

A DIFFERENTIAL SEMBLANCE ALGORITHM FOR THE INVERSE PROBLEM OF REFLECTION SEISMOLOGY

WILLIAM W. SYMES*
Department of Mathematical Sciences
Rice University, Houston, Texas 77215-1892

Abstract — This paper presents an analysis of stability and convergence for a special case of *differential semblance optimization (DSO)*. This approach to model estimation for reflection seismology is a variant of the *output least squares inversion* of seismograms, enjoying analytical and numerical properties superior to those of more straightforward versions. We study a specialization of DSO appropriate to the inversion of convolutional-approximation planewave seismograms over layered constant-density acoustic media. We prove that the differential semblance variational principle is locally convex in suitable model classes for a range of data noise. Moreover, the structure of the convexity estimates suggest a family of quasi-Newton algorithms. We describe an implementation of one of these algorithms, and present some numerical results.

1. INTRODUCTION

Algorithms for solution of the seismic reflection inverse problem (“seismic waveform inversion”) have been discussed extensively over the past several years; see, e.g., [1-7], for a small sample. All of the work just cited is based on the *output least-squares principle*, according to which a subsurface model is required to generate (synthetic) data which fit given data in the least mean-square error sense. This approach does not require picked travel times, unlike reflection tomography [8,9], and, in principle, could extract an optimal distribution of seismic velocities, as well as reflection amplitudes, unlike linearized inversion [10-13]. In addition, any desired level of detail concerning the physics of seismic wave propagation may be built into the output-least-squares principle, and it may also incorporate non-seismic constraints [14].

In practice, it has proven difficult to realize the apparent promise of least-squares inversion even when applied to synthetic data sets. Since the problems are computationally large, only iterative methods are feasible, and these require for their efficient convergence a measure of convexity often not possessed by the mean-square seismogram error (see, e.g., [1,15,16]).

The cause of this non-convexity is the extreme sensitivity of the synthetic seismogram to changes in the slowly-varying components of the velocity, as will be reviewed in Section 2. Thus, velocity trends emerge slowly or not at all as iteration proceeds, and the reflectivities are correspondingly degraded. For layered medium problems, Kolb and others have developed a number of continuation and reparameterization strategies which render the optimization more tractable [5,17]. Only limited tests of these devices (and no theoretical justification) have been reported, however, and it is difficult to understand how lateral heterogeneity might be accommodated by these techniques. Straightforward least-squares inversion of two-dimensional acoustic and elastic models has appeared to require *a priori* information of the gross model (velocity) features of very high quality and, when such information is provided, the results closely resemble those of carefully designed amplitude-preserving migration—unsurprisingly, as these are essentially equivalent [2,18,19]. In fact, the principal tangible result of the work on least-squares

*Research supported in part by the Office of Naval Research under contract N00014-89-J-1115, the Air Force Office of Scientific Research under contract AFOSR89-0056 and by the National Science Foundation under grant DMS89-05878.

inversion so far has been to provide a version of migration which may conceivably yield estimates of *elastic* reflection amplitudes [19–21], hence, a rational basis for amplitude versus offset analysis. While this is an important step, it appears that least-squares inversion has not so far advanced the estimation of velocities in regions of complex structure, and without such velocity information its depth-migration function is disabled as well [16,22].

It is important to understand that the difficulty is *not* due to inadequate modeling of seismic wave propagation, or to the features of field data (beyond specifying the general character of any useful model). That is, least-squares inversion fails for essentially *mathematical*, rather than (*geo-*)*physical*, reasons. These reasons are explored in depth in the monograph [16].

The purpose of this paper is to analyse another approach to full-waveform seismic reflection inversion, in which the coherence of multiple reflectivity estimates is optimized. We have called this approach *differential semblance optimization* (DSO), and have presented theoretical and numerical evidence of its efficacy in a number of recent papers. While the basic ideas behind this approach are quite old, our use of them in designing an optimization algorithm for inversion seems new. Moreover, DSO appears to avoid the mathematical pitfalls which often prove fatal to output-least-squares inversion while producing the same type of subsurface estimate.

Minimal requirements for a useful model-based inversion algorithm might be stated as follows:

- (a) *Stability: at least for data which are nearly model-consistent, the estimates of subsurface properties should “degrade gracefully” in the presence of data noise;*
- (b) *Computability: for iterative methods, this means that convergence should occur at a reasonable rate, and for “poor” initial estimates, i.e., convergence should not require knowing the answer beforehand;*
- (c) *Completeness: the algorithm should extract “most” of the information about the model implicit in the data, and the user should be able to have confidence that it has done so.*

Of course, an algorithm which meets these conditions may still fail: besides these mathematical requirements, one must also impose the (geo-)physical requirement that the underlying model faithfully reflect the physics of the reflection seismology experiment, at least at the level of detail desired in the subsurface estimate. While this physical consistency is the ultimate limiting factor in the utility of any inversion algorithm, our focus in this paper will be limited to the necessary preconditions (a)–(c) above.

The differential semblance approach to seismic reflection inversion was introduced in [23,24]. Theoretical and numerical studies verifying the properties (a)–(c) were presented in [25–27]. The approach was successfully applied to field data in [28,29]. All of these works concern the layered problem discussed here, for which DSO provides the only method known to the author capable of reliable estimation of layered velocities from poor initial estimates. An extension to nonlayered problems is explained in [24,28] and preliminary numerical results given in [26].

In the remainder of this introduction, we shall outline the coherency optimization algorithm in general terms. In succeeding sections we shall:

Section 2: formulate differential semblance principle precisely for the simplest interesting model: layered, constant-density acoustics;

Section 3: analyse the variational principle of Section 2 in some detail, demonstrate its smoothness and local convexity;

Section 4: show, by numerical examples, that a suitable quasi-Newton algorithm will converge to the global optimum for nearly model-consistent data without detailed *a priori* knowledge; thus supporting the contention that (a)–(c) above hold for DSO.

In the concluding Section 5 we recapitulate the properties of DSO and its comparison with other approaches, and offer a few more observations concerning the range of models to which the technique might be applied.

DSO is easiest to describe in the context of the linearized seismogram, in which the model is regarded as the sum of smooth background model and a highly oscillatory (“rough”) perturbation

(reflectivity section). Note that the smooth background velocities are still regarded as part of the model, so it is still nonlinear. With some high-frequency approximations (“WKB Seismogram”) the inversion of each common-source (or common-receiver or common offset) gather *for a given background velocity* may be accomplished semi-analytically (perhaps the best reference is [13]), by what amounts to an amplitude-preserving before-stack migration. If the background model is correct, presumably the images from the various inverted gathers will line up. If not, the discrepancy indicates that the background (“migration”) velocity ought to be changed.

So far, this is hardly novel: a process called “iterative before-stack migration” is described in just this way in [30], for instance. The refinement leading to a feasible algorithm is the introduction of a well-behaved *cost function*, which we dub the *differential semblance*, which both measures the discrepancy between the various common gather inversions and indicates how the velocity fields should be changed to line them up (through its *gradient*). A natural choice is *stack power* or *semblance* [31], for example), but this leads directly back to the nonconvexity difficulties of output-least-squares: the two are very closely related [16, Appendix A]. Instead, we depend on the accurate inversion of amplitudes, and take the *mean-square of the differences* between successive inverted gathers. These differences should all vanish at the “correct” velocity estimate. Most important, as the background velocity is changed, this “differential semblance” (DS) changes smoothly, rather than abruptly, as is the case with stack power.

Two technical refinements are necessary to turn this idea into an algorithm. First, data noise which is uncorrelated from gather to gather will yield a very large contribution to the differences of the inverted gathers, completely distorting the DS and masking the correct choice of velocity. To avoid this oversensitivity to data noise, partly decouple the inverted gathers from the data: lump these reflectivities (one per gather) together with the background velocity field as the “model” to be determined, and redefine the *cost* to be the sum of:

- (a) *all of the mean-square data errors, i.e., differences between data seismogram and predicted seismogram based on the corresponding reflectivity section, and*
- (b) *the mean-square sum of pairwise differences of reflectivity sections (DS).*

Incoherent noise in the data gathers will be accounted for mostly in error-of-fit (i.e., (a)), as it should be, since it causes a much smaller increase in cost that way than when forced into the DS (i.e., (b)).

The second technical modification is required by the nature of the model-to-seismogram relation: as pointed out above, this relation is extremely sensitive to the background velocity component, when the model is defined in terms of depth (and offset). This oversensitivity is due to the *time shift* accompanying a change in background velocity, which causes the temporal location of a high-frequency reflection from a (depth-) fixed reflector to change. For a discussion of this “time-shift” disaster in the one-dimensional context see [32]. The remedy is clear: the reflectivity sections should be defined as *time sections* rather than as depth sections.

For the layered acoustic problem of this paper, it is rather trivial to accomplish this transformation. It is particularly convenient to work with plane wave sources, since the gather corresponding to a planewave source at definite slowness has only one independent trace. Thus a single reflectivity time trace at each slowness, together with the (single) background velocity depth profile, specifies the model. The reflectivity time traces are converted to depth traces through the (background) vertical travel-time-to-depth transformation appropriate to each slowness. The high-frequency approximation to the (p -tau) seismogram is simply the convolution of the time-trace reflectivities with the source wavelet, whereas the DS is essentially the derivative of the corresponding depth-trace reflectivities with respect to slowness.

For laterally heterogeneous models, the transformation from time to depth is less obvious. As explained in [25,29], the analogue of the travel-time-to-depth change of variables appropriate to plane waves in layered media is, in general, a before-stack Kirchhoff migration appropriate to the acquisition geometry being simulated, and the reflectivity gather is simply a time section. The differential semblance is a difference (or derivative) with respect to the shot location (for shot gathers) of the migrated reflectivity depth sections.

Two points remain to be emphasized. First, there is nothing to “cohere” unless reflectors are present, i.e., unless the reflectivity gathers are rich in high-frequency energy. As is the case with semblance, the differential semblance is a measurement of the portion of move-out in reflection phase not accounted for by the current velocity model. Therefore, the density of reflectors is a limiting factor in the recovery of accurate velocity models, as is the data aperture. These observations are quantified in Section 3.

Second, both DS and data misfit are summed in the cost functional for DSO. Thus both are minimized simultaneously. The major theoretical result of the paper (Section 3) is that, with appropriate weights on the two components, sufficiently dense reflectors and enough aperture, the cost functional is smooth and convex for near-consistent data. It also appears that due to the structure of these estimates, the selection of weights can be arranged to cause suitable quasi-Newton iterations to converge to the global optimum without accurate *a priori* knowledge. Numerical evidence for this assertion is given in Section 4; a corresponding theorem will be proved in another publication.

2. OUTPUT LEAST-SQUARES VS. DIFFERENTIAL SEMBLANCE OPTIMIZATION

Denote the plane-wave (“p-r”) seismogram corresponding to a velocity profile $c(z)$ by $S[c]$. The arguments in this paper are based on the well-known convolutional approximation, which is reasonably accurate when c may be re-written as a sum

$$c \approx c_s + c_r,$$

where

$c_s(z)$ is a slowly varying background velocity model

$c_r(z)$ is a rapidly varying “reflector sequence,” having locally zero mean on the length scale of significant change in the background velocity.

Thus the (two-way) travel-time to depth z of a precritical plane-wave at (horizontal) slowness p is determined with a small error by the background velocity $c_s(z)$, according to

$$\tau(z, p) = 2 \int_0^z dz' \left(\frac{1}{c_s^2(z')} - p^2 \right)^{1/2} = 2 \int_0^z \frac{1}{v_s},$$

where v is the vertical (plane-wave) velocity at slowness p :

$$\begin{aligned} v_s(z, p) &= \left(\frac{1}{c_s^2(z)} - p^2 \right)^{-1/2} = c_s(z) \Lambda(z, p), \\ \Lambda(z, p) &= (1 - c_s^2(z) p^2)^{-1/2}. \end{aligned}$$

The convolutional approximation with (isotropic) source wavelet $f(t)$ is then

$$\begin{aligned} S[c_s, c_r](t, p) &= v(0, p)^{-1} \left(f * \frac{\partial r}{\partial t} \right) (t, p) \\ &= \frac{1}{v(0, p)} \int_0^t dt' f(t - t') \frac{\partial r}{\partial t}(t', p), \end{aligned} \tag{1}$$

where the “reflectivity” $r(t, p)$ is given by

$$r(t, p) = \frac{v_r(\zeta(t, p), p)}{v_s(\zeta(t, p), p)}, \tag{2}$$

by means of the *inverse two-way travel-time* function ζ , defined implicitly by

$$t = 2 \int_0^{\zeta(t, p)} \frac{1}{v},$$

and the vertical velocity perturbation

$$v_r = c_r \cdot \Lambda^3.$$

Note that reflectivity conventionally means $\partial r / \partial t$; we shall confuse r and its t -derivative, calling both “reflectivity” as convenient.

Now S is clearly linear in c_r , but quite nonlinear in c_s . In fact, a change in c_s will typically result in a change in the “phase” ζ , and thus in a *shift* in the high-frequency components of S , which in turn derive from the high-frequency components in c_r . Since such a phase-shift may have a drastic effect on components of the appropriate (high) frequency, and since c_r must have a great deal of high-frequency content in order to model the dense distribution of reflectors in the typical sedimentary column, one expects S to be extremely sensitive to changes in the background velocity c_s .

This oversensitivity to background errors shows quite clearly in the expression derived for the perturbation δS in the seismogram, due to a perturbation δc_s in the background velocity (holding c_r fixed, and assuming $\delta c_s(0) = 0$):

$$\delta S(t, p) = v_s(0, p)^{-1} f * \frac{\partial}{\partial t} \left(-\frac{v_r}{v_s^2} \delta v_s + \frac{\partial}{\partial z} \left(\frac{v_r}{v_s} \right) \delta \tilde{\zeta} \right) (\zeta(t, p), p),$$

where

$$\delta \tilde{\zeta}(z, p) := \delta \zeta(\tau(z, p), p) = v_s(z, p) \int_0^z dz' v_s(z', p) c_s^{-3}(z') \delta c_s(z')$$

is the phase perturbation corresponding to c_r , referred to depth/slowness coordinates.

Thus, the z -derivative of c_r (or, of r) appears in the perturbations δS associated with a background velocity perturbation δc_s . On the other hand, a perturbation δc_r in c_r simply results in

$$\delta S \sim S[c_s, \delta c_r],$$

as S is linear in c_r —thus, no derivative of c_r is involved.

Again, c_r must be highly oscillatory to model the typical reflector distribution, so the derivative of c_r is typically much bigger, in any reasonable sense, than is c_r itself. Thus, a perturbation of the background velocity c_s has a much larger effect on S than does a perturbation of c_r : in the language of linear algebra, the linear perturbation map (“Jacobian”)

$$\delta c_s, \delta c_r \mapsto \delta S$$

is *ill-conditioned*.

Even worse, a straightforward extension of this reasoning shows that the difference between the perturbed section

$$S[c_s + \delta c_s, c_r + \delta c_r]$$

and its linear prediction

$$S[c_s, c_r] + \delta S$$

can easily be on the order of S itself, even for quite small δc .

To summarize: under realistic assumptions on c_s and c_r ,

- (i) the Jacobian δS is ill-conditioned;
- (ii) S is poorly approximated by its linearization.

Consequently, the output least-squares objective function

$$J[c_s, c_r; S_{\text{data}}] := \int \int dp dt |S[c_s, c_r] - S_{\text{data}}|^2$$

is highly non-convex, with rapidly changing gradient, and the optimization problem

$$\min_{c_s, c_r} J[c_s, c_r; S_{\text{data}}] \quad (3)$$

is extremely difficult to solve by means of any variant of Newton's method. For extensive discussion and illustration of these points, consult [16].

The crux of the difficulty is the interaction between the background c_s and the reflectivity r through the definition (2). Accordingly, it is tempting to *decouple* c_s and r by treating r , rather than c_r , as the "other" component of the model: thus define

$$\tilde{S}[c_s, r](t, p) = v_s(0, p)^{-1} f * \frac{\partial r}{\partial t}(t, p).$$

Certainly, if r and c_r are related by (2), then

$$S[c_s, c_r] = \tilde{S}[c_s, r].$$

In fact, apart from the surface normalization, the background velocity c_s enters the definition of \tilde{S} only implicitly, through the condition (2). If we are to use r as one of the independent variables, instead of c_r , we must develop a condition, phrased only in terms of c_s and r , which guarantees that (2) holds. Fortunately, this is rather easy: from the useful identity

$$\frac{v_r}{v_s^3} \equiv \frac{c_r}{c_s^3}, \quad (4)$$

it follows that, if (2) is satisfied, then

$$c_s^{-3}(z) c_r(z) = v_s^{-2}(z, p) r(\tau(z, p), p). \quad (5)$$

The notable quality of (5) is that the left-hand side is independent of p . Thus, differentiation with respect to p eliminates c_r altogether:

$$0 = \frac{\partial}{\partial p} [v_s^{-2}(z, p) r(\tau(z, p), p)]. \quad (6)$$

It is easy to reverse this reasoning. Thus:

A section $r(t, p)$ is the reflectivity corresponding to $c_r(z)$ if and only if (6) is satisfied, in which case c_r is given in terms of $r(t, p)$ and $c_s(z)$ by (5).

In formulating the constraint given by (6), it is advantageous to return to (t, p) coordinates (the reason will become apparent below). Thus define the quantity $W[c_s, r]$ (the *differential semblance*) by

$$W[c_s, r](t, p) = \left\{ \frac{\partial}{\partial p} [v_s^{-2}(z, p) r(\tau(z, p), p)] \right\} \Big|_{z=\zeta(t, p)}.$$

Then

$$S[c_s, c_r] = S_{\text{data}},$$

if and only if

$$\begin{aligned} \tilde{S}[c_s, r] &= S_{\text{data}}, \text{ and} \\ W[c_s, r] &\equiv 0, \end{aligned}$$

with c_r and r related by (2) or (5).

Since the problem is clearly overdetermined, it is commonplace to replace the exact fit of predicted to measured seismograms by a best-fit condition. We shall follow the lead of most authors on this subject (e.g., [33]) and use the mean-square error measure, i.e., L^2 norms. Since derivatives are also involved in the definitions of important quantities, we also introduce some norms from the Sobolev scale, which measure the size of derivatives. Some notation for this is desirable: for a single z - or t -trace, e.g., $f(t)$, we write

$$\begin{aligned} \|f\| &= \left(\int |f(t)|^2 \right)^{1/2}, \\ \|f\|_k^2 &= \sum_{j=0}^k \|f^{(j)}\|^2, \end{aligned}$$

whereas for a (t, p) or (z, p) section, e.g., $F(t, p)$, we write

$$\begin{aligned} \|F\| &= \left(\int dt \int dp p |F(t, p)|^2 \right)^{1/2}, \\ \|F\|_1^2 &= \|F\|^2 + \left\| \frac{\partial F}{\partial t} \right\|^2 + \left\| \frac{\partial F}{\partial p} \right\|^2. \end{aligned}$$

Integrations in all cases are over appropriate sets.

The weight p in the integral defining $\|F\|$ is a residue of polar coordinates, and is used to make the norm $\|F\|$ as close as possible to equivalent to the ordinary mean-square of an $x-t$ section. See [16, Appendix B]; also [34].

In view of the equivalence noted above, it is natural to pose the constrained least-squares problem:

$$\begin{aligned} \text{minimize}_{c_s, r} \quad & \| \tilde{S}[c_s, r] - S_{\text{data}} \|^2, \\ \text{subject to} \quad & W[c_s, r] \equiv 0. \end{aligned} \tag{7a}$$

This problem is closely related to the output least-squares problem (3). In fact, the relation is too close. The two problems have the same solution, and, in fact, are mathematically quite similar as well, so that (7a) is also poorly suited to computation. (For a detailed analysis of a simpler situation with similar features, see [25].) To obtain a problem which is not too "stiff," we "relax" (7a) by making the constraint $W = 0$ soft:

$$\text{minimize}_{c_s, r} \quad \tilde{J}[c_s, r, S_{\text{data}}], \tag{7b}$$

where

$$\tilde{J}[c_s, r, S_{\text{data}}] = \| \tilde{S}[c_s, r] - S_{\text{data}} \|^2 + \sigma^2 \| W[c_s, r] \|^2.$$

We shall call (7b) the *differential semblance optimization problem*.

From the equivalence above, we see that if S_{data} is *consistent*, i.e., $S_{\text{data}} = S[c_s, c_r]$, then the problem (7b) has $[c_s, r]$ amongst its solutions, for which r and c_r are related by (5). That is, for consistent data, (7b) has "the same solutions" as the output least-squares problem (3). We shall show, however, that (7b) is far better suited to numerical computation by Newton's method and its relatives. We shall also show that the solution of (7b) is stable, i.e., "degrades gracefully" in the presence of data error, for reasonable choices of the penalty parameter σ .

These conclusions are true, provided that S_{data} is near (in the mean-square sense) to some "exact" or consistent data $S[c_s, c_r]$, and provided that c_s, c_r satisfy certain conditions. "Physical," or poetic, statements of these conditions are:

- (i) c_r should be "rough," i.e., contain significant variation (reflectors), on a length scale dictated both by the wavelet (f) passband and by the smoothness (characteristic length) of c_s ;
- (ii) The range of slownesses p available in the data ("aperture") should be sufficiently large relative to both the degree of roughness mentioned in (i) and the amount of data error, so that the move-out of reflections clearly discriminates the velocities.

It is also necessary that the penalty parameter σ be chosen appropriately. "Appropriate" means, so that the desired properties are obtained, i.e., so that Newton's method works (from a poor initial guess) and so that the solution obtained is relatively insensitive to data error. Note that in the limit $\sigma \rightarrow \infty$, (7b) becomes (7a), i.e., the constraint becomes hard. Previous remarks have indicated that (7a) is not susceptible to numerical solution through Newton's method or its relatives. Thus the possibility of choosing σ "not too large" is crucial to the construction of algorithms which converge from a poor initial guess.

In the following section we will carefully quantify (i) and (ii) and justify these conclusions through an analysis of (7b).

3. LOCAL CONVEXITY OF DSO

The local analysis of optimization problems amounts to the verification of several conditions concerning the first and second derivatives of the objective function. These conditions are imposed near a particular solution (i.e., local minimum) and guarantee that this solution is stable under data perturbations. The same conditions ensure rapid local convergence of Newton's method. We state the verbal description of these conditions here, together with the interpretation in terms of the quantities introduced in the last section:

(i) *(Regularity) The objective function should be twice differentiable near the solution.*

This is completely obvious for the seismogram error term $||\tilde{S}[c_s, r] - S_{\text{data}}||^2$ as it is independent of c_s and quadratic in r . For the DS, the regularity is suggested by the identity

$$\begin{aligned} W[c_s, r](t, p) &= \left. \frac{\partial}{\partial p} (v_s(z, p)^{-2} r(\tau(z, p), p)) \right|_{z=\zeta(t, p)} \\ &= -2p r(t, p) + v_s^{-2}(\zeta(t, p), p) \left\{ \frac{\partial r}{\partial p}(t, p) - p \frac{\partial r}{\partial t}(t, p) \int_0^{\zeta(t, p)} v_s \right\}. \end{aligned} \tag{8}$$

Clearly, varying c_s will have the effect of differentiating v_s —but c_s , hence v_s , is presumed sufficiently smooth that its derivative is not significantly larger than c_s itself (we make this assumption precise below). On the other hand, no additional derivatives of r (the locus of high-frequency energy) result from varying c_s . Thus W may be regarded as regular. Note that the perturbation of W would involve further derivatives of r if we had left off the final referral back to time/slowness coordinates.

(ii) *(Second-order sufficiency) The objective Hessian is positive-definite.*

Most of this section is devoted to verifying this last condition; it is the source of the requirements mentioned at the end of the last section.

The Hessian is the second-order coefficient $\delta^2 \tilde{J}$ in the power series expansion

$$\tilde{J}[c_s + \epsilon \delta c_s, r + \epsilon \delta r, S_{\text{data}}] = \tilde{J}[c_s, r, S_{\text{data}}] + \epsilon \delta \tilde{J} + \frac{\epsilon^2}{2} \delta^2 \tilde{J} + \dots$$

In view of the goal, enunciated at the end of the last section, to examine the perturbation of a consistent solution due to perturbing the data, we shall assume that

$$\tilde{J}[c_s, r, S_{\text{data}}] = 0.$$

That is, that

$$\tilde{S}[c_s, r] = S_{\text{data}} \text{ and } W[c_s, r] = 0. \tag{9}$$

Let us state very carefully the meaning of the condition (ii) in terms of $\delta^2 \tilde{J}$: i.e., $\delta^2 \tilde{J}$ is positive. Condition (ii) actually requires that $\delta^2 \tilde{J}$ be "positive relative to" $\delta c_s, \delta r$: for some $L > 0$,

$$||\delta^2 \tilde{J}|| \geq L (||\delta c_s||^2 + ||\delta r||^2). \tag{10}$$

Indeed, as follows from the proof of the implicit function theorem, as $L \rightarrow 0$ in (10), the size of the region of convexity of \tilde{J} goes to zero, and the possible ratio of solution error to data error goes to infinity. Thus, we need not only $L > 0$, but some positive control over L , to ensure stability (and computability) of the solution.

An easy calculation gives

$$\delta^2 \tilde{J} = 2 v(0, p) \left\| \left\| f * \frac{\partial}{\partial t} \delta r \right\| \right\|^2 + 2\sigma^2 ||\delta W||^2.$$

Thus as a first requirement, we must have (for each p)

$$\left\| f * \frac{\partial}{\partial t} \delta r \right\| \geq K_1 \left\| \frac{\partial}{\partial t} \delta r \right\|, \quad (11)$$

for some constant $K_1 > 0$. Unless K_1 is to be uselessly tiny, this means, in effect, that each trace δr ($p = \text{const.}$) (hence, eventually, r) must have most of its energy in the passband of the source f , or else, that the out-of-band components must be constrained *a priori*. We adopt the first option: i.e., that we shall estimate only passband reflectivities. In quantitative terms, K_1 should be at least some substantial fraction, say one-half, of the absolute moment of the wavelet (this choice is justified by the theory of convolution operators, and is essentially optimal):

$$K_1 \approx .5 \int |f|.$$

Inequality (11) only shows that $\delta^2 \tilde{J}$ majorizes $\|(\partial \delta r)/(\partial t)\|^2$. Reference to (10) shows that this is inadequate—unsurprisingly, since we have not made full use of the hypotheses of (ii). To go further, we must introduce the linearized constraints (since otherwise δc is unconstrained). As we have assumed c_s, r, S_{data} consistent, i.e., the system of equations (9), this simplifies considerably. The calculation is displayed in the Appendix: with c_r given by (11), we obtain the identities

$$\delta W = \delta_c W + \delta_r W, \quad (12a)$$

$$\begin{aligned} \delta_c W = & -2p \left\{ 2 \left[(Q \delta c) \Lambda c^{-1} + \left(\int_0^z Q \delta c \right) p^2 \Lambda^3 c' \right] \left(\frac{c_r}{c} \right) \right. \\ & \left. + \left[\left(\int_0^z Q \delta c \right) c^{-1} \Lambda \right] \frac{\partial}{\partial z} \left(\frac{c_r}{c} \right) \right\} \Big|_{z=\zeta}, \end{aligned} \quad (12b)$$

$$\delta_r W = W[c, \delta r] = \left\{ \frac{\partial}{\partial p} (v^{-2} \delta r \circ \tau) \right\} \circ \zeta, \quad (12c)$$

Here the quantity $Q \delta c$ is defined by

$$Q \delta c := [\delta(v \circ \zeta)] \circ \tau = \Lambda^3 \left(\delta c + c' v \int_0^z v \frac{\delta c}{c^3} \right).$$

(We shall suppress the subscript “ s ” on c_s whenever no confusion is possible.) Also, we have used the useful notation

$$g \circ \tau(z, p) = g(\tau(z, p), p)$$

for (any) function $g(t, p)$ and, similarly, for the symbol $\circ \zeta$.

The further analysis of this condition in general is a little involved. We consider in detail only the special case

$$\begin{aligned} c & \text{ (hence } v \text{) constant for } z \geq z_0 \\ \delta c & \text{ (hence } \delta v \text{) } \equiv 0 \text{ for } z \leq z_0. \end{aligned} \quad (13)$$

A few remarks about the general case appear near the end of this section. See Figure 1.

Under this restriction, $\delta_c W$ simplifies considerably: $Q \delta c = \Lambda^3 \delta c$, and

$$\delta_c W \circ \tau = -2p \Lambda^4 c^{-2} \left\{ 2\delta c c_r + \left(\int_0^z \delta c \right) \left(\frac{\partial c_r}{\partial z} \right) \right\}.$$

Now

$$\frac{\partial}{\partial p} \Lambda^2 = 2p c^2 \Lambda^4.$$

Thus what remains of (12a) is an exact derivative in p , so we integrate in p from p_1 to p_2 to obtain

$$\begin{aligned} c^{-4} \left\{ 2\delta c c_r + \left(\int_0^z \delta c \right) \left(\frac{\partial c_r}{\partial z} \right) \right\} \Lambda^2 \Big|_{p_1}^{p_2} &= v^{-2} (\delta r \circ \tau) \Big|_{p_1}^{p_2} + \int_{p_1}^{p_2} dp (\delta_c W \circ \tau) \\ &= c^{-2} \Lambda^{-2} (\delta r \circ \tau) \Big|_{p_1}^{p_2} + \int_{p_1}^{p_2} dp (\delta_c W \circ \tau). \end{aligned} \quad (14)$$

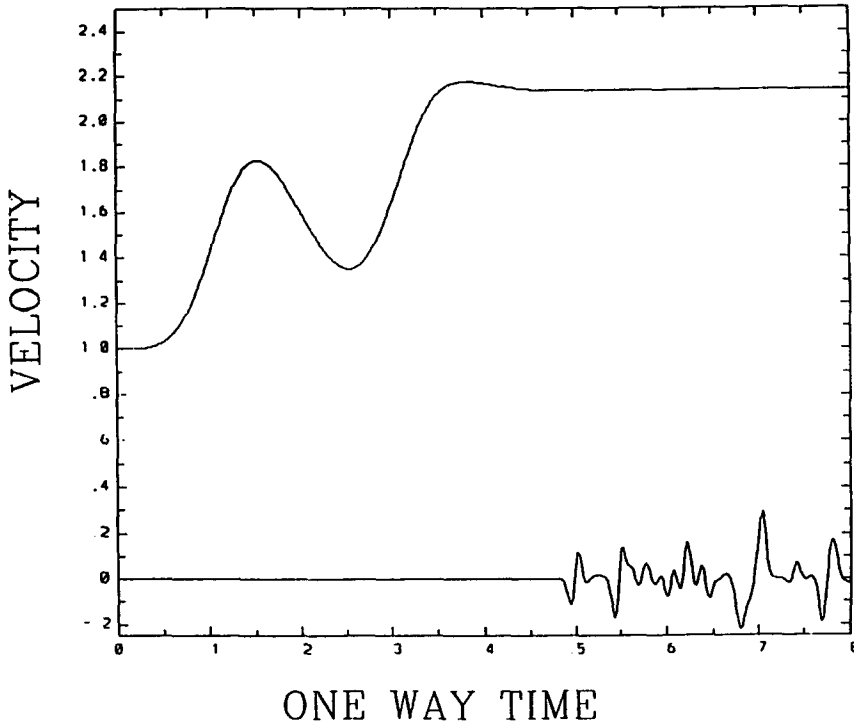


Figure 1. Smooth velocity model (upper curve) and perturbation (lower curve): the latter is non-zero only where the former is constant (this is the situation analysed in Section 3). The horizontal scale is one-way time at surface velocity.

Equation (14) will be the “hook” to get $\delta^2 \tilde{J}$ into an explicit relation with δc . This is accomplished in three steps. First, we make use of an important by-product of the assumed bandlimited feature of f :

Since f , hence $r, \delta r$, has its energy concentrated away from 0 Hertz, the t -derivative of δr is at least as big as (probably much larger than) δr itself: for each p ,

$$K_2 \|\delta r\|^2 \leq \left\| \frac{\partial \delta r}{\partial t} \right\|^2,$$

for a constant K_2 depending on the lowest passband frequency.

Together with (11), this gives

$$K_2 K_1^2 \|\delta r\|^2 \leq \delta^2 \tilde{J}. \tag{15}$$

The second step is to manipulate (14) into an inequality with $\|\delta r\|^2$ on the right-hand side. This is accomplished with the aid of several identities relating time and depth, which are massively simplified by assuming that $c \equiv \text{const.}$, where $\delta c \neq 0$, as we have done.

According to the hypothesis (13), for $z > z_0$,

$$\begin{aligned} c(z) &= c_1, \quad \text{for } z > z_0, \\ \tau(z, p) &= \tau_0(p) + \frac{2(z - z_0)}{c_1 \Lambda_1(p)}, \end{aligned}$$

where

$$\tau_0(p) = \tau(z_0, p),$$

and

$$\Lambda_1(p) = (1 - c_1^2 p^2)^{-1/2}.$$

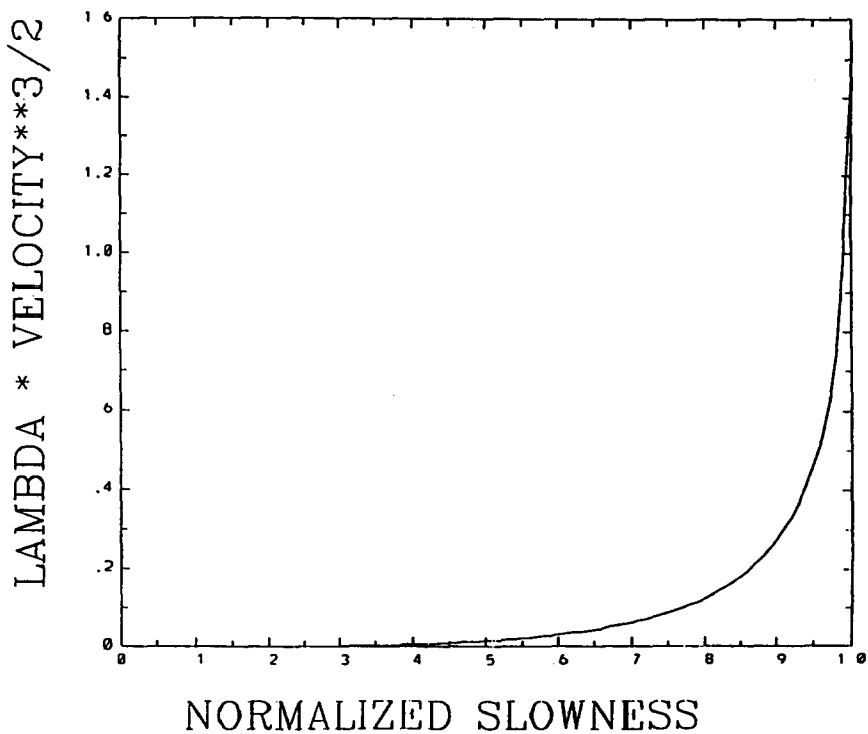


Figure 2. The coefficient λ (display (18)), normalized to remove velocity dependence.

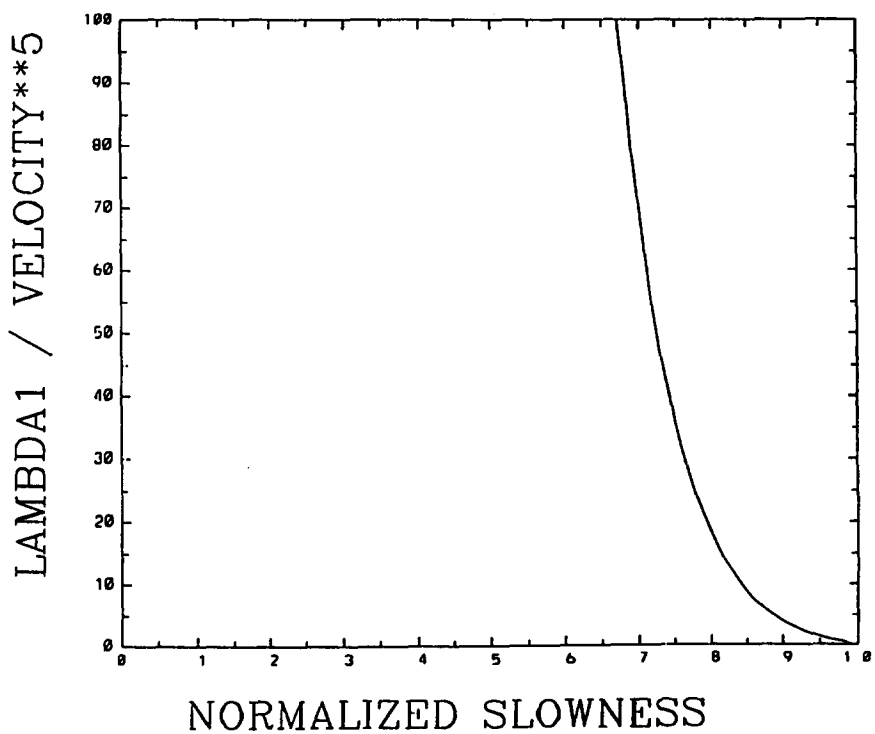


Figure 3. The coefficient λ_1 (display (19)), normalized to remove velocity dependence.

Thus (since (14) implies $\delta r \circ \tau = 0, z < z_0$):

$$\begin{aligned} \|\delta r \circ \tau\|^2 &= \int dz |\delta r(\tau(z, p), p)|^2 = \frac{1}{2} c_1 \Lambda_1(p) \int dt |\delta r(t, p)|^2 \\ &= \frac{1}{2} c_1 \Lambda_1(p) \|\delta r\|^2. \end{aligned} \tag{16}$$

Now set $p_1 = p, p_2 = \frac{1}{2} p_{\max} + p$, and integrate \sqrt{p} times the first summand on the r.h.s. of (14), from $p = 0$ to $p = \frac{1}{2} p_{\max}$, (for fixed z) neglecting the factor of c^{-2} :

$$\begin{aligned} \pm \int_0^{\frac{1}{2} p_{\max}} dp \sqrt{p} (\Lambda_1^{-2} \delta r \circ \tau) \Big|_p^{p + \frac{1}{2} p_{\max}} &\leq \int_0^{\frac{1}{2} p_{\max}} dp \sqrt{p} \{ \Lambda_1^{-2} |\delta r \circ \tau| \Big|_{p + \frac{1}{2} p_{\max}} + \Lambda_1^{-2} |\delta r \circ \tau|_p \} \\ &\leq \int_0^{p_{\max}} dp \sqrt{p} \Lambda_1^{-2} |\delta r \circ \tau| \\ &\leq \left(\int_0^{p_{\max}} dp \Lambda_1^{-3} \right)^{\frac{1}{2}} \left(\int_0^{p_{\max}} dp p \Lambda_1^{-1} |\delta r \circ \tau|^2 \right)^{\frac{1}{2}}. \end{aligned} \tag{17}$$

Applying the same step to the absolute value of the l.h.s. of (14) gives

$$c_1^{-4} \left\{ 2\delta c c_r + \left(\int_0^z \delta c \right) \left(\frac{\partial c_r}{\partial z} \right) \right\} \lambda(c_1, p_{\max}), \tag{18}$$

where

$$\lambda(c_1, p_{\max}) := \int_0^{\frac{1}{2} p_{\max}} dp \sqrt{p} \left(\Lambda_1^2 \left(p + \frac{1}{2} p_{\max} \right) - \Lambda_1^2(p) \right).$$

The modulus $\lambda(c_1, p_{\max})$ satisfies $\lambda(0) = 0$ and $\lambda(c_1, p_{\max}) \rightarrow \infty$ as $p_{\max} \rightarrow 1/c_1$ (since the integrand has a nonintegrable singularity at $p_{\max} = 1/c_1$). Otherwise, λ is best investigated numerically; a plot of λ vs. c_1, p_{\max} appears as Figure 2. Clearly, in order that $\lambda \gg 0$, it is necessary that $c_1 p_{\max}$ be relatively close to 1.

Since (17) is valid with either sign, we can square both sides of the inequality we have just derived from (14), (17), (18), and make an obvious estimate of the second term on the r.h.s. of (14) to get:

$$\begin{aligned} \left\{ 2\delta c c_r + \left(\int_0^z \delta c \right) \left(\frac{\partial c_r}{\partial z} \right) \right\} &\leq \lambda(c_1, p_{\max})^{-2} c_1^2 \left[\left(\int_0^{p_{\max}} dp \Lambda_1^{-3} \right) \int_0^{p_{\max}} dp p \Lambda_1^{-1} |\delta r \circ \tau|^2 \right. \\ &\quad \left. + \frac{1}{4} c_1^2 p_{\max}^3 \int_0^{p_{\max}} dp p |\delta W \circ \tau|^2 \right], \end{aligned}$$

integrate both sides in z , and make use of (16):

$$\left\| 2\delta c \Delta c + \left(\int_0^z \delta c \right) \left(\frac{\partial \Delta c}{\partial z} \right) \right\|^2 \leq \lambda_1(c_1, p_{\max}) \|\delta r\|^2 + \lambda_2(c_1, p_{\max}) \|\delta C\|^2, \tag{19}$$

where λ_1 is given by

$$\lambda_1(c_1, p_{\max}) = \frac{1}{2} c_1^3 \lambda(c_1, p_{\max})^{-2} \int_0^{p_{\max}} \Lambda_1^{-3},$$

and

$$\lambda_2(c_1, p_{\max}) = \frac{1}{4} c_1^5 \lambda(c_1, p_{\max})^{-2} \int_0^{p_{\max}} \Lambda_1^{-3}.$$

A plot of λ_1 appears as Figure 3. Evidently:

In order that λ_1 be small, it is necessary that $c_1 p_{\max}$ be relatively large (near 1).

This is still not quite what is required, as we have an estimate on the left-hand side of (18), not on δc itself. The third step amounts to the observation that, if δc is smooth enough, and c_r rough enough, then this quantity cannot be small without δc itself being small.

This observation is the most physically-meaningful, and the most difficult to quantify, of the points presented in this paper. As illustrated thoroughly in [16], the local behaviour of the seismogram function near a rough model is quite different from that near a smooth model, and it is this property that is being used implicitly at this point.

A rigorous semiquantitative treatment of roughness is found in [27], and the present problem can be treated along the same lines. Such rigorous argument is neither particularly enlightening, however, nor very precise. For the present, a numerical illustration of this property will suffice. The smoothness of the velocity model c and its perturbation are guaranteed, for example, by insisting that

$$c(z) = c_0 \exp \left\{ \sum_{j=1}^N x_j \psi_j(z) \right\},$$

$$\delta c(z) = \sum_{j=1}^N \delta x_j \psi_j(z) \cdot c(z),$$

where $\{\psi_j\}$ is a finite, smooth set of basis functions. In the tests reported in Section 4, cubic b -spline integrals are selected for the $\{\psi_j\}$: these have associated with them a definite length scale (“width”); see Figure 4. Then in vector notation,

$$\|\delta c_{c_r}\|^2 = \delta x^T R^0 \delta x$$

where

$$R_{ij}^{(0)} = \int dz \psi_i(z) \psi_j(z) c^2(z) c_r^2(z)$$

is the $N \times N$ symmetric positive-semidefinite “roughness matrix” associated with c_r and the basis $\{\psi_j\}$.

Similarly,

$$\|\delta c\|^2 = \delta x^T M \delta x$$

and

$$\left\| \left(\int_0^z \delta c \right) \frac{\partial c_r}{\partial z} \right\|^2 = \delta x^T R^{(1)} \delta x,$$

where

$$R_{ij}^{(1)} = \int dz \left(\int_0^z \psi_i \right) \left(\int_0^z \psi_j \right) c^2(z) \left(\frac{\partial c_r}{\partial z}(z) \right)^2$$

and M is the symmetric positive-definite “mass matrix”

$$M_{ij} = \int dz \psi_i(z) \psi_j(z) c^2(z).$$

It follows that

$$\mu_{\max}^{(0)} \|\delta c\|^2 \geq \|\delta c_{c_r}\|^2, \tag{20a}$$

where $\mu_{\max}^{(0)} \geq 0$ is the largest eigenvalue of the N -dimensional generalized eigenvalue problem

$$R^{(0)} \mathbf{y} = \mu M \mathbf{y}. \tag{20b}$$

Likewise,

$$\mu_{\min}^{(1)} \|\delta c\|^2 \leq \left\| \left(\int_0^z \delta c \right) \frac{\partial c_r}{\partial z} \right\|^2, \tag{20c}$$

where $\mu_{\min}^{(1)}$ is the least eigenvalue of the generalized eigenvalue problem

$$R^{(1)} \mathbf{y} = \mu M \mathbf{y}. \tag{20d}$$

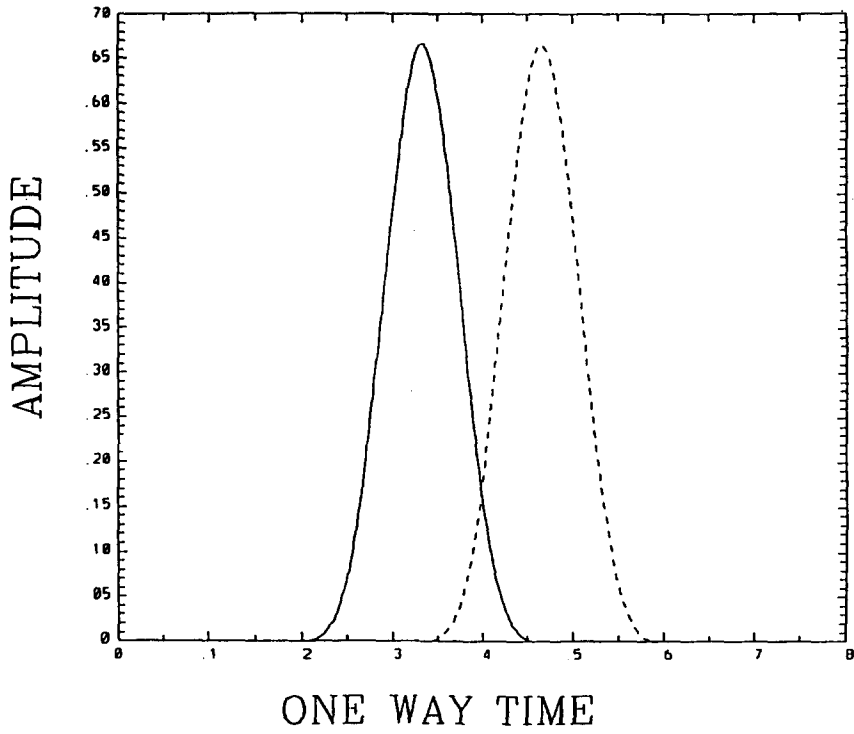


Figure 4a. Cubic *b*-splines: 5th (solid) and 7th (dashed) of 12 nodes in the interval.

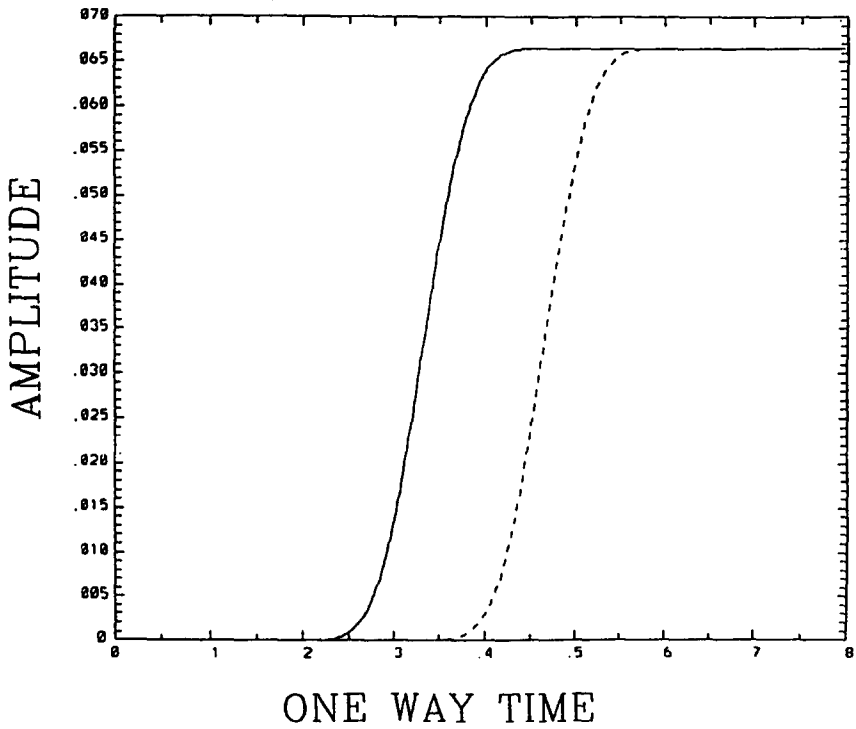


Figure 4b. Indefinite integrals of splines in Figure 4a.

Finally,

$$\mu_{\min} \|\delta c\|^2 \leq \left\| 2\delta c c_r + \left(\int_0^z \delta z \right) \frac{\partial c_r}{\partial z} \right\|^2, \quad (20e)$$

where μ_{\min} is the least eigenvalue of the generalized eigenvalue problem

$$R y := \left(2R^{(0)} + R^{(1)} \right) y = \mu M y. \quad (20f)$$

Combining (19), (20a), (20c) and (20d) we get

$$\mu_{\min} \|\delta c\|^2 \leq \lambda_1(c_1, p_{\max}) \|\delta r\|^2 + \lambda_2(c_1, p_{\max}) \|\delta W\|^2 \quad (21a)$$

and

$$\mu_{\min} \geq \left(\mu_{\min}^{(1)} - 4\mu_{\max}^{(0)} \right). \quad (21b)$$

Obviously, inequality (21b) has force only if $\mu_{\min}^{(1)} \gg \mu_{\max}^{(0)}$, that is, $\partial c_r / \partial z$ is much bigger than c_r , whence c_r must be *rough*. In fact, c_r must be *uniformly rough* on the length scale of significant variation in c . To see this, note that the worst possible situation is when $R^{(0)}$ and $R^{(1)}$ have a common null vector, since then the left-hand side of (19) constrains the corresponding component not at all. This can indeed occur. Note that the integrated spline basis $\{\psi_j\}$ is so constructed that $\psi_j - \psi_{j+1}$ vanishes outside an interval I_j , encompassing five spline nodes: see Figure 5. Suppose that for some j , $c_r \equiv 0$ in the interval I_j : that is, there are no reflectors in I_j . Set $\delta x_j = 1$, $\delta x_{j+1} = -1$, $\delta x_i = 0$ for $i \neq j, j+1$. Then $R^{(j)} \delta x = 0$, $j = 0, 1$, so $\mu = 0$ with eigenvector δx for both matrices. Thus:

In order that $\mu_{\min}^{(1)} > 4\mu_{\max}^{(0)} > 0$, significant reflectors must be present in every depth interval of the characteristic length scale of the smooth model class.

In fact, it turns out that this condition is also necessary in order that the “combined” eigenvalue μ_{\min} be reasonably large, as we shall see below.

Note also the connection with the wavelet passband. In order for the convolutional model to be accurate, c must have almost all of its energy concentrated below the passband (or rather its spatial analogue): that is, c must be smooth on the spatial wavelength scale. Suppose that c is chosen to attain, roughly, the maximum degrees of freedom permitted by this constraint. Thus, the length scale associated with the background model is roughly the largest spatial wavelength in the data, and we can re-phrase the above conclusion as:

In order that (19) above constrains δc , significant reflectors must be present in every depth interval longer than the longest spatial wavelength in the data.

This necessary condition is actually also sufficient, when made appropriately precise ([27]; see also [16]). We illustrate the sufficiency by introducing the reflector sequence c_r figuring in the experiments of Section 4. See Figure 6. We have computed the extreme eigenvalues of the problems, (20b), (20d), (20e). For an N -dimensional integrated spline space, the characteristic length scale (i.e., the length of I_j) is proportional to $1/N$. In Figure 7, we tabulate $\mu_{\max}^{(0)}$, $\mu_{\min}^{(1)}$ and μ_{\min} against N , along with the characteristic length scale; for the background velocity profile we have taken the profile in Figure 6 also. We have not restricted the non-zeros of δc to the $c = \text{const.}$ segment, so this is a more severe test than warranted by the present discussion. Evidently, the criterion above is indeed sufficient to guarantee $\mu_{\min}^{(1)} - 4\mu_{\max}^{(0)} > 0$ in this case and appears necessary to have $\mu_{\min} \gg 0$.

We now combine the conclusions of three steps: (13), (19), (20e), to get

$$\frac{\mu_{\min}}{\gamma} \|\delta c\|^2 \leq \delta^2 \tilde{J}, \quad \gamma = \frac{\lambda_1}{K_2 K_1^2} + \frac{\lambda_2}{\sigma^2},$$

and, using (15) again,

$$\min \left(\frac{1}{2} K_2 K_1^2, \frac{\mu_{\min}}{2\gamma} \right) (\|\delta c\|^2 + \|\delta r\|^2) \leq \delta^2 \tilde{J}, \quad (22)$$

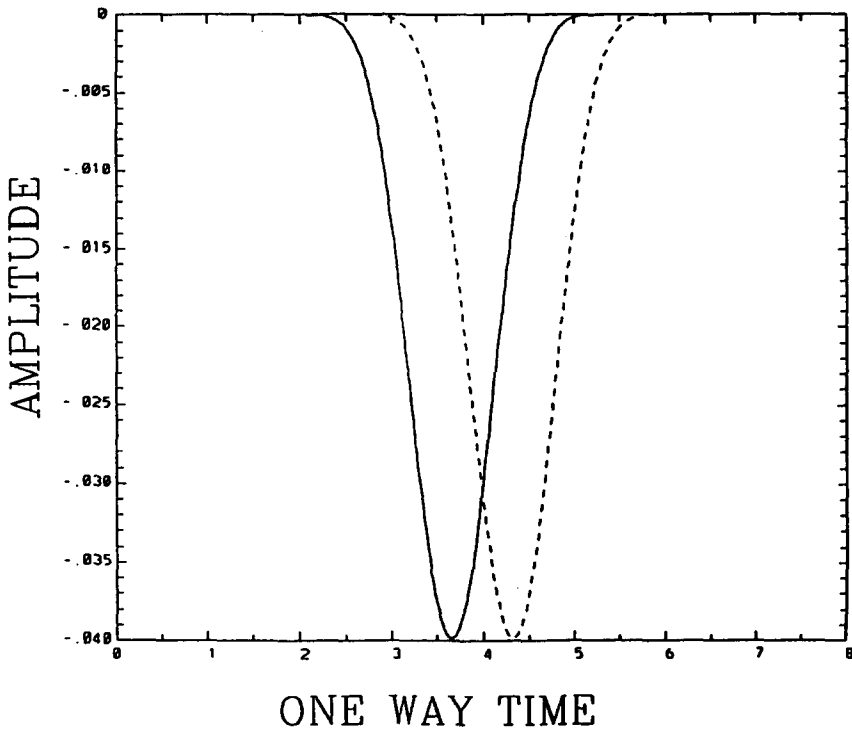


Figure 5. Differences: 6th-5th (solid), 7th-6th (dashed) spline integrals.

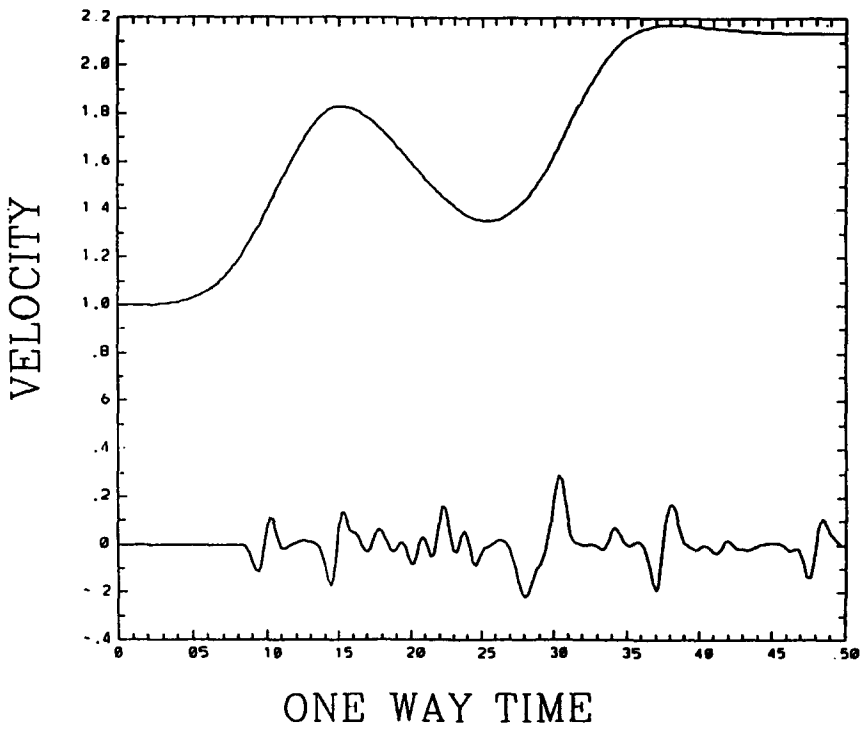


Figure 6. Smooth velocity model (upper curve) and velocity perturbation (lower curve) used to generate data for coherency optimization experiments.

| Nodes | $\mu_{\max}^{(0)}$ | $\mu_{\min}^{(1)}$ | μ_{\min} |
|-------|--------------------|--------------------|--------------|
| 5 | 0.8465e-02 | 0.1096e+01 | 0.1111e+01 |
| 6 | 0.1305e-01 | 0.5551e+00 | 0.5741e+00 |
| 7 | 0.1211e-01 | 0.3650e+00 | 0.3872e+00 |
| 8 | 0.1644e-01 | 0.1900e+00 | 0.2192e+00 |
| 9 | 0.1798e-01 | 0.5334e-01 | 0.8287e-01 |
| 10 | 0.1753e-01 | 0.3068e-01 | 0.4868e-01 |

Figure 7. Eigenvalues $\mu_{\max}^{(0)}$, $\mu_{\min}^{(1)}$, and μ_{\min} as in displays (20) and (21).

which gives an explicit estimate for the constant L in (10), as required. Note that this constant of proportionality is neatly separated into μ_{\min} , which depends only on the roughness of c_r , relative to c_s , and γ , which depends on aperture and passband amplitude, and on the penalty parameter σ .

Note the role of σ in (22): it is only the quotient λ_2/σ^2 which influences the lower bound. The aperture-dependent λ_1 and λ_2 are of roughly the same size, so the requirement that the lower bound be reasonable imposes a (rather loose) relation between the aperture, the passband amplitudes (K_1, K_2), and σ : in effect, a (soft) floor is placed under σ . On the other hand, the *upper* limit of the spectrum of $\delta^2 J$ clearly grows with σ^2 . Thus for “moderate” values of σ , the $\delta^2 J$ is as well-conditioned as possible. Several possibilities exist for pinning down this range: guessing (the option pursued in numerical work reported here), rigorous theoretical estimation (far too conservative), or adaptive estimation during iterative solution of the linear stage of Newton’s method (a project for the future).

Recall that the argument leading to this conclusion was based on the restriction (13). Without this restriction, the argument becomes more involved, but the conclusion is qualitatively the same: for sufficiently rough c_r , an inequality like (22) holds. It is evident from the form of the expression $Q \delta c_s$ that $c'_s \neq 0$ will degrade the contribution of large- p traces to the lower bound, and this is indeed the main quantitative effect. A rigorous argument of this nature, in the context of the fully nonlinear problem, can be found in [27].

4. NUMERICAL EXPERIMENTS

In this section, we report an implementation of the algorithm suggested in Section 2, and the results of some numerical experiments which establish its feasibility. More extensive tests of the method will be reported elsewhere.

The implementation requires (1) a choice of discretization of the parameter space and operators; (2) a choice of optimization strategy. We give sufficient detail in both areas that the interested reader should be able to reproduce our results.

4.1. Discretization

Both to maintain the applicability of the convolutional model, and for the more subtle reasons given in Section 3, it is necessary that the background velocity c be quite smooth—much smoother than the reflectivity, for example. A simple explicit way (though certainly not the only way) to ensure a given degree of smoothness is to choose c from a finite-dimensional function space spanned by smooth functions.

For the space velocity profiles, we took a manifold of exponentiated integrated b -splines (see Figure 4):

$$c_s(z) = c_0 \exp \left\{ \int_0^z dz' \sum_{i=1}^{nmod} x_i \psi_i(z') \right\},$$

where $\psi_i(z) = \Psi(nspl \cdot (z - z_{i+1})/zspl)$, Ψ is a standard b -spline, and $z_i = (i/nspl) \cdot zspl$, $i = 0, \dots, nspl$ are $nspl+1$ evenly spaced nodes. We set $nmod = nspl-3$, so that all summands vanish at $z = 0$ and $z = zspl$. See Figure 4. The surface velocity c_0 was regarded as a fixed parameter.

The perturbations δc have the form

$$\delta c_s(z) = \left(\sum_{i=1}^{nmod} \delta x_i \int_0^z \psi_i \right) c_s(z).$$

When values $z > zspl$ are needed, both c and δc are regarded as constant and equal to their values at $zspl$.

The depth functions occurring in the various formulae are sampled on a fixed grid $\{j dz : 0 \leq j \leq nz\}$, where $nz \cdot dz =: z_{max} \geq zspl$ ("the z -grid"). Routines were written which convert sampled z -grid functions ($c_s, \delta c_s$) into spline coefficients ($x, \delta x$) and vice versa, and which satisfy certain adjointness conditions detailed below.

The time-slowness sections are regarded from the outset, as sampled on a fixed grid of size $(nt+1) \times np$, at given sample intervals dt and dp . We used *one-way time* as the time parameter throughout.

4.2. Operators

The formulas for the incoherency C and its perturbation δC from the Appendix and Section 2 require the travel-time change-of-variable and its inverse. These were accomplished *via* interpolation. For example, the integral

$$\tau(z, p) = \int_0^z \frac{dz'}{v_s(z', p)}$$

is approximated using the trapezoidal rule on the z -grid, yielding an unequally spaced set $\{\tau_i\}$ of travel-times. The time-function to be converted to a depth function must then be evaluated at the τ_i , which is done *via* piecewise-linear interpolation. The total process has a truncation error on the order of dz^2 .

The derivatives with respect to z , t and p are replaced by simple 3-point centered difference formulas, maintaining the truncation order. Discretization of W and δW was accomplished by means of the formulas indicated in the Appendix.

The synthetic p -tau seismogram $\tilde{S}[c, r]$ was computed using a centered 3-point t -difference and the trapezoidal rule approximation of the convolution integral.

4.3. Norms

The definition of the cost function \tilde{J} involves the L^2 -norm. As noted in Section 2, this should really include a factor of p , to most closely maintain the relation with mean-square error in (x, t) domain. To simplify our calculations, however, we ignored this point and defined the section L^2 norms by

$$\|F\|^2 = \sum_{j=0}^{nt} \sum_{k=1}^{np} dt dp w_t(j) w_p(k) |F_{jk}|^2,$$

where

$$w_t(j) = \begin{cases} 1/2, & \text{if } j = 0, nt, \\ 1, & \text{else,} \end{cases}$$

$$w_p(k) = \begin{cases} 1/2, & \text{if } k = 1, np, \\ 1, & \text{else,} \end{cases}$$

i.e., the trapezoidal rule. With this choice, \tilde{J} is completely defined.

The importance of choice of norms in the *model* space $[c_s, r]$ cannot be overemphasized. Even though the *definition* of \tilde{J} is independent of this choice, it is the main factor affecting the *efficiency* of the optimization. This is principally because of the role of the model norm in the definition of the gradient. We examine this point for the summand $\sigma^2 |||W[c, r]|||^2$ of \tilde{J} , since it involves both c and r . By definition, the gradient of this term is the (unique) model vector $[\dot{c}, \dot{r}]$ for which for any $[\delta c, \delta r]$,

$$\lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} (\sigma^2 |||W[c_s, +\epsilon \delta c_s, r + \epsilon \delta r]|||^2 - \sigma^2 |||W[c_s, r]|||^2) = \langle (\dot{c}_s, \dot{r}), (\delta c_s, \delta r) \rangle_M, \quad (23)$$

where $\langle \cdot, \cdot \rangle_M$ is the scalar product in model space corresponding to the norm:

$$||[\delta c_s, \delta r]||_M^2 = \langle [\delta c_s, \delta r], [\delta c_s, \delta r] \rangle_M.$$

(We assume that the model space is a Hilbert space, since all efficient smooth optimization methods are predicated on this assumption.)

A principal requirement ((i) in Section 3) is that the function $[c_s, r] \rightarrow |||W[c, r]|||^2$ is regular, i.e., differentiable. Examining (A.2), we see that *derivatives* of δr are involved in δW . Since $r, \delta r$ are to be allowed to be arbitrary (grid-representable) functions, δW can only be continuously dependent on δr , as is required by regularity, if the model norm includes explicit control over derivatives of δr . The obvious choice for the section part of the model norm is

$$||\delta r||_1^2 = \sum_{j=0}^{nt} \sum_{k=1}^{np} dt dp w_t(j) w_p(k) \{ |\delta r_{jk}|^2 + |D_t \delta r_{jk}|^2 + |D_p \delta r_{jk}|^2 \},$$

where D_t and D_p are 2-point one-sided difference approximations for $\partial/\partial t$ and $\partial/\partial p$. The subscript "1" stands for "first derivatives"; this is the discrete version of the first in the Sobolev scale of norms, a basic tool in modern analysis of partial differential equations.

For the velocity profile part, i.e., δc , we can make use of the fact that c is required to belong to a space of smooth splines.

We enforce the membership of c in the spline manifold by parameterizing the model in terms of the spline coefficients x_i themselves, rather than the z -grid values of c . Also, we tacitly use $\log c_s$ and its perturbation, $\delta c_s/c_s$, as fundamental quantities, rather than c_s and δc_s ; this nondimensionalizes that part of the model (note that r is already non-dimensional, by definition). Thus, we need to express the norm of $\delta c_s/c_s$ in terms of the δx_i ; this is easily accomplished via the *mass matrix*

$$M_{ij} = \int dz \psi_i \psi_j.$$

This is computed *via* the trapezoidal rule, of course. Then

$$\left\| \frac{\partial}{\partial z} \frac{\delta c_s}{c_s} \right\|_{L^2}^2 = \delta x^T M \delta x.$$

Note that we are measuring $(\partial/\partial z)(\delta c_s/c_s)$ rather than $(\delta c_s/c_s)$ itself. Since $\delta c_s(0) = 0$, a bound on the former implies a bound on the latter, but the choice we have made here weights the more oscillatory velocity perturbations more highly, and, thus, tends to rotate the gradient in the direction of smoother δc_s 's, with favorable computational consequences.

Now regarding the model as the pair $[x, r]$ (rather than $[c, r]$), the model norm is taken to be

$$||[\delta x, \delta r]||_M^2 = \mu_c \delta x^T M \delta x + \mu_r ||\delta r||_1^2.$$

Adjustment of the weights μ_c, μ_r allows the gradient to be rotated in the x - or r -directions; this scaling-preconditioning turns out to be important in achieving rapid convergence.

4.4. Gradients, Hessians

First examine the incoherency component of \tilde{J} , as before. To write the result in a revealing way, recall that δW is linear in $[\delta x, \delta r]$, and write $\delta W \cdot [\delta x, \delta r]$ for the value. Then the limit on the l.h.s. of (23) can be carried out to give

$$2\sigma^2 \langle \delta W \cdot [\delta x, \delta r], W[x, r] \rangle_{L^2} = \langle [\delta x, \delta r], 2\sigma^2 \delta W^* \cdot W[x, r] \rangle_M, \quad (24)$$

where the *adjoint operator* δW^* is defined by the condition

$$\langle \delta W[\delta x, \delta r], F \rangle_{L^2} = \langle [\delta x, \delta r], \delta W^* \cdot F \rangle_M, \quad (25)$$

which is to hold for arbitrary model perturbations $[\delta x, \delta r]$ and (t, p) -sections F .

Comparison of (23) and (24) reveals that

$$\begin{aligned} \text{grad}(\sigma^2 |||W|||^2) &= [\dot{x}, \dot{r}] \\ &= 2\sigma^2 \delta W^* \cdot W. \end{aligned} \quad (26a)$$

Likewise, the Gauss-Newton approximate Hessian operator [35, Section 10.2] is given by

$$\text{Hess}(\sigma^2 |||W|||^2) \cdot [\delta x, \delta r] = 2\sigma^2 \delta W^* \cdot \delta W \cdot [\delta x, \delta r]. \quad (26b)$$

Similar formulas hold for the term $|||S - S_{\text{data}}|||^2$.

The calculations (26) are the principal parts of the quasi-Newton methods to be introduced below. Thus efficient and accurate calculation of the adjoints δW^* , δS^* are essential to a successful optimization.

4.5. Adjoint

The definition (25) of δW^* must be taken absolutely seriously. That is, even though δW is given (in principle) by an enormous matrix, δW^* is *not* simply the operator defined by the matrix transpose of δW . In fact, δW^* is *related* to the matrix transpose and this relation provides a convenient avenue for computing δW^* .

The scalar products involved in (25) may be written symbolically in the form

$$\langle X, Y \rangle_G = X^T G Y,$$

where X and Y are parameter vectors and G is the *Gram matrix* of $\langle \cdot, \cdot \rangle_G$. Thus, G is a positive-definite symmetric matrix of size equal to the dimension of the parameter space.

In the cost of the functional \tilde{J} , two essentially different inner products are involved, on two different parameter spaces (models $[\delta x, \delta r]$, sections F), as well as a linear transformation (δW) mapping one parameter space into the other. Accordingly, consider two inner products of the form given above, with Gram matrices G_ν , $\nu = 1, 2$, and a linear transformation A , given by a matrix of appropriate dimensions, mapping one parameter space into the other. The *adjoint* of A is defined by:

$$\langle A X_1, X_2 \rangle_{G_1} = \langle X_1, A^* X_2 \rangle_{G_2}$$

for arbitrary X_ν in the ν^{th} parameter space, $\nu = 1, 2$. Written in matrix form,

$$(A X_1)^T G_1 X_2 = X_1^T G_2 A^* X_2,$$

from which it is clear that, as matrices,

$$A^* = G_2^{-1} A^T G_1. \quad (27)$$

To see what (27) implies for (25), write δC in components:

$$\delta W = [\delta_x W, \delta_r W].$$

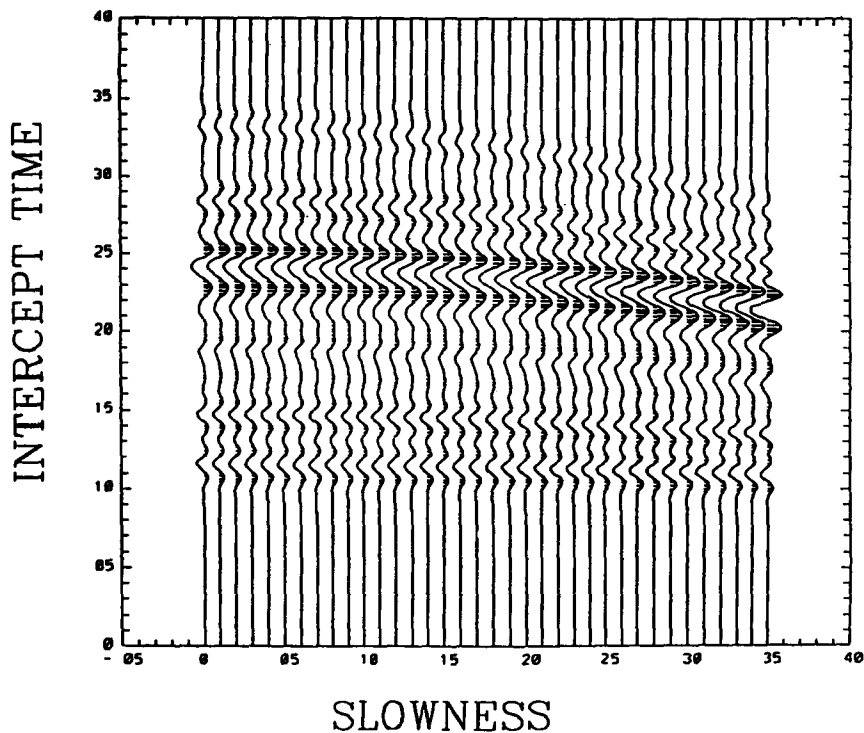


Figure 8. Plane wave section generated from the model displayed in Figure 6, using a Ricker wavelet peaked at 20 Hz.

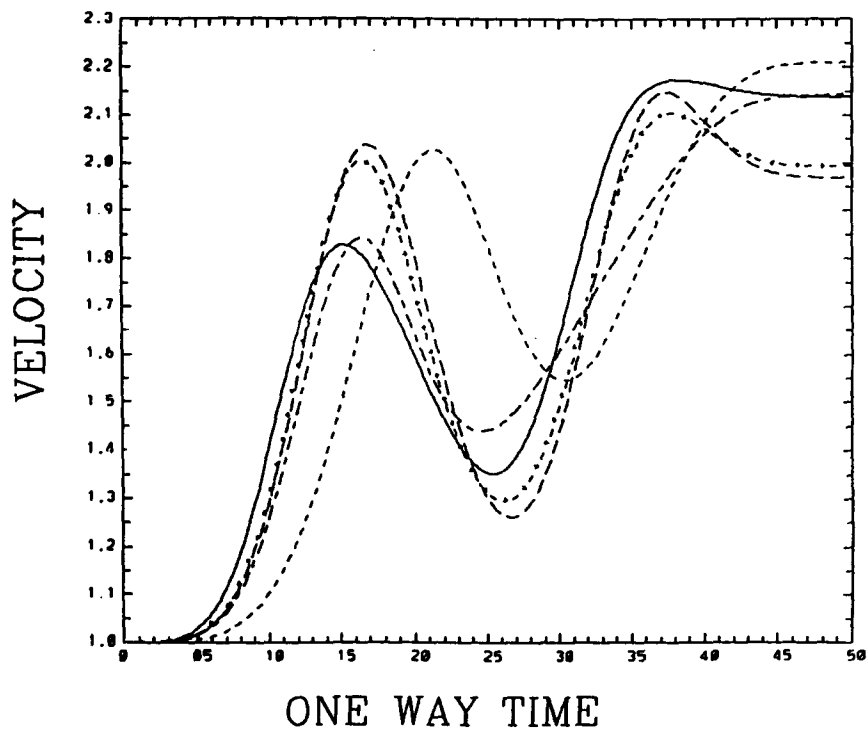


Figure 9. (0) target model, from Figure 6 (solid curve); (1) a non-constant initial model (short dashes); (2) result of 5 Gauss-Newton steps starting at constant model (long dashes); (3) result of 10 Gauss-Newton steps starting at constant model (dash/cross); (4) result of 5 Gauss-Newton steps starting at model (1) (short/long dashes).

This is the correct matrix representation if the model perturbation is written as a column vector:

$$\delta W \cdot \begin{bmatrix} \delta x \\ \delta r \end{bmatrix} = \delta_x W \cdot \delta x + \delta_r W \cdot \delta r.$$

Thus,

$$\delta W^* = \begin{pmatrix} \delta_x W^* \\ \delta_r W^* \end{pmatrix},$$

where

$$\langle \delta_x W \cdot \delta x, F \rangle_{L^2} = \delta x^T M (\delta_x W^* \cdot F) \quad (28a)$$

and

$$\langle \delta_r W \cdot \delta r, F \rangle_{L^2} = \langle \delta_r, \delta_r W^* \cdot F \rangle_1. \quad (28b)$$

Identify a section F with a vector in any convenient fashion, e.g., by listing the traces (columns) sequentially. Then the L^2 -inner product, discretized by the trapezoidal rule, is realized by the scaling matrix S (which scales the edge entries by $1/2$, the corner entries by $1/4$, and everything else by 1):

$$\langle F_1, F_2 \rangle_{L^2} = F_1^T S F_2$$

for any sections F_1, F_2 .

Thus (27) applied to (28a) gives

$$\delta_x W^* = M^{-1} \delta_x W^T S. \quad (29a)$$

For (28b), it is necessary to write the ‘‘Sobolev’’ inner product $\langle \cdot, \cdot \rangle_1$ in the canonical form given above; its Gram matrix turns out to be *exactly the matrix of the discrete Neumann problem* for the usual five-point discretization of the Laplace operator, which we shall denote by N . Thus

$$\delta_r W^* = N^{-1} \delta_r W^T S. \quad (29b)$$

In principle, this completes the calculation of the adjoints, hence of the gradient and Hessian. In practice, inspection of (26a,b) shows that we need only routines which apply δW^* to a vector $[\delta x, \delta r]$, *not* the entire matrix of δW^* . This extremely important observation saves much computational effort and storage. In fact, application of the trapezoidal scaling operation represented by S is trivial, and the transpose operations $\delta_x W^T, \delta_r W^T$ are relatively easy to work out, as the same sort of recurrence rules that form the ‘‘forward’’ calculations of $\delta_x W, \delta_r W$. Thus we represent (29a,b) in the alternate forms: for an arbitrary section F ,

$$\begin{aligned} \delta_x W^* \cdot F &= \dot{x}, \\ M \dot{x} &= \delta_x W^T S F, \end{aligned} \quad (30a)$$

$$\begin{aligned} \delta_r W^* \cdot F &= \dot{r}, \\ N \dot{r} &= \delta_r W^T S F. \end{aligned} \quad (30b)$$

We have just described how to compute the right-hand sides of the second equations in each of these pairs. To solve the linear system with the spline mass matrix M , we used a standard linear equation solver (LINPACK: SPOFA, SPOSL), as the spline space is small-dimensional—i.e., the background model has few degrees of freedom, ≤ 20 in all of our experiments. To solve the discrete Neumann problem, which is quite large ($nt = 300$, $np = 40$ in some experiments), we took advantage of explicit knowledge of the discrete Neumann eigenfunctions (tensor-product cosines) to design an FFT-based discrete Neumann solver, which solves the second member of (30b) very efficiently.

This step would have been more difficult if the factor p had been included in the integral defining $\langle \cdot, \cdot \rangle_1$, as it should be. Then N would include a p -difference approximation to the Bessel operator of order zero, with Neumann condition at $p = p_{\max}$ (a residue of the cylindrical geometry implicit in the proper definition of the Radon transform). Thus, fast Bessel transform software would be required.

4.6. Interface with the Optimizer

Since the optimizer described below accepts standardized n -vector arguments, it was necessary to “bundle” the computations just described into procedures with standard calling sequences. We used a pack/unpack routine, which collapses the spline/section pair $[\delta x, \delta r]$ into a vector of length $n_{spl} + (nt + 1) * np$, and vice versa.

4.7. Choice of Optimizer

The tests reported below were made using a so-called truncated Newton code. This code is based on the model trust region principle [35, Section 6.4] and on the extensions to it introduced by Steihaug in his Yale thesis [36]. Essentially, the Gauss-Newton linear step

$$\text{Hess } \tilde{J} \cdot [\delta x, \delta r] = -\text{grad } \tilde{J}$$

is solved by a conjugate residual iteration [37, Chapter 10], which is terminated when the step estimate exits a ball about the current solution estimate, the radius of which is determined by a simple and robust updating strategy. This expedient avoids expensive conjugate residual steps taken outside the region in which the linearized model can be “trusted,” hence the name. A more lengthy description of the code can be found in [16, Chapter 9], where the same codes were used in solving the nonlinear output least-squares problem, using finite difference synthetic seismograms instead of the convolutional model.

An important amendment of the trust region idea is natural in this problem. All models in the iteration are supposed to have a fixed rectangle $[0, t_{\max}] \times [0, p_{\max}]$ as a precritical set. This may cease to be true during an update step, if the velocity is increased by too much at some depth. The computation of the gradient simply flags this occurrence, and the algorithm attempts a smaller step in the same direction. Thus the trust region, for problems like the present one, may involve in a natural way constraints on the validity of the model itself.

4.8. Numerical Experiments

We performed a number of numerical experiments using the velocity profile c (upper curve) and perturbation c_r (lower curve) exhibited in Figure 6 to generate the p -tau convolutional model data of Figure 8, by convolving with a Ricker wavelet with center frequency 20 Hz. A target background velocity with a lower velocity zone was chosen because the structure of such a zone is impossible to determine from refraction arrival times, and intrinsically more difficult for least-squares methods—see [16]. The velocity, hence the slowness, were normalized against the surface velocity, by changing the measure of depth to normal-incidence time, for a constant background velocity equal to the surface velocity. This step also normalizes the slowness to the range $0 \leq p \leq 1$. A happy side-effect of this normalization was to reduce substantially the numerical imprecision resulting from mis-scaling inherent in the use of physical units.

The algorithm explained in the preceding subsections was used to extract estimates of c_s and r from the data of Figure 8. Parameters common to all experiments were

$$\sigma = 1, \quad \mu_c = 10^{-4}, \quad \mu_r = 1.$$

We found the small value of μ_c necessary to rotate the gradient of \tilde{J} toward the “ c_s ” direction.

In all cases, we observed the same pattern. We began with the simple estimate $c_{\text{initial}} = \text{const.} (= 1)$, $r_{\text{initial}} \equiv 0$. The first Newton step did not change the estimate of c_s , since the incoherence of $r \equiv 0$ vanishes for any velocity model. Otherwise put, since there are initially no reflectors, there is initially no move-out information in the reflectivity with which to update the velocity model. The first iteration is thus devoted entirely to minimizing

$$\left\| \left\| f * \frac{\partial r}{\partial t} - S_{\text{data}} \right\| \right\|^2$$

which amounts to deconvolving the data in a least-squares sense to find an initial (nontrivial) estimate for r . Unless otherwise noted, each Newton step (including the first) is approximated by five conjugate-residual iterations.

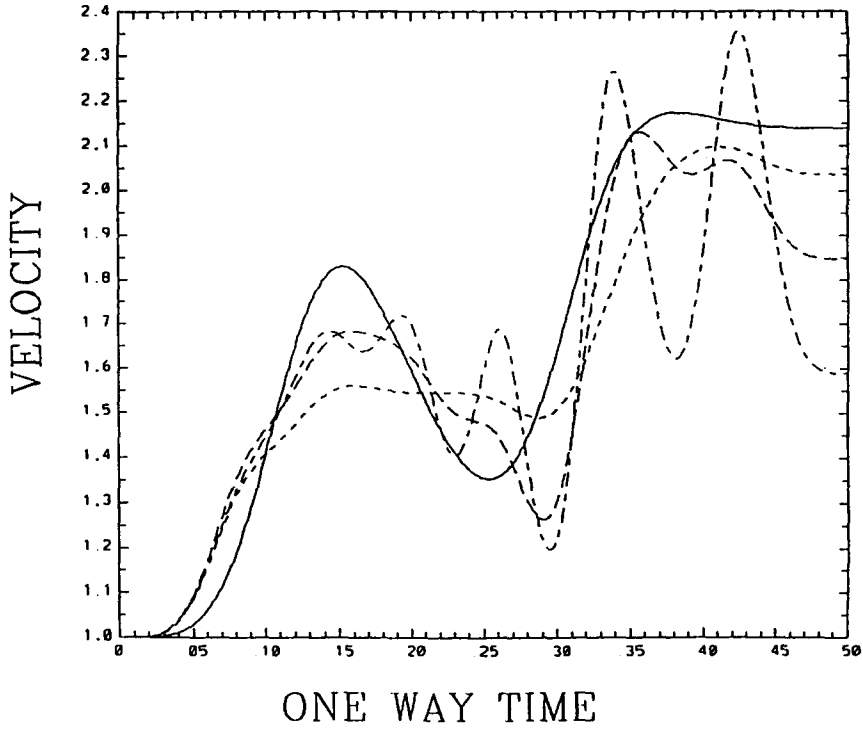


Figure 10. Results using too many degrees of freedom in the model (16 nodes): (0) target model (solid curve); (1) Gauss-Newton steps (short dashes); (2) 10 Gauss-Newton steps (long dashes); (3) 30 Gauss-Newton steps (short/long dashes).

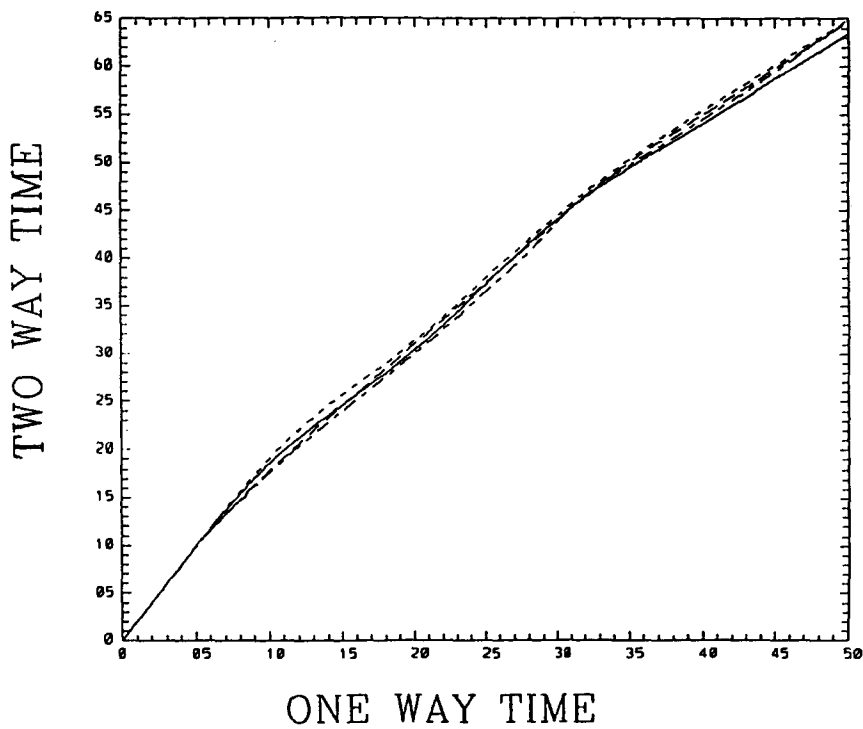


Figure 11. Two-way travel time curves at normal incidence: (0) target model (solid curve); (1) curve 2, Figure 9 (short dashes); (2) curve 1, Figure 10 (long dashes); (3) curve 2, Figure 10 (short/long dashes).

In the second and subsequent iterations, the velocity model is improved, and data-noise-generated incoherency in the reflectivity is reduced.

Figure 9 shows the velocity-estimate results of five and ten Newton steps from both the constant velocity model (curve 0) and five Newton steps from the velocity model identified as the curve 1. The results are labeled as curves 2, 3 and 4, respectively. In all cases we used 8 spline nodes (thus the velocity is determined by 5 parameters). This result is evidence for the independence of initial estimate of the final estimates. In particular, the results of ten iterations from a constant initial model, and five iterations from the "incorrect trend" initial model (curve 1) are virtually identical. The error is quite stable and is due to the fact that the target is not a member of the space of velocities defined by the eight-node spline basis—i.e., we get the "closest" eight-node velocity estimate.

These estimates are reasonably accurate, especially considering the computational work required. Apparently, the incoherence resulting from the erroneous basement velocity was insufficient to cause further corrections, or possibly, the calculation of the incoherence is substantially inaccurate there—see comments below. Most of the reduction in \bar{J} (about an order of magnitude) came in the first iteration, in which the data is deconvolved. Yet another order of magnitude is gained in the remaining iterations, in which the incoherence is reduced.

Two major points have emerged from the experimental work conducted thus far. The first concerns the number of spline nodes: the outcome is in some ways quite sensitive to this number. For example, the experiment of Figure 9 was repeated with 16 nodes instead. Five, ten, and thirty iterations of the Gauss-Newton process produced the curves labeled 1, 2, and 3 in Figure 10. These look quite "wild," and certainly the mean-square error is much greater than is the case with those in Figure 9. Recall, however, that the principal role of the background model is to supply travel-times. A quite different picture emerges when the travel-times are plotted against the "true" travel time curve (from the velocity profile of Figure 6). In Figure 11 are displayed (normal-incidence) travel-time curves from the "exact" velocity, curve 1 from Figure 9, and curves 1 and 2 from Figure 10. In fact, the latter two curves are closer to the "true" travel-time than is the former, despite their correspondence with velocity estimates having larger L^2 -error. This relation emerges more clearly when the (normal incidence) travel-time errors are plotted: see Figure 12.

Apparently, the result of increasing the number of degrees of freedom in the model is to allow a closer fit to the travel-time, at least at points corresponding to major reflectors, but at the cost of an oscillatory error which may be large in the mean-square sense. This is easy to understand: the errors oscillate on a length scale too short to affect the travel-times between major reflectors, hence correspond to *small eigenvalues* of the incoherence Hessian. While the effect on travel-times of this sort of error is *a priori* small, it does produce irritating ambiguities in the velocity estimate, and (more important) has a negative impact on the convergence of the iterative scheme. Several approaches to the removal of this ambiguity suggest themselves. Trial-and-error determination of the optimal spacing for spline nodes, as has been done here, is clearly unsatisfactory. Systematically increasing the number of nodes until a good fit is obtained requires some notion of an "acceptable" level of fit, and such information may itself only be obtainable by trial-and-error. A more satisfactory approach might be adaptive estimation of small Hessian eigenvalues, through the close relation of conjugate-residual iteration with the Lanczos algorithm [37, Chapter 10], and penalization of the corresponding eigenvector components. For the application of such "iterative deflation" to linear systems, see [38] (also [39]). Since the number of small eigenvalues associated with velocity perturbations is small, and since their characterization is somewhat independent of the current velocity estimate, this deflation strategy should work rather well in combination with the Gauss-Newton iteration. Finally, since the culprit is the oscillatory error, penalization of a velocity derivative may regularize this problem satisfactorily. Computational trials are in progress; results will be reported elsewhere.

A second point concerned the density of p -samples. Inspections of Figure 8 clearly show that for large p and t , the move-out difference in neighboring traces may be a substantial fraction of a wavelength. As we have based our difference approximations to the incoherency on centered difference approximations to the coordinate derivatives $(\partial/\partial t)$, $(\partial/\partial p)$, the possibility exists of severe undersampling in p . In fact, when we increased Δp to .02 (from .01 as in Figure 8), the computation analogous to that for Figure 9 gave completely erroneous results for the deeper part

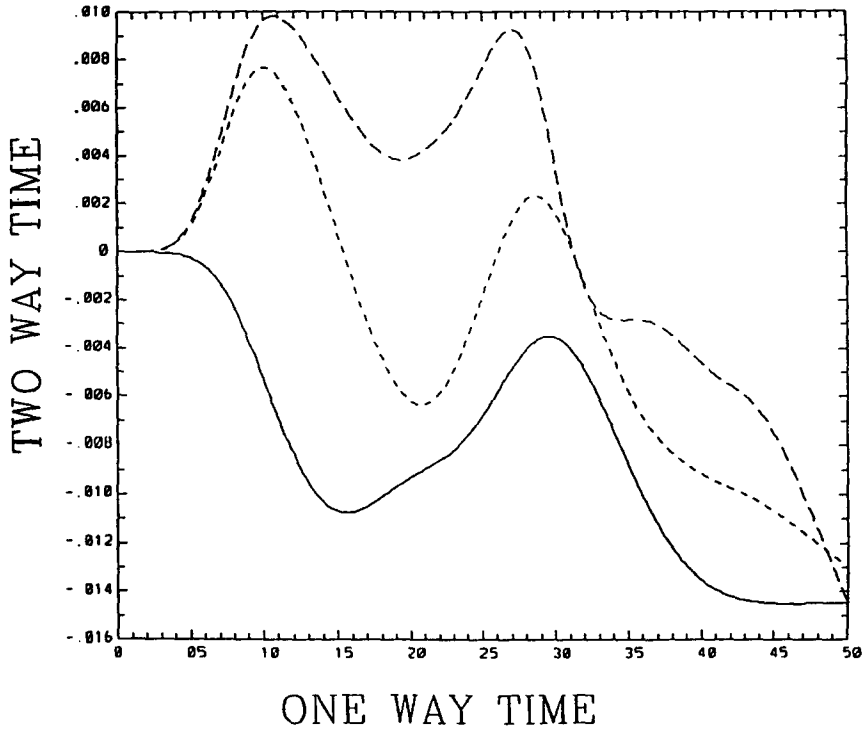


Figure 12. Two-way travel-time errors: (0) between curves (0) and (1), Figure 11 (solid line); (1) between curves (0) and (2), Figure 11 (short dashes); (2) between curves (0) and (3), Figure 11 (long dashes).

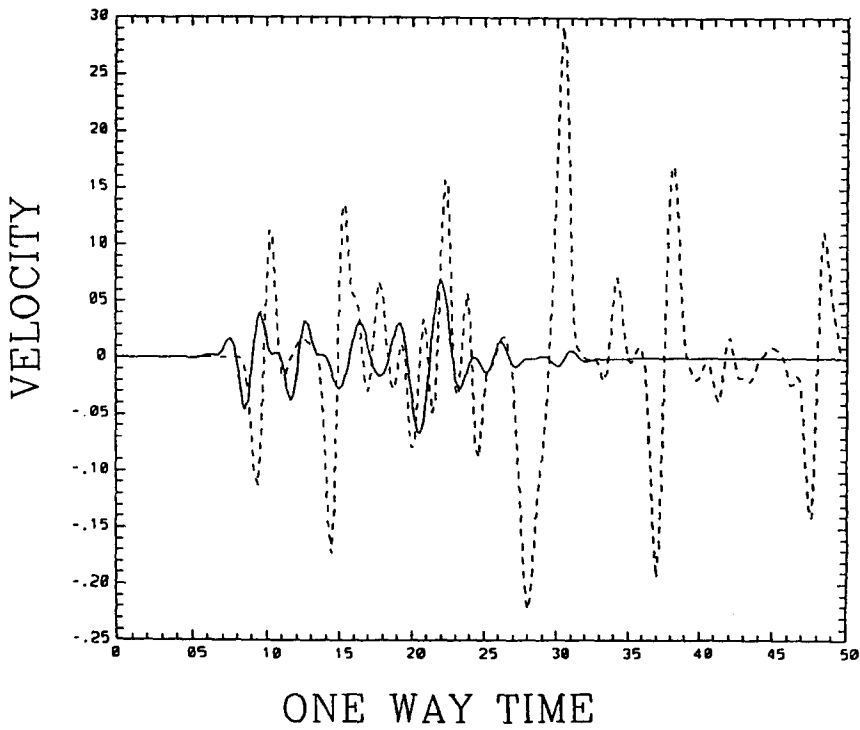


Figure 13. Stack (solid line) of the reflectivity estimate corresponding to Figure 9, curve 2, at constant velocity, compared with target velocity perturbation (dashed line).

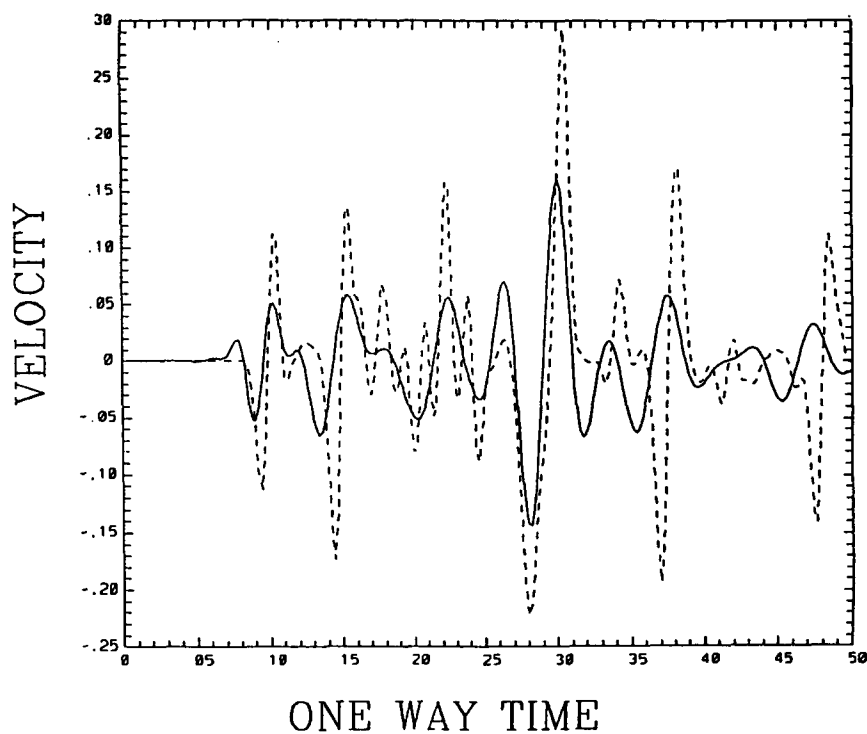


Figure 14. Stack (solid line) of the reflectivity estimate corresponding to Figure 9, curve 2, at a velocity given by Figure 9, curve 2, compared with the target velocity perturbation (dashed line).

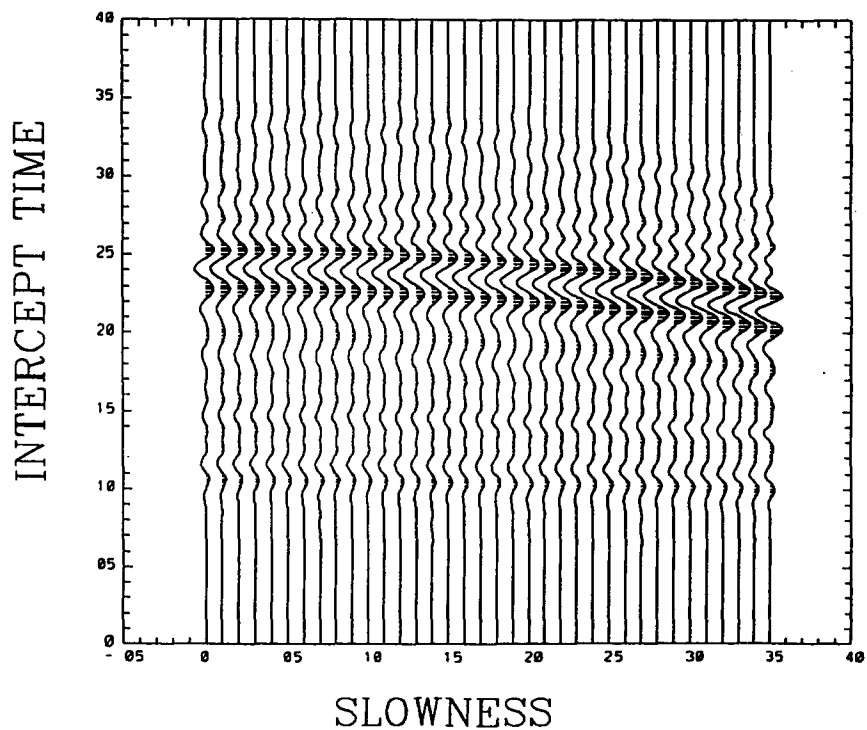


Figure 15. "Synthetic" section produced using the velocity from Figure 9, curve 2, with a stacked velocity perturbation estimate from Figure 14.

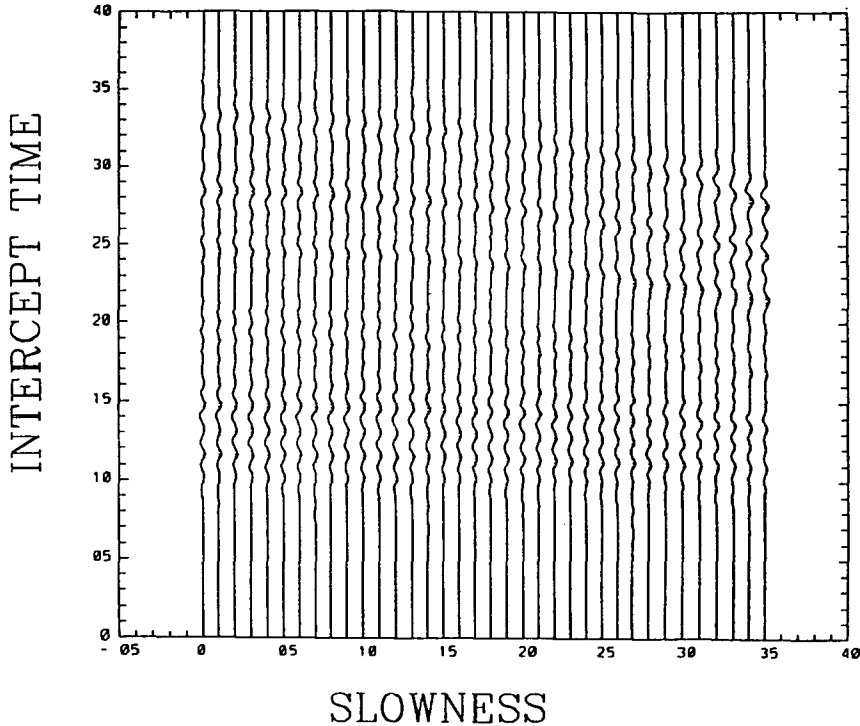


Figure 16. Difference of Figures 15 and 8.

of the velocity profile, apparently because the part of the incoherency due to deeper reflectors was grossly underestimated. We suspect that the residual inaccuracy in the deeper parts of the curves in Figure 9 is due to a milder version of the same effect.

Besides finer sampling, methods to overcome errors in incoherency due to undersampling include higher-order difference formulas and difference formulas better adapted to the move-out. Indeed, low-order differences along even a crude approximation to the correct move-out curve should produce better results at coarser sampling than the coordinate derivatives used in our present code. These ideas are also under investigation.

It might be objected that, while the output includes an estimate of the travel-time reflectivity section $r(t, p)$, no estimate of the corresponding velocity perturbation $c_r(z)$ is provided. The final reflectivity is not necessarily entirely coherent, and so does not correspond to any $c_r(z)$, strictly speaking. Nonetheless, an estimate may be produced by stacking $r(t, p)$ on the basis of equation (5), i.e.,

$$c_r(z) \sim \frac{c_s^3(z)}{p_{\max}} \int_0^{p_{\max}} dp v_s^{-2}(z, p) r(\tau(z, p), p).$$

This output might truly be regarded as the final image produced by an iterative before-stack migration, specialized to constant-density acoustics.

We have implemented this post-inversion stack of the final reflectivity estimate, and display the results for the reflectivity corresponding to curve 3 in Figure 9. When stacked with constant-velocity move-out, the estimate of $c_r(z)$ is completely erroneous in phase and wrong by a factor of perhaps 3 in amplitude (Figure 13). When stacked with the final velocity estimate from Figure 9 (curve 3), the estimated $c_r(z)$ has essentially correct phases and amplitudes for the major events (Figure 14), as compared to the $c_r(z)$ actually used to generate the data. Of course, none of the subwavelength-scale variations in the true $c_r(z)$ could appear in the reflectivity $r(t, p)$ or in the stacked estimate, so a more interesting comparison is the p -tau section generated using the velocity from Figure 9 and the stacked $c_r(z)$ from Figure 14. This is displayed in Figure 15, and the difference of Figure 15 and Figure 8, plotted on the same scale, in Figure 16.

4.9. Summary

We have described a preliminary implementation of the differential semblance optimization method, exhibited the results achievable with this rather crude code, and suggested some issues worthy of further examination. Given the rather large discretization errors in our present implementation, which play the role (at least) of data noise, the stability of the final estimates and the rate of convergence both appear quite satisfactory: to a limited extent, we appear to have satisfied the criteria (a)–(c) stated in Section 2.

Extensive noise studies, tuning, modifications along the lines suggested above, and direct comparison with output least-squares optimization are reported in [25,27–29]

5. DISCUSSION AND CONCLUSION

5.1. The Scope of the Coherency Approach

From a theoretical point of view, and perhaps from a practical one as well, the chief defect in the results of Sections 2 and 3 on the layered acoustic problem is the absolute lack of any provision for multiple reflections. This defect is cured in the companion paper [27], in which the fully nonlinear bandlimited layered velocity inversion problem is treated with full mathematical rigor. We reach a qualitatively identical conclusion about the appropriate version of differential semblance optimization: that is, it gives a (regularized) solution of the inverse problem stably dependent on near-consistent data, provided that sufficiently many reflectors are present, i.e., that the target profile is sufficiently *rough*. We give a precise sense for “rough,” and a relation emerges between stability, aperture, bandlimits, and roughness (reflector density) very similar to that explained in Section 3.

Both the convolutional model and the fully nonlinear versions of other layered inverse problems should succumb to the same approach. We mention, specifically, nonconstant-density acoustics, the “marine” elastic model (with sources and receivers in an overlying fluid layer), and either of these with the source time-dependence and directivity also regarded as unknowns. Some idea of the novel features of these problems in convolutional approximation may be gleaned from [40,41]. No technical obstacles appear to lie in the way of an analogous treatment of these problems.

Note that the density, regarded as independent of velocities, will not be recovered with trend, in any of these problems, as density trend perturbations do not affect ray geometry. This gross density ambiguity is well-known [1] and has been observed in output-least-squares results [17].

The papers [25,28] present a variant of DSO suitable for nonlayered (laterally homogeneous) models using the linearized seismogram (i.e., the higher dimensional analogue of the convolutional approximation). For some initial numerical results, see [26]. On the other hand, while an analogous differential semblance optimization principle can be formulated for fully nonlinear laterally heterogeneous models, its analysis will require fundamental advances in the understanding of wave propagation in rough media.

5.2. Conclusion

We have presented a novel approach to the reflection seismic inverse problem, which has its roots in utterly commonplace concepts in seismic data processing. We have formulated this differential semblance optimization principle precisely for the convolutional approximation to the layered constant-density acoustic model, and suggested both analytically and numerically that its solution yields accurate and stable estimates of both velocity trends and reflectivities, to the extent that these are determined by the precritical plane-wave data set used. Our analysis indicated that differential semblance optimization should require markedly less computational effort than output-least-squares inversion. Finally, we formulated an analogous principle for laterally heterogeneous velocity inversion, a problem for which output-least-squares inversion is so inefficient as to be infeasible. It remains to be seen whether differential semblance optimization yields a computationally tractable approach to such several-dimensional problems.

REFERENCES

1. A. Tarantola, A strategy for nonlinear elastic inversion of seismic reflection data, *Geophysics* **51**, 1893–1903 (1986).
2. O. Gauthier, A. Tarantola and J. Virieux, Two-dimensional nonlinear inversion of seismic waveforms, *Geophysics* **51**, 1387–1403 (1986).
3. Mcaulay, Prestack inversion with plane-layer point source modeling, *Geophysics* **50**, 77–89 (1985).
4. P. Mora, Nonlinear 2-D elastic inversion of multi-offset seismic data, *Expanded abstract, 56th Annual International Meeting, Society of Exploration Geophysicists*, Houston, pp. 533–537 (1986); also, *Geophysics*, **52**.
5. P. Kolb, F. Collino and P. Lailly, Prestack inversion of a 1D medium, *Proc. IEEE* **74**, pp. 498–506 (1986).
6. L. Lines and S. Treitel, A review of least-squares inversion and its application to geophysical problems, *Geophysical Prospecting* **32**, 159–186 (1984).
7. A. Bamberger, G. Chavent, C. Hemon and P. Lailly, Inversion of normal incidence seismograms, *Geophysics* **47**, 757–770 (1982).
8. K. Bube, D.B. Jovanovich, R.T. Kagan, J.R. Resnick, R.T. Shuey and D.A. Spindler, Well-determined and poorly determined features in seismic reflection tomography: Part II, *55th Annual Meeting, SEG*, Washington, DC, (1985).
9. R. Bording, A. Gersztenkorn, L. Lines, J. Scales and S. Treitel, Applications of Seismic Traveltime Tomography, *Geophys. J. Roy. Astr. Soc.* **90**, 285–304, (1987).
10. L. Ikelle, J. Diet and A. Tarantola, Linearized inversion of multioffset seismic data in the omega- k domain: Depth-dependent reference medium, *Geophysics* **53**, 50–64 (1988).
11. J.K. Cohen and N. Bleistein, Velocity inversion procedure for acoustic waves, *Geophysics* **44**, 1077–1085 (1979).
12. R. Clayton and R. Stolt, A Born-WKB inversion method for acoustic reflection data, *Geophysics* **46**, 1559–1567 (1981).
13. 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).
14. L. Lines, A. Schultz and S. Treitel, Cooperative inversion of geophysical data, *Geophysics* **53**, 8–20 (1988).
15. Y. Hadjee and F. Collino, A geometrical approach to the *a priori* study of the 1-D inverse problem, (IFP preprint), (1988).
16. F. Santosa and W. Symes, An analysis of least-squares velocity inversion, *Geophysical Monograph* **4**, Society of Exploration Geophysicists, Tulsa, (1989).
17. G. Canadas and P. Kolb, Least-squares inversion of prestack data: simultaneous identification of density and velocity of stratified media, *Expanded Abstract, 56th Annual International Meeting, Society of Exploration Geophysicists*, Houston, pp. 604–607 (1986).
18. P. Lailly, Migration methods: Partial but efficient solutions to the seismic inverse problem, In *Inverse Problems of Acoustic and Elastic Waves*, Santosa *et al.*, Eds., SIAM, Philadelphia, 182–214, (1984).
19. P. Mora, Nonlinear 2-D elastic inversion of real data, *Expanded abstract, 57th Annual International Meeting, Society of Exploration Geophysicists*, New Orleans, pp. 430–432 (1987).
20. A. Tarantola, The seismic reflection inverse problem, In *Inverse Problems of Acoustic and Elastic Waves*, Santosa *et al.*, Eds., SIAM, Philadelphia, (1984).
21. G. Beylkin and R. Burridge, Linearized inverse scattering problems of acoustics and elasticity, *Expanded Abstract, 57th Annual International Meeting, Society of Exploration Geophysicists*, pp. 747–749, New Orleans, (1987).
22. S. Spratt, Effect of normal moveout errors on amplitude versus offset-derived shear reflectivity, *Expanded Abstract, 57th Annual International Meeting, Society of Exploration Geophysicists*, New Orleans, pp. 634–637 (1987).
23. W. Symes, Stability and instability results for inverse problems in several-dimensional wave propagation, In *Computing Methods in Applied Science and Engineering VI*, R. Glowinski and J.-L. Lions, Eds., North-Holland, (1985).
24. W. Symes, Velocity inversion by coherency optimization, Technical Report 88-4, Department of Mathematical Sciences, Rice University, Houston, TX, (1988).
25. W. Symes, Velocity inversion: A case study in infinite-dimensional optimization, *Math. Prog.* **48**, 71–102 (1990).
26. W. Symes, Non-interactive estimation of the Marmousi velocity model by differential semblance optimization: Initial trials, In *The Marmousi Experience: Proceedings of the EAEG Workshop on Practical Aspects of Inversion*, (G. Grau and R. Versteeg, Eds.), pp. 125–138, EAEG, The Hague, (1991).
27. W. Symes, Velocity inversion: A model problem from reflection seismology, *SIAM J. Math. Anal.*, **22**, 680–716, (1991).
28. 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*, J.B. Bednar, Ed., SIAM.

29. W. Symes and J. Carazzone, Velocity inversion by differential semblance optimization, *Geophysics* **56**, 654–663, (1991).
30. A.H. Kleyn, *Seismic Reflection Interpretation*, Applied Science Publishers, New York, (1983).
31. P. Fowler, Migration velocity analysis by optimization: Linear theory, *Expanded Abstract, 56th Annual International Meeting, Society of Exploration Geophysicists*, Houston, pp. 660–662 (1986).
32. W. Symes, Stability properties for the velocity inversion problem, In *Proc. of Conference on Seismic Exploration and Reservoir Modeling*, W. Fitzgibbon, Ed., Philadelphia, SIAM (1986).
33. A. Tarantola and B. Vallette, Inverse problems: Quest for information, *J. Geophysics* **50**, 159–170 (1982).
34. H. Brysk and D. McCowan, A slant-stack procedure for point-source data, *Geophysics* **51**, 1370–1386 (1986).
35. J.E. Dennis Jr. and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, (1983).
36. T. Steihaug, Quasi-Newton methods for large-scale nonlinear problems, Ph.D. Thesis, Yale, (1981).
37. G. Golub and C. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, (1983).
38. T. Chan, Deflated Lanczos procedures for solving nearly singular linear systems, Research Report YALEU/DCS/RR-403, Department of Computer Science, Yale University, (1985).
39. J. Meza and W. Symes, Deflated Krylov subspace methods for nearly singular linear systems, Technical Report 87-3, Department of Mathematical Sciences, Rice University, Houston, TX, (1987).
40. P. Sacks and W. Symes, Recovery of the elastic parameters of a layered half-space, *Geophys. J. Roy. Astr. Soc.* **88**, 593–620 (1987).

APPENDIX

Computation of a Derivative

This appendix details the calculation of the derivative of the incoherency

$$\begin{aligned} W[c, r] &= \left[\frac{\partial}{\partial p} (v^{-2} r \circ \tau) \right] \circ \zeta = -2pr + \left[v^{-2} \frac{\partial}{\partial p} (r \circ \tau) \right] \circ \zeta \\ &= -2pr + v^{-2} \circ \zeta \left(\frac{\partial r}{\partial p} - p \frac{\partial \tau}{\partial t} \int_0^\zeta v \right) \end{aligned} \quad (\text{A.1})$$

in which have been used the identities

$$\frac{\partial}{\partial p} v^{-2} = -2p, \quad \frac{\partial v}{\partial p} = pv^3, \quad \frac{\partial \tau}{\partial p} = -p \int_0^z v.$$

Clearly,

$$\delta W = \delta_c W + \delta_r W,$$

with

$$\delta_r W = W[c, \delta r].$$

On the other hand, an easy calculation shows that

$$\delta \zeta = v \circ \zeta \int_0^\zeta v \frac{\delta c}{c^3}.$$

So, from the second line in (A.1),

$$\begin{aligned} \delta_c W &= \delta(v^{-2} \circ \zeta) \left[\frac{\partial}{\partial p} (r \circ \tau) \right] \circ \zeta - p v^{-2} \circ \zeta \frac{\partial r}{\partial t} \delta \int_0^\zeta v \\ &= -2 \left[\frac{\delta c}{c^3} + \frac{1}{c^3} \frac{\partial c}{\partial z} c \int_0^z v \frac{\delta c}{c^3} \right] \circ \zeta \left[\frac{\partial}{\partial p} (r \circ \tau) \right] \circ \zeta - p \left[\int_0^z v^{-2} \Lambda \delta c + v^{-2} \int_0^z \Lambda \delta c \right] \circ \zeta \frac{\partial r}{\partial t}. \end{aligned} \quad (\text{A.2})$$

An alternate form of (A.2) follows from the identity

$$\delta \int_0^\zeta v = \delta \left(\int_0^\zeta (v \circ \zeta)^2 \right) = 2 \int_0^\zeta (v \circ \zeta) \delta(v \circ \zeta) = 2 \int_0^\zeta \delta(v \circ \zeta) \circ \tau.$$

So, using the notation

$$Q \delta c = \delta(v \circ \zeta) \circ \tau = \Lambda^3 \left(\delta c + \frac{\partial c}{\partial z} v \int_0^z v \frac{\delta c}{c^3} \right)$$

the first line in (A.2) may be re-written

$$\delta_c W = -2 \left[v^{-3} Q \delta c \frac{\partial}{\partial p} (r \circ \tau) \right] \circ \zeta - 2p v^{-2} \circ \zeta \frac{\partial r}{\partial t} \int_0^\zeta Q \delta c. \quad (\text{A.2}')$$

For Section 3, we require an evaluation of $\delta_c W$ when r is *coherent*, i.e., (5) is satisfied. Then, for a suitable Δc ,

$$r = \left(\Lambda^2 \frac{c_r}{c} \right) \circ \zeta.$$

Moreover, from (A.1),

$$\frac{\partial}{\partial p}(r \circ \tau) \circ \zeta = 2p(v^2 \circ \zeta) \tau,$$

so (A.2') becomes

$$\begin{aligned} \delta_c W &= -2p \left\{ 2 [Q \delta c \cdot v^{-1}] \circ \zeta \cdot r + \left[v^{-2} \int Q \delta c \right] \circ \zeta \frac{\partial r}{\partial t} \right\} \\ &= -2p \left\{ 2 (Q \delta c) v c^{-3} c_r + \left(\int Q \delta c \right) v^{-1} \frac{\partial}{\partial z} \left(\Lambda^2 \frac{c_r}{c} \right) \circ \right\} \\ &= -2p \left\{ 2 (Q \delta c) \Lambda c^{-2} c_r + \left(\int Q \delta c \right) \left(2 \Lambda^3 c' p^2 \left(\frac{c_r}{c} \right) + c^{-1} \Lambda \frac{\partial}{\partial z} \left(\frac{c_r}{c} \right) \right) \right\} \circ \zeta \\ &= -2p \left\{ 2 \left[(Q \delta c) \Lambda c^{-1} + \left(\int Q \delta c \right) p^2 \Lambda^3 c' \right] \left(\frac{c_r}{c} \right) + \left[\left(\int Q \delta c \right) c^{-1} \Lambda \right] \frac{\partial}{\partial z} \left(\frac{c_r}{c} \right) \right\} \circ \zeta. \end{aligned} \tag{A.3}$$

For our computations, we used the formulas (A.1) and (A.2') with

$$\left(\frac{\partial}{\partial p}(r \circ \tau) \right) \circ \zeta$$

computed as in the second line of (A.1).