Introduction

Solving FWI requires numerous computations of wavefields over large models. Since wave equation is usually solved with a finite difference schemes, computational grid has to be fine enough to support frequencies present in data. For this reason, global optimization methods are not used to find the global minimum of FWI objective function. The most feasible alternative is to use iterative optimization methods. We compare the performance of the nonlinear conjugate gradient (NCG) method and the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method. We present results of inversion sensitivity analysis for the two aforementioned optimization schemes, analyzing the effects of step lengths, initial models and acquisition geometries.

Elastic full waveform inversion

We consider the following isotropic elastic wave equation:

$$\rho\ddot{\mathbf{u}} - \lambda[\nabla(\nabla\cdot\mathbf{u})] - \mu\left[\nabla\cdot\left(\nabla\mathbf{u}+\nabla\mathbf{u}^T\right)\right] = \mathbf{f}, \tag{1}$$

where $\mathbf{u}(e,\mathbf{x},t)$ is the elastic wavefield, $\mathbf{f}(e,\mathbf{x},t)$ is the source function, $\lambda(\mathbf{x})$ and $\mu(\mathbf{x})$ are the Lamé parameters, $\rho(\mathbf{x})$ is the density and e, $\mathbf{x}$ and t are, respectively, the experiment index, space coordinates and time. Considering $\mathbf{u}_s$ as the source wavefield, the waveform inversion problem is solved by minimizing an objective function $\mathscr{J}(\mathbf{u}_s,\lambda,\mu)$ consisting of a term that evaluates the misfit between the simulated and observed data. We use a gradient-based method to update the model, and we compute the gradient of $\mathscr{J}$ with respect to λ and μ using the adjoint-state method (Plessix, 2006).

For the implementation of NCG as the numerical optimization strategy to solve the inversion, we follow the adaptation proposed by Fletcher and Reeves (1964). The NCG method is a good choice for large scale problems such as FWI because it only requires evaluation of the objective function and its gradient. The storage requirement is also minimal (Nocedal and Wright, 2006). One important consideration for this method is for the choice of a step length α. If the line search is not exact, estimated α may be an ascent direction. This can be avoided by imposing strong Wolfe conditions.

Alternatively, limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm is a quasi-Newton method and takes advantage of the second order information for the updates. Since the FWI uses large models, the computation and storage of the full Hessian matrix is prohibitively expensive. The limited-memory version of BFGS algorithm avoids the problem of computing and storing the full Hessian by only saving m vector pairs $\{s_i, y_i\}$, which are later used to compute inverse Hessian approximation. The model update equation then takes the form:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha_k \mathbf{H}_k \mathbf{g}_k, \tag{2}$$

where H_k denotes Hessian inverse approximation derived above. The step length α_k should satisfy the Wolfe conditions or strong Wolfe conditions for the updates to be stable (Nocedal and Wright, 2006).

Examples

We illustrate the elastic full waveform inversion method with four synthetic examples and compare inversions using 1- Nonlinear conjugate gradient and L-BFGS as the optimization methods; 2- different step-length for the optimization methods; 3- different initial models; 4- different acquisition geometry. All synthetic examples use a model with negative and positive Gaussian anomalies centered at $(1.25, 0.75)$ and $(1.25, 1.75)$ km, respectively (Figure 1). We use the number of iterations as a stopping criterion for all inversions due to the computational cost of FWI. For the first, second and third examples, we simulate the acquisition geometry with 20 evenly spaced sources located at $x = 0.1$ km and a vertical line of receivers at $x = 2.4$ km.

In the first example, we choose the constant background of the true models (Figure 1) as the initial

models for parameters μ and λ. We use quadratic interpolation line search and we set the test step-length to 0.1. The inversion results are shown in Figures 2a and 3a with the corresponding objective functions in Figures 2b and 3b. The tested trial step length is not optimal for the NCG algorithm, as shown by the oscillatory behavior of the objective function and resulting models. L-BFGS updates are stable; however, inversion is stuck at local minimum after just a few iterations.

The second experiment uses the same initial models as the first example, but a much smaller test step-length equal to 0.001. The inversion results are shown in Figures 4a and 5a with the corresponding objective functions in Figures 4b and 5b. By selecting smaller trial step length parameter, we are able to improve NCG updates and converge to a local minimum. The models recovered with L-BFGS remain largely unaffected by the change and the objective function behavior is also similar as in the first example.

In the third example, we use the same test step-length as the second experiment, but a different initial model (Figure 6), which are much closer to the true models (Figure 1). The inversion results are shown in Figures 7a and 8a with the corresponding objective functions in Figures 7b and 8b. Note that both recovered models and the behavior of the objective function is similar for NCG and L-BFGS. Recovered anomalies are much closer to the true model than in a previous example, but still suffer from acquisition-related artifacts.

The last example uses the same initial models and test step-length as the third experiment, but a different acquisition geometry. For this experiment, we simulate the data with only ten shots. The inversion results are shown in Figures 9a and 10a with the corresponding objective functions in Figures 9b and 10b. Just as in the previous example, NGC and L-BFGS yield similar results. The recovered anomalies are elongated along the x axis and the acquisition footprint is stronger than before (the pattern resembling wave interference in the background model).

Conclusions

The choice of the step length parameter is crucial: choosing a step too small can significantly slow the convergence of an algorithm and lead to a local minimum far from the true solution. However, a step too big can prevent the algorithm from converging at all. The two point quadratic interpolation strategy we use for picking the step length relies critically on the trial step length. It is a parameter that requires some testing before an optimal choice can be made for running the full inversion. We find that L-BFGS is more robust to the choice of the trial step length. Additionally, FWI requires an initial model which is highly accurate kinematically. Since building such an accurate model from field data can pose a challenge, strategies to decrease the inversion sensitivity to the initial model are actively researched. Inversion with constant background model yields results that resemble a true model to some degree, but neither shape nor the magnitude of the anomalies is truly recovered. Our examples also reveal that even starting with a very good initial model cannot recover the true model. There are a couple of reasons for that. First, seismic acquisition has always limited aperture and as such certain regions of the model have more waves passing through them than others. This results in parts of the model being well constrained by data while other parts are barely constrained. Second, the parameters of the inversion are not completely independent and thus some information interchange between them results in smearing in the recovered models. Finally, the geometry of sources and receivers plays an important role in FWI. In a scenario with field data, one has to start with building a good low frequency initial model and in order to do that, long offsets (source - receiver distances) are required. The spatial distribution of sources and receivers also affects the illumination of subsurface targets. Regular downsampling of the sources and/or receivers results in the artifacts which look like interference pattern; this interference is mostly cancelled when sampling is dense enough.

References

Fletcher, R. and Reeves, C.M. [1964] Function minimization by conjugate gradients. *The computer journal*, **7**(2), 149–154.

Nocedal, J. and Wright, S. [2006] *Numerical optimization*. Springer Science & Business Media.
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**, 495–503.

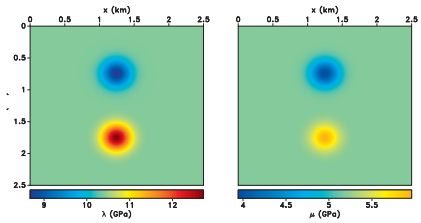
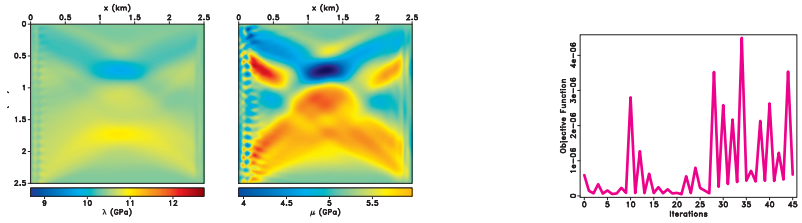
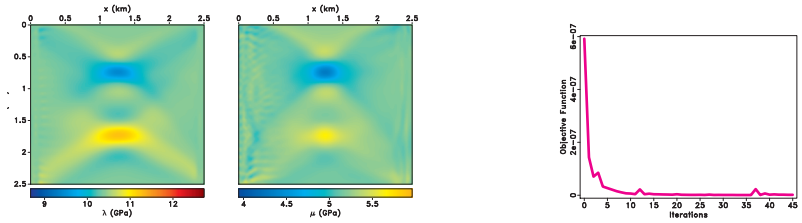


Figure 1 *True λ and μ models with a negative and a positive Gaussian anomalies.*



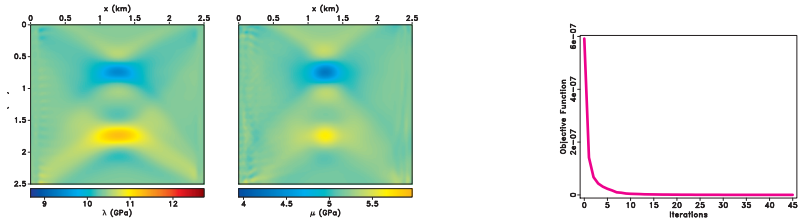
(a) (b)

Figure 2 *Example 1: (a) The models recovered with NCG and the (b) inversion objective function.*



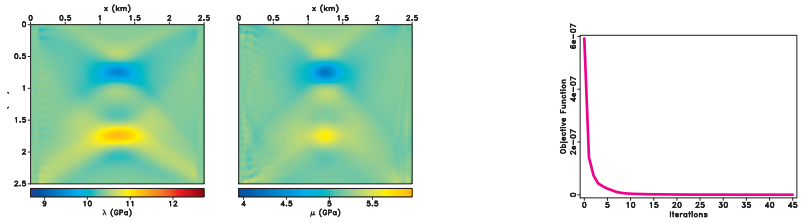
(a) (b)

Figure 3 *Example 1: (a) The models recovered with L-BFGS and the (b) inversion objective function.*



(a) (b)

Figure 4 *Example 2: (a) The models recovered with NCG and the (b) inversion objective function.*



(a) (b)

Figure 5 *Example 2: (a) The models recovered with L-BFGS and the (b) inversion objective function.*

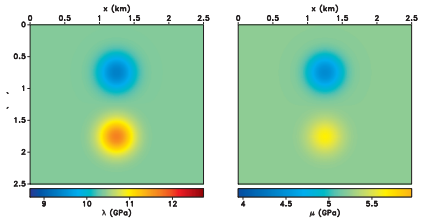
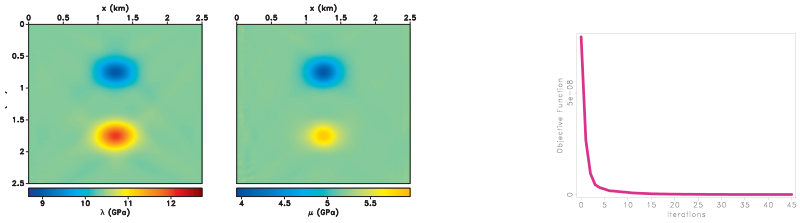
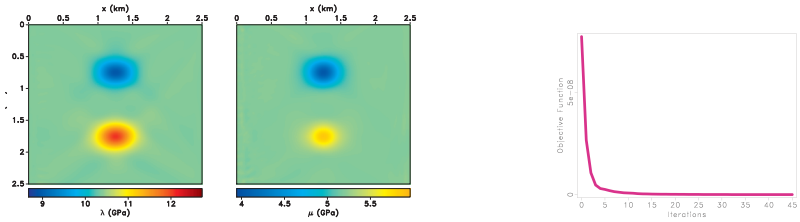


Figure 6 *Initial λ and μ models for the third and fourth example.*



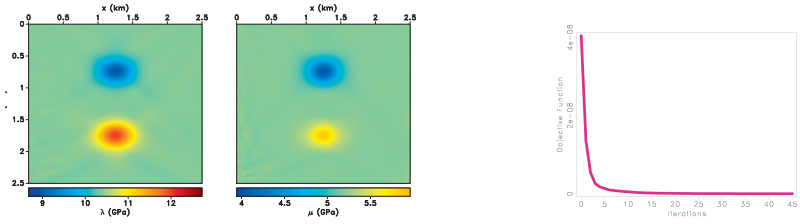
(a) (b)

Figure 7 *Example 3: (a)The models recovered with NCG and the (b) inversion objective function.*



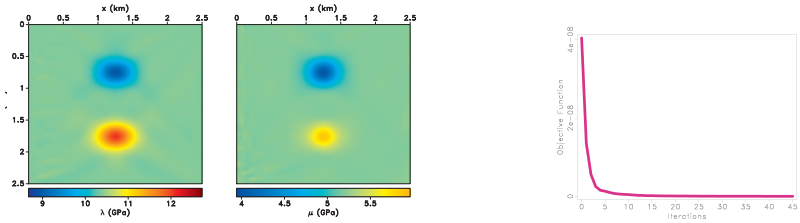
(a) (b)

Figure 8 *Example 3: (a) The models recovered with L-BFGS and the (b) inversion objective function.*



(a) (b)

Figure 9 *Example 4: (a)The models recovered with NCG and the (b) inversion objective function.*



(a) (b)

Figure 10 *Example 4: (a) The models recovered with L-BFGS and the (b) inversion objective function.*