Least Squares Pre-Stack Time Migration for Imaging and Reconstruction of Converted Waves

Aaron Stanton, Department of Physics, University of Alberta, aaron.stanton@ualberta.ca

Summary

Converted waves suffer from a low signal to noise ratio and stronger aliasing than P-waves. These characteristics are problematic when an adjoint migration operator is applied to the data. In this article we apply preconditioned least squares migration using a pre-stack Kirchhoff time migration operator to converted wave data. To precondition the inversion we apply a five-point triangle smoothing operator on the offset dimension. We show a synthetic data example in which the input data have been regularly decimated. We find several advantages to our approach when compared with Fourier regularization plus migration.

Introduction

Migration is an important step in the processing of converted waves. Asymmetric source and receiver side ray-paths make binning and stacking these data particularly inapt. Kirchhoff migration, using the adjoint operator, aims to solve the migration problem by integration. Because we record discrete data with a finite aperture and aim to solve for a discrete earth model integration is replaced with finite discrete summation. Another problem with Kirchhoff methods relates to the artifacts created by spreading amplitudes over travel time curves. While typically the bulk of the energy constructively sums in place of true reflectors, a great deal of energy persists in the form of "smiles" or migration artifacts. Least squares migration attempts to deal with these problems. Instead of simply using the adjoint operator, the inverse is sought by attempting to match the data predicted by the model with the observed data. Significant work has been done in the field of least squares migration. Nemeth et al. (1999) use a Kirchhoff operator to perform least squares migration of incomplete surface seismic and Vertical Seismic Profile (VSP) data. They find that the migration results are more focussed and suffer less from acquisition footprint compared with the adjoint operation. Kuehl and Sacchi (2003) use least squares migration with a Double Square Root (DSR) operator to express reflectivity as a function of subsurface position and ray parameter to be used for further Amplitude Versus reflection Angle (AVA) inversion. Rebollo and Sacchi (2010) use a Kirchhoff time migration operator with dip-oriented filtering to attenuate artifacts, while Guitton et al. (2007) use least squares migration to attenuate low frequency Reverse Time Migration (RTM) artifacts due to crosscorrelation of diving-waves, head-waves, or backscattered waves. The attenuation of artifacts is achieved via preconditioning using prediction-error filters.

While prior efforts have focussed primarily on P-wave seismic data, this article considers the application of least squares migration to converted waves. Converted waves suffer from some additional complications compared to P-waves. The signal to noise ratio can be significantly lower, and the lower velocity receiver side ray path makes the aliasing frequency of converted waves lower than their P-wave counterpart. One way to deal with these issues is to regularize the data prior to imaging (Cary and Zhang, 2011b). In this contribution we try to deal with these issues in the imaging step via preconditioned least squares migration. We show that the method can be used to either regularize the data, or to migrate the data. For regularization, the algorithm has some unique advantages over other methods, namely, it can handle irregular trace spacing (binning is not required), it can reconstruct weak diffraction energy accurately, and it can reconstruct regular or irregularly missing traces equally well. To illustrate these advantages we compare the least squares migration approach with Minimum Weighted Norm Interpolation (MWNI) for a synthetic data example where the input data have been regularly decimated.

Theory

Pre-Stack Time Migration of converted waves

The Kirchhoff operator is currently the most widely used migration operator. One of the main reasons for the algorithm's popularity lies in its simplicity. Kirchhoff migration involves integrating trace amplitudes over a reflectivity model. After travel times and Kirchhoff weights have been calculated, the migration process can be written as a trace by trace process (Audebert, 2001). This means that a small amount of memory is required for this process, and perhaps more importantly the task is highly parallelizable. The speed of this algorithm also makes it possible to perform iterative velocity analysis.

In the case of least squares migration the forward and adjoint migration operators are required. Given the simplicity of the Kirchhoff adjoint operator, the forward operator is straightforward to define.

In addition to its simplicity, the Kirchhoff operator also suffers from artifacts that could benefit from LS migration. Preconditioning that aims to mitigate these artifacts will be shown later in this paper.

The basis of Kirchhoff migration is the Huygen-Fresnel principle which states when a disturbance reaches a point this point behaves as a secondary source and that summing the waves from this source can be used to determine the form of the wave at a later time.

The forward Kirchhoff operator can be written

$$\mathbf{d}(s_x, g_x, t_d) = \sum_x \sum_t \mathbf{m}(t_m, x, h)W \tag{1}$$

where $\mathbf{d}(t_d, s_x, g_x)$ are the data, $\mathbf{m}(t_m, x, h)$ is the model, and W are the 2.5D converted wave Kirchhoff PSTM weights. Here t_m is the time from zero offset to the reflector in the model space, and t_d is the corresponding travel time that is the sum of two terms: t_s, the time a ray takes to travel from the source to the scanned model position (using the P-wave velocity), and t_g, the time a ray takes to travel back to the receiver (using the S-wave velocity). The adjoint, or migration operator can be written

$$\mathbf{m}(t_m, x, h) = \sum_{N_{trace}} \mathbf{d}(s_x, g_x, t_d)W. \tag{2}$$

To calculate $t_d = t_s + t_g$, we use $t_s = \sqrt{t_{s0}^2 + (\frac{x - s_x}{v_p})^2}$ and $t_g = \sqrt{t_{g0}^2 + (\frac{x - g_x}{v_s})^2}$, where $t_{s0} = \frac{t_m}{(1+\gamma)}$, $t_{g0} = \frac{t_m \gamma}{(1+\gamma)}$, and γ is the v_p/v_s ratio. In the case of anisotropic wave propogation higher order terms are added to these equations (Wang et al., 2013), but here we assume isotropy.

The 2.5D converted wave Kirchhoff weights are calculated via $W = \frac{z(1+cos\theta)}{2v_p^3 \sqrt{1 + 2\gamma cos\theta + \gamma^2}} (\frac{t_g}{\gamma^2 t_s} + \frac{\gamma^2 t_s}{t_g}) \sqrt{\frac{\gamma^2 t_s + t_g}{t_s t_g}}$ (Cary and Zhang, 2011a), where z is the image depth (estimated using v_p and t_{s0}), and θ is the opening angle of the downgoing P-wave and upgoing S-wave raypaths. The cosine of the opening angle, θ, can be calculated using the sum of the angle of the source-side ray with the vertical, θ_s, and the angle of the receiver side raypath path with the vertical, θ_g as $cos(\theta_s + \theta_g) = cos\theta_s cos\theta_g - sin\theta_s sin\theta_g$, where $cos\theta_s = \frac{z}{v_p t_s}$, $cos\theta_g = \frac{z}{v_s t_g}$, $sin\theta_s = \frac{x - s_x}{v_p t_s}$, and $sin\theta_g = \frac{g_x - x}{v_s t_g}$. If v_s is replaced with the P-wave velocity these equations break down to the conventional P-wave Kirchhoff migration equations.

In this paper we implement parallelization over offsets classes using a cluster with shared memory architecture.

In figure 1 the forward operator, $\mathbf{L}$ is applied to an impulse in model space (a). The forward operator diffracts the model into data space (b), while the adjoint collapses the diffractions back into model space. It is apparent that result of the adjoint operation (c) is not equivalent to the original reflectivity (a). This is because the Kirchhoff operator is not orthogonal ($L^T \neq L^{-1}$).
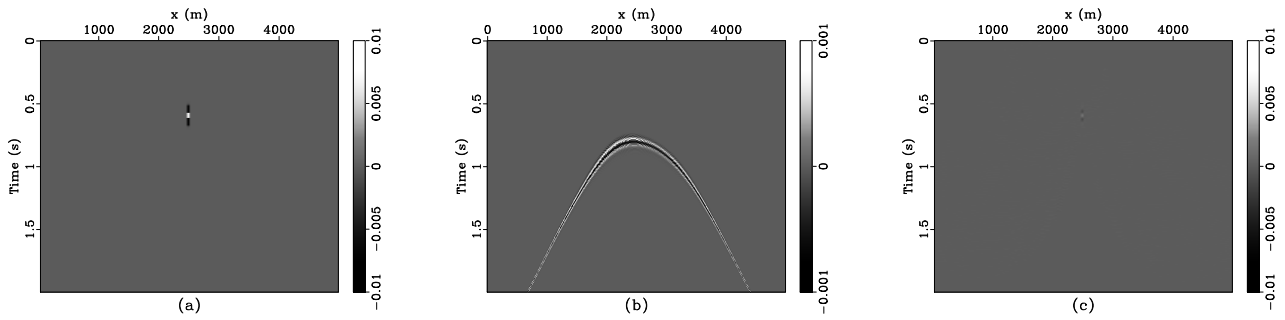
Figure 1 *An impulse in model space (a), the effect of spraying the model into the data domain using the forward operator (b), and migration of the data using the adjoint operator (c). Notice that the adjoint operation does not recover the original impulse. This is because the Kirchhoff operator is not orthogonal.*

Least squares migration

Given the forward Kirchhoff operator, $\mathbf{L}$, data can be generated from a reflectivity model. In seismic migration the goal is to collapse diffractions to the position where reflection occurred in the subsurface. To do this we should use the inverse of the forward operator, $\mathbf{L}^{-1}$. Typically for imaging the inverse is approximated by the adjoint operator, $\mathbf{L}^T$. While the adjoint can produce reasonable structural images, the recovered model is not true amplitude in the sense that applying the forward operator will not recover the recorded data to a reasonable degree, especially if the acquisition geometry contains significant holes, or irregular geometry (Nemeth et al., 1999). Least squares migration aims to recover the model, $\mathbf{m}$, such that the application of the forward operator reproduces the recorded data. The system of equations for least squares migration are $\mathbf{Lm} = \mathbf{d} + \mathbf{n}$, where $\mathbf{d}$ are the observed data and $\mathbf{n}$ is noise. The minimum norm solution comes from minimizing the cost function $J = ||\mathbf{Lm} - \mathbf{d}||_2^2 + \lambda^2 ||\mathbf{m}||_2^2$, where $\mathbf{m}$ is the reflectivity model to be recovered, and λ is a trade-off parameter which controls the under-fitting or over-fitting to the observed data, $\mathbf{d}$.

The minimum norm solution to this cost function is $\hat{\mathbf{m}} = (\mathbf{L}^T\mathbf{L} + \lambda^2\mathbf{I})^{-1}\mathbf{L}^T\mathbf{d}$. Typically the size of the model is greater than the size of the data which makes this an underdetermined problem. The next section discusses methods to precondition the problem in order to narrow down the range of solutions.

Preconditioning

To obtain a smooth solution a derivative operator can be applied to the model. Sharp changes in the model are then penalized in the cost function resulting in a smoother set of solutions. This is written $J = ||\mathbf{Lm} - \mathbf{d}||_2^2 + \lambda^2 ||\mathbf{Dm}||_2^2$, where $\mathbf{D}$ can be referred to as a "bad pass" operator, such as a derivative across the offset dimension of the migrated gathers.

An alternative way to write this is via a change of variables $\mathbf{z} = \mathbf{Dm}$, giving $J = ||\mathbf{LD}^{-1}\mathbf{z} - \mathbf{d}||_2^2 + \lambda^2 ||\mathbf{z}||_2^2$. If the derivative operator, $\mathbf{D}$, acts as a high-pass operator then its inverse, $\mathbf{D}^{-1}$ acts as a low-pass operator (Wang, 2005). It becomes clear that operators placed into the right hand side (model-space part) of the cost function are "bad-pass" operators that emphasize parts of the model we wish to penalize, while terms on the left hand side (data-space) are "good-pass" operators that emphasize parts of the model we wish to reinforce.

This cost function can be incorporated into a Conjugate Gradients framework such as given by (Scales, 1987), by simply concatenating the preconditioning operator, $\mathbf{D}$, with the migration operator, $\mathbf{L}$, to form the new operator $\widetilde{\mathbf{L}} = \mathbf{LD}^{-1}$, with adjoint $\widetilde{\mathbf{L}}^T = (\mathbf{D}^{-1})^T\mathbf{L}^T$.

The good pass operator used in this study is a five-point triangle filter, $\mathbf{P}$, which is applied across offsets within each migrated gather. Unlike a rectangular filter, a triangle filter gives higher weight to the central point on the filter compared to the edges. When smoothing a signal of length 8 the forward and adjoint operators can be

written in matrix form as

$$\mathbf{P} = \frac{1}{9} \cdot \begin{bmatrix} 3 & 4 & 2 & 0 & 0 & 0 & 0 & 0 \\ 2 & 3 & 2 & 2 & 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 3 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 & 2 & 3 & 2 \\ 0 & 0 & 0 & 0 & 0 & 2 & 4 & 3 \end{bmatrix}, \text{ and } \mathbf{P}^T = \frac{1}{9} \cdot \begin{bmatrix} 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 4 & 3 & 2 & 1 & 0 & 0 & 0 & 0 \\ 2 & 2 & 3 & 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 2 & 3 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 3 & 2 & 2 & 0 \\ 0 & 0 & 0 & 1 & 2 & 3 & 2 & 2 \\ 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 \end{bmatrix}.$$

Note the action of the forward operator on the edges of the gather (low and high offsets). At the edge of the gather the offsets on the missing side are replicated before smoothing to minimize edge effects. This operator has very few non-zero entries making it computationally inexpensive to implement in operator form. To ensure that the forward adjoint pair are written correctly, the dot product test is performed. Two vectors of random numbers are created, one with the dimension of the data, $\mathbf{d}_1$, and one with the dimension of the model, $\mathbf{m}_1$. $\mathbf{d}_1$ is passed through the concatenation of adjoint operators, $\mathbf{P}^T\mathbf{L}^T$, to generate $\mathbf{m}_2$, while $\mathbf{m}_1$ is passed through the concatenation of forward operators, $\mathbf{LP}$, to generate $\mathbf{d}_2$. The inner products $< \mathbf{d}_1, \mathbf{d}_2 >$ and $< \mathbf{m}_1, \mathbf{m}_2 >$ are compared and found to agree within machine precision.

Examples

In this article we consider a synthetic data example generated using the "sigmoid" model (Claerbout, 1992) and the forward converted wave PSTM operator, $\mathbf{L}$. Data are generated with a 4ms sample rate, and a 10m Asymptotic Conversion Point (ACP) increment. The data contain offsets from 100-2500m with a 100m offset increment. Note that data are shown at a constant offset of 1000m.

In figure 2 the input data (a) are regularly decimated on the ACP and offset axes by zeroing three traces between every live trace (b). The data are then reconstructed using a Fourier reconstruction method (Minimum Weighted Norm Interpolation (MWNI)) (c). In this case Fourier reconstruction struggles as the aliasing created by the zeroing pattern has stronger energy than desired signal. The diffractions and conflicting dips appear to be better reconstructed by synthesizing the data after least squares migration (d).

Figure 3 shows the reflectivities corresponding to the data shown in figure 2. Applying migration directly to the decimated data (b), or to the interpolated data (c) does not properly resolve the structure, while least squares migration (figure 3(d)) does a better job of recovering the true reflectivity. The inversion converges after approximately 14 iterations of Conjugate Gradients.

Conclusions

In this article we apply least squares migration to converted wave seismic data. We use a converted wave pre-stack time demigration operator and its adjoint in a Conjugate Gradients algorithm to solve for the reflectivity that best fits the data, while minimizing a norm on the model (migrated gathers). We precondition the problem by including a five-point triangle smoothing operator which acts on the offset dimension of the migrated gathers. We consider a synthetic data example in which traces are regularly decimated in data space. We find an advantage both in the reconstruction of data and the migration of the data when using the least squares migration algorithm compared with Fourier reconstruction plus migration using the adjoint operator. Following the advice of Gray et al. (2001), namely, that a more accurate forward modelling operator should be used compared to the migration operator being tested, we plan to test this algorithm on data generated using elastic finite differences to further evaluate its merit.

Acknowledgements

References

Audebert, F., 2001, 3-D prestack depth migration: why Kirchhoff?: Technical Report 80, Stanford Exploration Project.
Cary, P. W., and A. Zhang, 2011a, Aspects of amplitude preserving converted-wave prestack time migration:
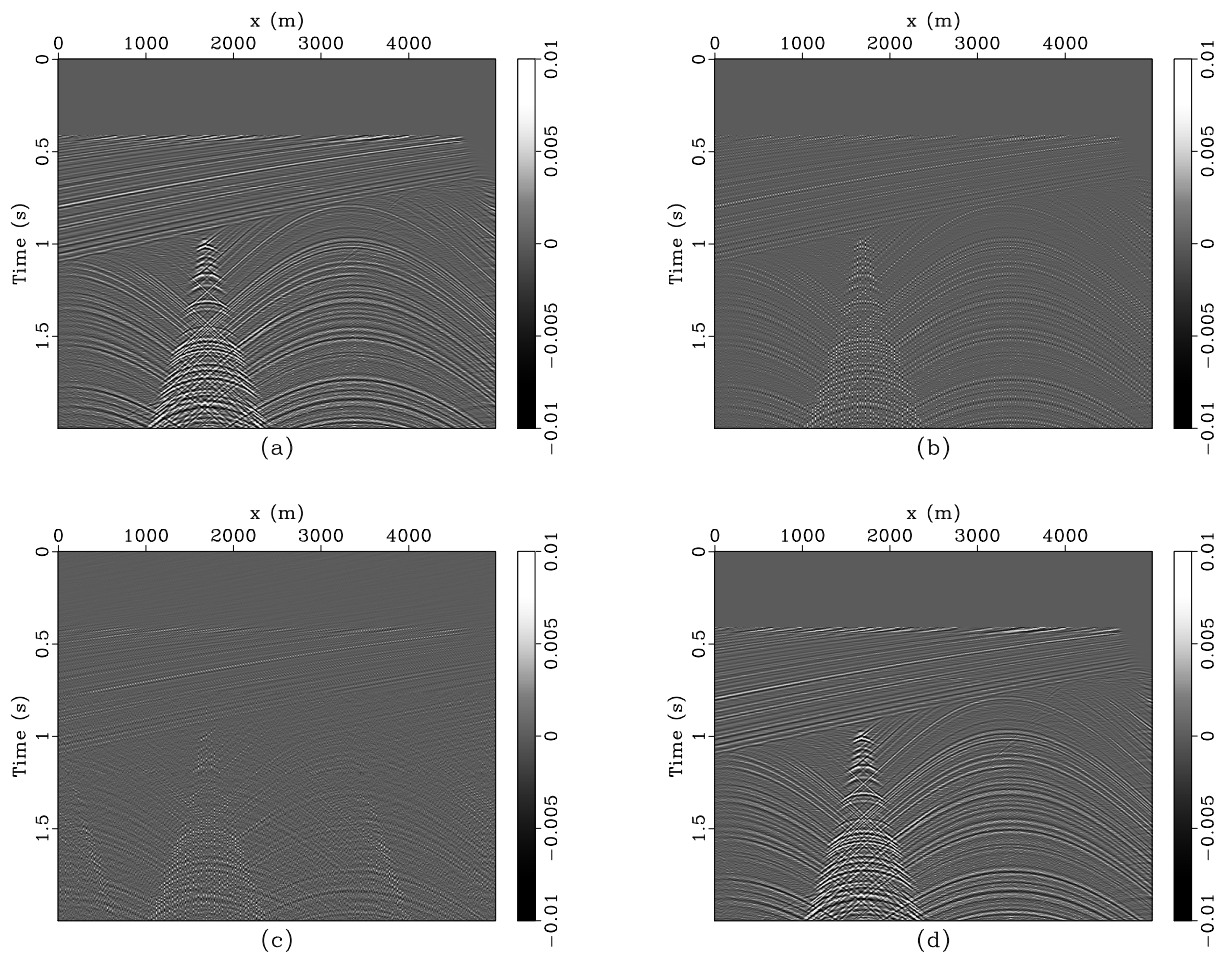
Figure 2 *Reconstruction of regularly decimated data. True data (a), regularly decimated data (b), reconstructed data using MWNI (c), and data reconstructed using least squares migration (d).*

Presented at the 73rd EAGE conference and exhibition.

———, 2011b, True-amplitude ps prestack time migration via 5D interpolation: Presented at the CSEG conference and exhibition.

Claerbout, J. F., 1992, Earth soundings analysis: Processing versus inversion: Blackwell Scientific Publications, **6**.

Gray, S. H., J. Etgen, J. Dellinger, and D. Whitmore, 2001, Seismic migration problems and solutions: Geophysics, **66**, 1640.

Guitton, A., B. Kaelin, and B. Biondi, 2007, Least-squares attenuation of reverse-time-migration artifacts: Geophysics, **72**, S19–S23.

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

Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221.

Rebollo, R., and M. D. Sacchi, 2010, Time domain least-squares prestack migration: Presented at the CSEG conference and exhibition.

Scales, J. A., 1987, Tomographic inversion via the conjugate gradient method: Geophysics, **52**, 179–185.

Wang, J., 2005, 3-d least-squares wave-equation avp/ava migration of common- azimuth data: PhD thesis, University of Alberta.

Wang, S., H. Hollis, and W. Nowry, 2013, Converted-wave time imaging- nuts and bolts: Presented at the CSEG conference and exhibition.
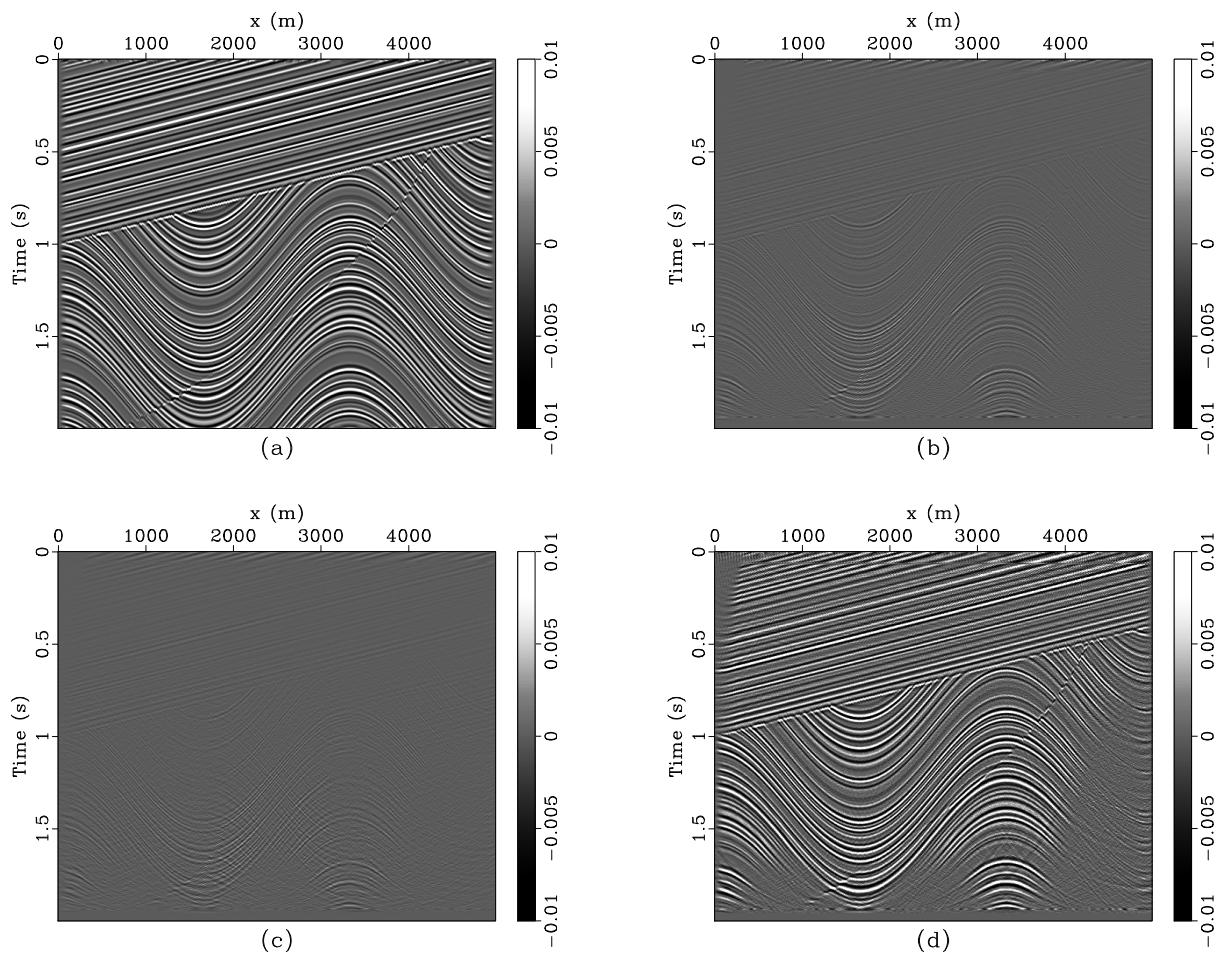
Figure 3 *Migration of regularly decimated data. True reflectivity model (a), migration of regularly decimated data (b), migration of data that has been reconstructed using MWNI (c), and least squares migration (d).*