Multioffset Inversion by Differential Semblance Optimization

William Symes

CRPC-TR90177
December, 1990

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

Multioffset Inversion by Differential Semblance Optimization

W.W. Symes
Department of Mathematical Sciences
Rice University
Houston, TX 77251-1892

Abstract

A still-unresolved problem in reflection seismology is the estimation of wave velocities directly from waveform data, bypassing expensive manual steps such as event-picking in before-stack data. This paper suggests a waveform inversion algorithm, *differential semblance optimiztion* for shot-gather data, designed to yield accurate veloctiy estimates even in the absence of accurate *a priori* information. The algorithm described here is a variant of *least-squares inversion by gradient optimization*, which has proven incapable of estimating large-scale velocity features. We explain the obstacles to success of least-squares velocity inversion, and offer both theoretical and numerical evidence that our modification should overcome these obstacles. An appendix describes our approach to accurate calculation of the gradient of the modified cost functional.

1 Introduction

Least-squares (or more generally, least-error or maximum-likelihood) inversion has been advocated over the last decade as a general approach to the inversion of reflection seismograms, for several reasons. It is capable of reflecting directly almost any physics of seismic wave propagation; it can be

modified to incorporate nonseismic constraints; and it possesses an elegant statistical justification (Tarantola, 1987).

Unfortunately, the promise of least squares inversion using gradient (quasi-Newton) optimization has been largely unrealized, principally because of its failure to yield large-scale velocity trends. This failure, first noticed in numerical experiments (see Kolb *et al.* 1986, Gauthier *et al.* 1986), has attracted widespread comment. For an attempt at an explanation, see Santosa and Symes 1989.

The purpose of this note is to present a modification of least-squares inversion which appears to retain some of its advantages while yielding velocity trend estimates as well. The main step in this modification is the inclusion in the least-squares principle of a differential measure of *event semblance*, so we have called this modification the *differential semblance method.*

The *differential semblance* method actually encompasses a number of algorithms. In previous papers (Symes 1988, 1990a, Symes and Carazzone 1989, 1990) we have explored algorithms in this class appropriate for plane-wave data and layered media. We have given both theoretical and numerical evidence that the plane-wave differential semblance method yields efficient and accurate velocity and reflectivity estimates. The numerical evidence includes successful treatment of field data sets.

In the following sections we will formulate a simple acoustic model of reflection seismology; explain the difficulty faced by conventional least squares inversion in context; present a version of the differential semblance method appropriate for shot-record data; explain how the differential semblance method avoids the least-squares pitfalls; and show some preliminary numerical evidence of its effectiveness. This paper aims only to introduce the differential semblance method in the shot-record context. Some mathematical details of the gradient calculation are given in an Appendix; a more detailed treatment of mathematical and numerical aspects will appear elsewhere. A first crude attempt at velocity inversion for a complex model appears in Symes 1990b.

2 The Acoustic Model

The constant-density linear acoustic model connects the (spatially varying) sound velocity $v(x)$, the pressure field $p(x,t)$, and a body force divergence

("source") $f(x,t)$ through the acoustic wave equation

$$\frac{1}{v^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = f\ .$$

We assume that f is transient and known and p is causal: $p \equiv 0$ for $t << 0$. We also assume (for simplicity) that the fluid occupies the half-space $\{z > 0\}$, and that the surface $\{z_3 = 0\}$ is free, i.e. $p \equiv 0$ there.

It is very important in what follows that f depends on another parameter x_s, so we write: $f(x, x_s)$. Thus p depends on x_s as well. For common-shot seismic data x_s is naturally the source position; however x_s could also be construed as offset for application to common-offset data, or as slowness for application to plane wave data.

The seismogram is the sampling of the pressure field at a set $\{x_r\}$ of receiver points. It depends functionally on the velocity distribution v. We base this discussion on the well-known perturbational approximation to the seismogram, in which v is split as $v = v_b + v_r$, with v_b a smooth background velocity and v_r a rough or oscillatory perturbation. Using regular first-order perturbation theory we write

$$S = S_b + S_r$$

where S_b is the background seismogram and S_r is the perturbation due to v_r. If v_b is sufficiently smooth, which we assume, then S_b consists of the direct wave, plus possibly refractions. We limit our attention to reflections here; so we assume that S_b has been subtracted or muted out (a nontrivial step in practice!) and identify S with S_r.

S is thus the sampling at the receiver locations of the pressure field perturbation δp, which solves

$$\frac{1}{v_b^2}\frac{\partial^2 \delta p}{\partial t^2} - \nabla^2 \delta p = \frac{2v_r}{v_b}\nabla^2 p$$

plus appropriate side conditions. Note the appearance of the *reflectivity* $r = v_r/v_b$; in the sequel we shall use r rather than v_r to represent the rough part of the model.

As is well-known (Cohen and Bleistein 1977, Beylkin 1985, Rakesh 1988, Percell 1989) S can be approximated rather effectively as an oscillatory integral of the form

$$S[v_b]r(x_s; x_r, t)$$

$$\cong \quad f * \int d\xi A[v_b](x_s, x_r, t, \xi) e^{i\phi[v_b](x_s, x_r, t, \xi)} \cdot \hat{r}(\xi) \ .$$

The notation is chosen to emphasize the following points:

(1) The seismogram is a function of the source parameter x_s, the receiver coordinate x_r, and time t;

(2) The *amplitude* or *symbol* A and the *phase* ϕ also depend on a wave vector ξ of the same dimensionality as the space coordinates.

(3) A and ϕ depend further on a source coordinate x and are convolved in x and t against the source distribution $f(x, t, x_s)$.

(4) The seismogram S, the amplitude A, and the phase function ϕ depend functionally on v_b. The phase function ϕ is related to the travel time of certain reflected rays; see Percell 1989 for details.

(5) The seismogram S depends linearly on the reflectivity r.

For space dimension n, the symbol A behaves for large $|\xi|$ like

$$A_0(x_s, x_r, t, \xi/|\xi|)|\xi|^{\frac{n-1}{2}}$$

for a suitable smooth function A_0 (the "principal symbol"). A_0 is non-zero over a sector in $\xi/|\xi|$ determined by the ray geometry (hence by v_s). The phase function of ϕ is positively homogeneous of degree 1 in ξ.

3 Mathematical and Scientific Difficulties of Least Squares Inversion

With the conventions established so far, we can state a simple version of the least-squares inversion problem:

Find v_b, r to minimize

$$J_{LS}\left[v_b, r; S_{\text{data}}\right] = \frac{1}{2} \int\int\int dx_s dx_r dt \left| S[v_b] \cdot r - S_{\text{data}} \right|^2$$

Here we understand the integral sign to mean integral or sum, as appropriate. To avoid writing an excessive number of integral signs in the sequel, we introduce the standard notations for the L^2 inner product and norm:

$$\begin{aligned} \langle \psi, \phi \rangle &= \int \bar{\psi} \phi \\ \|\psi\| &= \langle \psi, \psi \rangle^{\frac{1}{2}} \end{aligned}$$

where the integration variables are understood from context, and replaced by scaled sums in case the functions ψ, ϕ are discretely sampled in one or more variables.

In this notation,

$$J_{LS}[v_b, r; S_{\text{data}}] = \frac{1}{2} \| S[v_b] r - S_{\text{data}} \|^2 .$$

In a typical reflection seismic model in 2D, v_b might be represented by a few tens to a few hundreds of parameters, while r requires perhaps $10^5 - 10^6$ parameters for a useful degree of resolution. Thus the least-squares problem is computationally very large, and efficient minimization algorithms are required. By far the most efficient numerical optimization techniques are the descent methods related to Newton's method — when they work. These iterations take steps predicted by the linearized model/data relation so rely for their effectiveness on a close relation between the cost function and its quadratic approximation. Accordingly, we now examine (somewhat formally) the response of S to perturbations in v_b and r.

From the oscillatory integral expression above the perturbation of S due to a change δv_b in v_b is

$$\delta_{v_b} S \cdot r = \int d\xi (i \delta\phi \cdot A + \delta A) e^{i\phi} \hat{r} .$$

This is an oscillatory integral of the same form as that approximating S, with a different symbol. In fact $\delta\phi$ is also homogeneous of order 1, exactly as is ϕ. Therefore the symbol (amplitude) in the above integral grows as $|\xi|^{\frac{n+1}{2}}$ as $|\xi| \to \infty$, i.e. at a more rapid rate than A. It follows that for at least some oscillatory r, smooth δv_b

$$\| \delta_{v_b} S \| >> \| S \| .$$

Taking this reasoning one step further, one sees immediately that

$$\| \delta^2_{v_b} S \| >> \| \delta_{v_b} S \| ,$$

that is, that S is very nonlinear in v_b.

On the other hand S is linear in r. Since

$$\begin{aligned}
\delta_{(v_b,r)} J_{LS} &= \langle \delta_{(v_b,r)}[S[v_b]\cdot r], S[v_b]r - S_{\text{data}} \rangle \\
&= \langle \delta_{v_b} S[v_b]r + S[v_b]\delta r, S[v_b]r - S_{\text{data}} \rangle \\
\delta^2_{(v_b,r)} J_{LS} &= \|\delta_{v_b} S[v_b]r + S[v_b]\delta r\|^2 \\
&\quad + \langle \delta^2_{v_b} S[v_b]r + 2\delta_{v_b} S[v_b]\delta r, S[v_b]r - S_{\text{data}} \rangle
\end{aligned}$$

one might well expect that

$$\left|\delta^2_{v_b} J_{LS}\right| >> \left|\delta^2_r J_{LS}\right|$$

and this is indeed the case for some oscillatory $r, \delta r$ of the same magnitude and smooth $v_b, \delta v_b$. Thus the Hessian is extremely ill-conditioned.

Moreover, it also follows that the growth rate of J_{LS} as one moves v_b away from the minimizer must be many times the overall size of J_{LS} itself. Therefore the growth cannot be sustained over a large change in v_b, and J_{LS} *saturates*. Consequently J_{LS} tends to be very non-convex, with a very small region of convexity near the global optimum model. See Symes and Carazzone 1989, Figure 4 for an actual picture of J_{LS} illustrating these features.

The highly non-quadratic nature of J_{LS} explains the great difficulty of recovery of v_b by least-squares inversion reported frequently in the literature (Gauthier *et al.* 1986, Kolb *et al.* 1986, Mora 1987, for example — see also Santosa and Symes 1989 for an extensive explanation of these ideas in the context of a simpler model). Both numerical obstacles and difficulties of principle arise. First, one can say with confidence that extraction of the global minimizer of J_{LS} by means of least-squares inversion and local, Newton-type optimization is *impossible*, unless the initial estimate if v_b is quite accurate. The necessary degree of accuracy in the initial estimate appears very difficult to predict.

Random or systematic search has been suggested as alternative to gradient-based methods (Cao *et al.* 1990, Tarantola 1987, Tarantola *et al.* 1990, and many references cited there). Such methods may work well when the background velocity may be represented by a few parameters in a known way. In general, severely parsimonious parameterization is likely to introduce unjustified bias, that is, to fail to sample the model space sufficiently to well-approximate the optimal v_b. On the other hand, refined parameterization generates extremely large search tasks.

It has also been suggested that *all* of the local minima of J_{LS} should be viewed collectively as the "solution" of the estimation problem. If one employs local optimization techniques to minimize J_{LS}, one is *de facto* forced to this point of view. Tarantola 1987, Tarantola *et al.* 1990 has argued that the various local minima of J_{LS} (perhaps biased toward an "a priori" model by a penalty term) represent statistically important models, relatively more likely than the surrounding models — and in fact $\exp\{-J_{LS}\}$, regarded as an un-normalized probability density, is really the "solution" of the inversion problem. The viability of this "standard inverse theory" viewpoint is delicate, and in our opinion unresolved. In practice, only simple estimators of the "solution" density are sought, e.g. maximum likelihood points, leading directly back to output least-squares. For highly nonlinear problems like velocity estimation, however, even the density itself is *extremely* sensitive to the noise model ("state of *a priori* information"), which is in turn selected almost entirely on the basis of heuristics and mathematical convenience.

In sum, estimation of v_b *via* the least-squares principle is unlikely to yield useful results in general, or reliable inversion methods.

4 The Differential Semblance Method

Our resolution of the difficulties outlined in the preceding paragraphs begins with two observations:

(i) For *fixed* v_b, J_{LS} is perfectly convex — in fact, quadratic!

(ii) If the set of shot parameter values $\{x_s\}$ reduces to a *singleton*, e.g. only one point source record is used, the minimum value of J_{LS} is essentially independent of v_b.

That is, the inversion of a single shot record is feasible, and constrains *only* r, *not* v_b. Since this task is practical, it suggests the expedient of viewing r as a function of the shot parameter x_s

$$r = r(x, x_s) .$$

Of course, if S_{data} is noise free,

$$S_{\text{data}} = S[v_b^*]r^*$$

then $r(x, x_s) \equiv r^*(x)$ is amongst the minimizers of

$$\left\| S[v_b^*] r(\cdot, x_s) - S_{\text{data}}(\cdot, x_s) \right\|^2$$

and has the addition property of *coherence*, or independence of x_s, which we can express as

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

Only coherent reflectivity estimates have any ultimate meaning, since there is only one earth!

The above two conditions can be combined into a single cost functional, for instance:

$$\| S[v_b] r - S_{\text{data}} \|^2 + \sigma^2 \| \partial r / \partial x_s \|^2$$

where r is now allowed to depend explicitly on x_s — with such dependence penalized by the second term, weighted by a parameter σ^2.

This functional is quadratic in r, so the minimization with respect to r presents no difficulties, in principle. On the other hand, as a functional of *both* v_b and r, it is still quite non-convex, for the same reasons as is J_{LS}. Together these two observations suggest *elimination* of r: that is, we define a functional of v_b only by

$$\boxed{J_\sigma \left[v_b ; S_{\text{data}} \right] = \min_r \frac{1}{2} \left\{ \left\| S[v_b] \cdot r - S_{\text{data}} \right\|^2 + \sigma^2 \left\| \frac{\partial r}{\partial x_s} \right\|^2 \right\} \, .}$$

It is a remarkable fact that this functional is smooth — in fact, nearly quadratic — in its dependence on v_b, despite its rather close relation with the least squares functional! Minimization of J_σ over a smooth class of background velocities v_b is the *differential semblance optimization problem.* Note that for noise-free data $S_{\text{data}} = S[v_b^*] r^*$, J_σ attains the value 0 for $v_b = v_b^*$, which is clearly its global minimum, and that this minimum is reached by setting $r = r^*$ on the right-hand side. That is, the global minimum is achieved at the correct velocity — and, implicitly, at the correct reflectivity. Since J_σ is smooth, it is necessarily convex near a consistent global minimizer. That is, if the data is noise-free, then J_σ is necessarily convex near v_b^*, and remains convex when S_{data} is perturbed by small amounts of noise. We conjecture

that J_σ is strongly convex for near-consistent data S_{data} and proper choice of σ^2, over a large subset of background velocity models. We will present some numerical evidence for this later, and have given a proof for the related (but simpler) plane/wave layered medium case (see Symes 1988, 1990; Symes and Carazzone 1989).

In the remainder of this section, we will outline the reasons for the smoothness of J_σ, and our reasons for thinking that J_σ might be minimized quite efficiently. We give only the formal skeletons of arguments here; precise statements and proofs will be presented elsewhere. Since we propose to use Newton-type ("gradient") optimization methods to minimize J_σ, accurate calculation of the gradient is important. The derivation of an accurate technique for gradient calculation is relegated to the Appendix, as it is somewhat more technical in nature than the following material.

Before starting we take care of a few technical details. The first is that the normal operator

$$S^T S$$

is a *pseudodifferential operator of order* $n-1$ *in dimension* $n(=1, 2,$ *or* $3)$ if the source is impulsive, $f(x,t) = \delta(x - x_s)\delta(t)$, under some ray-geometric restrictions (no caustics in the incident wavefront). That is, $S^T S$ is given by an oscillatory integral of the form

$$S^T S\, r(x_s, x) = \int d\xi\; b(x_s, x, \xi) e^{ix\cdot\xi} \hat{r}(x_s, \xi)$$

where $b \sim b_0(x_s, x, \xi/|\xi|)|\xi|^{n-1}$ as $|\xi| \to \infty$, and $\hat{r}(x_s, \xi)$ is the Fourier transform in x of $r(x_s, x)$. In fact $S^T S$ is a family of pseudodifferential operators in x, parameterized by x_s, of order $n-1$. It is a slight technical headache that such a family of pseudodifferential operators in x is *not* a pseudodifferential operator in x *and* x_s; however this is not an essential complication (e.g., Taylor 1975, Appendix) and we shall ignore it here. This pseudodifferential representation of $S^T S$ is an immediate consequence of the FIO calculus (Duistermaat 1973), and is mentioned explicitly in Beylkin 1985, Symes 1986, and Rakesh 1988 for example. As shown in Percell's thesis 1989, this conclusion is *false* when caustics are present in the incident wave-front — a generic occurrence in heterogeneous media. It is possible to recover the pseudodifferential nature of $S^T S$ by modifying the definition of S by a microlocal (phase space) suppression or muting of the reflectivity r. Without going into

details, we assume that this has been done. Then S^TS operator is *elliptic*, i.e. acts as an invertible Fourier multiplier at high spatial frequencies, over a conic sector of wave vectors (the "reflection aperture") determined by the relative positions of sources and receivers and the ray geometry of the background velocity field. Outside of the reflection aperture, which varies with location in the subsurface, S^TS suppresses high-frequency components (these correspond to *off-cable reflections*). That is, the symbol b satisfies

$$\begin{aligned} b_0(x_s, x, \xi/|\xi|) &> 0 \text{ within the reflection aperture} \\ b_0(x_s, x, \xi/|\xi|) &= 0 \text{ outside of the reflection aperture .} \end{aligned}$$

The theory of pseudodifferential operators is an indispensible tool in developing a precise and effective understanding of the reflection seismic inverse problem. Good references are Taylor 1980 and Hörmander 1983. For the following discussion, the reader needs to understand that

(a) A pseudodifferential operator of order k (roughly) scales the Fourier transform of the function to which it is appliedy by $|\xi|^k$. That is, it "acts" like a k^{th} derivative operator. Here k may be negative or even non-integral. A pseudodifferential operator of order zero does not alter the rate of growth of the Fourier components, so acts locally as a bounded operator on L^2.

(b) If the symbol b of a pseudodifferential operator

$$B[v]u(x) = \int d\xi b(v, x, \xi) e^{i\xi \cdot x} \hat{u}(\xi)$$

depends smoothly on parameters v, then so does the operator $B[v]$, and all derivatives with respect to v are operators of the same order. That is, differentiation with respect to parameters does not result in increased weighting of high-frequency components. Note the contrast with the behaviour of oscillatory integrals of the type defining S, as presented in the last section, in which the phase also depends on parameters. This contrast underlies the entire theory developed in these pages.

Because of the aperture limitation mentioned above, the high-frequency components of r outside the inversion aperture must be constrained *a priori* in solving equations involving S^TS. To accomplish this goal in a well-scaled

way, we first modify the definition of S: we assume that the source has point-support, and in its time dependence is a low frequency perturbation of the $(\frac{3-n}{2})$-th derivative of $\delta(t)$: thus

$$f(x,t) = \text{const.} \quad \begin{cases} \delta(x-x_s)t_+^{-\frac{3}{2}}\,, & n=2 \\ \delta(x-x_s)\delta(t)\,, & n=3 \end{cases} + f_0(x,t)$$

(the distribution $t_+^{-\lambda}$ is defined in Gel'fand and Shilov 1958, for example), where f_0 is a smooth function. This amounts to assuming that f, while bandlimited below, behaves as $t_+^{-\frac{3}{2}}$ $(n=2)$ or $\delta(t)$ $(n=3)$ across the upper part of the passband of the seismic signals. Practically, this assumption is realized by preprocessing the data to re-scale it in the frequency domain.

With this modification, S^TS is a *pseudodifferential operator* of order 2, independently of the dimension.

We chose a regularizing operator W, pseudodifferential of order 2 in x and depending parametrically on x_s, so that

$$S^TS + \lambda^2 W$$

is *uniformly elliptic* for each x_s as long as $\lambda^2 > 0$. A simple choice is

$$W = I - \nabla_x^2\,.$$

This choice is suboptimal, as it also affects the components within the reflection aperture, but for small λ^2 this is probably of little consequence. It will be important in the sequel to write $W = C^TC$, with C a pseudodifferential operator of order 1. This is certainly possible for the simple choice just given with $C = (I - \nabla_x^2)^{\frac{1}{2}}$.

The (regularized) differential semblance functional J_σ is defined by minimizing over r the (regularized) quadratic

$$\frac{1}{2}\left\{ \left\| Sr - S_{\text{data}} \right\|^2 + \lambda^2 \langle r, Wr\rangle + \sigma^2 \left\| \frac{\partial r}{\partial x_s} \right\|^2 \right\}\,.$$

A minimizer of the above quadratic is a solution of the normal equations

$$Nr \;:=\; [S^TS + \lambda^2 W - \sigma^2\,\partial^2/\partial x_s^2]r = S^TS_{\text{data}}\,.$$

The operator N is (essentially) an elliptic pseudodifferential operator of order 2 in x, x_s. Standard techniques show that N is invertible, under reasonable restrictions on r, and that r depends stably on S_{data} in suitable norms.

Since $S = S[v_b]$, the solution of the normal equations also depends on v_b: $r = r[v_b, S_{\text{data}}]$ also. The dependence of r on v_b is quite erratic — this is another consequence of our analysis of the least-squares problem in the last section. In any case we can write $r = N^{-1}S^T S_{\text{data}}$ and obtain

$$\begin{aligned} J_{CM} &= \frac{1}{2}\langle S^T S_{\text{data}}, (I - N^{-1})S^T S_{\text{data}}\rangle \\ &= \frac{1}{2}\langle S_{\text{data}},\ S(I - N^{-1})S^T S_{\text{data}}\rangle\ . \end{aligned}$$

The operator $S(I - N^{-1})S^T$ is — essentially — pseudodifferential, for the same reasons as is the normal operator. Its symbol is a smooth function of v_b, whence follows the smoothness of J_σ, by virtue of the property of pseudodifferential operators depending on a parameter, noted earlier.

Note that the quantity $S^T S_{\text{data}}$ is *not* a smooth functional of v_b: it is given by an oscillatory integral with phase depending nontrivially on v_b. The same is true of $N^{-1}S^T S$. We can write

$$\begin{aligned} \langle S^T S_{\text{data}}, N^{-1}S^T S_{\text{data}}\rangle &= \frac{1}{4}\left\{\|(I + N^{-1})S^T S_{\text{data}}\|^2\right. \\ &\quad \left. - \|(I - N^{-1})S^T S_{\text{data}}\|^2\right\}\ . \end{aligned}$$

Thus the smoothness of J_σ is an amusing instance of the observation that a non-smooth function taking values in a Hilbert space may have a smooth norm.

We assume that we apply to the normal equations an algorithm yielding, after a finite number of steps, an estimate of r. We will consider only Krylov space methods such as conjugate gradient iteration. The normal equations are very similar in nature to the Laplace equation, and in particular preconditioning is required to produce rapid convergence of the Krylov sequence. Preconditioning amounts to replacing the normal equations by the *preconditioned normal equations*

$$Mr = S^* S_{\text{data}}$$

where $M = GN$, $S^* = GS^T$, and G is the solution operator of the Laplace equation in (x, x_s) with suitable boundary conditions. (In fact, S^* is the

adjoint of S with respect to a so-called Sobolev scalar product.) This is the system solved in the numerical work reported in the next section. The operator M is now of order zero, i.e. does not enhance high-frequency components (nor does it suppress them). In a qualitative sense at least, M is well-conditioned.

5 Numerical Exploration

In this section we present some preliminary numerical results, in which we have evaluated the coherency optimization functional over line segments in the space of background velocity models, to display directly its convex nature.

As is evident from the expression

$$\begin{aligned} J_\sigma &= \|Sr - S_{\text{data}}\|^2 + \lambda^2\langle r, Wr\rangle + \sigma^2\|\partial r/\partial x_s\|^2 \\ N &= S^TS + \lambda^2 W - \sigma^2\,\partial^2/\partial x_s^2 \\ M &= GN \end{aligned}$$

it is necessary to approximate

(i) the seismogram ("forward map") S;

(ii) its adjoint S^T;

(iii) the regularizing operator W;

(iv) $\partial/\partial x_s$;

(v) the solution operator G of the Laplace equation in x and x_s, with suitable boundary conditions.

We shall examine each of these approximations briefly. We were incautious in the work reported here, and set $\lambda^2 = 0$. Thus step (iii) (the regularizing operator) was neglected.

We computed 2D shot-record seismograms of duration 1.6 s, over a model 3 km deep and 7.5 km wide. Velocities ranged from 1.5 to 2.5 km/s. Shot locations were confined to the interval 3.5–4.5 km from the left edge of the model. The cable stretched from 150 to 1950 m to the right of each shot, with 60 receivers (modeled as points) spaced 30 m apart. The shot depth

was 8 m, and the shot was modeled as a point source with Ricker wavelet time dependence. The receiver depth was 15 m.

With this geometry, no reflections from the edge of the model could arrive back at the cable within the time gate of the simulation. Accordingly we used Dirichlet boundary conditions on all four sides of the computational domain.

In order to keep the computational cost low for these exploratory experiments, we chose to use a very low-frequency source, with peak frequency at 10 Hz. We could then use sample rates $\Delta t = 4$ ms, $\Delta x = \Delta z = 15$ m, and obtained stable and reasonably nondispersive simulation with the commonplace second order centered difference scheme.

The adjoint S^T was computed with the adjoint state technique. The x_s derivative was approximated by the midpoint rule. The boundary conditions for the Laplace operator were periodic in x and z and Neumann in x_s. The solution operator G was implemented *via* Fourier transform in x and z, and tridiagonal solve in x_s.

Computational modules written in FORTRAN incorporating these choices were coupled to a conjugate residual procedure for solution of the normal equations. The calculations were performed on a Stardent Titan 1500 Series superworkstation. The Titan spent the bulk of its time in the finite difference solution of the wave equation, achieving roughly 15 M flops in parallel/vector loops.

We present the results of two experiments, each involving 10 shots. The first shot was located at $x_s = 3600$ m and the shot interval was 100 m. Each reflectivity was estimated via 10 conjugate residual iterations. The evaluation of J_σ for each model required approximately 4 hours of Stardent CPU time.

For both experiments we sample J_σ at the sequence of velocity models shown in Figure 1. The "target" velocity v_1 rises linearly from 1.5 km/s at $z = 0$ m to 2 km/s at $z = 500$ m, and was used to generate the data for both experiments. The "initial" velocity v_0 is constant, $\equiv 1.5$ km/s. The other velocities are convex combinations of these two,

$$\begin{aligned} v &= (1-h)v_0 + hv_1 \\ h &= 0,\ 0.25,\ 0.5,\ 0.75,\ 1.0,\ 1.25\ . \end{aligned}$$

In both experiments σ^2 was adjusted so that the two terms in the cost function would have roughly similar magnitudes. This was done by comparing

the mean-square residual term at $r = 0$ (i.e. the data mean-square) to the incoherence term after ten CR iterations with $\sigma^2 = 0$, in both cases with $v = v_0$. Then σ^2 was chosen to equilibrate (roughly) these two values. In both experiments this procedure resulted in the choice $\sigma^2 = 1000$.

In the first experiment, the target reflectivity was a single flat reflector located at $z = 750$ m. That is, r is a step function jumping from $r = 0$ to $r = r^*$ at $z = 750$ m. Figure 2 shows the cost function and its two components. The value at the target is nonzero, suggesting that the reflectivity is not accurately reconstructed in 10 CR iterations. Nonetheless the curve is smooth and convex and has the target model at its minimum, suggesting that this type and degree of inaccuracy will not prevent effective estimation of velocities. Common x gathers ("image gathers") for the reflectivity estimates at v_0 and v_1 are offered in Figures 3a and 3b as further confirmation that the computed reflectivities are sufficiently accurate to detect moveout error. These gathers correspond to a well-illuminated location below $x = 4400$ m. Further from the center of the line, CR was not able to overcome aperture truncation effects, which account for most of the remaining incoherency.

The second set of experiments was based on the reflectivity model of Figure 4. In this model a sequence of flat-lying beds are truncated by underlying curved beds dipping to the left. A pair of shot gathers from near the center of the line appear as Figure 5. Figure 6 displays the cost functional, which is once again without any secondary minima. Figure 7 gives the common x gathers for a number of locations near the center of the line. The residual aperture effects are evident, but also the dip of the principal reflector, and the flatness of the overlying layer, are clearly reproduced. Figure 7 uses a common-x sort of the reflectivity estimated after ten conjugate residual iterations with the target velocity v_1.

Discussion and Conclusion

Let us state precisely the sense in which we hope to solve the velocity inversion problem *via* differential semblance optimization. First, for noise-free data, as noted above, the "exact" model is amongst the minima of J_σ. Thus the velocity estimator obtained by minimization of J_σ is consistent with the model. Second, if our convexity conjecture is correct, then the velocity estimate is a *stable* function of the data; in particular it is well-defined.

Third, because of the smoothness (proved) and convexity (conjectured) of J_σ, rapidly convergent Newon-type iteration will produce the estimator: it is efficiently computable.

We believe that this triple of properties just stated (consistency, stability, and computability of the velocity estimator) form a satisfactory substitute for the statistical inverse theory (Tarantola 1987). While the latter appears not to apply to velocity estimation, the former are *mathematical* properties, which are effective when the noise is small in a well defined sense. Smallness of the noise, both random and model-error, can be verified — by solving the minimization problem! In contrast, the statistical assumptions underlying standard inverse theory are unverifiable, probably wrong, and force the investigator into computationally intractable problems.

We have described a modified least-squares principle and velocity and reflectivity estimation. We have explained how this principle avoids the main pitfalls of least-squares inversion, and described an accurate calculation of the gradient of a computable approximating family of functionals. Finally we have offered some preliminary numerical evidence that this functional is smooth and has no secondary minima over a large domain in model space. Should this conjecture hold, quasi-Newton methods should be adequate to find the global minimum, at a kinematically correct velocity model. We hope to report the implementation of such a quasi-Newton velocity estimator in a future article.

Appendix. Accurate Calculation of the Differential Semblance Gradient

Since the differential semblance

$$J_\sigma[v_b] = \min_r \frac{1}{2}\left\{\|S[v_b]r - S_{\text{data}}\|^2 + \lambda\langle r, Wr\rangle + \sigma^2\|\partial r/\partial x_s\|^2\right\}$$

is itself the solution of a minimization problem, the calculation of its gradient is not straightforward. The complication, to which this appendix is devoted, arises from the infeasibility of the exact minimization of the right-hand side. Instead of J_σ, we can have computational access to

$$\tilde{J}_\sigma[v_b] = \frac{1}{2}\left\{\|S[v_b]\tilde{r} - S_{\text{data}}\|^2 + \lambda^2\langle \tilde{r}, W\tilde{r}\rangle + \sigma^2\|\partial\tilde{r}/\partial x_s\|^2\right\}$$

where $\tilde{r}$ is an approximate solution of the normal equations:

$$\begin{aligned} N[v_b]r &= (S[v_b]^T S[v_b] + V)r \\ &= S[v_b]^T S_{\text{data}} \\ \text{where}\quad V &= \lambda^2 W - \sigma^2\,\partial^2/\partial x_s^2\,. \end{aligned}$$

As mentioned in the text, we have chosen to produce $\tilde{r}$ *via* a fixed, finite number of steps of a Krylov subspace iteration, e.g. the conjugate gradient algorithm, applied to the preconditioned normal equations

$$Mr = S^* S_{\text{data}}\,, \quad M = GN\,, \quad S^* = GS^T$$

where G is the solution operator of the Laplace equation in x and x_s, with suitable boundary conditions. Such an algorithm produces $\tilde{r}$ in the form

$$\tilde{r} = P(M)S^* S_{\text{data}}$$

where P is a polynomial.

In order to state the sense in which convergence $\tilde{r} \to r$ occurs, introduce the Laplace operator H (in x and x_s) with the boundary conditions previously mentioned: thus $GH = HG =$ identity. Also, e is the normal residual:

$$e = M\tilde{r} - S^* S_{\text{data}}$$

which measures directly the error in the solution of the normal equations. Then S^* is the adjoint of S in the sense of the form defined by

$$\|r\|_1^2 = \langle r, Hr \rangle$$

(a version of the Sobolev 1-norm — we assume that the boundary conditions on r are chosen so that this is indeed a norm). Sufficiently many steps of a Krylov space method will drive both the error

$$\|\tilde{r} - r\|_1$$

and the normal residual

$$\|M\tilde{r} - S^* S_{\text{data}}\|_1 = \|e\|_1$$

below any prescribed tolerance.

We will first investigate the approximation of the directional derivative δJ_σ in the direction δv_b by $\delta \tilde{J}_\sigma$. Note that

$$\begin{aligned} \delta J_\sigma &= \langle \delta S \cdot r, Sr = S_{\text{data}} \rangle \\ &\quad + \langle \delta r, (S^T S + V)\, r - S^T S_{\text{data}} \rangle \\ &= \langle \delta S \cdot r, Sr - S_{\text{data}} \rangle \\ &\quad + \delta r, Nr - S^T S_{\text{data}} \rangle \,. \end{aligned}$$

Here δr is the (implicit) derivative of $r = r[v_b, S_{\text{data}}]$ with respect to v_b. This could be computed by differentiating the normal equations, but fortunately this effort is unnecessary: because of the normal equations the second term drops out. Thus

$$\delta J_\sigma = \langle \delta S r, Sr - S_{\text{data}} \rangle \,.$$

Similarly,

$$\begin{aligned} \delta \tilde{J}_\sigma &= \langle \delta S \cdot \tilde{r}, S\tilde{r} - S_{\text{data}} \rangle + \langle \delta \tilde{r}, N\tilde{r} - S^T S_{\text{data}} \rangle \\ &= \langle \delta S \cdot \tilde{r}, S\tilde{r} - S_{\text{data}} \rangle + \langle \delta \tilde{r}, He \rangle \end{aligned}$$

but now the second term does not drop out, since in general $e \neq 0$. We can write

$$\delta \tilde{r} = \delta P G S^T S_{\text{data}} + P G \delta S^T S_{\text{data}} \,.$$

Now N and G are formally self-adjoint, i.e. $N = N^T$, etc. Since P is a polynomial in $M = GN$, PG is also formally self-adjoint. Thus

$$\begin{aligned}\delta\tilde{J}_\sigma &= \langle \delta S \cdot \tilde{r}, S\tilde{r} - S_{\text{data}} \rangle \\ &\quad + \langle S_{\text{data}}, SG\delta P^T He \rangle \\ &\quad + \langle S_{\text{data}}, \delta S \cdot Pe \rangle \, .\end{aligned}$$

To go further, we must introduce some properties of S. First, as one might guess from the fact that S^TS is or order two (after the normalization introduced in the text), S itself is of order one in the sense that

$$\|Su\| \leq C\|u\|_1$$

for a suitable constant $C > 0$.

Next we introduce the factorization

$$\delta_{v_b} Sr = \int (\delta A + i\delta\phi \cdot A) e^{i\phi} r = SQ_1 r$$

which follows from the calculus of Fourier integral operators (e.g. Duistermaat 1973). Here $Q_1[v_b, \delta v_b]$ is a pseudodifferential operator of order 1 (i.e. its symbol grows like $|\xi|$ for large $|\xi|$) and depends smoothly on v_b, linearly on δv_b. Moreover, Q_1 is *essentially skew-adjoint*: there is another pseudodifferential operator Q_0 of order zero (i.e. whose symbol q_0 is bounded as $|\xi| \to \infty$) so that

$$Q_1 + Q_1^T = Q_0 \, .$$

Using this factorization, we can write

$$\begin{aligned}\delta J_\sigma &= \langle \delta S \cdot r, Sr - S_{\text{data}} \rangle \\ &= \langle Q_1 r, S^T(Sr - S_{\text{data}}) \rangle \\ &= \langle Q_1 r, -Vr \rangle\end{aligned}$$

where we have used the normal equations. Also,

$$\begin{aligned}\delta\tilde{J}_\sigma &= \langle Q_1\tilde{r}, -V\tilde{r} + He \rangle \\ &\quad + \langle S_{\text{data}}, SG\delta P^T He \rangle + \langle S_{\text{data}}, SQ_1 Pe \rangle \, .\end{aligned}$$

Note that

$$\begin{aligned}\langle S_{\text{data}}, SQ_1Pe\rangle &= \langle S_{\text{data}}, S[Q_1,P]e\rangle + \langle S_{\text{data}}, SPGHQ_1e\rangle \\ &= \langle S_{\text{data}}, S[Q_1,P]e\rangle = \langle PGS^TS_{\text{data}}, HQ_1e\rangle \\ &= \langle S_{\text{data}}, S[Q_1,P]e\rangle + \langle \tilde{r}, [Q_1H]e\rangle + \langle Q_1^T\tilde{r}, He\rangle\ .\end{aligned}$$

Thus

$$\begin{aligned}\delta\tilde{J}_\sigma - \delta J_\sigma &= \langle Q_1\tilde{r}, -V\tilde{r}\rangle - \langle Q_1r, -Vr\rangle \\ &\quad + \langle (Q_1+Q_1^T)\tilde{r}, He\rangle \\ &\quad + \langle S_{\text{data}}, SG\delta P^THe\rangle \\ &\quad + \langle S_{\text{data}}, S[Q_1,P]e\rangle \\ &\quad + \langle \tilde{r}, [Q_1,H]e\rangle\ .\end{aligned}$$

We claim that the sum on the r.h.s. is bounded by $\|r-\tilde{r}\|_1$ and $\|e\|_1$, which as noted before can be made as small as you like.

For the first two summands, note that

$$\begin{aligned}\langle Q_1r, Vr\rangle &= \langle r, Q_1^TVr\rangle \\ &= \langle r, (VQ_1^T + [Q_1,V])r\rangle \\ &= \langle Q_1^Tr, Vr\rangle + \langle r, [Q_1^T,V]r\rangle\end{aligned}$$

since V is self-adjoint. So

$$\begin{aligned}2\langle Q_1r, Vr\rangle &= \langle (Q_1+Q_1^T)\,r, Vr\rangle \\ &\quad + \langle r, [Q_1^T,V]r\rangle \\ &= \langle r, Q_0V + [Q_1^T,V]r\rangle\end{aligned}$$

where $Q_1 + Q_1^T = Q_0$ is of order zero (as noted above) and necessarily self-adjoint.

Now we employ the L^2 bounds from the calculus of pseudodifferential operators. Since Q_1^T is of order 1 and V of order 2, $[Q_1^T, V]$ is of order 2, and hence defines a bounded quadratic form on the subspace of H^1 reserved for reflectivity sections r, as does Q_0V. Thus

$$|\langle Q_1\tilde{r}, -V\tilde{r}\rangle - \langle Q_1r, -Vr\rangle|$$

$$
\begin{aligned}
&= \frac{1}{2}\left|\left\{\langle \tilde{r}, (Q_0 V + [Q_1^T, V])\tilde{r}\rangle - \langle r, (Q_0 V + [Q_1^T, V])r\rangle\right\}\right| \\
&= \frac{1}{2}\left|\langle (\tilde{r} - r), (Q_0 V + [Q_1^T, V])\tilde{r}\rangle + \langle r, (Q_0 V + [Q_1^T, V])(\tilde{r} - r)\rangle\right| \\
&\leq C(\,\|r\|_1 + \|\tilde{r}\|_1)(\,\|r - \tilde{r}\|_1)
\end{aligned}
$$

for some $C > 0$.

Similar but simpler arguments show that the other terms are bounded by multiples of $\|\tilde{r}\|_1\ \|e\|_1$ or $\|S_{\text{data}}\|\ \|e\|_1$. Since $\|r\|_1$ and $\|\tilde{r}\|_1$ are both controlled by $\|S_{\text{data}}\|$, the final result is an estimate of the form

$$\|\delta J_\sigma - \delta \tilde{J}_\sigma\| \leq C(\,\|r - \tilde{r}\|_1 + \|e\|_1)\|\delta v_b\|$$

where C depends on v_b and S_{data}. The linear dependence on δv_b follows from the linear dependence of the symbols of Q_1 and δP on δv_b.

Since the error in directional derivatives can be made as small as one likes, the same is true of the error in the gradients, which are defined by

$$
\begin{aligned}
\delta J_\sigma &= \langle \text{grad } J_\sigma, \delta v_b\rangle \\
\delta \tilde{J}_\sigma &= \langle \text{grad } \tilde{J}_\sigma, \delta v_b\rangle\ .
\end{aligned}
$$

Note however that gradient error depends on the error in the reflectivity estimate $\|r - \tilde{r}\|_1$, not just on the normal residual — or, otherwise put, on both $\|e\|_1$ and the coercivity of the preconditioned normal operator M.

There remains the question of a feasible computational procedure: the formula given above for $\delta \tilde{J}_\sigma$ does not define such a procedure, as it turns out. It is tempting to drop the second term (involving $\delta\tilde{r}$) entirely, thus

$$\delta \tilde{J}_\sigma \approx \langle \delta S \cdot \tilde{r}, S\tilde{r} - S_{\text{data}}\rangle$$

i.e., to use the functional form of the exact δJ, but with the approximate $\tilde{r}$. Unfortunately, this approximation is not guaranteed to converge as the Krylov iteration proceeds. In fact, the "cancellation" $Q_1 + Q_1^T = Q_0$ does not occur in this expression, leaving a residual term not dominated by $\|r - \tilde{r}\|_1$, $\|e\|_1$, but only by norms of $r - \tilde{r}$, e involving second derivatives. After discretization, these norms are equivalent, but we lose all control over the rates. Convergence degrades as the grid is refined, since in the continuum limit the size of second derivatives of e may remain bounded away from zero, or even be undefined.

Note however that if we drop the summand involving δP *only*, we obtain an approximation

$$\delta\tilde{J}_\sigma \approx \langle \delta S\tilde{r}, S\tilde{r} - S_{\text{data}}\rangle + \langle S_{\text{data}}, \delta S \cdot Pe\rangle = D$$

in which *only computable quantities occur*, and to which the same analysis as before applies to yield convergence to δJ_σ. (Note that Pe can be computed by applying to e the Krylov polynomial constructed during computation of $\tilde{r}$, e.g. by saving the conjugate gradient coefficients and "replaying" the CG iteration). The essential thought behind this manipulation is: since $\tilde{J}_\sigma$ is only an approximation to J_σ, it is not necessary to compute $\delta\tilde{J}_\sigma$ precisely — only to ensure that it converges to δJ_σ as the Krylov iteration proceeds!

In fact the term involving δP is exactly

$$\langle S_{\text{data}}, SG\delta P^T He\rangle$$

in which $G\delta P^T H$ is a pseudodifferential operator of order 0. Since S is of order 1, this quantity is bounded by a multiple of $\|e\|_1$, as claimed before. Thus

$$|\delta\tilde{J}_\sigma - D| \leq C\|e\|_1$$

and in particular this error is independent of the lower bound for M.

It remains only to extract a practical gradient procedure from these formulae. It would appear necessary to compute δS, which is really the second derivative of the nonlinear forward map (S itself being the linearized forward map). Denote by S_0 the nonlinear forward map: then

$$\begin{aligned}
\delta S \cdot \tilde{r} &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}(S[v_b + \epsilon\delta v_b]\tilde{r} - S[v_b]\tilde{r}) \\
&= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left\{\lim_{h\to 0}\frac{1}{h}(S_0[v_b + \epsilon\delta v_b + h\tilde{r}] - S_0[v_b + \epsilon v_b])\right. \\
&\quad \left. - \lim_{h\to 0}\frac{1}{h}(S_0[v_b + h\tilde{r}] - S_0[v_b])\right\} \\
&= \lim_{h\to 0}\frac{1}{h}\{S[v_b + h\tilde{r}]\delta v_b - S[v_b]\delta v_b\}\ .
\end{aligned}$$

Assume that the interchange of limits is OK (sufficient conditions are known in $1D$, but only conjectured at this point for $nD, n > 1$). Since the source

is assumed oscillatory and v_b and δv_b both smooth, the summand $S[v_b]\delta v_b$ should be negligible. Thus

$$\delta S \cdot \tilde{r} \approx \frac{1}{h} S[v_b + h\tilde{r}]\delta v_b$$

for sufficiently small h. Using this approximation we get

$$\begin{aligned} \delta \tilde{J}_\sigma \approx D \;\; &\approx \;\; \frac{1}{h}\langle S[v_b + h\tilde{r}]\delta v_b \, , S[v_b]\tilde{r} - S_{\text{data}}\rangle \\ &\quad + \frac{1}{h}\langle S_{\text{data}}, S[v_b + hPe]\delta v_b\rangle \; . \end{aligned}$$

In this formula, δv_b is independent of x_s , so that the adjoint operator to S in this context includes summation over x_s , i.e. a stack. We signify this "stacked migration" with an overbar: $\bar{S}^T = \sum_{x_s} S^T$. Then we obtain

$$\text{grad } \tilde{J}_\sigma \approx \frac{1}{h}\left\{\bar{S}^T[v_b + h\tilde{r}](S[v_b]\tilde{r} - S_{\text{data}}) + \bar{S}^T[v_b + hPe]S_{\text{data}}\right\} \; .$$

This "raw" gradient involves only simulation (S) and stacked migration ($\bar{S}^T$), but with a twist: migration is carried out at oscillatory reference states $v_b + h\tilde{r}$, $v_b + hPe$. This accounts entirely for the trend information carried in grad $\tilde{J}_\sigma$, and contrasts with the least-squares gradient

$$\bar{S}[v_b]^T(S[v_b]\tilde{r} - S_{\text{data}}) \; .$$

To be consistent with the assumed smoothness of the velocities, the "raw" gradient sections must be projected into a space of smooth velocity perturbations V; denote by Π_V the projection operator. The implementable approximation to the gradient is then

$$\text{grad}_V \tilde{J}_\sigma \;\; = \;\; \Pi_V \frac{1}{h}\left\{\bar{S}^T[v_b + h\tilde{r}](S[v_b]\tilde{r} - S_{\text{data}}) + \bar{S}^T[v_b + hPe]S_{\text{data}}\right\} \; .$$

1. Velocities used in scan experiments

2. Evaluation of J_{CM} at the velocities of Figure 1, for data from a flat reflector at 750m

3a. Common-x gather, $x = 4400$m, for $v = v_0$

3b. Common-x gather, $x = 4400$ m, for $v = v_1$

4. Reflectivity model for second set of experiments

5. Shot records, $x_s = 3900$ m, 4000 m

6. Evaluation of J_{CM} at velocities of Figure 1, data from reflectivity of Figure 4

7. Common x gathers, $x = 4300$–4700 m, for reflectivity estimated by 10 CR iterations with $v = v_1$

References

[1] G. BEYLKIN. Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized radon transform. *J. Math. Phys.*, 26:99–108, 1985.

[2] D. CAO, S. SINGH, and A. TARANTOLA. Simultaneous inversion for background velocity and impedance maps. *Geophysics*, 55:458–469, 1990.

[3] J.K. COHEN and N. BLEISTEIN. An inverse method for determining small variations in propagation speed. *SIAM J. Appl. Math.*, 32:784–799, 1977.

[4] J. DUISTERMAAT. Fourier integral operators. Lecture notes, Courant Institute, New York, 1973.

[5] O. GAUTHIER, A. TARANTOLA, and J. VIRIEUX. Two-dimensional nonlinear inversion of seismic waveforms. *Geophysics*, 51:1387–1403, 1986.

[6] I.M. GEL'FAND and G.E. SHILOV. *Generalized Functions*, volume I. Academic Press, New York, 1958.

[7] L. HORMANDER. *The Analysis of Linear Partial Differential Operators*, volume I. Springer Verlag, Berlin, 1983.

[8] P. MORA. Nonlinear 2-d elastic inversion of multi-offset seismic data. *Geophysics*, 52:1211–1228, 1986.

[9] KOLB P., F. COLLINO, and P. LAILLY. Prestack inversion of a 1d medium. In *IEEE 74*, pages 498–506, 1986.

[10] C. PERCELL. The effect of caustics in acoustics inverse scattering experiments. Technical Report 89-3, Department of Mathematical Sciences, Rice University, Houston, Texas, U.S.A, 1989.

[11] RAKESH. A linearized inverse problem for the wave equation. *Comm. on P.D.E.*, 13:573–601, 1988.

[12] F. SANTOSA and W. SYMES. An analysis of least-squares velocity inversion. Geophysical Monograph 4, Soc. of Expl. Geophys., Tulsa, 1989.

[13] W. SYMES. Stability and instability results for inverse problems in several-dimensional wave propagation. In R. Glowinski and J.-L. Lions, editors, *Computing methods in applied science and engineering VI.* North-Holland, 1986.

[14] W. SYMES. Velocity inversion by coherency optimization. Technical Report 88-4, Department of Mathematical Sciences, Rice University, Houston, TX, 1988.

[15] W. SYMES. Non-interactive estimation of the Marmousi velocity model by differential semblance optimization: Initial trials. Technical Report 90-36, Department of Mathematical Sciences, Rice University, Houston, TX, 1990. To appear in The Marmousi Experience: Proceedings of the EAEG Workshop on Practical Aspects of Inversion, ed. G. Gran and R. Versteeg, EAEG, The Hague, 1991.

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

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

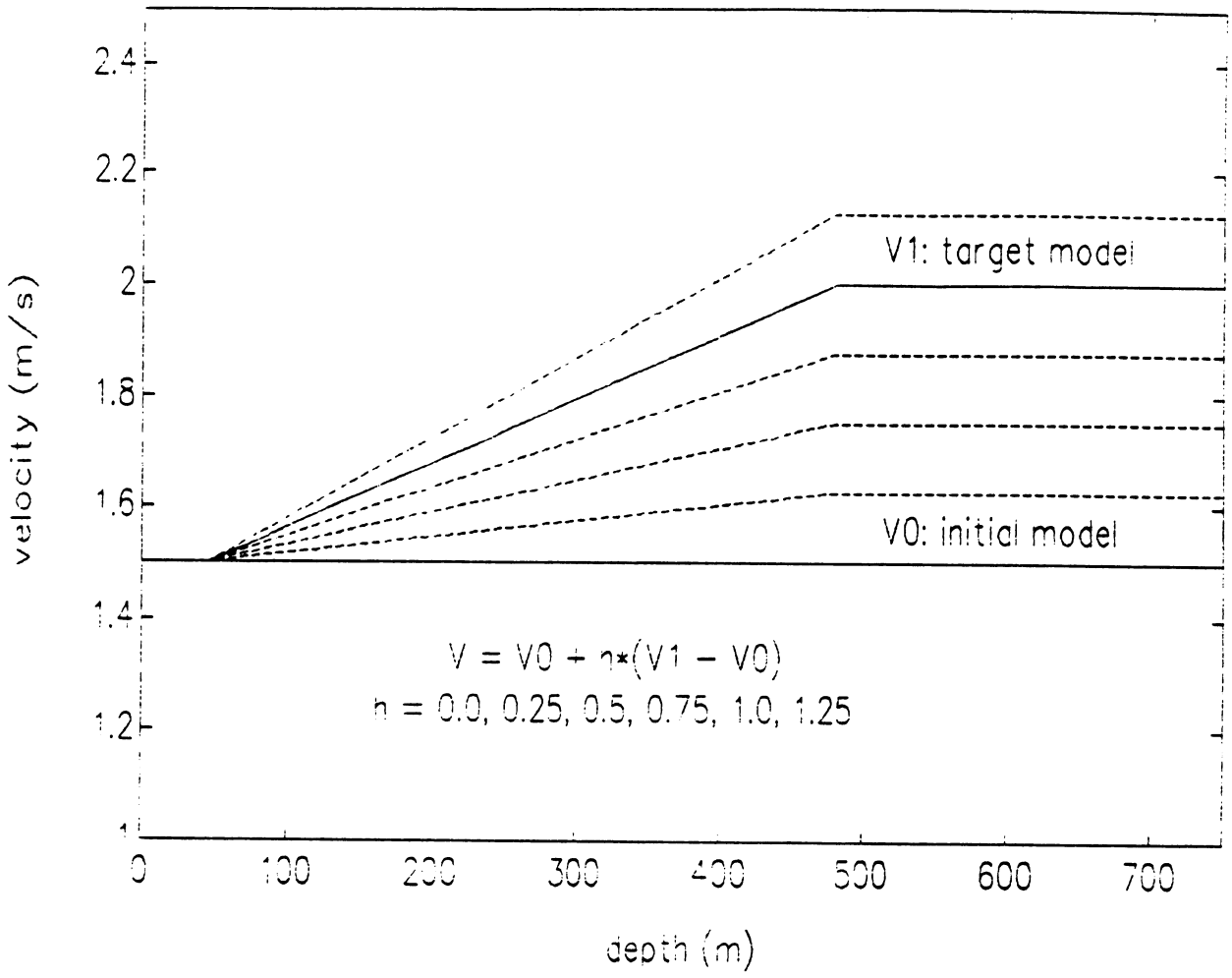
[18] W. SYMES and J. CARAZZONE. Velocity inversion by differential semblance optimization. *Geophysics*, in press, 1990.

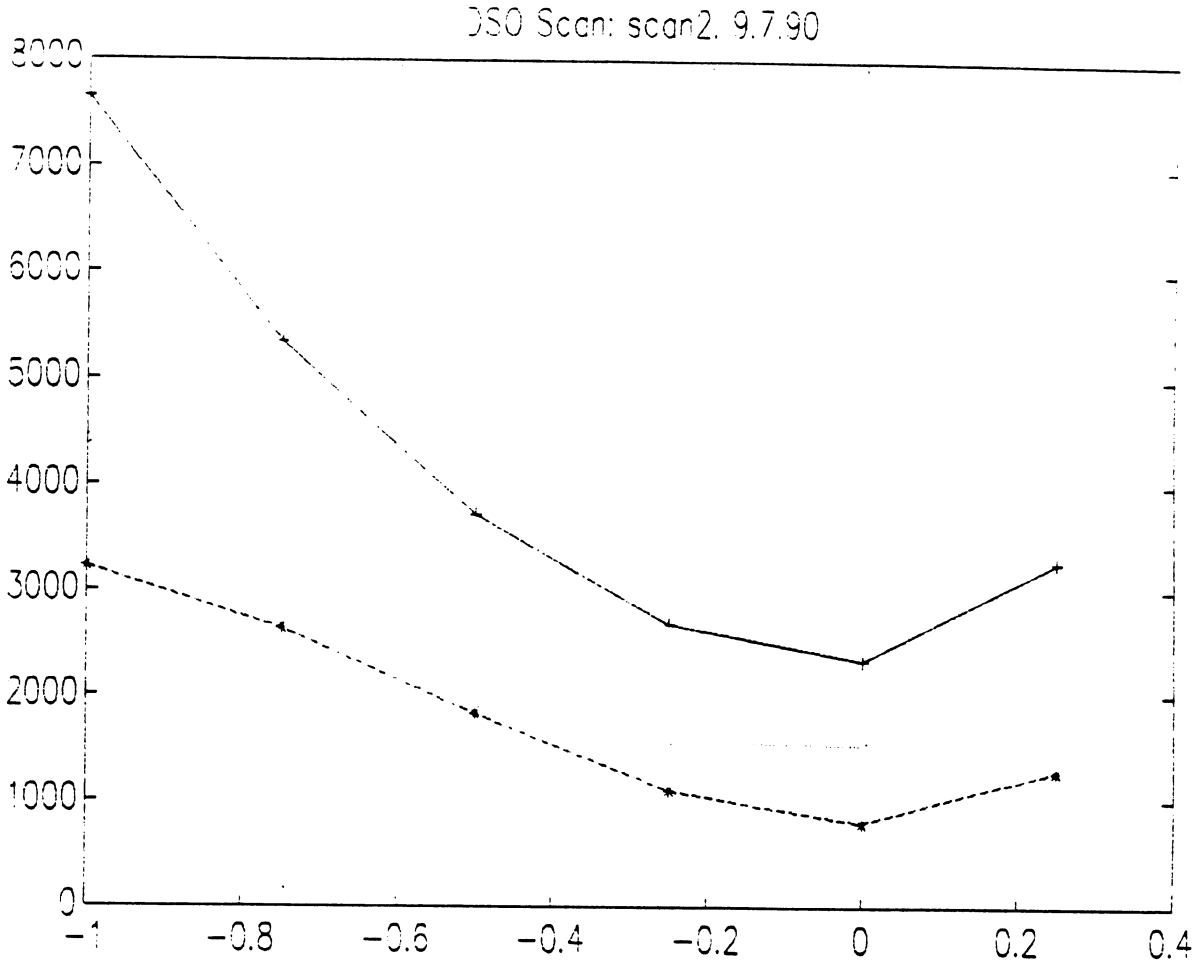
[19] A. TARANTOLA. *Inverse Problem Theory.* Elsevier, 1987.

[20] A. TARANTOLA, E. CRASE, M. JERVIS, Z. KONEN, J. LINDGREN, K. MOSEGARD, and M. NOBLE. Nonlinear inversion of seismograms: State of the art. In *Proc. 60th Annual International Meeting and Exposition*, San Francisco, 1990. Society of Exploration Geophysicists. Expanded abstract S13.7.

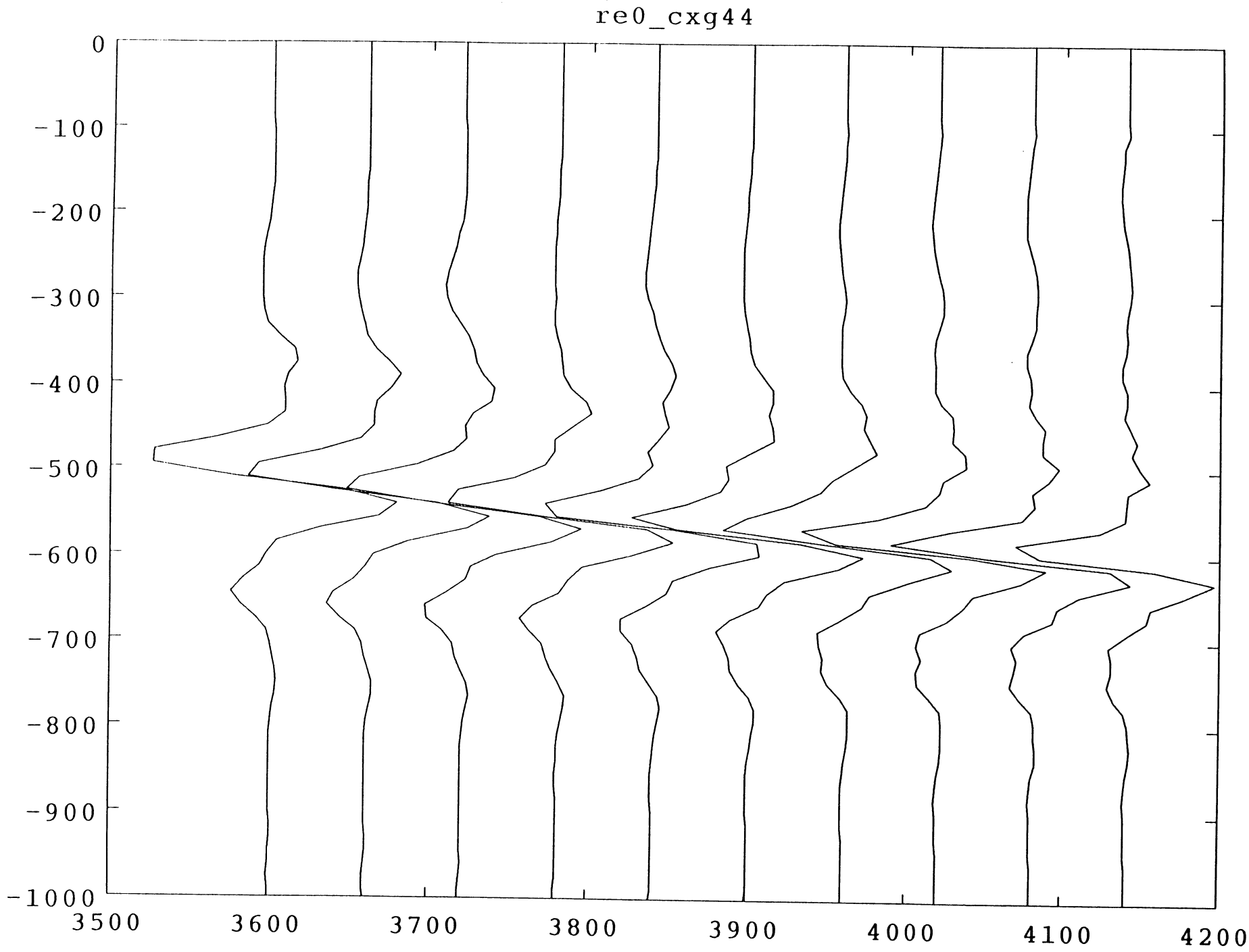
[21] M. TAYLOR. Reflection of singularities of solutions to systems of differential equations. *Comm. on Pure and Applied Math.*, 28:457–478, 1975.

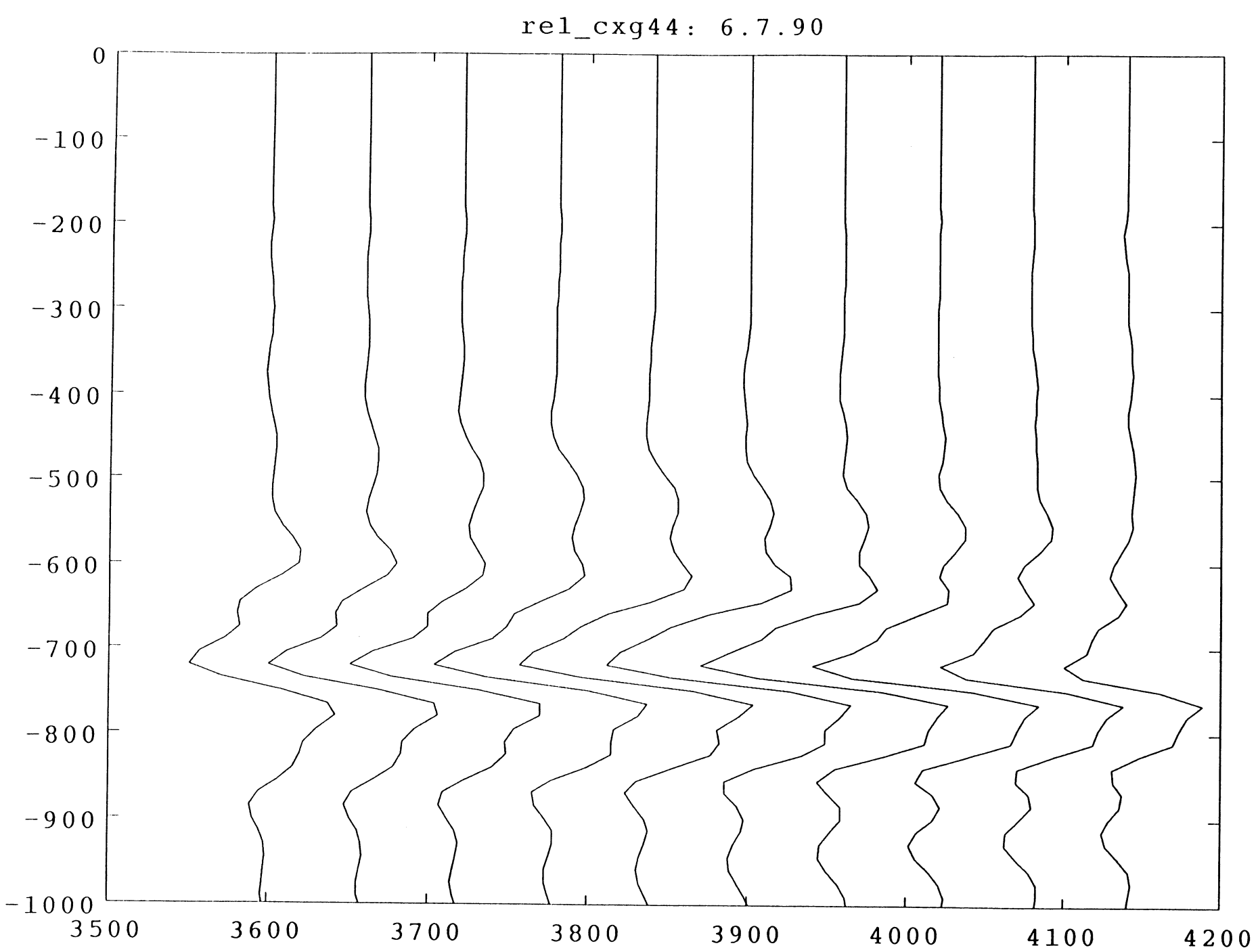
[22] M. TAYLOR. *Pseudodifferential Operators.* Princeton University Press, Princeton, New Jersey, 1980.

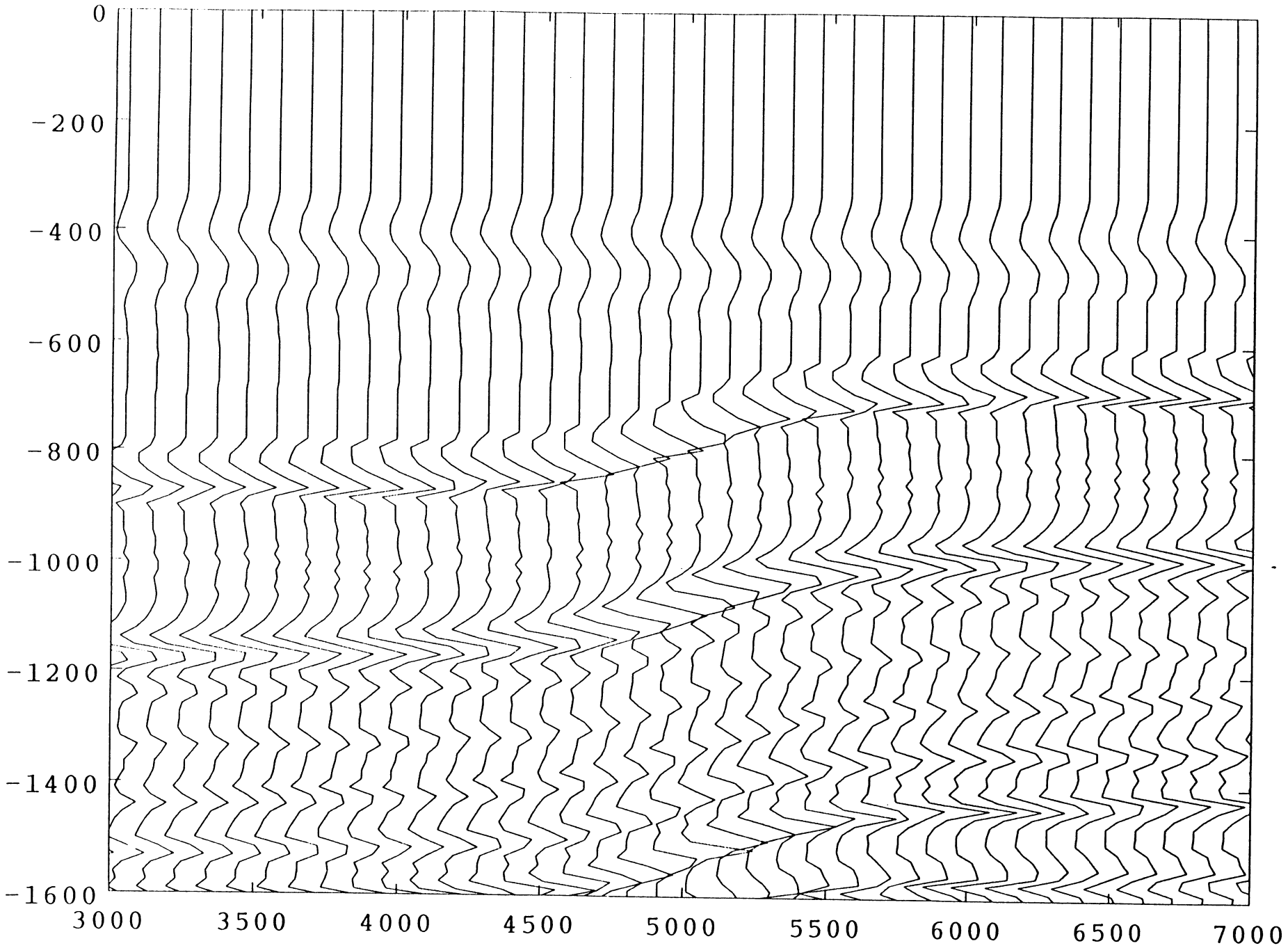

Velocity Models for Scan Experiments: 9.7.90
velocity (m/s)
depth (m)
V1: target model
V0: initial model
V = V0 + h*(V1 − V0)
h = 0.0, 0.25, 0.5, 0.75, 1.0, 1.25
2.4
2.2
2
1.8
1.6
1.4
1.2
0
100
200
300
400
500
600
700

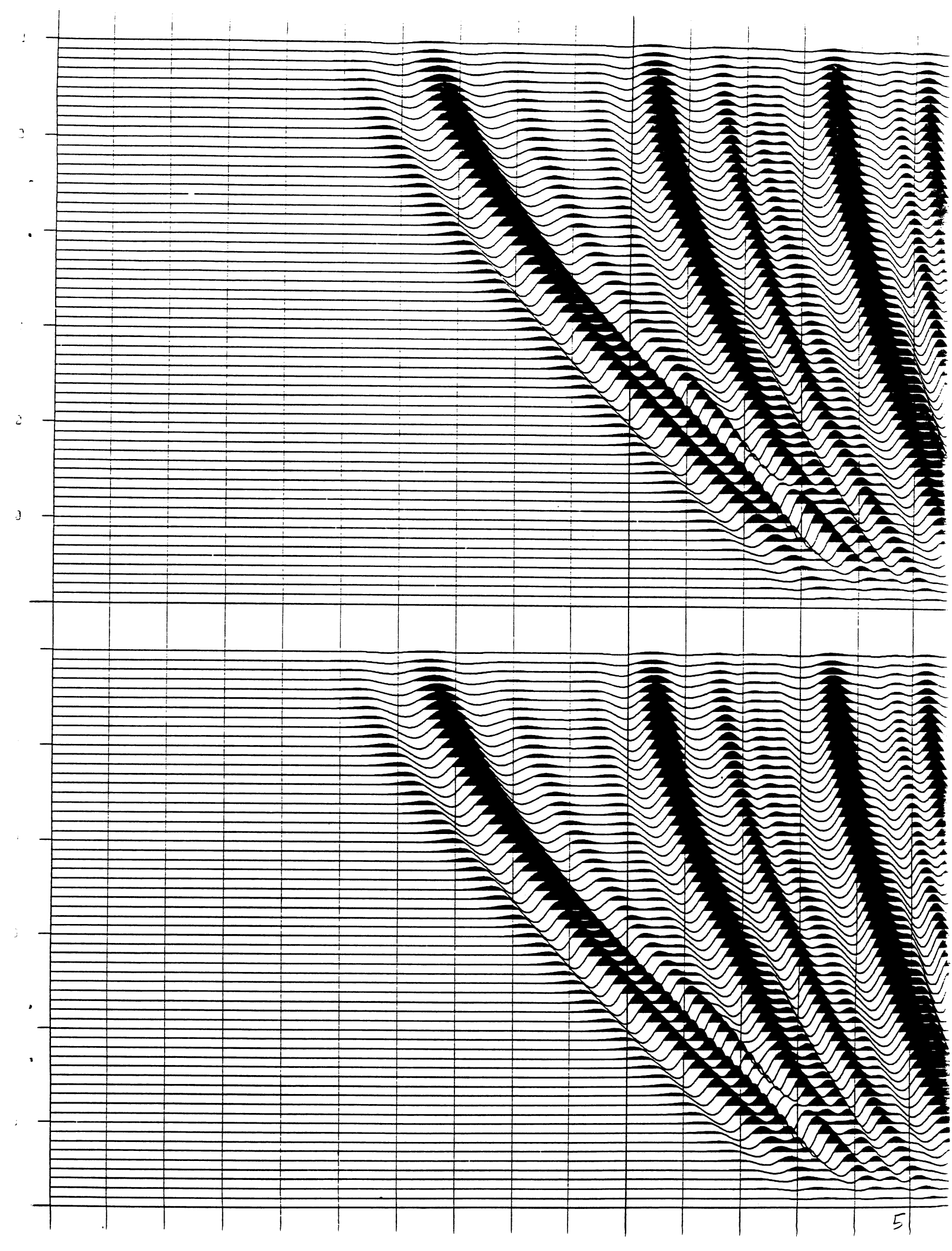


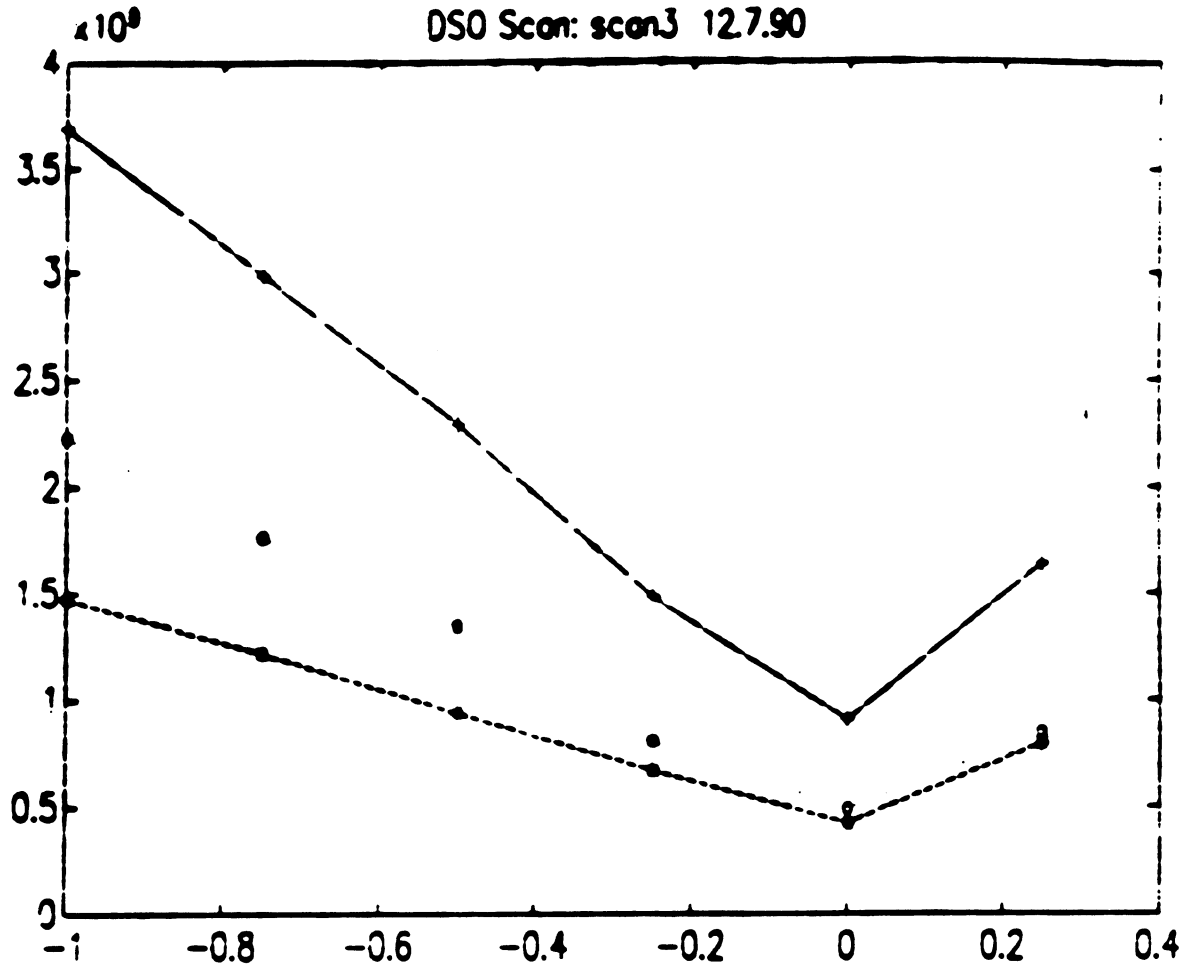
o = ms error; * = incoherency*1000; + = diffsemb


re0_cxg44
0
-100
-200
-300
-400
-500
-600
-700
-800
-900
-1000
3500
3600
3700
3800
3900
4000
4100
4200

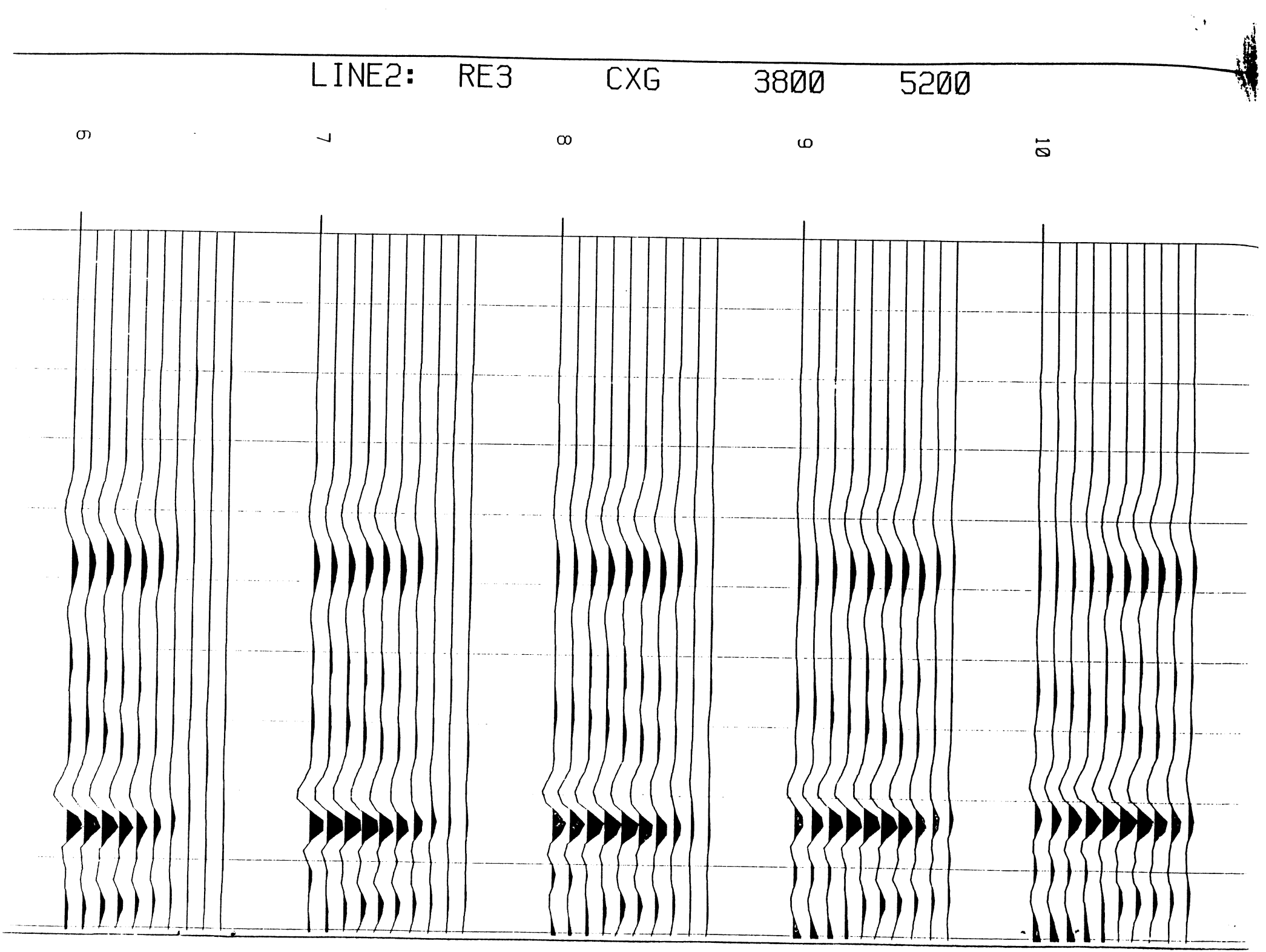

rel_cxg44: 6.7.90
0
-100
-200
-300
-400
-500
-600
-700
-800
-900
-1000
3500
3600
3700
3800
3900
4000
4100
4200


DVEL3 7.7.90
0
-200
-400
-600
-800
-1000
-1200
-1400
-1600
3000
3500
4000
4500
5000
5500
6000
6500
7000





o = ms error; • = inconerency; + = diffsemb


LINE2: RE3 CXG 3800 5200
6
7
8
9
10