Automatic Velocity Analysis by Differential Semblance Optimization

W.A. Mulder and A.P.E. ten Kroode, Shell EP Technology Applications & Research*

Summary

We present a method for automatic velocity analysis of seismic data based on Differential Semblance Optimization (DSO). The data are mapped per offset from the time domain to the depth domain by a Born migration scheme using ray tracing with the efficient wavefront construction method. The DSO cost functional is evaluated by taking differences of the migration images for neighboring offsets. The gradient of this functional with respect to the underlying velocity model is obtained by a first-order approximation of the adjoint-state method, leading to an optimal complexity: the cost of evaluating the gradient is about the same as that of evaluating the functional. A gradient optimization method minimizes the functional. The method has been applied to a marine line. Multiples turned out to be a problem, but were handled effectively by incorporating a multiple filter inside the DSO cost functional.

Introduction

The goal of seismic full waveform inversion is the automatic construction of a subsurface model that reproduces seismic data measured at the surface. In the least-squares approach, the difference between measured and modeled data is minimized in the quadratic or root-mean-square norm (Tarantola, 1986). This method, however, may fail if the initial model is too far from the correct model because the least-squares functional has many local minima. If a gradient method is used to solve the optimization problem, chances are that the iteration process will end up in a local minimum that may be far from the absolute minimum. To avoid this, global optimization methods such as simulated annealing or genetic algorithms can be used. However, the associated computational cost is huge.

An alternative is Differential Semblance Optimization (DSO) proposed by Symes et al. (1991). This method exploits the redundancy in the data by requiring that the earth be invariant for different seismic experiments. If the data are partitioned into "minimal" data sets, each of these data sets can provide a subsurface model independently of the other subsets. DSO attempts to minimize the differences between the various subsurface models. A partitioning and ordering based on shots is natural, but leads to illumination problems. These are less severe for a common depth point (CDP) sorting of the data. In that case, offset (source-receiver distance) is the redundant parameter and DSO attempts to flatten common image gathers (CIGs) – a well-known concept in exploration geophysics.

The computation of the DSO cost functional is described in the next section. To perform the optimization with respect to the velocity model, a gradient optimization method is used (Zhu et al., 1994). The computation of the gradient (sensitivity w.r.t. the model) is carried out by a first-order approximation to the adjoint-state equations (Ciarlet, 1989; Plessix et al., 2000).

A real-data example on a marine line is included. Multiples turn out to be a problem. A modification of the DSO functional is proposed to better handle multiples. With this modification, DSO shows a good convergence behavior, even when started far away from the final velocity model.

Method

Common image gathers are computed by ray-tracing and Born migration. We trace downward from sources and receivers by using the wavefront construction method (Vinje et al., 1993). The slowness model is represented by bicubic splines. If the spline nodes are spaced at a sufficiently large distance, no additional smoothness constraint for regularization is needed.

Let the result of the migration operator that maps data from the time to the depth domain be denoted by $f(h,x,z)$. Here h is the offset (source-receiver distance) and (x,z) denotes the CDP. For fixed x, $f(h,x,z)$ represents a CIG. We use a DSO cost functional of the form

$$J^{\mathrm{DSO}} = \int dx \left[\frac{\int dz \int dh\, (\partial_h f)^2}{\int dz \int dh\, f^2} \right]. \qquad (1)$$

The scaling by $1/\int dz \int dh\, f^2$ has been recommended by Chauris et al. (2001) and makes the functional less sensitive to the data and migration amplitudes. Note the similarity with the stacking power functional

$$J^{\mathrm{STACK}} = 1 - \int dx \left[\frac{\int dz \left(\int dh f \right)^2}{\int dz \int dh\ f^2} \right], \qquad (2)$$

which takes the integral of f over offset rather than the derivative with respect to offset. The important difference between the two functionals lies in there behavior away from the minimum. In that case, a highly not-flat event in the CIG may effectively stack to almost nothing and the stacking power functional will not be able to make a suitable correction on the velocity. This will show up as local minima in the cost functional. The DSO functional, however, will get a strong contribution from the curved event and make the appropriate model correction. For this reason, the DSO functional is expected to provide good results even when the starting model is not close to the true model.

The actual expression for J^{DSO} is more complicated and involves tapers, weights, and filters:

$$J^{\mathrm{DSO}} = \sum_{k=1}^{N_x} (f^{\mathrm{T}} \mathcal{A} f)_k / (f^{\mathrm{T}} \mathcal{B} f)_k, \qquad (3)$$

where k runs over a set of horizontal positions $\{x_k\}$, which may be spaced fairly sparse, but at less than half the horizontal spline node distance. The operator $\mathcal{A} = \mathcal{T}^{\mathrm{T}} \mathcal{F}^{\mathrm{T}} D_h^{\mathrm{T}} \mathcal{S}^{2\nu} D_h \mathcal{F} \mathcal{T}$ resembles an approximation to $-\partial_{hh}$. Here $\mathcal{T}$ is a taper that smoothly reduces the amplitudes towards the smallest and largest offsets to avoid edge effects. Also, $\mathcal{T}$ removes the larger offsets at smaller depths to suppress the contribution of refraction events. A filter $\mathcal{F}$, to be specified later on, is applied next. $\mathcal{S}$ is a smoother that is applied ν times, as is its transpose, which is the same as $\mathcal{S}$. For this smoother, we simply take convolution by $\{\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\}$ in each coordinate. A weighted difference operator is used,

$$(D_h r)_{l\,m} = \frac{(w_{l\,m} f_{l+1\,m} - w_{l+1\,m} f_{l\,m})}{(w_{l\,m} + w_{l+1\,m})}, \qquad (4)$$

where the subscript k has been suppressed. The weights $w_{k;l\,m}$ are computed during the migration by simply counting the number of contributing ray pairs in each point. For the scaling we have adopted $\mathcal{B} = \mathcal{T}^{\mathrm{T}} \mathcal{F}^{\mathrm{T}} \mathcal{F} \mathcal{T}$.

The minimization of the DSO cost functional is carried out by a limited-memory BFGS method (Zhu et al., 1994). This requires the computation of the gradient (sensitivity) of the cost functional with respect to the model slowness values at the spline nodes. The adjoint-state approach (Ciarlet, 1989) is used, implying that most of the computations should be done in reversed order. In particular, backward ray tracing needs to be done. We have simplified this part of the computation by a first-order approximation to the linearized backward ray equations, which happens to be a discretization of the perturbed ray equation, given by for instance Schnieder and Sambridge (1992):

$$\frac{\partial \tau}{\partial \sigma_{ij}} = -\int_0^{\tau} dt \, \frac{1}{\sigma} \frac{\partial \sigma}{\partial \sigma_{ij}}. \qquad (5)$$

Here σ denotes the interpolated slowness and σ_{ij} the slowness values at the spline nodes. The same technique without this approximation is described by Plessix et al. (2000). The use of the adjoint-state approach results in an optimal complexity: the cost of evaluating the functional is about the same as the cost of evaluating the gradient of this functional with respect to the slowness model.

Example

We have applied Differential Semblance Optimization to a marine seismic line. Some preprocessing was applied to the data: the first arrival was blanked out and a gapped decon operation was applied to remove the sea-bottom reverberations. Offsets ranged from 50 to 2250 m with 25 m spacing. There are 480 shots spaced at 25 m, covering a distance of about 10 km. A shortest-offset display of the data is shown in Fig. 1, with shot positions on the horizontal axis and time in msec on the vertical. There are clear events up till about 1 second. At later times, the amplitudes are rather weak, even with the compensation for spherical spreading applied in the figure. This is due to the hard layer between 0.8 and 1.0 seconds. Most of the events below this layer are due to surface and interbed multiples and converted waves.

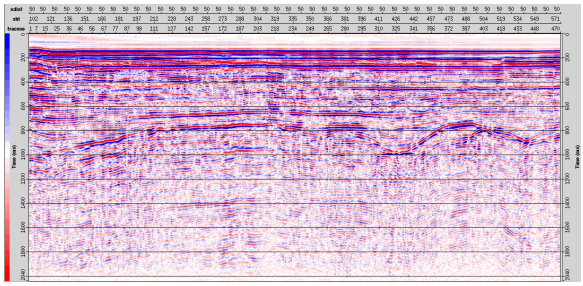

Fig. 1: Shortest-offset data panel.

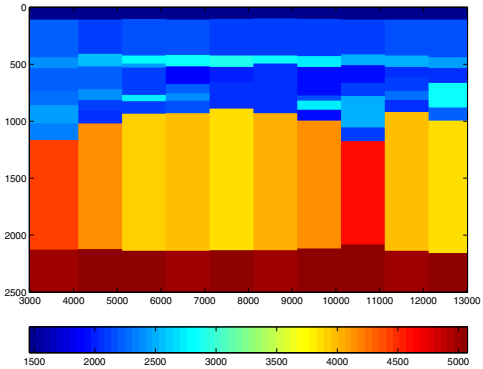

Fig. 2: NMO velocity model.

A Normal Move-Out (NMO) velocity model obtained for this data set is shown in Fig. 2. The true-amplitude migration image for a smoothed version of the NMO model using the migration kernel of the DSO code is displayed in Fig. 3.

Next we ran the DSO code on these data, starting from a *constant* velocity model at the sea water velocity (1.5 km/s). Initially, a one-parameter search for the best model with a velocity linear in depth was performed. From there on, the gradient method was applied. After 82 iterations (or about one night on a 500 Mhz Pentium PC), the velocity model shown in Fig. 4 was obtained. The grid of spline nodes had spacings of 1000 m in x and 100 m in z, the grid for the CIGs had 250 m in x and 10 m in z. For the resulting velocity model, a true-amplitude migration image was computed on a finer grid (25 m in x, 5 m in z). The result is shown in Fig. 5. By comparing the velocity model and migration image in Figs. 4 and 5, it is obvious that there is no correlation between the reflectors and the velocity model in the deeper parts below 1 km. This strongly suggests that most of the structures beyond 1 km depth are artifacts due to multiples or per-

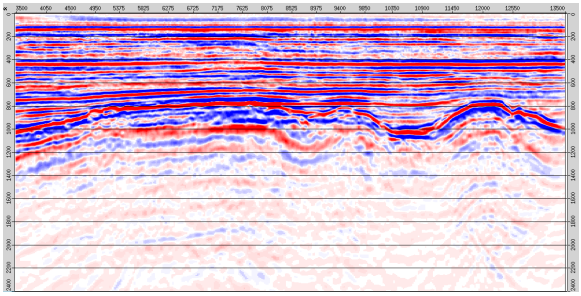

Fig. 3: True-amplitude migration image for NMO model.

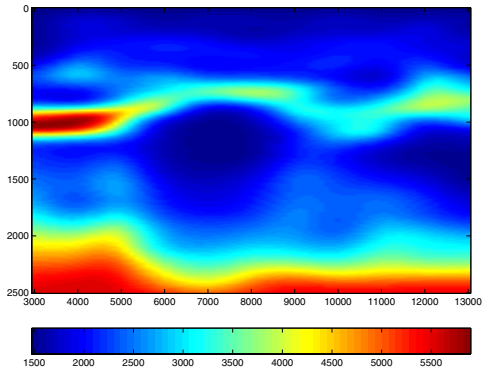

Fig. 4: Velocity model obtained by DSO.

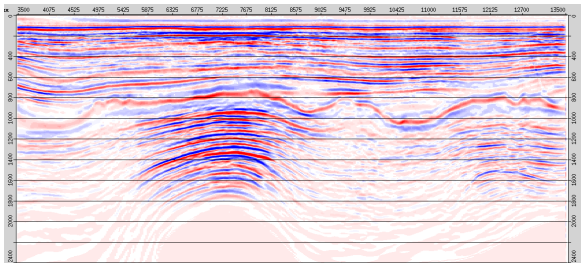

Fig. 5: True-amplitude migration image for DSO model.

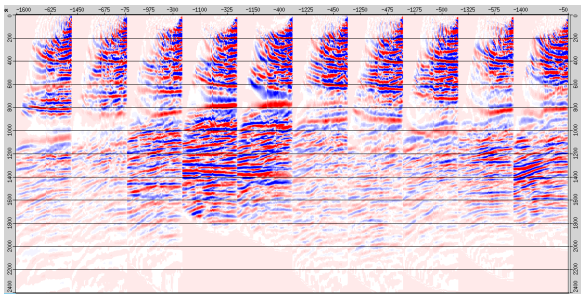

Fig. 6: A number of Common Image Gathers.

haps converted waves. CIG panels for a subset of lateral positions x_k are displayed in Fig. 6. Indeed, conflicting events are present beyond a depth of 1 km. Some of these events are probably longer-period surface multiples.

As standard multiple elimination procedures proved to be insufficient, a multiple filter was incorporated into the DSO cost functional. This is the $\mathcal{F}$ in Eq. (3). The idea is to introduce a bias that ignores the undermigrated (downward curving) events in the CIGs, which are due to surface multiples that have seen a lower effective velocity than the primary reflections. $\mathcal{F}$ is implemented by taking a 2D fast Fourier transform the CIG, reducing the amplitude of down-dipping events, and transforming back. This procedure is outlined in Fig. 7. At the top left, the original CIG is shown. The modulus of its Fourier transform is displayed at the bottom left. The zero horizontal wavenumber is centered; the vertical wavenumber is zero at the top and increases downwards. The downward dipping events lie in the right half of the panel. Most of the peaks occur at zero horizontal wavenumber and correspond to events that have already been flattened. After reducing the amplitude in the right half of the panel by tapering with half a Gaussian, we obtain the right bottom figure which produces the top right result after an inverse Fourier transform. Note that a number of downward curving events have been removed and the remaining events stand out more clearly.

With this filter, the DSO code produced a velocity model in 34 iterations, starting from a constant initial model. The model and the migration image on a finer grid are shown in Figs. 8 and 9, respectively. Some of the CIGs are displayed in Fig. 10. Drawn are the panels *before* filtering, so the downward curving events are still there. The same CIGs *after* the application of the multiple filter are displayed in Fig. 11. Note that the resulting velocity model bears more resemblance to the original NMO model that the previous one.

Conclusions

We have described an efficient DSO code that uses wavefront construction for fast ray tracing and has a gradient computation of optimal complexity. DSO provides highly coherent migration images for a velocity model that may be incorrect in the presence of strong coherent noise. The word "noise" is used here to denote any data component that cannot be explained by the model. In the high-frequency Born approximation employed here, this includes multiples, refraction and diffraction events, shear waves, and so on. We removed the effect of short-period sea bottom multiples by preprocessing, whereas the longer-period surface multiples were suppressed by an embedded filter that weakens undermigrated events, thus introducing a bias towards higher velocities. In the marine example presented here, the DSO cost functional converged to an acceptable velocity model, starting from a constant model at the sea-water velocity. This illustrates the strength of the DSO cost functional over those based on stacking power or least-squares fitting of the data, which are known to suffer from local minima.

Multi-valued arrivals have been suppressed by only accepting the shortest travel times. In principle, all arrivals can be taken into account in the code but this has not been investigated. The DSO cost functional, if

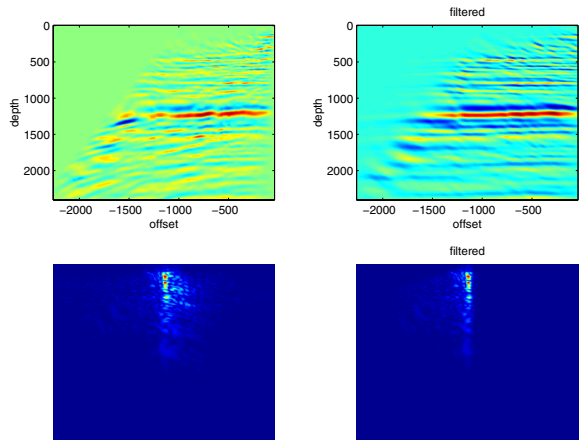

Fig. 7: Filter that reduces the effect of multiples.

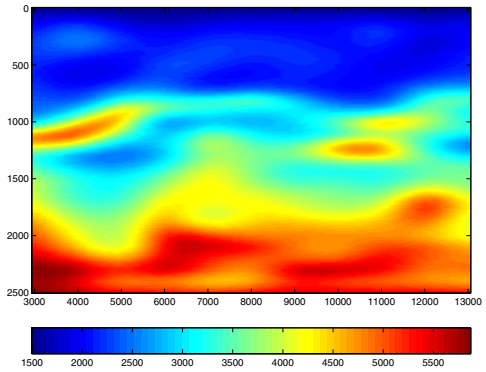

Fig. 8: Model obtained with the multiple filter.

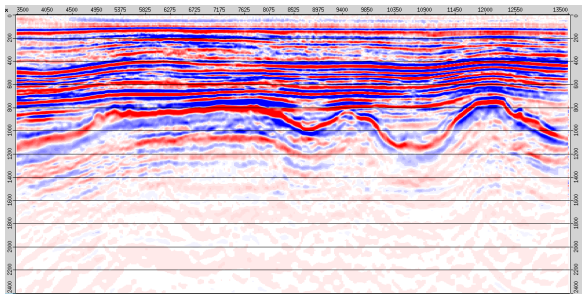

Fig. 9: True-amplitude migration image using the multiple filter.

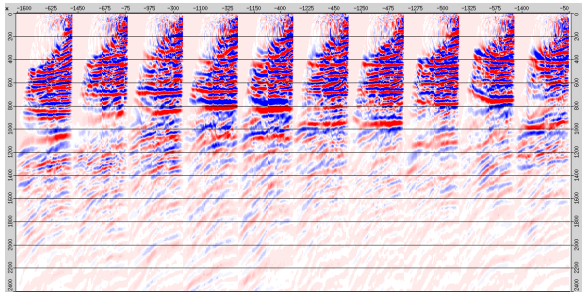

Fig. 10: A number of Common Image Gathers.

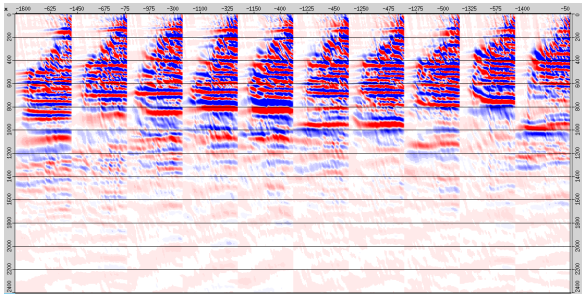

Fig. 11: CIGs after applying the multiple filter.

parametrized by offset, may loose its global convexity in the presence of multiple arrival times and exhibit local minima. If offset is replaced by reflection angle, this problem might disappear (Brandsberg-Dahl et al., 1999). A correct treatment of multi-valued arrivals is expected to be crucial for handling complex subsurface models.

References

Brandsberg-Dahl, S., de Hoop, M., and Ursin, B., 1999, Velocity analysis in the common scattering-angle/azimuth domain: 69th Ann. Internat. Mtg., Soc. Expl. Geophys., 1715–1718.

Chauris, H., and Noble, M., 2001, 2D velocity macro model estimation from seismic reflection data by local differential semblance optimization: application on synthetic and real data: Geophys. J. Internat., **144**, 14–25.

Ciarlet, P., 1989, Introduction to numerical linear algebra and optimisation: Cambridge University Press, Cambridge.

Plessix, R.-E., Mulder, W., and ten Kroode, A., 2000, Automatic crosswell tomography by semblance and differential semblance optimization: theory and gradient computation: Geophys. Prosp., **48**, 913–935.

Snieder, R., and Sambridge, M., 1992, Ray perturbation theory for traveltimes and ray paths in 3-D heterogeneous media: Geophys. J. Internat., **109**, 294–322.

Symes, W., and Carazzone, J., 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 2061–2073.

Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic data: Geophysics, **51**, 1893–1903.

Vinje, V., Iversen, E., and Gjøystdal, H., 1993, Traveltime and amplitude estimation using wavefront construction: Geophysics, **58**, 1157–1166.

Zhu, C., Byrd, R., and Nocedal, J. L-BFGS-B - FORTRAN subroutines for large-scale bound constrained optimization:. Report NAM-11, EECS Department, Northwestern University, 1994.