Using band-tridiagonal preconditioners for 1D seismic waveform inversion

Milton J. Porsani and Saulo Pomponet de Oliveira CPGG/UFBA, Brasil.

This paper was prepared for presentation at the 9^{th} International Congress of The Brazilian Geophysical Society held in Salvador, Brazil, 11-14 September 2005.

Abstract

Geophysical inversion is quite a challenge. Many different methods and strategies have been developed in the past decades. Global methods like simulated annealing and genetic algorithms do not depend on an initial model, although they may be time consuming using a huge number of fitness evaluations. Local methods such Gauss-Newton require a 'good' start model and may be not convergent to the global minimum of the objective function. In this paper we evaluate the use a band-tridiagonal preconditioner in the multiple reweighted least square inversion method (MRLS) to the seismic waveform inversion. The MRLS inversion method is a type-gradient method that generates many candidate models for each sensitivity matrix, exploring the model space in a very effective way. A numeric example shows that the inverted model has better fitness and may be obtained with less computational CPU time, when compared to the results given by the Gauss-Newton approach.

Introdução

Due to the nonlinear nature of the seismic waveform inversion problem, global optimization methods such as simulated annealing and genetic algorithm have been applied to these problems (Porsani et al, 2000, Sambridge and Drikoningen, 1992, Sen and Stoffa, 1991). Here we evaluate the use of the preconditioners in the linearized inversion method named as the multiple reweighted least square method (MRLS) proposed by Porsani et al. (2001) to seismic waveform inversion with the aim of achieving greater accuracy and reducing computational cost.

Following we summarize the MRLS method.

Consider the geophysical model of N parameters represented by $\mathbf{m} = (m_1, \ldots, m_N)^T$. Let us write the deviations between the observed data and the computed one, $\Delta d(\mathbf{m}, x) = d(x) - F(\mathbf{m}, x)$, as a power function in Δd defined as,

$$f(\mathbf{m}, p, x) = |\Delta d(\mathbf{m}, x)|^{p/2} \; . \tag{1}$$

We now linearize $f(\mathbf{m}, p, x) \approx \tilde{f}(\mathbf{m}, p, x)$, by expanding in a Taylor series about the current model, $\mathbf{m}_k$ by retaining only the first derivative term. By using vector and matrix notation the resulting equation can be represented as,

$$\mathbf{f}(\mathbf{m}, p, \mathbf{x}) \approx \tilde{\mathbf{f}}(\mathbf{m}, p, \mathbf{x}) = \mathbf{f}_{pk} - \mathbf{W}_{pk}\mathbf{G}_k\Delta\mathbf{m} \, , \tag{2}$$

where $\mathbf{G}_k$ is the sensitivity matrix, $\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}_k$ is the parameter vector correction, $\mathbf{f}_{pk} = \mathbf{f}(\mathbf{m}_k, p, \mathbf{x})$ is the vector of the data deviation at the observation positions $x_i \, , i = 1, \cdots, M$, and $\mathbf{W}_{pk}$ is a weighting diagonal matrix whose elements may be computed as,

$${}^k w_{ii} = \frac{p}{2} sgn(\Delta d(\mathbf{m}_k, x_i))|\Delta d(\mathbf{m}_k, x_i)|^{p/2-1} \; . \tag{3}$$

We note that $\tilde{\mathbf{f}}(\mathbf{m}, p, \mathbf{x})$ in Eq. (2) is a linear function of $\Delta\mathbf{m}$, and a solution can be obtained by solving the linear system of equation in the LS sense,

$$(\mathbf{W}_{pk}\mathbf{G}_k)\Delta\mathbf{m}_k = \mathbf{f}_{pk} \, . \tag{4}$$

We note that for small Δd and $p < 2$, ${}^k w_{ii}$ may become very large. To avoid the implied singularity in the computational algorithm a dumped LS approach may be used. For any value of p ($p \neq 0$) the dumped LS solution can be obtained and the expression for updating the current model may be written,

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \mathbf{A}^{-1}\mathbf{b} \, , \tag{5}$$

where,

$\mathbf{A} = \mathbf{A} + \lambda\mathbf{I} \quad , \quad \mathbf{A} = \mathbf{G}_k^T\mathbf{W}_{pk}^T\mathbf{W}_{pk}\mathbf{G}_k$

$\mathbf{b} = \mathbf{G}_k^T\mathbf{W}_{pk}^T\mathbf{f}_{pk}$

λ is a regularization factor and $\mathbf{I}$ is the identity matrix

The LS approach based on Eq. (5) can be seen as an iterative procedure where the row i of the sensitivity matrix is weighted by the factor ${}^k w_{ii}$ (Eq. 3) which is a function of the deviation between the observed geophysical data and the ones computed from the current model $\mathbf{m}_k$. By setting $p = 1$, Eq. (5) is reduced to an iterative equation for the LS method based on the L_1 norm. By setting $p = 2$ into the Eq. (5) $\mathbf{W}_{pk}^T\mathbf{W}_{pk} = \mathbf{I}$, $\mathbf{I}$ being the identity matrix,

$\mathbf{W}_{pk}^T\mathbf{f}_{pk} = \Delta d(\mathbf{m}, x)$ and it is reduced to the classical L_2 norm solution for the LS method,

$$\mathbf{m}_{k+1} = \mathbf{m}_k + (\mathbf{G}_k^T\mathbf{G}_k)^{-1}\mathbf{G}_k^T\Delta d(\mathbf{m}_k, \mathbf{x})\,.$$

Eq. (5) provides more flexibility to the LS method. During the inversion it allows us to weight the deviation by using a fixed or variable exponent p in each iteration. For a given sensitivity matrix the update of the current model depends on p. In general we do not know the probability distribution function of the error in the data and consequently there is no reason to prefer one to another norm, say L_2 ($p = 2$) or L_1 ($p = 1$) or any other fixed value of p to drive the inversion method. For many inversion problems in geophysics the computational cost for solving the linear system is significantly smaller than the one for generating the sensitivity matrix $\mathbf{G}_k$. For those inverse problems the use of Eq. (5) with multiple values of p could be of extraordinary importance once the sensitivity matrix does not depend on p, and once formed cam be used for many values of p.

The inversion conducted by means of Eq. (5) by using multiple values of p in each step was named as the multiple reweighted LS (MRLS) inversion method.

From Eq. (5) we can observe that the value p only affects the $\mathbf{W}_{pk}$ matrix and the vector $\mathbf{f}_{pk}$. For each current model and for each value of p a different solution can be derived. This means that a different point in the model space can be found to update the current model. By using N_p different values of p at each iteration, and by solving Eq. (5) N_p times, N_p different possibilities to update the current model can be calculated. By evaluating the fitness value for each candidate model we can select as the current model the one which presents the best performance. The full steps of the MRLS algorithm and numerical examples are presented in Porsani et al. (2001), Santos, et al.(2005).

Fitness Evaluation

In the numerical examples presented, for the fitness evaluation we have used the function given below.

$$\Phi(\mathbf{m}) = \frac{2\sum_{j=1}^{N} d(x_i)F(\mathbf{m}, x_i)}{\sum_{j=1}^{N} d(x_i)^2 + \sum_{j=1}^{N} F(\mathbf{m}, x_i)^2}. \qquad (6)$$

Band-tridiagonal preconditioners

Let us to consider the system of normal equation to be solved,

$$\mathbf{A}\mathbf{h} = \mathbf{b}. \qquad (7)$$

Matrix $\mathbf{A}$ may be written as $\mathbf{A} = \mathbf{T} + \mathbf{R}$ where $\mathbf{T}$ is the band-tridiagonal matrix that has a non-zero main diagonal principal $\{\mathrm{a}\}$ e non-zero co-diagonals and $\{\mathrm{c}\}$ defined as follows:

$$\mathbf{T}_{i,j} = \begin{cases} A_{i,j} & , \quad |i-j| = 0, \\ A_{i,j} & , \quad |i-j| = m \\ 0 & , \quad \text{otherwise.} \end{cases}$$

The Levinson recursion may be exploited to develop efficient algorithm even when the coefficient is a full matrix (Porsani and Ulrych, 1991) without Toeplitz structure. By using an efficient and recursive Levinson-type algorithm (Porsani and Oliveira, 2005) we can find a triangular matrix $\mathbf{F}$ such that,

$$\mathbf{F}^T\mathbf{T}\mathbf{F} = \mathbf{D}$$

where $\mathbf{D}$ is a diagonal matrix. The matrix $\mathbf{F}$ may be used to preconditioner the original system as following,

$$\begin{aligned}(\mathbf{F}^T\mathbf{A}\mathbf{F})(\mathbf{F}^{-1}\mathbf{h}) &= \mathbf{F}\mathbf{b} \\ \mathbf{A}_1\mathbf{h}_1 &= \mathbf{b}_1\end{aligned} \qquad (8)$$

where,

$$\begin{cases} \mathbf{A}_1 = \mathbf{F}^T\mathbf{A}\mathbf{F} \\ \mathbf{h}_1 = \mathbf{F}^{-1}\mathbf{h} \\ \mathbf{b}1 = \mathbf{F}\mathbf{b} \end{cases}$$

The Jacob normalization

A diagonal matrix $\mathbf{E}$ may be formed taking the square root of the elements of the diagonal of the matrix $\mathbf{A}_1$. Analogously to Eq. (8), matrix $\mathbf{E}$ may be used to normalize the system $\mathbf{A}_1\mathbf{h}_1 = \mathbf{b}_1$, generating a new system,

$$\widehat{\mathbf{A}}\widehat{\mathbf{h}} = \widehat{\mathbf{b}} \quad , \quad \begin{cases} \widehat{\mathbf{A}} = \mathbf{E}^T\mathbf{A}_1\mathbf{E} \\ \widehat{\mathbf{h}} = \mathbf{E}^{-1}\mathbf{h}_1 \\ \widehat{\mathbf{b}} = \mathbf{E}\mathbf{b}_1 \end{cases} \qquad (9)$$

We remark that the matrix $\widehat{\mathbf{A}}$ has unitary elements into the main diagonal. By solving equation (9) the solution vector $\mathbf{h}$ of the original system may be obtained,

$$\mathbf{h} = \mathbf{F}\mathbf{E}\widehat{\mathbf{h}} \qquad (10)$$

The preconditioned conjugate gradient algorithm

The preconditioned conjugate gradient (PCG) algorithm, used to solve Eq. (10), is defined by the following iterative equations,

$$\begin{aligned}\mathbf{h}_{j+1} &= \mathbf{h}_j + \alpha\mathbf{v}_j \\ \mathbf{v}_{j+1} &= \widetilde{\mathbf{A}}^{-1}\mathbf{g}_{j+1} + \beta\mathbf{v}_j\end{aligned}$$

where $\mathbf{v}_j$ is an auxiliary vector and $\mathbf{g}_{j+1} = -\widehat{\mathbf{b}} + \widehat{\mathbf{A}}\mathbf{h}_{j+1}$ is the residual of the approximate solution $\mathbf{h}_{j+1}$. The parameter α on equation for $\mathbf{h}_{j+1}$ is calculated such that $\mathbf{g}_{j+1}^T\mathbf{v}_j = 0$. The matrix $\widetilde{\mathbf{A}}^{-1}$ represents the approximate inverse of the matrix $\widehat{\mathbf{A}}$. The parameter β on equation for $\mathbf{v}_{j+1}$ ensures that $\mathbf{v}_{j+1}^T\widehat{\mathbf{A}}\mathbf{v}_j = 0$. One may show that if $\widetilde{\mathbf{A}}^{-1} = \widehat{\mathbf{A}}^{-1}$, the PCG method converges in a single step. On the other hand, if $\widetilde{\mathbf{A}}^{-1} = \mathbf{I}$ then the algorithm PCG reduces to the conjugate gradient (CG) algorithm, where the solution of a linear system of size N can be found in N steps.

Iterative refinement into the PCG algorithm

As previously mentioned, the number de steps of the PCG algorithm is directly related with the degree of approximation used in the solution of the system

$$\mathbf{A}\mathbf{q}_{j+1} = \mathbf{g}_{j+1}.$$

Considering $\widehat{\mathbf{A}} = \mathbf{I} + \widehat{\mathbf{R}}$, the inverse of $\widehat{\mathbf{A}}$ may be approximated by using polynomial division. The following iterative equation may be be derived,

$$\mathbf{q}_{j+1,k+1} = \mathbf{g}_{j+1} - \mathbf{q}_{j+1,k} - \widehat{\mathbf{R}}\mathbf{q}_{j+1,k} \qquad (11)$$

$k = 0, \ \ldots$ Niter. Note that for $k = 0$, $\mathbf{q}_{j+1} = \mathbf{g}_{j+1}$ and the algorithm reduces to the CG algorithm. This procedure fits in the framework of the Iterative Refinement (Kincaid and Cheney, 1996). To get convergence the radius of $\widehat{\mathbf{R}}$ should be less than 1.

Methodology

The inversion was performed using the MRLS method testing three types os preconditioners to solve the systems of NE generated at each step of the algorithm:

- Direct solution of the NE without preconditioning (indicated by CG in figure 2);
- Solution of the NE preconditioned with the Jacob normalization (indicated by CG(1);
- Solution of the NE with the band-tridiagonal preconditioner for m=1 and m=10 (indicated by CG(1,2) and CG(1,10), respectively).

The implementation of the MRLS algorithm is summarized below:

(i) We start with a initial model and set the initial range for the p values, $1 < p_i < 4$;

(ii) generate the sensitivity matrix;

(iii) for each p_i value we form the corresponding linear system and solve it in the LS sense;

(iv) use the solution to generate a candidate model to substitute the current ones;

(v) evaluate the fitness of the candidate models;

(vi) substitute the current model by the candidate of best fitness;

(vii) use the value of p associated to the best candidate model (p_c) to update the range of p values $0.8p_c < p_i < 1.2p_c$;

(viii) restart from step (iii) and continue up to a convergence or a stop criterium is satisfied.

In the numeric example presented we have used a fixed number of 7 values of p. At each step of the MRLS method the system of NE, $\mathbf{A}\mathbf{h} = \mathbf{b}$ is transformed into $\widehat{\mathbf{A}}\widehat{\mathbf{h}} = \widehat{\mathbf{b}}$ by applying the Jacob or the preconditioners:
band-tridiagonal plus the Jacob. The modified system is solved using the CG algorithm. The LS desired solution for the original system is obtained using Eq. (10).

Numerical results

In this optimization problem plane wave seismograms are modelled (Kennett, 1983) and the model parameters to be found are the P wave velocity, impedance and Poisson's ratio.

Figure 1 shows a synthetic seismogram associated with a 10 layers model (total of 30 parameters), corresponding to the 'observed data' to be inverted. Forty plane wave seismograms for ray parameters of 0.0 to $0.39\ s/km$ every $0.01\ s/km$ with a bandwidth of 8 to $80Hz$ were used in the inversion. The modelling included only primary compressional wave reflections but did include loss due to shear conversion. No internal multiples or converted phases were included as observed arrivals.

The effect of the preconditioners into the CG algorithm, during the solution of a particular system of NE, is illustrated in the Figure 2. The reduction of the norm of the residual vector $||\mathbf{g}_j||^2$, is displayed as function of the iterations of the CG algorithm. The iterations stops when the norm of the residua reach 10^{-20}. The solution using the brute CG algorithm (without preconditioner) require more then 100 iterations and the convergence curve is very unstable. The other curves shows the stability and the efficacy of the preconditioners, and the solution is obtained with less than 15 iterations.

The effect of the number of iteration used in Eq. (11) (degree of refinement) coupled with the use of the preconditioners is illustrated by figure 3. The number of iteration in the PCG algorithm decrease when we increase iteration (refinement) to calculate the auxiliary vector $\mathbf{q}_{j+1}$, which is equivalent to improve the approximation of the inverse.

A full inversion was performed using the MRLS algorithm exploring 7 differents candidates models for step as described previously. The convergence is compared with the conventional Gauss-Newton (GN) approach using fixed value of p ($p = 2$). Figure 4 shows the convergence plots for the best candidate models generated by the MRLS algorithm (L_p) and the GN (L_2) approach. The fitness value (eq. (6)) of the best model is plotted for each step of the MRLS algorithm. Note the improved performance of the L_p fitness compared to L_2.

The CPU time spent on the MRLS inversion is presented in table 1. The use of preconditioners reduces the computational cost in all three cases tested, compared with the case where no preconditioner was not used, i.e.

the CG case.

METHOD	PRECONDITIONER	CPU TIME (s)
CG		286.33
CG(1)	Jacob normalization	280.27
CG(1,2)	Tridiagonal (m=1) + Jacob	277.95
CG(1,10)	Tridiagonal (m=10) + Jacob	282.59

Table 1: CPU time spent on the MRLS inversion

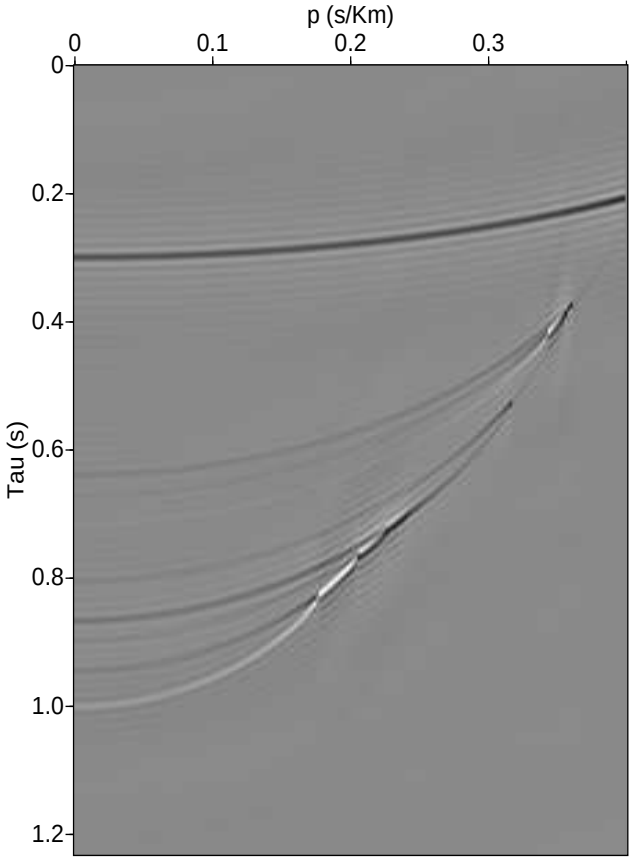

Figura 1: Synthetic plane wave seismograms used as the 'observed' data.

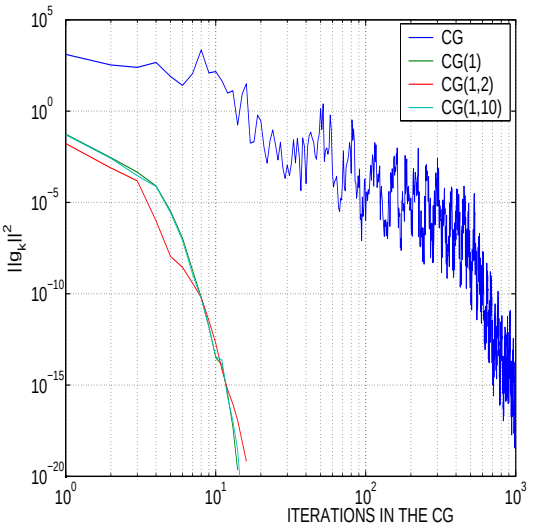

Figura 2: The effect of the preconditioners into the CG algorithm, during the solution of a particular system of NE. The norm of the residual vector is plotted as function of the iteration of the CG algorithm.

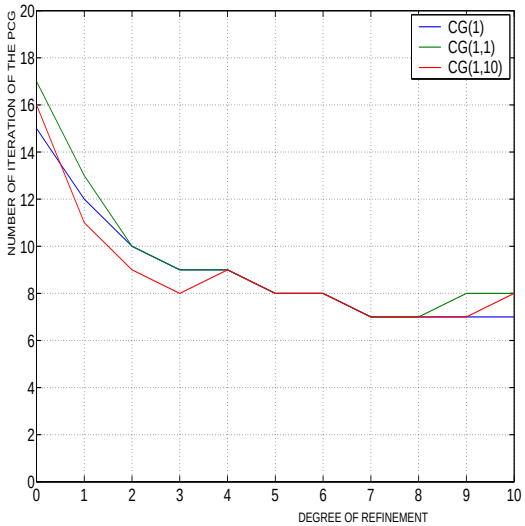

Figura 3: Reduction of the number of iterations of the PCG algorithm as function of the degree of refinement used to evaluate Eq. (11). Note that for $k = 0$ we have the CG algorithm.

Conclusions

We evaluate the use of band-tridiagonal preconditioners in the solution of systems of normal equations associated with the 1D seismic waveform inversion based on the multiple reweighted least-square method. The numeric experiment demonstrate that the preconditioners gives numeric stability to the solution and reduce the number of iteration in the conjugate gradient algorithm. The full inversion shows a small computational reduction in the CPU time. The results are promising to encourage additional research.

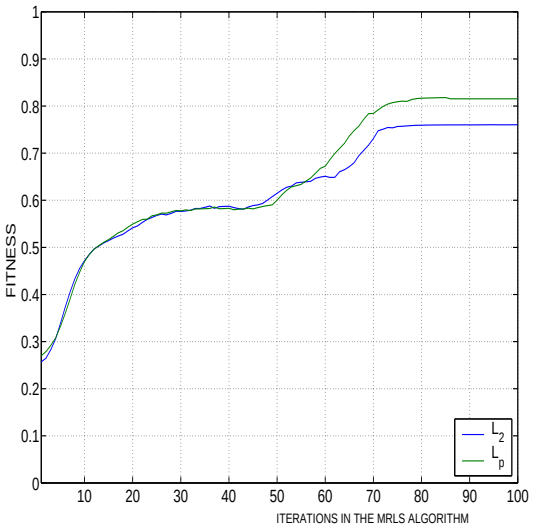

Figura 4: Convergence plots of the fitness values for the best candidate models obtained with the MRLS algorithms for the L_p and L_2 fitness. Note the improved performance of the L_p fitness compared to L_2.

Acknowledgements

The authors thank FINEP, PETROBRAS, CNPQ, PGS and CAPES.

References

Kennett, B. L., N., 1983, *Seismic wave propagation in stratified media*, Cambridge University Press.

Kincaid, D. and Cheney, W. 1996. Numerical analysis, 2nd. ed., Brooks/Cole, Pacific Grove, USA.

Porsani, M. J. and Oliveira, P. S., 2005, Modified Incomplete Cholesky preconditioners for 2D resistivity modeling (accepted to be presented at the eaeg-Madri, june).

Porsani, M. J. and Ulrych, T. J., 1991, Levinson-type extensions for non-Toeplitz systems. IEEE Transaction on Signal Processing, v. 39, n. 2, p. 366-375.

Porsani, M. J. and Oliveira, S. P., 2005, The use of band-tridiagonal preconditioners for seismic waveform inversion (being prepared).

Porsani, M. J. and Stoffa, P. L, Sen, M. Chunduru, R. K., 2000, Fitness functions, genetic algorithms and hybrid optimization in seismic waveform inversion, Journal of Seismic Exploration, 143-164 , **9**.

Sambridge, M., and G. G. Drikoningen, 1992, Genetic algorithms in seismic waveform inversion, Geophys. J. International, 109, 323-342.

Santos, A. B., Sampaio, E. E. S., and Porsani, M. J., 2005, A robust two-step inversion of complex magnetotelluric apparent resistivity data. Stud. Geophys. Geod., V. 49, pp. 109-125.

Sen, M. K., and Stoffa, P. L., 1991, Non-linear one-dimensional seismic waveform inversion using simulated annealing, *Geophysics,* **56**, 1624-1638.