DATA REGULARIZATION AND DATUMING BY CONJUGATE GRADIENTS

DANIEL R. SMITH[1], MRINAL K. SEN[2] and ROBERT J. FERGUSON[3]

[1] *Hess Corporation, 500 Dallas St., Houston, TX 77002, U.S.A.*
[2] *John A. and Katherine G. Jackson School of Geosciences, The University of Texas at Austin, University Station, Box X Austin, TX 78713-8972, U.S.A.*
[3] *Department of Geoscience, University of Calgary, 2500 University Drive N.W., Calgary, Canada T2N 1N4. rjfergus@ucalgary.ca*

(Received May 5, 2010; accepted July 2, 2010)

ABSTRACT

Smith, D.R., Sen, M.K. and Ferguson, R.J., 2010. Data regularization and datuming by conjugate gradients. *Journal of Seismic Exploration*, 19: 321-347.

Irregular spacing of seismic sources and receivers, and strong topographic variations plus velocity heterogeneity, cause spatial and temporal irregularity in seismic data. Because so much of seismic processing, imaging, and inversion relies on the Fast Fourier transform for efficiency, and because seismic modelling requires continuous reflectors for analysis, seismic regularization is desirable. Here, we address spatial and temporal irregularity simultaneously. We use weighted, damped least-squares to extrapolate data from an irregularly sampled, topographic surface to a regularly sampled datum. This process requires an accurate velocity model of the near-surface, and it returns seismic traces with a constant trace-to-trace distance and more continuous reflection events. As an inverse problem, the Hessian in process is costly to compute, so the method of conjugate gradients (CG) are employed so that the required matrix-matrix multiplication is reduced to two matrix-vector multiplications. We find that use of the CG method reduces the total number of multiplication operations from $O(n^3)$ for the direct solution to $O(n^2)$ where n is the number of trace locations.

KEY WORDS: conjugate gradients, data regularization, seismic interpolation, damped least-squares, seismic inversion, wave equation statics, wave equation redatum.

0963-0651/10/$5.00 © 2010 Geophysical Press Ltd.

INTRODUCTION

There have been many methods presented in the literature for dealing with spatial aliasing and near-surface effects. Spatial anti-aliasing methods use some form of interpolation such as using filters (Spitz, 1991), wave-equation based methods (Ronen, 1987), or Fourier reconstruction (Duijndam et al., 1999). Near-surface effects are addressed with either statics (Cox, 1999) or wave-equation based methods (Berryhill, 1979). The quest for fast, cheap, and accurate methods inspired the present work here which uses a wave-equation based method and smoothing to achieve the desired results.

Spatial aliasing is commonly compensated for with interpolation. Spitz (1991) uses prediction filters in the f-x domain so as to avoid having to calculate true dips. He uses an interpolation operator to decrease the spatial sampling interval of the recorded data. The operator is made up of a linear combination of the Fourier coefficients and phase shifts for each frequency which is related to the time shifts of the event from trace to trace. This method, however, assumes that all the events are linear. Sacchi and Liu (2005) also try to reconstruct the wavefield in the spatial coordinate by estimating the Fourier components of the interpolated traces as an inversion problem while using regularization to obtain a stable and unique solution. It takes the input data and reconstructs the missing traces as if they were recorded with a denser array. Ronen (1987) also uses an inversion approach to the problem but incorporates wave propagation theory to help interpolate. The operator used to relate the coarsely sampled data to the model or the more finely sampled data is constructed as a block matrix of inverse DMO operators based on the wave-equation. Zwartjes and Sacchi (2004) modify a Fourier reconstruction algorithm to use only the unaliased lower part of the Fourier spectrum to predict the higher frequency components.

Cox (1999) gives a comprehensive look at statics and how to remove the time shifts caused by topography and complex near-surface velocities. Taner and Koehler (2002) lay out a method for correcting for the near-surface effects by correcting the amplitudes rather than applying time shifts. Berryhill (1979) gives a wave-equation based method for dealing with near-surface effects. He proposes to use wavefield extrapolation operators to downward continue the data to remove the phase distortion. The operators are derived from the Kirchhoff integral formulation of the scalar wave-equation. Bevc (1997) suggests upward continuing the data to "flood the topography" with a replacement velocity to remove the phase distortions caused by rugged topography. This method does not require prior knowledge of the near-surface velocity. Shtivelman and Canning (1988) compare statics and wave-equation based methods to show the superiority of the wave-equation based methods when there is near-surface complexity. Taner et al. (1982) and Kelamis et al. (2002) outline methods that

do not require a knowledge of the near-surface velocities, but rely on the arrival time discrepancies and the healing of the down-going wavefield. Zhu et al. (1992) show the superiority in using turning ray tomography to model near-surface velocities then the more traditional refraction analysis.

THEORY

A scalar, monochromatic wavefield at depth z,ψ_z, is estimated using a linear operator for wavefield extrapolation, $\mathbf{U}_{-\Delta z}$, and nearby wavefield, $\psi_{z+\Delta z}$ according to

$$\psi_z \approx \mathbf{U}_{-\Delta z}\psi_{z+\Delta z} \quad , \tag{1}$$

where ψ_z and $\psi_{z+\Delta z}$ depend on the x, y, z, ω coordinates, and subscript $-\Delta z$ indicates extrapolation upward towards the recording surface. We would like to find the model, or nearby wavefield $\psi_{z+\Delta z}$, that minimizes the energy of the squared residuals for this problem. The model is estimated by minimizing the square of the l_2 norm, **e**, [Ferguson (2006), see also Kühl and Sacchi (2003)]

$$E_d = \|\mathbf{e}\|^2 = \|\mathbf{W}_d[\psi_z - \mathbf{U}_{-\Delta z}\psi_{z+\Delta z}]\|^2 \quad . \tag{2}$$

where E_d is the prediction error, and weighting matrix $\mathbf{W}_d$ gives unit weight to live traces and zero weight to null traces. Finding $\psi_{z+\Delta z}$ that minimizes E_d determines the desired extrapolated-wavefield $\psi_{z+\Delta z}$.

Since the number of estimated traces exceed the number of actual traces, the regularization problem is underdetermined so damping is needed to establish uniqueness. This is done by adding constraints to the residual norms that penalize the objective function for excess roughness. The constraints take the form of

$$E_m = \|\mathbf{W}_m\psi_{z+\Delta z}\| \quad , \tag{3}$$

where E_m is the model norm and $\mathbf{W}_m$ is a spatial derivative which smooths to ensure minimum roughness is invoked to help find the simplest model. The cost function E to be minimized, which consists of the prediction error E_d and the model norm E_m, is

$$\phi(\psi_{z+\Delta z}) = E = E_d + \varepsilon^2 E_m \quad , \tag{4}$$

where ε is a scalar that controls how much the solution is penalized for roughness. Small values for ε results in a wavefield estimate that appears noisy, and large values give a smooth appearance. A qualitative value for ε, then, is found by experiment.

Minimization of cost function E with respect to desired wavefield $\psi_{z+\Delta z}$ gives the following least-squares solution (Ferguson, 2006)

$$\psi_{z+\Delta z} = [\mathbf{U}^{A}_{-\Delta z}\mathbf{W}_d\mathbf{U}_{-\Delta z} + \varepsilon^2\mathbf{W}_m]^{-1}\mathbf{U}_{-\Delta z}\mathbf{W}_d\psi_z \quad , \tag{5}$$

where $\mathbf{U}^A$ is the matrix adjoint of $\mathbf{U}$. Note, use of a least-squares estimate of an extrapolated wavefield is an implicit form of wavefield extrapolation (Berkhout, 1981).

From Margrave and Ferguson (1999),

$$\psi(x,z+\Delta z,\omega) = [\mathbf{U}_{\Delta z}\psi_z](x,z+\Delta z,\omega)$$

$$= [1/(2\pi)^2] \int \varphi(k_x,z,\omega)\alpha[v(x),k_x,\omega]e^{-ik_x x}\, dk_x \quad , \tag{6}$$

where nonstationary phase-shift operator α is given by

$$\alpha[v(x),k_x,\omega] = e^{ik_z[v(x),k_x,\omega]\Delta z} \quad , \tag{7}$$

and spatially-variable wavenumber k_z is

$$k_z[v(x),k_x,\omega] = \sqrt{\{[\omega/v(x)]^2 - k_x^2\}} \quad , \tag{8}$$

for $0 \leq [\omega/v(x)]^2 - k_x^2$ and

$$k_z[v(x),k_x,\omega] = i\,\mathrm{sign}(\Delta z)\sqrt{\{[\omega/v(x)]^2 - k_x^2\}} \quad , \tag{9}$$

otherwise, where 'sign' is the signum function. Eq. (9) ensures rapid numerical decay of energy in the evanescent region. The adjoint linear operator, $\mathbf{U}^A_{\Delta z}$, is given by (Margrave and Ferguson, 1999)

$$\psi(x,z-\Delta z,\omega) = [\mathbf{U}^A_{\Delta z}\psi_z](x,\Delta z,\omega)$$

$$= [1/(2\pi)^2] \int \varphi(k_x,z,\omega)\alpha^*[v(x),k_x,\omega]e^{-ik_x x}\, dk_x \quad , \tag{10}$$

where α^* is the complex conjugate of α or that $\alpha^*(\Delta z) = \alpha(-\Delta z)$.

The structure of the extrapolation matrices, $\mathbf{U}$ and $\mathbf{U}^A$, is given in Fig. 1 for a frequency of 16 Hz and a velocity model that is linearly increasing from 2000 to 4000 m/s. With this velocity model, $\mathbf{U}$ and $\mathbf{U}^A$ are sparse with most non-zero elements confined to the trace of each matrix and near-trace diagonals. In general, however, $\mathbf{U}$ and $\mathbf{U}^A$ are non-sparse, and any limited-diagonal approximation (Ferguson, 2006) will break down in strongly heterogeneous media.

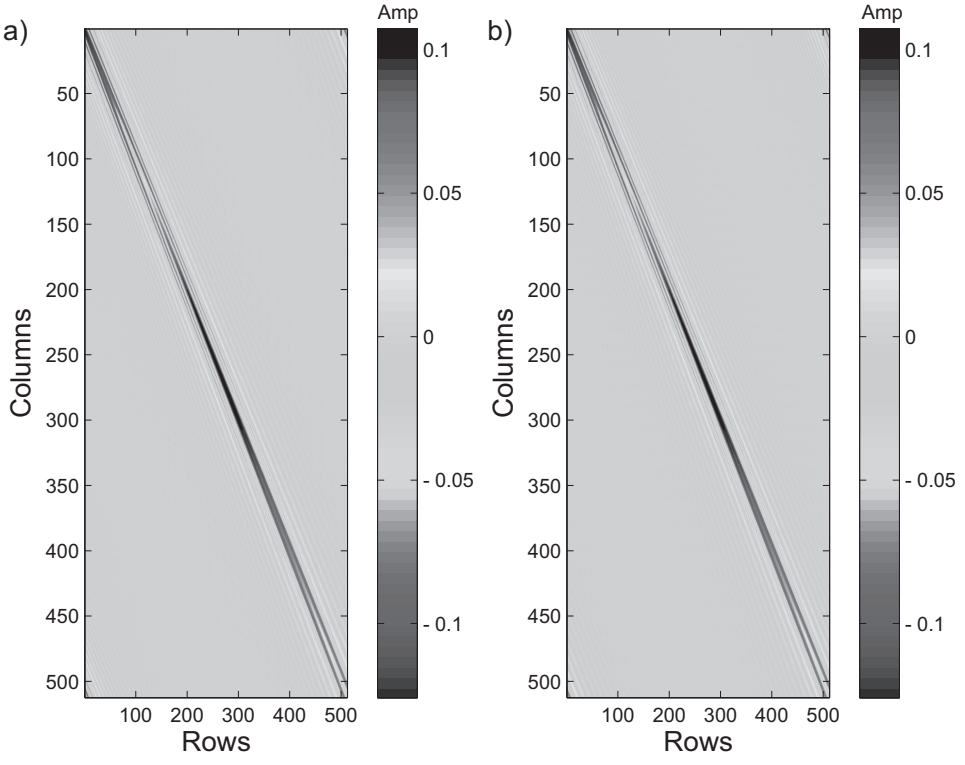

Fig. 1. Extrapolation operators. a) $\mathbf{U}_{\Delta z}$, and b) $\mathbf{U}^{A}_{\Delta z}$ its adjoint.

The solution in eq. (5) is costly to compute. It is dominated by the matrix-matrix multiplication of two dense matrices in the form of the composite operator $\mathbf{U}^{A}_{-\Delta z}\mathbf{W}_{d}\mathbf{U}_{-\Delta z}$. The number of floating point operations is $O(n^3)$ where n is the number of traces and the order is dependent on the rate of convergence to the solution. Ferguson (2006) overcomes this expensive cost by developing an approximate analytic solution to the Hessian. This, however, adds a dip limitation to the solution by filtering out the off-diagonal components and restricting the matrices to diagonal and near-diagonal entries. The next costly operation is the inversion of the Hessian and the number of operation is also $O(n^3)$ dependent on the numerical method used to invert the matrix. For example, LU decomposition and Gaussian elimination uses $\frac{2}{3}n^3$ operations if higher order terms are ignored (Golub and van Loan, 3 1996). The gradient $\mathbf{U}_{-\Delta z}\mathbf{W}_{d}\psi_{z}$ is the least dominate cost and if the weighting matrix is diagonal, then the cost is $O(n^2)$ with numerous efficient numerical techniques to carry this out.

The method of conjugate gradients (CG) is used to solve sparse linear systems such as equation 5 (Tarantola, 1987). It is widely used in the geosciences to solve boundary value problems. Yang et al. (2006) uses a preconditioned conjugate gradient algorithm for atmospheric tomography. Mackie and Madden (1993) and Zhang et al. (1995) use a conjugate gradient relaxation method for three dimensional magnetotelluric forward modelling and inversion. Mou-Yan and Unbehauen (1995) incorporates conjugate gradients in an image restoration technique for use in medical instruments and Youmaran and Adler (2004) for deblurring noisy data. Ronen (1987) uses conjugate gradients to perform a spatial spectral balancing in order to interpolate missing traces in seismic data. The conjugate gradient algorithm finds the optimal solution out of the possible solutions in an iterative fashion by starting at a point on the objective function representing the starting model, then moving from point to point while reducing the objective and updating the model estimate and search direction each time. This is done by using the fact that the Hessian, $\mathbf{U}^{A}_{-\Delta z}\mathbf{W}_{e}\mathbf{U}_{-\Delta z}$, is a positive definite matrix. When the Hessian is a positive definite matrix, the solution is the minimum of a quadratic form (Fig. 2) and the residuals are the direction of steepest descent and are orthogonal to each other.

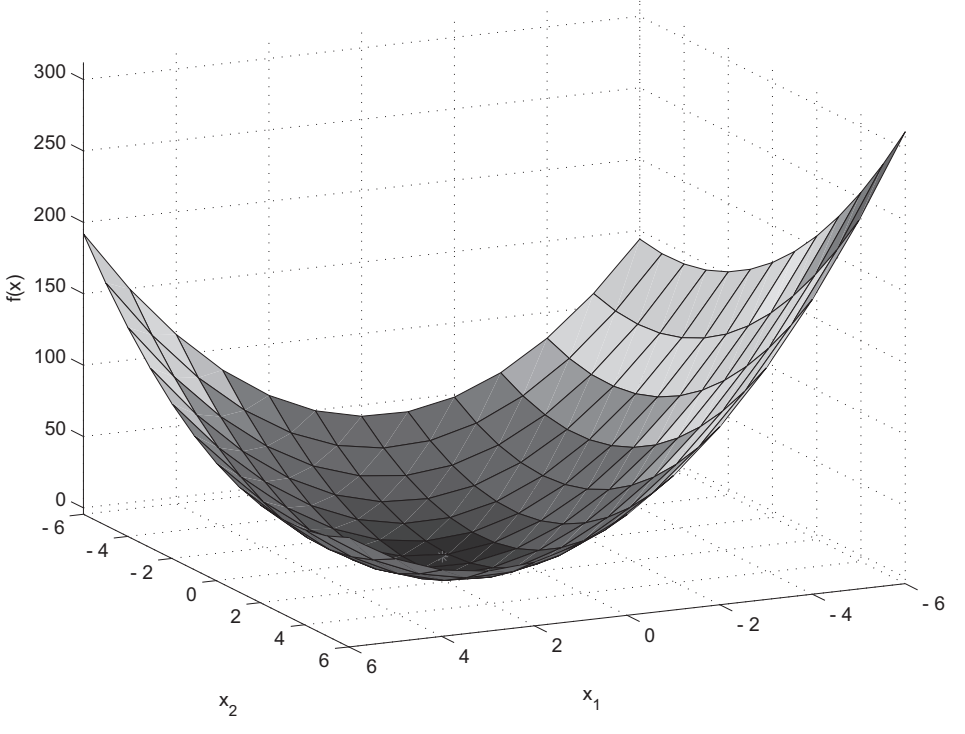

Fig. 2. Plot of a quadratic form. The minimum is the solution of $\mathbf{Ax} = \mathbf{b}$.

The method searches in the direction of the conjugation of these residuals. The true residuals are $\mathbf{r} = \psi_z - \mathbf{U}_{-\Delta z}\psi_{z+\Delta z}$ (Ferguson, 2006), whereas in the conjugate gradient method the residual polynomials are constructed (Nocedal and Wright, 1999) by a linear combination of the search directions and residuals. This works by realizing that the majority of the signal lies in the lower dimensional subspace then the higher dimensional data. If the assumption that most of the signal lies in the subspace spanned by $r_0, r_1, \ldots, r_k$ holds then the approximation x_k is a good approximation of the data. A line search is then conducted to find the step length γ by minimizing the function along one dimension or in other words a straight line along the search direction. If the start point is ψ_0, which can be completely arbitrary, each update thereafter to the model is

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \gamma_i \mathbf{p}_i \quad , \tag{11}$$

where $\mathbf{p}_0 = \mathbf{r}_0$. The step length γ is

$$\gamma_i = \mathbf{r}^t\mathbf{r}_i / \mathbf{p}_i^t\mathbf{A}\mathbf{p}_i \quad , \tag{12}$$

as derived by Shewchuk (1994) and $\mathbf{A}$ is the augmented matrix

$$\begin{matrix} \mathbf{W}_d\mathbf{U}_{\Delta z} \\ \varepsilon\mathbf{W}_m \end{matrix} \quad .$$

The new residual is constructed as

$$\mathbf{r}_{i+1} = \mathbf{r}_i - \gamma\mathbf{A}\mathbf{p}_i \quad , \tag{13}$$

The new direction is made as a conjugation of this new residual being

$$\mathbf{p}_{i+1} = \mathbf{r}_{i+1} + \beta_{i+1}\mathbf{p}_i \quad , \tag{14}$$

where β_{i+1} is a scalar multiplier derived from the conjugate Gram-Schmidt process. It is defined as

$$\beta_{i+1} = \mathbf{r}_{i+1}\mathbf{r}_{i+1} / \mathbf{r}_i^t\mathbf{r}_i \quad , \tag{15}$$

taken from Shewchuk (1994). The final model update $\mathbf{x}_k$ is the regularized and redatumed spectrum for a single frequency $\psi_{z+\Delta z}$.

The solution should converge in no more than n iterations. CG will choose the coefficients of the polynomial that will minimize the error. With each iteration, the polynomial tries to reach zero at each eigenvalue while attacking the larger eigenvalues which are associated with the larger errors first

[(Shewchuk, 1994), see also (Nocedal and Wright, 1999)]. The rate of convergence is determined on how fast these eigenvalues reach zero. If there are clusters of eigenvalues or eigenvalues with multiplicity > 1, it makes it easier to minimize the error as the polynomial can pass through the points at a lower degree and it should converge in iterations $< n$. Also, if the eigenvalues are greater than one, the solution will not converge. As an operator whose eigenvalues are greater than one is repeatedly applied to a vector within the conjugate gradient scheme, the vector grows in magnitude. Since these vectors are the residuals, the error is in turn growing. A singular value decomposition analysis is helpful in determining the stability of the wavefield operators but will not be explored here (Etgen, 1994).

The method can be made to terminate at a given tolerance, d, which can be defined as

$$\delta \leq \mathbf{r}_k/(\mathbf{r}_k - 1) \quad , \tag{16}$$

where $\mathbf{r}$ is the residual as described in eq. (13). The stopping criteria can be chosen in several different ways. Here, the data fit is used as the measure of tolerance. The minimal error conjugate gradient method (Hanke, 1995) uses the iteration error rather than the data fit to stop the iterations.

The construction of the residual and directional polynomials can lead to possible round-off errors due to finite precision. This can lead to non-convergence. With each iteration, the residuals become less orthogonal and the conjugate gradient method breaks down. Roundoff errors also cause the search directions to lose their conjugacy. van den Eshof and Sleijpen (2004), Schneider and Willsky (2001), and Bobrovnikova and Vavasis (2001) present ways to compensate for these errors. van den Eshof and Sleijpen (2004) made a slight modification to the CGLS algorithm (Hansen, 1994) to account for the round-off errors and the modified version of CGLS is the mcgls algorithm. The modification is in two parts. The first part is an added perturbation term to the formation of $\mathbf{Ap}_k$. This term improves the structure of the problem to avoid the ill-conditioned problem which presents itself as an elongated quadratic form. This elongation increases the area around the solution minimum that is within machine precision which makes it hard for the algorithm to find a suitable solution. The second part is to correct the recurrences by scaling them according to the perturbation term. This helps in the stability and robustness of the CGLS algorithm which is imperative in the regularization problem addressed in this paper.

The cost of the least squares conjugate gradient method improves upon the direct solution of eq. (5) by avoiding the direct computation of the Hessian. The most costly operations in the conjugate gradients are two matrix-vector

multiplications when multiplying the forward and reverse operators, $\mathbf{A}$ and $\mathbf{A}^T$, respectively, with the search direction $\mathbf{p}_k$. This takes n^2 floating point operations. There are several inner products to compute which take $O(n)$ operations. As long as the solution converges with a reasonable amount of iterations, i.e., $< n$, then the conjugate gradient algorithm greatly improves on the efficiency.

SYNTHETIC EXAMPLE

A synthetic data example is constructed to apply the method of inversion by least squares using conjugate gradients in order to test the viability of this method for regularization and datuming. The synthetic data are generated by a ninth-order finite differences modelling algorithm with a line source to simulate a flat reflector and point sources at depth to simulate seismic events with steep, conflicting dips. This allows the algorithm to be tested for its ability to reconstruct the wavefield with a wide range of dips. A laterally heterogeneous velocity model, shown in Fig. 3, is used to add the effects of a heterogeneous

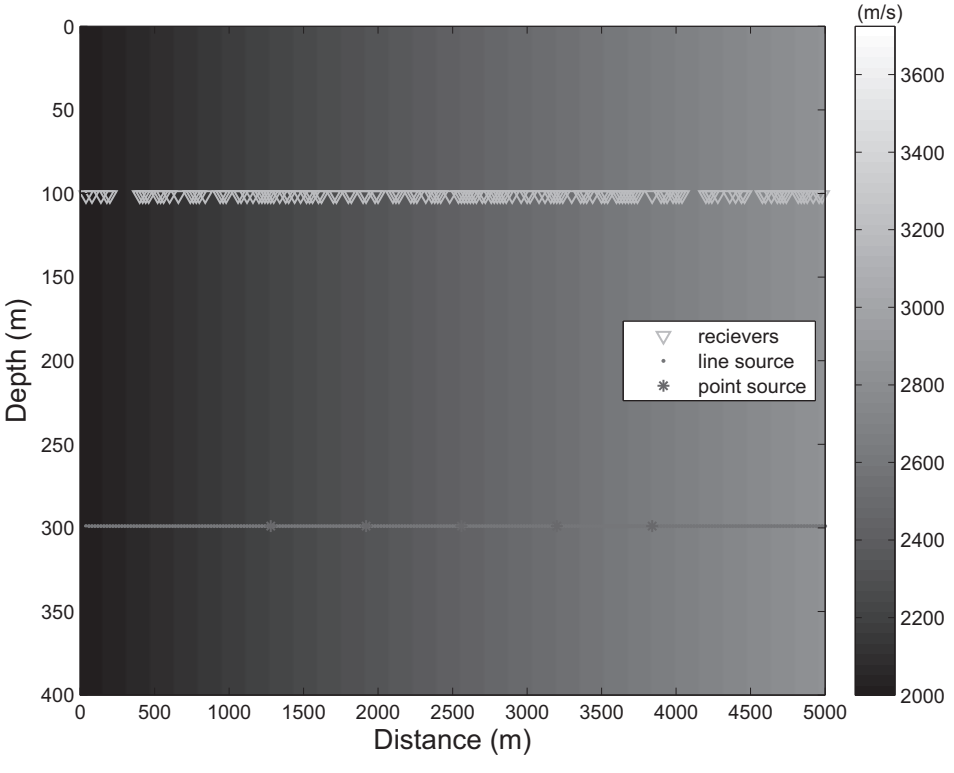

Fig. 3. The model setup used in the synthetic example. It simulates a line source (diamonds) and five point sources (large diamonds) and a highly irregular receiver array (triangles) (Ferguson, 2006).

near-surface layer. The velocity is linearly increasing from the left to right from 2000 to 4000 m/s. The experimental setup is an exploding reflector model with a flat line of weak sources and five strong point sources at a depth of 300m. The receiver array (green triangles in Fig. 3) consists of the live and dead receivers spaced at 20 m and is set at a depth of 100 m. The line of receivers are irregularly spaced to simulate a highly irregular data set. The depths of the receivers and sources were chosen to demonstrate the stability of the method over large extrapolation distances, which in this case is 200 m. Using the finite difference algorithm for the forward wavefield modelling, the sources are then exploded and the data are recorded at the receivers.

The data generated by this model are in Fig. 4a. Fig. 5b is a magnified view showing the flat reflector and apexes of the point diffractions. The record length is just over one second with a temporal sampling interval of 4 ms. The data are padded with zeros so that the total number of traces is 512 in order to suppress operator wrap-around and to get a power of 2 in the number of traces to accommodate the FFT algorithm. A 1D Fourier transform from t - ω using the FFT was applied to the data in Fig. 4 to decompose the wavefield into its planewave components. Each row of the transformed data is a monochromatic wavefield spectrum and the range of frequencies is user defined depending on

Table 1. The Least-Squares Conjugate Gradient algorithm following van den Eshof and Sleijpen (2004).

$\mathbf{s}_o = \mathbf{b}; \mathbf{r}_o = \mathbf{A}^T\mathbf{s}_o$

$\mathbf{p}_o = \mathbf{r}_o; \mathbf{x}_o = 0$

for $k = 1 : k_{max}$

$\mathbf{q}_{k-1} = \mathbf{A}\mathbf{p}_{k-1}$

$\gamma_{k-1} = \|\mathbf{r}_{k-1}\|^2/[\|\mathbf{q}_{k-1}\|^2 + \varepsilon\|\mathbf{p}_{k-1}\|^2]$

$\mathbf{x}_k = \mathbf{x}_{k-1} + \gamma_{k-1}\mathbf{p}_{k-1}$

$\mathbf{s}_k = \mathbf{s}_{k-1} - \gamma_{k-1}\mathbf{q}_{k-1}$

$\mathbf{r}_k = \mathbf{A}^T\mathbf{s}_k - \gamma_{k-1}\mathbf{q}_{k-1}$

$\beta = \|\mathbf{r}_k\|^2/\|\mathbf{r}_{k-1}\|^2$

$\mathbf{p}_k = \mathbf{r}_k + \beta_{k-1}\mathbf{p}_{k-1}$

end

$\psi_{z+\Delta z} = \mathbf{x}_k$

---

the bandwidth within the data that needs to be extrapolated or the bandwidth that contains the greater part of the signal. For this data set, the bandwidth selected from the data is from 12 to 40 Hz and the frequency sampling, Δf, is 0.9766 Hz. Each monochromatic wavefield is then fed into the least squares conjugate gradient algorithm. A pseudo-code is displayed in Table 1. The modifications to the algorithm in order to improve convergence are found in van den Eshof and Sleijpen (2004) algorithm 2. No a priori information is needed so the starting model for the conjugate gradient scheme is $x_0 = \mathbf{0}$. Each monochromatic wavefield is extrapolated to collapse the point diffractions and to move the line source to the depth of the sources at 300 m. After each wavefield is extrapolated, an isotropic phase-shift extrapolation operator is applied to the solution to back extrapolate the wavefield to the desired depth using a constant velocity. In this case, it was extrapolated back to the depth of 100 m or the depth of the receiving array for comparison with the original data.

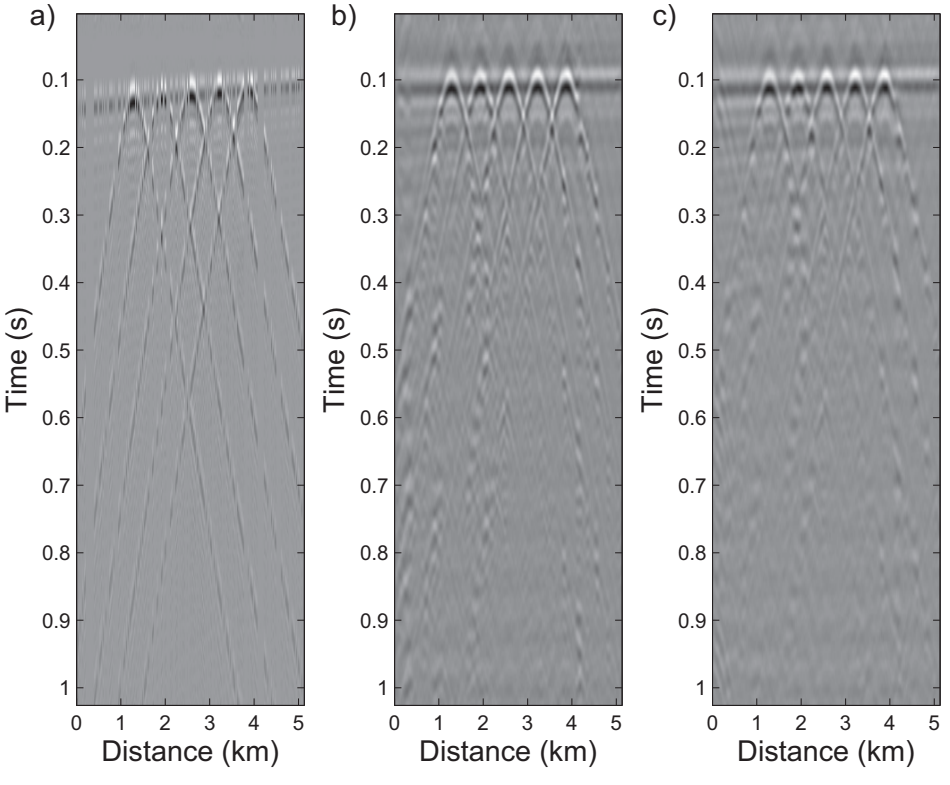

Fig. 4. Plot of the results for the synthetic seismic data set. a) The input data, b) solution using the conjugate gradient scheme and c) Newton's least square solution.

The velocity used was the mean of the velocity array which in this case was 2862 m/s. A band-limited 1D inverse Fourier transform was applied to get the results displayed in Fig. 4b with a magnified view in Fig. 5b. The results show the line source is continuous as well as the steep dipping events and apexes of the point diffractions. The velocity effects are removed from the data. This can be seen by the repositioning of the flat reflector to a horizontal direction from a sub-horizontal direction.

The smoothing operator, ε in eq. (5), is found by trial and error and a value of 0.5 was applied to produce the desired results, reducing the noise sufficiently but not oversmoothing the result. The tolerance level, defined in eq.(16), was chosen to be 1% of the magnitude of the original residual. The convergence histories for a low (15 Hz), moderate (25 Hz), and high (39 Hz)

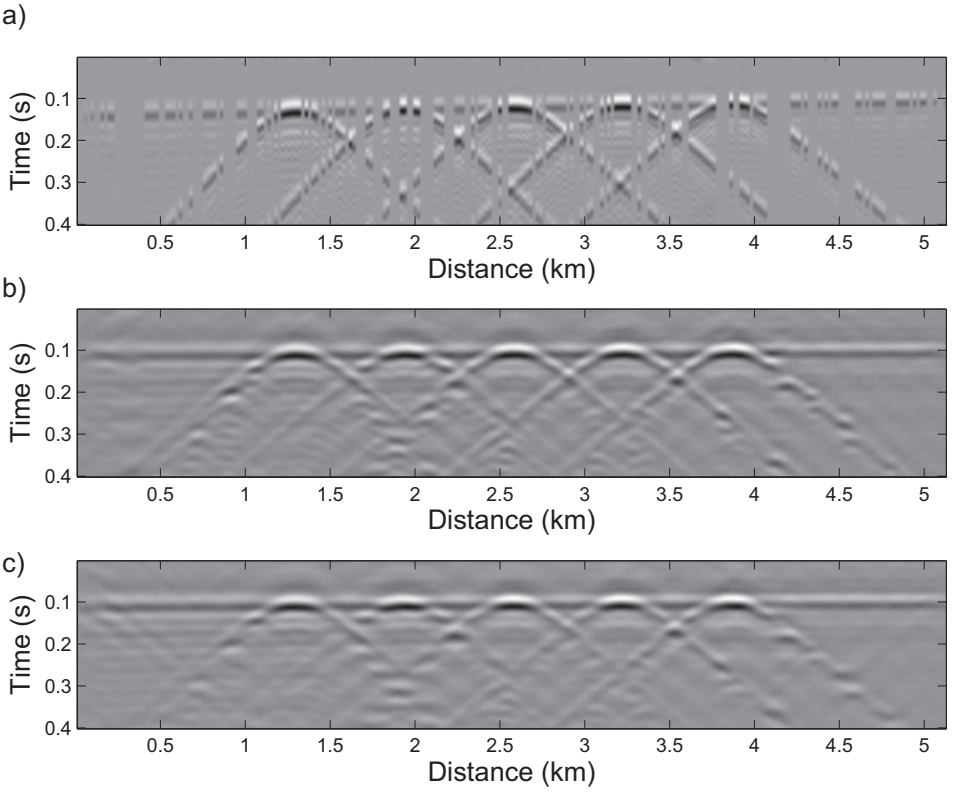

Fig. 5. A magnified view of the results in Fig. (4) for the synthetic seismic data set. a) The input data, b) the solution using the conjugate gradient scheme and c) Newton's least-square solution.

frequency are shown in Fig. 6. The error reduction within the first few iterations is greater than in later iterations. This is expected as the conjugate gradient method attacks the eigenvalues associated with the larger error first and then moves to reduce the smaller errors. For the histories of 15 and 25 Hz in Fig. 6, the error reduction slows and then there is a sharp drop. This could be attributed to clusters of eigenvalues where with each iteration the error reduction associated with distinct eigenvalues is small, but with a group of eigenvalues the error reduction is much larger in a single iteration. The average number of iterations for all frequencies is 19 with the higher frequencies requiring, for the most part, more iterations to converge.

The exact Newton's method solution is displayed for comparison in Fig. 4c with a magnified view in Fig. 5c. These results are achieved by the direct solution of eq. (5) for each monochromatic wavefield. In comparing the conjugate gradient method results to that of the Newton's method, there is no discernible difference. However, due to the direct computation of the Hessian

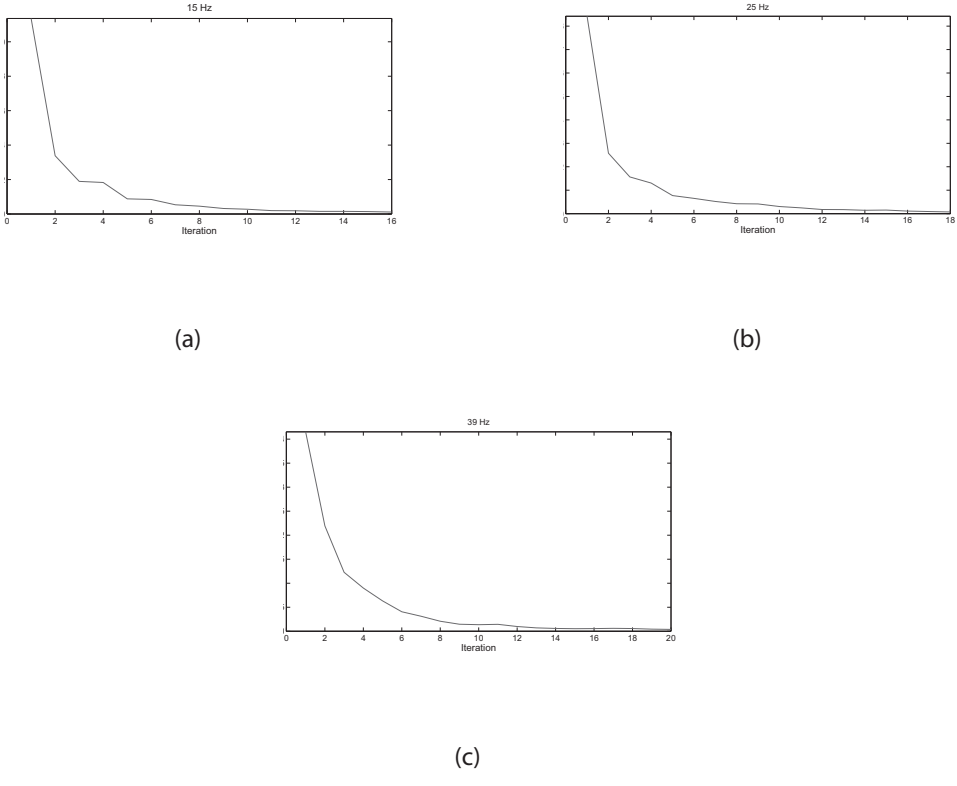

Fig. 6. Plot of normalized error vs. iteration for various frequencies. a) 15 Hz, b) 25 Hz, and c) 39 Hz.

within the least-squares solution, the exact Newton is computationally costly. An approximation can be applied to make the Hessian more efficient as done by Ferguson (2006) but this adds a dip limitation onto the outcome. The computations for each monochromatic wavefield are $O(n^3)$ as previously discussed.

Fig. 7 shows the data fit for a single frequency. The blue line is the input data or the monochromatic wavefield spectrum. The red and green lines are the result of the conjugate gradients and Newton's method respectively. The results from the conjugate gradient and Newton's method are similar and show the reconstruction of the missing traces. The methods reconstruct the zeros in the input data thereby filling in the missing traces as can be seen in Fig. 8a once the wavefield is extrapolated and transformed back to the time domain. The results are also smoother than the input data due to the constraints imposed on the inversion by the operator $\mathbf{W}_m$. The velocity effects are removed as seen in Fig. 8a. The conjugate gradients and Newton's method both shifted the trace to correct for the low-velocity effects.

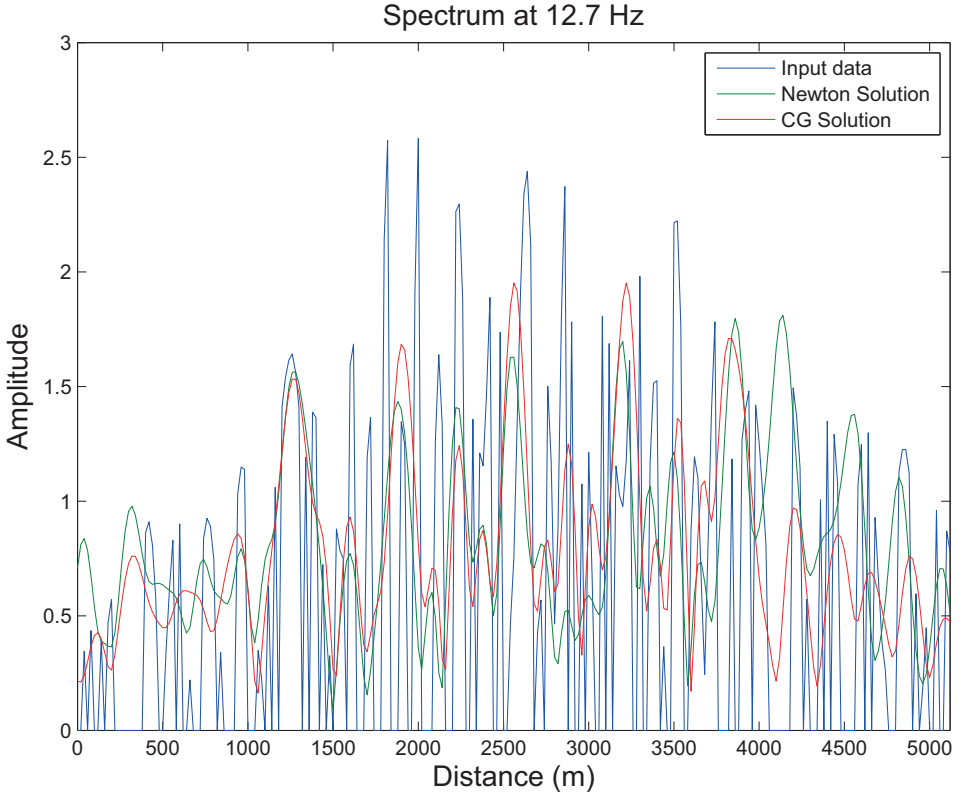

Fig. 7. Plot of the results of the extrapolation of a single monochromatic wavefield spectrum input with the blue line as the input data. The green line is the Newton's least square solution and the red line is the conjugate gradient solution.

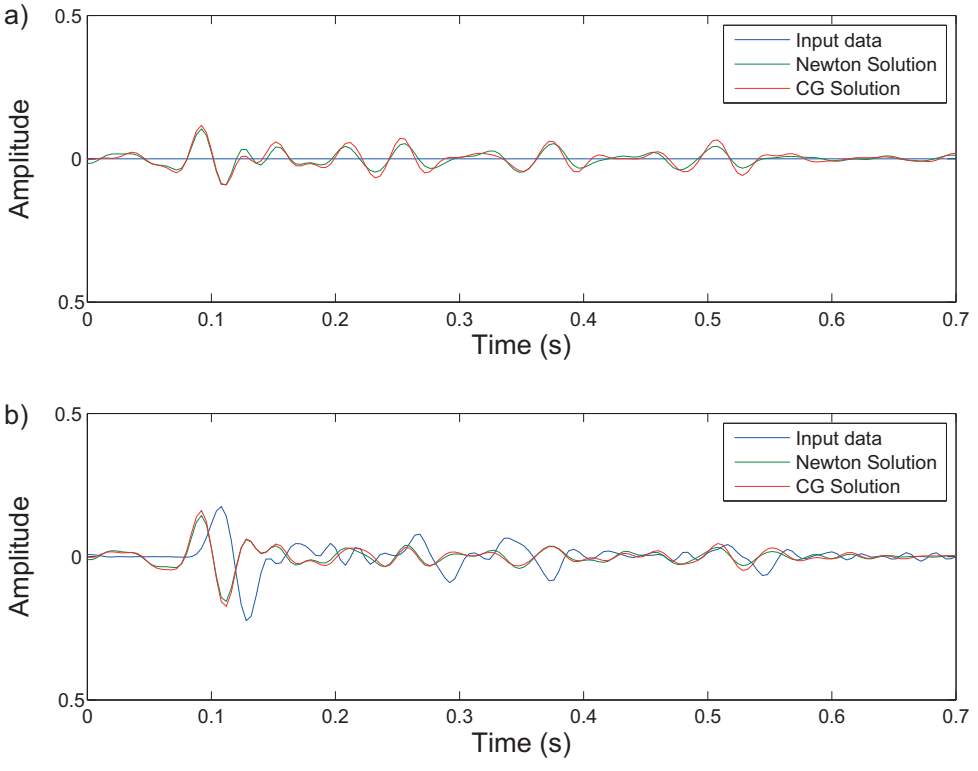

Fig. 8. Plot of the results for a single reconstructed trace and a redatumed trace with blue line as the input data, the red line is the CG result and green is the Newton's result. a) The reconstructed trace and b) the redatumed trace.

Results on this data set from other common methods for redatuming and trace interpolation are shown in Fig. 9. The PSPI method (Gazdag and Sguazzero, 1984) is in Fig. 9a. This method takes and extrapolates the data for various reference velocities and then interpolates to get the full data section. An f-x operator as derived by Spitz (1991) is used to get the results in Fig. 9b. This operator tries to interpolate traces through a filtering approach where the filter is derived from a spatial prediction filter. The assumption though is that the events are linear. The flat reflector is discontinuous and the diffraction events are not fully reconstructed. Fig. 9c is similar to the Newton's method but it uses an approximation for the Hessian in the Newton's method (Ferguson, 2006). The data is reconstructed well but there is a dip limitation due to the approximation.

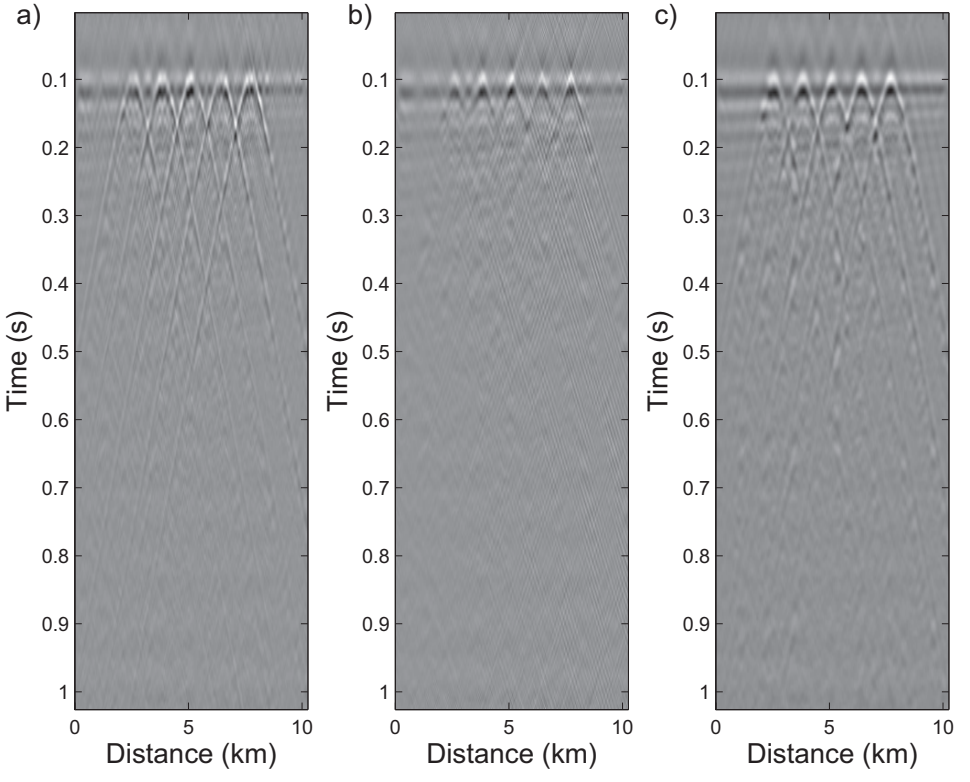

Fig. 9. Results for using alternate methods on the synthetic data set. a) The PSPI method following Gazdag and Sguazzero (1984), b) an f-x operator as derived by Spitz (1991), and c) using an approximation for the Hessian in the Newton's method (Ferguson, 2006).

CANADIAN FOOTHILLS EXAMPLE

The synthetic data example was designed to demonstrate the stability of the method over large extrapolation distances. The large extrapolation steps can only be used when the velocity is constant in the vertical direction within the large extrapolation step. In a real world example, the velocity usually varies significantly in the vertical direction over similar distances and so the large extrapolation steps cannot be used. When the near-surface is highly heterogeneous, the extrapolation steps need to be small enough so that the assumption that the velocity is constant in the vertical direction over the extrapolation step holds. To demonstrate this method for a real world example, a real seismic data set of source and receiver pairs are continued downward in

order to redefine the reference datum on which the pairs are located. The wavefield $\psi(x,z = 0,t)$ is extrapolated down to a new depth so that the new wavefield is now $\psi(x,z = z_1,t)$ where z_1 is the new depth and use z_1 as the new reference surface. This can then be done in a recursive process where the extrapolation is done over several steps and the output wavefield of the previous is used as the input wavefield of the current step.

Wavefield extrapolation can also be done recursively when there is irregular topography using a filter as shown in (Reshef, 1991). The column and row positions in the filter coincide with the columns and rows of the velocity model. The filter applies zero weight to any receivers above the irregular topography and unity weight to the receivers at or below the irregular topography down to the depth of the current step. Therefore, the extrapolation process is started using the receiver at the highest elevation. This receiver is extrapolated down to the next depth using the velocity corresponding to that elevation and horizontal coordinate from the velocity model. At the next step, the receivers are activated that are at or above this new reference surface. This is continued until the minimum of the topography has been passed so the new datum is either flat or the process can continue until an irregular datum (i.e., weathering layer) is established by shutting off the receiver once it has passed the desired depth. After the wavefield extrapolation process of regularization and redatuming is done, an isotropic wavefield extrapolation can be applied to move the datum to the desired height as done in the synthetic data example. The real data example is from the Canadian foothills in Alberta and a typical shot gather which is used for an example is shown in Fig. 12a. The data were acquired by Husky Oil Ltd. and is a publicly accessible data set. The data set was made accessible by Husky Oil and Talisman Energy to be used as a benchmark example of a 2D seismic land data in an area with complex structure and topography. The area where the data were recorded is in the Canadian foothills overthrust belt region. The region is characterized by overthrust structures of various geometric complexity dominated by the typical fold-fault structure and stratigraphy units ranging from carbonates, shales as well as other clastics (Stork et al., 1995). Because of the complexity of the subsurface and the topography, this area has historically been a challenge for imaging. Traditional statics can have problems in this area due to the large variations in velocity and topography. The imaging of this area was sensitive to even small variations in the velocity ($< 5\%$) (Stork et al., 1995).

A common shot gather (Fig. 12a) is used to demonstrate the effectiveness of the method described in this paper on a real data example. The vertical axis is two-way traveltime. The horizontal axis is the distance along the entire survey line so the receiver array is approximately 6 km long starting for this shot at 7686 km and ending at 13684 km from the start of the survey line with a receiving spacing of 20 m. The source was excited at 10645 km and the duration of recording was just over 4 s with a temporal sampling interval of 4

ms. Reflection events in a shot gather should approximately follow a hyperbolic curve. The reflection events in Fig. 12a are discontinuous and do not follow such a curve. The steep, linear events present are caused from ground roll and also present problems in processing. The velocity model in Fig. 10 shows the strong lateral variations within the aperture of the seismic survey. The velocity model (Fig. 10) for this area was derived by turning-wave tomography (Stork, 1994) as the near-surface heterogeneity did not allow for refraction velocity analysis. The red line in Fig. 10 are the elevations of the receivers and shows the highly irregular topography along the line. The elevation changes approximately 300 m along the line. The effects of the near surface and topography can be seen in Fig. 12a by the distortions in the reflection events, especially the strong event at just over 3 seconds.

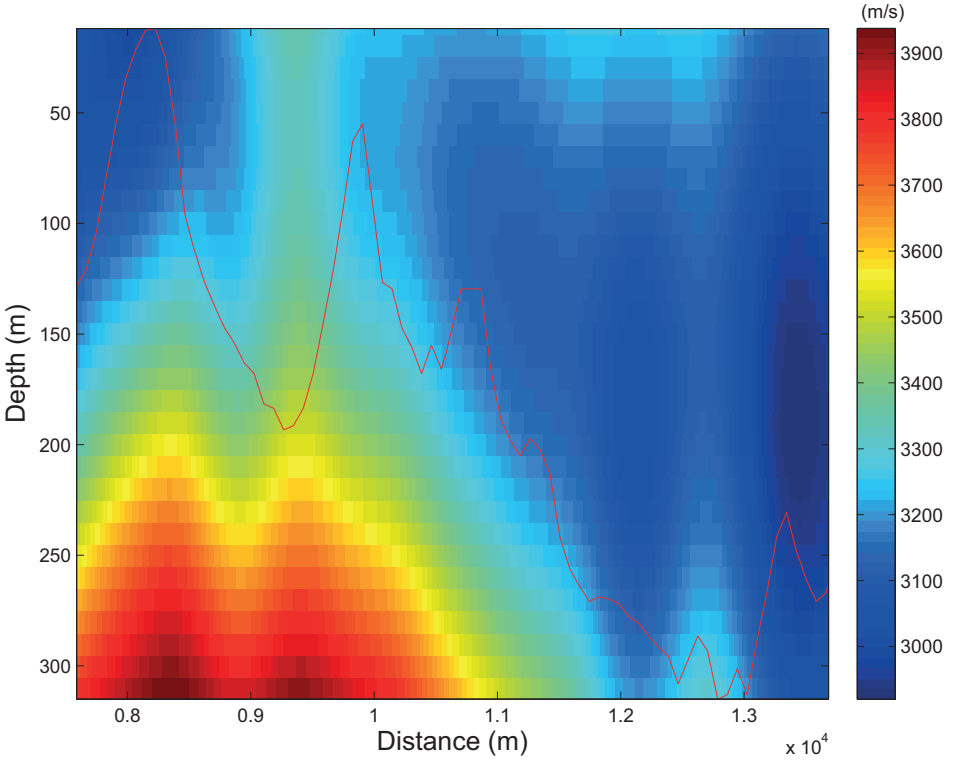

Fig. 10. Velocity model used to regularize and redatum the shot record. The topography is shown in red.

The receiver spacing for this line was fairly regular as shown by the blue circles in Fig. 11. Receivers are small and mobile and can be placed in areas of irregular topography. Shot sources or shot-hole drilling units for land can be large and cumbersome depending on the strength of the source needed to image the problem. The red diamonds in Fig. 11 show the locations of the shots taken along the seismic line. The irregular spacing of the shots is because of inaccessible areas to the sources due to the irregular topography. The common shot gathers (field records) as represented in Fig. 12a are for the most part regularly spaced. The common receiver gathers, however, are irregularly spaced due to the irregular shot spacing and therefore are sparsely sampled along the horizontal coordinate. This can present problems in processing steps such as migration by introducing artifacts and noise due to spatial aliasing.

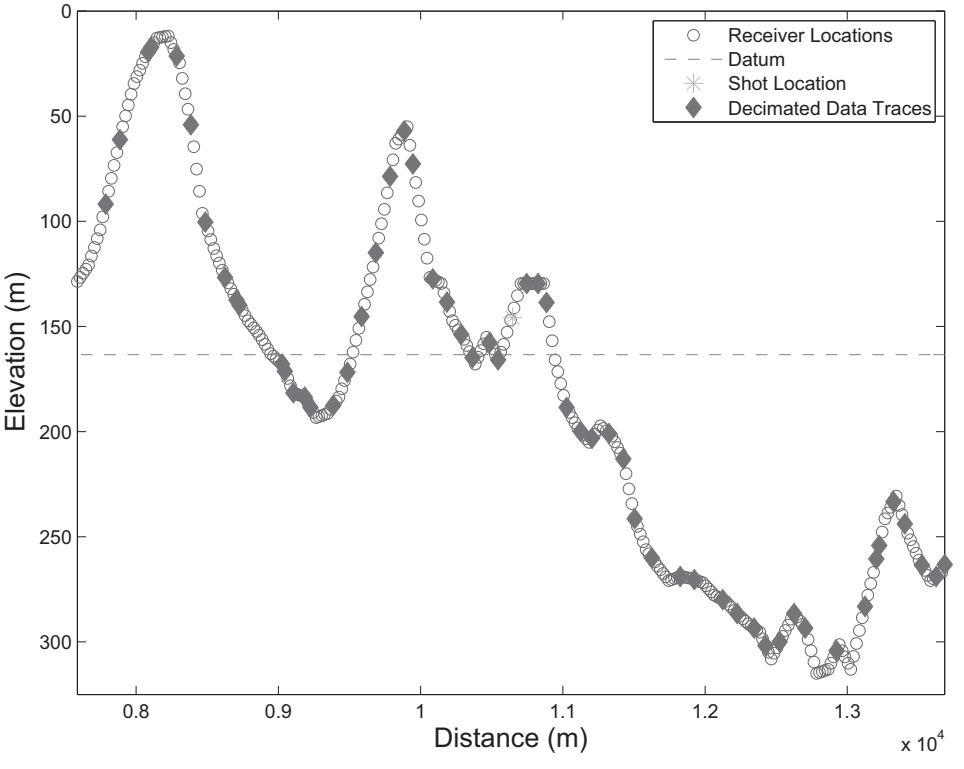

Fig. 11. Elevation profile shown by the positions of the receivers by blue circles. The red diamonds show the positions of the shots along the line with the green asterisk as the shot position of the input data in Fig. 12. The green dashed line shows the height of the final datum after the wavefield extrapolation.

The common shot gather in Fig. 12a was decimated from 300 traces to 56 traces to simulate a truly sparsely sampled common receiver gather (Fig. 13a). The common shot gather was taken and decimated to simulate a sparse gather instead of using a common receiver gather so we could see the quality of the reconstruction and redatuming by comparing with the original regularly sampled data. However, for complete redatuming of the seismic data, both common shot and common receiver gathers should be redatumed. Therefore, the conjugate gradient based method was applied recursively to the full and decimated gathers using the filter as described above incorporating the velocities in Fig. 10 in the computation of the wavefield extrapolation operators. The source or receiver gathers are then back extrapolated to the elevation of the green dashed line in Fig. 11 by an isotropic wavefield extrapolation operator using a constant velocity of 3184 m/s which is the mean of the first row of the velocity model. Figs. 14a-d shows the input and results for various values of ε as used in eq. (5). For $\varepsilon = 0.2$ (Fig. 14b), there is significant noise present due to the lack of smoothing caused by the low value of the smoothing parameter.

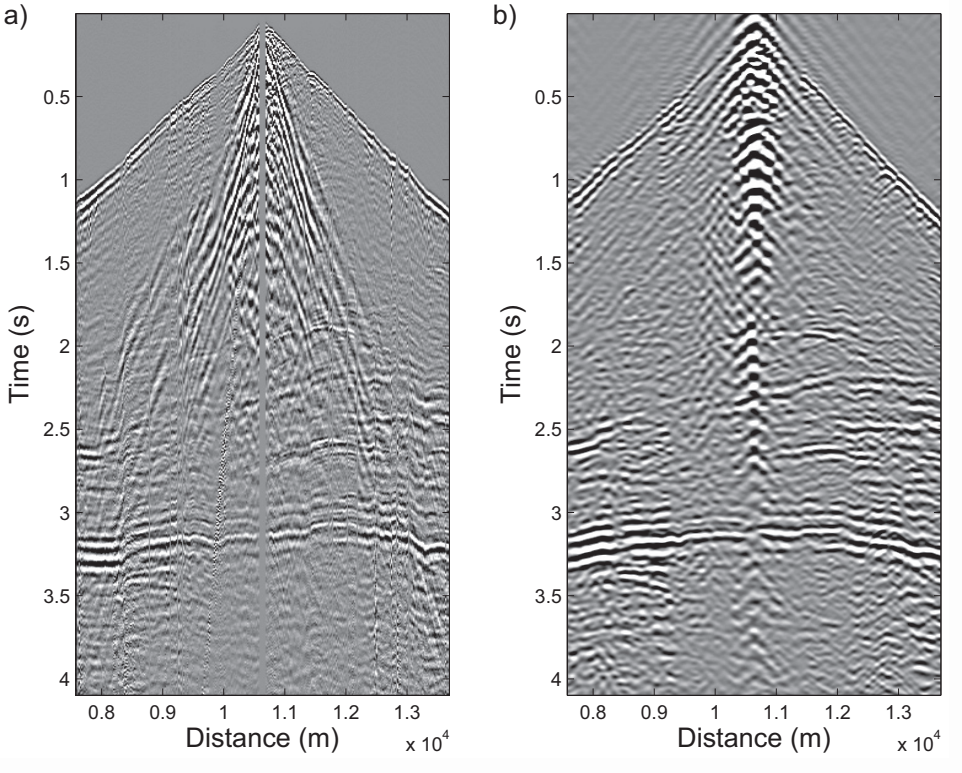

Fig. 12. Input data. a) Shot record, b) redatumed results of the input data in Fig. 12a using CG method.

The reconstructed data show significant diffraction off the edges of the original traces due to the extrapolation process which is in agreement with wave propagation theory. A spherical wave will be generated from the gap, and these waves present themselves as hyperbolic diffractions on seismic records. Greater smoothing is needed to rid the section of these diffractions. For $\varepsilon = 0.5$ (Fig. 14c), a good result is achieved and the diffractions have been smoothed over but still some noise is present and will be discussed below. For $\varepsilon = 0.75$ (Fig. 14d), the noise has been removed but the result seems to have an overly smooth appearance so this value is disregarded. A smoothing parameter of $\varepsilon = 0.5$ was determined to give the most desirable result and was used to obtain the results shown in Figs. 12b and 13b.

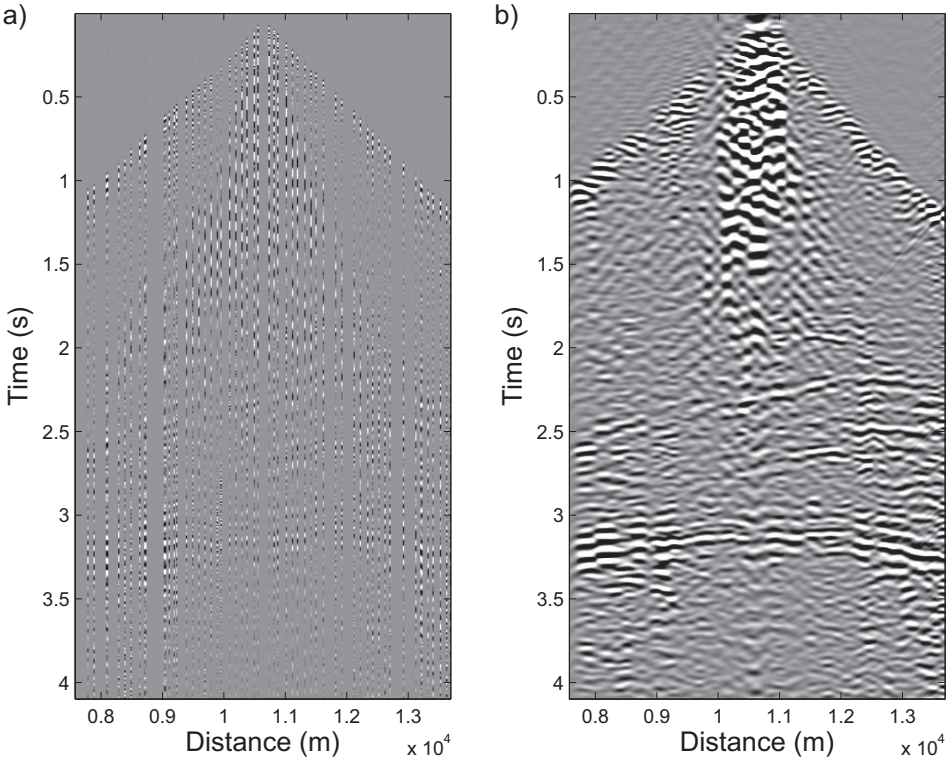

Fig. 13. Input data a) Decimated record, b) Regularized and redatumed results of the input data in Fig. 13a using CG method.

The results in Fig. 12b of extrapolating the gather from the surface topography to the datum show improved lateral continuity in the reflections with the input data in Fig. 12a for comparison. The regularized and redatumed gather in Fig. 13b shows the robustness of the method in that it successfully reconstructs the data and improves the lateral continuity of the reflectors through redatuming. For example, the strong reflector in both gathers at just over three seconds is laterally continuous. The missing traces have been successfully filled in for the reconstructed gather. The ground roll has also been suppressed due to the finite bandwidth that was extrapolated. Steep events are associated with high wavenumbers, k_z, the extrapolation process only extrapolates a certain spatial bandwidth. If these steep events are outside of the bandwidth, then these events are filtered out.

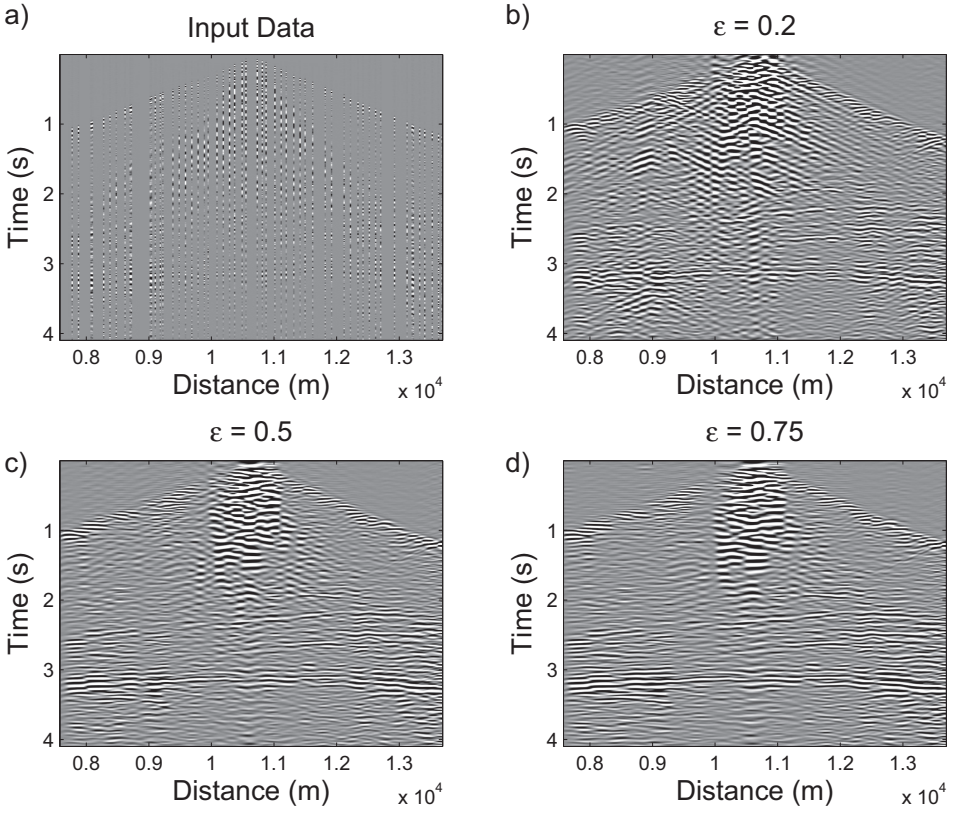

Fig. 14. Decimated shot gather after regularization and redatuming with varying values of the smoothing parameter ε. a) Decimated input data, b) $\varepsilon = 0.2$, c) $\varepsilon = 0.5$, and d) $\varepsilon = 0.75$. Notice the increase in noise with decreasing ε and increased smoothing with increasing ε.

Figs. 12b and 13b however still have some noise present. There are still some problems with spatial aliasing in the reconstruction. In the upper part of the section near the apexes of the reflection events, the spatial aliasing is evidenced by the upward change in direction of the flanks of the reflection events. This is also an issue with the spatial bandwidth that is being extrapolated. Events outside the bandwidth can be aliased and consequently treated as lower frequencies. Figs. 15a-d shows the f-k spectrum for the various gathers. Notice the aliased energy in Figs. 15a and 15c for the two input gathers. The aliased energy is suppressed in the resulting gathers of the regularization and redatuming process. The evanescent energy is also successfully filtered out as seen in the lack of energy outside the red lines which represent the boundary of the evanescent region. This decrease in the spatial bandwidth could be the cause of the problems in the treatment of higher spatial frequencies seen in the results. The aliased energy is represented as lower frequency content. This could also be associated with the instability of the operators in the lower frequencies (Etgen, 1994).

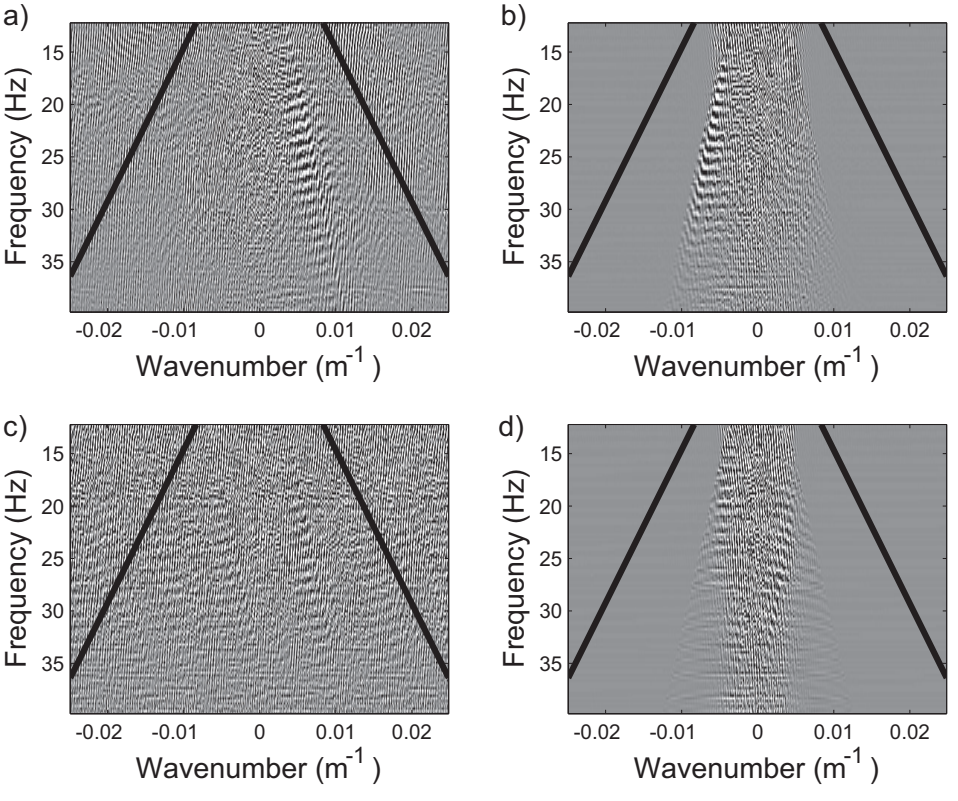

Fig. 15. The f-k spectrum of a) the shot gather, b) the redatumed shot gather, c) the decimated shot gather, and d) the redatumed and regularized shot gather. The red lines indicate the boundary of the evanescent region.

CONCLUSIONS

A method for regularization and redatuming is presented using conjugate gradients for the optimization of the solution (equation 5) for weighted damped least squares. The method is used to downward continue irregularly sampled data onto a regularly spaced grid while removing the low-velocity effects of the near-surface. This method is applied to highly irregularly sampled synthetic data as well as real data as well as decimated versions of real data to demonstrate its effectiveness. The synthetic data are generated to test the stability of the model with extreme conditions such as large extrapolation steps and conflicting, steep dips. The real data are taken from an area of complex structure where there are large distortions from the highly heterogeneous near-surface and irregular topography. The real data are also irregularly sampled in the horizontal coordinate as can be seen from the shot spacing. The data in this area require extensive processing to achieve moderate results in imaging. Removal of the near-surface effects and regularizing using a more accurate method could help solve some of the issues plaguing the imaging results in this region.

Weighted, damped least squares problems present themselves in many applications, especially with boundary value problems. The first step in solving these problems is developing an accurate forward model [eq. (1)]. With an accurate forward model, we can solve the inverse problem by finding a model estimate that best fits the forward relationship between model and data. The data are significantly undersampled causing an underdetermined problem where there are more model parameters than data available. When the problem is ill-posed as there is insufficient data to solve the problem, damping is needed to ensure a unique solution. Damping is simply adding equations to the system of equations in the forward model which puts constraints on the model. In this case, the operator $\mathbf{W}_m$ puts a penalty of minimum roughness on the solution. This ensures that the model is the simplest geologic model that solves the system. Repeated tests on the real data reveal that a smoothing parameter of 0.5 gave the best results, balancing the noise and oversmoothing.

With computational costs in mind, the solution needs to be reached within the least amount of time. Conjugate gradients is an optimization technique that is used to minimize the error between the model and data. Conjugate gradients were employed to improve efficiency of computing the Hessian by reducing the computation of the Hessian from matrix-matrix multiplication to vector-matrix multiplication within each iteration. In other words, we do not need to construct the Hessian. We only need to know the effect of the Hessian on an arbitrary vector. Convergence to a solution with iterative techniques are sometimes a challenge. When dealing with finite precision which is always the case with digital data, the algorithm used needs to be accurate, robust, and stable for the problem at hand. The structure of the problem greatly affects the algorithm's ability to achieve a solution. With an ill-posed problem such as this due to the

irregular sampling, the structure of this problem was not suited for a traditional approach using conjugate gradients. With smaller eigenvalues and large number of iterations, the round-off errors worsen. Algorithms need to be efficient but also need to converge in a suitable amount of iterations while also reducing the error associated with each eigenvalue sufficiently and effectively. Several algorithms were used to attain a solution with only one producing acceptable results. The others either did not converge to within machine precision or just simply diverged. The multishift version of conjugate gradients incorporates some modifications to the traditional least squares conjugate gradients method. These modifications compensated for the loss of orthogonality and conjugacy by adjusting for round-off errors. The method was able to improve the structure of the problem in order to achieve convergence within 1% error and an average of 19 iterations. Other means of compensating for these errors could be to apply a preconditioner such as linear combinations of Chebyshev polynomials (Akçadoğan and Dağ, 2003) or to apply some other means of reorthogonalization, however these are not within the scope of this paper.

This method successfully regularizes and removes the lateral velocity effects from a synthetic model. The line source is continuous and the steep dips are restored. The method also successfully reconstructs the real data with some artifacts and also removes the phase distortions caused by the near-surface which is apparent in the improved lateral continuity of the reflectors. The f-k spectrum of the data and results show an improved result in which the aliased energy has been significantly reduced. In addition, the evanescent energy has been successfully filtered out.

In the future, we would like to expand this method to handle 3D data. The cost using the conjugate gradients method has been reduced by an order of magnitude when compared to the direct solution in 2D data and we expect to see a greater cost savings in 3D data. As with all multichannel 3D data processing, it is considerably more complicated in application compared to 2D data processing. Berkhout and Verschuur (2005) describe how to arrange monochromatic 3D data into large 2D matrices. This could be a starting point for development of a 3D algorithm incorporating the principles used in this thesis such as conjugate gradients for the indirect building of the Hessian and its inverse. For a complete 3D survey the computational cost for the direct solution is $O(10^{17})$ operations. The computational cost for each trace using the conjugate gradient method is $O(\mathrm{n})$, where n is the number of traces. This could translate into a computational cost for a 3D survey to be $O(10^{15})$ operations.

Irregularly sampled data and near-surface effects as well as irregular topography present significant challenges to the geophysical acquisition and processing communities. The acquisition teams try to develop geometries that will compensate for the finite length and sampling within a survey. Processing groups use all the tricks in the book or at least what their managers will allow

to bring out the best image. The interpreters who in turn use all the tools available to them to interpret the best possible geology that science has allowed them. I hope that this method will be a meaningful tool in creating a better understanding of the subsurface and what it holds.

REFERENCES

Akçadoğan, Ç. and Dağ, 2003. A parallel implementation of Chebyshev preconditioned conjugate gradient method. Expanded Abstr., 2nd Internat. Symp. Parallel Distrib. Comput., IEEE: 1-8.

Berkhout, A.J., 1981. Wave-field extrapolation techniques in seismic migration, a tutorial. Geophysics, 46: 1638-1656.

Berkhout, A.J. and Verschuur, D.J., 2005. Inverse data processing, a paradigm shift?. Expanded Abstr., 75th Ann. Internat. SEG Mtg., Houston: 2099-2102.

Berryhill, J.R., 1979. Wave equation datuming. Geophysics, 44: 1329-1344.

Bevc, D., 1997. Flooding the topography: Wave-equation datuming of land data with rugged acquisition topography. Geophysics, 62: 1558-1569.

Bobrovnikova, E.Y. and Vavasis, S.A., 2001. Accurate solution of weighted least squares by iterative methods. J. Matrix Analys. Applic., 22: 1153-1174.

Cox, M., 1999. Static Corrections for Seismic Reflection Surveys. SEG, Tulsa, OK.

Duijndam, A.J.W., Schonewille, M.A. and Hindriks, C.O.H., 1999. Reconstruction of band-limited signals, irregularly sampled along one spatial direction. Geophysics, 64: 524-538.

Etgen, J.T., 1994. Stability of explicit depth extrapolation through laterally-varying media. Expanded Abstr., 64th Ann. Internat. SEG Mtg., Los Angeles: 1266-1269.

Ferguson, R.J., 2006. Regularization and datuming of seismic data by weighted, damped least squares. Geophysics, 71: U67-U76.

Gazdag, J. and Sguazzero, P., 1984. Migration of seismic data by phase shift plus interpolation. Geophysics, 49: 124-131.

Golub, G.H. and van Loan, C.F., 1996. Matrix Computations. John Hopkins Univ.

Hanke, M., 1995. Conjugate Gradient Type Methods for Ill-posed Problems. Longman Scientific and Technical, Essex.

Hansen, P.C., 1994. Regularization tools, a matlab package for analysis and solution of discrete ill-posed problems. Numeric. Algor., 6: 1-35.

Kelamis, P.G., Erickson, K.E., Verschuur, D.J. and Berkhout, A.J., 2002. Velocity-independent redatuming: A new approach to the near-surface problem in land seismic data processing. Geophysics, 21: 730-735.

Kühl, H. and Sacchi, M.D., 2003. Least-squares wave-equation migration for avp/ava inversion. Geophysics, 68: 262-273.

Mackie, R.L. and Madden, T.R., 1993. Three-dimensional magnetotelluric inversion using conjugate gradients. Geophys. J. Internat., 115: 215-229.

Margrave, G.F. and Ferguson, R.J., 1999. Wavefield extrapolation by nonstationary phase shift. Geophysics, 64: 1067-1078.

Mou-Yan, Z. and Unbehauen, R., 1995. A weighted space restoration algorithm using the aperiodic model of deconvolution. 8th Ann. IEEE Symp. Computer-Based Medic. Syst., CBMS: 308-314.

Nocedal, J. and Wright, S.J., 1999. Numerical Optimization. Springer Science & Business Media, Inc., New York.

Paige, C.C. and Saunders, M.A., 1982. LSQR: An algorithm for sparse linear equations and sparse least squares. ACM Transact. Mathem. Software, 8: 43-71.

Reshef, M., 1991. Depth migration from irregular surfaces with depth extrapolation methods. Geophysics, 56: 119-122.

Ronen, J., 1987. Wave equation trace interpolation. Geophysics, 52: 973-984.
Sacchi, M.D. and Liu, B., 2005. Minimum weighted norm wavefield reconstruction for ava imaging. Geophys. Prosp., 53: 787-801.
Schneider, M.K. and Willsky, A.S., 2001. Krylov subspace estimation. SIAM J. Scientif. Comput., 22: 1840-1864.
Shewchuk, J.R., 1994. An introduction to the conjugate gradient method without the agonizing pain. Technical paper, School of Computer Science, Carnegie Mellon Univ.
Shtivelman, V. and Canning, A., 1988. Datum correction by wave equation extrapolation. Geophysics, 53: 1311-1322.
Spitz, S., 1991. Seismic trace interpolation in the f-x domain. Geophysics, 56: 785-794.
Stork, C., 1994. Demonstration of mva tomography with controls and constraints for determining and accurate velocity model for prestack depth migration. Expanded Abstr., 64th Ann. Internat. SEG Mtg., Los Angeles: 1338-1342.
Stork, C., Welsh, C. and Skuce, A., 1995. Demonstration of processing and model building methods on a real complex structure data set. Workshop #6, Ann. Internat. SEG Mtg., Houston.
Taner, M.T., Berkhout, A.J., Treitel, S. and Kelamis, P.G., 1982. The dynamics of statics. The Leading Edge, 26: 396-402.
Taner, M.T. and Koehler, F., 2002. Surface consistent corrections. Geophysics, 46: 17-22.
Tarantola, A., 1987. Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation. Elsevier Science Publishers, Inc., New York.
van den Eshof, J. and Sleijpen, G.L.G., 2004. Accurate conjugate gradient methods for families of shifted systems. Appl. Numer. Mathemat., 49: 17-37.
Yang, Q., Vogel, C.R. and Ellerbroek, B.L., 2006. Fourier domain preconditioned conjugate gradient algorithm for atmospheric tomography. Appl. Optics, 45: 5281-5293.
Youmaran, R. and Adler, A., 2004. Combining regularization frameworks for image deblurring: Optimization of combined hyper-parameters. Expanded Abstr., Can. Conf. Electric. Comput. Engin., IEEE Comput. Soc.: 723-726.
Zhang, J., Mackie, R.L. and Madden, T.R., 1995. 3-D resistivity forward modeling and inversion using conjugate gradients. Geophysics, 60: 1313-1325.
Zhdanov, M.S., 2002. Geophysical Inverse Theory and Regularization Problems. Methods in Geochemistry and Geophysics, Vol. 36, 1st ed. Elsevier Science Publishers, Amsterdam.
Zhu, X., Sixta, D.P. and Angstman, B.G., 1992. Tomostatics: Turning-ray tomography + static corrections. The Leading Edge, 11: 15-23.
Zwartjes, P.M. and Sacchi, M.D., 2004. Fourier reconstruction of non-uniformly sampled, aliased data. Expanded Abstr., 74th Ann. Internat. SEG Mtg.,Denver: 1997-2000.