Hierarchical seismic imaging: A multiscale approach

Jean Virieux

Improved vision of the earth’s interior relies on remote-sensing approaches in which information sent from the surface of the earth is brought back to the surface after sampling the medium properties. The technique with the expected best resolution of these indirect techniques is based on seismic waves propagating inside the earth. Unfortunately, interpretation of the recorded seismic data is challenging because of the oscillating nature of seismic records, especially when the medium is complex, as it is, for example, in the upper crust.

The dramatic recent increase of acquisition configuration (from narrow-azimuth to full-azimuth coverage, longer and longer offset ranges, and broadband frequency range) and the massive increase of real or virtual recording points has opened new opportunities in the imaging of physical quantities (Figure 1).

Before the emergence of wave-azimuth, long-offset data, the short-offset range recorded by seismic-reflection surveys and the limited frequency bandwidth of the active sources made seismic imaging inadequate for intermediate wavelengths (Jannane et al., 1989). This led to a quite elaborate two-step workflow: (1) construction of the macromodel using kinematic information (essentially traveltimes of reflections) and (2) amplitude projection through different types of migrations (Claerbout and Doherty, 1972; Gazdag, 1978; Stolt, 1978; Baysal et al., 1983; Yilmaz, 2001; Biondi and Symes, 2004; Virieux and Operto, 2009).

This procedure has turned out to be efficient for relatively simple geologic targets in shallow-water environments, although more limited performances have been achieved for imaging structurally complex structures such as salt domes, subbasalt targets, thrust belts, and foothills. In such complex environments, building an accurate and quantitative depth-dependent velocity model is challenging. Various approaches for iterative updating of the macromodel reconstruction have been proposed (Snieder et al., 1989; Docherty et al., 1997), but they remain limited by poor sensitivity of the reflection-seismic data to the large and intermediate wavelengths of the subsurface (Virieux and Operto, 2009).

With these new data, one can exploit the entire information contained in these time series for improving the seismic-imaging procedure. This strategy has been termed full-waveform inversion (FWI) (Lailly, 1983; Tarantola, 1984). This approach is a challenging data-fitting procedure for extracting quantitative information from seismograms. High-resolution imaging at half the propagated wavelength is expected from this procedure as long as the illumination is enough (Virieux and Operto, 2009).

This least-square minimization method is based on the well-established local Newton optimization approach, which should reduce the sum of squared misfit sample differences over sources, receivers, and times or transformed quantities as wavenumbers or frequencies (Tarantola, 2005) with the following misfit expression:

$$\mathcal{E} = \sum (d_{\text{cal}}(m) - d_{\text{obs}})^2,$$

where the synthetic data are estimated from the model by the general expression $d_{\text{cal}} = G(m)$. The nonlinear operator $G$ links the model parameters to the synthetic data. The observed data are denoted by $d_{\text{obs}}$. One may consider recorded data in the field or any other transformed data as long as the transformation could be applied to both synthetic and observed data.

This method, which balances errors in data, errors in forward modeling, and limitations in the model description, is sensitive to outliers in spite of an increase in data redundancy. One can benefit from the local differential property for efficient convergence with estimation of the gradient operator $\gamma$ and eventually estimation of the effect of the Hessian operator $\mathcal{H}$. Other definitions of the misfit function can mitigate this sensitivity, although one should preserve the capacity of computing derivatives.

The deduced normal equation is expressed compactly as

$$\mathcal{H} \Delta m = -\gamma,$$

The model perturbation $\Delta m$ could be estimated locally, and therefore, we have to deal with a starting guess of the solution: The initial model which will be updated through iterations as the relation between seismic data and model parameters is a nonlinear relation. How to design this initial guess is still a debating issue, with different strategies such as traveltime tomography.

![Figure 1](http://www.iris.edu/data/distribution), accessed 1 October 2014. Used by permission.
(Luo and Schuster, 1991) with the difficult task of picking times (Hale, 2013; Ma and Hale, 2013), Fourier-Laplace transform (Shin and Cha, 2009), stereotomography (Prieux et al., 2011), misfit reshaping (van Leeuwen and Mulder, 2010; Luo and Sava, 2011; Warner and Guasch, 2014; Guasch and Warner, 2014), and other strategies.

The key ingredient of this optimization technique is an efficient forward-modeling engine devoted to the gradient estimation through the adjoint approach (Plessix, 2006). This engine should compute seismic waves efficiently in a heterogeneous medium for the incident and adjoint wavefields (two forward problems to be solved per source). This leads to the gradient field estimation in the model space, a necessary output of the forward-modeling engine (Figure 2) and an additional workout when designing modeling approaches.

This is the core of the FWI workflow, and the efficiency of the method will depend on how one designs this core structure. Separating gradient buildup (forward-problem dependent) and model updating (optimization dependent) is possible for the wave equation even for anisotropic viscoelastic modeling. It will provide an abstract level in FWI designing that one might favor for simplicity and maintenance of computer algorithms and codes. Applications-oriented investigation could be based on this FWI workflow with a need for understanding the following concepts without dealing with their implementation.

Updating model parameters from the gradient requires an estimation of the Hessian impact (Pratt et al., 1998), which can be achieved by different means from a single arbitrary factor when considering the steepest descent (Gauthier et al., 1986) or conjugate gradient (Mora, 1987), a diagonal Hessian estimation (Shin et al., 2001), a quasi-Newton Hessian estimation through l-BFGS technique (Brossier et al., 2009), or a full-Newton Hessian estimation such as the truncated Newton (TCN) through a matrix-free approach (Métivier et al., 2014).

The inverse Hessian operator acts as a deconvolution operator that accounts for the limited bandwidth of the seismic data, corrects for the loss of amplitude of poorly illuminated subsurface parameters, and reduces leakage between different parameters. In addition, it helps to remove artifacts that the second-order reflected waves might generate on the model update (Figure 3).

Figure 2. Estimation of the gradient through time computation. The gradient will be the sum of the last column of values. (a) The contribution at a short time when the incident field (left panel) and adjoint field (middle panel) do not crosscorrelate. (b) A later time when crosscorrelation provides a gradient contribution (right panel) while (c) at a later time, the crosscorrelation of fields contributing to the gradient estimation goes back to zero. Courtesy of S. Operto. Used by permission.

Figure 3. (a) A portion of the BP 2004 model in the left panel and the initial model in the right panel we start with. (b) FWI results for three Hessian-influence estimations. From left, conjugate gradient (CG) method, quasi-Newton method (l-BFGS), and full-Newton method (TCN). Please note differences in the deep canyon on the left among different schemes. Courtesy of L. Métivier and R. Brossier. Used by permission.
The running-model solution could be trapped into a local minima because of the limited accuracy of the starting model, the lack of low frequencies, the presence of noise, and the approximate modeling of the wave-physics complexity. Different hierarchical multiscale strategies are designed to mitigate the nonlinearity and ill-posedness of FWI (Virieux and Operto, 2009). The hierarchical strategy is based on the organized data structure and is therefore data driven.

We can consider specific phases of the data such as diving or early waves, limited frequency-band window of the data with the increasing introduction of higher frequencies during inversion, and so forth — any manipulation on real and synthetic data that we consider worthy of investigation. The final step should be performed with full data without any restriction. In this way, we hope to swing around local minima using all available intuition and expertise in seismic data analysis, making the FWI workflow not so trivial to apply.

Other alternatives for escaping local minima problems are related to model description, in which we can consider a progressive inclusion of shorter and shorter wavelengths to be reconstructed through iterations. The way we discretize the model and the way we select the parameter sets to be inverted will influence FWI behavior strongly. Moreover, model prior information, if available, will complement the data gradient nicely by the introduction of completely different spectral features. In addition, buildup of this model gradient is less complex to elaborate.

Simple synthetic studies illustrate the promise of this technique, which should address the reconstruction of various parameters, from P-wave and S-wave velocities to density, anisotropy, and attenuation. Scaling and cross-talk between parameters are the difficulties to face. Often, one can consider parameters which affect the kinematics of the data and parameters which affect the dynamics (amplitude) of the data — another illustration of the important hierarchical structure of seismic data.

For a glance at potentialities of this FWI application, illustrations on a real data set of Valhall oil field in the North Sea (Figure 4) will detail the dramatic increase of resolution one can expect when optimization converges. Challenging issues related to multiparameter reconstruction will be presented also (Gholami et al., 2013a; Gholami, 2013b; Prieux et al., 2013a, 2013b). Based on synthetic tests and data reconstruction, we shall discuss aspects of the FWI approach through this case study.

References


Acknowledgments: Jean Virieux would like to thanks all participants and sponsors of the consortium SEISCOPE for sharing their scientific investigations and studies devoted to various aspects of the FWI approach (http://seiscope2.osug.fr). Continuous discussions with R. Brossier, L. Combe, L. Métivier, S. Operto, B. Pajot, A. Ribodetti, and students of the consortium have made this presentation possible. Many thanks to BP Norway for providing the real data set from Valhall. Deep thanks to SEG for this opportunity to be an SEG Distinguished Lecturer and to Judy Wall of the SEG Professional Development Department for organization of this worldwide tour.