Non-interactive Estimation of the Marmousi Velocity Model by Differential Semblance Optimization: Initial Trials

W. W Symes

CRPC-TR90104
August, 1990

Center for Research on Parallel Computation
Rice University
P.O. Box 1892
Houston, TX 77251-1892

Non-interactive Estimation of the Marmousi Velocity Model by Differential Semblance Optimization: Initial Trials

W.W. Symes

Department of Mathematical Sciences

Rice University

Houston, Texas U.S.A.

August 1990

ABSTRACT

A subset of the Marmousi data is processed to yield an initial approximation to a kinematic velocity model. The process described is the first step in a gradient iteration scheme for a modified least-squares inversion method. The approach requires no picking or other extensive interaction with the data, and appears to avoid convergence difficulties reported for conventional least-squares inversion.

INTRODUCTION

A critical step in migrating (or "inverting") the Marmousi data set is the construction of the velocity model. It is not too strong to say that other aspects of data treatment stand or fall along with the success in velocity estimation. Presumably velocities are equally important in treating data from structurally complex areas in the real world and perhaps in extracting subtle features even in structurally simple areas.

The results presented below represent an initial attempt at noninteractive extraction of velocity models directly from 2-D waveform data. We invert only a small subset of the Marmousi dataset, namely ten shot records from the west end of the line. The structure there is "layered," and we extract mainly laterally homogeneous velocity models. We use only two steps of an extremely crude optimization method, and start with a substantially wrong initial model. Nonethelss, the algorithm moves the velocity in the right direction. It will be clear that larger datasets and more sophisticated optimization methods could be used, and laterally heterogeneous velocites obtained, though the quality of such results remains to be determined. Thus we have not addressed the central topic of the workshop — determination of complex velocity structure — in a definitive way. Nonetheless we believe that the present, preliminary results, though simplistic and crude, justify further investigation of our approach.

Our approach belongs to the "least-squares inversion" genre. It is by now well established that straightforward least-squares inversion (i.e. model-based data fitting *via* gradient-type optimization) of reflection seismograms is incapable of extracting the very important slowly varying trends in wave velocity. Therefore we have modified least-squares inversion in an essential way, to produce a class of algorithms which can, under some circumstances, successfully extract velocities. We call these algorithms "differential semblance optimization" (DSO), or "coherency optimization", for reasons that will become apparent below. Our previous papers on this topic have dealt with a version appropriate to plane-wave data sets and layered models (Symes 1988, 1990a, Symes and Carazzone 1989, 1990, Santosa and Symes 1989). In those papers we gave a complete

mathematical foundation for the DSO approach to layered velocity estimation, and showed by numerical experiments with both synthetic and field data that DSO produces reasonable velocity estimates (and, along with these, reasonable reflectivities), essentially without human intervention.

For this article we present a version of DSO adapted to 2-D shot gather inversion. We define an objective function of the velocity model which DSO is to minimize, and describe the calculation of this function and its gradient. We implemented this calculation, and applied it to a small part of the Marmousi data set near the well at 2700 meters. Our main result is a laterally homogeneous velocity model obtained by two steps of the steepest descent method. Steepest descent is the simplest of the gradient-based optimization methods. The satisfactory nature of the steepest-descent result suggests that the DSO objective function is actually as well-behaved as the theory indicates.

Before giving the details of our approach and results, we interject a few remarks about the nature of direct velocity estimates, and about the possible role of non-interactive techniques.

Kinematic Versus "Real" Velocities

The velocities constructed *via* DSO are designed to account for the moveout in the data: i.e. to explain the kinematic features of the data in an internally consistent way. In this respect, DSO is similar to other velocity analysis techniques, including reflection tomography.

It is important to understand that a velocity model may succeed in the kinematic sense without possessing many features, both obvious and subtle, of the true distribution of compressional velocity in the earth. Kinematic velocity models need not make geological sense, may lack identifiable horizons, and may vary sufficiently as inversion parameters are changed to prevent association with well-defined geological units. These points are well-understood in the context of stacking velocities. Our point here is that "inversion velocities" suffer from many of the same limitations. For example, our prior work on layered velocity estimation produced smoothly varying models, very different in appearance from typical blocked sonic logs or interpreters' facsimiles thereof. Moreover, our experimental codes allowed control over smoothing parameters, for instance; varying these parameters in experiments with "real" p-tau seismograms (processed from field data from structurally simple areas) yielded a range of smooth velocity models, differing substantially in point values but equally "valid", in the kinematic sense. Some examples with synthetic and field data appear in Symes 1990a and Symes and Carazzone 1989, 1990.

It is instructive to recall the state-of-the-art constructions of velocity models employed by other contributors to this workshop. These methods yield blocked models based on horizon picks from preliminary sections of various sorts. The blocks are filled in with sparsely parameterized (e.g. linear) velocity samples. Such models have intrinsic structural meaning, lacked by the smooth models produced by DSO. It is very difficult to see how an entirely automatic, pick-free process such as DSO could emulate the judgement inherent in blocked models.

Of course, structural information is not lost by resort to smooth, "non-geological" models, which can encode the kinematics of reflection just as successfully as can blocked structural models (again, this judgement is based on experience with the layered case). Any kinematically correct model produces a post-inversion stack in which any structural information present in the data is preserved, and available for interpretation. While the eventual role, if any, of noninteractive velocity estimation is far from clear, its successful implementation does not destroy geological information, at least. Moreover, while temporarily excluding informed human judgement from the data process, DSO for example also excludes inadvertent bias implicit in sparse parameterization, and (hopefully) yields an optimally consistent and "objective" account of moveout.

If we accept these arguments as justifying an interest in non-interactive estimation of *kinematic* velocities, there remains the question of its feasibility, to which we now turn.

Differential Semblance Optimization

We based our work on the linearized wave equation of linear acoustics, which relates the

(smooth) velocity field $v(x,z)$, the (oscillatory) reflectivity $r(x,z)$ $(= \delta v(x,z)/v(x,z))$, the source wavelet $f(t)$, and the seismogram (shot-gather) $S[v]r(x_s,x_r,t)$. Here x_s denotes shot position, x_r receiver position and x,z position coordinates in the earth. The notation is chosen to emphasize that S depends linearly on r, nonlinearly on v. The boundary value problem defining the relation $v,r \mapsto S$ has appeared in many other publications, and we do not repeat it here. We discretized the partial differential equations and boundary conditions using low-order finite differences in the usual fashion.

In physical terms, our use of the linearized wave equation amounted to neglect of multiple reflections, while our neglect of density variations restricted our ability to model the offset dependence of reflectivity. We felt that neither of these omissions would prevent a successful "first pass" at velocity analysis for the Marmousi data set. In a similar vein, we used a zero-phase Ricker wavelet in our modeling, in full awareness that the actual wavelet differed substantially in phase and amplitude. Finally, we used the full two-way wave equation, precluding the use of velocity models with sharp interfaces. As discussed above, we intended to use only smooth velocities anyway, and also stayed close to "reality" by using two-way propagation (that's how the data were generated).

As mentioned in the introduction, the straightforward application of least-squares inversion in this setting fails to produce useful velocity estimates, even of the kinematic type. Our monograph (Santosa and Symes 1989) provides detailed explanation of this failure, with many references. In Appendix E of that volume, we introduced "coherency optimization," since renamed differential semblance optimization, a modification of least-squares inversion, designed to remedy its defects as a velocity estimator.

DSO is based on the use of a *maximal model space*. The model explained above is *minimal* in the sense that it contains (hopefully!) precisely the physics needed to explain the data. A single reflectivity distribution, in particular, is required to fit the data at all shots. This is only possible for a very nearly correct velocity: it is this feature which accounts for the relative uselessness of straightforward least-squares inversion. It is also well-known that inversion of a *single shot gather* (in the linearized setting) is very easy, up to the noise level in the data. Therefore *any* mapping of a model space into the data space can be factored through the set of *shot-dependent reflectivity models*

$$\{v(x,z), r(x_s; x, z)\}$$

which is therefore a *maximal* model space. The minimal models, i.e. the shot-independent reflectivities, are characterized for instance by the equation

$$\frac{\partial r}{\partial x_s} \equiv 0 \,. \tag{1}$$

For the rest of this paper reflectivity distributions are presumed to be shot-dependent.

Of course, the earth is unique, not shot-dependent, so we want to enforce equation (1). The first main mathematical result of the theory is that this constraint can only be enforced in a soft sense, through a penalty term. (Otherwise a pathological optimization problem results.) This idea leads to the cost function

$$J_{1,\sigma}[v,r] = \frac{1}{2}\left[\sum_{x_r,x_s}\int dt\,(S[v]r - S_{\text{data}})^2 + \sigma^2 \sum_{x_s}\int\int dx\,dz\left(\frac{\partial r}{\partial x_s}\right)^2\right] . \tag{2}$$

The roles of the terms in J_1 are clear. The first serves to adjust r to fit the data S_{data}. The second forces the various reflectivities for neighboring shot locations to resemble each other, whence the name of the technique: differential semblance. The influence on v is indirect, through the second (semblance) term.

Once r is determined, the second term measure the "flatness" in the common image gather, i.e. $r(x_s,x,z)$ for fixed x, in a very strict sense: the traces in this gather must be *the same* to make the second term vanish. In other variants of migration velocity analysis, as described in other papers in this volume, common image gather traces are merely required to be similar. We are able to penalize the difference, rather than some looser measure of semblance, because we deal with the

(inverted) physical parameter $r = \delta v/v$, which ought to be shot independent, rather than the migrated image, which may have shot-dependent amplitude.

Since $J_{1,\sigma}$ contains (a simple version of) the conventional least-squares objective function, it must suffer from the same malady: extreme nonlinearity in v. The second main mathematical result is that *elimination of* r (a quadratic minimization!) yields a very smooth, nearly quadratic function of v:

$$J_\sigma[v] = \min_r J_{1,\sigma}[v, r] \,. \tag{3}$$

J_σ is the differential semblance functional; its minimization is DSO. Moreover, some evidence exists (even proof, in the layered case) that J_σ is strongly convex near its global minimizer, and that secondary minima occur only outside reasonable *a priori* definable model sets, for suitably chosen σ. For noise-free data, $S_{\text{data}} = S[v^*]r^*$, a global minimizer of J_σ is clearly v^*, with

$$J_\sigma[v^*] = J_{1,\sigma}[v^*, r^*] = 0 \,.$$

Thus we can expect that global minimization of J_σ is possible for at least low-noise data sets, and would yield accurate kinematic velocity models.

The current state of the theory of DSO for multidimensional shot gather inversion is reviewed in Symes 1990b. We end this section with a brief description of the gradient calculation, which is derived in detail in Symes 1990b; the derivation is sketched in the Appendix. Note that the right-hand side of (3) implicitly defines the reflectivity r as a function of the velocity v and the data; thus we write $r = r_\sigma[v]$.

The velocity models v in this approach must be restricted to a class V of smooth models. Denote by P the projection operator onto V, orthogonal in the sense of the usual (L^2) inner product. Then the gradient of J_σ is given approximately by the expression

$$\text{grad} J_\sigma[v] \cong$$
$$P\left\{\frac{1}{h}S[v + hr_\sigma[v]]^T \left(S[v]\cdot r_\sigma[v] - S_{\text{data}}\right)\right\} \tag{4}$$

This formula has several features worthy of comment. First, the expression inside the curly brackets does not depend on the choice of V, except that v should belong to V, at least in principle. That is, the projection onto the "feasible" velocities in V is a postprocess; once the quantity in the brackets is calculated, the gradient can be computed for any choice of V containing v. For instance, we can obtain both laterally homogeneous and heterogeneous versions of the gradient from the same "raw" section.

Second, the adjoint operator S^T represents a basic version of prestack migration (Lailly 1983). That is, application of S^T to a collection of shot records involves (before-stack) migration of each record, followed by a stack. This is so provided that the shot record migrations are carried out via two-way reverse-time extrapolation and cross-correlation with the incident field.

Third, the gradient of J is uncannily similar to the gradient of the usual least-squares error (Lailly 1983, Tarantola 1984). In fact, the only difference of substance is that S^T is evaluated at $v + hr_\sigma[v]$, rather than at v, for a suitable (small) choice of the scalar h. The inclusion of the high-frequency, shot-dependent term $hr_\sigma[v]$ in the background field for the migration is responsible for the presence of low-frequency trend information in the final output.

The accuracy of the approximation is determined in part by the choice of the scalar h.

In reality, we cannot access the inverted reflectivity $r_\sigma[v]$, but only an approximation. In the present implementation of our algorithm $r_\sigma[v]$ is approximated by a small, fixed number of conjugate gradient iterations applied to the quadratic minimization (3). We have shown elsewhere (Symes 1990(b)) how formula (4) can be corrected to compensate for the resulting inaccuracy in $r_\sigma[v]$.

Initial Application to the Marmousi Data Set

The calculations described in the preceding pages were implemented in Fortran and C. For a first trial with the Marmousi dataset, we selected shot records $1 + 4(i-1)$, $i = 1, \ldots 10$, to use as data for the inversion. That is, the shot spacing in our truncated dataset of ten shot records is $\Delta x_s = 100m$, with the first (westernmost) shot

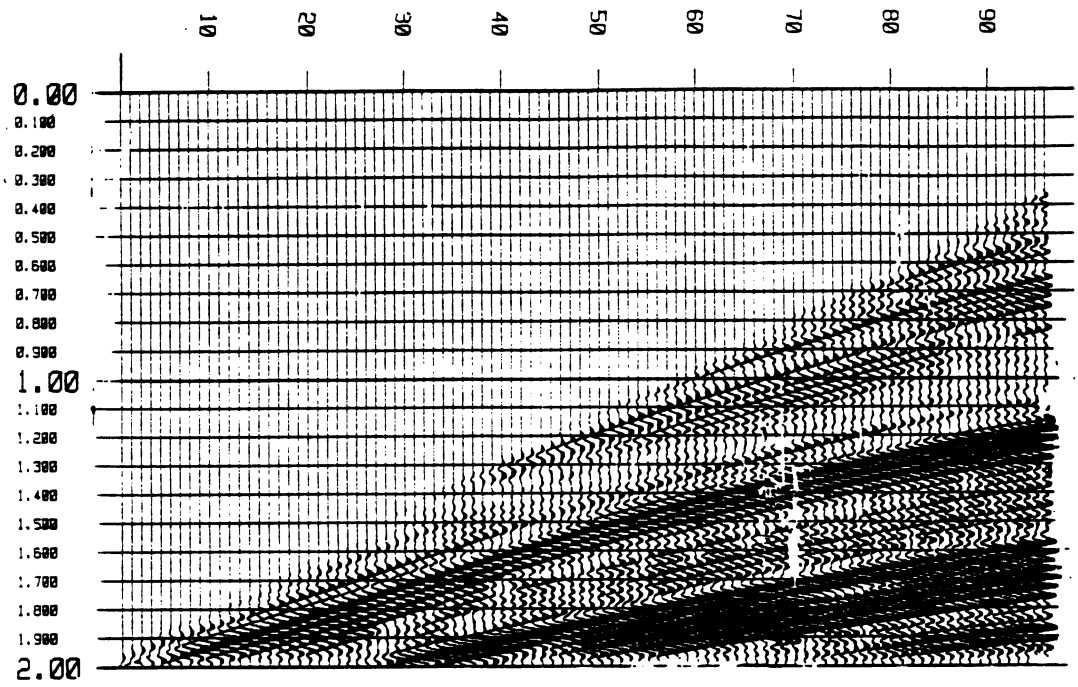

FIGURE 1 (a) Muted shot record 1.

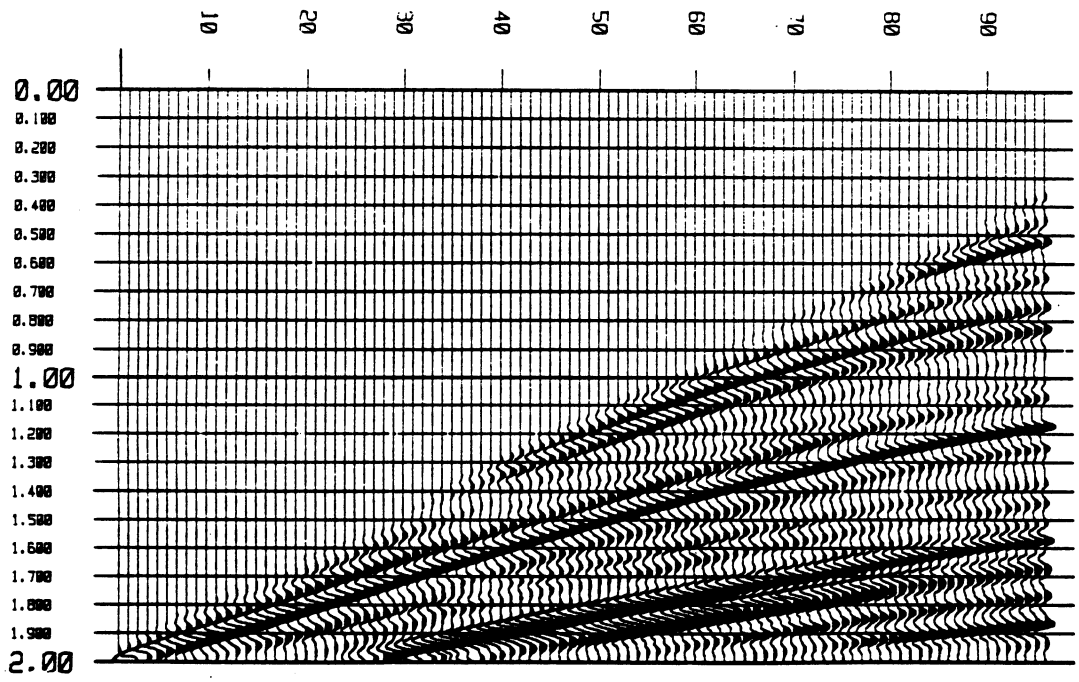

FIGURE 1 (b) Muted shot record 1, after filtering by 25 Hz high-cut Gaussian. First record of input data set for inversion.

being located at $3000m$ from the west edge of the model.

Our preprocessing of this small data subset was minimal. We applied a linear mute to each shot record, with mute $t_0 = 200m/s$ and mute velocity of 1500 m/s, to remove the water layer reverberations and direct wave. We also truncated the data at $2s$. The muted, truncated version of shot 1 is displayed as Figure 1(a). We then applied a Gaussian filter with high cut at 25 Hz to each trace; the filtered version of shot 1 appears as Figure 1(b).

The removal of the direct wave *via* the linear mute is a simple way to force the data to approximate that of a linearized model. We believe that relatively little multiply reflected energy is present in the Marmousi dataset, so that the direct wave is the principal difference between the actual (nonlinear) dataset and one that would have been produced by linearization about a suitable smooth reference model.

The truncation to $2s$ in the time domain, and 25 Hz in the frequency domain, were indicated by the limits of our computing environment, as was the overall size of the dataset employed (10 shots). All of the calculations reported here were carried out on a Stardent Titan 3000 series graphics super workstation, with four P3 processors and 64 Mb of core memory. The finite difference calculations were carefully parallelized and vectorized, so that the Titan yielded roughly 20 Mflops in the main part of the calculations. Nonetheless, even with $\Delta t = 4ms$ and $\Delta x = \Delta z = 15m$ (a very coarse grid), a single application of the normal operator (the main conjugate gradient calculation) required roughly 45 min. CPU. A single gradient estimation required roughly eight hours, and a single steepest-descent step two days. All results presented below were obtained in the first two weeks of November 1990.

We used a centered finite difference scheme of order 4 in both space and time. This scheme gives good results on the grid just described for solution components up to about 20 Hz, and reaches its stability limit at $v = 2300\, m/s$. This latter fact also restricted our attention to the upper part of the model.

We defined the space V of velocity models by means of a so-called multiresolution or wavelet basis of L^2 (Daubechies 1988, Meyer 1990), computed via the fast algorithm of Mallat 1987. We are grateful to R.M. Lewis for coding the Mallat algorithm. Multiresolution bases give a convenient way of selecting components of a signal according to scale, or correlation length, in a more localized fashion than is permitted by the Fourier transform. We implemented multiresolution projection in both 1-D (laterally homogeneous gradients) and 2-D. We used a correlation length of $300m$ in all results reported below.

While a detailed numerical analysis of our approach remains to be accomplished, an initial consideration is clearly the accuracy with which the x_s-derivative in the differential semblance condition (1) is approximated. We used the simplest possible approximations, viz.

$$\frac{\partial r}{\partial x_s} \cong \frac{1}{\Delta x_s}(r(x_s + \Delta x_s) - r(x_s)) .$$

For this approximation to succeed, it is essential that the traces at shot location $x_s + \Delta x_s$ have time shifts of somewhat less than a wavelength relative to those at shot location x_s. A quick estimation of this displacement is obtained by examination of migrated common image gathers (the (before-stack) migrations being the first step of the conjugate gradient process for reflectivity estimation). These are displayed for horizontal location $x = 3000m$ at reference velocities $v = 1500\, m/s$ and $v = 1750\, m/s$ as figure 2(a), (b). Clearly the traces in Figure 2(a) are completely aliased; to deal with a velocity range including $v \equiv 1500\, m/s$, a smaller Δx_s than $100m$ is required. On the other hand neighboring traces in Figure 2(b) exhibit relative time shifts well under a wavelength, for the most part — in fact the shallower events are lined up rather well. Therefore we used $\equiv 1750\, m/s$ as the initial velocity estimate.

The gradient calculation explained in the last section was carried out using the data set and reference velocity described above. The penalty parameter σ^2 was chosen to make the second term in the expression (2) less than 5% the size of the first term. The reflectivity was estimated using 10 conjugate gradient steps. The correction for re-

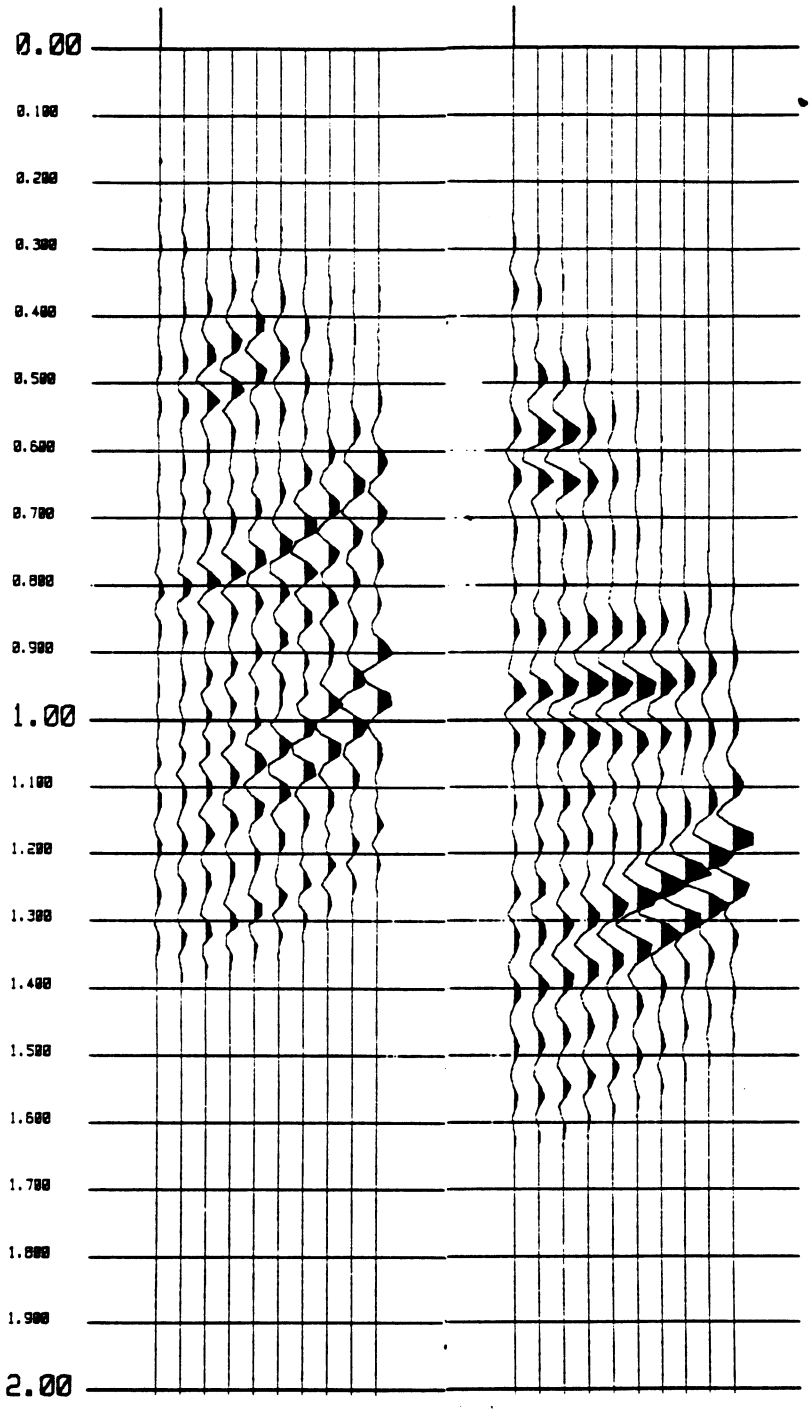

FIGURE 2. Left: Migrated common image gather at offset $3000m$ for migration velocity $1500\ m/s$.
Right: Migrated common image gather at offset $3000m$ for migration velocity $1750\ m/s$.

maining reflectivity errors, mentioned at the end of the last section, was not employed. Thus we expect the computed gradient to be only a crude approximation to the true gradient.

Figure 3 displays a mesh plot of the gradient in the space of laterally heterogeneous multiresolution series truncated at correlation length of $300m$ in each direction. Considerable horizontal oscillation is evident at this length scale, as are apparent aperture and truncation artifacts. In fact, the range of well-covered midpoints is so small for this data set that we do not believe any inference about lateral velocity variation. Instead we extract a layered velocity gradient, by projecting the raw section (inside brackets in (4)) onto the *layered* multiresolution series truncated at length scale $300m$. The resulting depth profile is displayed in Figure 4 (lower curve). We then approximated a step of the steepest descent algorithm by computing the approximate value of the objective function J_σ (3) along the line segment in (layered) velocity model space defined by adding multiples (0.0, 0.25, 0.5, 0.75, 1.0) of the negative computed gradient to the reference velocity (= 1750 m/s). The negative gradient was first scaled so that all velocities generated in this way did not exceed 2250 m/s, as dictated by the stability condition for our choice of space-time grid and difference scheme. The resulting values for the two summands in (2) and their sum are displayed in Figure 5.

We believe (based on experience with other synthetic data sets) that the mean-square error (first term in (2)) is heavily influenced by the error in our choice of wavelet. As mentioned in the introduction, we used a zero-phase Ricker wavelet with peak frequency of 10 Hz in our modeling and migration. This choice is certainly quite different from the "true" effective wavelet of our data set, and by itself prevents a good fit-to-data. Accordingly we used the differential semblance (second term in (2)) to govern our choice of steepest descent step. From Figure 5 one might guess that the optimum lies between 25% and 50% of the negative gradient step. The result of adding 37.5% of the scaled negative gradient to the reference velocity (1750 m/s) is the upper curve in Figure 4, which we will call "the first iteration".

This result is worth several comments. First, the DSO variational principle has moved the velocity in the right direction (increase), even though its parts are contaminated by an undetermined but probably large amount of noise. Second, the increase in velocity has been placed too shallow. That is easy to understand: the reference velocity is too low, so all aspects of the section are interpreted as occurring at incorrectly shallow depths. Third, the "un-geological" aspect of kinematic velocity estimates is obvious in this simple instance. The smoothness of the estimate prevents any obvious "geological" interpretation. Even worse, the velocity returns to its surface value at depth - a very unrealistic feature! On the other hand, the velocity below $2000m$ has no influence on the data set used here. Therefore, the deep part of velocity is not constrained by the data. It is certainly possible to build in constraints of various sorts on the basement velocity, by re-defining the projection operator in (4). We chose not to do so in order to emphasize the kinematic nature of our velocities.

To end the series of experiments, we computed the gradient of J_σ at the first iteration. The result, suitably scaled, is displayed as the lower smooth curve in Figure 6. We then guessed a suitable proportion of the negative gradient to add to the first iteration, and thus produced the second iteration. This is displayed as the upper smooth curve on Figure 6, along with the "log" velocity profile at 2700 meters, supplied by IFP. Note that the velocity increase has been moved closer to its proper position in depth.

The progress of the optimization can also be gauged by examination of common image gathers of the estimated reflectivity. In Figure 7 we display reflectivity common image gathers at offset $3000m$ for the (initial) constant reference velocity 1750 m/s (left plot), first iteration (middle plot), and second iteration (right plot). At the initial velocity, the shallow events are lined up but the deeper ones are undercorrected. At the first iteration the deeper events are improved but the shallow ones are now overcorrected. The second iteration removes some of the overcorrection of the shallow events and without significantly flattening the deeper ones. The gradient has detected the

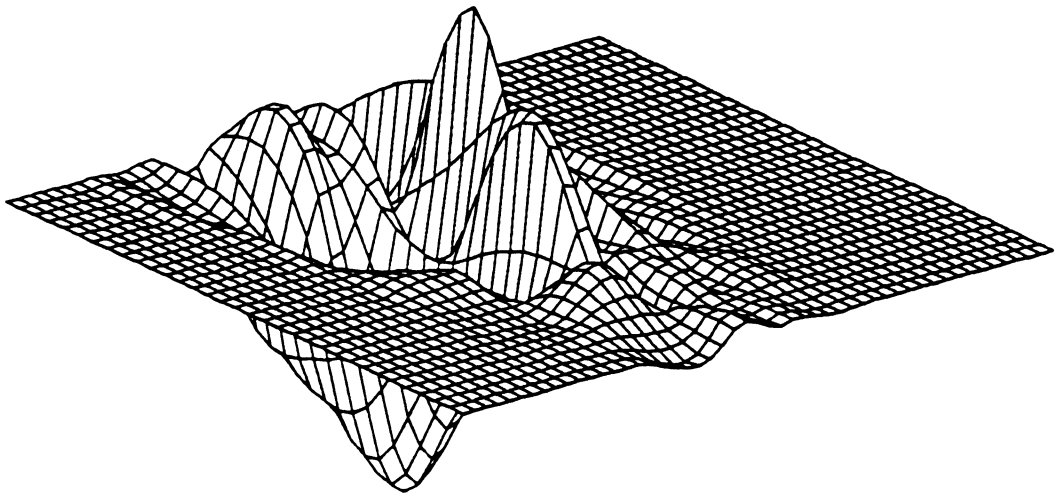

FIGURE 3. Mesh plot of 2D multiresolution gradient at reference velocity 1750 m/s, correlation lengths $300m$ in each direction.

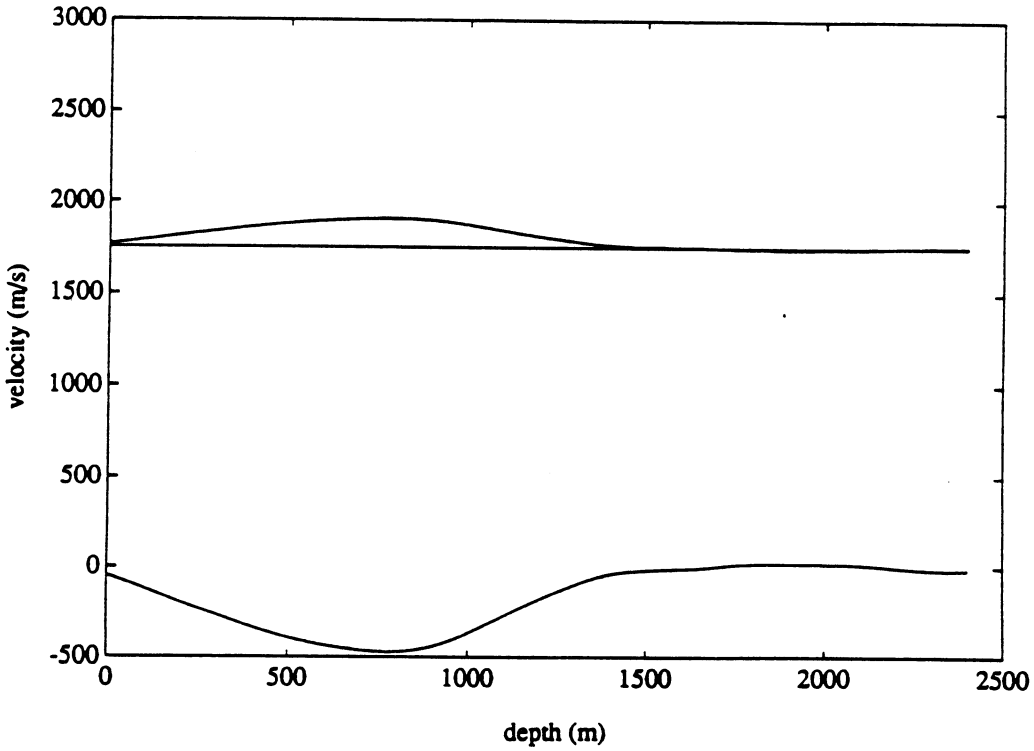

FIGURE 4. Lower curve: profile of 1D (layered) multiresolution gradient at reference velocity $1750m/s$, correlation length $300m$. Middle curve: constant reference velocity 1750 m/s.
Upper curve: result of first approximate steepest descent step from reference velocity 1750 m/s.

% of scaled negative gradient added	mean-square error	differential semblance	J_σ
0	0.5020	0.0158	0.5179
25	0.5030	0.0139	0.5169
50	0.5097	0.0148	0.5245
75	0.5385	0.0171	0.5556
100	0.6111	0.0220	0.6331

FIGURE 5. Optimization along the negative gradient direction.

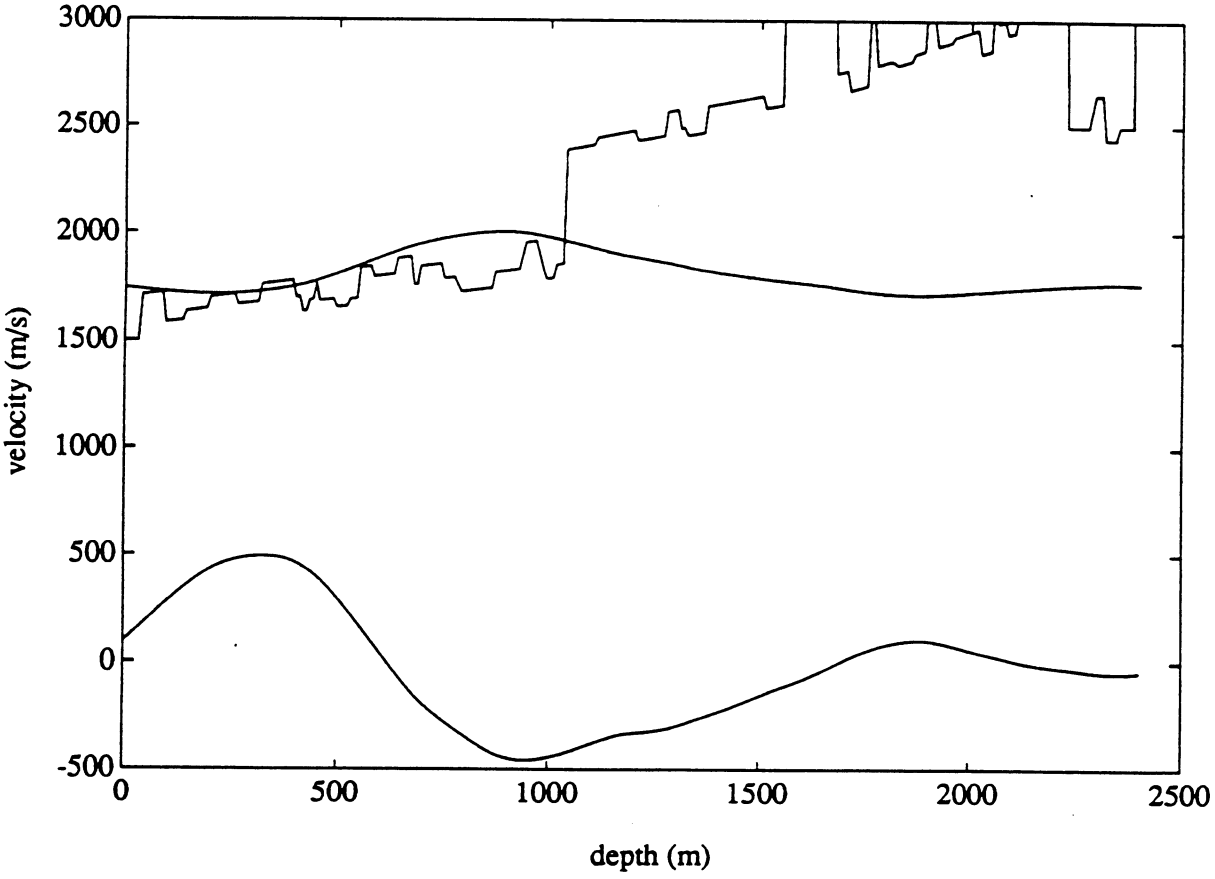

FIGURE 6. Lower smooth curve: profile of 1D (layered) multiresolution gradient at reference velocity given by upper curve in FIGURE 4.
Upper smooth curve: final layered velocity estimate, result of second approximate steepest descent step.
Upper rough curve: "log" velocity profile at offset 2700m.

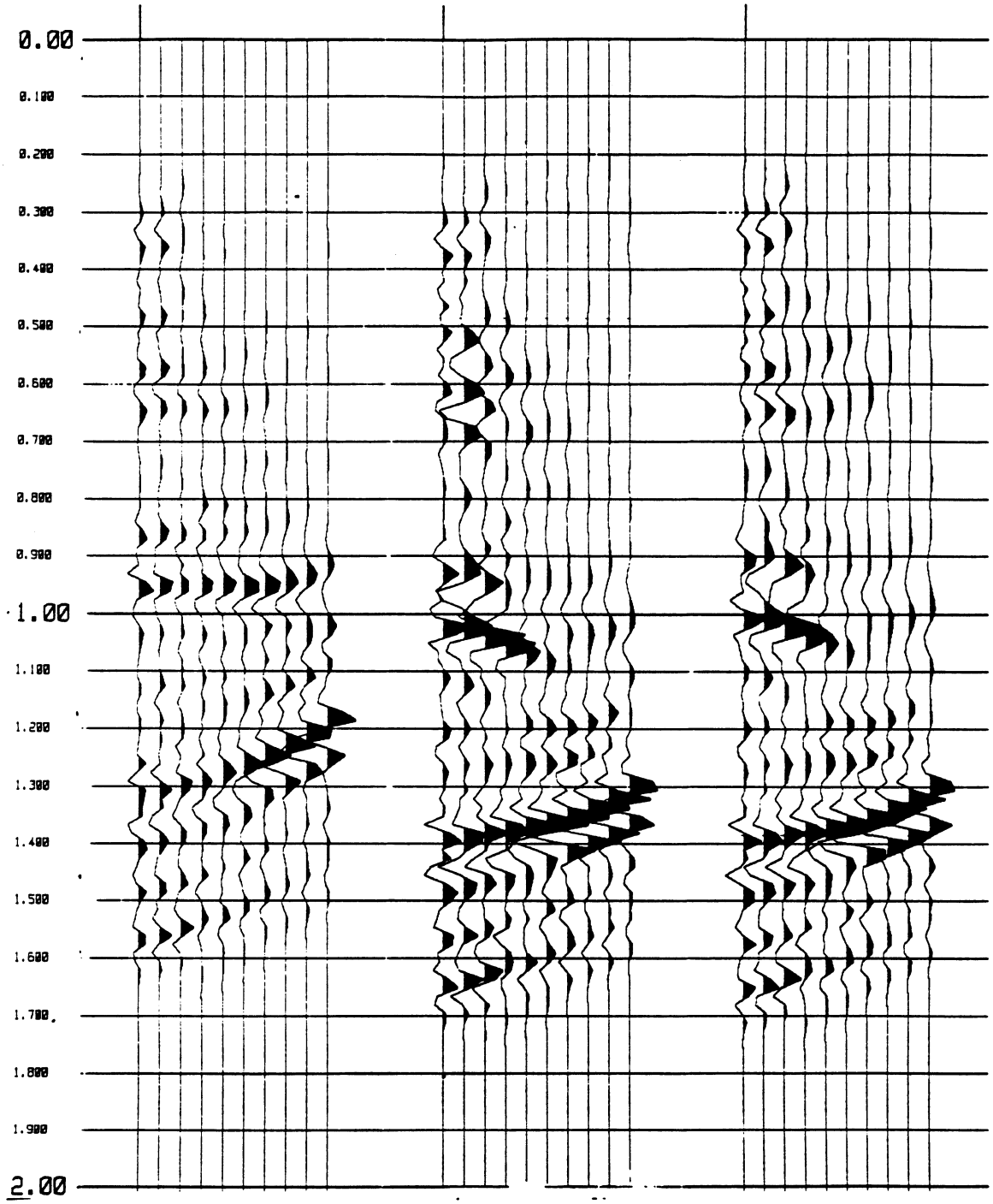

FIGURE 7. Common image gathers of reflectivity at offset $3000m$.
Left: reference velocity $1750\,m/s$.
Middle: reference velocity = first steepest descent iteration
= upper curve in FIGURE 4.
Right: reference velocity = second steepest descent iteration
= upper smooth curve in FIGURE 6.

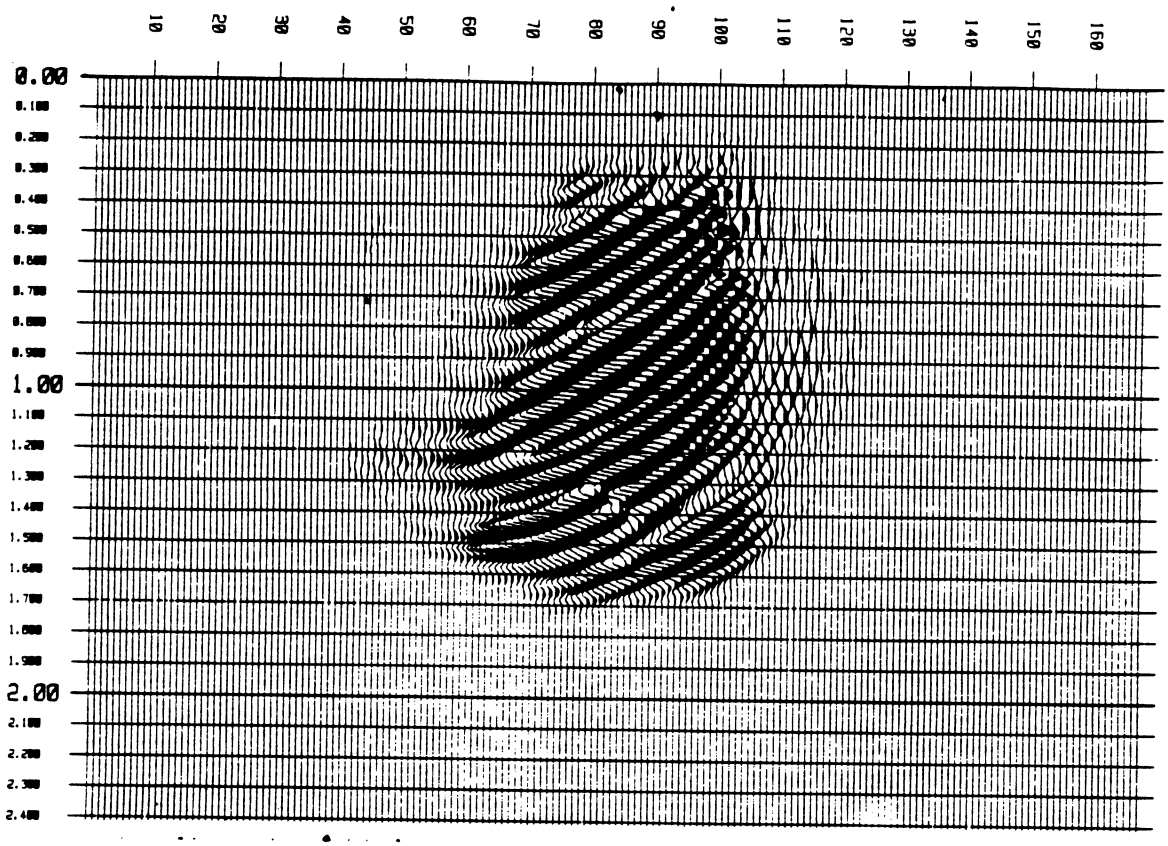

FIGURE 8. Stacked migrated section at reference velocity = second steepest descent iteration = upper smooth curve in FIGURE 6.

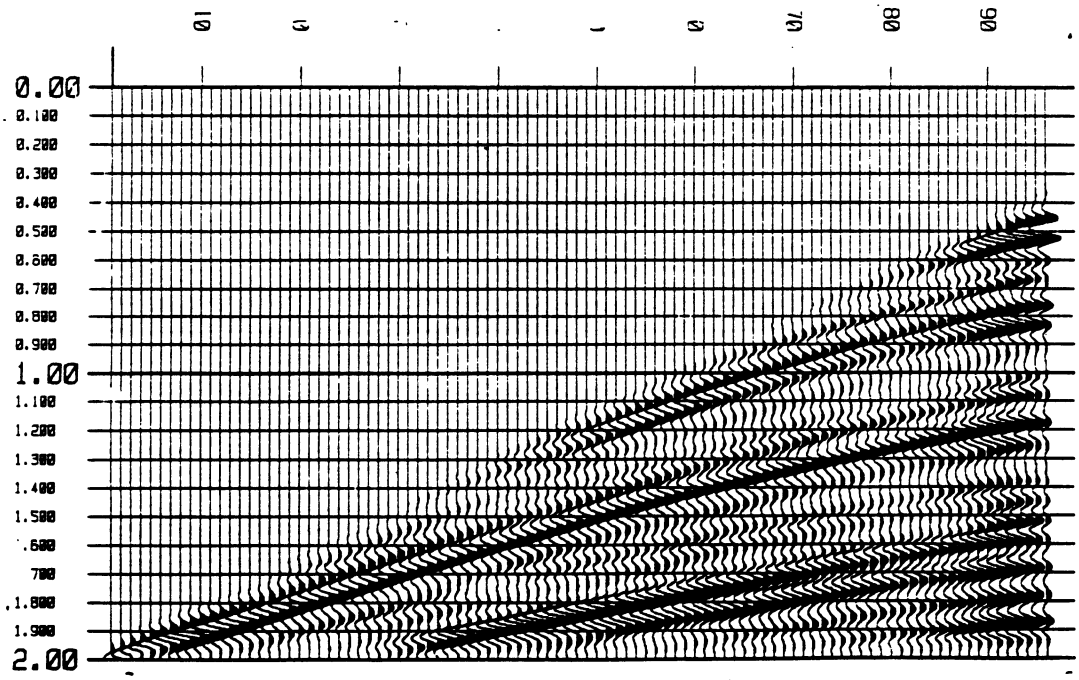

FIGURE 9. Prediction of shot record 1, based on second steepest descent iteration for velocity (FIGURE 6, upper curve) and reflectivity. Should be compared to FIGURE 1 (b).

distinction between shallow (overcorrected) and deeper (undercorrected) events; conceivably further iteration (or even a better choice of step at the second iteration) will tend to flatten further all of the events. Since the velocity corrections are layered, some residual moveout will remain in any case. The correct velocity is not layered, but increases substantially eastward, so that a layered approximation cannot completely flatten any of the common image gathers.

For completeness, we display the stacked migrated section with the final velocity estimate in Figure 8.

In Figure 9, we give the simulation of shot record 1 using the final velocity estimate and reflectivity from the second iteration as input. The discrepancies between this simulation and the input data (Figure 1 (b)) are presumably due to the causes mentioned before. The difference between this simulated data set and the input data is roughly 20% RMS.

Conclusion

We have explained a modified least-squares inversion procedure, differential semblance optimization (DSO). DSO is designed to overcome the difficulties experienced by standard least-squares inversion in estimating the long-wavelength components of velocity. In common with other automatic velocity estimation methods, DSO produces *kinematic* velocity models, which explain moveout in data while not necessarily admitting immediate geological interpretation. We stated the DSO variational principle in a form suitable for application to shot record inversion, and outlined the calculation of its gradient. The version of DSO explained here assumes constant density and absence of multiply reflected energy. These simplifications are incorrect for the Marmousi data set, but seem likely to allow at least a crude velocity estimate.

We carried out two steps of steepest descent minimization for DSO on a small subset of the Marmousi data (ten shot records from the west end of the line). We assumed a simple Ricker wavelet model of the source, and a layered velocity model (but laterally heterogeneous reflectivity). These latter restrictions are grossly wrong; nonetheless, DSO gave constructive velocity updates.

None of the restrictive assumptions made in this work are intrinsic to DSO. In future work we plan to allow density fluctuations and nontrivial wavelet updates, better near-surface modeling, and (most importantly) laterally heterogeneous velocity models. It will be important to employ larger data sets with better subsurface coverage, especially in order to constrain lateral velocity variations. The preliminary results reported here, while crude, appear to justify further investigation of the DSO approach.

Acknowledgement

I am pleased to thank Guy Chavent, James Carazzone, Patrick Lailly, Francois Chapel, Pierre Kolb, Francois Clement, and Albert Tarantola for their insights and advice, which materially advanced my work. Michael Lewis, Gang Bao, David Dobson, and Michael Pearlman helped with the programming. Alan Levander made available the Disco software package and graphics hardware used to produce the figures. Vivian Moser Choi typeset the manuscript. The work reported here was supported by the National Science Foundation, the Office of Naval Research, and the Air Force Office of Scientific Research.

References

Daubechies, I., 1988, Orthonormal bases of compactly supported wavelets, Comm. Pure Appl. Math. 51, 909–996.

Lailly, P., 1983, The seismic inverse problem as a sequence of before-stack migrations, in Conference on Inverse Scattering: Theory and Applications, J. Bednar et al., Eds. SIAM, Philadelphia, 206–220.

Mallat, S., 1987, Multiresolution approximation and wavelets, Tech. rep., Preprint GRASP Lab., Dept. of Computer and Information Science, U. Pennsylvania.

Meyer, Y., 1990, Ondelettes et Operateurs, Hermann, Paris.

Symes, W., 1988, Velocity inversion: a model problem from reflection seismology, Technical

Report 88-13, Department of Mathematical Sciences, Rice University, Houston, TX (SIAM J. Math. Anal., in press).

SYMES, W., 1990a, Velocity inversion: a case study in infinite-dimensional optimization, Math. Prog. 48, 71–102.

SYMES, W., 1990b, Multi-offset inversion by differential semblance optimization, preprint.

SYMES, W., AND CARAZZONE, J., 1989, Velocity inversion by coherency optimization, Technical Report 89-8, Department of Mathematical Sciences, Rice University, Houston, TX. (Proc. of Workshop in Geophysical Inversion, ed. J.B. Bednar, SIAM, in press).

SYMES, W., AND CARAZZONE, J., 1990, Velocity inversion by differential semblance optimization (Geophysics, in press).

TARANTOLA, A., 1984, Inversion of seismic reflection data in the acoustic approximation, Geophysics 49, 1259–1266.

APPENDIX: CALCULATION OF THE GRADIENT

Recall that

$$J_\sigma[v_b] = \tfrac{1}{2}\min_r\{\|S[v_b]r - S_{\text{data}}\|^2 + \sigma^2\|D_s r\|^2\} \tag{5}$$

where we have used the commonplace notations

$$\begin{aligned}\|u\|^2 &= \sum |u|^2 = \langle u, u\rangle\,;\\ \langle u, v\rangle &= \sum uv\,,\end{aligned}$$

the sums being over all indices (i.e. x_s, x_r, t in the first term, x_s, x, z in the second), and $D_s = \partial/\partial x_s$ or an approximation.

The minimizer of the quadratic form in r stated above is the solution of the normal equations

$$\{S[v_b]^T S[v_b] + \sigma^2 D_s^T D_s\}r = S[v_b]^T S_{\text{data}}\,. \tag{6}$$

On the other hand, first order perturbation theory applied to (5) gives for a perturbation δv_b

$$\begin{aligned}\delta J_\sigma &= \langle \delta S\cdot r, Sr - S_{\text{data}}\rangle\\ &\quad + \langle S\delta r, Sr - S_{\text{data}}\rangle + \sigma^2\langle D_s\delta r, D_s r\rangle\\ &= \langle \delta S\cdot r, Sr - S_{\text{data}}\rangle\\ &\quad + \langle \delta r, (S^TS + \sigma^s D_s^T D_s)r\\ &\qquad - S^T S_{\text{data}}\rangle\,.\end{aligned} \tag{7}$$

Now the second term in the right hand side of (7) vanishes because of the normal equations (6). On the other hand,

$$\delta S\cdot r = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\{S[v_b + \epsilon\delta v_b]r - S[v_b]r\}$$

while S is already the solution of a linearized problem:

$$S[v_b]r = \lim_{h\to 0}\frac{1}{h}\{S_0[v_b + hr] - S_0[v_b]\}$$

where S_0 is the solution operator of the *nonlinear* forward map. Combining these, we get

$$\begin{aligned}\delta S\cdot r &= \lim_{\epsilon,h\to 0}\frac{1}{h\epsilon}\{S_0[v_b + \epsilon\delta v_b + hr] - S_0[v_b + \epsilon\delta v_b]\\ &\qquad - S_0[v_b + hr] + S_0[v_b]\}\\ &= \lim_{h\to 0}\frac{1}{h}\{S[v_b + hr]\delta v_b - S[v_b]\delta v_b\}\end{aligned}$$

after interchanging the ϵ and h limits and letting $\epsilon \to 0$. (This is the argument for equality of mixed partial derivatives!) Since both v_b and δv_b are smooth, and the source is oscillatory, $S[v_b]\delta v_b$ should be negligible, and we obtain

$$\delta S\cdot r \cong \frac{1}{h}S[v_b + hr]\delta v_b \tag{8}$$

for suitably small $h > 0$.

Combining (7) and (8), we obtain the first equality in

$$\begin{aligned}\delta J_\sigma &\cong \langle\frac{1}{h}S[v_b + hr]\delta v_b\,,\ S[v_b]r - S_{\text{data}}\rangle\\ &= \langle\delta v_b\,, \frac{1}{h}S[v_b + hr]^T(S[v_b]r - S_{\text{data}})\rangle\\ &= \langle\delta v_b\,, \text{grad}J_\sigma\rangle\,.\end{aligned} \tag{9}$$

The second equality in (9) defines the adjoint or transpose operator S^T, while the third states the definition of the gradient. Since (9) must hold for all δv_b, the gradient formula (4) follows from the last two lines in (9).

In practice, we cannot actually solve the normal equations (6) but only produce an approximation to r. Then the effect of the remaining error on the computed gradient must be understood. For a treatment of one type of approximation (via Krylov subspace projection) see Symes 1990b.