Differential Semblance: Formulation and Implementation

William. W. Symes *

ABSTRACT

Differential semblance is a variational principle for determination of wave velocities in seismic data processing, whose analytical and numerical properties make it amenable to local optimization methods. All proper differential semblance formulations share the same abstract skeleton, whatever the underlying physical description of seismic waves. This common mathematical structure suggests that a common computational implementation is feasible. The Standard Vector Library facilitates almost direct translation of differential semblance into working code.

Introduction

All velocity estimation methods in reflection seismology are based on the redundancy of seismic data: data sets contain sufficient information to produce many partial, overlapping images, parametrized by one or more acquisition parameters. The parameters indexing partial images may be field-recorded trace headers or simple combinations (shot position, offset) or the result of considerable pre-processing (slowness or ray parameter, scattering angle and azimuth). The fundamental utility of data redundancy is expressed in the *semblance principle*: that choice of correct velocity will produce partial images whose structure does not depend (strongly) on the indexing acquisition parameter(s).

The differential semblance ("DS") principle for velocity estimation (Symes, 1986) specializes the semblance principle by requiring that the alignment of structure across acquisition parameters be *measured* by norming differences of images for *neighboring* parameter values. A consequence of the differential measurement of alignment is the asymptotic stability of the variational principle: its shape is essentially independent of data frequency content, so long as spatial and temporal sampling is adequate. In some cases, it has been established that the stationary points of the DS variational principle are global minima, at

*The Rice Inversion Project, Department of Computational and Applied Mathematics, Rice University, Houston TX 77251-1892 USA, email symes@caam.rice.edu

least approximately (Symes, 1991; Symes, 1999; Symes, 2001b), rigorously justifying the use of local optimization methods for velocity analysis. Numerical experiments with synthetic and field data have conformed to the theoretical conclusions. In some (though not all!) places the Earth is kind enough to conform sufficiently well to the theory underlying DS that its application produces usable velocities (Symes and Carazzone, 1991; Symes and Versteeg, 1993; Kern and Symes, 1994; Araya et al., 1996; Symes, 1997; Minkoff and Symes, 1997; Chauris et al., 1998; Chauris and Noble, 2001; Mulder and ten Kroode, 2001a). Thus DS may provide at least a starting point for automation of velocity analysis.

DS differs from standard approaches to velocity analysis by being quantitative and variational, rather than qualitative and interactive (Yilmaz, 1987; Taner and Koehler, 1969; Deregowski, 1990; Kleyn, 1983). It differs from other suggested quantitative approaches to inversion of velocity from reflection data (Al Yahya, 1989; Cao et al., 1990; Fowler, 1986; Frazer and Sun, 1998; Toldi, 1985; Plessix et al., 1995; Kolb et al., 1986; Beydoun et al., 1990; Gauthier et al., 1986; Ikelle et al., 1986; Tarantola, 1984; Tarantola, 1986; Tarantola, 1987; Tarantola, 1988; Crase et al., 1990; Mora, 1986b; Mora, 1986a; Mora, 1987; Jervis et al., 1993; Sen and Stoffa, 1991b; Sen and Stoffa, 1991a; Santosa and Symes, 1989) in that only images from nearby parameter values are compared to guage alignment. Finally, it differs from the various forms of reflection tomography (Bishop et al., 1987; Stork and Clayton, 1991; Stork and Clayton, 1992; DelPrat-Jannaud and Lailly, 1993; Biondi, 1992; Billette and Lambaré, 1998; Chauris, 2000) in that the input data are waveforms, rather than picked traveltimes.

The primary purpose of this report is to set out a proper formulation of differential semblance based on its abstract structure, so that an implementation following the template presented here will have the principal favorable mathematical and computational properties promised by the theory. DS applies to a wide variety of primary reflection models, from the simplest layered acoustic models using hyperbolic moveout to 3D finite difference anisotropic viscoelasticity. It accomodates multiple short scale parameter fluctuations, as occurs in primaries only modeling of elastic reflectivity. It also accomodates multiple reflection modes, eg. P-P, P-S, S-P and S-S together as predicted by isotropic elastic modeling, and even the more complicated anisotropic variants. All these variants of DS share the the same abstract structure, with the same key players in the same roles: model and data spaces, subdivision of the model into slowly varying background parameters and rapidly varying parameter perturbations, a modeling operator together with its derivatives and other associated operators, and a semblance operator. The abstract structure sketched below organizes these players according to role, rather than implementation.

A secondary goal is to illustrate the effectiveness of the object oriented programming paradigm in capturing, via reusable code, the features of DS and similar applications common to all instances. I use the C++ Standard Vector Library ("SVL") to encode the DS principle, as expounded here, in a natural way that is immediately accessible to modern optimization algorithms (also expressed in SVL).

The effectiveness of DS, and of the semblance principle, is circumscribed by theoretical constraints, some well-known, some discovered only recently. After a review of these caveats, I introduce the basic theoretical structure on which DS is based, and its formal definition. I include sketches of proofs of the basic properties of DS, reserving detailed mathematical derivations for another place. For single-mode modeling, an alternative derivation is possible, which was in fact the original definition to DS; I show that it is equivalent to the definition introduced here. DS velocities are quite sensitive to coherent noise; I briefly describe the *dual regularization* method, which renders DS insensitive to moderate amounts of coherent noise. Computation of the gradient is a critical step in any local optimization approach, and follows a standard pattern. Unlike stack power optimization, DS is essentially insensitive to event and migration amplitudes: I explain how DS vanishes on model-consistent data, regardless of amplitude treatment, to leading order in frequency. The translation into nearly functional code, lacking only C++ boilerplate and numerical kernels (written in anything appropriate) to be compilable, is quite straightforward, thanks to the structure provided by SVL. I display explicitly the core code of an SVL class implementing the DS operator, and discuss briefly opportunities for performance optimization present even at this very abstract code level.

Limitations

The semblance principle itself is not valid under all circumstances: in particular, it fails when multiple ray paths induce ambiguity in partial image structure (Nolan and Symes, 1997). Therefore DS shares this limitation, and the version developed in this paper presumes a strong ray-theoretic constraint which rules out image ambiguity. This constraint is generally valid locally, except for some quasi-shear modes, and always microlocally (i.e. after filtering to produce suitable source, receiver, or scattering angle-azimuth beams, see (Brandsberg-Dahl et al., 1999)). Development of DS modifications which do not suffer from this limitation is a subject of active research, and is outside the scope of this report.

Another limitation of DS, shared (implicitly) by all velocity analysis methods in current use, is the assumption that all energy in the data records may be regarded as primary reflection. Of course a great deal of effort in reflection seismic processing has gone into development of techniques for suppression of multiple reflections, often the most troublesome source of non-primary energy (see for example (Yilmaz, 1987)). DS can be used in conjunction with any combination of multiple-suppression technologies, as can any velocity analysis tool. The variational nature of DS even leads to a multiple-suppression approaches which would not necessarily be appropriate, or even possible, in an interactive setting (Gockenbach and Symes, 1999; Symes, 2001a; Mulder and ten Kroode, 2001b).

In principle, the alternative to suppression of multiple reflections - along with refractions, dispersion, and other features ignored in the theory underlying DS - is to *model* them, and account for their presence explicitly in forming images volumes. This idea is a recurring theme in the work cited above on other (data-fitting) variational approaches to velocity analysis, but research on differential semblance with fully nonlinear modeling is just beginning.

Model and Data Spaces and Operators

All of the vector spaces in this subject are Hilbert spaces. The inner product of space X is $<\cdot,\cdot>_X$, and the corresponding norm is $\|\cdot\|_X$.

The data space is divided into *bins* or *gathers*, representing various redundant views of Earth structure. Standard binning schemes are all fine: common shot, common offset, and plane wave for example. Common midpoint binning is actually a skewed view of common offset binning appropriate for (near) layered modeling as I will show in the last section. Therefore the data space is best viewed as a product

$$D = \Pi_{i=1}^I D_i$$

where I is the number of bins and D_i is the data space of the ith bin. Note that these bin data spaces are not *a priori* identical - they may differ in numbers of records (traces) and even in samples per record.

The design of the model space and operator components of the algorithm rests on three key assumptions underlying DS:

1. The model is naturally divided into slowly varying background and rapidly varying fluctuations. The forward model is linearized about the slowly varying background, with the rapidly varying fluctuations serving as model perturbation. Primary energy is modeled as the output of the linearized forward model.

2. The rapidly varying fluctuations = *reflectivity* are *a priori incoherent*, i.e. depend on data bin parameter: there is one reflectivity map per data bin. Of course a final model must consist of a single reflectivity map, but in DS this single map is a result of a postprocess combining the various reflectivity bins.

3. The (in)coherence or (dis)semblance of the model, as measured by the mean square differences of *neighboring* reflectivity bins, determines the model quality: the goal of differential semblance optimization is to adjust the model to reduce this differential measure of incoherence to its minimum, while maintaining fit to data.

Remark. In fact assumption (1) may be too restrictive - formally, differential semblance applies just as well to basic elastodynamic models, without a (necessarily artificial) division of the model into smooth background and oscillatory perturbation, and linearization of the response. However, with one rather special exception (Symes, 1991), nothing is known mathematically about this more general version of DS, and very little computational work has been done so far along these lines. So in this document I will stick with the explicit scale dichotomy and the linearized model of primary energy.

Remark. The semblance principle is valid only locally at best, as remarked in the introduction. "Locally" may mean "in the entire region of interest", of course, and for near-layered structures with gentle lateral and vertical velocity variations this it actually the case. This fortunate coincidence is responsible for the success of contemporary velocity analysis. A sufficient condition for validity of the semblance principle, amongst other things, is described in (Symes, 1995), where it is called the "simple geometry" assumption.

Taking each of these points in turn, first express the model space M as a product

$$M = V \times R$$

of background parameters (typically the important member is velocity, hence the notation) and fluctuations about them (reflectivity).

The reflectivity is itself a product, one factor for each data bin:

$$R = \Pi_{i=1}^{I} R_i$$

In contrast to the data, there is generally no reason for the reflectivity factor spaces to differ: they all represent sections of the same Earth model. A possible exception to this rule: since the various data factors (bins) may contain in some cases information about limited and changing regions in the subsurface, the reflectivity factors could be cropped or truncated for the sake of economy.

Each reflectivity factor may itself have several components: for example a model of isotropic elastic primary reflectivity might assign a maps of p-wave impedance, s-wave impedance, and density to each bin:

$$R_i = \Pi_{j=1}^{J} R_{j,i}$$

in which $R_{j,i}$ represents a single parameter (elastic tensor component, density, attenuation parameter,...) which is of course a function of the space variables.

The operators and functions to be defined below are functions of the background parameters, but linear or quadratic in the reflectivity and data variables. Therefore differentiability has a clear meaning for the data and reflectivity components as soon as a norm is defined in the corresponding spaces. Pinning down the sense of differentiability in the background components is a subtler task, which is eased by assuming that V is a finite dimensional submanifold of $C^\infty(\mathbf{R}^d)$ ($d = 1, 2$, or 3 being the space dimension of the model). For example compressional wave velocity $v_p(\mathbf{x})$ might be parametrized as

$$v_p((\mathbf{x})) = v_p^0, \exp(\sum_{n=1}^{N} v_p^i \phi_i(\mathbf{x}))$$

for suitable C^∞ shape functions ϕ_i. This assumption conforms to computational practice, although in actual computation the shape functions are typically splines with finite order of regularity (Symes, 1997; Gockenbach and Symes, 1999; Chauris and Noble, 2001; Mulder and ten Kroode, 2001b). This deviation presumeably is of little importance, as the computational results conform well to the theory developed here.

Like all of the vector spaces mentioned in this report, the bin spaces D_i are Hilbert spaces, with inner product $< \cdot, \cdot >_{D_i}$ and associated norm $\| \cdot \|_{D_i}$. The reflectivity spaces have similar structure. Generally the factor inner products are (possibly diagonally weighted) Euclidean (L^2 in the continuum limit), and the inner product on the product space D is the natural one induced from the factors, though other choices are possible.

A naive differential measure of semblance or coherence is an operator ∂ mapping the reflectivity space R into itself, approximating the derivative in the bin direction. If neighboring bin indices represent neighboring acquisition parameters, the simple choice

$$\partial r_i = \frac{1}{\Delta_i}(r_{i+1} - r_i),\ i = 1, ..., N-1;\ \partial_0 r_I = 0$$

works just fine. The bin lengths Δ_i need not be uniform, but must be small enough that ∂ approximates well the derivative in the bin parameter, for the frequency content present in image volumes. Extensions accomodating more bin parameters are necessary for some 3D acquisition geometries, but this report will treat the bin parameter as if it were scalar.

For theoretical purposes, it is *essential* that the operator used to measure differential semblance *avoid increasing high frequency content.* Of course a simple approximation to the derivative, like the one just described, does indeed increase high frequency content. Therefore it should be followed, in principle, by a smoothing operator, such as a $-\frac{1}{2}$ power H of the Laplace operator: that is,

$$\partial_0 = H\partial$$

so that ∂_0 is a *pseudodifferential operator* ("ΨDO") of order zero. ΨDOs are characterized by their *pseudolocal* property: their arguments have high frequency components in the *same* location as their inputs. ΨDOs play a very important role in the high frequency asymptotic theory of imaging (references), and in particular in the theory of differential semblance. Operators such as H may be defined by FFTs and/or sparse solvers, or through Radon filtering followed by trace integration.

Remark: The bin parameter is being treated as discrete, and of course in actual computation the other (space and time) variables will be discretized as well. Therefore it is strictly speaking incorrect to call ∂_0 a ΨDO: it is a discrete approximation of some sort to one. Throughout this discussion I will deliberately confound discrete and continuous operators, assuming that the behaviour of the continuum limit objects will be closely mimicked by their discretizations. The approximation theory necessary to justify this blasé attitude hardly exists, but numerical experience so far leads me to be cautiously optimistic that it is correct.

The forward modeling operator $F : M \rightarrow D$ is block diagonal in its second argument:

$$F = \mathrm{diag}\,(F_0, F_1, \ldots),\;\; F_i : V \times R_i \rightarrow D_i$$

As the operator F_i is linear in the second factor, write $(v, r_i) \mapsto F_i[v]r_i$. Each factor operator in turn has additional canonical structure, reflecting the number of parameters in the reflectivity and the number of propagating wave modes or phases:

$$F_i[v]r_i = \sum_{k=1}^{K} B_{i,k} \sum_{j=1}^{J} A_{j,i,k} r_{j,i}$$

where $B_{i,k}$ is a discretized *Generalized Radon Transform* ("GRT") for each mode or phase index k and bin index i, and the "amplitudes" $A_{j,i,k}$ are in general ΨDOs (Beylkin, 1985; Rakesh, 1988; Beylkin and Burridge, 1990; Nolan and Symes, 1997; deHoop and Bleistein, 1997; Stolk, 2000a; Stolk, 2000b).

Remark: This representation is incomplete, as it does not take into account varying ray multiplicity, which is normal rather than exceptional for anisotropic elastodynamics. I am aware of only one rigorous analysis of the forward scattering operator in the presence of characteristics of varying multiplicity (Nolan, 2001).

It is assumed in this paper that the definition of $r_{j,i}$ is such that F_i is an operator of order zero, i.e. $\|F_i[v]r_i\|_s = O(\|r_i\|_s)$, $\|\cdot\|_s = s$th Sobolev norm, for any $s \in \mathbf{R}$. This can always be arranged, though it may require that the components of r_i be merely related to, rather than identical with, fluctuations in mechanical parameters such as density, shear modulus, etc.

Note that ΨDO factors can be traded between $B_{i,k}$ and the $A_{j,i,k}$ modulo smoothing operators, and that other representations may be more natural; the cited references describe the forward map as

$$F_i[v]r_i = \sum_{k=1}^{K} \sum_{j=1}^{J} \tilde{B}_{j,i,k} r_{j,i}$$

where $\tilde{B}_{j,i,k} = B_{i,k}\, A_{j,i,k}$ is a GRT. However these representations are equivalent.

Assume throughout that ray geometry is such that the normal operator $B_{i,k}[v]^* B_{i,k}[v]$ is a ΨDO for all admissible v, all i, k. This is a very restrictive condition, amounting essentially to the requirement that there be no caustics in the incident or illuminating fields (Beylkin, 1985; Rakesh, 1988; Nolan and Symes, 1997; Stolk, 2000b; Stolk, 2000a). Generally caustic-free illumination occurs only locally, so robust velocity analysis techniques must deal with the appearance of caustics and multipathing.

Denote by $G_{i,k}$ a GRT with canonical relation inverse to that of $B_{i,k}$, and of order so that $B_{i,k}G_{i,k}$ and $G_{i,k}B_{i,k}$ are ΨDOs of order zero. $G_{i,k}$ should produce as much reflectivity as possible, in view of its role as a kinematic inverse of $B_{i,k}$. A quantitative statement of this condition is as follows: by standard GRT theory (Beylkin, 1985; Beylkin and Burridge, 1990), there exists a right parametrix GRT $G^{\dagger}_{i,k}$ so that

$$G_{i,k}G^{\dagger}_{i,k} = P_{i,k} \tag{1}$$

where $P_{i,k}$ is a *pseudodifferential projection* operator, ie. a ΨDO whose principal symbol is 1 in some conic set, and vanishes outside of larger conic set (in the cotangent bundle of the data manifold). Require that

$$B_{i,k}P_{i,k} \simeq B_{i,k} \tag{2}$$

That is, up to a smooth error, all reflectivity energy preserved by $B_{i,k}$ is representable as a migrated image under $G_{i,k}$. Standard constructions of GRT pairs as in the references cited achieve this property.

Set

$$G_k = \mathrm{diag}\,(G_{1,k}, G_{2,k}, ..., G_{I,k}) : V \times D \to R$$

and similarly for B_k. Define the partial differential semblance operator for the kth phase by

$$W_k = (I, B_k\partial_0 G_k) : V \times D \to V \times D \tag{3}$$

The differential semblance operator is

$$W = \Pi_{k=1}^{K} W_k : V \times D \to V \times D \tag{4}$$

The differential semblance variational principle is: choose v to minimize

$$J[v;d] = \frac{1}{2}\|\Pi_D W[v]d\|^2 \tag{5}$$

in which Π_D denotes projection on the D factor.

The most important property of the operators defined in the preceding paragraphs is their smoothness:

Theorem. The differential semblance functional J is *smooth*, i.e. possesses partial derivatives, of all orders in v and d, for finite energy data d.

Sketch of Proof: Assume (as I do throughout - see remark above) that sampling is sufficiently fine relative to data frequency content that both bin parameter and within-bin parameters $\mathbf{y}$ (time, spatial coordinates) are effectively continuous variables, and replace the discrete GRTs by continuous ones. The GRTs making up G_k and B_k have amplitudes and phases depending smoothly on the background parameters, thanks to

Hamilton-Jacobi theory. The basic composition result for GRTs then implies that the mode k differential semblance operator takes the form

$$W_k[v] = \tilde{\partial}_0 + P_k[v](\mathbf{y}, \partial_{\mathbf{y}})$$

in which $\tilde{\partial}_0$ is a zero-order ΨDO on D similar to ∂_0 on R (the bin derivative followed by an operator of order -1), and P_k is a ΨDO of order 0 in the within-bin variables. It's evident from its construction that the symbol of P_k depends smoothly on the background parameters. **QED**

Smoothness is a *very* restrictive requirement: it implies that

- The operator used to measure semblance (∂ above) *must* be an (approximate) derivative. A stack operator for example will not do, even though it also measures semblance. Only differential measures of semblance give smooth objectives for velocity estimation, with well-defined limits for high frequency data. Therefore only automatic velocity analysis methods based on DS can successfully *and robustly* use gradient based optimization to update velocities (Kim and Symes, 1998).
- G_k must be kinematically inverse to B_k, i.e. each component of G_k is a prestack migration operator corresponding to the modeling operator B_k.

Note that the GRTs B_k and G_k, and the modeling operator F are *not* differentiable, strictly speaking: if velocities are amongst the background model parameters, then perturbation of these typically results in time shifts in the predicted data, which in turn drastically promote high frequency components. Thus small perturbations in parameters give very large and nonlinear perturbations in output, which are poorly predicted by the linearization. These maps are differentiable if one assigns a weaker metric in their ranges, or a stronger metric in their domains ("differentiable with loss of a derivative"); the gradient computation explained below uses their derivatives in this sense below. Smoothness of W_k etc. therefore involves cancellations of large quantities, which suggests that care must be taken in numerical implementation.

Alternate Form for Single Wave Modes

For single-mode modeling ($K = 1$) another definition of DS is possible, which was in fact the definition originally suggested (Symes, 1986). Dropping the mode index k, write the forward map as

$$F_i[v]r_i = B_i[v] \sum_{j=1}^{J} A_{j,i}[v]r_{j,i}$$

or (suppressing the bin index i)

$$F[v]r = B[v] \sum_{j=1}^{J} A_j[v]r_j$$

Then a straightforward statement of the DS concept is the equality constrained optimization problem:

$$\min_{v,r} \sum_j \|\partial_0 r_j\|^2 \text{ subj } F[v]r = d \tag{6}$$

The same approach applies to multi-mode modeling when the modes can be extracted separately from the data; in this case, each mode can be treated independently in the fashion suggested here. For example, it is to some extent possible to extract from three-component seismograms separate P-P, P-S/S-P, and S-S reflection records; these can then be treated independently (so far as reflectivities are concerned - the background parameters (velocities) are of course entangled!).

Theorem: Suppose that $\sum_j A_j^2$ is elliptic. Then this variational principle is equivalent to the single mode version of the DS principle presented in the last section, after perturbation of constraint and objective. The perturbation is small when the data is highly oscillatory.

The meaning of "equivalent" is that, when r is eliminated via (a perturbation of) the first order necessary conditions for optimality, the resulting unconstrained problem for v is identical to that described in the last section.

Note that since the amplitude operators A_j are order zero and have real symbols, hence are self-adjoint modulo smoothing errors, the ellipticity assumption simply means that at no point does the model produce zero total reflectivity for all parameter combinations. This condition is satisfied by simple acoustic and elastic models, and seems minimal for successful inversion.

Sketch of Proof: The GRTs B are elliptic by standing assumption; specifically, assume that they possess left parametrices, denoted optimistically by B^{-1}. Therefore the constraint can be written as

$$\sum_{j=1}^{J} A_j[v]r_j \simeq B[v]^{-1}d$$

The operators A_j and ∂_0 are ΨDO of order zero, so the standard calculus would seem to imply that they commute modulo a relatively smoothing remainder. This conclusion is not quite correct, even ignoring discretization issues, as A_j is a ΨDO in the within-bin parameters (spatial location, as it applies to reflectivities), parametrized by the bin parameter, and is not in general a ΨDO in all parameters simultaneously - unlike ∂_0. However the well-known fix for this problem (Taylor, 1975) amounts to filtering away an unphysical sort of noise for this application, so ignore the issue. Thus

$$\partial_0 \sum_{j=1}^{J} A_j[v]r_j \simeq \sum_{j=1}^{J} A_j[v]\partial_0 r_j$$

with an error that should be small for oscillatory r_j. Introduce

$$s_j = \partial_0 r_j$$

Then for fixed v, the problem (6) is asymptotically equivalent to

$$\min_s \sum_j \|s_j\|^2 \text{ subj } \sum_j A_j s_j = \partial_0 B^{-1} d$$

Introduce the Lagrange multiplier $\lambda \in R$. The first order necessary conditions for this problem are

$$s_j + A_j \lambda = 0,\ j = 0, ...J;\ \sum_j A_j s_j = \partial_0 B^{-1} d$$

whence using the ellipticity assumption

$$\lambda = -(\sum_j A_j^2)^{-1} \partial_0 B^{-1} d$$

and so at the optimum s hence r,

$$\sum_j \|\partial_0 r_j\|^2 = \sum_j \|s_j\|^2$$

$$=< \partial_0 B^{-1} d, (\sum_j A_j^2)^{-1} \partial_0 B^{-1} d >_R$$

and optima of (6) are to be reached by further minimizing this quantity over v. The symbol calculus provides an order zero ΨDO square root C of $(\sum_j A_j^2)^{-1}$, with which you can rewrite the above quantity as

$$=< C\partial_0 B^{-1} d, C\partial_0 B^{-1} d >$$

$$\simeq< \partial_0 C B^{-1} d, (B^* B)^{-1/2} B^* B (B^* B)^{-1/2} \partial_0 C B^{-1} d >$$

$$\simeq \|B \partial_0 G d\|^2$$

in which $G \equiv (B^* B)^{-1/2} C B^{-1}$ is a GRT with canonical relation inverse to that of B. Comparison with (3), (5) shows that this remaining optimization over v is of exactly the form developed in the last section. **QED**

Dual Regularization

Since DS identifies velocity by detecting moveout of coherent events, it is natural to expect that coherent noise will degrade DS velocity estimates (Mulder and ten Kroode, 2001b; Gockenbach and Symes, 1999). *Dual regularization* relaxes the requirement that the model fit the data precisely, or alternatively that the reflectivity volume be the migrated data. This idea was introduced in Gockenbach's thesis (Gockenbach, 1994; Gockenbach et al., 1995).

For single mode modeling, a straightforward relaxation of the alternate form (6) is

$$\min_{v,r} \sum_j \|\partial_0 r_j\|^2 \text{ subj } \|F[v] r - d\| \leq \sigma \|d\| \tag{7}$$

This is equivalent, modulo the usual asymptotic errors, to

$$\min_{v,e} \sum_j \|W[v](d+e)\|^2 \text{ subj } \|e\| \leq \sigma\|d\| \tag{8}$$

and this latter form applies equally well to multi-mode modeling, with W defined by (4) in either case.

Dual regularization insulates DS velocity estimation to a certain extent from the requirement that the data consist of primary events. In particular, it permits accurate and automatic velocity analysis in the presence of some degree of coherent noise, such as multiple reflections. See (Gockenbach and Symes, 1999; Symes, 2001a) for information on the amount and type of noise accomdated by this relaxation of DS.

Gradients

From (3, 4), dropping the bin index i,

$$DJ[v,d]\delta v = \langle DW[v][\delta v, d], W[v]d\rangle_D$$

$$= \langle \sum_{k=0}^{K} \Pi_{k'<k} W_{k'}[v]\,[DB_k[v][\delta v, \partial_0 G_k[v]\Pi_{k''>k}W_{k''}[v]d]$$
$$+B_k[v]\partial_0 DG_k[v][\delta v, \Pi_{k''>k}W_{k''}[v]d]]\ , W[v]d\rangle_D$$

$$= \langle \delta v, \sum_{k=0}^{K} DB_k[v]^*[(\Pi^0_{k'=k-1}W^*_{k'}[v])\,W[v]d,\ \partial_0 G_k[v]\Pi^K_{k''=k+1}W_{k''}[v]d]]$$
$$+DG_k[v]^*[\partial_0^* B_k^*[v][(\Pi^0_{k'=k-1}W^*_{k'}[v])W[v]d, \Pi^K_{k''=k+1}W_{k''}[v]d]]\ \rangle_V$$

from which you can read off the gradient of J. Note what is involved in the computation of $\nabla_v J$:

- the linear operators $B_k[v]$ and $G[v]$ and their adjoints;
- the filtered bin derivative ∂_0 and its adjoint;
- the adjoints of the linear operators $\delta v \mapsto DG_k[v][\delta v, d]$ and $\delta v \mapsto DB_k[v][\delta v, r]$, of course with respect to the inner product on the velocity tangent space V.

The single mode case is worth displaying on its own: in that case,

$$\nabla_v J[v,d] = DB[v]^*[W[v]d, \partial_0 G[v]d] + DG[v]^*[\partial_0^* B[v]^* W[v]d, d] \tag{9}$$

Amplitudes

It has been observed that amplitude variations may strongly affect stack power velocity analysis (Sarkar et al., 2001). Perhaps surprisingly, amplitude choices have essentially no effect on DS - this is also observed in (Sarkar et al., 2001). This notion is implicit in the definitions (4,3) which call merely for migration and demigration operators without specifying any constraints on their effect on amplitudes. In fact, it is a trivial observation that any such choices lead to DS functionals which vanish at consistent data, to leading order in frequency:

Theorem: Suppose that $r \in R$ satisfies $\partial r = 0$. Then $J[v, F[v]r] = O(\|Kr\|^2)$ for a pseudodifferential operator K of order -1.

Remark: Since the entire theory is asymptotic in nature, this is the best possible result concerning the values of the DS functional - and it is independent of migration amplitudes. The location of stationary points, on the other hand, may well be influenced by the amplitudes used to define Gs and Bs. The global analysis of DS stationary points - which of course are the velocities found by gradient-based optimization methods - is in its infancy. The reports (Symes, 1991; Symes, 1999; Symes, 2001b) show that, for layered acoustics, stationary points are essentially global minima, which justifies local optimization for this restricted class of problems. No such result is currently known even for multidimensional acoustics, though numerical evidence is promising (Symes and Versteeg, 1993; Chauris et al., 1998; Chauris and Noble, 2001; Mulder and ten Kroode, 2001b).

Sketch of Proof: The single mode, single parameter case is typical:

$$J[v, F[v]r] = \frac{1}{2}\|B[v]\partial_0 G[v]B[v]A[v]r\|^2$$

$C[v] = G[v]B[v]$ is a ΨDO of order zero, with symbol depending smoothly on v, whereas $A[v]$ has appropriate order p so that $F[v]$ is of order zero, as was assumed above (in particular $B[v]$ is of order $-p$).

Since $\partial r = 0$, the above is

$$= \frac{1}{2}\|B[v][\partial_0, C[v]A[v]r\|^2 = \frac{1}{2}\langle r, [\partial_0, C[v]A[v]]^* B[v]^* B[v][\partial_0, C[v]A[v]]r\rangle$$

The commutator $[\partial_0, C[v]A[v]]$ is not a ΨDO, but via the argument due to Taylor mentioned above (Taylor, 1975) one sees that its effect on functions with essential support like that of r is equivalent to that of a ΨDO in the within-bin variables of order $p-1$. Since $B[v]^*B[v]$ is a ΨDO in the within-bin variables of order $-2p$, the proof (sketch) is complete. **QED**

Implementation

Object oriented programming provides a means to take advantage of the common structure described in the last few sections, by supplying abstract programming constructs which mimic the abstract mathematics as faithfully as possible. The Standard Vector Library ("SVL", (Gockenbach et al., 1999; Scott and Symes, 2001)) supplies computational objects corresponding closely to the mathematical players in my formulation of DS. In fact SVL was designed with exactly this kind of application in mind.

The key DS artifact to implement is the DS operator $W[v]d$ (4). I will describe an implementation this operator in the case of single-mode modeling; the multi-mode case follows exactly the same general pattern but is more compicated. I will display actual code rather than pseudocode, and hope that C++ syntax does not too greatly obscure the ideas.

The key SVL class for this application is `Operator`, which realizes the nonlinear operator concept. The sharp-eyed reader will have realized that the gradient formulas above really involved *partial* derivatives of the modeling operators with respect to both velocity and data or reflectivity - a fact disguised by my use of linear operator notation for the action on data or reflectivity. That is, the various operators are functions on *product spaces*, eg. $V \times D$, a notion also implemented in SVL. The class `OperatorProductDomain` defines the SVL interface for such operators. Therefore the class implementing W, which we shall call `DS`, is a *derived* or child class or subclass of `OperatorProductDomain`.

To implement `DS` as an `OperatorProductDomain`, the user must supply code for several functions (or *methods* in OO-speak), the chief of which are

```
virtual void apply(Vector & x,
                   Vector & y) = 0;

virtual void applyPartialDeriv(int i,
                               Vector & x,
                               Vector & dxi,
                               Vector & dy) = 0;

virtual void applyAdjPartialDeriv(int i,
                                  Vector & x,
                                  Vector & dxi,
                                  Vector & dy) = 0;
```

In these declarations, the names of the methods indicate their purposes, and the arguments are meant to suggest their roles also - `apply` applies the operator to vector `x` to produce output vector `y`, and the others are read similarly. The `Vector` class implements a straightforward notion of vector. Basic SVL classes are *templated* on the scalar type, ie. they are passed the type (`float` or `double`) as a parameter. Only `float` is interesting for seismic calculations, so I've suppressed the template parameter in the code to follow.

The formulas (3,9) involve the operators G, B, ∂_0 and their partial derivatives, the velocity v, also the data vector d, so these must be provided.

The pair (v, d) is encoded in the `Vector` object `x`, which represents a vector in the product space $V \times D$. In SVL, the components of a product vector are accessed through the `Components` class. Accordingly, if `x` is the name of the product vector containing v as its first factor and d as its second, you access v like so:

```
Components xc(x);
Vector & v = xc[0];
```

That is, a `Components` object has an indexing operator which returns references to the factors of a product vector.

The operators G, B and ∂_0 are represented by `OperatorProductDomain` objects `G` and `B` and a `LinearOp` object `D0`. SVL operators have the wherewithal to sanity-check their input and output vectors, and we assume that this is done in the implementations of G, B, and ∂_0.

A natural `apply` method would be

```
virtual void apply(Vector & x,
   Vector & y) {
  G.apply(x,r0);
  D0.apply(cr0[1],cr1[1]);
  B.apply(r1,y);
}
```

The `Vector` objects `r0`, `r1` are work vectors in $V \times D$ - how they come to be available will be explained later. The objects `cr0`, `cr1` are their `Components` objects, which access the 2nd component (index 1) which is of course the reflectivity. Note that I have left out here some standard C++ boilerplate, such as exception handling, which must be included to obtain robust or even compilable code. The code you see above is the gist.

Access to the `apply...` methods of these operators is *protected*, meaning that in general code outside of the class methods cannot call the `apply...` methods. In order to make these methods available to the DS operator code, the implementations of G, B and ∂_0 must declare `DS` to be a *friend* class.

Remark: Protected access to the apply methods is responsible for a considerable computational economy. In general these methods are accessible to non-class code only through the OperatorEvaluation class, which pairs (an independent copy of) the operator object with a vector in the domain. G.apply(x,r0) may create much intermediate data (eg. traveltime tables). Since these calls are mediated by OperatorEvaluation object, which fixes x, it completely safe for G to create any such intermediate data when first needed in any method of G, then preserve the data as persistent internal state of G. Since the OperatorEvaluation object preserves the identity of x between calls, no possibility exists of using internal data inconsistent with the argument x. Evaluation at another input vector will create another OperatorEvaluation object for DS, containing a completely independent copy of G, which will have its own internal data, ditto for B. Thus the OperatorEvaluation construction permits optimal re-use of intermediate computational results - that is why it was introduced (Gockenbach et al., 1999).

Code for application of the partial derivative branches on the component index:

```
virtual void applyPartialDeriv(int i,
                               Vector & x,
                               Vector & dxi,
                               Vector & dy) {
  if (i == 0) {
    G.apply(x,r0);
    D0.apply(cr0[1],cr1[1]);
    B.applyPartialDeriv(0,r1,dxi,dtmp);
    G.applyPartialDeriv(0,x,dxi,r0);
    D0.apply(cr0[1],cr1[1]);
    B.apply(r1,dy);
    Components cdy(dy);
    dy[1].linComb(1.0,dy[1],1.0,cdtmp[1]);
  }
  else if (i == 1) {
    G.applyPartialDeriv(1,x,dxi,r0);
    D0.apply(cr0[1],cr1[1]);
    B.apply(r1,dy);
  }
  else {
    (throw an exception)
  }
}
```

Code for application of the adjoint partial derivatives is very similar - left to the reader as an exercise!

The operator objects `G`, `B`, and `D0` and the various work vectors (`r0` etc.) get attached to and object of the `DS` class at construction, which takes place in special methods called *constructors*:

```
class DS: public OperatorProductDomain {

private:

  OperatorProductDomain G;
  OperatorProductDomain B;
  LinearOp D0;

  Vector r0;
  Vector r1;
  Vector dtmp;

  Components cr0;
  Components cr1;
  Components cdtmp;

public:

  // copy constructor
  DS(const DS & W)
  : G(W.G), B(W.B), D0(W.D0),
    r0(G.getRange()), r1(G.getRange()), dtmp(G.getDomain()),
    cr0(r0), cr1(r1), cdtmp(dtmp) {}

  // basic constructor
  DS(OperatorProductDomain & _G,
     OperatorProductDomain & _B,
     LinearOp & _D0)
  : G(_G), B(_B), D0(_D0),
    r0(G.getRange()), r1(G.getRange()), dtmp(G.getDomain()),
    cr0(r0), cr1(r1), cdtmp(dtmp) {}

protected:

  Space & getRange()  { return G.getDomain(); }
  ProductSpace & getProductDomain() { return G.getProductDomain(); }

  virtual void apply(Vector & x,
     Vector & y);
```

```
  ...
};
```

The `getProductDomain` and `getRange` methods return references to SVL `ProductSpace` and `Space` objects respectively. An operator should know its domain and range, but there is no particular reason for it to divulge this information to external code, so these methods are protected. They are accessible to objects of child and friend classes, which is how the calls to the corresponding methods of `G` and `B` succeed.

Construction of the work vectors `r0`, `r1`, and `dtmp` uses the standard SVL `Vector` constructor, which takes an SVL `Space` as argument ("let x be a member of X..."). These, and the other data members of the class, are destroyed when the DS operator object is destroyed; they are allocated statically, on the stack, and do not require explicit memory management.

The first constructor is a *copy constructor*: it simply copies the internal data of the argument, thus creating a functionally identical copy of the `DS` object. The `OperatorEvaluation` constructor uses this copy constructor (through a *virtual constructor* interface) to produce its internal independent copy of the operator. Note that the internal data of the new object is independent (but identical) to that of the old object. Each internal datum (`G` etc.) is regarded as a new object: any of *its* internal data not initialized on construction is regarded as uninitialized, not as identical to that of the source object. For example a Kirchhoff migration operator using a gridded traveltime map might initialize the grid on construction, but not the traveltime values - these are initialized (via ray tracer, eikonal solver,...) when the Kirchhoff operator is applied for the first time.

Note that in fact the statements which initialize the internal data (`G(W.G)` etc) are themselves calls to copy constructors of the data objects. For this reason the *allocation of memory* for any data not initialized on construction, not just its initialization, should be postponed until the first time the data is needed. Such a design avoids unnecessary allocation of memory for objects used only as prototypes for other objects. Such memory allocation must necessarily be dynamic, so requires the class code to manage the memory explicitly. However this sacrifice in simplicity will be worth the gain in efficiency for larger problems.

The second constructor implements *ab initio* construction of a `DS` object from its constituent parts, and might be used in a driver program.

The attentive reader will have realized that several computations are shared by *all* of the `apply...` methods. For example the reflectivity DS volume $\partial_0 G[v]d$ is a component in all of the computations, and this computation could be amortized over all calls to any of the `apply...` methods by carrying it out as part of a call to any of them, then storing the result as part of the persistent state of the DS operator. Since each work vector is a copy of either the reflectivity or data volume, it is also reasonable to do whatever can be done to eliminate the need for such workspace. At least one of the work vectors used above can be eliminated by rearrangement of the computations; these rearrangements also eliminate passes through the data. A complete discussion of the possible code optimizations available even at this very abstract level of programming differential semblance is beyond the scope of this report.

Discussion

This paper has presented an abstract framework for differential semblance velocity analysis, and overviewed the core properties of the method. The definition of DS used here, unlike older definitions, accomodates multiple wave modes and multiple reflectivity parameters - the general form of the DS operator is independent of these details, and indeed of the internal details of the migration andemigration operators out of which it is built. Dual regularization, to desensitize the method to coherent noise, and computation of the gradient can be described at this level of abstraction, as can the insensitivity of DS to amplitude variation, both along events and in migration operators. Perhaps surprisingly, OO programming methodology makes coding these abstract structures possible.

DS as formulated here shares a principal drawback with all contemporary velocity analysis methods: it is based on primaries only modeling, which is in turn essentially equivalent to linearization and high frequency asymptotic perturbation theory. Of course much, perhaps most, seismic data does not conform well to this model: multiple reflections and other nonlinear and anelastic effects are evident in many field surveys. Moreover, even within these theoretical constraints, the semblance principle is invalid for strongly refracting structures: image gathers may not be flat even at a correct velocity model. These two limitation stand in the way of robust automatic velocity analysis. Recent work on strongly refraction in the primaries only context gives some hope that the second limitation may be overcome. Practical use of nonlinear and anelastic modeling in velocity estimation and imaging is then the next frontier to be explored.

Acknowledgements

This work was supported in part by National Science Foundation grants DMS-9973423, DMS-9973308, and EAR-9977697, by the Los Alamos National Laboratory Computer Science Institute (LACSI) through LANL contract number 03891-99-23, by the Department of Energy EMSP grant DE-FG07-97 ER14827, and by The Rice Inversion Project (TRIP). TRIP sponsors for 2001 were Amerada Hess Corp., Conoco Inc., ExxonMobil Upstream Research Co., Landmark Graphics Corp., Shell International Research, and Western Geco.

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. In *Mathematical Methods in Geophysical Imaging*. Society of Photooptical Instrumentation Engineers.

Beydoun, W., Mendes, M., Blanco, J., and Tarantola, A. (1990). North sea reservoir description: benefits of an elastic migration/inversion applied to multicomponent vertical seismic profile data. *Geophysics*, 55:209–217.

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

Beylkin, G. and Burridge, R. (1990). Linearized inverse scattering problem of acoustics and elasticity. *Wave Motion*, 12:15–22.

Billette, F. and Lambaré, G. (1998). Velocity macro-model estimation from seismic reflection data by stereotomography. *Geophys. J. Int.*, 135:671–680.

Biondi, B. (1992). Velocity estimation by bean stack. *Geophysics*, 57(8):1034–1047.

Bishop, T., Bube, K., Cutler, R., Laingan, R., Love, P., Resnick, J., Shuey, R., Spindler, D., and Wyld, H. (1987). Tomographic determination of velocity and depth in laterally varying media. *Geophysics*, 50:903–923.

Brandsberg-Dahl, S., deHoop, M., and Ursin, B. (1999). Velocity analysis in the common scattering angle - azimuth domain. In *EAGE Expanded Abstracts*, pages 49–53.

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

Chauris, H. (2000). *Analyse de vitesse par migration pur l'imagerie des structures complexes en sismique réflexion.* PhD thesis, École des Mines de Paris.

Chauris, H. and Noble, M. (2001). Two-dimensional velocity macro model estimation from seismic reflection data by local differential semblance optimization: applications synthetic and real data sets. *Geophys. J. Int.*, 144:14–26.

Chauris, H., Noble, M., and Podvin, P. (1998). Testing the behaviour of differential semblance for velocity estimation. In *68th Annual International SEG Meeting*, New Orleans. Society of Exploration Geophysicists. Expanded Abstract.

Crase, E., Pica, A., Noble, M., McDonald, J., and Tarantola, A. (1990). Robust elastic nonlinear waveform inversion: Application to real data. *Geophysics*, 55:527–538.

deHoop, M. and Bleistein, N. (1997). Generalized Radon transform inversions for reflectivity in anisotropic elastic media. *Inverse Problems*, 16:669–690.

DelPrat-Jannaud, F. and Lailly, P. (1993). Ill posed and well posed formulation of the reflection traveltime tomography problem. *J. Geophys. Res.*, 98:6589–6605.

Deregowski, S. (1990). Common offset migrations and velocity analysis. *First Break*, 8:225–234.

Fowler, P. (1986). Migration velocity analysis by optimization: linear theory. In *Proc. 56th Annual International Meeting*, pages 660–662, Houston. Society of Exploration Geophysicists. Expanded abstract.

Frazer, N. and Sun, X. (1998). New objective functions for waveform inversion. *Geophysics*, 63(1):213–222.

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

Gockenbach, M. (1994). *An abstract analysis of differential semblance optimization*. PhD thesis, Computational and Applied Mathematics Department, Rice University, Houston, Texas, U.S.A.

Gockenbach, M. and Symes, W. (1999). Coherent noise suppression in velocity inversion. In *Expanded Abstracts, Society of Exploration Geophysicists, 69th Annual International Meeting*, Tulsa. SEG.

Gockenbach, M., Symes, W., and Tapia, R. (1995). The dual regularization approach to seismic velocity inversion. *Inverse Problems*, 11(3):501–531.

Gockenbach, M. S., Petro, M. J., and Symes, W. W. (1999). C++ classes for linking optimization with complex simulations. *ACM Transactions on Mathematical Software*, 25:191–212.

Ikelle, L., Diet, J., and Tarantola, A. (1986). Linearized inversion of multioffset seismic data in the omega-k domain. *Geophysics*, 51:1266–1276.

Jervis, M., Sen, M., and Stoffa, P. (1993). Optimization methods for 2-D pre-stack migration velocity estimation. In *Proc. 63rd Annual International Meeting*, pages 965–968, Washington, D.C. Society of Exploration Geophysicists.

Kern, M. and Symes, 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.

Kleyn, A. (1983). *Seismic Reflection Interpretation*. Applied Science Publishers, New York.

Kolb, P., Collino, F., and Lailly, P. (1986). Prestack inversion of a 1D medium. *Proceedings of the IEEE*, 74:498–508.

Minkoff, S. E. and Symes, W. W. (1997). Full waveform inversion of marine reflection data in the plane-wave domain. *Geophysics*, 62(2):540–553.

Mora, P. (1986a). Elastic wavefield inversion of reflection and transmission data. *Geophysics*, 53:750–759.

Mora, P. (1986b). Nonlinear 2-d elastic inversion of multi-offset seismic data. *Geophysics*, 52:1211–1228.

Mora, P. (1987). Nonlinear 2-d elastic inversion of real data. In *Proc. 57th Annual International Meeting*, pages 430–432, New Orleans. Society of Exploration Geophysicists. Expanded abstract. Also *Geophysics* 52.

Mulder, W. and ten Kroode, A. (2001a). Automatic velocity analysis by differential semblance optimization. In *Expanded Abstracts, Society of Exploration Geophysicists, 71st Annual International Meeting*, Tulsa. SEG.

Mulder, W. and ten Kroode, A. (2001b). Automatic velocity analysis by differential semblance optimization. *Geophysics.* (submitted).

Nolan, C. (2001). Permittivity recovery from multiple characteristic waves. *SIAM J. Appl. Math.*, 62:448–461.

Nolan, C. and Symes, W. (1997). Global solution of a linearized inverse problem for the wave equation. *Commun. in PDE*, 22(5&6):919–952.

Plessix, R. E., de Roeck, Y. H., and Chavent, G. (1995). Automatic and simultaneous migration velocity analysis and waveform inversion of real data using a MBTT/WKBJ formulation. In *Expanded abstracts from the 65th Ann. Internat. Mtg. of Soc. Expl. Geophys.*, pages 1099–1102.

Rakesh (1988). A linearized inverse problem for the wave equation. *Comm. on P.D.E.*, 13(5):573–601.

Santosa, F. and Symes, W. (1989). *An Analysis of Least-Squares Velocity Inversion*, volume 4 of *Geophysical Monographs.* Soc. of Expl. Geophys., Tulsa.

Sarkar, D., Castagna, J., and Lamb, W. (2001). Avo and velocity analysis. *Geophysics*, 66:1284–1293.

Scott, S. D. and Symes, W. W. (2001). Vector class design: a proposed standard. In *The Rice Inversion Project: Annual Report 2001*, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005-1892. (available electronically at `www.trip.caam.rice.edu`).

Sen, M. and Stoffa, P. (1991a). Nonlinear multiparameter optimization using genetic algorithms: Inversion of plane wave seismograms. *Geophysics*, 56:1794–1810.

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

Stolk, C. (2000a). Microlocal analysis of a seismic linearized inverse problem. *Wave Motion*, 32:267–290.

Stolk, C. (2000b). *On the modeling and inversion of seismic data.* PhD thesis, Universiteit Utrecht.

Stork, C. and Clayton, R. (1991). Linear aspects of tomographic velocity analysis. *Geophysics*, 56:483–495.

Stork, C. and Clayton, R. (1992). Using constraints to address the uncertainties of automated prestack velocity analysis. *Geophysics*, 57:404–419.

Symes, W. (1986). Stability and instability results for inverse problems in several-dimensional wave propagation. In Glowinski, R. and Lions, J., editors, *Proc. 7th International Conference on Computing Methods in Applied Science and Engineering*. North-Holland.

Symes, W. (1991). Layered velocity inversion: a model problem from reflection seismology. *SIAM J. Math. Anal.*, 22:680–716.

Symes, W. (1995). Mathematical Foundations of Reflection Seismology. Technical report, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.

Symes, W. and Carazzone, J. (1991). Velocity inversion by differential semblance optimization. *Geophysics*, 56(5):654–663.

Symes, W. and Versteeg, R. (1993). Velocity model determination using differential semblance optimization. In *Proc. 63rd Annual International Meeting*, pages 696–699, Washington, D.C. Society of Exploration Geophysicists. Expanded abstract.

Symes, W. W. (1997). High frequency asymptotics and velocity estimation. In *Depth Imaging Symposium*, Tulsa, Oklahoma, U.S.A. Geophysical Society of Tulsa.

Symes, W. W. (1999). All stationary points of differential semblance are asymptotic global minimizers: Layered acoustics. In *The Rice Inversion Project: Annual Report 1999*, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005-1892. also appeared in SEP-95.

Symes, W. W. (2001a). Dual regularization and coherent noise suppression in velocity analysis. In *The Rice Inversion Project: Annual Report 2001*, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005-1892.

Symes, W. W. (2001b). Stationary points of differential semblance for layered acoustics. In *The Rice Inversion Project: Annual Report 2001*, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005-1892.

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

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49:1259–1266.

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

Tarantola, A. (1987). *Inverse Problem Theory*. Elsevier.

Tarantola, A. (1988). Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation. In Aki, A. and Wu, R. S., editors, *Scattering and attenuation of seismic waves, PAGEOPH*. Birkhäuser Verlag, Basel.

Taylor, M. (1975). Reflection of singularities of solutions to systems of differential equations. *Comm. on Pure and Applied Math.*, 28:457–478.

Toldi, J. (1985). Velocity analysis without picking. In *Proc. 55th Annual Meeting, SEG*, Washington, DC.

Yilmaz, O. (1987). Seismic data processing. In *Investigations in Geophysics No. 2*. Society of Exploration Geophysicists, Tulsa.