

High-resolution wave-equation amplitude-variation-with-ray-parameter (AVP) imaging with sparseness constraints

Juefu Wang¹ and Mauricio D. Sacchi²

ABSTRACT

We propose a new scheme for high-resolution amplitude-variation-with-ray-parameter (AVP) imaging that uses nonquadratic regularization. We pose migration as an inverse problem and propose a cost function that uses a priori information about common-image gathers (CIGs). In particular, we introduce two regularization constraints: smoothness along the offset-ray-parameter axis and sparseness in depth. The two-step regularization yields high-resolution CIGs with robust estimates of AVP. We use an iterative reweighted least-squares conjugate gradient algorithm to minimize the cost function of the problem. We test the algorithm with synthetic data (a wedge model and the Marmousi data set) and a real data set (Erskine area, Alberta). Tests show our method helps to enhance the vertical resolution of CIGs and improves amplitude accuracy along the ray-parameter direction.

INTRODUCTION

It has been shown (Nemeth et al., 1999; Duquet et al., 2000; Kuehl and Sacchi, 2002, 2003) that seismic resolution can be improved by inverting the demigration/migration kernel and by enforcing a regularization constraint, for example, by introducing smoothness in the solution. However, as the results of these methods show, there are many artifacts present in the solution because of operator mismatch, wavefield sampling, and noise.

One possible way to further enhance the resolution and attenuate artifacts is by taking advantage of the solution itself. Iteratively using the result as a model-space regularization can lead to high-resolution, artifact-free seismic images. This idea has been used in many fields of signal and image processing (Sacchi and Ulrych, 1995; Charbonnier et al., 1997; Youzwishen, 2001; Sacchi et al., 2003; Trad et al., 2003; Downtown and Lines, 2004). In this paper, we utilize a model-dependent sparse regularization and a model-indepen-

dent smoothing regularization to estimate amplitude-variation-with-ray-parameter (AVP) common-image gathers (CIGs). Model-dependent sparse regularization is introduced via a nonquadratic norm (Cauchy norm). Smoothing, on the other hand, is implemented via a convolutional operator applied to AVP CIGs along the ray-parameter direction. This idea is used to develop an algorithm to simultaneously improve the structural interpretability and amplitude accuracy of seismic images.

It is important to point out the similarities between our algorithm and methods for impedance inversion based on sparse spike deconvolution of poststack cubes (Oldenburg et al., 1983; Debeye and van Riel, 1990). In principle, we are using very similar concepts to find a solution that exhibits predefined properties such as sparseness and smoothness. The main difference of our method from sparse spike inversion strategies is that our operator is a one-way wave-equation forward-modeling operator rather than a convolutional kernel. In addition, our inversion results are in depth, and the input data are prestack volumes as opposed to time-domain reflectivity estimates and poststack volumes, respectively. We believe our method provides a unifying thread between convolution-based sparse spike inversion and regularized migration/inversion methods.

METHODOLOGY

One advantage of imaging via regularized inversion is that we can use a priori information about the unknown image model (Prucha and Biondi, 2002). Robust inversion algorithms can be developed by properly honoring such information. For example, Kuehl and Sacchi (2002, 2003) show that applying smoothing regularization on the ray-parameter axis helps to remove artifacts introduced by missing information, aliasing, noise, and operator mismatch. The scheme is based on minimizing a quadratic cost function. In addition, Sacchi et al. (2003) show that higher resolution can be acquired by solving a nonquadratic problem.

In this paper, we reformulate the cost function for the least-squares wave-equation AVP/amplitude-variation-with-angle-of-incidence (AVA) migration problem as follows:

Manuscript received by the Editor March 20, 2006; revised manuscript received September 4, 2006; published online November 16, 2006.

¹Formerly University of Alberta; presently Divestco Inc. and Alberta Ingenuity, Calgary, Alberta, Canada. E-mail: juefu.wang@divestco.com.

²University of Alberta, Institute for Geophysical Research, Department of Physics, Edmonton, Alberta, Canada. E-mail: sacchi@phys.ualberta.ca.

© 2007 Society of Exploration Geophysicists. All rights reserved.

$$J(\mathbf{m}) = \|\mathbf{W}(\mathbf{L}\mathbf{m} - \mathbf{d})\|_2^2 + \lambda^2 F(\mathbf{S}\mathbf{H}(\mathbf{m})), \quad (1)$$

where \mathbf{m} is the earth model in terms of AVP CIGs, \mathbf{L} is a 2D/3D wave-equation modeling operator that transforms the model to prestack seismic data, \mathbf{d} is the seismic data, and \mathbf{W} is a sampling matrix used to accommodate missing data in the inversion. The modeling operator is synthesized via the double square root upward-continuation operator with split-step corrections in conjunction with a radial transform that converts ray-parameter-dependent reflectivity to local wavefields (Kuehl and Sacchi, 2003). The operator \mathbf{H} is a model-independent high-pass filter that we use to penalize non-smooth solutions along ray-parameter direction, \mathbf{S} is a stacking operator that converts CIGs to a stacked image, F is a model-dependent functional used to enforce sparseness, and λ is a trade-off parameter that controls the amount of regularization. By using the Cauchy norm (Sacchi and Ulrych, 1995), the sparse regularization operator F is given by

$$F(\mathbf{m}) = \sum_{i=1}^n \ln\left(1 + \frac{m_i^2}{\sigma^2}\right), \quad (2)$$

where σ^2 is a scale parameter of the Cauchy distribution, n is the number of elements of the stacked image, and m_i is the amplitude of the i th element of the stacked image. By adopting a preconditioning strategy (Prucha and Biondi, 2002; Fomel and Claerbout, 2003; Trad et al., 2003; Wang et al., 2004), the cost function can be expressed as

$$J = \|\mathbf{W}(\mathbf{L}\mathbf{P}\mathbf{z} - \mathbf{d})\|_2^2 + \lambda^2 F(\mathbf{S}\mathbf{z}), \quad (3)$$

where \mathbf{P} is a preconditioning matrix and \mathbf{z} is the model modified by the preconditioner. It is clear that $\mathbf{m} = \mathbf{P}\mathbf{z}$. In our implementation, we use a low-pass filter as the preconditioner operator \mathbf{P} . In particular, applying \mathbf{P} entails convolution along the ray-parameter axis with a Hamming window (Wang et al., 2004). The adjoint operator \mathbf{P}' is a simple crosscorrelation procedure (Claerbout, 1992). To minimize the cost function J , we solve the problem $dJ/d\mathbf{m} = \mathbf{0}$. The latter leads to the following nonlinear system of equations (proven in Appendix A):

$$(\mathbf{P}'\mathbf{L}'\mathbf{W}'\mathbf{W}\mathbf{L}\mathbf{P} + \mu^2\mathbf{S}'\mathbf{Q}\mathbf{S})\mathbf{z} = \mathbf{P}'\mathbf{L}'\mathbf{W}'\mathbf{d}, \quad (4)$$

where $\mu = \lambda/\sigma$ and \mathbf{Q} is a diagonal matrix defined by a vector $\mathbf{s} = \mathbf{S}\mathbf{z}$. The i th diagonal element of \mathbf{Q} can be expressed as

$$Q_{ii} = \frac{1}{1 + \left(\frac{s_i}{\sigma}\right)^2}, \quad (5)$$

where s_i is the i th element of \mathbf{s}

The nonlinear system given by expression 4 has been solved in the context of high-resolution Radon transforms (Sacchi and Ulrych, 1995) using the iterative reweighted least-squares (IRLS) method (Scales and Smith, 1994). The IRLS algorithm is generalized as below:

1. Choose the hyperparameters μ and σ , and initialize the algorithm with $k = 0$, $\mathbf{z}^k = \mathbf{0}$.
2. For $k = 1, 2, \dots$
 - Compute $\mathbf{Q}^k = \mathbf{Q}(\mathbf{S}\mathbf{z}^{k-1})$
 - Solve $(\mathbf{P}'\mathbf{L}'\mathbf{W}'\mathbf{W}\mathbf{L}\mathbf{P} + \mu^2\mathbf{S}'\mathbf{Q}^k\mathbf{S})\mathbf{z}^k = \mathbf{P}'\mathbf{L}'\mathbf{W}'\mathbf{d}$
 - $\mathbf{m}^k = \mathbf{P}\mathbf{z}^k$
 - Compute the predicted data $\mathbf{d}^k = \mathbf{L}\mathbf{m}^k$ and residuals $\mathbf{d} - \mathbf{d}^k$ to monitor convergence
 - End

This procedure requires about three or four updates (iterations) to obtain a solution that is sparse in depth and smooth with respect to the ray parameter. In addition, as discussed by Kuehl and Sacchi (2003), the constrained linear system of equations at the second step within the loop can be solved efficiently with the method of conjugate gradients (CG). In this method, we do not need to configure operators in explicit matrix form.

Finally, it is important to define a strategy to provide the scale parameter σ . We have obtained encouraging results by setting σ to some percentage of the maximum amplitude of the vector $\mathbf{S}\mathbf{z}^{k-1}$. In other words, we introduce a new parameter δ such that $\sigma = \delta \times \max(|\mathbf{S}\mathbf{z}^{k-1}|)$. The problem is then reduced to finding two parameters, δ and μ . Based on our experience, pairs of δ and μ with a constant product lead to similar solutions. In addition, large values of δ yield low-resolution results (Wang, 2005). Therefore, we usually set δ to an arbitrary small number (in our tests, we set it to 0.02) and adjust the trade-off parameter μ to obtain a satisfactory fitting. We first try $\mu = 0.1$. If the solution is not sparse enough, we increase the parameter ten times ($\mu = 1.0$). If the solution is too sparse, we decrease the parameter ten times ($\mu = 0.01$). We repeat this trial-and-error routine until we find a good trade-off parameter.

One may wonder why we can introduce sparseness in the solution without destroying the AVA features of the CIGs. The reason is that we apply the regularization in two steps. First, we smooth the CIGs in the prestack ray-parameter direction. Then we apply diagonal weighting operator \mathbf{Q} to the stacked image. In other words, the sparseness constraint is applied to the stacked image and not directly to the prestack CIGs. In this way, image points at a given depth receive the same weight, thereby preserving the amplitude ratio of different ray parameters.

Note that when we turn off the sparse regularization by setting λ to zero in equation 3, the problem is reduced to the preconditioned least-squares migration (PLSM) (Wang et al., 2004). To glean the benefits and shortcomings of various imaging methods, we compare their image quality and accuracy in the following synthetic and real data examples.

EXAMPLES

Multichannel deconvolution

The combination of sparseness and smoothness constraints is an interesting idea. Before applying it to least-squares (LS) migration, we would rather validate its efficiency using a much cheaper operator — a multichannel deconvolution. This is an unrealistic scenario in seismic deconvolution, yet it is a fast way to test the algorithm. The procedure, however, may be used to deconvolve time-migrated CIGs. Figure 1a is a time-domain model with 20 offsets. We convolve the model with a zero-phase wavelet and remove three offsets (trace number equals 2, 6, and 9) to test the procedure in situations of missing information. The data are portrayed in Figure 1b.

We compared two methods of inversion, preconditioned LS inversion (Wang et al., 2004) and sparse LS inversion. Figure 1c is the result of the preconditioned LS inversion after 50 iterations of the CG algorithm. It is evident that the inversion successfully fills the gaps in the incomplete data. However, the vertical resolution is unsatisfactory. Spurious side lobes are present in the inverted reflectivity model. On the other hand, the sparse inversion provides a superior result (Figure 1d). It is almost identical to the real reflectivity model (Figure 1a); the wavelet is properly compressed. In addition, the amplitude variation with offset (AVO) signature is preserved.

Figure 1e compares the spectra of the results obtained via the LS and sparse LS deconvolutions. The results are quite similar at low and middle frequencies (0–80 Hz). The sparse inversion, however, provides a much broader amplitude response with an important amplitude gain for frequencies above 80 Hz.

Figure 1f shows a potential shortcoming of sparse inversion methods. When the trade-off parameter μ is too large, we may lose valuable information. In other words, if the solution becomes too sparse, small reflections can be annihilated. This problem is also encountered in the application of sparse deconvolution (Oldenburg et al., 1983) and high-resolution Radon transforms (Sacchi and Ulrych, 1995).

Wedge model

A 2D synthetic data set was used to test the algorithm. We prepared the data by applying the forward operator L to a constant-velocity wedge model, represented by a set of AVP CIGs. Ideally, the inversion should be able to reconstruct the CIGs, and the stacked image of these CIGs should clearly portray the modeled structure.

We processed the data by three methods: conventional migration (the adjoint of the modeling operator), PLSM (Wang et al., 2004), and sparse least-squares migration (SLSM) proposed in this paper. Figure 2 shows the stacked images. The result of the adjoint is quite blurry because the algorithm is incapable of reconstructing high frequencies not present in the data. On the other hand, both PLSM and SLSM are able to recover high frequencies in the structural images. In addition, the SLSM algorithm produces a highly resolved CIG.

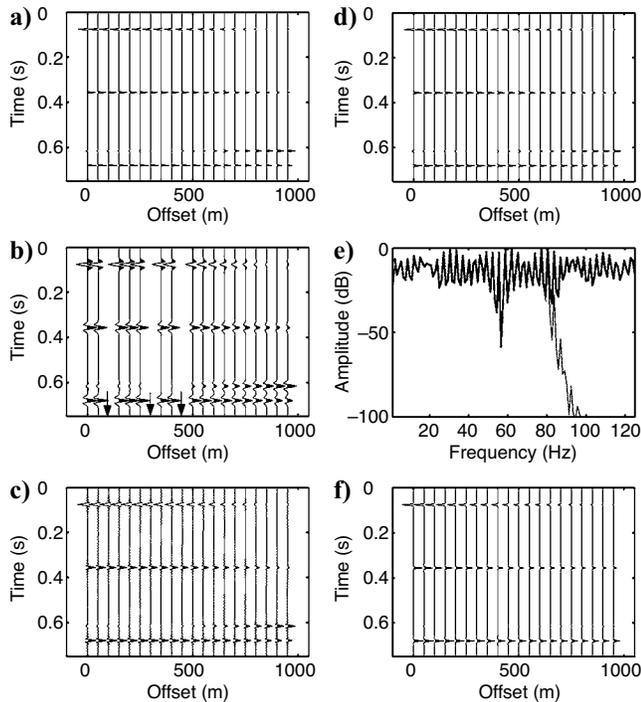


Figure 1. A multichannel deconvolution example to compare the preconditioned least-squares (LS) inversion and the sparse LS inversion. (a) Reflectivity model. (b) Incomplete multichannel data. Arrows mark the locations of the muted traces. (c) Result of the preconditioned LS inversion. (d) Result of the sparse LS inversion ($\mu = 10.0$). (e) Average spectra of (c) (thin curve) and (d) (bold curve). (f) Result of the sparse LS inversion ($\mu = 1000.0$).

This is a consequence of using a sparseness constraint that attempts to collapse the band-limited seismic wavelet into a broadband impulsive signal.

Figure 3a–c displays a zoomed view of three CIGs produced by these methods. The SLSM method has the ability of suppressing the side lobes introduced by the band-limited wavelet. To complete our analysis, we extracted the amplitude of the tilted event and plotted AVA curves for the three methods in Figure 3d. Both PLSM and SLSM are able to preserve the amplitude response of the reflection.

Figure 4 compares the data misfit of the two inversion methods. The PLSM method starts to converge at the seventh iteration. As shown in Figure 3d, the method provides accurate amplitude for an angle range between 0° and 40° . However, the structural image is unsatisfactory because of the tuning effects. The SLSM method converges to a sparse solution in four iterations of the IRLS algorithm. In this test, the estimated amplitude response completely fits the theoretical AVA curve. In addition, the SLSM method has completely

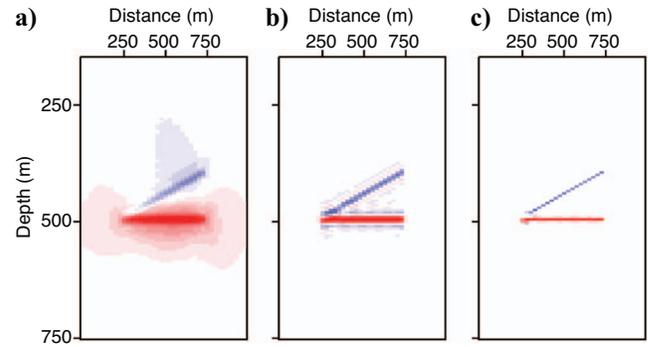


Figure 2. Stacked images of the wedge model, obtained with (a) migration, (b) PLSM, and (c) SLSM.

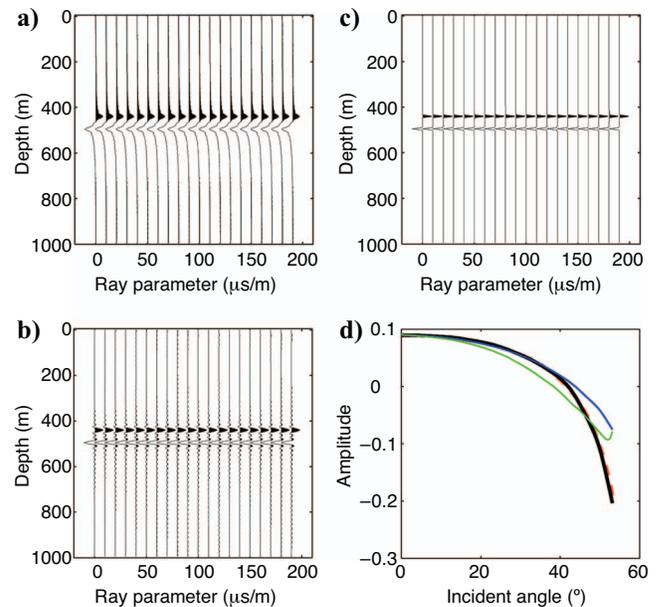


Figure 3. CIGs and AVA curves at $x = 500$ m for the wedge model. (a) Migration. (b) PLSM. (c) SLSM. (d) AVA curves for the first event: red dashed curve — the theoretical curve, green curve — migration, blue curve — PLSM, black curve — SLSM.

compressed the seismic wavelet. Figure 5 compares the original data, the reconstructed data by PLSM, and the residual panel. The data fitting for the SLSM method is portrayed in Figure 6. Both methods fit the input data quite well.

Note that least-squares migration is much more expensive than conventional migration. Each CG iteration involves one migration and one forward modeling (demigration). Usually the PLSM algorithm starts to converge in less than 10 CG iterations. The cost of the SLSM algorithm is usually three or four times that of the PLSM algorithm.

The Marmousi data set

We also applied the algorithm to the Marmousi data set. We randomly removed 70% of the traces to simulate a sparse data acquisition. Figure 7 compares the stacked images obtained by migration, PLSM, and SLSM. Both PLSM and SLSM provide images with higher resolution than conventional migration. SLSM cleans up the image further than PLSM because of the sparse regularization.

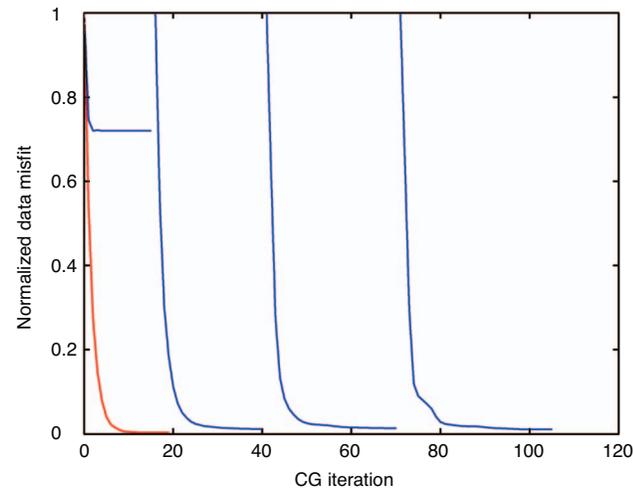


Figure 4. Data misfit of two inversion methods. Red curve — PLSM. Blue curves — SLSM. The four blue curves represent four iterations of the IRLS algorithm. From left to right, the first blue curve represents the first iteration of the IRLS algorithm, the second represents the second iteration, and so on. At the beginning of each IRLS iteration, the model is initialized with zeros. Therefore, within each iteration, the data misfit always decreases from the same maximum value until convergence.

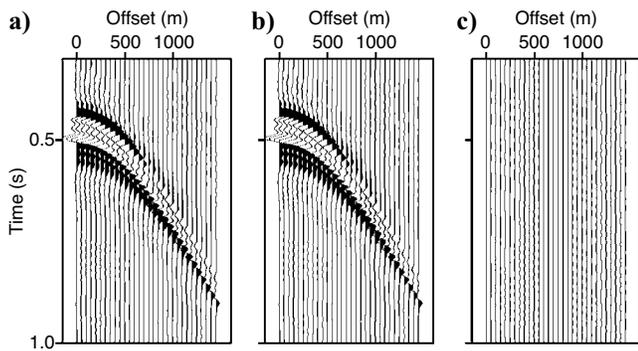


Figure 5. (a) The original data, (b) data reconstructed by PLSM, and (c) the residual at $x = 500$ m.

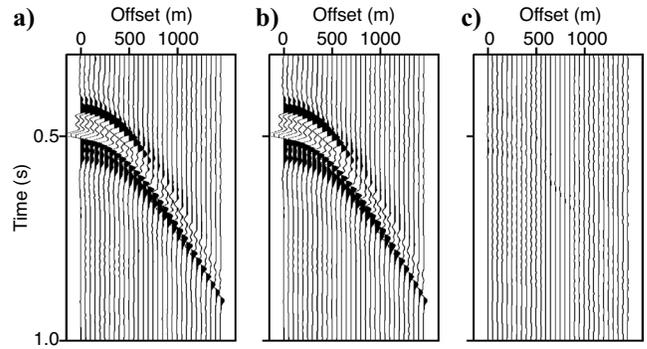


Figure 6. (a) The original data, (b) data reconstructed by SLSM, and (c) the residual at $x = 500$ m.

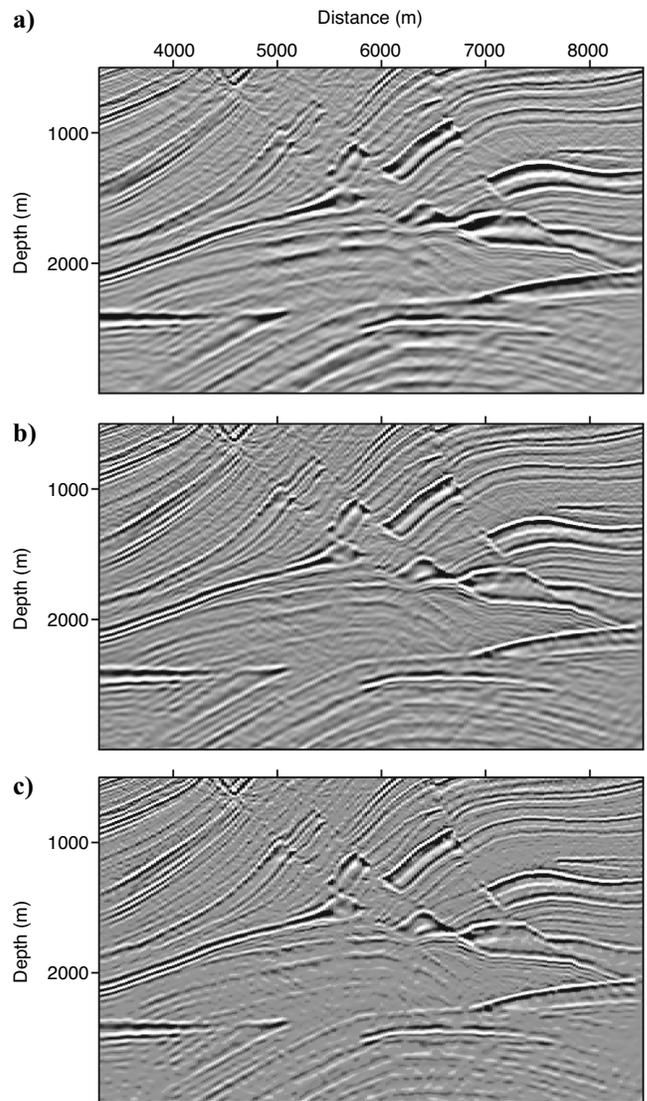


Figure 7. Stacked images of the Marmousi model: (a) migration, (b) PLSM (four iterations), (c) SLSM (three IRLS iterations). We use the same clipping percentage for these figures.

Figure 8 shows the CIGs calculated by these three methods. It is clear that many artifacts are present in the CIG obtained with the migration algorithm. These artifacts are substantially removed from the images obtained with PLSM and SLSM.

For comparison, we calculated the reflectivity series by using the true velocity and density model. A side-by-side comparison confirms that the SLSM has properly reconstructed the model. We observe again, as in our previous examples, an important attenuation of ringing arising from the band-limited wavelet in the data. To evaluate the amplitude-preserving properties of our algorithm, we obtained AVA curves for the event at depth $z = 800$ m. The amplitude response obtained with the migrated image is difficult to extract because of sampling artifacts. The inverted AVA responses (PLSM and SLSM), on the other hand, are in good agreement with the theoretical value.

Erskine data set (WCSB)

As an experiment, we also processed a field data set with the SLSM algorithm. The data set corresponds to a 3D orthogonal survey acquired in the Erskine area (Alberta, Canada); the seismic data are typical of exploratory plays in the Western Canadian Sedimentary Basin (WCSB). To save computing time and avoid overfitting the noisy data, we only ran the IRLS program for three iterations.

A comparison of the stacked images of an inline of the data (see Figure 9) shows that both PLSM and SLSM provide higher resolution than conventional migration. For example, PLSM and SLSM help to separate the overlapping events above 1500-m depth. As a benefit, SLSM provides higher resolution than PLSM. One example is the event below 2000-m depth (marked with hollow arrows). Clearly, the wavelet is suppressed better by SLSM.

Figure 10 compares the CIGs at midpoint position 1155 m. Both PLSM and SLSM greatly improve the coherence of the CIG. Again, we can see that SLSM resolves events better by suppressing the wavelet.

To complete our analysis, we compare the spectra of the stacked images in Figure 11. We observe that both PLSM and SLSM enhance the high-frequency components of the solution. As expected, SLSM increases the amplitude response at large vertical wavenumbers (or high frequencies in the temporal-frequency domain) better than the PLSM method. We need to clarify that if a full-band forward-modeling operator is applied to the inverted sparse solution, the modeled data will not fit the original band-limited seismic data. However, in our inversion, both the forward operator and its adjoint (migration) are applied within a limited frequency band. The latter allows us to honor the observed data in a limited seismic bandwidth. The effects of wavelet mismatch within the limited frequency band are examined by Wang (2005). His tests on synthetic data show that when the wavelet is not precisely known, the SLSM algorithm can recover a highly resolved image of the subsurface. However, the solution is not as good as that inverted with a precisely known wavelet.

There is one caveat in the application of SLSM. Overregularization can lead to too sparse solutions and therefore loss of our ability to image weak reflections. Processors and interpreters

should compare the result obtained with SLSM to that with conventional migration to avoid this problem. In the Marmousi and Erskine examples, we tried different parameters μ and chose a conservative one to avoid overregularization.

INFLUENCE OF VELOCITY ERRORS

Providing a good velocity model is fundamental to achieving optimal migration results. In reality, one can never expect to have access to the true velocity. The practical idea is to estimate a velocity model close to the true one. Wave-equation migration techniques still perform well when the input velocity is smoothed (Duquet et al., 2002). To test whether this is also true for PLSM and SLSM, we use the Marmousi data set (after removing 70% of the traces) to show that our algorithm can tolerate reasonable velocity errors.

In the first test, we smoothed the velocity model using a 10-point moving average. Then we ran migration, PLSM, and SLSM; the results are shown in Figure 12. Again, PLSM and SLSM improve the quality of the CIGs. Smoothing the velocity field is acceptable not only in standard migration techniques but also in least-squares migration methods.

In the second test, we perturbed the velocity field by 5%. Figure 13 portrays the results of migration, PLSM, and SLSM. As in the previous example, the data were severely decimated (70% of the trace were removed). The migrated CIG is dominated by artifacts. On the other hand, these artifacts have been alleviated by PLSM and SLSM. We notice, in addition, that residual moveout has not been destroyed by PLSM and SLSM. However, we do not recommend relying on PLSM or SLSM to handle velocity errors. Instead, one should minimize velocity errors before migration. Our examples are provided for the purpose of examining the robustness of PLSM and SLSM when dealing with smooth and/or inaccurate velocity fields.

DISCUSSION

Our algorithm obtains high-resolution AVA gathers. The algorithm removes spurious artifacts by constraining the solution to be smooth in the ray-parameter direction and sparse in depth.

Our tests show that overregularization leads to loss of valuable information that is often contained in events with small amplitudes.

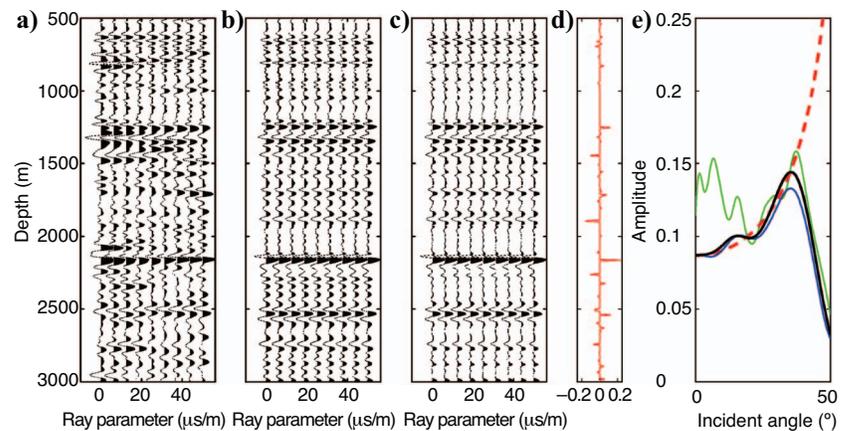


Figure 8. CIGs and AVA curves at $x = 7500$ m for the Marmousi data: (a) migration, (b) PLSM, (c) SLSM, (d) zero-offset reflectivity from the density and velocity models, (e) AVA curves for the event at 800-m depth (red dashed curve — theoretical curve, green curve — migration, blue curve — PLSM, black curve — SLSM).

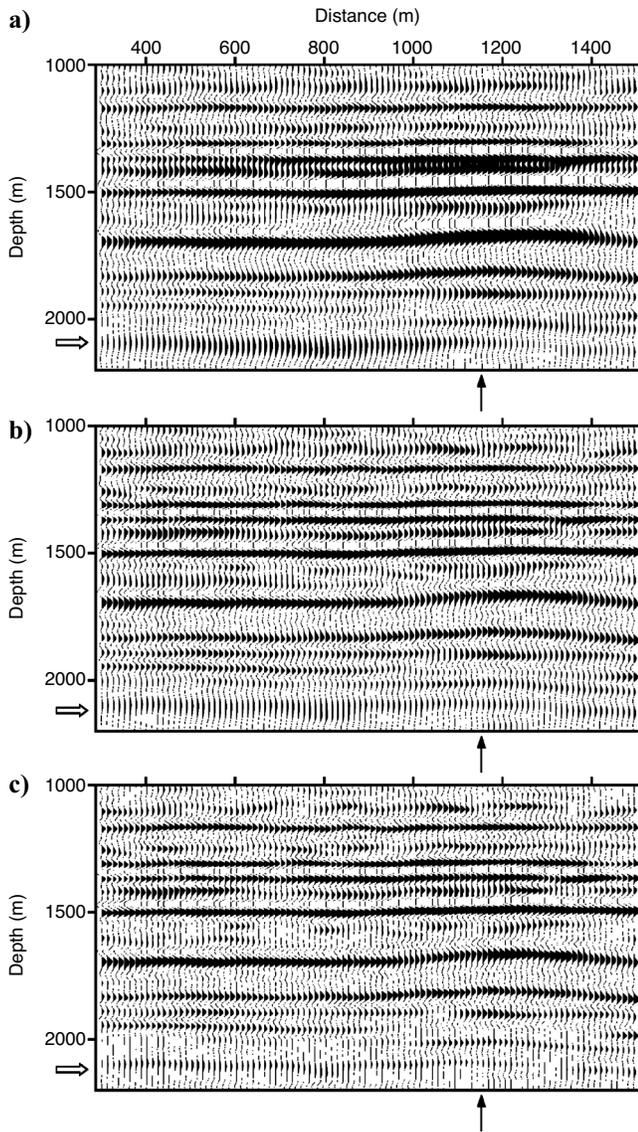


Figure 9. Stacked images of the Erskine data set, obtained with (a) migration, (b) PLSM, and (c) SLSM. Solid arrows mark where CIGs are extracted and displayed in Figure 10. Open arrows point to the event for vertical resolution comparison.

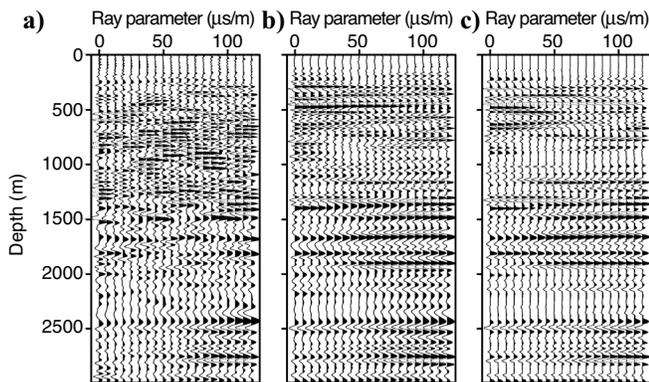


Figure 10. CIGs calculated by (a) migration, (b) PLSM, and (c) SLSM.

This problem is also encountered in techniques for poststack inversion of seismic data that are based on the sparse reflectivity assumption.

Imaging/inversion with the introduction of quadratic and nonquadratic constraints could lead to a new class of imaging algorithms where the resolution of the inverted image can be enhanced beyond the limits imposed by the data (bandwidth and aperture). This is not a completely new idea. Geophysicists have been using similar concepts to invert poststack data (sparse spike inversion) to construct highly resolved impedance profiles; in addition, similar concepts are used in high-resolution Radon transforms for multiple elimination and data reconstruction. What constitutes an optimal regularization strategy for imaging problems is a topic of current study. Presently,

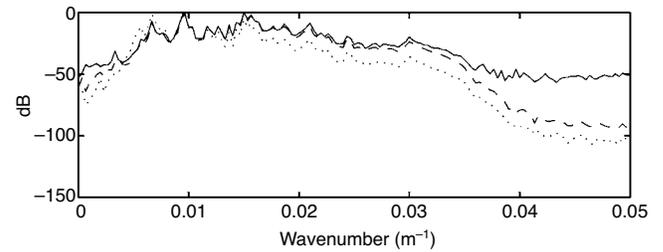


Figure 11. Average spectra of the stacked images of the Erskine data set, obtained with migration, PLSM, and SLSM. Dotted curve — migration, dashed curve — PLSM, and solid curve — SLSM. The migration depth step is 10 m; therefore, the Nyquist wavenumber is 0.05 m^{-1} .

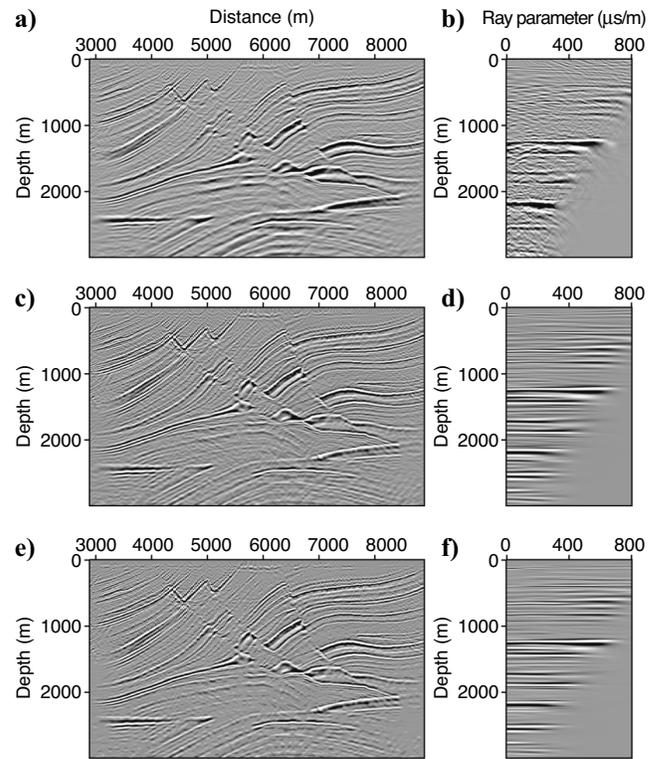


Figure 12. Results of migration, PLSM, and SLSM using a smoothed velocity model: (a) Stacked image (migration), (b) CIG (migration) at 7500 m, (c) stacked image (PLSM), (d) CIG (PLSM) at 7500 m, (e) stacked image (SLSM), and (f) CIG (SLSM) at 7500 m.

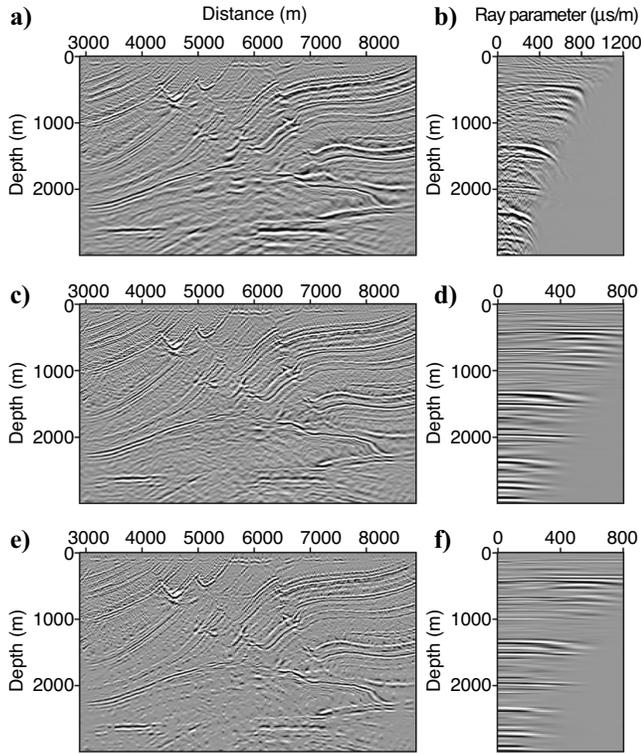


Figure 13. Results of migration, PLSM, and SLSM using a velocity model with 5% error: (a) Stacked image (migration), (b) CIG (migration) at 7500 m, (c) stacked image (PLSM), (d) CIG (PLSM) at 7500 m, (e) stacked image (SLSM), and (f) CIG (SLSM) at 7500 m.

smoothing the CIGs (in the offset- or ray-parameter axis) plus vertical sparseness appears to be a regularization goal that is simple and consistent with the estimation of high-resolution CIGs for subsequent studies such as estimating AVO signatures.

In the current implementation of our SLSM, we are not considering the source wavelet. Therefore, operator mismatch exists in the data domain. Some of our tests (Wang, 2005) show that including the source wavelet in the modeling routine can help to acquire a sparser solution. However, as we know, estimation of the source wavelet is a difficult problem. For prestack data, the wavelet can be offset dependent, which makes the problem more complicated (B. Biondi, 2005, personal communication). An alternative approach is to use half-migration methods as proposed by Zhang and Wapenaar (2006). In this method, the image is converted to the time domain, and the wavelet can be handled in an easier way.

Future research directions call for methods to mitigate operator mismatch, efficient numerical optimization methods for large-scale problems, regularization methods capable of incorporating the reflectivity character (color) into our subsurface estimates, and extensive field data tests to properly access the benefits of regularized migration.

ACKNOWLEDGMENTS

We acknowledge the financial support of the sponsors of the Signal Analysis and Imaging Group at the University of Alberta. We are grateful to Biondo Biondi for discussions on common azimuth migration, which is used to implement our 3D inversion. We also thank

three anonymous reviewers and associate editor Yu Zhang for their valuable suggestions.

APPENDIX A

COST FUNCTION FOR LEAST-SQUARES MIGRATION WITH SMOOTHNESS AND SPARSENESS CONSTRAINTS

The transition from equation 3 to equation 4 is provided in this appendix. We first note that the cost function for sparse least-squares migration is given by the expression

$$J = J_M + \lambda^2 J_R, \quad (\text{A-1})$$

where the misfit and regularization terms are given by

$$J_M = \|\mathbf{W}(\mathbf{L}\mathbf{P}\mathbf{z} - \mathbf{d})\|_2^2 \quad (\text{A-2})$$

and

$$J_R = F(\mathbf{S}\mathbf{z}) = F(\mathbf{s}) = \sum_i \ln\left(1 + \frac{s_i^2}{\sigma^2}\right). \quad (\text{A-3})$$

The minimum of J satisfies

$$\frac{dJ}{d\mathbf{z}} = \frac{dJ_M}{d\mathbf{z}} + \lambda^2 \frac{dJ_R}{d\mathbf{z}} = 0. \quad (\text{A-4})$$

The first term in the gradient can be shown to be

$$\frac{dJ_M}{d\mathbf{z}} = 2(\mathbf{P}'\mathbf{L}'\mathbf{W}'\mathbf{W}\mathbf{L}\mathbf{P})\mathbf{z} - 2\mathbf{P}'\mathbf{L}'\mathbf{W}'\mathbf{d}. \quad (\text{A-5})$$

The second term requires a bit more work. We first express $\mathbf{s} = \mathbf{S}\mathbf{z}$ in element form, $s_i = \sum_j S_{ij}z_j$, and notice that

$$\frac{dJ_R}{dz_k} = \sum_i \frac{2s_i}{\sigma^2\left(1 + \frac{s_i^2}{\sigma^2}\right)} \frac{ds_i}{dz_k} = \frac{2}{\sigma^2} \sum_l \left(\sum_i S_{ik}Q_{il}S_{il} \right) z_l. \quad (\text{A-6})$$

The latter is the k th element of the gradient of the regularization term, which can be written in vector form as follows:

$$\frac{dJ_R}{d\mathbf{z}} = \frac{2}{\sigma^2} \mathbf{S}'\mathbf{Q}\mathbf{S}\mathbf{z}, \quad (\text{A-7})$$

where the elements of the diagonal operator \mathbf{Q} are given by equation 5. Combining equations A-4, A-5, and A-7 leads to

$$\frac{dJ}{d\mathbf{z}} = 2(\mathbf{P}'\mathbf{L}'\mathbf{W}'\mathbf{W}\mathbf{L}\mathbf{P})\mathbf{z} - 2\mathbf{P}'\mathbf{L}'\mathbf{W}'\mathbf{d} + \frac{2\lambda^2}{\sigma^2} \mathbf{S}'\mathbf{Q}\mathbf{S}\mathbf{z} = 0. \quad (\text{A-8})$$

After rearranging terms and defining $\mu^2 = \lambda^2/\sigma^2$, equation A-8 is equal to expression 4 in the main text.

REFERENCES

- Charbonnier, P., L. Blanc-Feraud, G. Aubert, and M. Barlaud, 1997, Deterministic edge-preserving regularization in computed imaging: *IEEE Transactions on Image Processing*, **6**, 298–311.
- Claerbout, J., 1992, *Earth soundings analysis: Processing versus inversion*: Blackwell Scientific Publications, Inc.
- Debye, H. W. J., and P. van Riel, 1990, LP-norm deconvolution: *Geophysical Prospecting*, **38**, 381–404.
- Downtown, J., and L. Lines, 2004, Three term AVO waveform inversion: Annual Convention, Canadian Society of Exploration Geophysicists, Abstracts on CDROM.
- Duquet, B., K. J. Marfurt, and J. A. Dellinger, 2000, Kirchhoff modeling, inversion for reflectivity, and subsurface illumination: *Geophysics*, **65**, 1195–1209.
- Duquet, B., S. Xu, and G. Lambare, 2002, 3D multi-arrival Kirchhoff versus wave equation migration: Application to the 3D SEG/EAGE salt model: 72nd Annual International Meeting, SEG, Expanded Abstracts, 1364–1367.
- Fomel, S., and J. F. Claerbout, 2003, Multidimensional recursive filter preconditioning in geophysical estimation problems: *Geophysics*, **68**, 577–588.
- Kuehl, H., and M. D. Sacchi, 2002, Robust AVP estimation using least-squares wave-equation migration: 72nd Annual International Meeting, SEG, Expanded Abstracts, 281–284.
- , 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.
- Oldenburg, D. W., T. Scheuer, and S. Levy, 1983, Recovery of the acoustic impedance from reflection seismograms: *Geophysics*, **48**, 1318–1337.
- Prucha, M., and B. Biondi, 2002, Subsalt event regularization with steering filters: 72nd Annual International Meeting, SEG, Expanded Abstracts, 1176–1179.
- Sacchi, M. D., C. M. Constantinescu, and J. Feng, 2003, Enhancing resolution via nonquadratic regularization — Next generation of imaging algorithms: Annual Convention, Canadian Society of Exploration Geophysicists, Abstracts on CDROM.
- Sacchi, M. D., and T. J. Ulrych, 1995, High resolution velocity gathers and offset space reconstruction: *Geophysics*, **60**, 1169–1177.
- Scales, J. A., and M. L. Smith, 1994, *Introductory geophysical inverse theory*: Samizdat Press.
- Trad, D., T. Ulrych, and M. D. Sacchi, 2003, Latest views of the sparse Radon transform: *Geophysics*, **68**, 386–399.
- Wang, J., 2005, 3-D least-squares wave-equation AVP/AVA migration of common azimuth data: Ph.D. dissertation, University of Alberta.
- Wang, J., H. Kuehl, and M. D. Sacchi, 2004, Preconditioned least-squares wave-equation AVP migration: Annual Convention, Canadian Society of Exploration Geophysicists, Abstracts on CDROM.
- Youzwishen, C., 2001, Non-linear sparse and blocky constrains for seismic inverse problems: M.S. thesis, University of Alberta.
- Zhang, J., and K. Wapenaar, 2006, High-resolution depth imaging with sparseness-constrained inversion: *Geophysical Prospecting*, **54**, 49–62.