Inverse Problems, Design and Optimization Symposium
Rio de Janeiro, Brazil, 2004

THE PERFORMANCE OF GRADIENT TYPE ALGORITHMS APPLIED TO SEISMIC WAVEFORM INVERSION

Sérgio Adriano Moura Oliveira
Laboratory of Petroleum Engineering and Exploration , LENEP/CCT
North Fluminense State University, UENF
Macaé, RJ, Brazil
Moura@lenep.uenf.br

ABSTRACT

The objective of this work is to analyze the performance of three first derivative optimization methods (the gradient, conjugate gradient and quasi Newton method) when applied in the solution of seismic waveform inversion. Important aspects as scaling, step length calculation, preconditioning and model parametrization are evaluated as well to define the best numerical implementation for this optimization based inversion. Examples involving the seismic wave velocity inversion of vertically inhomogeneous medium is set up to illustrate the convergence properties of the selected methods.

INTRODUCTION

Seismic waveform inversion is a technique designed to estimate petrophysical parameters. The data space for this method is the seismic data itself, so it takes into account information from amplitude, travel time and phase of the recorded signal. It has high potential to resolve the subsurface in great detail and is conceptually superior to the conventional seismic processing methods that are currently been used in the industry. This methodology is an optimization-based inversion that tries to obtain a model, which adequately describes the data set. It is achieved by the minimization of a cost or objective function that measures, in a given norm, the distance between calculated and observed data. This technology received great contributions from works from Lally, Tarantola and Mora who derived general procedures to obtain the derivatives of the cost function with respect to the model parameters [7, 9 and 11].

Since then, many nonlinear seismic waveform inversion algorithms and examples of applications have appeared in the exploration geophysical literature, most of then are based on the gradient and conjugate gradient methods. References to quasi Newton implementations are very rare [6] and until now, there is not a common sense about which of this classical first derivative optimization methods is the best for this particular application. Another important implementation topic is the step length calculation. In applications, it is usually done by one step, simplified routines. The main reason for this is the high computer cost for waveform inversion that typically involves thousands of unknowns (for 2-D case). However, the use of these routines may lead to an unsatisfactory convergence. Others alternatives to accelerate the method and save memory are the use of preconditioning and the representation of the unknowns parameters as combination of some convenient basis functions, what will be discussed along this work.

THE OPTIMIZATION APPROACH

The seismic waveform inversion method is based in the minimization of a cost function, which measures the distance among the calculated and observed data. This function may be defined using the common L_2, least square criteria. In a continuos functional space it is given as:

$$S(m) = \frac{1}{2}\int dx_s \int dx_r \int dt \left[P_c(x_r,t:x_s) - P_o(x_r,t:x_s)\right]^2 \quad (1)$$

Where $P_c(x_r,t:x_s)$ is the field calculated at the position x_r at the time t, due to the presence of a source at the position x_s and P_o is the observed field. The parameter $m(x)$ represents physical characteristics of the medium as density, porosity or P wave impedance. In the present work, we will concentrate in the acoustic case, so P_c and P_o represents pressure, and will obey the following equation:

$$\nabla^2 P_c - \frac{1}{c^2(x)}\frac{\partial^2 P_c}{\partial t^2} = f(x, t : x_s), \qquad (2)$$

Which is valid for a constant density medium. The velocity *c(x)* is the parameter to be estimated. In seismic exploration, the field is known along the earth surface. The source function f is also a parameter to be estimated in practical situations but here, it will be given. The gradient of S may be obtained by the formula:

$$\frac{\partial S}{\partial m} = \frac{1}{c^3(x)} \int dx_s \int_0^T dt\ \partial_t P_c(x,t : x_s)\ \partial_t y\,(x,t : x_s) \quad (3)$$

The function y is defined as:

$$y\,(x,t : x_s) = \int drG(x,-t : x_r,0) * dp(x_r, t : x_s) \qquad (4)$$

Where dp is the residual, that is, the difference between the calculated and observed seismic wavefield: $dp = P_c - P_o$. T is the final record time, G is the acoustic Green's function and * means convolution. The function y is required to obey the following final conditions (see [12] for details):

$$y\,(x,t : x_s) = 0 \qquad (5)$$
$$y\,(x,T, x_s) = 0$$
$$\partial_t y\,(x,T : x_s) = 0$$

NUMERICAL IMPLEMENTATION

Some considerations of the numerical implementation of nonlinear waveform inversion algorithms will be done in this section. The first is related to the choice of the type of first derivative optimization technique. Conjugate gradient is supposed to gives better result than the simple gradient, without any significant increase of the computations. However some results seems to contradict it (see [3], for example). Quasi Newton methods may converge faster but requires extra memory for vector storage. Step length calculations is also a topic that deserves attention. The question here is the effectiveness of approximate procedures commonly used to save computations. Other topics are related to scaling, preconditioning, parametrization and numerical solution of equation (1).

Discretizing the problem

Practical applications demand the problem to be solved in a discrete space for computations. A fundamental step of its solution is the calculation of the wavefield. If we think in a rectangular mesh for the subsurface discretization, the method of finite differences is a good option for this task. Here the problem will be restrict to 2-D and the following difference scheme will be used for solution of equation (1):

$$
\begin{aligned}
p^{l+1}_{m,n} &= (2 - 5a^2_{m,n})\,p^{l}_{m,n} \qquad (6)\\
&+ \frac{a^2_{m,n}}{12}\{16[p^l_{m+1,n} + p^l_{m-1,n} + p^l_{m,n+1} + p^l_{m,n-1}]\\
&- [p^l_{m+2,n} + p^l_{m-2,n} + p^l_{m,n+2} + p^l_{m,n-2}]\}\\
&- p^{l-1}_{m,n} + f^{\,l} + O(h^4, dt^2)
\end{aligned}
$$

Where $x=m.h$, $z=n.h$, $t=l.dt$ and $a_{m,n}=c_{m,n}dt/\ h$. The scheme is stable provided $a_{m,n}<(3/8)^{1/2}$ [8]. The calculation of y may be done using the same scheme, but at each receiver point x_r we set a source whose time function is the reversal of the data residual dp. Each iteration of the waveform inversion requires equation (1) to be solved many times for the evaluation of the cost function, gradient and step length calculations. So the computational cost becomes very high if there are many variables and parameters to be inverted in the problem.

Gradient method and preconditioning

The gradient methods provides an iterative solution that may be given by the formula:

$$m_{i+1} = m_i + a_i A_i s_i \qquad (7)$$

Where s_i is a descent direction, a_i is the step length and A_i is a positive definite matrix, known as the preconditioning matrix, that is used to accelerate the convergence. In the classical Gradient and Conjugate gradient this matrix is keep constant along the iterations. In this work, it is a diagonal matrix, which acts to compensate the amplitude decay caused by geometrical spread. In the quasi Newton approach, A is modified at each iteration in order to approximate the Hessian, or it's inverse. Here the quasi Newton method was implemented according to the BFGS formula, due to Broyden, Fletcher, Goldfarb and Shanno [1,2]. Although, this method is supposed to achieve the fast convergence of Newton's method as A approximate the inverse Hessian, it requires storage of, at least, a gradient and a direction vector s after each iteration to do the actualization of A. It becomes a serious difficult for large scale applications. This problem may be partially circumvented by the use of the limited memory BFGS method [10].

Step length calculations

The step length may have a considerable effect on the performance of the gradient optimization method. The strategy adopted here is to use a line search routine based on cubic approximations for S. A reasonable minimum point is found, only if it reaches the Wolf-Powell conditions:

$$I)\quad S(a) \leq S(0) + a r \left.\frac{dS}{da}\right|_{a=0} \qquad r \in (0, \frac{1}{2})$$

$$II)\quad \left|\frac{dS}{da}\right| \leq -s \left.\frac{dS}{da}\right|_{a=0} \qquad s \in (r, 1)$$

Condition I, exclude the right-hand extreme of the search interval and condition II impose a sufficient decrease on the slop. The precision of the line search is controlled by the parameters r and s. Before the line search routine, it is necessary to implement a *bracket* procedure to find an interval that contains an acceptable point. The rigorous routine described above requires many evaluations of the cost function and its gradient. Simple procedures for fast evaluation of the optimum step length may be derived, by one step line search based on a quadratic approximation for the objective function along the search direction (see [3] for example). These procedures are usual in the implementation of seismic waveform inversion algorithms. It saves a lot of computation, but may lead to negligible reduction of the cost function at each iteration. A comparison among algorithms with rigorous and approximate step length calculation is made in the numerical test section.

Scaling

Another important practical aspect for implementation of optimization algorithms is the correct scaling of the variables. This is fundamental to avoid work with very small or very big numerical values, what may cause truncation errors. Since the final result is not supposed to be very far from the initial model, it is possible to establish reasonable bounds for the variables: $(b_k\,^3c_k\,^3a_k)$. Then the following scaling procedure is adopted in this work:

$$C_k = \frac{2c_k}{b_k - a_k} - \frac{a_k + b_k}{b_k - a_k} \tag{8}$$

So the scaled variable C_k will always belong to the interval $(1\,^3C_k\,^3\text{-}1)$. This procedure showed to be very important for the conjugate gradient method and fundamental to the quasi Newton.

Parametrization

The numerical solution of equation (1), by the scheme (6), requires the discretization of the subsurface in a set of nodal points, where the velocity should be given. So, it seems to be natural to use the present inversion method, to update the velocity at each of these nodal points used in the modeling step. However, this procedures generates a lot of variables and leads to non-uniqueness for the inverse problem, since the grid size is normally smaller than the details that we hope to reconstruct from the data. So it is not necessary to use the nodal modeling parameters as inversion parameters. This may be defined as combination of some convenient basis functions:

$$m_j = m(z_j) = \sum_{i=1}^{N} n_i\, a^i(z_j) \tag{9}$$

Where $a^i(z_j)$ is one of the N basis function and n_i are the new parameters of the problem. The gradient vector in the new parametrization is calculated by the simple procedure:

$$\frac{\partial S}{\partial n} = \frac{\partial S}{\partial m}\frac{\partial m}{\partial n}, \tag{10}$$

Where ¶m/¶n is a matrix, whose elements are c_{ij}=¶m_j/¶n_i. In this work a block parametrization was adopted, in this case a^i are simple box functions. This kind of parametrization is convenient if we think that the earth is a 1-D layered medium, where the velocity is constant at each layer. Other kinds of parametrization as splines, if ones want smooth results, may be implemented using equation (10).

NUMERICAL TESTS

Numerical tests are set up in order to check the procedures described in the last sections and the performance of the classical gradients methods. These tests are based in the inversion of multichannel synthetic data (only one shot), acquired over a layered medium (figure 1). This data consists of a 48 traces separated by 24 meters, with 800 samples generated by a 0.0015 seconds sampling interval. An initial model is necessary in order to initialize the iterative minimization. This initial model is also shown in figure (1), it preserves the tendency of the increase of the velocity with depth present in the first five layers. The model was discretized by 170 velocity samples, with 8 meters interval. In the first set of tests, parametrization was not used, so each one of these sample is an inversion variable to be estimated.

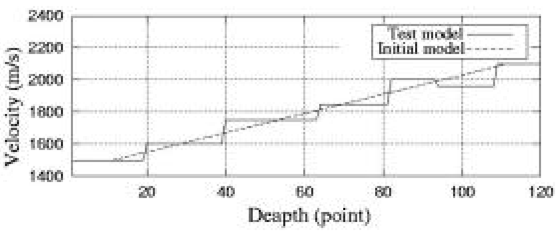

Figure 1- Test model and the initial model used in the iterative minimization algorithms.

The first test shows the advantage of the use of preconditioning. Figure (2) shows the convergence histories of the gradient method implemented without and with preconditioning to compensate amplitude decay caused by geometrical spreading. This simple procedure is fundamental to increase the convergence speed.

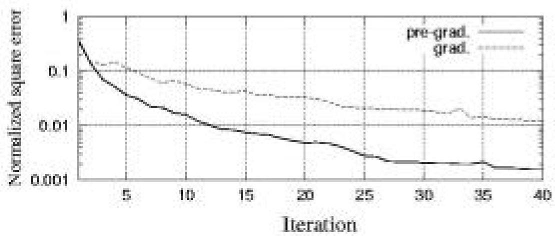

Figure 2 – Convergence of the Gradient method and preconditioned Gradient method.

The second test compares rigorous and approximated step length strategies. For the rigorous one, we adopted r=*0.001* and s=*0.1* , these values provide a good precision in the line search routine. The approximated strategy consists of only one step based on a quadratic interpolation for the objective function along the search direction. It is clear in figure (3) that this approximate step length routine results in a very slow convergence when compared to the more rigorous procedure.

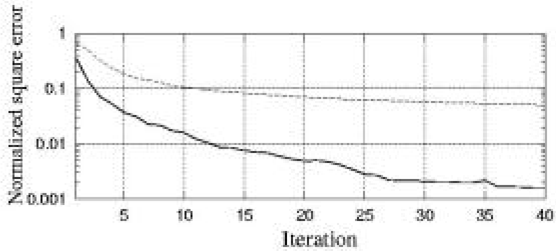

Figure 3 – Convergence of the preconditioned Gradient method with rigorous (solid line) and approximate step length calculations.

Some implementation details of the gradient algorithms should be explained before its comparison. The Polak-Ribiere formula (see [4] for example) was used for the conjugate Gradient method, with preconditioning. The quasi Newton method needs an initial approximation for the inverse Hessian matrix A_0 . It is usual to take the identity matrix for this, but here the same preconditioning matrix described early will be used for this approximation since it showed to be a good practice along the numerical experiments. The convergence history of the three methods is depicted in figure (4).

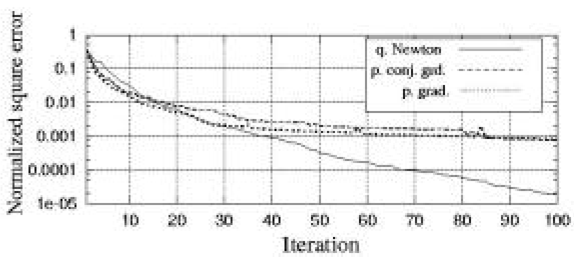

Figure 4 – Convergence history of the preconditioned Gradient, preconditioned Conjugate Gradient and BFGS quasi Newton Methods.

It is interesting to see that the preconditioned Conjugate Gradient method did not improve the results when compared to the simple preconditioned Gradient. These methods did not make any considerable progress after iteration 40. The BFGS quasi Newton, instead, continued to make consistent progress until when it was tested in iteration 150, when the normalized error achieved 10^{-6}. The final results obtained by the Gradient and Conjugated Gradient methods are shown in figure (5).

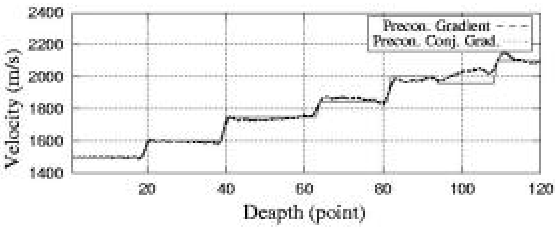

Figure 5 – Final results of the preconditioned Gradient and conjugate Gradient method.

These results, in general, are in good agreement with the true model, although they fail after layer 5. This occurs because the inversion in the velocity growth tendency given by the initial models is not preserved in layer 6. Reflection seismic waveform inversion using gradients methods, is known to be very sensitive to local minimums, especially if low frequencies are absent of the data [5]. So it depends on an accurate initial model, that contains at least the smooth character of the true model, in order to give a good result. However, the BFGS quasi Newton method seems to be more robust then the Gradient and Conjugate Gradient. Its velocity reconstruction is very good for all layers. The inversion process recovers the high frequency characteristics of the model first, so the interfaces are the first feature that appears in the solution. The convergence velocity decreases in the reconstruction of the smooth (low frequency) features of the model. Some oscillation around the interfaces may also be saw in the solutions, this resembles the Gibbs Phenomenon from Fourier analysis.

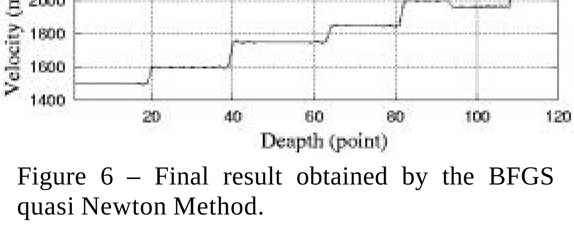

Figure 6 – Final result obtained by the BFGS quasi Newton Method.

The last numerical testes show the advantage of the use of the parametrization. Figure (7) shows that only 35 iterations of the BFGS algorithm using block parametrization are sufficient to recover the main features of the model. For this test, each block has 24 meter, so the number of inversion variables was reduced from 170 to 56.

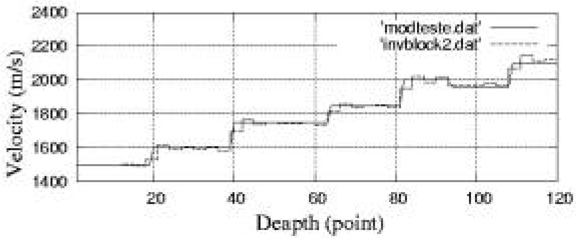

Figure 7 – Result of 35 iterations of the BFGS quasi Newton method, using block parame-trization.

The advantage of this type of parametrization for layered medium becomes more evident in the numerical test shown in figure (8). It is a reconstruction of model with only one thin layer. For this example, a constant velocity medium (*c=2000 m/s*), was taken as initial model. The inversion procedure using the modeling variables as inversion variables had to deal with 90 variables, this number was reduced to 18 in the block parametrization scheme. After 20 iterations, inversion using point parametrization is far from the real model, while the result obtained using blocks gave an almost perfect result. Block parametrization also tends to eliminate the oscillations around interfaces present in the solutions obtained using point parametrization.

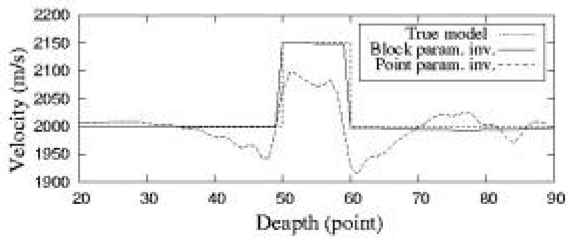

Figure 8 – Reconstruction of a thin layer model using 20 iteration of the point and block parametrization .

CONCLUSIONS

The BFGS quasi Newton method seems to be the better first derivative optimization method for the non-linear seismic waveform inversion problem. For this method, the preconditioning matrix is a better option for the initial inverse

Hessian then the Identity matrix. Approximate routines for step length calculations, although commonly used in practical applications, should be avoided due to the very slow convergence achieved due to under estimate of the steps. Rigorous step length calculation is preferable even it becomes the heaviest computational task of the algorithm. Parametrization by a combination of orthogonal or interpolating functions is a better option then to use the model variables as the inversion variables. The block parametrization, using simple box functions adopted in this work, showed to be very convenient for the layered models used as examples. This procedures acts to regularize the inversion, decreasing the number of variables and iterations necessary to achieve a good reconstruction.

REFERENCES

1. R. Fletcher, *A new approach to variable metric algorithms*. Comput. J.,13, 317-322 (1970)
2. R. Fletcher, *Practical methods of optimization*. Wiley Inter. Science publication, 1987.
3. O. Gauthier, J. Virieux and A. Tarantola, *Two-dimensional nonlinear inversion of wave forms: Numerical results*. Geophysics, 51, 1387-1403 (1986)
4. P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*. Academic Press, 1986.
5. M. Jananne, W. Beudoun, E. Crase, D. Cao, Z. Koren, E. Landa, M. Mendes, A. Pica, M. Noble, G. Roeth, S. Singh, R. Snieder, A. Tarantola, D. Trezeguet and M. Zie, *Wavelentghs of the earth structures that can be resolved from seismic reflection data.* Geophysics, 54, 906-910 (1989)
6. P. Kolb, *Pre-stack inversion of a 1-D medium*. Proc. IEEE, 74, 498-508 (1986)
7. P. Lally, *The seismic inverse problem as a sequence of before stack migrations*. Inverse scattering, Theory and applications, Bednar, Robinson and Weglein, Eds. Philadelphia, PA: SIAM, 1984.
8. A. R. Mitchell, *Computational methods in partial differential equations* . Jhon Wiley & Sons, 1969
9. P. Mora, *Nonlinear two-dimensional elastic inversion of multoffset seismic data*. Geophysics, 52, 1211-1228 (1987)
10. J. Nocedal, *Updating the quasi-Newton matrices with limited storage*. Math. Comp, 95, 339-353, (1980).
11. A. Tarantola, *Inversion of seismic reflection data in the acoustic aproximation*. Geophysics, 49, 1259-1266, (1984).
12. A. Tarantola, *Inverse problem theory-methods for data fitting and model parameter estimation*. Elsvier Science Publ. Co., 1987