},t)$, where $\mathscr{F}_\omega$ is the Fourier operator to extract frequency ω. Considering these frequency-domain solutions available, we can rely on a frequency-domain FWI gradient, which is equivalent to equation (3) if steady-state has been reached :

$$\mathscr{G}(\mathbf{x}) = \sum_\omega -i\omega \boldsymbol{\lambda}^T(\mathbf{x},\omega) \frac{\partial N^{-1}(\mathbf{m}(\mathbf{x}))}{\partial \mathbf{m}} \mathbf{u}(\mathbf{x},\omega). \tag{4}$$

Algorithm 1 Algorithm of the time-domain gradient computation

for $it = 1$ to nt **do**
  update incident field $\mathbf{u}(\mathbf{x},t)$ in forward time + store (in memory) fields on boundaries
  extract solution at receivers through Hicks (2002) interpolation
end for
Misfit function and residuals computations
for $it = 1$ to nt **do**
  compute the adjoint source term through Hicks (2002) interpolation
  update adjoint field $\lambda(\mathbf{x},t)$ in forward time
  restore fields on boundaries + update incident field $\mathbf{u}(\mathbf{x},t)$ in backward time
  update the gradient
end for

For both time and frequency-domains FWI formulations, data at receiver positions are required. This extraction can be accurately achieved using the cardinal sine approximation of Hicks (2002) for receivers at arbitrary positions in the FD grid.

Algorithms

Time-domain FWI gradient

Equation (3) requires to multiply the incident and adjoint fields at each time step. As the adjoint field is a final time field, it is generally solved as an initial time problem through a change of variable on time. However, this means that gradient requires fields at time steps that are not computed simultaneously. Three strategies can be used to overcome this issue

(1) Store the whole incident wavefield on disks when it is computed, to be able to read it at the time of adjoint computation. This strategy allows to model viscous media, does not require recomputation of wavefield, but requires intensive I/Os for large size data volume.

(2) Store partially the incident wavefield in memory at specified time steps through a checkpointing strategy (Symes, 2007; Anderson et al., 2012), and recompute the missing time-steps when required at the time of adjoint computation. This strategy allows to model viscous media but requires some recomputation of the incident wavefield (typically between 1 and 4 times).

(3) Store the incident wavefield on the boundary points, or at the point before the absorbing layer, at all time steps when it is computed and the final time field. When computing the adjoint field, recompute the incident field backward in time from the final step and the boundaries. This strategy does not allow to simulate viscous media (as it assumes reversible equation) but requires only 1 recomputation of the incident field.

Algorithm 1 shows the required steps to compute the time-domain gradient, using the third strategy used in our implementation.

Frequency-domain FWI gradient

The frequency-domain gradient implementation appears simpler to compute (algorithm 2), as soon as the frequency-domain wave-equation solutions are available. One incident wavefield computation is saved, compared to time-domain formulation, but the Fourier transform is required based on discrete Fourier transform (Sirgue et al., 2007) or phase-sensitive detection (Nihei and Li, 2007).

The data volume in frequency-domain FWI is also significantly reduced, as the number of frequencies considered in the inversion is much smaller than the number of time-step.

Complexities

In table 1, we review the computational complexities of the major steps required by time and frequency-domain gradients computation. Some steps, as incident and adjoint fields computation or data extraction and adjoint source computation, are common to both strategies and improvement on implementation or specific developments done for hardware specificity would benefit the two approaches. Several steps are, however, specific to each approach, as the gradient step for time-domain FWI and Fourier transform step for frequency-domain FWI. Based on these theoretical complexities, it is clear that the forward-problem

Algorithm 2 Algorithm of the frequency-domain gradient computation. Note that $nt2 > nt$ (nt from algorithm 1) as we need to reach steady-state.

for $it = 1$ to $nt2$ **do**
    update incident field $\mathbf{u}(\mathbf{x},t)$ in forward time
    update the frequency domain incident field $\mathbf{u}(\mathbf{x},\omega)$ for frequencies of interest
    extract solution at receivers through Hicks (2002) interpolation
end for
Misfit function and residuals computations
for $it = 1$ to $nt2$ **do**
    compute the adjoint source term through Hicks (2002) interpolation
    update incident field $\lambda(\mathbf{x},t)$ in forward time
    update the frequency-domain adjoint field $\lambda(\mathbf{x},\omega)$ for frequencies of interest
end for
compute the gradient for each frequency and summation

Table 1 *Theoretical computational complexities of the different steps required for time and frequency-domain FWI gradients computation. We consider here a 3D grid of size N^3, with N_{rec} receivers, nt time steps, nf frequencies in inversion (for frequency-domain) and a sinc operator for Hicks (2002) interpolation of size N_{hicks}. To be able to compare steps, the third column is built considering $N_{rec} \approx N^2$, $nt \approx N$, $nf \approx 1$ and $N_{hicks}^3 \approx N$, which seems reasonable value for realistic 3D data-sets.*

Steps	computational complexity	rough computational complexity
Forward simulation	$\mathcal{O}(N^3 \times nt)$	$\mathcal{O}(N^4)$
Adjoint simulation	$\mathcal{O}(N^3 \times nt)$	$\mathcal{O}(N^4)$
Extract solution at receiver	$\mathcal{O}(N_{rec} \times N_{hicks}^3 \times nt)$	$\mathcal{O}(N^4)$
Compute adjoint source term	$\mathcal{O}(N_{rec} \times N_{hicks}^3 \times nt)$	$\mathcal{O}(N^4)$
Compute gradient (time-domain)	$\mathcal{O}(N^3 \times nt)$	$\mathcal{O}(N^4)$
Fourier transform (frequency-domain)	$\mathcal{O}(N^3 \times nf \times nt)$	$\mathcal{O}(N^4)$
Compute gradient (frequency-domain)	$\mathcal{O}(N^3 \times nf)$	$\mathcal{O}(N^3)$

steps (namely update of stress and velocity by FD kernel) represent only a part of the elapsed time for FWI in both domains. Specific developments of FD kernel represents therefore only a limited part of potential improvement of FWI implementations for realistic 3D projects.

Application on the Valhall dataset

In order to illustrate the theoretical complexities evaluated in the previous section, we present here some performance results obtained on a similar 3D application on the Valhall dataset. We consider here the frequency-domain FWI code *GeoInv3D* and the time-domain code *TOYxDAC_TIME*, both developed in the SEISCOPE group. Order 4 finite difference is considered here. The computational grid is $71 \times 140 \times 240$, including PML points, the number of time-steps is 1022 and the number of receivers 49500. The frequency-domain FWI considers 5 frequencies. The Hicks (2002) interpolation used for data extraction and adjoint source building considers 8 points in each direction.

Tests have been performed on the IBM BlueGene/Q from IDRIS/CNRS HPC center. Both implementations rely on a two-level parallelism: one MPI parallelism over seismic sources, and one OpenMP parallelism for each source. We consider here 2 sources per node, meaning that 8 cores and 32 threads (using hyperthreading) are associated to each seismic source.

Figure 1 sums up performance on the case-study. We can see that FD kernel computation represents only 36% and 44% of the whole gradient building elapsed time for frequency and time-domain gradient, respectively. As expected from theoretical complexities, all the significant steps identified shows a non-negligible computational effort and both gradients finally required a similar elapsed-time. All these steps are here parallelized with a similar shared-memory OpenMP implementation in both codes. This

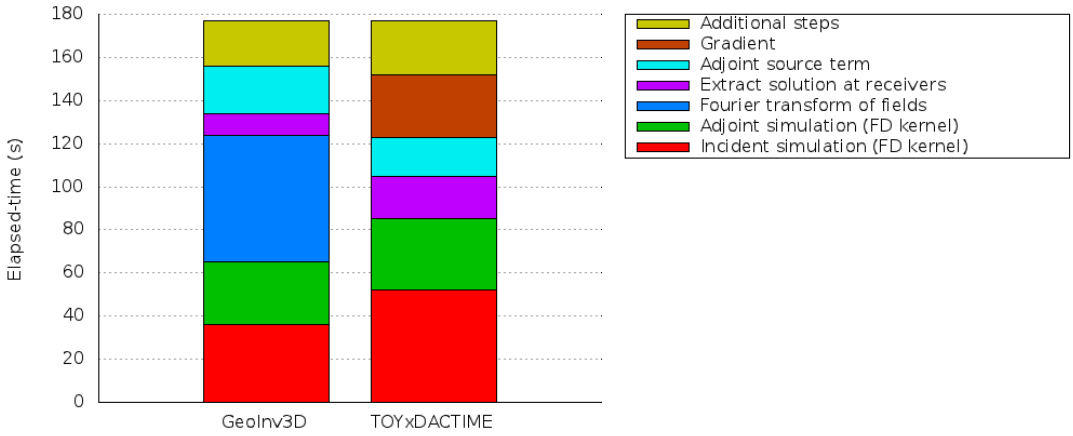

Figure 1 *Elapsed time of major steps of gradient building for our frequency-domain and time-domain implementations applied on the Valhall model.*

study shows that specific hardware-oriented implementation and porting of FWI codes on many-cores architectures should take into account all these steps for an efficient overall speed-up.

Conclusions

In this study, we present time and frequency-domain FWI formulations based on time-domain solver for the forward problem. We discuss theoretical complexity of major steps required for both implementations and show experimental values coming from a realistic 3D case. This analysis shows that, in addition to the forward problem FD kernels, other steps require a significant elapsed time of computation that should be considered for global efficient implementation of FWI on modern HPC platforms.

Acknowledgments

This study was funded by the SEISCOPE II consortium (*http://seiscope2.osug.fr*), sponsored by BP, CGG, CHEVRON, EXXON-MOBIL, JGI, PETROBRAS, SAUDI ARAMCO, SHELL, SINOPEC, STATOIL, TOTAL and WESTERNGECO. This study was granted access to the HPC facilities of CIMENT (Université Joseph Fourier Grenoble), and of GENCI-CINES under Grant 046091 of GENCI (Grand Equipement National de Calcul Intensif).

References

Anderson, J.E., Tan, L. and Wang, D. [2012] Time-reversal checkpointing methods for RTM and FWI. *Geophysics*, **77**, S93–S103.

Burridge, R. [1996] Elastic waves in anisotopric media, schlumberger-Doll Research.

Fornberg, B. [1988] Generation of finite difference formulas on arbitrarily spaced grids. *Mathematics of Computation*, **51**, 699–706.

Hicks, G.J. [2002] Arbitrary source and receiver positioning in finite-difference schemes using kaiser windowed sinc functions. *Geophysics*, **67**, 156–166.

Komatitsch, D. and Martin, R. [2007] An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation. *Geophysics*, **72**(5), SM155–SM167.

Nihei, K.T. and Li, X. [2007] Frequency response modelling of seismic waves using finite difference time domain with phase sensitive detection (TD-PSD). *Geophysical Journal International*, **169**, 1069–1078.

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.

Sirgue, L., Etgen, T.J., Albertin, U. and Brandsberg-Dahl, S. [2007] System and method for 3D frequency-domain waveform inversion based on 3D time-domain forward modeling. *US Patent Application Publication*, **US2007/0282535 A1**.

Symes, W.W. [2007] Reverse time migration with optimal checkpointing. *Geophysics*, **72**(5), SM213–SM221, doi:10.1190/1.2742686.

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