

Feasibility of solving least squares Gazdag migration using method of multigrid.

Abdolnaser Yousefzadeh and John C. Bancroft

ABSTRACT

Methods of multigrid have been widely used in solving partial differential equations in physics and mathematics. With the ability of faster recovery of the low frequency components of the solution, they have been used in solving some problems in the exploration geophysics.

Our previous studies showed that the standard method of multigrid is not viable to solve the Kirchhoff least squares prestack migration equation for two reasons. First, the kernel of the main problem is not a diagonally dominant matrix and solvable by the typical multigrid solvers, Jacobi and Gauss Seidel. Second, kernel matrices are extremely large to work with.

This study investigates the feasibility of using multigrid methods in solving a system of Gazdag least squares migration. It is shown by doing least squares migration for each temporal frequency at the time the kernel matrix become smaller, but it is a diagonally non-dominant matrix. By implementing least squares for each temporal and spatial frequency separately, the kernel matrix remains non-diagonal dominant. Best scenario is performing least squares migration for each temporal frequency and depth separately at a time. The kernel matrix reduces to a diagonal and easily invertible matrix.

1. INTRODUCTION

Stacked seismic data suffers from the presence of the diffracted energies and incorrect positioning of the dipping layers. Migration is a method to move dipping reflection events to their true geological locations and collapse the unwanted diffracted energy back to the scatterpoint locations.

Diffraction summation (Hagedoorn, 1954) is the first computational method of migration. Diffraction summation then developed to the Kirchhoff migration. Claerbout and Doherty (1972) showed how migration is an approximate solution to the wave equation. The integral formulation of the Kirchhoff migration was introduced later by Schneider (1978). Gazdag (1978) and Stolt (1978) developed the migration methods using the Fourier transform domain (Gary et al., 2000). Reverse time migration, the most expensive wavefield continuation method of migration was introduced later (Baysal et al., 1983; Whitmore, 1983). Hill (1990) introduced the Gaussian beam migration based on the decomposition of data and source function in Gaussian beams.

With a simple geology structure and in the case of regularly and completely sampled data, Kirchhoff, Stolt, or Gazdag migration produce a reasonable estimate of the underground reflectivity. However, they may not be able to reveal a complex subsurface geology with the strong lateral velocity variations. Kirchhoff migration with ray tracing, some extensions to the Gazdag migration for example Phase Shift Plus Interpolation

(PSPI) migration (Gazdag and Sguazzero, 1984) and Non-Stationary Phase Shift (NSPS) migration (Margrave and Ferguson, 1999), are able to consider the lateral velocity variations.

These methods need a dense and regularly sampled seismic data to succeed. However, it is impractical or expensive to acquire a dense and regularly sampled seismic data. Even if for instance, Kirchhoff migration is able to handle incomplete or irregularly seismic data, it may produce migration artifacts. This is explained in more detail in the previous CREWES report (Yousefzadeh and Bancroft, 2009). Augmenting migration by a generalized inverse (Tarantola, 1984) and using least squares migration or inversion instead, reduces the migration artifacts (Nemeth et al., 1999; Duquet et al., 2000; Kuehl and Sacchi, 2001).

In case of Gazdag migration in a $v(z)$ media, the steep dip components are not being migrated properly. Least squares migration can be used for recovery of these parts of the migrated images too (Ji, 1997).

Least Squares Conjugate Gradient (LSCG) (Scales, 1987) has been used or recommended for solving the least squares migration equation by many researchers (Nemeth et al., 1999, Duquet et al., 2000, Kuehl and Sacchi, 2001, and Yousefzadeh, 2008). Yousefzadeh and Bancroft (2009) investigated the feasibility of solving least squares Kirchhoff prestack migration using multigrid methods. In the current study, the feasibility of using multigrid method for solving the Gazdag least squares migration is investigated. This is an ongoing research and authors are extending the study to solve least squares phase shift prestack migration in the future.

2. GAZDAG MIGRATION

Phase shift or Gazdag migration was introduced by Gazdag (1978). Here, for simplicity, the Claerbout definition of the Gazdag migration (Claerbout, 2005) is explained. Gazdag migration may be well explained with the downward continuation concept.

2.1. Downward continuation

Considering $P(x, z = 0, t)$ to be a recorded compressional wavefield on the earth surface, migration is just the extrapolation of this wavefield to the zero time: $P(x, z, t = 0)$. For simplicity, let assume that the subsurface velocity is horizontally invariant, $v = v(z)$, and there is no multiple reflections. If the upcoming wave consists of only a vertically upcoming plane wave as $u(x, t, z = 0) = u(t) \text{const}(x)$, it can be downward continued back to the earth by a simple time shifting which is equal to the convolution with an impulse function (Claerbout, 2005):

$$u(t, z) = u(t, z = 0) * \delta(t + z/v), \quad (1)$$

where $*$ denote the convolution operator, v is compressional velocity and δ is the Dirac delta function. In the frequency domain this can be done by the multiplication with a complex exponential:

$$U(\omega, z) = U(\omega, z = 0) e^{-i\omega z/v}. \quad (2)$$

where ω is the angular temporal frequency and $i = \sqrt{-1}$.

Instead of vertically traveling wave, if the plane wave arrives at the surface with an angle θ , the angle between the waveform and the earth surface, and with the assumption of stationary waveform, it can be downward continued into the earth. In such cases, the time shift, Δt , is $\Delta z \cos \theta / v$.

The downward continuation function to an interval Δz is:

$$u(t, \theta, z + \Delta z) = u(t, \theta, z) * \delta(t + \Delta z/v \cos \theta), \quad (3)$$

and in the frequency domain is expressed by (Claerbout, 2005):

$$U(\omega, \theta, z + \Delta z) = U(\omega, \theta, z) \exp(-i\omega \Delta z/v \cos \theta). \quad (4)$$

This is the equation for the downward continuation of any wave in a constant velocity media. By considering the earth subsurface as a constitution of horizontal layers with constant velocity inside each layer, this equation is extendable to a media with $v = v(z)$ (Claerbout, 2005). In the case of downward continuation depth steps must be smaller than the layer thickness.

Using the Snell's parameter, $p = \sin \theta / v$, θ can be eliminated from equation (4) :

$$U(\omega, p, z + \Delta z) = U(\omega, p, z) \exp\left(\frac{-i\omega\Delta z}{v(z)} \sqrt{1 - p^2 v(z)^2}\right), \quad (5)$$

which states the downward continuation for each Snell's parameter. Since any real waveform can be considered as the summation of sinusoids with different frequencies and amplitudes, the seismic data can also be decomposed to Snell waves of all values of p s. Then each one downward continued separately and summed again. This is downward continuation using Fourier transform.

It can be shown that a plane wave is a point in the (ω, k_x, k_z) space, where k_x and k_z are wave numbers or spatial frequencies on the x and z axis and defined by (Claerbout, 2005):

$$k_x = \frac{\omega}{v} \sin \theta, \quad k_z = \frac{\omega}{v} \cos \theta, \quad (6)$$

respectively.

Replacing equations (6) in the equality $\sin^2 \theta + \cos^2 \theta = 1$, results the dispersion relation:

$$k_x^2 + k_z^2 = \frac{\omega^2}{v^2}. \quad (7)$$

Replacing k_x from equation (6) into the Snell's parameter, $p = \sin \theta / v$, results in:

$$p = \frac{k_x}{\omega}, \quad (8)$$

which helps to eliminate p from equation (5) (Claerbout, 2005):

$$U(\omega, k_x, z + \Delta z) = U(\omega, k_x, z) \exp\left(-\frac{i\omega\Delta z}{v(z)} \sqrt{1 - \frac{v(z)^2 k_x^2}{\omega^2}}\right), \quad (9)$$

or

$$U(\omega, k_x, z + \Delta z) = U(\omega, k_x, z) e^{ik_z \Delta z}. \quad (10)$$

This is the equation for the downward continuation of the surface recording of an upcoming wave.

2.2. Gazdag migration

Gazdag migration begins with the 2D Fourier transformation of seismic data to (ω, k_x) domain. Then, for each depth interval, Δz , the Fourier transformed data is multiplied by $e^{ik_z \Delta z}$. The resulted wavefield is evaluated at $t = 0$, by summation on all ω s. Since equation (9) is for the downward continuation of upgoing waves, to incorporate the downgoing wave as well, we need to multiply the time delay by two or equivalently divide velocity by two:

$$e^{ik_z \Delta z} = \exp\left(-i \omega \frac{2}{v(z)} \sqrt{1 - \frac{v(z)^2 k_x^2}{4\omega^2}} \Delta z\right). \quad (11)$$

The last step is the inverse Fourier transformation from k_x to x . This algorithm is explained in more detail elsewhere in this report.

3. LEAST SQUARES MIGRATION EQUATION

Convolution, Kirchhoff and phase shift seismic modelling are some geophysical problems that can be formulated in a general linear form:

$$\mathbf{d} = \mathbf{G}\mathbf{m}. \quad (12)$$

In our case, \mathbf{d} is the observed seismic data, \mathbf{m} is the earth reflectivity model, and \mathbf{G} is an operator acting on \mathbf{m} to produce \mathbf{d} .

The inversion process,

$$\mathbf{m} = \mathbf{G}^{-1}\mathbf{d}, \quad (13)$$

recovers the earth model or reflectivity from the seismic data. Since matrix \mathbf{G} may not be square or non-invertible or it may be extremely large, calculating the inverse of \mathbf{G} may be difficult or impossible. Thus, approximations to the inversion are used. The first approximation uses the conjugate transpose of \mathbf{G} :

$$\hat{\mathbf{m}} = \mathbf{G}'\mathbf{d}. \quad (14)$$

which is the same as Gazdag migration. In equation (14), $\hat{\mathbf{m}}$ is the migrated image and \mathbf{G}' is the migration operator, the conjugate transpose of \mathbf{G} .

By defining poststack Gazdag seismic modelling as the forward process and Gazdag migration as its adjoint (conjugate-transpose) operator, seismic imaging becomes an inversion problem. Substitution of \mathbf{d} from equation (12) into equation (14) results in:

$$\hat{\mathbf{m}} = \mathbf{G}'\mathbf{G}\mathbf{m}. \quad (15)$$

Hessian matrix, $\mathbf{G}'\mathbf{G}$, is different from an identity matrix, \mathbf{I} (Nemeth et al., 1999). Therefore, migration images are different from the real earth subsurface reflectivity with having some artifacts. This difference can be reduced by minimizing the difference between the observed data, \mathbf{d} , and the modeled data, $\mathbf{G}\mathbf{m}$, expressed by $|\mathbf{G}\mathbf{m} - \mathbf{d}|$. Since data include some errors, trying to find a model to fit data perfectly is not recommended. Therefore, the exact fitting will be replaced by:

$$\mathbf{e} = \mathbf{G}\mathbf{m} - \mathbf{d}, \quad (16)$$

where \mathbf{e} is the error vector (Sacchi, 2005). Minimum norm solution includes finding a model, \mathbf{m} , that minimizes the following cost function:

$$J(\mathbf{m}) = \|\mathbf{m}'\mathbf{m}\|^2, \quad (17)$$

subject to the data constraint:

$$\|\mathbf{G}\mathbf{m} - \mathbf{d}\|^2 = \epsilon. \quad (18)$$

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

$$J(\mathbf{m}) = \|\mathbf{G}\mathbf{m} - \mathbf{d}\|^2 + \mu^2\|\mathbf{m}\|^2. \quad (19)$$

An objective function which reduces migration artifacts may be written in the general form of (Nemeth, 1999):

$$J(\mathbf{m}) = \|\mathbf{G}\mathbf{m} - \mathbf{d}\|^2 + \mu^2\mathcal{R}(\mathbf{m}), \quad (20)$$

where \mathbf{d} is the observed data which may be spatially incomplete or irregular. The first term in the right hand side of equation (20) is “data misfit”. In minimizing the cost function, $J(\mathbf{m})$, this term recovers model in the way to fit the data. Second term on the right hand-side of equation (20) is “regularization term” and μ is the “regularization weight”. $\mathcal{R}(\mathbf{m})$ is a linear operator acting on \mathbf{m} and is different for each purpose and uses some priori information about model.

To solve equation (20) for simplicity, if we may ignore the regularization term, $\mu = 0$. By taking first derivative of the remaining cost function and setting it equal to zero, the following normal equation is achieved:

$$\mathbf{G}'\mathbf{G}\mathbf{m}_{LS} - \mathbf{G}'\mathbf{d} = \mathbf{0}. \quad (21)$$

This gives the least squares solution, \mathbf{m}_{LS} ,

$$\mathbf{m}_{LS} = (\mathbf{G}' \mathbf{G})^{-1} \mathbf{G}' \mathbf{d}. \quad (22)$$

An objective function with the minimum norm or Euclidian norm, the simplest form of the regularization function: $\mathcal{R}(\mathbf{m}) = \|\mathbf{m}\|_2^2$, leads to the damped least squares solution, \mathbf{m}_{DLS} , of the problem:

$$\mathbf{m}_{DLS} = (\mathbf{G}' \mathbf{G} + \mu^2 \mathbf{I})^{-1} \mathbf{G}' \mathbf{d}. \quad (23)$$

Since the images resulted from the least squares migration have higher resolution than the images from migration, the high resolution images can then be used in a forward problem to reproduce or interpolate the missing data (Nemeth, 1999). Examples for data reconstruction and comparison between images from migration and least squares migration can be found at Yousefzadeh (2008) and Yousefzadeh and Bancroft (2009).

There are some issues associated with the replacement of the migration with the least squares migration. The main problem is that the convergence of the method to a solution depends on the accuracy of the background velocity model. Without accurate background velocity, least squares migration does not converge to the desired solution. This is because least squares migration is more sensitive to the accuracy of the velocity information than the migration itself (Yousefzadeh, 2008).

The second problem is that least squares migration is more computer time and memory consuming procedure than the migration. As an example, in solving the equation with LSCG method, each iteration in the CG requires two migration/modelling running time where migration is a time and memory consuming process.

3.1. Solving least squares migration equation

Method of CG (Hestenes and Steifel, 1952) which is an extension to Steepest Descent method has been the typical solver for solving the seismic inversion problems. CG requires that the kernel matrix be positive definite. LSCG, a modified version of Conjugate Gradient (CG) method, does not require this condition and directly works with the \mathbf{G} and \mathbf{G}' matrices (Scales, 1987). If equation $\mathbf{G}\mathbf{m} = \mathbf{d}$ is an overdetermined problem, then $\mathbf{G}'\mathbf{G}$ is nonsingular and LSCG converges to solve equation $\mathbf{G}'\mathbf{G}\mathbf{m}_{LS} = \mathbf{G}'\mathbf{d}$. However, by replacing the method of CG with LSCG the multiplication of the matrices \mathbf{G} or \mathbf{G}' with vectors are replaced with the applying forward (seismic modelling or demigration) or adjoint (seismic migration) operators on the model or data, respectively. This procedure reduces required memory to load big matrixes \mathbf{G} and \mathbf{G}' into the computer and also avoids big matrix-vector multiplications.

In this study, feasibility of using multigrid properties in solving Gazdag least squares migration in order to reduce the computational cost or enhance the resolution of the resulted image is investigated. To understand the multigrid methods, method of Jacobi iterations and its properties is explained in detail in the section four. Then the feasibility of using multigrid methods on solving least squares migration is discussed in the last section.

4. MULTIGRID METHODS

There are many different methods to solve a linear system of equation:

$$\mathbf{A}\mathbf{u} = \mathbf{b}, \quad (24)$$

where \mathbf{A} is a $m \times n$ matrix, \mathbf{b} is a vector with m elements and \mathbf{x} is the unknown to be found. Gaussian elimination is a method to find the exact solution to this equation. However, this method may not work efficiently for a large system of equations. For such problems, it is more efficient to use an iterative method. An iterative method starts with an initial guess as the solution and retrieves a reasonable approximation to the solution. Jacobi, Gauss-Seidel, CG and Krylove are examples of the iterative methods.

Multigrid is not an iterative (or non-iterative) method of solving equation (24). It is a method of solving a problem in different grid sizes. Multigrid uses an iterative solver to solve the equation on different grid sizes. This method is well explained in the previous CREWES report by authors (Yousefzadeh and Bancroft, 2009). However, in order to have the current paper stand alone, a brief explanation of multigrid method and solver is presented here.

With the ability of faster recovery of the low frequency content of the solution, multigrid methods are used to solve many problems in the exploration seismology. Multigrid method helped performing seismic waveform velocity inversion on the Marmousi data set (Bunks et al., 1995). Bunks et al. (1995) eliminated the local minima of the objective function by solving problem on a coarser grid in order to guarantee convergence to the global minimum.

Millar and Bancroft (2004) used multigrid method to enhance the resolution of the seismic data during deconvolution. They showed better recovery of the reflectivity and damping the low frequencies than the Gauss-Seidel method. They showed that the good estimation of the wavelet and also the frequency content of the data is necessary for the method to succeed (Millar and Bancroft, 2004).

Plessix (2007) studied the effects of using multigrid cycles for the 3D frequency domain wave equation migration. He considered the result of using multigrid on the undamped wave equation at seismic frequencies.

Yousefzadeh and Bancroft (2010) examined using different approaches of multigrid method to solve the least squares Kirchhoff prestack migration.

They showed why the standard method of the multigrid is not viable to solve the Kirchhoff least squares prestack migration equation for at least two reasons: first, the kernel of the main problem, is not a diagonally dominant matrix, therefore, Jacobi or Gauss-Seidel iterations, the standard iterative methods in multigrid, are not effective, second, matrices are too large and dense to be loaded in computers' memory (Yousefzadeh and Bancroft, 2010) (Figure 1).

In this study, we attempt to use multigrid method to solve phase shift least squares migration in order to reduce the computational costs or enhance the resolution of the resulted image.

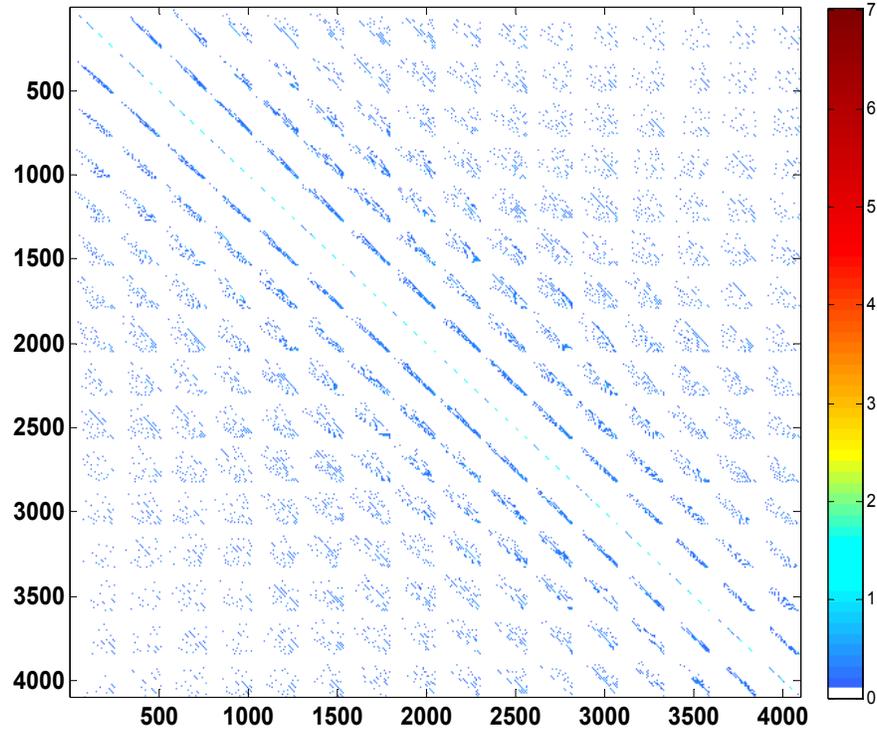


FIG. 1. Non-zero elements of the matrix $\mathbf{G}'\mathbf{G}$ for the least squares Kirchhoff prestack migration. Matrix is not diagonally dominant and savable by Jacobi or Gauss Seidel methods (Yousefzadeh and Bancroft, 2009).

4.1. Multigrid Solvers

Jacobi and Gauss Seidel are typical multigrid solvers. They have an especial property, which made them the best known candidate as the multigrid solvers. This property helps multigrid methods to find the solution faster than using the alternate iterative methods alone.

Idea of solving a problem in different grid sizes returns back to a well known property of the Jacobi method. In equation $\mathbf{A}\mathbf{u} = \mathbf{b}$, where \mathbf{u} is the unknown desired exact solution and \mathbf{b} and \mathbf{A} are known vector and matrix, respectively, if \mathbf{v} is considered as an approximation to the exact solution, then the algebraic error, \mathbf{e} , is the difference between these two solutions:

$$\mathbf{e} = \mathbf{u} - \mathbf{v}. \quad (25)$$

Since \mathbf{u} is unknown, \mathbf{e} is not directly computable. Therefore, residual, a measurable version of the error, may be defined by:

$$\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{v}. \quad (26)$$

Replacing \mathbf{b} from equation (24) gives:

$$\mathbf{r} = \mathbf{A}\mathbf{u} - \mathbf{A}\mathbf{v} = \mathbf{A}(\mathbf{u} - \mathbf{v}), \quad (27)$$

or

$$\mathbf{r} = \mathbf{A}\mathbf{e}, \quad (28)$$

which called the “residual equation” (Briggs et al., 2000). With \mathbf{v} as an approximation to \mathbf{u} , \mathbf{r} is computable from equation (26). Solving residual equation for the \mathbf{e} , gives a new approximate solution using equation (25) in the form of $\mathbf{u} = \mathbf{v} + \mathbf{e}$.

Substitution of the residual equation into equation (25) gives:

$$\mathbf{u} = \mathbf{v} + \mathbf{P}^{-1}\mathbf{r}, \quad (29)$$

where $\mathbf{P} \cong \mathbf{A}$, is a preconditioner. This suggests iterations in the form of (Briggs et al., 2000):

$$\mathbf{v}^{k+1} = \mathbf{v}^k + \mathbf{P}^{-1}\mathbf{r} \quad (30)$$

where k is the iteration number.

Estimation of \mathbf{v} can be better improved by choosing \mathbf{P} as close as possible to \mathbf{A} . In the method of Jacobi, preconditioner \mathbf{P} is chosen to be the diagonal matrix of \mathbf{A} : $\mathbf{P} = \mathbf{D}$. By splitting matrix \mathbf{A} to \mathbf{D} , a diagonal matrix, and $-(\mathbf{L} + \mathbf{U})$, summation of the lower and upper triangle matrices, (or $\mathbf{A} = \mathbf{D} - \mathbf{L} - \mathbf{U}$), equation (24) can be written as:

$$\mathbf{D}\mathbf{v} = (\mathbf{L} + \mathbf{U})\mathbf{v} + \mathbf{b}. \quad (31)$$

This leads to the following equation:

$$\mathbf{v} = \mathbf{D}^{-1}(\mathbf{L} + \mathbf{U})\mathbf{v} + \mathbf{D}^{-1}\mathbf{b}. \quad (32)$$

Choosing \mathbf{D} , the diagonal matrix of \mathbf{A} , as the preconditioner has this advantage that it is easily invertible. Inverse of a diagonal matrix can be simply found by only inverting the diagonal elements of that matrix (Strang, 1986). Equation (32) suggests the Jacobi iterations in the form of (Briggs et al., 2000):

$$\mathbf{v}^{k+1} = \mathbf{D}^{-1}(\mathbf{L} + \mathbf{U})\mathbf{v}^k + \mathbf{D}^{-1}\mathbf{b}. \quad (33)$$

Jacobi method can also be expressed by (Briggs et al., 2000):

$$\mathbf{v}^{k+1} = \mathbf{R}_j\mathbf{v}^k + \mathbf{D}^{-1}\mathbf{b}, \quad (34)$$

where $\mathbf{R}_j = \mathbf{D}^{-1}(\mathbf{L} + \mathbf{U})$ is the Jacobi iteration matrix.

Weighted Jacobi method is a modification to the Jacobi method in the form of (Briggs et al., 2000):

$$\mathbf{v}^{k+1} = ((1 - w)\mathbf{I} + w\mathbf{R}_J) \mathbf{v}^k + w\mathbf{D}^{-1}\mathbf{b}, \quad 0 < w < 2, \quad (35)$$

where $w \in \mathcal{R}$ is weighting factor. By defining weighted Jacobi iteration, $\mathbf{R}_w = (1 - w)\mathbf{I} + w\mathbf{R}_J$, Jacobi method has the following matrix shapes (Briggs et al., 2000):

$$\mathbf{v}^{k+1} = \mathbf{R}_w \mathbf{v}^k + w\mathbf{D}^{-1}\mathbf{b}, \quad (36)$$

and

$$\mathbf{v}^{k+1} = \mathbf{v}^k + w\mathbf{D}^{-1}\mathbf{r}^k. \quad (37)$$

It also can be written in the component form as (Saad, 2000):

$$v_i^{k+1} = \frac{1}{a_{ii}} \left(b_i - \sum_{\substack{j=1 \\ j \neq i}}^N a_{ij} v_j^k \right), \quad i = 1, 2, \dots, N, \quad (38)$$

where k is the iteration number and i is the component number of the vectors \mathbf{b} and \mathbf{v} . Starting from an initial value as \mathbf{v}^0 , in each iteration all components of \mathbf{v}^{k+1} are calculated, then \mathbf{v}^k is replaced by \mathbf{v}^{k+1} . This procedure repeats until the desired convergence is achieved.

If each component of the solution is replaced as soon as it updated, leads to Gauss-Seidel method (Saad, 2000):

$$v_i^{k+1} = \frac{1}{a_{ii}} \left(-\sum_{j=1}^{i-1} a_{ij} v_j^{k+1} - \sum_{j=i+1}^N a_{ij} v_j^k - b_i \right), \quad i = 1, 2, \dots, N, \quad (39)$$

This reduces not only the necessary memory to keep all components of \mathbf{v}^{k+1} before updating but also the number of iterations for the same rate of convergence.

Gauss-Seidel iteration can be written as:

$$\mathbf{V} \leftarrow \mathbf{R}_G \mathbf{V} + (\mathbf{D} - \mathbf{L})^{-1}\mathbf{b}, \quad (40)$$

where “ \leftarrow ” shows the displacement of elements and $\mathbf{R}_G = (\mathbf{D} - \mathbf{L})^{-1}\mathbf{U}$, is the Gauss-Seidel iteration matrix.

It can be shown that the convergence of the Jacobi (and the Gauss-Seidel) iterations guaranteed if and only if the magnitude of all eigenvalues of \mathbf{R}_J be less than 1 (Strang, 1986):

$$|\lambda(\mathbf{R}_J)| < 1 \quad (41)$$

Spectral Radius of a matrix, ρ , is the maximum amount of its eigenvalues: $\rho(\mathbf{R}_J) = \max|\lambda(\mathbf{R}_J)|$. The Jacobi method converges if and only if $\rho(\mathbf{R}_J) < 1$. The speed of convergence depends on the amount of $\rho(\mathbf{R}_J)$. Smaller $\rho(\mathbf{R}_J)$ causes faster convergence to the solution (Strang, 1986).

In another point of view Jacobi method converges to the solution if matrix \mathbf{A} (in equation (24)) be diagonally dominant. Otherwise, these methods do not converge to the solution. A matrix \mathbf{A} is strictly diagonally dominant if for all elements a_{ij} of the matrix: $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$, where i and j are the number of rows and columns, respectively.

In order to explain the role of the Jacobi iterations in the multigrid method, let consider a system of equation (24) where \mathbf{A} is the second difference matrix:

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 0 & 0 & \cdots & \\ -1 & 2 & -1 & 0 & \cdots & \\ 0 & -1 & 2 & -1 & \cdots & \\ & & \vdots & & \ddots & \vdots \\ \cdots & & & & \cdots & 2 \end{bmatrix}. \quad (42)$$

Therefore, Jacobi iteration matrix is:

$$\mathbf{R}_J = \begin{bmatrix} 0 & 1/2 & 0 & 0 & \cdots & \\ 1/2 & 0 & 1/2 & 0 & \cdots & \\ 0 & 1/2 & 0 & 1/2 & \cdots & \\ & & \vdots & & \ddots & \vdots \\ \cdots & & & & \cdots & 0 \end{bmatrix}. \quad (43)$$

The eigenvalues of \mathbf{A} are $\lambda_j(\mathbf{A}) = 2 - 2 \cos j\theta$, where $\theta = \frac{\pi}{N+1}$. Thus, $\lambda_j(\mathbf{R}_J) = \lambda(I - 1/2 \mathbf{A}) = \cos j\theta < 1$ and the convergence is guaranteed. For example if $N = 4$, then \mathbf{R}_J has four eigenvalues as (Briggs et al., 2000):

$$\lambda_j = \cos \frac{\pi}{5}, \cos \frac{2\pi}{5}, \cos \frac{3\pi}{5} \text{ and } \cos \frac{4\pi}{5} (= -\cos \frac{\pi}{5}). \quad (44)$$

The λ s become larger for the smaller angels. It means that the convergence is slower for lower frequencies of the solution. This is an important and general property of the Jacobi iterations.

This property causes the removing of the high frequency contents from residuals in the Jacobi (and the Gauss-Seidel) first few iterations. As a result it produces a smooth (includes mostly low frequency contents) error vector. This “*smoothing*” property is shown in the following example.

Suppose that we want to solve the system of linear equation:

$$\mathbf{A}\mathbf{u} = \mathbf{0}, \quad (45)$$

where \mathbf{A} is the second difference matrix in equation (42), with $n = 64$. The trivial exact solution is $\mathbf{u} = \mathbf{0}$, therefore, $\mathbf{e} = -\mathbf{v}$. Let apply the weighted Jacobi method (with $w = 2/3$) to solve this equation with $\mathbf{u}0_j = \sin\left(\frac{jk\pi}{n}\right)$, $0 \leq j \leq n, 1 \leq k \leq n - 1$, the Fourier modes with the frequency k , as the initial guess.

Figure 2a shows the convergence rate of the weighted Jacobi method for different initial values: $k = 1, 4, 8$ and 16 . There is a faster convergence to the solution by

choosing an initial guess with the higher frequency contents (larger k s). It can be shown that the weighted Gauss-Seidel behaves similarly.

If the initial guess be a superposition of the four previous Fourier modes, $u_0j = \frac{1}{4} \left[\sin\left(\frac{j\pi}{n}\right) + \sin\left(\frac{4j\pi}{n}\right) + \sin\left(\frac{8j\pi}{n}\right) + \sin\left(\frac{16j\pi}{n}\right) \right]$, the convergence rate is different. The convergence for the weighted Jacobi iterations starting with this initial guess is shown in Figure 2b. In first five iterations the error decreases rapidly. Then, the convergence becomes slower. The faster decrease corresponds to the presence of the high frequency components in the initial value and less rapid decrease is due to the lower frequency components of the initial value (Strang, 1986).

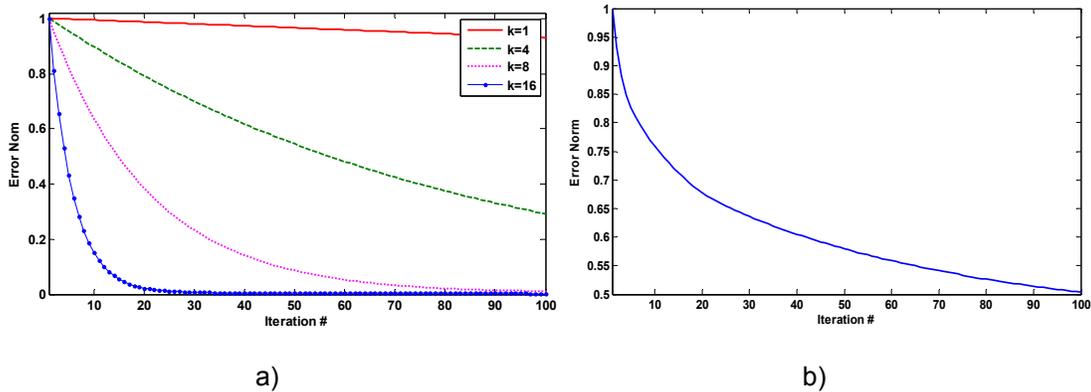


FIG. 2. a) The maximum norm error versus number of iterations is plotted for the weighted Jacobi iterations for different initial values to solve equation (45). b) Convergence when initial values are superposition of the four Fourier modes with both low and high frequencies.

Faster convergence for the higher frequency content is the key to the multigrid idea. In the method of multigrid, an iterative solver which has the smoothing property (Jacobi or Gauss-Seidel, generally), produces low frequency contents in the residual after a few iterations (for example four iterations) on equation (24). By *restriction*, kernel of the main problem and its residual are transferred (restricted/decimated) to a coarser grid (scale), where the low frequency components act as the high frequency components. Problem is being solved in the coarser grid and a solution is being achieved. Then the solution is interpolated to the main grid size.

Solving the original equation with this interpolated solution as the initial guess, gives a solution to the main problem which also contains more low frequency components than the solving equation with a vector of zeros as the initial guess (Strang, 1986). This is the simplest shape of using multigrid and called v-cycle (with the lowercase “v”) multigrid.

A v-cycle multigrid starts with a few iterations on the fine grid, then the error is transferred to a coarser grid by a restriction process, iterations are performed on the coarse grid and the interpolated results used as the starting point in the Jacobi method on the main grid size.

There are two processes in multigrid in order to transfer the problem to the coarser or finer grid (Strang, 1986). Multiplying Restriction matrix, \mathbf{R} , transfer problem from finer grid to the coarse grid and Interpolation matrix, \mathbf{I} , return problem back to the finer grid.

A v-cycle multigrid includes only two grids: a fine grid (with the size of the main problem) and a coarse grid (Figure 3). Iterations start with zero as the initial model to solve equation $\mathbf{A}^h \mathbf{u} = \mathbf{b}^h$ on the fine grid where h corresponds to the size of main grid which is finest grid. After a few iterations, residual to the equation $\mathbf{A}^h \mathbf{u} = \mathbf{b}^h$, calculated as $\mathbf{r}^h = \mathbf{b}^h - \mathbf{A}^h \mathbf{u}^h$. Then multiplication of the Restriction matrix transfers \mathbf{r}^h to the coarser grid \mathbf{r}^{2h} , where $2h$ corresponds to the first coarser grid with size equal to half of the original grid size. Solving $\mathbf{A}^{2h} \mathbf{e}^{2h} = \mathbf{r}^{2h}$ for \mathbf{e}^{2h} on the coarse grid requires a few more iteration and then the solution, \mathbf{e}^{2h} , must be interpolated to the fine grid as \mathbf{e}^h . Finally, iterations to solve $\mathbf{A}^h \mathbf{u} = \mathbf{b}^h$ starts with the improved initial value $\mathbf{u}^h + \mathbf{e}^h$. A few iterations on this problem size/grid returns a solution which include both low and high frequency contents (Strang, 1986; Briggs et al., 2000).

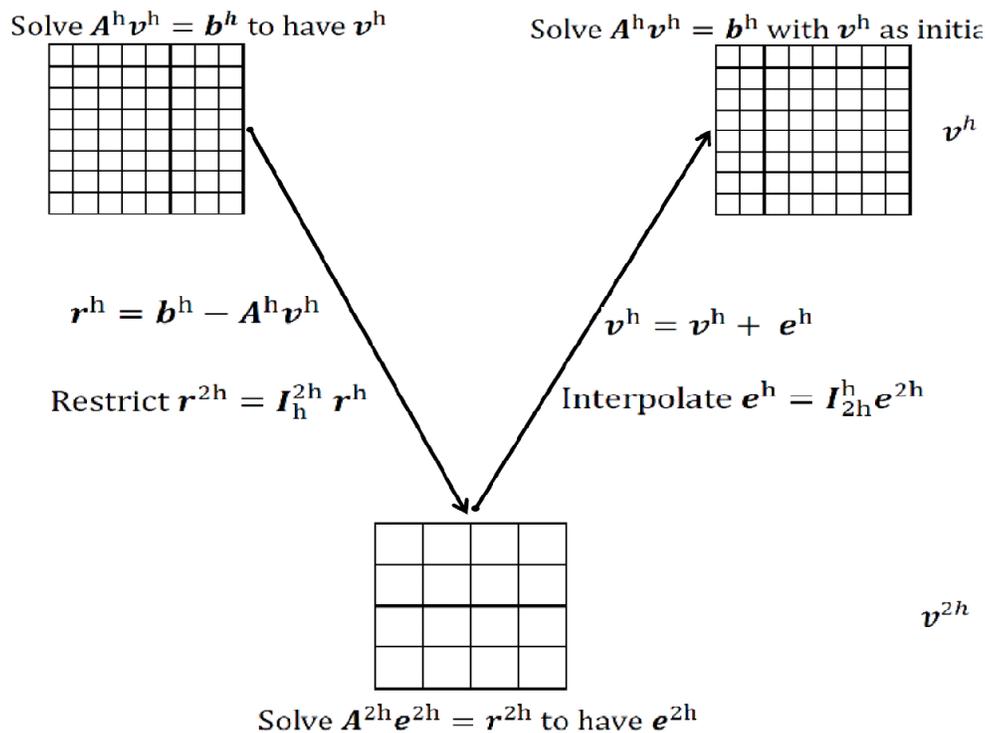


FIG. 3. Schematic v-cycle multigrid. Main problem is solved on the fine grid and residuals which are restricted to the coarse grid with two grids are used to solve equation in the coarse grid. Solution are interpolated to the fine grid and used as the initial value for iterations on the main grid.

It is possible to calculate the residuals in the coarse grid, $\mathbf{r}^{2h} = \mathbf{b}^{2h} - \mathbf{A}^{2h} \mathbf{e}^{2h}$, and restrict it to a coarser grid $4h$ and repeat the procedure to a very coarse grid, $nh!$. This algorithm is known as V-cycle (with the uppercase “V”). W-cycle algorithm performs more iteration on the coarser grids (Strang, 1986; Briggs et al., 2000).

In the full multigrid (FMG), iteration starts on the coarsest grid, the solution is interpolated and used as the initial value for one step finer grid. A v-cycle improves the result. Then, result will be used for the finer grid, a V-cycle improves this result and process continues to arrive to the finest grid which is the size of the initial problem (Figure 4) (Strang, 1986; Briggs et al., 2000).

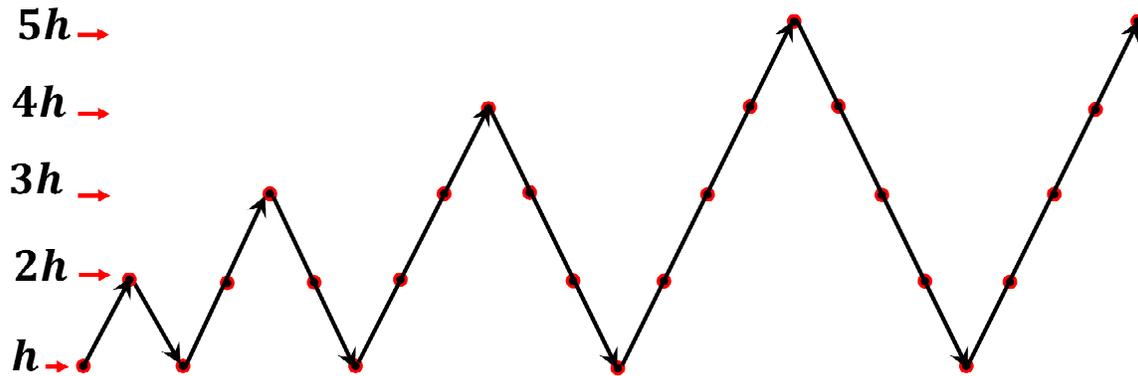


FIG. 4. Schematic FMG for a 5×5 matrix. It starts with the coarsest grid size, h , then the results being interpolated to the one step finer size. A V-cycle improves the result and interpolates it to a finer grid size and procedure continues to the finest grid size.

5. FEASIBILITY OF USING MULTIGRID METHOD TO SOLVE GAZDAG LSM

5.1. First Scenario: LSM for migration and modeling as the forward and adjoint

A MATLAB-like subroutine for the Gazdag migration in a $v(z)$ media includes three loops on temporal frequency, ω , special frequency, k_x , and depth, z is shown in Figure 5 (Claerbout, 2005). These loops are interchangeable. The “if” condition is to prevent the evanescent waves.

$$U(\omega, k_x) = U(\omega, k_x, z = 0) = 2D \text{ FFT } u(t, x)$$

$$\text{image}(Nz, Nk_x) = 0$$

for all ω s

for all k_x s

for all depths $iz = 1:nz$

$$C = \exp\left(-i\omega\Delta z \sqrt{1 - \frac{v(z)^2 k_x^2}{4\omega^2}}\right)$$

if $4\omega^2 > v(z)^2 k_x^2$

$$U(\omega, k_x) = U(\omega, k_x) \times C$$

```

                                 $image(z, k_x) = image(z, k_x) + U(\omega, k_x)$ 
                                end if
                            end loop on depths
                        end loop on  $k_x$ s
                    end loop on  $ws$ 
                 $image = 1D\ IFFT(image)$ 

```

FIG. 5. A subroutine for Gazdag migration in a $v = v(z)$ media (Claerbout, 2005).

Algorithm starts with the 2D Fourier transformation of data. Then for each depth, wavefield is extrapolated. Migration is simply this wavefield for each depth, summed over all temporal frequencies. Resulted images must be transformed to the (x, t) domain by 1D inverse Fourier transform over horizontal spatial frequencies, k_x s. It is possible to make migration faster by using the symmetry properties of the Fourier transform.

For the Gazdag modeling a MATLAB-like subroutine is shown in Figure 6. Modeling starts from the last row of the Fourier transformed image and continues to the earth surface. Data must be 2D inverse Fourier transformed from (ω, k_x) domain to the (t, x) domain.

```

 $image(Nz, Nk_x) = 1D\ FFT(image)$ 
 $U(N\omega, Nk_x) = 0$ 
for all  $ws$ 
    for all  $k_x$ s
        for all depths  $iz = nz: -1: 1$ 
             $C = \exp(+i\omega\Delta z \sqrt{1 - \frac{v(z)^2 k_x^2}{4\omega^2}})$ 
            if  $4\omega^2 > v(z)^2 k_x^2$ 
                 $U(\omega, k_x) = U(\omega, k_x) \times C + image(z, k_x)$ 
            end if
        end loop on depths
    end loop on  $k_x$ s
end loop on  $ws$ 

```

```

end loop on depths

end loop on  $k_x$ s

end loop on  $w$ s

 $U = 2D\ IFFT\ U(\omega, k_x)$ 

```

FIG. 6. A subroutine for Gazdag modeling in a $v = v(z)$ media (Claerbout, 2005).

For a $v(z)$ media, the mentioned migration and modelling subroutines do not pass the dot product test. It means that they are not exactly conjugate transpose of each other. Consequently, it is not possible to use them directly as the \mathbf{G}' and \mathbf{G} operators in a LSCG algorithm to perform a least squares migration solution.

The other option is constructing \mathbf{G} and \mathbf{G}' matrices explicitly. The \mathbf{G} will be a matrix with $N\omega \times Nk_x$ rows and $Nz \times Nk_x$ columns. Hence, $\mathbf{G}'\mathbf{G} + \mu^2\mathbf{I}$ will be a $Nz \times Nk_x$ by $Nz \times Nk_x$ matrix.

Suppose the stacked seismic data includes 100 traces with 1000 samples per trace. To use FFT in the migration or modeling algorithms, the data or model must be zero padded to number of samples be a power of two. Therefore, in the frequency domain k_x has 128 samples and $N\omega = 1024$. If we consider the migrated image with the same size as the data, then $Nz = 1000$. Therefore \mathbf{G} is a $131,072 \times 128,000$ matrix. Consequently $\mathbf{G}'\mathbf{G} + \mu^2\mathbf{I}$ is a matrix with $128,000 \times 128,000 = 16,384,000,000$ elements. This is an extremely large and dense matrix. This matrix is not a diagonally dominant matrix as well. Hence, it is not solvable by the Jacobi method and consequently by the method of multigrid in its typical definition.

Our studies show that in the case of constant velocity media, $v = const.$, the migration and modeling subroutines pass the dot product test by adding an additional line to the migration algorithm as shown in Figure 7 by “(■)”. In this case it is possible to use the migration and modeling subroutine as the \mathbf{G}' and \mathbf{G} matrix multiplication and insert them directly to the LSCG for solving the least squares migration.

```

 $U(\omega, k_x) = U(\omega, k_x, z = 0) = 2D\ FFT\ u(t, x)$ 

image ( $Nz, Nk_x$ ) = 0

for all  $w$ s

for all  $k_x$ s

```

```

C = exp (-iωΔz √(1 - (v²k_x²)/(4ω²))
for all depths iz = 1:nz
    if 4ω² > v²k_x²
        U(ω, k_x) = U(ω, k_x) × C
        image(z, k_x) = image(z, k_x) + U(ω, k_x)
    end if
end loop on depths
image(:, k_x) = image(:, k_x)/C (■)
end loop on k_x's
end loop on ws
image = 1D IFFT (image)

```

FIG. 7. A subroutine for Gazdag migration in a $v = \text{const.}$ media.

5.2. Second Scenario: LSM for migration and modeling for each ω

The second scenario is performing the least squares migration for each temporal frequency of the data. In this scenario, data 2D Fourier transformed into the (ω, k_x) domain. Then, for each temporal frequency, the corresponding image is calculated, then, all images added together which is equal to a summation on the all temporal frequencies.

By taking the least squares migration inside the most outer loop, loop on ω s, the size of data inside least squares migration reduces to Nk_x samples. The size of \mathbf{G} matrix also reduces to a matrix with Nk_x rows and $Nz \times Nk_x$ columns. For our example it is 1024 times smaller. However, size of $\mathbf{G}'\mathbf{G} + \mu^2\mathbf{I}$ matrix remains unchanged. The Hessian is still a large, non sparse and non diagonally dominant matrix. Half of the Hessian matrix elements are nonzero. Hence, it is not solvable by Jacobi method and consequently by the method of multigrid. For our synthetic example and for one temporal frequency, nonzero elements of the first 3500 rows and columns of matrix $\mathbf{G}'\mathbf{G} + \mu^2\mathbf{I}$ is shown in Figure 8.

It can be shown that choosing a reasonably large μ does not change the diagonally nondominancy of the Hessian matrix.

5.3. Third Scenario: LSM for migration and modeling for each ω and k_x

Other option to consider is to perform least squares migration for each temporal frequency and wave number of data. In this scenario the \mathbf{G} matrix is a $1 \times Nz$ matrix. $\mathbf{G}'\mathbf{G} + \mu^2\mathbf{I}$ is a matrix with Nz rows and also Nz columns. This is a relatively small matrix, but as shown in Figure 9 it is not sparse and also it is not a diagonally dominant matrix. This matrix is solvable by the method of CG. However, its condition number is too large and CG is not a fast solver for that.

5.4. Last Scenario: LSM for migration and modeling for each ω and depth

The last scenario in this study is performing least squares migration for each temporal frequency of the data and each depth of the image. By this procedure, the size of \mathbf{G} matrix reduces to a matrix with Nk_x rows and Nk_x columns. $\mathbf{G}'\mathbf{G} + \mu^2\mathbf{I}$ matrix has the same size as \mathbf{G} . The Hessian has a smaller diagonal matrix and therefore is easily invertible. Inverse of a diagonal matrix is achievable just by inverting its diagonal elements. This matrix is shown in Figure 10.

Although from first to the fourth scenario the size of $\mathbf{G}'\mathbf{G} + \mu^2\mathbf{I}$ matrix is decreasing, it is necessary to mention that the number of least squares migration equation to solve is increasing. For the first scenario, there is just one least squares problem to be solved, where in the second scenario it increases by the number of temporal frequencies, 1024 in our example. For the third scenario, this number increase to $N\omega \times Nk_x$, 1024×128 in our example. Finally, the last scenario needs that least squares migration equation to be solved $N\omega \times Nz$ times.

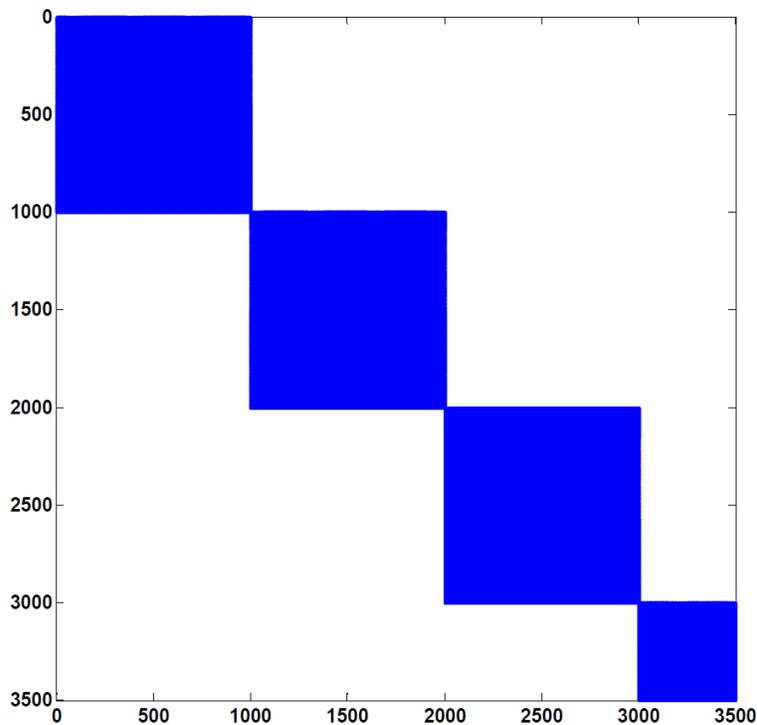


FIG. 8. First 3500 rows and columns of matrix $\mathbf{G}'\mathbf{G} + \mu^2\mathbf{I}$ for the mentioned example.

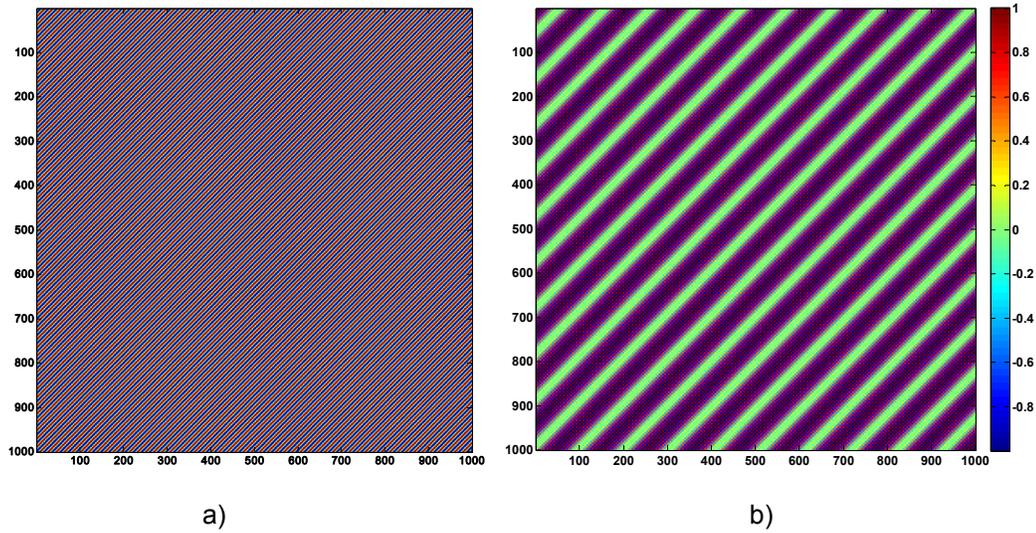


FIG. 9. Real part of elements in matrix $G'G + \mu^2 I$, when $G'G + \mu^2 I$ is constructed for one ω and one k_x . a) $\omega = -1558.5$; b) $\omega = -193.3$.

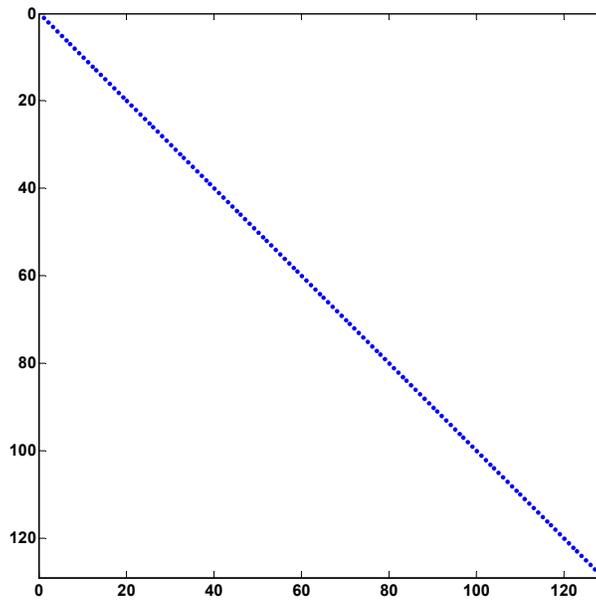


FIG. 10. Nonzero elements of the matrix $G'G + \mu^2 I$, when $G'G + \mu^2 I$ is constructed for one ω and one depth.

6. CONCLUSION

Replacing Gazdag migration with the Gazdag least squares migration results in images with higher resolution than migration images. However, this is a costly procedure. This

study was looking for a faster and less costly solver for the Gazdag least squares migration.

Current study shows that implementing the inversion procedure can be done on each temporal frequency; consequently the Hessian matrix that must be inverted becomes smaller. However, matrix is still large and since it is not a diagonally dominant, is not solvable by the Jacobi or Gauss Seidel method which are typical solvers of the multigrid method. Therefore, multigrid is not able to solve the least squares migration for each temporal frequency. The Hessian matrix remains diagonally nondominant even if the least squares migration applied on each temporal and spatial frequency, separately.

Finally, it is shown that the Hessian matrix is a diagonal matrix if least squares migration performed for each temporal frequency and depth. This is the fastest procedure, since it does not need any solver.

The number of least squares migration to be solved on each scenario is another important factor that must be considered as well.

ACKNOWLEDGEMENTS

Authors wish to acknowledge the sponsors of CREWES project for their continuing support. They especially express their acknowledgment to Dr. Rob Ferguson, Dr. Gary Margrave, Dr. Mostafa Naghizadeh and Marcus Wilson for their technical opinions and help. They also appreciate other faculty, staff and computer technicians in the CREWES project.

REFERENCES

- Baysal, E., D. D. Kosloff, and J. W. C. Sherwood, 1983, Reverse time migration: *Geophysics*, 48, 1514-1524.
- Briggs, W. L., V. E. Henson, and S. F. McCormick, 2000, *A multigrid tutorial*, second edition, SIAM.
- Bunk, C., F. M. Saleck, S. Zaleski, and G. Chavents, 1995, Multiscale seismic waveform inversion. *Geophysics*, 60, 1457-1473.
- Claerbout, J. F., and S. M. Doherty, 1972, Downward continuation of moveout corrected seismograms: *Geophysics*, 37, 741-768.
- Claerbout, J. F., 2005, *Basic Earth Imaging (Version 2.4)*: Stanford Exploration Project, <http://sepwww.stanford.edu/sep/prof/>.
- Duquet, B., K. J. Marfurt, and J. A. Dellinger, 2000, Kirchhoff modeling, inversion for reflectivity, and subsurface illumination: *Geophysics*, 65, 1195-1209.
- Gazdag, J., 1978, Wave equation migration with the phase-shift method: *Geophysics*, 43, 1342-1351.
- Gazdag, J., and P. Sguazzero, 1984, Migration of seismic data by phase-shift plus interpolation: *Geophysics*, 49, 124-131.
- Hagedoorn, J. G., 1954, A process of seismic reflection interpretation: *Geophysical Prospecting*, 2, 85-127.
- Hestenes, M., and E. Steifel, 1952, Methods of conjugate gradients for solving linear systems: *Nat. Bur. Standards J. Res.*, 49, 403-436.
- Hill, N. R., 1990, Gaussian beam migration: *Geophysics*, 55, 1416-1428.
- Ji, J. 1997, Least squares imaging, datuming, and interpolation using the wave equation: Stanford Exploration Project.
- Kuehl, H., and M. D. Sacchi, 2001, Split-step WKB least-squares migration/inversion of incomplete data: 5th Soc. Expl. Geophys. of Japan Internat. Symp. Imaging Technology.
- Margrave, G. F., and R. J. Ferguson, 1999, Wavefield extrapolation by nonstationary phase shift: *Geophysics*, 64, 1067-1087.
- Millar, J., and J. C. Bancroft, 2004, Multigrid Deconvolution of seismic data, CSEG Conference Abstract.

- Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: *Geophysics*, 64, 208-221.
- Plessix, R. E., 2007, A Helmholtz iterative solver for 3D seismic-imaging problems: *Geophysics*, 72, P.SM185-SM194.
- Saad, Y., 2000, *Iterative methods for sparse linear systems*, second edition.
- Sacchi, M. D., 2005, *Inverse problems in exploration seismology: Short Course note*, MITACS 2005, University of Alberta.
- Schneider, W. A., 1978, Integral formulation for migration in two-dimensions and three-dimensions: *Geophysics*, 43, 49-76.
- Stolt, R. H., 1978, Migration by Fourier transform: *Geophysics*, 43, 23-48.
- Strang, G., 1986, *Introduction to applied mathematics*: Wellesley-Cambridge Press.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, 49, 1259-1266.
- Whitmore, N. D., 1983, Iterative depth migration by backward time propagation: 53rd Annual International Meeting, Society of Exploration Geophysicists, Expanded Abstract, 382-385.
- Yousefzadeh, A. 2008, *Seismic Data Reconstruction with Regularized Least-Squares Prestack Kirchhoff Time Migration*, M. Sc. thesis, University of Alberta.
- Yousefzadeh, A., and J. C. Bancroft, J. C., 2009, Feasibility of solving least squares prestack Kirchhoff migration using multigrid methods: CREWES Research Report, Volume 21.
- Yousefzadeh, A., and J. C. Bancroft, J. C., 2010, Solving least squares Kirchhoff migration using multigrid methods: 80rd Annual International Meeting, Society of Exploration Geophysicists, Expanded Abstract, 3135-3139.