Solving least squares Kirchhoff migration using multigrid methods

*Abdolnaser Yousefzadeh * and John C. Bancroft, CREWES, Department of Geoscience, University of Calgary.*

Summary

In order to attenuate the migration artifacts and increase the special resolution of the subsurface reflectivity, conventional migration may be replaced by the least squares migration (LSM). However, this is a costly procedure. To reduce the cost, the feasibility of using the multigrid methods in solving the linear system of prestack Kirchhoff LSM equation is investigated. This study showed that the conventional method of multigrid is not viable to solve Kirchhoff LSM equation for at least two reasons. The main reason is that the Hessian matrix is not a diagonally dominant matrix. Therefore, the conventional iterative solvers of the multigrid are not effective. The performance of Conjugate Gradient (CG) multigrid is discussed. It is shown that since CG does not have a smoothing property, it should not be considered as an effective multigrid iterative solver. Using the CG as an iterative solver for the multigrid may slightly reduces the number of iterations for the same rate of convergence in the CG itself. However, it does not reduce the total computational cost.

Introduction

The ability to handle incomplete and irregular seismic data is probably the main advantages of the Kirchhoff to the other methods of migration. However, incomplete data produce migration artifacts and may give a blurred image of the earth reflectivity. Let Kirchhoff seismic modeling be defined as a forward process by:

$$\boldsymbol{d} = \boldsymbol{Gm}, \qquad (1)$$

where $\boldsymbol{d}$ is the seismic data, $\boldsymbol{m}$ is the earth reflectivity, and $\boldsymbol{G}$ is Kirchhoff forward modeling operator, a matrix containing diffraction hyperbolas. The transpose of $\boldsymbol{G}$, may be used as an approximation to the inverse of modeling:

$$\widehat{\boldsymbol{m}} = \boldsymbol{G}'\boldsymbol{d}, \qquad (2)$$

where $\widehat{\boldsymbol{m}}$ is the migrated image. Substitution of $\boldsymbol{d}$ from Equation 1 into Equation 2 results in:

$$\widehat{\boldsymbol{m}} = \boldsymbol{G}'\boldsymbol{Gm}. \qquad (3)$$

Since the Hessian matrix, $\boldsymbol{G}'\boldsymbol{G}$, is not equal to unity, Kirchhoff migration is not able to reconstruct the true model of the subsurface reflectivity (Nemeth et al., 1999).

In order to overcome the migration artifacts, Kirchhoff migration can be augmented by a generalized inverse as an approximation to the exact inverse (Tarantola, 1984). This approach is called LSM (Nemeth et al., 1999; Chavent and Plessix, 1999; Duquet et al., 2000; Kuehl and Sacchi, 2001). In the LSM, migration artifacts are attenuated by minimizing the difference between the observed data, $\boldsymbol{d}$, and the modeled data, $\boldsymbol{G}\widetilde{\boldsymbol{m}}$, expressed by $|\boldsymbol{G}\widetilde{\boldsymbol{m}} - \boldsymbol{d}|$, where $\widetilde{\boldsymbol{m}}$ is an approximation to $\boldsymbol{m}$. Since data include some errors, trying to find a model to fit the data perfectly will be replaced by:

$$\boldsymbol{e} = \boldsymbol{G}\widetilde{\boldsymbol{m}} - \boldsymbol{d}, \qquad (4)$$

where $\boldsymbol{e}$ is an error vector (Sacchi, M. D., 2005, course note in MITACS 2005, University of Alberta). Minimum norm solution includes finding a model, $\widetilde{\boldsymbol{m}}$, that minimizes the following cost function:

$$J(\widetilde{\boldsymbol{m}}) = \left\|\widetilde{\boldsymbol{m}}'\widetilde{\boldsymbol{m}}\right\|^2, \qquad (5)$$

subject to the data constraint:

$$\|\boldsymbol{G}\widetilde{\boldsymbol{m}} - \boldsymbol{d}\|^2 = \epsilon. \qquad (6)$$

These two equations, together implies the minimization of a cost function in the form of:

$$J(\widetilde{\boldsymbol{m}}) = \|\boldsymbol{G}\widetilde{\boldsymbol{m}} - \boldsymbol{d}\|^2 + \mu^2\|\widetilde{\boldsymbol{m}}\|^2. \qquad (7)$$

In a more general cases, an objective function to reduce the migration artifacts can be written as (Nemeth, 1999):

$$J(\widetilde{\boldsymbol{m}}) = \|\boldsymbol{G}\widetilde{\boldsymbol{m}} - \boldsymbol{d}\|^2 + \mu^2\mathcal{R}(\widetilde{\boldsymbol{m}}). \qquad (8)$$

The first term in the right hand side of Equation 8 is data misfit. In the minimization of $J(\widetilde{\boldsymbol{m}})$, this term recovers model to fit the data. Second term on the right hand side of Equation 8 is a regularization term and μ is a regularization weight. Euclidian norm is the simplest form of the regularization term: $\mathcal{R}(\widetilde{\boldsymbol{m}}) = \|\widetilde{\boldsymbol{m}}\|_2^2$, which leads to the damped least squares solution, $\widetilde{\boldsymbol{m}}_{\boldsymbol{DLS}}$, of the problem obtained by solving following equation:

$$\left(\boldsymbol{G}'\boldsymbol{G} + \mu\boldsymbol{I}\right)\widetilde{\boldsymbol{m}}_{\boldsymbol{DLS}} = \boldsymbol{G}'\boldsymbol{d}. \qquad (9)$$

Many other regularization functions may be used. For example, the solution to a LSM problem with smoothing in the offset direction, $\widetilde{\boldsymbol{m}}_{\mathbf{SLS}}$, is obtained when $\mathcal{R}(\widetilde{\boldsymbol{m}}) = \|\boldsymbol{D}_h\widetilde{\boldsymbol{m}}\|_2^2$, where $\boldsymbol{D}_h$ is the first derivative in the offset direction.

The higher resolution images of the LSM can then be used in a forward manner in order to reproduce or interpolate the missing traces (Nemeth, 1999). However, there are two issues associated with replacing conventional migration with LSM. The first problem is that the convergence of the method to the correct solution strongly depends on the accuracy of the background velocity information as shown by Yousefzadeh (2008). Second, LSM consumes more computer time and memory than the migration. The performance of least squares seismic migration usually requires solving a large system of linear equations in the general form of:

$$\boldsymbol{Au} = \boldsymbol{b}, \qquad (10)$$

where $\boldsymbol{A}$ is a general $M \times N$ matrix, for example $\boldsymbol{G}'\boldsymbol{G} + \mu^2\boldsymbol{I}$ in Equation 9, $\boldsymbol{b}$ is a vector with M known elements, i.e. migration image, and $\boldsymbol{u}$ is the unknown solution. The Hessian is a large matrix which is too difficult to be solved using direct methods. Size of $\boldsymbol{G}'\boldsymbol{G}$ equals to the square of number of grids in the model. Iterative methods replace the direct methods to solve Equation 9.

CG and CG least squares (CGLS) (Hestenes and Steifel, 1952; Scales, 1987) has been widely used as a solver for the LSM equation (Nemeth et al., 1999; Duquet et al., 2000; Kuehl and Sacchi, 2001; Yousefzadeh, 2008). However, this is an expensive method. In solving the equation with CGLS method, each iteration requires more than two migration running time.

Multigrid methods may be another choice. Some PDEs are being solved faster and with the better recovery of the low frequency contents by multigrid than many other methods such as Successive Over Relaxation (SOR) and CG (Stuben, 2002). Using multigrid methods for solving problems in seismic exploration is not a new idea (Bunks et al., 1995; Millar and Bancroft, 2004; Plessix, 2007).

In this study, the feasibility of using multigrid properties for solving prestack Kirchhoff time LSM in order to reduce the computational cost or enhance the resolution of the resulted image is investigated. The same study could be done for Kirchhoff prestack depth migration. However, in order to compare different solvers of the prestack Kirchhoff migration the simpler and faster algorithm of time migration is chosen. The same results can be achieved using Kirchhoff depth migration.

Multigrid Theory

Equation 10 may be solved by weighted Jacobi iterations expressed by (Briggs et al., 2000):

$$\boldsymbol{v}^{k+1} = ((1-w)\boldsymbol{I} + w\boldsymbol{R}_J)\,\boldsymbol{v}^k + w\boldsymbol{D}^{-1}\boldsymbol{b},\ \ 0 < w < 2, \quad (11)$$

where $\boldsymbol{v}$ is an approximation to the exact solution and matrix $\boldsymbol{A}$ is splitteed to $\boldsymbol{D}$, a diagonal matrix, and $-(\boldsymbol{L} + \boldsymbol{U})$, summation of the lower and upper triangle matrices, $(\boldsymbol{A} = \boldsymbol{D} - \boldsymbol{L} - \boldsymbol{U})$, $\boldsymbol{R}_J = \boldsymbol{D}^{-1}(\boldsymbol{L} + \boldsymbol{U})$ is Jacobi iteration matrix, and $w \in \mathcal{R}$ is weighting factor. Starting from an initial value for $\boldsymbol{v}^0$, in each iteration all components of $\boldsymbol{v}^{k+1}$are calculated, then $\boldsymbol{v}^k$ is replaced by $\boldsymbol{v}^{k+1}$. This procedure repeats until the desired convergence is achieved. In the Gauss-Seidel method, each component is replaced as soon as it updated, thus, it needs less memory and converges faster than the Jacobi.

It is shown that the Jacobi and Gauss-Seidel methods converge to the solution if matrix $\boldsymbol{A}$ in Equation 10 be diagonally dominant. The convergence rate is slower for the lower frequencies of the solution (Strang, 1986). This is a general property of the Jacobi iterations. Removing high frequency contents from residuals in the Jacobi (or the Gauss-Seidel) first few iterations produces a smooth error vector including mostly low frequency contents (Briggs et al., 2000).

This property, leads to the multigrid idea. In the method of multigrid, an iterative solver (Jacobi or Gauss-Seidel, generally), produces low frequency contents in residual after a few iterations on Equation 10. By restriction, the kernel of the main problem and its residual are transferred to a coarser grid, where the low frequency components act as the high frequency components. Solving the original equation with a solution to the problem in the coarse grid as the starting point returns an answer which contains more low frequency contents than the solving equation with a vector of zeros as the initial guess (Strang, 1986). This algorithm called a v-cycle multigrid (Figure 1). It is possible to calculate the residuals in the coarse grid and restrict it to a coarser grid and repeat the procedure to a very coarse grid. This algorithm is known as V-cycle. W-cycle algorithm performs more iteration on the coarser grids. In the full multigrid, iteration starts on the coarsest grid, then each interpolation to the finer grid size will be improved by a V-cycle (Strang, 1986; Briggs et al., 2000).

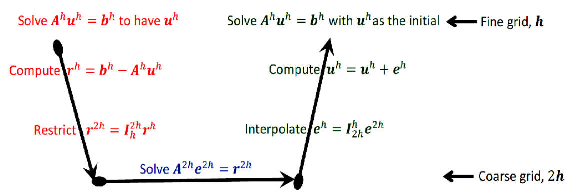


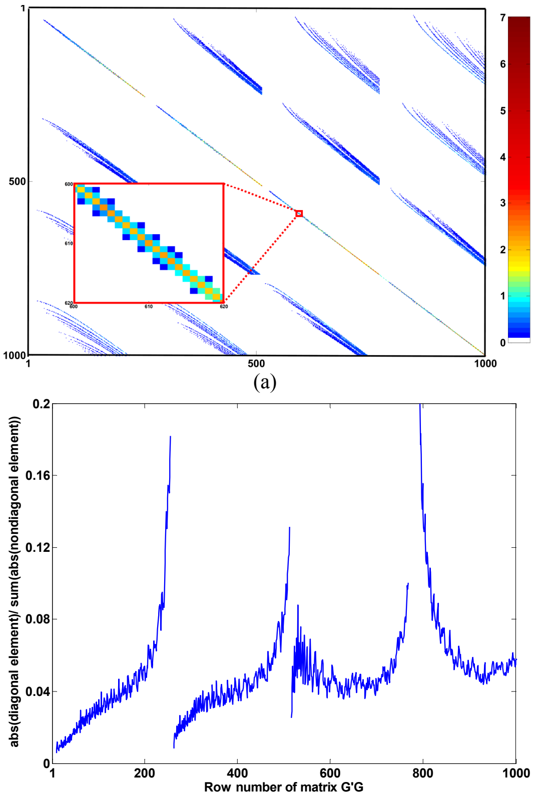
Figure 1. Schematic v-cycle multigrid. The main problem is solved on the fine grid and residuals which are restricted to the coarser grid are used to solve equation in the coarse grid. Solution are interpolated to the fine grid and used as the initial value for the iterations on the main grid. $\boldsymbol{I}_h^{2h}$ and $\boldsymbol{I}_{2h}^h$ are restriction and interpolation matrices, respectively.

Solving LSM using Conventional Multigrid

Jacobi and Gauss-Seidel are conventional multigrid solvers. In order to apply multigrid to the LSM, it is necessary to have its Hessian matrix $\boldsymbol{G}$ (and $\boldsymbol{G}'\boldsymbol{G}$) in the explicit form. However, due to the large size and non-sparseness of $\boldsymbol{G}'\boldsymbol{G}$ matrix multiplication of $\boldsymbol{G}$ or $\boldsymbol{G}'$ to the vectors are more costly than applying migration and modelling operators. The size of matrix $\boldsymbol{G}$ equals to the size of data multiply by the size of the migration image. Therefore, $\boldsymbol{G}$ can be large enough to be impossible to be loaded in the memory of today's computers.

Ignoring this technical problem, the experiences with the explicit form of the Hessian matrices, for different data acquisition geometries show that shows that they are relatively dense and diagonally non-dominant matrices. For instance, for a modeling operator with two sources with

180 m interval spacing and five receivers per source with 72 m interval spacing and with a model 632 m long distance and 0.512 seconds depth, matrix $\boldsymbol{G'G}$ has more than 10% nonzero elements as shown in Figure 2a. Figure 2b shows the ratio of absolute values of diagonal elements to the sum of absolute values of nondiagonal elements for each row of the $\boldsymbol{G'G}$ matrix.



(b)

Figure 2. Elements of matrix $\boldsymbol{G'G}$ (a). The ratio of absolute values of diagonal elements to the sum of absolute values of nondiagonal elements for each row of matrix $\boldsymbol{G'G}$. Values higher than one are diagonally dominant rows (b).

Many examples show that adding a reasonably large amount to the diagonal elements of the Hessian matrix or applying restriction to that does not change it to a diagonally dominant matrix. Applying the restriction operator several times on the Hessian matrix, $\boldsymbol{G'G}$, and converting it to a very coarse matrix does not change the diagonally non-dominancy of the matrix, as well. Another possibility would be reducing the size of $\boldsymbol{G}$ (and $\boldsymbol{G'G}$) by solving for each column of the model at each time ($ix = 1, \ldots, Nx$), where Nx is number of model grid in the horizontal (usually the number of CMPs). However, matrix $\boldsymbol{G_{ix}}'\boldsymbol{G_{ix}}$, is not a diagonally dominant matrix and solvable by Jacobi or gauss Seidel methods. Therefore, the main precondition for Kirchhoff LSM problem to be solvable by the conventional solvers of the multigrid methods is violated. Therefore, at least at its conventional shape, multigrid is not an effective solver for the LSM problem.

Solving LSM Using CG Multigrid

When multigrid is not effective using its conventional iterative methods, other iterative methods may be considered. CG is the usual and powerful method for solving a system of linear equation. Since there is not any decomposition of $\boldsymbol{G}$ or $\boldsymbol{G'}$ in the CG algorithm, it is possible to use operators instead of multiplication of explicit forms of $\boldsymbol{G}$ or $\boldsymbol{G'}$ matrices with vectors. Where CG requires that $\boldsymbol{G'G}$ be symmetric and positive definite matrix, CGLS does not require this condition (Scales, 1987). Therefore, without requirement of a large memory size, in a few iterations, CGLS retrieves a high resolution image of the earth subsurface reflectivity. If equation $\boldsymbol{Gm} = \boldsymbol{d}$ is an overdetrermined problem, then $\boldsymbol{G'G}$ is nonsingular and CGLS converges to solve equation $\boldsymbol{G'G\widetilde{m}_{LS}} = \boldsymbol{G'd}$ in n iterations.

In order to be effective, multigrid requires an iterative method to be a smoother when used as the solver. The smoother must is able to find the high frequency contents of the solution and leave the low frequency contents in the residuals after a few iterations. However, CG does not have a smoothing property. In fact, CG methods are roughers and not smoothers (Shewchuk, 1994). This property is investigated on a LSM problem. Figure 3 shows the convergence rate of CGLS for a synthetic model with different dominant frequencies in the data. This figure shows no better convergence when data include higher frequency content in the damped LSM. Figure 4 shows

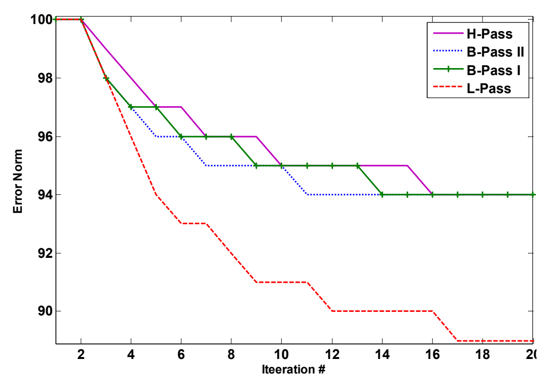


Figure 3. Convergence of CGLS to solve LSM for the filtered version of the Marmousi data set. L-Pass: $f < 15Hz$, B-Pass I: $15Hz < f < 30Hz$, B-Pass II: $30Hz < f < 45Hz$, H-Pass: $45Hz < f$.

same results with the convergence of LSM of the Marmousi data filtered by different band-pass filters. Same conclusion is achieved by solving LSM with smoothing in the offset direction as the regularization.

Multigrid least squares migration

When an iterative method dose not converge for data with higher frequency contents faster than the data containing lower frequency contents, it is not able to leave low frequency contents in the residuals and act as the smoother. Therefore, CG methods should not be an effective solver for the multigrid.

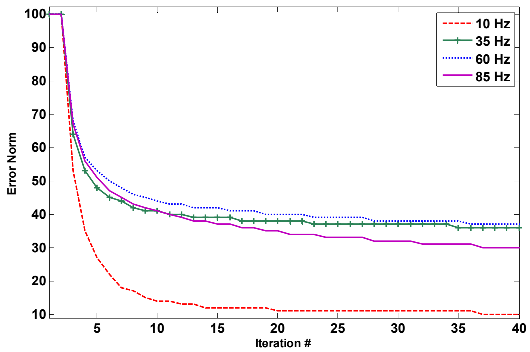


Figure 4. Convergence of CGLS to solve damped LSM for a synthetic model with wavelets having different dominant frequencies: 10, 35, 60, and 85 Hz.

However, three main approaches to applying multigrid to a LSM equation may be examined. To transfer the problem to a higher or lower grid sizes, restriction and interpolation of a LSM problem can be applied in each, horizontal or distance, vertical or time, or both directions of the model. Applying multigrid in the vertical (time) direction should not improve the performance of LSM since the convergence is not faster for the lower frequency components of the data. The analyses of using multigrid CGLS with restriction and interpolation in the distance direction is shown by comparison between full multigrid CGLS and CGLS (Figure 5). Restriction to a coarser grid is performed by deleting half of the traces (leavening one trace and removing next one) from the migration image. Figure 5a is the true model, Figure 5b shows the image of Kirchhoff LSM with five iterations on the CGLS, and same results is obtained from performing a full multigrid CGLS with five iterations on each grid (Figure 5c).

With CG as a rougher (Shewchuk, 1994), a reverse v-cycle, may be examined. In this method, the problem may be solved in the main grid size and then, results which include mostly high frequency contents, interpolated to a finer grid where high frequency contents act as low frequency contents. In this new grid size, the problem is solved and restricted solution added to the solution in the main grid size to be used as the new initial value in the main problem. However, this method is not effective in LSM for two reasons: first, solving LSM in a grid finer than the main problem is very costly; second, roughness property of CG is not as robust and regular as the smoothness property of the Jacobi method. It seems that the convergence of CG is faster only with an initial value with very low frequency contents in the initial model. For higher frequencies, there is not any clear relationship between rate of convergence and the frequency contents of the initial value.

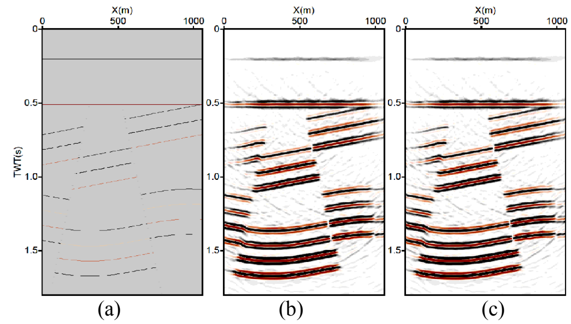


Figure 5. Comparison between true model (a), CGLS (b), and Multigrid CGLS (c) of a synthetic seismic dataset.

However, the experiments show when there is not any idea about the starting point for the CG iterations, CG full multigrid may be used instead. This may eliminate the possible local minima of the objective functions by solving the problem on a coarser grid in order to guarantee convergence to the global minimum and avoid local minima for the nonlinear problems. This also reduces the number of iteration for the same rate of convergence. However, this reduction is not so significant that can reduce the total computational cost.

Conclusions

Solving inverse problems requires solving a large linear equation. Numerical examples show that time domain LSM problem is not solvable by the Jacobi or Gauss-Seidel iterations. Consequently, the conventional multigrid is not viable to the mentioned problem. Requirement of large memory size is another problem with this method.

CG is an effective solver. However, CG is not a smoother. Therefore, using CG as the multigrid solver does not increase the speed of convergence or provide a better recovery of the low frequency contents. Using multigrid with CG as the iterative solver may slightly reduces the number of iterations for the same rate of convergence in comparison to the CGLS by introducing an initial value. However, it may not reduce the total computational cost. It may still being used to avoid local minima in the cost functions where problems become nonlinear.

Acknowledgments

Authors wish to acknowledge the sponsors of Consortium for Research in Elastic Wave Exploration Seismology (CREWES) project for their continuing support. They also appreciate the computer technicians in the CREWES project.