G019

Nesterov's Method and L-BFGS Minimisation Applied to Acoustic Migration

W.A. Mulder* (Shell International Exploration & Production BV), R.H. Steenweg (Delft University of Technology) & C. Roos (Delft University of Technology)

SUMMARY

Seismic full waveform inversion based on the acoustic wave equation attempts to find the acoustic parameters of the subsurface from seismic data. Because the least-squares error between observed and modelled data has local minima, a good initial velocity model is
required. We therefore prefer the name nonlinear migration. We compared three gradient-based minimisation methods on a two-dimensional acoustic test problem: the limited-memory BFGS method with or without bounds on the model parameters, and Nesterov's method.

We found that the L-BFGS method without bound constraints performs best on the test problem, followed by Nesterov's method. However, the cost of decreasing the error by a given factor increases dramatically when the error becomes smaller.

Introduction

Acoustic full waveform inversion of seismic data using finite-difference modelling was pioneered by Albert Tarantola [1] in the 1980s. The method attempts to minimise the least-squares error between observed and modelled data. The problem size dictates the use of gradient-based rather than global optimisation methods. The least-squares error functional has local minima, causing this procedure to end up in the minimum nearest to the initial model. Therefore, the name *full waveform inversion* may be misleading. The method closely resembles migration, where first an initial velocity model is constructed and the data are then mapped to the depth domain in that given model. We therefore prefer the name *nonlinear migration*. Recently, interest in acoustic inversion has been revived by the observation that diving waves can be used for tomographic inversion of the velocity in shallow parts of the model [2, 3]. For the deeper parts, however, the problem with local minima remains.

For migration in the linear case, where the wave propagation is decoupled from the presence of reflectors by the Born approximation, convergence can be obtained in just a few iterations by using the conjugate gradient method with a proper preconditioning [2, 4]. In the nonlinear case, the inversion is more difficult as velocity, density, and wavelet need to be found simultaneously [2]. In this paper, we study the performance of three optimization methods on a synthetic nonlinear migration problem. The first one is Nesterov's method [5, 6], which is relatively unknown. The second and third are the limited-memory BFGS method [7] applied with or without bounds.

Methods

Nesterov's method is summarized in Fig. 1. The author proved that the method converges for convex functions that have a Lipschitz continuous gradient [5, 6]. The rate of convergence after k iterations is given by $|f(\mathbf{x}_N) - f(\mathbf{x}^{min})| \leq O(k^{-2})$, for a function $f(\mathbf{x})$ with a minimum at $\mathbf{x}^{min}$. The Lipshitz constant L is estimated by the algorithm. Its update, Step 2 in Fig. 1, takes the place of the more familiar line-searches in other gradient-based methods. The choice of the step-length parameter α_i is dictated by the convergence proof for convex functions. Note that α_i approaches 1 for large i. The original method started with $j = -1$ in Step 2 and used restarts each time the norm of the gradient was halved, but we found that this approach was slower for the current problem.

The BFGS method attempts to estimate the inverse of the hessian from earlier updates. For problems with a large number of unknowns, this quickly leads to storage problems. A limited-memory variant, L-BFGS, can be obtained by only keeping the last M updates (here we used 7). The price paid is a slower convergence rate for most problems, compared the the full BFGS method.

We have compared Nesterov's method to the L-BFGS algorithm for a number of convex and non-convex functions. In general, L-BFGS outperformed Nesterov method by far with convex functions, but the difference in performance becomes less pronounced for non-convex functions. Note that the convergence estimates for both methods are only valid for convex problems. Nevertheless, both still provided results for the non-convex problems we studied. In the case of multiple minima, both methods converged to the nearest local minimum for the simple test problems that we considered. The nonlinear migration problem is more challenging. There are certainly local minima and the problem may not be convex.

Examples

We present convergence results for Nesterov's method and L-BFGS with and without constraints for a nonlinear migration problem. We have created synthetic acoustic data for the 2D marine example displayed in Fig. 2 using a finite-difference time-domain method with 8^{th} order accuracy in space and 2^{nd} order in time. Data for 251 shots were generated at a 25-m spacing between a

Initialisation: set $i = 1$, take $\mathbf{y}_0$ as the initial starting point and let $t_0 = t_{-1} = 1$. Calculate $\mathbf{y}_1 = \mathbf{y}_0 + \nabla f(\mathbf{y}_0)$. Set the first approximation L_1 equal to

$$L_1 = \frac{|\nabla f(\mathbf{y}_1) - \nabla f(\mathbf{y}_0)|}{|\mathbf{y}_1 - \mathbf{y}_0|}.$$

Step 1: set $\alpha_i = \frac{t_{i-2}-1}{t_{i-1}}$ and compute $\mathbf{x}_i = \mathbf{y}_i + \alpha_i(\mathbf{y}_i - \mathbf{y}_{i-1})$, $f(\mathbf{x}_i)$ and $\nabla f(\mathbf{x}_i)$. Stop if the stopping criterion is satisfied.

Step 2: sequentially test the values $\ell = 2^j L_{i-1}$, j = 0,1,..., to find the first ℓ obeying

$$f(\mathbf{x}_i - \frac{1}{\ell}\nabla f(\mathbf{x}_i)) \leq f(\mathbf{x}_i) - \frac{1}{2\ell}|\nabla f(\mathbf{x}_i)|^2.$$

Set L_i equal to the resulting value of ℓ.

Step 3: set $\mathbf{y}_{i+1} = \mathbf{x}_i - L_i^{-1}\nabla f(\mathbf{x}_i)$ and $t_i = \frac{1}{2}(1+\sqrt{1+4t_{i-1}^2})$. Then let $i = i+1$ and continue with Step 1.

Figure 1: Nesterov's algorithm for gradient-based minimisation of a functional $f(\mathbf{x})$.

distance of 6050 to -2000 m and at a depth of 5 m below the free surface. Receivers were placed to the right of the source at offsets between 100 and 4000 m at a 25-m interval and 5 m depth.

A frequency-domain code was used for the nonlinear migration [2]. The optimisation problem is

$$\min_{\boldsymbol{\nu}} \left[\sum_{\omega} \min_{w} \left\{ \tfrac{1}{2} \sum_{s,r} |wp_{r(s)} - p_{r(s)}^{\text{obs}}|^2 \right\} + \text{penalty terms} \right]. \tag{1}$$

Here ω is the angular frequency, s enumerates the shots and $r(s)$ the receivers for a shot s. The observed pressure at receiver $r(s)$ is $p_{r(s)}^{\text{obs}}$. The modelled pressure is $wp_{r(s)}$, where $w(\omega)$ is the source signature. The unknowns are the wavelet $w(\omega)$ and the model parameters $\boldsymbol{\nu}(\mathbf{x})$. In the code, they are represented by $1/\rho(\mathbf{x})$ and $1/v(\mathbf{x})$ with ρ the density and v the velocity as a function of position $\mathbf{x}$. Penalty terms are based on Gardner's rule [8] that describes a crude relation between velocity and density, and on total-variation smoothness, cf. [9]. The minimisation with respect to the wavelet was carried out explicitly:

$$w(\omega) = < p, p^{\text{obs}} > / < p, p >, \tag{2}$$

where $< \cdot, \cdot >$ denotes the complex scalar product, involving a summation over all shots and receivers. Details can be found in [2].

The numerical tests were started from the smooth model shown in Fig. 3. Frequencies between 5 and 15 Hz with an increment of 0.5 Hz were selected for the migration. Only shots and receivers between distance of -1000 and 7100 m were used. First, ten iterations with the L-BGFS method were carried out without penalty terms. The reason for this is that we have implemented the scaling factors for the penalty terms as factors relative the leasts-squares error functional. Because the initial model is very smooth and Gardner's rule is used for the initial density model, these scaling factors are chosen way too large in the first iteration. After these ten iterations, the ratio of the least-squares functional to the data energy was 0.30. The data energy is defined by $\frac{1}{2}\sum_{\omega}\sum_{s,r}|p_{r(s)}^{\text{obs}}|^2$.

Convergence histories for the three methods are displayed in Fig. 4. The least-squares functional, normalised by its initial value, is plotted as a function of iteration cost. Here the cost of evaluating the functional is counted as one, where as the computation of the gradient costs about three times as much due to our implementation which recomputes rather than stores intermediate results. Clearly, the L-BFGS method without bounds performs best, followed by Nesterov's method which requires less storage. Note that the plot has a log-log scale. This means that the cost of reducing the error by a given factor increases dramatically when the functional becomes smaller.

The solution obtained with the unconstrained L-BFGS method is shown in Figs 5 and 6. The convergence history suggests that the least-squares error has not converged to zero. Apparently, we have reached a local minimum. The density is different from the true model, which does not obey Gardner's rule. This behaviour is typical for nonlinear migration. Because the wavelet is estimated as part of the problem, there is no absolute scaling for the density. One way to determine the scaling is by setting the density to $1\ \mathrm{g/cm^3}$ at small water depths. Adding a penalty term based on Gardner's rule is another way.

Conclusions

We have compared three optimisation methods on an acoustic nonlinear migration problem and found that limited-memory BFGS algorithm without bounds performs best, followed by Nesterov's method. L-BFGS with bounds is slowest. Nesterov's method uses the smallest amount of storage. All three methods, however, show a dramatic increase of the cost for reducing the least-squares error by a given factor when the error becomes smaller.

References

[1] Tarantola, A., 1984. *Inversion of seismic reflection data in the acoustic approximation*, Geophysics, **49**, 1259–1266.

[2] Mulder, W.A., and Plessix, R.-E., 2004. *A comparison between one-way and two-way wave-equation migration*, Geophysics, **69**, 1491–1504.

[3] Sirgue, L., and Pratt, R., 2002. *Can the sub-basalt imaging problem be solved by gradient-based waveform inversion?*, 72nd Ann. Internat. Mtg: Soc. of Expl. Geophys., 930–933.

[4] Plessix, R.-E., and Mulder, W.A., 2004. *Frequency-domain finite-difference amplitude-preserving migration*, Geophys. J. Intl., **157**, 975–987.

[5] Nesterov, Yu. E., 1983. *A method of solving a convex programming problem with convergence rate* $0(1/k^2)$, Sov. Math., Dokl. **27**, 372–376; translation from Dokl. Akad. Nauk SSSR **269**, 543–547.

[6] Nemirovski, Y., 1994. Efficient methods in Convex Programming, Chapter 10, Technion – The Israel Institute of Technology (http://iew3.technion.ac.il/Home/Users/Nemirovski.html#part4).

[7] Byrd, R.H., Lu, P., and Nocedal, J., 1995. *A limited memory algorithm for bound constrained optimization*, SIAM J. Scientific and Stat. Comput., **16**, 1190–1208.

[8] Gardner, G. H. F., Gardner, L. W., and Gregory, A. R., 1974, *Formation velocity and density – The diagnostic basics for stratigraphic traps*, Geophysics, **39**, 770–780.

[9] Vogel, C.R., 2002. Computational methods for inverse problems, SIAM, Philadelphia.

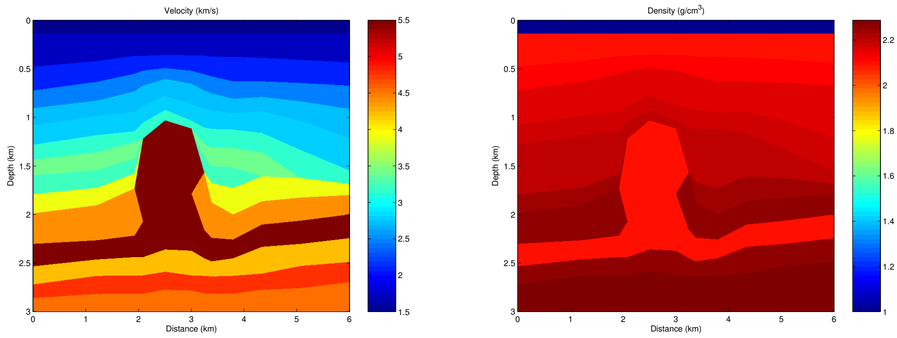

Figure 2: Model used to generate synthetic data. The velocity is shown at the left and the density at the right.

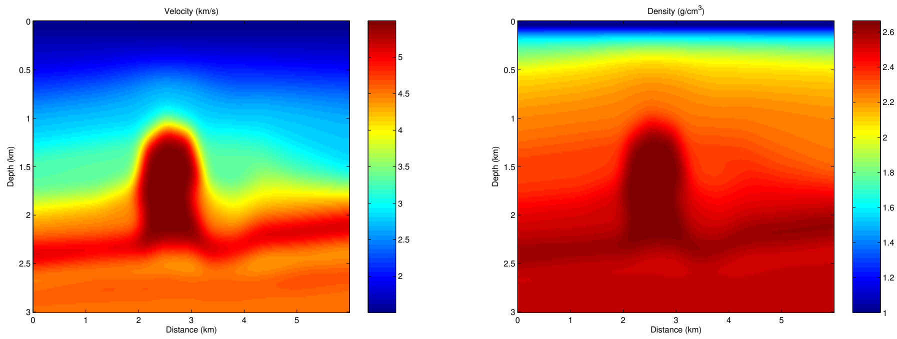

Figure 3: Initial guess of the velocity model (left) and density (right).

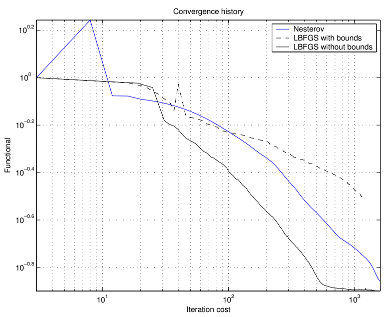

Figure 4: Convergence histories for the various methods.

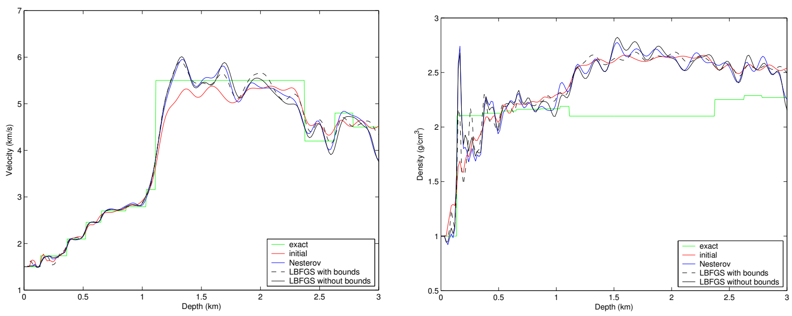

Figure 5: Cross sections of the velocity model (left) and density (right) at 3 km distance obtained by the three methods.

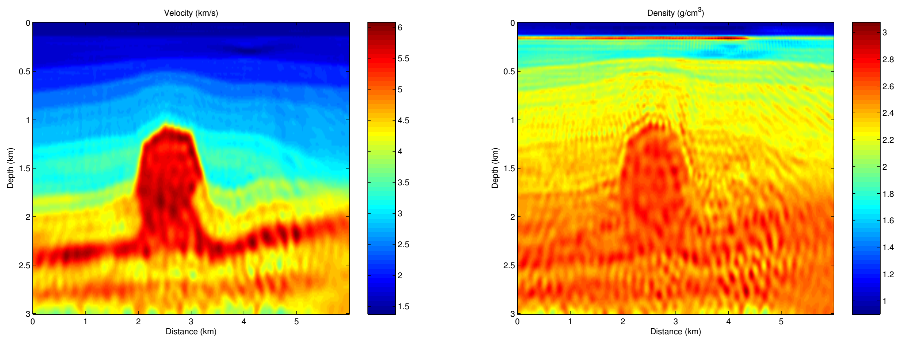

Figure 6: Solution obtained by L-BFGS minimisation without bounds.