All stationary points of differential semblance are asymptotic global minimizers: Layered acoustics

William W. Symes[1]

keywords: *velocity, semblance, inversion*

ABSTRACT

Differential semblance velocity estimators have well–defined and smooth high frequency asymptotics. A version appropriate for analysis of CMP gathers and layered acoustic models has no secondary minima. Its structure suggests an approach to optimal parametrization of velocity models.

INTRODUCTION

The core problem of primaries-only (linearized, Born approximation) modeling, imaging, and inversion is that of finding an accurate reference velocity. Since the typical survey is highly redundant, predictions of reflectivity are redundant, and unlikely to be consistent (or *flat* in common image panels) unless the velocity field used to make them is essentially correct. This concept of *semblance of redundant images* underlies velocity estimation methods in widespread use (Taner and Koehler, 1969; Yilmaz and Chambers, 1984; Reshef, 1997).

A number of researchers have cast velocity analysis as an *optimization* problem: that is, they propose an objective function to be minimized or maximized at a correct velocity model (Toldi, 1985; Al Yahya, 1989; Fowler, 1986; Kolb et al., 1986; Cao et al., 1990; Clément and Chavent, 1993; Sevink and Herman, 1993; Martinez and McMechan, 1991; Sen and Stoffa, 1991). Extremization of the objective is then an automatic process, to be accomplished through numerical optimization algorithms. The most widely investigated objectives - variants of stack power or RMS data fit error ("output least squares") - are velocity dependent quadratic forms in the data. These functions are believed to be multimodal and very ill-conditioned (Gauthier et al., 1986; Scales et al., 1991). The presumed existence of many local minima appears to mandate global search methods such as simulated annealing. These usually require orders of magnitude more function evaluations than do gradient-based methods which find local minima. The computational cost of global search methods renders them unsuitable for industrial scale velocity estimation.

[1]**email**: symes@caam.rice.edu

The mechanism underlying these features of stack power and similar objectives is *asymptotic instability*: none of these functions have limiting shapes as source and data bandwidth become infinite. Besides accounting for multimodality, saturation, ill conditioning, and other undesirable mathematical properties, asymptotic instability of stack power, output least squares, and similar objectives also inhibits analysis of local and global features *via* of high frequency asymptotics.

These observations lead to the question: do there exist velocity dependent quadratic forms in the data, extremized by the correct velocity with model-consistent data, which also have stable high frequency asymptotics? The answer is "yes", and the nature of such forms is specified completely as part of the answer: they express semblance through comparison of *neighboring traces*. In contrast, stack power and output least squares objectives measure semblance (explicitly or implicitly) by comparing of traces with widely differing offset and/or midpoint, and this fact accounts for the asymptotic instability of these functions. In the ideal limit of continuous sampling, traces to be compared should be infinitely near, so I have used the phrase *differential semblance* to describe these asymptotically stable forms. For mathematical details of the connection between differential semblance and stable asymptotics in the context of the simpler but similar *plane wave detection* problem, see (Kim and Symes, 1998), also (Claerbout, 1992), pp. 93 ff.

The asymptotic stability of differential semblance opens up the possibility of analysing its global shape by asymptotic methods. This paper presents such an analysis for a simple special case applicable to field data, based on the convolutional model of CMPs for layered acoustic Earth response. The layered medium assumption leads to simple explicit expressions for all quantities figuring in this approach to velocity estimation. The analysis shows that, for noise-free (model consistent) data in the continuous sampling limit and velocities limited to natural admissible sets, the length of the gradient bounds the objective, up to an error which vanishes in the high frequency limit. Therefore every stationary point is asymptotically a global minimum of the differential semblance objective function in this case. That is, differential semblance does not suffer from the local minima which plague other optimization formulations of velocity inversion.

Simple estimates bound the effect of noise. Numerical experiments have shown that random noise has virtually no effect on the location of DS stationary points, whereas strong coherent noise, such as multiple reflections, has a maximal effect. In any case the influence of noise is bounded, i.e. the differential semblance velocity estimate "degrades gracefully" as noise of any sort is added to the data.

These conclusions - stable asymptotics, unimodality, bounded influence of noise, significance of coherent noise - conform to the results of many numerical experiments with field and synthetic data. The present paper concerns only analysis: details of computational implementation and results appear elsewhere (Araya et al., 1996; Gockenbach and Symes, 1997; Symes, 1997, 1998; Chauris et al., 1998a,b).

The ubiquitous presence of various "multis" - multiple reflection, multiple refraction (transmission caustics), multiple wave modes, and of course multidimensional geometry - necessarily limits the practical importance of this or any other technique for velocity estimation (or imaging) based on primaries only layered acoustic modeling. Note that the differential semblance concept is *not* in any way limited to layered medium models or 1D velocity functions, any more than stack power is limited to NMO-based velocity analysis. The abstract (Chauris et al., 1998b) and earlier work of this author (Symes, 1993; Kern and Symes, 1994) present examples of multidimensional velocity estimation by differential semblance optimization.

I begin with an abstract definition of differential semblance. After defining the convolutional model for layered acoustics, I discuss various types of error inherent in this approximation, the construction of mutes, and natural admissible model sets. This groundwork supports an asymptotic analysis of the differential semblance objective, which reveals that in the case of noise free data it is essentially a data-weighted mean square error in RMS square slowness. This observation leads directly to the main result, and to some convenient estimates of the influence of noise in general data. It also suggests an approach to optimal parametrization of velocity profiles.

AN ABSTRACT FORMULATION OF DIFFERENTIAL SEMBLANCE

The general definition of differential semblance presented here owes much to ideas introduced by Hua Song and Mark Gockenbach in their theses (Song, 1994; Gockenbach et al., 1995).

The (reference or background) velocity v includes the slowly varying components of velocity and perhaps other fields. The reflectivity r is a field (or vector of fields) encompassing the rapidly varying components of the model. Linearized scattering treats r as a perturbation of v. Thus in the forward modeling operator $F[v]r$ the dependence on r is linear, whereas the dependence on v is (quite!) nonlinear.

Minimal data sets are those on which the kinematic relations in the data are bijective. Minimal data sets include common shot and common offset gathers, and - for layered models - single traces. For a few models, such as constant density acoustics, the forward modeling operator $F[v]$ is invertible (modulo smoothing operators) on minimal data sets. This paper deals only with constant density acoustics.

Denote by $G[v]$ an approximate inverse operator for $F[v]$ on each (minimal) data bin (common source, common offset, single trace,...). Thus $G[v]$ applied to the data produces a *prestack reflectivity volume*. Similarly, understand by $F[v]$ the application of forward modeling independently for each reflectivity bin.

Each binning scheme also implies a notion of *neighboring bins*: that is, neighboring source positions, offsets,... Denote by W an operator approximating the derivative or gradient in the bin parameter(s). Generally the definition of $F[v]$ necessarily incorporates a cutoff or mute, as does that of $G[v]$. Differentiation in the bin direction across this mute produces edge artifacts. To control these, introduce an additional mute ϕ slightly more severe than the mutes built into F and G. Since the edge effects are localized, application of this secondary mute ϕ eliminates them.

Differentiation enhances high frequency content. To keep the spectrum of the differential semblance output comparable to that of the data, employ a smoothing operator H. An appropriate choice is the inverse square root of the Helmholtz operator $(I - \nabla^2)^{-\frac{1}{2}}$ in all of the variables on which the data depends, i.e. both within-bin and cross-bin variables.

With these notations, define differential semblance $J_0[v]$ by:

$$J_0[v] = \frac{1}{2} \|H\phi F[v]WG[v]S\|^2$$

Here S denotes the data, and the vertical double bars denote the L^2 norm or root mean square, i.e. summation of the square of the quantity inside over all variables, followed by square root.

THE CONVOLUTIONAL MODEL FOR LATERALLY HOMOGENOUS ACOUSTICS

Linearization of the acoustic model for a layered fluid and application of high frequency asymtotics leads to the *convolutional model* of primaries-only reflection seismograms. The convolutional model of offset traces is one of the simplest models of the reflection process within which to pose the velocity analysis problem. A similar model for plane wave traces is almost equally simple, and was the subject of earlier work on differential semblance (Symes and Carazzone, 1991; Minkoff and Symes, 1997). However synthesis of accurate plane wave traces is a nontrivial task. Accordingly the version of the model developed here uses offset domain data.

A natural binning scheme for this model is the common midpoint gather. Since all midpoint gathers are in principle the same for a layered model, the data consists of a single CMP. The bins contain single traces, parameterized by offset x.

The velocity parameter is simply the interval velocity $v(z)$, whereas the reflectivity is $r = \frac{\delta v}{v}$ and is regarded as bin-dependent, i.e. $r = r(z, x)$; this section plays the role of a common image gather, as every trace represents reflectivity below the same midpoint. Thus successful velocity estimation will produce a "flat" (x-independent) $r(z, x)$.

The simple version of DS presented here will assume that source signature deconvolution has been applied to the data, so that it is essentially impulsive.

It will be convenient to parametrize velocity and reflectivity by *vertical two-way time*

$$t_0 = 2\int_0^z \frac{dz}{v}$$

rather than depth: thus $v = v(t_0)$, $r = r(t_0, x)$.

With these conventions, the forward modeling operator is

$$F[v]r(t, x) = a(t, x)r(T_0(t, x), x)$$

where a is the geometric amplitude and $T_0(t, x)$ is the inverse function of the *two-way traveltime function* $T(t_0, x)$.

ERROR, ERROR EVERYWHERE!

As an approximate predictor of seismic traces, the convolutional model exhibits several types of error:

- physics error: seismic waves are not small amplitude pressure waves in a fluid;
- linearization error: neglect of multiple reflections and other nonlinear effects;
- deconvolution error: complete removal of the source signature is not possible;
- asymptotic error: the convolutional model becomes more accurate as the frequency content of $r(t_0, x)$ moves away from zero Hz.

The practical meaning of asymptotic error is that the convolutional model predicts the higher frequency components of the data more accurately, so that the prediction error can be reduced by more aggressive low-cut filtering. Of course this discarding of low-frequency data is only possible to a limited extent as actual data is bandlimited.

The following computations will introduce yet more sources of asymptotic error - and, with one exception, only asymptotic error. Therefore I will identify asymptotic error explicitly, and treat other types of modeling error as data noise. It is possible to estimate every asymptotic error explicitly, but experience suggests that these explicit estimates are not particularly useful. So instead I will use the symbol "$O(\lambda)$" to suggest proportionality of the asymptotic error to a dominant wavelength in the data. Thus

$$F[v]r(t, x) = a(t, x)r(T_0(t, x), x) + O(\lambda)$$

The single important lesson to learn from the explicit error estimates of geometric optics is that they are uniform over C^∞-bounded sets of coefficients (meaning in this case the velocity v). Therefore the velocities appearing in the sequel are restricted to vary over such a C^∞-bounded set. A byproduct of the analysis will suggest explicit finite dimensional subspaces of smooth functions in which it is advantageous to seek v.

MUTES

The linearized model accurately predicts only precritical primary reflections. For layered media, precritical reflections have downgoing incident rays. Along downgoing rays, time is an increasing function of depth. It follows that if t_0 is to be a depth variable, then T must be an increasing function of t_0. This is generally true only in a subset of the t, x plane, i.e. only part of this plane contains data accurately modeled by linearized acoustics. Therefore the rest of the data must be muted out.

Define the *stretch factor*

$$s(t, x) = \frac{\partial T_0}{\partial t}(t, x) = \left(\frac{\partial T}{\partial t_0}(T_0(t, x), x)\right)^{-1}$$

Then the condition that $T(t_0, x)$ be monotone increasing as a function of t_0 is equivalent to demanding that for large enough t

$$0 < s(t, x) < C_{\rm stretch}$$

where $C_{\rm stretch}$ is a user-specified parameter larger than one. Define $T_{\rm mute}(x)$ (the *mute boundary*) to be the infimum of all t for which the above inequality is satisfied on the interval $(t, T_{\rm max})$. Then the support of the mute function ϕ should be contained in the set $\{(t, x) : t \geq T_{\rm mute}(x)\}$.

Define a corresponding t_0, x domain mute by $\phi_0(t_0, x) = \phi(T(t_0, x), x)$.

ADMISSIBLE MODELS

In this section I introduce *admissible sets* $\mathcal{A}$ of models, on which the convolutional model as defined above is reasonably well behaved. Note that the constraints imposed on the models by membership in the admissible sets are very natural from the physical or geological point of view.

First of all, the velocity must be smooth, as noted above in the section on errors. The restriction of v to a bounded subset of C^∞ implies bounds (maximum absolute value, mean square,...) on any derivative of v.

Second, impose smooth upper and lower "envelope" velocities as hard constraints: $v_{\rm min}(t_0) \leq v(t_0) \leq v_{\rm max}(t_0)$. It is natural to assume that the velocity is known at the surface, so assume that $v_{\rm min}(0) = v_{\rm max}(0)$. These bounds derive from geophysical measurements and general knowledge about rock physics, so should be regarded as distinct from the bounds implied by the first condition (membership in a bounded set in C^∞).

The set of velocities satisfying the constraints just outlined form the *admissible set* $\mathcal{A}$.

An important consequence is that the mute $\phi \in C_0^\infty(\mathbf{R}^2)$ may be chosen uniform over $\mathcal{A}$, as uniform bounds then exist for every value of the stretch factor $s(t, x)$. These bounds follow from the equations of geometric optics. However they are even more simply derived for the hyperbolic moveout approximation to traveltime, which I will eventually adopt, so I do not give a derivation here.

ASYMPTOTIC APPROXIMATION OF DIFFERENTIAL SEMBLANCE

The convolutional offset trace model is one of those for which the forward modeling operator on a minimal gather, ie. a single trace, is invertible. The inverse operator is

$$G[v]S(t_0, x) = \frac{S(T(t_0, x), x)}{a(T(t_0, x), x)}$$

The operator measuring semblance differentially is

$$W = \frac{\partial}{\partial x}$$

Then

$$F[v]WG[v]S(t,x) = a(t,x)\left[\frac{\partial}{\partial x}\frac{S(T(t_0,x),x)}{a(T(t_0,x),x)}\right]_{t_0=T_0(t,x)}$$
$$= \left(\frac{\partial T}{\partial x}(T_0(t,x),x)\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right)S(t,x) + ...$$
$$= \left(p(t,x)\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right)S(t,x) + ...$$

where

$$p(t,x) = \frac{\partial T}{\partial x}(T_0(t,x),x)$$

is the arrival (horizontal) slowness of the ray passing offset x at time t, and the elided terms involve the amplitude a, but do *not* involve derivatives of the data S. Thus these terms are of lower frequency content than the leading term (explicitly displayed), and are of the same relative order in frequency as terms neglected in the derivation of the convolutional model from the acoustic wave equation. Therefore they can be dropped: this leads to the remarkable conclusion that the differential semblance objective is *independent of the amplitude* at least to leading order in frequency.

This observation is due to Hua Song. As a result, within accuracy limitations already built into the asymptotic linearized model, a might as well be replaced by 1!. That is, to leading order in frequency, differential semblance is insensitive to wave dynamics (amplitude), and responds only to kinematic model changes, i.e. changes in traveltime. Thus minimization of differential semblance will amount to a sort of traveltime tomography.

Fons ten Kroode (personal communication) has pointed out that replacement of $G[v]$ by an asymptotically unitary operator with the same kinematics also yields an asymptotocally identical objective without leading order amplitude dependence, and without application of the forward modeling operator, thus at lower computational cost.

The computations above are correct when the map $(t_0,x) \mapsto (t,x)$ is smooth and invertible. This is so inside the mute zone defined above, uniformly for $v \in \mathcal{A}$. Therefore application of the inverse square root Helmholtz operator following will bring the spectral content back into alignment with that of the data, uniformly over $v \in \mathcal{A}$. Thus

$$H\phi F[v]WG[v]S = H\phi\left(\frac{\partial S}{\partial x} + p\frac{\partial S}{\partial t}\right) + O(\lambda)$$

The ray slowness p is locally a smooth function of the velocity v in any fixed open subset of the mute zone, hence J_0 (which is the mean square of the above expression) is a smooth function of $v \in \mathcal{A}$ as well.

NOISE FREE DATA

Assume that the data S are model-consistent, that is

$$S(t,x) = r^*(T_0^*(t,x)) + O(\lambda)$$

for target offset independent reflectivity $r^*(t_0)$ and velocity $v^*(t_0)$. [Since differential semblance does not depend to leading order on the amplitude, as noted above, I set the amplitude to 1 in the following, for simplicity - it can be reintroduced with almost no change in the results to follow.]

Note that

$$0 = \frac{\partial}{\partial x} T(T_0(t,x),x) = \frac{\partial T}{\partial t_0}(T_0(t,x),x)\frac{\partial T_0}{\partial x}(t,x) + \frac{\partial T}{\partial x}(T_0(t,x),x)$$

so

$$\frac{\partial T_0}{\partial x}(t,x) = -s(t,x)p(t,x)$$

(s being the stretch factor, defined above). Thus

$$\frac{\partial}{\partial x} r^*(T_0^*(t,x)) = -s^*(t,x)p^*(t,x)(\frac{\partial r^*}{\partial t_0})(T_0^*(t,x))$$

(s^* is the stretch factor belonging to v^*) whence

$$F[v]WG[v]S(t,x) = \left(\frac{\partial}{\partial x} + p(t,x)\frac{\partial}{\partial t}\right) r^*(T_0^*(t,x))$$

$$= s^*(t,x)(p(t,x) - p^*(t,x))\frac{\partial r^*}{\partial t_0}(T_0^*(t,x))$$

According to the calculus of pseudodifferential operators,

$$H\phi F[v]WG[v]S =$$

$$(I - \nabla^2)^{-\frac{1}{2}}\phi\left(s^*(p-p^*)\frac{\partial r^*}{\partial t_0}(T_0^*)\right)$$

$$= (I - \nabla^2)^{-\frac{1}{2}}\phi\left(s^*(p-p^*)\frac{\nabla T_0^* \cdot \nabla}{\nabla T_0^* \cdot \nabla T_0^*} r^*(T_0^*)\right)$$

$$= \phi\frac{s^*(p-p^*)}{\sqrt{1 + s^{*,2}(1+p^{*,2})}} r^*(T_0^*) + O(\lambda)$$

where you get from the next to the last line to the last by substituting ∇T_0^* for ∇, and using previously derived formulas for the partial derivatives of T_0.

Thus

$$J_0[v] = \int\int dt\,dx\; B^*(t,x)(p(t,x) - p^*(t,x))^2[r^*(T_0^*(t,x))]^2 + O(\lambda)$$

where

$$B^*(t,x) = \phi^2\frac{s^{*,2}}{1 + s^{*,2}(1+p^{*,2})}$$

is independent of v, i.e. depending only on v^* and $\mathcal{A}$.

In the next section I introduce the so called hyperbolic moveout approximation to traveltime. Note that up to this point the development is entirely independent of this approximation. In particular the formulas worked out in this section have precise analogues for versions of differential semblance based on multidimensional seismic models.

HYPERBOLIC MOVEOUT

Claim: To good approximation, for "small" offsets,

$$T(t_0, x) = \sqrt{t_0^2 + \frac{x^2}{v_{\rm RMS}^2(t_0)}}, \;\; T_0(T(t_0, x), x) \equiv t_0$$

where the RMS velocity is

$$v_{\rm RMS}(t_0) = \sqrt{\frac{1}{t_0}\int_0^{t_0} v^2}$$

Justification [Continuum derivation of the hyperbolic moveout approximation]: The 2-way traveltime $T(t_0, x) = \tau_2(z, x)$ from the surface at $z = 0$ to depth z and back at offset x is related to the solution of the eikonal equation $\tau(z, x)$ with point source at $z = x = 0$ by

$$\tau_2(z, x) = 2\tau\left(z, \frac{x}{2}\right)$$

Thus

$$\frac{1}{4}\left(\frac{\partial \tau_2}{\partial z}\right)^2 + \left(\frac{\partial \tau_2}{\partial x}\right)^2 = \frac{1}{v}^2$$

Differentiate this twice with respect to x and use the vanishing of odd-order x derivatives at $x = 0$ (implied by symmetry) to conclude that the second x derivative

$$q(z) = \frac{\partial^2 \tau}{\partial x^2}(z, 0)$$

satisfies

$$\frac{1}{2v}\frac{dq}{dz} + q^2 = 0$$

Introduce temporarily a new depth coordinate

$$\sigma(z) = 2\int_0^z v$$

Then in terms of σ, q satisfies the Ricatti equation

$$\frac{dq}{d\sigma} + q^2 = 0$$

The solution which is singular at $\sigma = 0$, i.e. $z = 0$, is

$$q(\sigma) = \frac{1}{\sigma} = \frac{1}{2\int_0^z v}$$

Since $dz = \frac{1}{2} v dt_0$, you can also write this as

$$q(\sigma) = \frac{1}{\int_0^{t_0} v^2}$$

Thus

$$T(t_0,x) = t_0 + \frac{x^2}{2}\frac{\partial^2 T(t_0,0)}{\partial x^2} + ...$$

$$= t_0 + \frac{x^2}{2\int_0^{t_0} v^2} + ...$$

Since

$$\frac{\partial^2}{\partial x^2}(T(t_0,x))^2_{x=0} = 2t_0\frac{\partial^2 T}{\partial x^2}(t_0,0)$$

the above can be rewritten

$$T(t_0,x)^2 = t_0^2 + \frac{x^2}{\frac{1}{t_0}\int_0^{t_0} v^2} + ...$$

$$= t_0^2 + \frac{x^2}{v_{\mathrm{RMS}}^2(t_0)} + ...$$

which reveals that the hyperbolic moveout approximation is just the second order Taylor expansion of T^2 in x, which should be good for "small" x.

This report adopts the hyperbolic moveout approximation, i.e. truncate the Taylor expansion above and take

$$T(t_0,x)^2 = t_0^2 + \frac{x^2}{v_{\mathrm{RMS}}^2(t_0)}$$

This amounts to assuming that all events in the data have precisely hyperbolic moveout. Of course this assumption is not entirely consistent with geometric optics. It has been suggested that the deviation of actual two-way time from the hyperbolic moveout approximation may be mistaken for evidence of anisotropy in some cases. In any case the error caused by replacing actual two way time by its hyperbolic moveout approximation is *not* an asymptotic error in the sense of the last section, so I will treat it as a component of data noise.

The reciprocal square RMS velocity, or *RMS square slowness* is the primary expression of velocity in the above formula. It occurs so often as to warrant its own notation:

$$u(t_0) \equiv (v_{\mathrm{RMS}}(t_0))^{-2}$$

The conditions defining the mute can be restated: since

$$\frac{\partial T}{\partial t_0}(t_0,x) = \frac{t_0 + \frac{x^2}{2}\frac{\partial u}{\partial t_0}(t_0)}{T(t_0,x)}$$

the quantity on the right hand side of this equation must be bounded away from zero. Since v generally increases with depth, hence u decreases, such a lower bound will only be possible for t_0 exceeding a threshold for each x, which is the mute boundary mentioned before. In the data, i.e. (t,x), coordinates, the stretch factor condition becomes

$$s(t,x) = \frac{\partial T_0}{\partial t}(t_0,x) = \left(\frac{\partial T}{\partial t_0}(T_0(t,x),x)\right)^{-1} = \frac{t}{T_0(t,x) + \frac{x^2}{2}\frac{\partial u}{\partial t_0}(T_0(t,x))} < C_{\mathrm{stretch}}$$

and as before the mute ϕ must be supported in the set specified by this condition.

The upper and lower velocity envelopes implied by membership of the velocity in $\mathcal{A}$ imply corresponding envelope mean square slownesses ($u_{\min}$ corresponding to $v_{\max}$ and vis-versa) so that $u_{\min}(t_0) \le u(t_0) \le u_{\max}(t_0)$.

It is usually reasonable to assume the lower velocity bound to be constant (independent of t_0) - for example, equal to sound velocity in water, or close to it. Then $u_{\max}$ is also constant, so you can explicitly estimate a lower bound for T_0:

$$T^2(t_0, x) = t_0^2 + x^2 u(t_0) \le t_0^2 + x^2 u_{\max}$$

so

$$T_0(t, x) \ge \sqrt{t^2 - x^2 u_{\max}}$$

The velocity bounds also imply a bound on the derivative of u:

$$\frac{du}{dt_0} = -\frac{u^2}{t_0}\left(v^2 - \frac{1}{u}\right) = -\frac{u^2}{t_0}\left(v^2 - v^2(0) - \frac{1}{t_0}\int_0^{t_0} (v^2 - v^2(0))\right)$$

The bounds on v, the known value of v at the surface, and the maximum two way time imply bounds on the slope

$$\frac{v^2(t_0) - v^2(0)}{t_0}$$

whence a bound $u'_{\max}$ on the derivative of u follows immediately.

Since both the lower bound on T_0 and the uppoer bound on the derivative of u are uniform over $\mathcal{A}$, a $\mathcal{A}$-uniform bound on the stretch factor follows:

$$s(t, x) \le \frac{t}{\sqrt{t^2 - x^2 u_{\max}} - \frac{x^2}{2} u'_{\max}}$$

From this you can derive a $\mathcal{A}$-uniform mute boundary. Therefore assume henceforth that ϕ is a $\mathcal{A}$-uniform mute.

GLOBAL ANALYSIS OF STATIONARY POINTS IN HYPERBOLIC MOVEOUT APPROXIMATION

Until further notice regard F etc. as depending on RMS square slowness u rather than on interval velocity v. Dependence on v, through the relatively easily analyzed map $v \mapsto u$, will be reintroduced at the end.

A short calculation shows that

$$p(t, x) = \frac{x}{t} u(T_0(t, x)).$$

Introduce the quantity Γ, with units of time:

$$\Gamma(t_0, x) = T_0(T^*(t_0, x), x)$$

That is, $\Gamma(t_0, x)$ is the zero offset time for which the time at offset x is the same in the slowness u as the time one obtains for t_0, x in slowness u^*.

Then introducing the expression for p, and changing variables from t to t_0 in the integral above, yields

$$J_0[u] = \int\int dt_0\, dx\, B_0(t_0, x)(u(\Gamma(t_0, x)) - u^*(t_0))^2 (r^*(t_0))^2 + O(\lambda)$$

where

$$B_0(t_0, x) = \phi(T^*(t_0, x), x) B^*(T^*(t_0, x), x) \left(\frac{x}{T^*(t_0, x)}\right)^2$$

depends only on u^* and $\mathcal{A}$.

It is now straightforward to compute the first order perturbation δJ_0 of J_0 with respect to u. First,

$$\delta T(t_0, x) = \frac{\frac{x^2}{2}\delta u(t_0)}{T(t_0, x)}$$

$$0 = \delta(T(T_0(t, x), x)) = \delta T(T_0(t, x), x) + \frac{\partial T}{\partial t_0}(T_0(t, x), x)\delta T_0(t, x)$$

$$= \frac{\frac{x^2}{2}\delta u(T_0(t, x))}{t} + \frac{\delta T_0(t, x)}{s(t, x)}$$

so

$$\delta T_0(t, x) = -\frac{\frac{x^2}{2} s(t, x)\delta u(T_0(t, x))}{t}$$

whence

$$\delta\Gamma(t_0, x) = \delta T_0(T^*(t_0, x), x) = -\frac{\frac{x^2}{2} s(T^*(t_0, x), x)\delta u(\Gamma(t_0, x))}{T^*(t_0, x)}$$

and

$$\delta(u(\Gamma(t_0, x))) = \delta u(\Gamma(t_0, x)) + \frac{du}{dt_0}(\Gamma(t_0, x))\delta\Gamma(t_0, x))$$

$$= \delta u(\Gamma(t_0, x)) \left(1 - \frac{\frac{x^2}{2} s(T^*(t_0, x), x)\frac{du}{dt}(\Gamma(t_0, x))}{T^*(t_0, x)}\right)$$

Recall that

$$s(t, x) = \frac{t}{T_0(t, x) + \frac{x^2}{2}\frac{\partial u}{\partial t_0}(T_0(t, x))}$$

so that

$$s(T^*(t_0, x), x) = \frac{T^*(t_0, x)}{\Gamma(t_0, x) + \frac{x^2}{2}\frac{\partial u}{\partial t_0}(\Gamma(t_0, x))}$$

so

$$\delta(u(\Gamma(t_0, x))) = \delta u(\Gamma(t_0, x)) \left(1 - \frac{\frac{x^2}{2}\frac{du}{dt}(\Gamma(t_0, x))}{\Gamma(t_0, x) + \frac{x^2}{2}\frac{\partial u}{\partial t_0}(\Gamma(t_0, x))}\right)$$

$$= \frac{\delta u(\Gamma(t_0, x))}{1 + \frac{x^2}{2\Gamma(t_0,x)}\frac{\partial u}{\partial t_0}(\Gamma(t_0, x))} = \frac{\Gamma(t_0, x)}{T^*(t_0, x)} s(T^*(t_0, x), x)\delta u(\Gamma(t_0, x))$$

Putting this all together,

$$\delta J_0[u] = \int\int dt_0\, dx\; B_1(t_0, x)(u(\Gamma(t_0, x)) - u^*(t_0))(r^*(t_0))^2 \delta u(\Gamma(t_0, x)) + O(\lambda)$$

where

$$B_1(t_0, x) = \frac{B_0(t_0, x)}{1 + \frac{x^2}{2\Gamma(t_0,x)}\frac{\partial u}{\partial t_0}(\Gamma(t_0, x))} = \frac{\Gamma(t_0, x)}{T^*(t_0, x)} B_0(t_0, x) s(T^*(t_0, x), x)$$

depends on u, u^*, and $\mathcal{A}$. To compute the gradient, change variables again to $t_0' = \Gamma(t_0, x)$ for each x. Since

$$\frac{\partial \Gamma}{\partial t_0}(t_0, x) = \frac{\partial}{\partial t_0}(T_0^*(T(t_0, x), x))$$

$$= \frac{\partial T_0^*}{\partial t}(T(t_0, x), x)\frac{\partial T}{\partial t_0}(t_0, x) = \frac{s^*(T(t_0, x), x)}{s(T(t_0, x), x)}$$

and so

$$\frac{\partial \Gamma^{-1}}{\partial t_0}(t_0, x) = \frac{s(T^*(t_0, x), x)}{s^*(T^*(t_0, x), x)}$$

you get

$$\delta J_0[u] = \int\int dt_0\, dx\; B_1^*(t_0, x)(u(t_0) - u^*(\Gamma^{-1}(t_0, x)))(r^*(\Gamma^{-1}(t_0, x)))^2 \delta u(t_0)) + O(\lambda)$$

with

$$B_1^*(t_0, x) = B_1(\Gamma^{-1}(t_0, x), x)\frac{s(T^*(t_0, x), x)}{s^*(T^*(t_0, x), x)}.$$

Thus the L^2 gradient of J_0 is

$$\nabla J_0[u](t_0) = \int dx\; B_1^*(t_0, x)(u(t_0) - u^*(\Gamma^{-1}(t_0, x)))(r^*(\Gamma^{-1}(t_0, x)))^2 + O(\lambda)$$

Both expressions for J_0 and its gradient suggest that these quantities are comparing the trial square slowness u and the target square slowness u^* at different points (eg. t_0 vs. $\Gamma^{-1}(t_0, x)$), and this in turn makes understanding of the implications for determination of u difficult. Fortunately this is not really the case:

Key Lemma: There exists a function $h(t_0, x)$, depending on velocity v (or slowness u) and also on u^* and $\mathcal{A}$, having the following properties:

- $h(t_0, x) > 0$ over the mute zone, and $\log h(t_0, x)$ is uniformly bounded for t_0, x in the mute zone and $v \in \mathcal{A}$;
- $u(t_0) - u^*(\Gamma^{-1}(t_0, x)) = h(t_0, x)(u(t_0) - u^*(t_0))$

Proof of Key Lemma: Note first that since

$$T^*(T_0^*(t,x),x) = t$$

$$\frac{\partial T^*}{\partial x}(T_0^*(t,x),x) + \frac{\partial T^*}{\partial t_0}(T_0^*(t,x),x)\frac{\partial T_0^*}{\partial x}(t,x) = 0$$

$$= \frac{xu^*(T_0^*(t,x))}{t} + \frac{1}{s^*(t,x)}\frac{\partial T_0^*}{\partial x}(t,x)$$

so

$$\frac{\partial T_0^*}{\partial x}(t,x) = -s^*(t,x)\frac{xu^*(T_0^*(t,x))}{t}.$$

It follows that since

$$\Gamma^{-1}(t_0,x) = T_0^*(T(t_0,x),x),$$

$$\frac{\partial \Gamma^{-1}}{\partial x}(t_0,x) = \frac{\partial T_0^*}{\partial t}(T(t_0,x),x)\frac{\partial T}{\partial x}(t_0,x) + \frac{\partial T_0^*}{\partial x}(T(t_0,x),x)$$

$$= s^*(T(t_0,x),x)\frac{xu(t_0)}{T(t_0,x)} - s^*(T(t_0,x),x)\frac{xu^*(T_0^*(T(t_0,x),x))}{T(t_0,x)}$$

$$= \frac{xs^*(T(t_0,x),x)}{T(t_0,x)}(u(t_0) - u^*(\Gamma^{-1}(t_0,x)))$$

Thus

$$u(t_0) - u^*(\Gamma^{-1}(t_0,x)) = u(t_0) - u^*(t_0) + \int_0^x dx' \, \frac{\partial}{\partial x'}(u(t_0) - u^*(\Gamma^{-1}(t_0,x')))$$

$$= u(t_0) - u^*(t_0) - \int_0^x dx' \, \frac{\partial u^*}{\partial t_0}(\Gamma^{-1}(t_0,x'))\frac{\partial \Gamma^{-1}}{\partial x'}(t_0,x')$$

$$= u(t_0) - u^*(t_0) - \int_0^x dx' \, \frac{\partial u^*}{\partial t_0}(\Gamma^{-1}(t_0,x'))\frac{x's^*(T(t_0,x'),x')}{T(t_0,x')}(u(t_0) - u^*(\Gamma^{-1}(t_0,x')))$$

$$= u(t_0) - u^*(t_0) + \int_0^x dx' g(t_0,x')(u(t_0) - u^*(\Gamma^{-1}(t_0,x')))$$

where

$$g(t_0,x) = -\frac{\partial u^*}{\partial t_0}(\Gamma^{-1}(t_0,x'))\frac{x's^*(T(t_0,x'),x')}{T(t_0,x')}$$

This simple integral equation has the solution

$$u(t_0) - u^*(\Gamma^{-1}(t_0,x')) = h(t_0,x)(u(t_0) - u^*(t_0))$$

where

$$h(t_0,x) = \exp\left(\int_0^x dx' g(t_0,x')\right)$$

has the properties claimed for it in the statement of the lemma. **Q.E.D.**

Now changing variables in the asymptotic formula for J_0, and applying the above relation to both this and the formula for ∇J_0, you obtain

$$J_0[u] = \int\int dt_0(u(t_0) - u^*(t_0))^2\, dx\, B_0^*(t_0, x)h(t_0, x)(r^*(\Gamma^{-1}(t_0, x)))^2 + O(\lambda)$$

$$\nabla J_0[u](t_0) = (u(t_0) - u^*(t_0)) \int dx\, B_1^*(t_0, x)h(t_0, x)(r^*(\Gamma^{-1}(t_0, x)))^2 + O(\lambda)$$

where

$$B_0^*(t_0, x) = B_0(\Gamma^{-1}(t_0, x), x)\frac{s(T^*(t_0, x), x)}{s^*(T^*(t_0, x), x)}.$$

Now B_0^* and B_1^* differ at each point in the mute zone by factors or divisors of s, s^*, T, and the like, and these are bounded over the mute zone uniformly in $v \in \mathcal{A}$. Therefore there exists a constant $C > 0$ depending only on $\mathcal{A}$ for which

$$J_0[u] \le C \int dt_0\, (u(t_0) - u^*(t_0))\nabla J_0[u](t_0) + O(\lambda)$$

and we have proved the

Theorem: If u, the RMS square slowness for $v \in \mathcal{A}$, is a stationary point of $J_0[u]$, then $J_0[u] = O(\lambda)$.

That is, for noise free data, any stationary point of J_0 is a global minimizer, up to an asymptotically vanishing error.

NOISE: GENERAL CASE

Suppose that the data $S(t, x)$ is the sum of model-consistent data and another field, regarded as noise or error:

$$S(t, x) = S^*(t, x) + E(t, x)$$

where "model-consistent" means as before

$$S^*(t, x) = r^*(T_0^*(t, x)) + O(\lambda)$$

and $E(t, x)$ is arbitrary (but finite "energy" = mean square).

Since there are several data running around in this part of the discussion, include the name of the data in the notation for the differential semblance objective:

$$J_0[v, S] = \frac{1}{2}\|H\phi F[v]WG[v]S\|^2$$

etc. Then

$$J_0[v, S] = J_0[v.S^*] + J_0[v, E] + K[v, S^*, E]$$

where

$$K[v, S^*, E] =< H\phi F[v]WG[v]S^*, H\phi F[v]WG[v]E >$$

$$= \int\int dx\, dt\, \phi \left(\frac{\partial}{\partial x} + p\frac{\partial}{\partial t}\right) S \left[(I - \nabla^2)^{-1}\phi \left(\frac{\partial}{\partial x} + p\frac{\partial}{\partial t}\right) E\right]$$

satisfies

$$|K[v, S^*, E]| \le C\|S^*\|\|E\|$$

Here and in the following, C will stand for a constant uniform over $v \in \mathcal{A}$ (though the precise value may vary from display to display).

Likewise,

$$J[v, E] \le C\|E\|^2$$

Similarly, the gradients *with respect to RMS square slowness* u satisfy

$$\nabla J_0[v, S] = \nabla J_0[v.S^*] + \nabla J_0[v, E] + \nabla K[v, S^*, E]$$

and

$$\|\nabla K[v, S^*, E]\| \le C\|S^*\|\|E\|$$
$$\|\nabla J_0[v, E]\| \le C\|E\|^2$$

Suppose that u (or its corresponding v) is a stationary point of $J_0[v, S]$, i.e. $\nabla J_0[v, S] = 0$. Then

$$J_0[v, S] \le J_0[v, S^*] + C\|E\|(\|S^*\| + \|E\|)$$
$$\le C(< u - u^*, \nabla J_0[v, S^*] > +\|E\|(\|S^*\| + \|E\|)) + O(\lambda)$$
$$= C(< u - u^*, \nabla J_0[v, S] - \nabla J_0[v, E] - \nabla K[v, S^*, E] > +\|E\|(\|S^*\| + \|E\|)) + O(\lambda)$$
$$\le C\|E\|(\|S^*\| + \|E\|) + O(\lambda)$$

If you presume that the data error is less than 100%, i.e. $\|E\| \le \|S^*\|$, which seems reasonable (or pick any other fixed percentage, if 100% seems wrong to you - just absorbs in C), then this becomes

$$J_0[v, S] \le C\|E\| + O(\lambda)$$

That is,

Theorem: At a stationary point of the differential semblance objective, its value is bounded by a $\mathcal{A}$-uniform multiple of the distance of the data to the set of model-consistent data.

Thus for a family of data converging to model-consistent data, any set of corresponding stationary points of J_0 must have J_0 values which converge to zero, modulo asymptotic errors.

This result may well *not* imply that stationary points for noisy data are global minima. Indeed, substitute the "target" velocity v^* in the expression for $J_0[v, S]$: from the expansion and estimates above you easily see that

$$J[v^*, S] \le C(\|E\|^2 + O(\lambda)\|E\|)$$

Certainly one hopes that the asymptotic error is no worse than other errors, in particular than the data error E, so this inequality effectively implies that the global minimum value of $J_0[\cdot, S]$ is proportional to $\|E\|^2$ for near consistent data, whereas the theorem shows only that the stationary values are proportional to $\|E\|$, so presumeably larger at least in some cases.

In the next section I will show that when the differential semblance minimization is supplemented with proper constraints on the velocity model, in addition to those already imposed, the error in the *RMS square slowness* is proportional to the error in the data. It then follows from the estimates above that stationary values conforming to these constraints are indeed proportional to the square of the error level, hence essentially global minima. It would be interesting to know whether relaxing these constraints actually permits anomalously large stationary values.

DATA DRIVEN MODEL PARAMETRIZATION AND OPTIMAL ERROR ESTIMATES

Up to this point I have imposed only minimal constraints on the RMS velocity, namely those necessary to justify use of the convolutional model. Most velocity analysis imposes far more stringent constraints, either explicitly or implicitly, in the form of *parsimonious parametrization* or *regularization*. In the former case, the choice of parameters (eg. how many spline nodes, where to place them) is *ad hoc*. In the latter, the type of regularization (first derivative, second derivative,...) and the choice penalty weight are also obscure.

In this section I suggest that the differential semblance objective itself supplies a mechanism for constraining the velocity to a parsimoniously parametrized space. I'll propose a choice of subspace within which

- the global minimum is unique for noise free data;
- the error in RMS square slowness is proportional to the error in the data, and so
- any stationary values are proportional to the square of the data error energy, so essentially global minima.

Assume until further notice that the data is free of noise:

$$S(t,x) = r^*(T_0^*(t,x)) + O(\lambda)$$

The Key Lemma proved in the last section then implies that the Hessian $\nabla\nabla J_0$ takes the form

$$\nabla\nabla J_0[u]\delta u(t_0) = \delta u(t_0) \int dx\; B_1^*(t_0,x)h(t_0,x)(r^*(\Gamma^{-1}(t_0,x)))^2 + O(\lambda, \|u-u^*\|)$$

$$= \tilde{R}[u](t_0)\delta u(t_0)$$

While the expression for $\tilde{R}$ above is not easily computable, the approximation

$$R[u](t_0) = \int dx\; (r^*(\Gamma^{-1}(t_0,x)))^2$$

is simply the stack of the squared prestack reflectivity estimates, and therefore an inexpensive byproduct of the computation. At $u = u^*$,

$$\nabla\nabla J_0[u^*]\delta u(t_0) = b[u](t_0)R[u^*](t_0)\delta u(t_0)$$

i.e. the Hessian is actually the approximation followed by a positive diagonal scaling.

Now suppose that u^* differs from a reference square slowness u^0 (in practice, an initial estimate) by a member of a space W. Introduce an inner product in the space of W by

$$\langle w_1, w_2\rangle_2 = \int \frac{d^2 w_1}{dt_0^2}\frac{d^2 w_2}{dt_0^2}$$

To make this inner product positive definite, thus defining a Hilbert space structure, assume furthermore that

$$w(0) = \frac{dw}{dt_0}(0) = w(t_0^{\max}) = \frac{dw}{dt_0}(t_0^{\max}) = 0$$

Thus W is a subspace of the Sobolev space $H_0^2([0, t_0^{\max}])$.

Since the interval velocities, hence the RMS square slownesses, are supposed to vary over a bounded set in C^∞, membership in $\mathcal{A}$ entails a bound on the W norm of $u - u^0$.

Let $g(t_0^1, t_0^2)$ be the Green's function for the operator

$$\frac{d^4}{dt_0^4}$$

with the boundary conditions stated above. Then the W gradient of J_0 restricted to $u^0 + W$ is

$$\nabla_W J_0[u] = \mathcal{G}\nabla J_0[u]$$

in which $\mathcal{G}$ denotes the operator with kernel g. Similarly,

$$\nabla_W\nabla_W J_0[u] = \mathcal{G}\nabla\nabla J_0[u]$$

Next suppose that $H[u]$ is uniformly positive definite for all $u \in \mathcal{A}$. That is, there exist $0 < h_* \le h^*$ for which

$$h_*\|w\|^2 \le \langle w, H[u]w\rangle_2 \le h^*\|w\|_2^2$$

for all $w \in W, u \in \mathcal{A}$.

Then there exists a similar uniform bound for $\nabla_W\nabla_W J[u]$, since the latter differs from $H[u]$ by a diagonal scaling operator with uniform upper and lower bounds over $\mathcal{A}$. For the same reason,

$$\|\nabla_W J_0[u]\|_2 = \|\mathcal{G}b[u]H[u](u - u^*)\|_2 \ge l_*\|u - u^*\|_2$$

for a suitable $l_* > 0$.

That is: within $u^0 + W$, u^* is the unique stationary point of J_0.

Moreover, consulting the estimates of the last section, you see that if the search is limited to $u^0 + W$, then at a stationary point u,

$$J_0[u, S] \leq C(\|E\|^2)$$

as claimed, since the cross-term $K[u, S^*, E]$ in the notation of that section is bounded by a multiple of $\|u - u^*\|_2$.

Finally, how does one lay hands on such a paragon of a function space as W with the properties supposed here? The operator $H[u]$ is symmetric positive semidefinite on $H_0^2([0, t_0^{\rm max}])$. An optimal choice for W is the direct sum of eigenspaces of $\nabla_W \nabla_W J_0[u^*]$ corresponding to the eigenvalues above the cutoff level h_*. A computable estimate of this space is the corresponding direct sum of eigenspaces of $H[u]$. A basis consists of eigenfunctions of the Sturm-Liouville problem

$$\frac{d^4 w}{dt_0^4} = \frac{R[u]}{\lambda} w,$$

$$w(0) = \frac{dw}{dt_0}(0) = w(t_0^{\rm max}) = \frac{dw}{dt_0}(t_0^{\rm max}) = 0$$

To construct W, find the eigenfunctions of this problem, and choose those whose eigenvalues lie above a "suitable" cutoff.

Note that if there is little data in a t_0 interval, $R[u]$ will be small in that interval and eigenfunctions of the 4th derivative operator will smoothly interpolate values to either side. Thus my suggested space implicitly "picks events" with significant energy, pins the RMS velocity down at those places, and interpolates between "events" - just as a human velocity analyst would.

It remains to analyse this "picking" effect, and to devise good algorithms for choosing the eigenvalue cutoff as a function of data quality and success in fitting moveout (i.e. minimizing J_0), so as to justify the assumption that $u^* \in u^0 + W$. But that's another story...

ACKNOWLEDGEMENT

The author performed part of the work reported here while a guest of the Stanford Exploration Project. The author thanks the Director of SEP, Prof. Jon Claerbout, and the other project members for their hospitality and for the stimulating intellectual atmosphere of SEP.

This work was partially supported by the National Science Foundation under grant numbers DMS 9404283 and DMS 9627355; the Office of Naval Research under grant number N00014-96-1-0156; the U. S. Department of Energy under grant number DE-FG07-97 ER14827; and The Rice Inversion Project. TRIP Sponsors for 1998 are Amerada Hess, Amoco Production Co., ARCO Oil and Gas, Conoco Inc., Discovery Bay, Exxon Production Research Co., Mobil Research and Development Corp., and Shell International Research.

REFERENCES

Al Yahya, K., 1989, Velocity analysis by iterative profile migration: Geophysics, **54**, 718–729.

Araya, K., Kim, S., Nolan, C. J., and Symes, W. W., 1996, Velocity inversion and high frequency asymptotics: Mathematical Methods in Geophysical Imaging: Society of Photooptical Instrumentation Engineers.

Cao, D., Singh, S., and Tarantola, A., 1990, Simultaneous inversion for background velocity and impedance maps: Geophysics, **55**, 458–469.

Chauris, H., Noble, M., and Podvin, P., 1998a, Testing the behaviour of differential semblance for velocity optimization: European Association for Geoscientists and Engineers Annual Meeting.

Chauris, H., Noble, M., and Podvin, P., 1998b, Testing the behaviour of differential semblance for velocity estimation: 68th Annual International Meeting and Exposition, Society of Exploration Geophysicists, 1305–1308.

Claerbout, J. F., 1992, Earth sounding analysis: Processing versus inversion: Blackwell, Boston.

Clément, F., and Chavent, G., 1993, Waveform inversion through MBTT formulation *in* Kleinman, E., Angell, T., Colton, D., Santosa, F., and Stakgold, I., Eds., Mathematical and Numerical Aspects of Wave Propagation:: SIAM.

Fowler, P., 1986, Migration velocity analysis by optimization: linear theory: Proc. 56th Annual International Meeting, Society of Exploration Geophysicists, 660–662.

Gauthier, O., Tarantola, A., and Virieux, J., 1986, Two-dimensional nonlinear inversion of seismic waveforms: Geophysics, **51**, 1387–1403.

Gockenbach, M. S., and Symes, W. W., 1997, Duality for inverse problems in wave propagation *in* Biegler, L., Coleman, T., Santosa, F., and Conn, A., Eds., Large Scale Optimization:: Springer Verlag.

Gockenbach, M. S., Symes, W. W., and Tapia, R. A., 1995, The dual regularization approach to seismic velocity inversion: Inverse Problems, **11**, no. 3, 501–531.

Kern, M., and Symes, W. W., 1994, Inversion of reflection seismograms by differential semblance analysis: Algorithm structure and synthetic examples: Geophysical Prospecting, **99**, 565–614.

Kim, S., and Symes, W. W., 1998, Smooth detectors of linear phase: Inverse Problems, **14** **(1)**, 101–112.

Kolb, P., Collino, F., and Lailly, P., 1986, Prestack inversion of a 1D medium: Proceedings of IEEE, **74**, 498–506.

Martinez, R., and McMechan, G., 1991, τ-p seismic data for viscoelastic media – Part I: Modelling; Part II: Linearized Inversion: Geophysical Prospecting, **39**, 141–181.

Minkoff, S. E., and Symes, W. W., 1997, Full waveform inversion of marine reflection data in the plane-wave domain: Geophysics, **62**, no. 2, 540–553.

Reshef, M., 1997, The use of 3D prestack depth imaging to estimate layer velocities and reflector positions: Geophysics, **62**, 206–210.

Scales, J., Smith, M., and Fischer, T., 1991, Global optimization methods for highly nonlinear inverse problems *in* Cohen, G., Halpern, L., and Joly, P., Eds., Mathematical and Numerical Aspects of Wave Propagation Phenomena:: SIAM, 434–444.

Sen, M., and Stoffa, P., 1991, Nonlinear one-dimensional seismic waveform inversion using simulated annealing: Geophysics, **56**, 1624–1636.

Sevink, A., and Herman, G., 1993, *in* Kleinman, R., Angell, T., Colton, D., Santosa, F., and Stakgold, I., Eds., Mathematical and Numerical Aspects of Wave Propagation:: SIAM.

Song, H., 1994, On a transmission inverse problem: Ph.D. thesis, Computational and Applied Mathematics Department, Rice University, Houston, Texas, U.S.A.

Symes, W. W., and Carazzone, J., 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, no. 5, 654–663.

Symes, W. W., 1993, A differential semblance criterion for inversion of multioffset seismic reflection data: J. Geoph. Res., **98**, 2061–2073.

Symes, W. W., 1997, High frequency asymptotics and velocity estimation: Depth Imaging Symposium: Geophysical Society of Tulsa.

Symes, W. W., 1998, Differential semblance estimation of velocities: Progress report: TRIP 1998 Annual Report (`http://www.trip.caam.rice.edu`).

Taner, M., and Koehler, F., 1969, Velocity spectra: digital computer derivation and application of velocity functions: Geophysics, **34**, 859–881.

Toldi, J., 1985, Velocity analysis without picking: Proc. 55th Annual International Meeting, Society of Exploration Geophysicists.

Yilmaz, O., and Chambers, R., 1984, Migration velocity analysis by wavefield extrapolation: Geophysics, **49**, 1664–1674.