Migration Velocity Analysis and Waveform Inversion

William W. Symes [*]

ABSTRACT

Waveform (output least squares) inversion of seismic reflection data can reconstruct remarkably detailed models of subsurface structure, and take into account essentially any physics of seismic wave propagation that can be modeled. However the waveform inversion objective has many spurious local minima, hence convergence of descent methods (mandatory because of problem size) to useful Earth models requires accurate initial estimates of long-scale velocity structure. Migration velocity analysis, on the other hand, is based on the Born approximation but is capable of correcting substantially erroneous initial velocities. Appropriate choice of objective (differential semblance) turns migration velocity analysis into an optimization problem, for which Newton-like methods exhibit little tendency to stagnate at nonglobal minima. The *extended modeling* concept links these two apparently unrelated approaches to estimation of Earth structure: from this point of view, migration velocity analysis is a solution method for the linearized (single scattering, Born) waveform inversion problem. Extended modeling also provides a basis for a nonlinear generalization of migration velocity analysis. Preliminary numerical evidence suggests that this new approach to nonlinear waveform inversion may combine the global convergence of velocity analysis with the physical fidelity of least squares.

Key words. Velocity analysis, waveform inversion, least squares

INTRODUCTION

Seismic waveform inversion is a synonym for model-based fitting of waveform (sampled hydrophone or geophone) data. It is often formulated as the least squares problem of reducing the energy in the difference between predicted and observed data to a minimum, by varying the model on which the prediction is based. This *output least squares* ("OLS") formulation will be taken as a synonym for waveform inversion throughout the following discussion.

The model-based data-fitting approach to inference of Earth structure from geophysical data is conceptually attractive, and has a long and productive history in the Earth

[*]The Rice Inversion Project, Department of Computational and Applied Mathematics, Rice University, Houston,TX 77005-1892 USA, email `symes@caam.rice.edu`

sciences. Its influence on reflection seismology has nonetheless been limited by two major impediments. The first is the computational intensity of reflection seismogram modeling, especially in 3D. Wavefield modeling and various associated computational tasks are required by OLS inversion, and until recently these computations were beyond reach at industrially or scientifically relevant scales. This computational obstacle is gradually fading, due to continued steady advances in computer hardware performance and algorithmic improvements in modeling.

The second impediment is more fundamental. Least-squares estimates, or rather their *prestack depth migration* ("PSDM") approximations, have become the preeminent tool for estimating *short, or wavelength, scale* Earth structure given a *long scale* model. However the least squares principle has proven poorly suited to inference of long scale structure (often called the *macromodel*), as we shall explain below. Since accurate knowledge of the long scale structure is prerequisite to successful estimation of short scale structure, this circumstance has strictly limited the straightforward application of waveform inversion. Instead, industrial seismology has developed a suite of *migration velocity analysis* ("MVA") techniques, apparently completely independent of waveform inversion, for the estimation of Earth macromodels.

The central thesis of this paper is that MVA is in fact an approach to solving a *partially linearized* variant of the waveform inversion problem (linear in short scales, nonlinear in long scales). The conceptual link between the MVA and waveform inversion is the notion of *extended model*, in which the modeling of wavefields is extended to nonphysical models depending on redundant parameters. I will show that both major variants of MVA, based on the "Kirchhoff" and "wave equation" variants of PSDM, amount to the solution of waveform inversion problems for appropriate extended models. These models use linearized or Born modeling at short scales, but include the full nonlinear (kinematic) effect of the macromodel - hence "partially linearized". This observation provides a conceptual framework for recent attempts to quantify and automate MVA. It also suggests that the extended modeling principle might be applied directly to nonlinear waveform inversion, and that this extended formulation might overcome the fundamental obstacle to effective waveform inversion in reflection seismology. I will present some preliminary evidence concerning the behaviour of nonlinear extended inversion, suggesting that a nonlinear analogue of the *imaging conditions* of MVA may distinguish data-consistent velocity models. I will also discuss the remaining challenges, both conceptual and computational, to full-scale implementation of this approach to nonlinear inversion.

Except for a few allusions, this paper does not include any discussion of *traveltime tomography*. In fact, reflection tomography is also widely used to constrain velocity models of the sedimentary Earth. The focus of this paper is however methods using waveform data directly, rather than through the data reduction that is the necessary first stage for any version of traveltime tomography.

Much of this paper concerns mathematical formulations of various versions of waveform inversion. I shall minimize mathematical formalism in this discussion, subject to introducing sufficient concepts and notation to support the main ideas. Extensive references provide the reader with access to many mathematical details not developed with full precision in the following pages.

The first two sections present an overview of OLS inversion and MVA. These subjects have extensive literatures, from which I have tried to cite a representative sample. The fourth section describes the extended modeling concept and its partially linearized variant, and establishes the link between MVA and OLS inversion. The fifth section describes the extended model approach to fully nonlinear OLS inversion and describes some initial numerical experiments which illustrate some of its properties. The paper ends with a discussion of prospects for further development of this nonlinear migration velocity analysis - waveform inversion hybrid.

WAVEFORM INVERSION

An abstract setting for geophysical inversion relates the *model space* M - a set of possible models of Earth structure - to the *data space* D through the *forward map* or prediction operator $F : M \rightarrow D$. The simplest version of data fitting inversion asks that $m \in M$ be chosen to minimize the mean square data misfit between the forward map output $F[m]$ and an observed datum $d \in D$: that is,

$$\min_m J_{\mathrm{OLS}}[m,d] \equiv \frac{1}{2}\|F[m]-d\|^2 \tag{1}$$

in which the symbol $\|\cdot\|$ stands for a (Hilbert) norm in the data space D. Generally M is actually an *admissible subset* of a vector space of functions, incorporating bounds on values (velocity is positive, etc. etc.) and perhaps other physically or mathematically motivated constraints. I will use the notation M to denote both this set and its ambient vector space, relying on context to distinguish the two, and refer to both as "model space".

In practice, geophysical inverse problems tend to be both overdetermined and underdetermined: that is, the minimum value of J_{OLS} is unlikely to be zero (data is inconsistent), and the models m which come within a reasonable tolerance of the best fit level achievable are likely to comprise a very large set (data is inadequate).

Data fitting inversion has a long and productive history, and geoscientists have developed considerable sophistication in understanding the implications of data incompleteness and inconsistency. The excellent papers of Backus and Gilbert (Backus and Gilbert, 1968; Backus and Gilbert, 1970), Jackson (Jackson, 1972; Jackson, 1976; Jackson, 1979) and Robinson and Treitel (Robinson and Treitel, 1980) continue to reward the reader. Parker (Parker, 1977), Tarantola (Tarantola, 1987), and Lines and Treitel (Lines and Treitel, 1984) provide excellent overviews of theory and application.

The application of data fitting inversion to reflection seismology dates at least to the pioneering work of Bamberger, Chavent, and Lailly in the late 70's (Bamberger et al., 1977; Bamberger et al., 1979) on the one-dimensional model problem. In contrast to much subsequent work, this study illustrated the physical and mathematical consequences of model space metric definition.

The simplest model which exhibits the basic kinematic and dynamic complexity of field reflection seismograms is perhaps *constant-density acoustics*, based on the wave equation

$$\left(\frac{1}{v^2(\mathbf{x})}\frac{\partial^2}{\partial t^2} - \nabla^2\right) u(t,\mathbf{x};\mathbf{x}_s) = f(t,\mathbf{x};\mathbf{x}_s);\ u(t,\cdot;\cdot) \equiv 0, t << 0 \tag{2}$$

in which $\mathbf{x}$ denotes position within a model of the Earth, $v(\mathbf{x})$ is the acoustic velocity field, $u(t,\mathbf{x};\mathbf{x}_s)$ is the acoustic potential, and $f(t,\mathbf{x};\mathbf{x}_s)$ is a body force representation of the energy source, indexed by *source location* $\mathbf{x}_s$.

The model space M is a set of possible velocity fields v, i.e. $m = \{v(\mathbf{x})\}$. The data space D consists of samples of pressure at a collection of *receiver locations* $\mathbf{x}_r$ possibly depending on the source location $\mathbf{x}_s$, over a time interval $t_{\text{init}} \le t \le t_{\text{final}}$. The energy source (RHS in the wave equation (2) may be regarded as known, or estimated along with the velocity. D is regarded as a Hilbert space, equipped with some version of the L^2 norm. Thus the forward map $F : M \to D$ is defined by

$$F[v](t,\mathbf{x}_r;\mathbf{x}_s) = \frac{\partial u}{\partial t}(t,\mathbf{x}_r;\mathbf{x}_s). \tag{3}$$

Throughout this paper, I shall alternate between notation appropriate to the acoustic model ($v \in M, F[v], ...$), when discussing concrete features of acoustic scattering and data processing concepts based on it, and the general inverse theory framework introduced in this section ($m \in M, F[m], ...$), when more generality is appropriate.

In order to resolve features of geologic interest, the dimension of the model space M must be rather large, on the order of $10^4 - 10^6$ in 2D, one or more orders of magnitude greater for 3D. Accordingly only iterative optimization methods with convergence rates more or less independent of model space dimension are feasible for these problems. In practice, this means Newton's method and various relatives. Chavent had already shown in the 1970's how to compute the gradient of functionals of solutions of time dependent problems, using the so-called adjoint state method borrowed from control theory (Chavent and Lemmonier, 1974). Applied to the wave equation (2), this principle becomes a version of *prestack reverse time migration* (Lailly, 1983; Lailly, 1984; Tarantola, 1984).

Application of Newton-like optimization to this acoustic inverse problem, and to more complex problems of which it is a special case, has revealed a surprising obstacle: under prototypical conditions of acquisition geometry and bandwidth, the output least squares objective (1) appears to possess many stationary points ("local minima"), most of them quite far from the global minimum even for noise-free data (in which case the global minimum occurs at the model generating the data). Tarantola and coworkers provided an early example of this pathology (Gauthier et al., 1986), and many others have since been constructed (Kolb et al., 1986; Santosa and Symes, 1989; Shin et al., 2001). While many attempts have been made to work around this obstacle (Tarantola et al., 1990; Mosegard and Tarantola, 1991; Sen and Stoffa, 1991b; Sen and Stoffa, 1991a; Clément and Chavent, 1993; Jin and Madariga, 1994; Bunks et al., 1995; Plessix et al., 1995; Plessix et al., 1999; Shin et al., 2001), all results to date are consistent with the conclusion that

Under prototypical conditions of acquisition geometry and bandwidth, the OLS objective function is strongly multimodal, and unsuitable for global minimization via Newton-like optimization methods.

The factors which appear to drive the pathological behaviour of output least squares for the acoustic inverse problem are (1) reflection geometry, i.e. sources and receivers separated from the scattering region by a hyperplane, and (2) bandlimitation of the recorded signals, in particular the absence of very low frequency energy. Transmitted wave data tends to reduce somewhat the proclivity of local minima to appear, as Tarantola et al. already observed (Gauthier et al., 1986). Mora (Mora, 1988) gave a partial explanation of this tendency, and Pratt and coworkers (Pratt, 1999; Pratt and Shipp, 1999; Sirgue and Pratt, 2004; Brenders and Pratt, 2006b; Brenders and Pratt, 2006a) have exploited this observation to devise functional OLS inversion for diving wave and crosshole geometries.

The oscillatory nature of field data, i.e. its lack of low frequencies, appears to pose a very serious obstacle to successful OLS inversion in the reflection configuration. It has been known for decades that 1D models could be reconstructed from their impulse responses, including DC components, (Goupillaud, 1961; Symes, 1980; Santosa and Schwetlick, 1982; Bube and Burridge, 1983; Symes, 1983; Symes, 1986a; Sacks and Santosa, 1987; Carazzone and Srnka, 1989), and that exclusion of DC components renders the reconstruction ambiguous (Pao et al., 1984; Gray and Symes, 1985; Santosa and Symes, 1989). For several dimensional problems, numerical explorations indicate that impulse responses (i.e. sampling of Green's functions) appear to uniquely determine acoustic models (velocity fields) via OLS inversion (Bunks et al., 1995; Shin et al., 2001), although no rigorous mathematical argument to that effect has yet been put forth.

Numerical evidence of spurious stationary points in the absence of low frequency energy ($< 3 - 5$ Hz, other scales being typical of exploration seismology) is quite strong, both direct (plots of objective vs. velocity, eg. (Symes and Carazzone, 1992)) and indirect (failure of Newton-llike optimization algorithms to achieve convergence to global minimum, (Gauthier et al., 1986; Tarantola et al., 1990; Shin et al., 2001)). While no proof exists, in the mathematical sense, of the existence of spurious minima, Chauris has given a satisfying explanation in his PhD thesis (Chauris, 2000). The mechanism identified by Chauris is the tendency of unrelated events in predicted and observed seismograms to be *tangent* near some source-receiver pairs: the tangency causes anomalously good data fit, which is destroyed by velocity perturbation.

Spurious stationary points do not exist for models whose kinematics are sufficiently close to those of the data. Thus a sufficiently good initial guess, providing the predicted data with events whose geometry closely matches that of the observed data, should permit convergence of Newton-like iteration to a satisfactory global minimum. Such convergence has been observed many times (Cao et al., 1990; Crase et al., 1990; Bunks et al., 1995; Plessix et al., 1999; Shin et al., 2001; Brenders and Pratt, 2006b). In some cases the initial estimate has been provided by traveltime tomography, in others by migration velocity analysis (described in the next section). The multiscale inversion of Bunks et al. (Bunks et al., 1995) can even be viewed as providing a reasonable initial guess for seismic band inversion by first inverting very low frequency data (significant energy as low as 0.25 Hz, in the example of Bunks et al. (Bunks et al., 1995)). Unfortunately no *a priori* assessment of initial estimate adequacy seems possible: it is quite common to fail to converge to a useful final estimate from an apparently reasonable initial estimate.

Granted that OLS inversion by Newton-like optimization converges to a best-fitting model, provided that a sufficiently good initial guess may be procured, it is natural to ask whether in fact (field) seismic data can be fit by such a model-driven process. Few studies of waveform inversion have modeled enough of the important physics of seismic wave generation and propagation to answer this question. Minkoff's 1995 thesis (Minkoff and Symes, 1997) is one exception to this pattern. Minkoff used a migration velocity analysis derived initial model estimate together with model-based attenuation parameter estimate and linearized viscoelastic modeling to estimate directionally dependent source wavelet and perturbational elastic parameters by OLS inversion. By taking all apparently relevant physics into account, she was able to fit 90% of the energy in the small plane wave data set used in the study. Furthermore, only when all relevant physics was incorporated in the model did the inversion yield lithologically correct predictions, validated by logging (Minkoff and Symes, 1997).

While more studies are needed incorporating this degree of physical realism and falsification (validation by independent criteria) design, Minkoff's thesis results give some limited confidence that waveform data fitting can indeed yield valid and highly detailed inference about the Earth's interior. The principal obstacle to use of this technique is provision of a reliable starting model, in the reflection configuration and with typical seismic bandwidth. One of the industry techniques for constructing seismic propagation models, migration velocity analysis, appears to be able to produce large velocity updates directly from data with these characteristics.

LINEARIZATION

The industry standard approach to determination of Earth structure from seismic reflection data relies upon the *linearized* model of acoustic scattering (single scattering, Born approximation), which results from applying first order perturbation theory to equation (2):

$$\left(\frac{1}{v^2(\mathbf{x})}\frac{\partial^2}{\partial t^2} - \nabla^2\right)\delta u(t,\mathbf{x};\mathbf{x}_s) = \frac{2\delta v(\mathbf{x})}{v^3(\mathbf{x})}\frac{\partial^2 u}{\partial t^2}(t,\mathbf{x};\mathbf{x}_s), \quad \delta u(t,\cdot;\cdot) \equiv 0, t << 0. \tag{4}$$

The linearized forward map or Born scattering operator DF maps the *tangent space* $M \times M$ of (reference model, model perturbation) pairs to the data space D. For acoustic modeling, DF is defined by

$$(DF[v]\delta v)(t,\mathbf{x}_r;\mathbf{x}_s) = \frac{\partial \delta u}{\partial t}(t,\mathbf{x}_r;\mathbf{x}_s). \tag{5}$$

The reference velocity model v is supposed to account for the kinematics of data, hence the large scale stucture of the Earth. Thus it is appropriate to assume v to be *smooth* or slowly varying, or at least piecewise smooth, on the scale of a wavelength. The perturbation δv is presumed to carry the wavelength scale features of Earth structure, and so is oscillatory. The notation is chosen to remind the reader that DF depends nonlinearly on the reference model (v in this case) and linearly on the model perturbation (δv).

Many of the statements to be made in the sequel about the behavior of the linearized operator $DF[v]$ rely on additional constraints being imposed on the velocity model v, such as explicit upper and lower bounds and bounds on spatial gradients. Together with the smoothness constraints, these additional requirements define appropriate admissible sets of velocity models. I will not specify admissible sets explicitly in this paper, but will tacitly assume that restriction of DF to an admissible set of velocities is in force.

Apparently the first order Taylor polynomial remainder $F[v+\delta v]-F[v]-DF[v]\delta v$ is anomalously small when v is sufficiently smooth (bounds on derivatives) and δv oscillatory (sufficiently small low-order moments). Considerable numerical evidence supports this point of view (Symes, 1995), but theoretical justification exists only for the special case of 1D modeling (Lewis and Symes, 1991; Symes, 1991b). Existence of the derivative DF under very general circumstances, as a bounded operator on suitable Hilbert spaces, is established by Stolk (Stolk, 2000b), see also Fernandez-Berdaguer et al. (Fernandez-Berdaguer et al., 1996).

The linearized or Born scattering operator $DF[v]$, for smooth background velocity v, has been studied extensively over the last twenty years, beginning with the pioneering work of Beylkin (Beylkin, 1985), Bleistein (Bleistein, 1987), and Rakesh (Rakesh, 1988). As the wave operators appearing in (2) and (4) then have smooth coefficients, oscillatory solutions are well approximated by geometric optics, and $DF[v]$ by an oscillatory integral operator (generalized Radon transform, (Beylkin, 1985)). These approximations underly an analysis of the linearized OLS problem

$$\text{minimize}\,\frac{1}{2}\|DF[v]\delta v-(d-F[v])\|^2\,\text{over}\,\delta v\in M. \tag{6}$$

A minimizer of the function (6) of minimal length is a so-called *pseudoinverse* solution, written $\delta v = DF[v]^{\dagger}d$ and well-approximated by the regularized least-squares solution

$$\delta v \simeq (DF[v]^*DF[v]+\lambda I)^{-1}DF[v]^*(d-F[v])$$

for small values of λ.

The essential properties of $DF[v]$ and its least squares problem follow from the modern theory of oscillatory integrals, and are easiest to state when it is assumed acquisition geometry is *complete*, i.e. that essentially all source-receiver azimuths are available, that sampling is dense enough to be neglected, and that the source is impulsive: $f(\mathbf{x},t;\mathbf{x}_s)=\delta(t)\delta(\mathbf{x}-\mathbf{x}_s)$. Then

> *$DF[v]$ is "almost unitary", in the sense that its adjoint or transpose $DF[v]^*$ differs from its (pseudo)inverse by dip-dependent scaling and filtering;*
>
> *If the data d is nearly consistent with the Born scattering model, $d\simeq F[v]+DF[v]\delta v$ for smooth v and oscillatory δv, then the least squares gradient*

> $DF[v]^*(d - F[v])$ *is an* image *of* δv*, in the sense of having the same oscillatory components or locations of rapid change ("reflectors"), except for dip-dependent scaling and filtering.*

The almost-unitary property of $DF[v]$ on oscillatory data was established by Beykin and Rakesh in the cited references, under the assumption that all source/receiver are connected to all possible scattering points (support of δv) by unique rays of geometric optics. This latter assumption was gradually relaxed through the work of Smit et al. (Ten Kroode et al., 1998), Nolan (Nolan and Symes, 1997), and Stolk (Stolk, 2000b; Stolk, 2000a). In particular, Stolk shows that the almost-unitary property of $DF[v]$ is a *generic*: if v doesn't have it, then arbitrarily small perturbations of v do (for 2D acoustics - the analogous property is conjectured to hold in 3D, but has not yet been proven). The normal operator $DF[v]^*DF[v]$ acts by dip-dependent scaling and filtering: it is a so-called *pseudodifferential* operator, for generic smooth background velocities v. In particular, the normal operator does not affect the location or orientation of highly oscillatory or short-scale features of its argument. This fact is responsible for the almost-unitary nature of the linearized map, and implies that its adjoint $DF[v]^*$ is a *structural imaging* operator, in fact a version of *prestack depth migration* ("PSDM"). This relation of linearized inversion to PSDM was noted by Lailly (Lailly, 1983; Lailly, 1984) and Tarantola (Tarantola, 1984).

Nolan (Nolan and Symes, 1997) also shows how the above statements must be modified when the acquisition geometry is incomplete, i.e. idealized narrow azimuth surveys. A number of researchers have generalized these results to various elastic settings of the seismic inverse problem (Beylkin and Burridge, 1990; deHoop and Bleistein, 1997; Burridge et al., 1998; de Hoop and Stolk, 2002). The details are naturally more complex, but the gist is as called out above: the linearized scattering operators are almost unitary in general, and their normal operators act by dip-dependent scaling and filtering, provided full bandwidth and spatially complete data is available. approximate solutions of linearized OLS problem can be written as oscillatory integrals, convenient for computation.

MIGRATION VELOCITY ANALYSIS

Solution of the linearized inverse problem, or approximate solution (prestack depth migration, PSDM), requires a velocity (reference) model v, which must also be estimated somehow from the data. The geophysical prospecting industry has developed a set of techniques for velocity estimation, called *migration velocity analysis* ("MVA") (Yilmaz, 2001). In contrast to PSDM itself, MVA appears at first glance to have little to do with OLS inversion.

MVA consists of two steps: given a velocity model,

- produce an *image volume*, containing the image (approximate linearized inverse) but also depending on one or more redundant parameters;
- apply an *imaging condition* to produce an image and, as a by-product, an updated velocity.

In principle, these steps may be carried out a number of times.

Image Volume Construction

Two distinct definitions of image volume are in common use. Both are most easily formulated with the aid of the solution operator of the linearized acoustic problem (4). Using the Green's function G of the reference medium, i.e. the causal solution of equation (2) with $f(t,\mathbf{x};\mathbf{x}_s) = \delta(\mathbf{x}-\mathbf{x}_s)\delta(t)$, one can write an integral representation for the solution of (4), evaluated at a receiver position $\mathbf{x}_r$:

$$\delta u(t,\mathbf{x}_r;\mathbf{x}_s) = \frac{\partial^2}{\partial t^2}\int d\mathbf{x}\int d\tau G(t-\tau\mathbf{x};\mathbf{x}_r)G(\tau,\mathbf{x};\mathbf{x}_s)\frac{2\delta v(\mathbf{x})}{v^3(\mathbf{x})} \tag{7}$$

In view of the definition (5), the integral representation (7) shows $DF[v]$ to be an integral operator with kernel

$$\frac{\partial^3}{\partial t^3}\int d\tau G(t-\tau,\mathbf{x};\mathbf{x}_r)G(\tau,\mathbf{x};\mathbf{x}_s)\frac{2}{v^3(\mathbf{x})}.$$

Thus the adjoint linearized operator, a version of PSDM, has the integral representation

$$DF[v]^*d(\mathbf{x}) = -\frac{2}{v^3(\mathbf{x})}\int d\mathbf{x}_s\int d\mathbf{x}_r\int dt\int d\tau G(t-\tau,\mathbf{x};\mathbf{x}_r)G(\tau,\mathbf{x};\mathbf{x}_s)\frac{\partial^3 d}{\partial t^3}(t,\mathbf{x}_r;\mathbf{x}_s). \tag{8}$$

As noted above the output of $DF[v]^*$ is an image of the subsurface, having approximately correct short-scale features as δv when $d \simeq DF[v]\delta v$. I will denote the output by $I(\mathbf{x}) \equiv DF[v]^*d$.

The *surface oriented* image volume definition introduces a surface acquisition parameter as the redundant degree of freedom, and limits the integration in (8) to common value (level) sets of this parameter (i.e. to gathers). For example, the *common offset* image volume I_{CO} is constructed by restricting (8) to common offset gathers: for (vector) half-offset $\mathbf{h}$, define

$$I_{\mathrm{CO}}(\mathbf{x},\mathbf{h}) = -\frac{2}{v^3(\mathbf{x})}\int d\mathbf{x}_s\int dt\int d\tau G(t-\tau,\mathbf{x};\mathbf{x}_s+2\mathbf{h})G(\tau,\mathbf{x};\mathbf{x}_s)\frac{\partial^3 d}{\partial t^3}(t,\mathbf{x}_s+2\mathbf{h};\mathbf{x}_s). \tag{9}$$

In practice, this integral is further averaged over bins in $\mathbf{h}$, or over an annuli in $\mathbf{h}$ to create a function of (scalar) offset $h = |\mathbf{h}|$.

Synthesis of plane wave response leads to a similar definition of plane (or conical) wave image volume (Treitel et al., 1982; Duquet and Lailly, 2006). An even less straightforward variant of this concept is the common *scattering angle* volume (Xu et al., 2001; Stolk, 2000b; Stolk, 2000a; Brandsberg-Dahl et al., 2003), in which the integral (8) is recast as an integral in *phase space* and subintegrals formed over data subsets (in phase space) characterized by common scattering angle.

All of the surface oriented image volume constructions share a common defect, in that *kinematic artifacts* may form when the velocity structure is sufficiently complex that significant energy can propagate along multiple paths connecting source, receiver, and scattering points (Nolan and Symes, 1997; Brandsberg-Dahl et al., 2003; Stolk and Symes, 2004). These artifacts arise because the restricted integrals in volume formation rules such as (9) do not restrict the slowness component of rays contributing to the image output, corresponding to the gather direction. For example, integration over a common offset gather implicitly combines image contributions propagating along all rays sharing a common midpoint ray parameter, but with possibly different offset ray parameters (difference of source, receiver horizontal ray parameters). Therefore energy arriving in the data along one pair of rays can migrate to a different point in the image volume along a different pair of rays, with different offset ray parameter. Similar image ambiguity may arise in any of the surface-oriented image volumes defined above, including common scattering angle volumes. The artifact events produced in this way can easily be as energetic as the events corresponding to actual reflectors. Their presence has serious consequences for the use of this type of image volume in velocity analysis, as explained below.

The second type of image volume definition introduces additional degrees of freedom by means of *spatial and/or temporal shifts* applied to the integrands in (8). The version of this *depth oriented* image volume appropriate to *subsurface* vector half-offset $\mathbf{h}$ is

$$I_{\mathrm{SS}}(\mathbf{x},\mathbf{h}) = \frac{2}{v(\mathbf{x}+\mathbf{h})v^2(\mathbf{x}-\mathbf{h})}\int d\mathbf{x}_s \int d\mathbf{x}_r \int dt \int d\tau G(t-\tau,\mathbf{x}+\mathbf{h};\mathbf{x}_r)G(\tau,\mathbf{x}-\mathbf{h};\mathbf{x}_s)\frac{\partial^3 d}{\partial t^3}(t,\mathbf{x}_r;\mathbf{x}_s). \tag{10}$$

Claerbout (Claerbout, 1971; Claerbout and Doherty, 1972) introduced essentially this definition, with the exceptions that (i) the factors of v and the t derivatives were missing - the reasons for their presence in our formulation will become apparent below; (ii) $\mathbf{h}$ is constrained to be horizontal, and that remains the most common variant. Indeed, Claerbout proposed viewing this image volume as the result of *sinking the survey*, with "sunken" sources and receivers ($\mathbf{x}-\mathbf{h}$ and $\mathbf{x}+\mathbf{h}$ respectively) on planes of increasing depth (Claerbout, 1985). Claerbout also showed how to build this image volume by means of one-way wavefield extrapolation from surface data (as opposed to evaluation of an integral like the above). It is also possible to use two way wavefield extrapolation, i.e. solution of the acoustic wave equation (2) and related computations, to accomplish this task, via a variant of *reverse time migration* (Biondi and Shan, 2002; Symes, 2002). Nonhorizontal subsurface offset may find constructive uses (Biondi and Symes, 2004; Biondi and Shan, 2002; Symes, 2002), as can time (as opposed to spatial offset) shifts (Sava and Fomel, 2005). Scattering angle can also be used as the redundant parameter (Prucha et al., 1999; Rickett and Sava, 2002; Sava and Fomel, 2003).

The computational cost of depth oriented image volume formation can be considerable.The integral on the RHS of equation (10) can be interpreted as the crosscorrelation at a range of spatial shifts of the field produce by the source ($G(\tau,\mathbf{x}-\mathbf{h};\mathbf{x}_s)$, under the integral sign), and the field produced by propagating the recorded data d backwards in time (integral over $\mathbf{x}_r, t$ of the remaining factors), followed by a final summation over sources. This *shot-profile* profile organization of the computation is the only one available for the reverse time approach, and is one of the two common techniques employing one-way propagators (Biondi and Palacharla, 1996; Biondi, 2003). The additional computational load due to the cross-correlations can easily dwarf that due to propagation. A commonly used workaround is to avoid computing the full image volume. For velocity analysis, as reviewed below, a relatively small fraction of the image volume suffices to constrain the velocity estimate. Thus much of the crosscorrelation load can be avoided, and the cost of the partial image volume formation approaches that of image formation alone.

Surface oriented image volume formation is often termed "Kirchhoff prestack depth migration". It is usually carried out via the formula (9) with the Green's function G replaced by its asymptotic (ray-theoretic) approximation. The resulting diffraction sum integral is then reminiscent of the Kirchhoff integral in electromagnetic scattering (Sommerfeld, 1964). The differential equation approach introduced by Claerbout (Claerbout, 1971) is the predominant method for depth oriented image volume production, hence the common usage "prestack wave equation migration" for this technique. Both imaging methodologies have huge literatures, overviewed for example in (Gardner, 1985; Yilmaz, 2001).

Imaging Conditions

The phrase "imaging condition" has come to signify two related but distinct concepts:

- a relation between the image *volume* and the image;
- a quality possessed by the image volume when velocity is chosen correctly, and which is characteristic of correct velocity.

The first of these concepts provides a mechanism for producing an image of the subsurface, given a suitable velocity model; the second can lead to analysis and correction of velocity defects.

For the two versions of image volume introduced above, the relation with the image, i.e. with the output of the adjoint operator $DF[v]^*$, is clear from comparison of (9) and (10) with (8). For the surface oriented offset image volume, one obtains the image by integrating over $\mathbf{h}$ ("stacking"):

$$I(\mathbf{x}) = \int d\mathbf{h} I_{\mathrm{CO}}(\mathbf{x}, \mathbf{h}).$$

For the depth oriented offset image volume, on the other hand, one simply sets $\mathbf{h} = 0$ ("extract the zero offset section"), which causes (10) to become identical to (8):

$$I(\mathbf{x}) = I_{\mathrm{SS}}(\mathbf{x}, \mathbf{0}).$$

The second point, on the nature of the image volume for a consistent velocity, is more subtle. It can be expressed as the requirement that the image volume be an eigenvector of the mapping from image volume to image, with the largest possible eigenvalue.

For surface oriented offset, this means that $I(\mathbf{x}, \mathbf{h})$ should be *constant* in offset. In practice, since amplitudes are not controlled in this process (and in practice because the Earth does not conform to the constant density acoustic model), constant phase is an acceptable substitute. The conventional tool with which to judge this quality is the *image gather*, a 2D display in which the horizontal midpoint coordinates are fixed and the image amplitudes displayed as function of depth and offset (or shot, or scattering angle,...). If depth is the vertical coordinate in such a plot, its appearance should resemble horizontal stripes: the gather should be *flattened*. This criterion is a generalization of the conventional quality assurance metric of standard NMO-based velocity analysis (Yilmaz, 2001). An equivalent quality is strength of the stack, ie. final image: it is strongest when all interference in the sum over $\mathbf{h}$ is constructive (Taner and Koehler, 1969).

Depth oriented offset volumes should on the other hand be *focused* at zero offset when the velocity is compatible with the data: that is, almost all energy in the volume should be concentrated near $\mathbf{h} = 0$. The scattering angle variant should produce flat depth vs. angle gathers.

Velocity Analysis

Since the image volume constructed by prestack migration should satisfy one of the conditions mentioned in the preceding paragraphs provided that the velocity is consistent with the data, the failure of these conditions to be satisfied provides both evidence that the velocity is not consistent with the data and means to update it to improve its data-consistency. For the surface oriented offset volume, the well-known "smiles and frowns" rule from standard processing velocity analysis carries over almost unaltered, provided that the lateral heterogeneity of the structure to be imaged is not too great:

> If an event in an image gather are convex upwards (downwards), then the average velocity above the depth of the event is too high (low).

Similar rules may be devised for updating velocity from depth oriented offset image volumes (Claerbout, 1985).

This sort of rule is qualitative, and is often applied interactively to build a velocity model layer by layer. Considerable effort has been expended on software tools to ease this task. Some attempts to quantify the velocity update process will be described below.

At this point, the significance of kinematic artifacts in surface oriented image volumes is clear. These artifacts do not conform to the nominal conditions for image volumes produced from correct velocities, even for synthetic data. That is, when the velocity structure is sufficiently complex that multipathing is important, image gathers are not in general flat even for correct velocity models (Nolan and Symes, 1997; Brandsberg-Dahl et al., 2003; Stolk and Symes, 2004). Even worse, nothing in the appearance of the artifact events distinguishes them in any obvious way from the actual reflector images (Nolan and Symes, 1997). Therefore surface oriented image volumes are not a suitable tool for velocity updating in complex Earth structure.

On the other hand, the landmark work of de Hoop and Stolk (de Hoop and Stolk, 2005; de Hoop and Stolk, 2006) established that no such artifacts are present in depth oriented image volumes, even in the presence of considerable multipathing, under the sole restriction that rays carrying significant energy do not turn horizontal. This result goes a long way towards explaining the widely reported superiority of "wave equation" migration for prestack imaging in regions of complex velocity structure. In particular, the use of the wave equation, in its depth extrapolation form, has nothing to do *per se* with the artifact-free nature of the image volumes constructed by this style of migration - the same results could be obtained with a suitable computational asymptotic representation of the integral formula (10). The restriction to vertically propagating energy may be important however, see (Symes, 2002).

EXTENDED MODELING: A UNIFYING CONCEPT

The extension concept links the image volume constructions of the last section with modeling and inversion: the imaging operators exemplified by the formulae (9) and (10) are in fact adjoints of linearized (Born) extended modeling.

An *extension* of model $F : M \rightarrow D$ consists of

- an *extended* model space $\bar{M}$;
- an extension operator $E : M \rightarrow \bar{M}$;
- an extended modeling operator $\bar{F} : \bar{M} \rightarrow D$ satisfying $F[m] = \bar{F}[E[m]]$ for any $m \in M$.

The extension operator E should be injective, i.e. one-to-one, hence enable one to view the model space M as a subset $E[M] \subset \bar{M}$ of the extended model space. Since the extended models will be in some sense unphysical, I will refer to $E[M]$ as the "physical models".

For an extended model as above, the extended inverse problem is: given $d \in D$, find $\bar{m} \in \bar{M}$ so that $\bar{F}[\bar{m}] \simeq d$. Since extended model space has more degrees of freedom, ambiguity is more likely. A solution $\bar{m}$ is physically meaningful only if it is physical, i.e. $\bar{m} = E[m]$. In that case, m is a solution of the original inverse problem, i.e. $F[m] \simeq d$, because of the extension relation between F and $\bar{F}$. That is,

> *To solve the inverse problem (approximately), find a solution of an extended inverse problem that lies in the range of the extension map.*

Application of this concept to MVA is based on linearized (tangent) extended modeling. Given an extension as above, its *tangent extension* consists of

- the tangent extended model space $T\bar{M}$ of pairs $(\bar{m}_0, \delta\bar{m}) \in \bar{M} \times \bar{M}$;

- E is typically linear, so its linearization $DE : TM \mapsto \bar{M}$ is defined by $DE[m_0]\delta m = E[\delta m]$.
- linearized (Born) extended modeling: $D\bar{F} : T\bar{M} \rightarrow D$.

From the chain rule, it follows that $D\bar{F}$ is an extension of DF, i.e. $D\bar{F}[E[m_0]]E[\delta m] = DF[m]\delta m$.

It turns out to be reasonable to restrict tangent extensions to perturbations of physical models, i.e. to assume always that $\bar{m}_0 = E[m_0]$. I shall assume this restriction implicitly in the following discussion, and confound reference models m_0 with their images $E[m_0]$ under the extension map.

An extension is *approximately (linearly) invertible* on the data space D if its pseudoinverse is an approximate right inverse:

$$D\bar{F}[m_0]D\bar{F}[m_0]^\dagger d \simeq d$$

for all $d \in D$. This property is equivalent to approximate surjectivity of the tangent extended operator: that is, for any $d \in D$ and admissible v, one can find an extended model perturbation $\delta\bar{v}$ so that $D\bar{F}[v]\delta\bar{v} \simeq d$.

The definition of the data space D may need to be adjusted to satisfy this surjectivity requirement. In the acoustic example, event dips must be limited to those which correspond to non-evanescent waves at the surface. Dip-limiting can be built into the definition of the data space, for example by redefining the norm in D to heavily weight evanescent wave data. Such weighting is straightforward if the admissible class of models includes constraints on wave velocities near the surface

The basic modeling operators considered in this paper are generally not (approximately) surjective. In particular, the linearized modeling operator of $DF[v]$ of the canonical acoustic problem is not surjective in general, unless the data is consistent kinematically with the velocity model v. This occurs generically only for very special acquisition geometries, for example single shot or offset gathers. In general a poorly chosen velocity prevents accurate fitting of multioffset data. Put another way, the *trivial extension*, defined by $\bar{M} = M, E = I, \bar{F} = F$, is almost never invertible for acoustic modeling and its generalizations.

Surface Oriented Offset Extension

Take for the extended model space $\bar{M}_{\rm SO}$ a set of positive functions $\bar{v}(\mathbf{x}, \mathbf{h})$ of position $\mathbf{x}$ *and* offset $\mathbf{h}$. The extension map $E_{\rm SO}$ simply views a physical velocity (positive function of $\mathbf{x}$) as a function of $\mathbf{x}$ and $\mathbf{h}$, i.e. as constant in $\mathbf{h}$: $E_{\rm SO}[v](\mathbf{x}, \mathbf{h}) = v(\mathbf{x})$.

The extended modeling operator $\bar{F}_{\rm SO}$ computes the traces for offset $\mathbf{h}$ by using the extended velocity model for offset $\mathbf{h}$, i.e. by solving (2) with $v = \bar{v}(\cdot, \mathbf{h})$.

Tangent extended modeling is similar: $D\bar{F}_{\rm SO}$ computes the perturbations of all traces with offset $\mathbf{h}$ by solving (4) with $\delta v = \delta\bar{v}(\cdot, \mathbf{h})$. Consistent with the remark above, only physically consistent reference velocities $\bar{v}(\mathbf{x},\mathbf{h}) = v(\mathbf{x})$ will be considered. Accordingly, the tangent extended modeling operator $D\bar{F}_{\rm SO}$ has an integral representation similar to that of DF:

$$D\bar{F}_{\rm SO}[v]\delta\bar{v}(t,\mathbf{x}_s+2\mathbf{h};\mathbf{x}_s) = \frac{\partial^3}{\partial t^3}\int d\mathbf{x}\int d\tau G(t-\tau,\mathbf{x};\mathbf{x}_s+2\mathbf{h})G(\tau,\mathbf{x};\mathbf{x}_s)\frac{2\delta\bar{v}(\mathbf{x},\mathbf{h})}{v^3(\mathbf{x})} \quad (11)$$

which is almost the same as equation (7), in particular G is the same Green's function. The difference is that now the velocity perturbation is allowed to depend on $\mathbf{h}$.

The tangent extended model $D\bar{F}_{\rm SO}[v]$ appears to be surjective in the sense defined above, i.e. any non-evanescent data can be well approximated by its range for reasonable definitions of the admissible model set. Some theoretical evidence exists for this presumption (Sacks, 1988).

The adjoint $D\bar{F}_{\rm SO}[v]^*$ may be read off from the integral expression (11): it is exactly the mapping defined by the integral (9)! That is,

$$I_{\rm CO} = D\bar{F}_{\rm SO}[v]^*d$$

As proven by Beylkin (Beylkin, 1985) and Rakesh (Rakesh, 1988), in the absence of multipathing, the linearized extended map $D\bar{F}_{\rm SO}[v]$ is almost unitary: its adjoint differs from its inverse only by dip-dependent scaling and filtering. Thus the common offset image volume is actually seen to be an approximate tangent extended inversion.

The adjoint of $E_{\rm SO}$ is integration over $\mathbf{h}$, i.e. stacking, so one recovers the relation between image and image volume from the chain rule and the composition relation $F = \bar{F}_{\rm SO}[E_{\rm SO}]$. That is, the imaging condition is simply application of $E^*_{\rm SO}$:

$$I = E^*_{\rm SO}I_{\rm CO}$$

The other aspect of the imaging condition - the quality of the image volume which signifies data-consistent velocity - is also inherent in the extension structure: it simply membership in the range of $E_{\rm SO}$. This range consists precisely of extended (linearized) models with $\delta\bar{v}$ independent of $\mathbf{h}$. As noted above, for reasons of modeling inaccuracy, use of the adjoint of $D\bar{F}_{\rm SO}[v]^*$ rather than the (pseudo)inverse ("true amplitude migration") operator $D\bar{F}_{\rm SO}[v]^\dagger$, aperture deficits, etc. it is usually necessary to relax this requirement, to membership in the larger subspace obtained from the range of $E_{\rm SO}$ by slowly-varying amplitude scaling and muting.

Note that when multipathing with significant energy occurs, $D\bar{F}_{\rm SO}[v]$ is no longer almost unitary: reflector energy may move to the locations of kinematic artifacts, so the adjoint operator is not an approximate inversion up to dip-dependent scaling and filtering (Nolan and Symes, 1997; Brandsberg-Dahl et al., 2003; Stolk and Symes, 2004). Otherwise put, the normal operator is not pseudodifferential, the economical construction of approximate inverses pioneered by Beylkin (Beylkin, 1985) is no longer possible, and PSDM does not produce image volumes remotely close to the range of E, even when the velocity is data-consistent. This defect is not limited to offset extensions, but also affects those based on other surface acquisition parameters, even scattering angle in the surface oriented definition (Stolk and Symes, 2004).

Depth Oriented Offset Extension

The nonlinear extended model relevant to wave equation migration requires a bit of a technical digression. I will sketch this reasoning here, both because it will figure in the discussion of the next section and because it establishes the clear-cut route between waveform inversion and MVA for the "wave equation" approach.

Denote by $\bar{M}_{\rm DO}$ the bounded positive selfadjoint operators on $L^2(\mathbf{R}^3)$, the Hilbert space of square-integrable functions. Given $\bar{v} \in \bar{M}_{\rm DO}$, the corresponding generalized acoustic potential field $\bar{u}(t,\mathbf{x};\mathbf{x}_s)$ with point source at $\mathbf{x}_s$ and wavelet $w(t)$ satisfies

$$\left(\bar{v}^{-2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)\bar{u}(t,\mathbf{x};\mathbf{x}_s) = w(t)\delta(\mathbf{x}-\mathbf{x}_s) \tag{12}$$

The extended modeling operator is defined by

$$\bar{F}_{\rm DO}[\bar{v}](\mathbf{x}_r,t;\mathbf{x}_s) = \bar{u}(t,\mathbf{x}_r;\mathbf{x}_s)$$

Define the extension map $E_{\rm DO} : M \rightarrow \bar{M}_{\rm DO}$ by

$$E_{\rm DO}[v]f(\mathbf{x}) = v(\mathbf{x})f(\mathbf{x}),\;\; f \in L^2(\mathbf{R}^3)$$

$E_{\rm DO}$ is continuous, eg. with the L^∞ norm in domain and the operator norm in the range.

Evidently $\bar{F}_{\rm DO}[E_{\rm DO}] = F$, so the foregoing actually does define an extension of F.

Physically, acoustic wave velocity is a combination of density and bulk modulus; the bulk modulus gives the volumetric strain response of the material to dilatational stress. Continuum mechanics mandates that this response be local, i.e. intermediated by a diagonal operator. The extension proposed here models *action at a distance*: the bulk modulus is represented by a general symmetric positive definite operator, not necessarily diagonal, so that the strain response is not confined to the location of the imposed stress. Physical models, i.e. those in the range of $E_{\rm DO}$, do not permit action at a distance, hence are modeled by diagonal operators (multiplication by functions). [I am indebted to Scott Morton (personal communication) for this observation.]

The problem defined in (12) is of course a bit unusual, as the coefficients in the wave equation have become operators, rather than scalars. This problem is mathematically well-behaved, however, with unique finite energy solutions for finite energy wavelet $w(t)$, as follows from slight generalizations of arguments given in Chapter 2 of Stolk's thesis (Stolk, 2000b).

The tangent extended modeling operator $D\bar{F}_{\rm DO}$ is given by

$$D\bar{F}_{\rm DO}[E[v_0]]\delta\bar{v}(t,\mathbf{x}_r;\mathbf{x}_s) = \delta\bar{u}(t,\mathbf{x}_r;\mathbf{x}_s)$$

where

$$\left(v^{-2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)\delta\bar{u}(t,\mathbf{x};\mathbf{x}_s) = 2v^{-1}\delta\bar{v}\left[v^{-2}\frac{\partial^2 u}{\partial t^2}\right](t,\mathbf{x};\mathbf{x}_s). \tag{13}$$

Presume as in the preceding section that the source is impulsive, so that $u = G$. Since we are perturbing only about physical models, the operator on the LHS is the ordinary wave operator. Write the action of the operator $\delta\bar{v}$ as if it were an integral operator, with kernel also denoted $\delta\bar{v}$. Then we can express the solution of (13) as

$$\delta\bar{u}(t,\mathbf{x}_r;\mathbf{x}_s) = \int d\mathbf{x} \int d\tau G(\tau,\mathbf{x};\mathbf{x}_r)\frac{2}{v(\mathbf{x})}\int d\mathbf{x}'\delta\bar{v}(\mathbf{x},\mathbf{x}')\frac{1}{v^2(\mathbf{x}')}\frac{\partial^2 G}{\partial t^2}(t-\tau,\mathbf{x}';\mathbf{x}_s).$$

Change variables of integration from $(\mathbf{x},\mathbf{x}')$ to $\mathbf{h} \leftarrow (\mathbf{x}'-\mathbf{x})/2$ and $\mathbf{x} \leftarrow (\mathbf{x}'+\mathbf{x})/2$. Then the above integral is equal to

$$= \frac{\partial^2}{\partial t^2}\int d\mathbf{x}\int d\mathbf{h}\int d\tau G(\tau,\mathbf{x}+\mathbf{h};\mathbf{x}_r)G(t-\tau,\mathbf{x}-\mathbf{h};\mathbf{x}_s)\frac{2}{v(\mathbf{x}+\mathbf{h})v^2(\mathbf{x}-\mathbf{h})}\delta\bar{v}(\mathbf{x},\mathbf{h}).$$

Considerable numerical evidence points to surjectivity of $D\bar{F}_{\rm DO}[v]$, provided that v is confined to a suitable admissible set and D is defined properly. No theoretical studies of this question have appeared, to the author's knowledge.

The adjoint tangent extended modeling operator is evidently

$$D\bar{F}_{\rm DO}[v]^* d(\mathbf{x},\mathbf{h}) = \frac{2}{v(\mathbf{x}+\mathbf{h})v^2(\mathbf{x}-\mathbf{h})}$$

$$\times \int d\mathbf{x}_s\int d\mathbf{x}_r\int dt\int d\tau G(t-\tau,\mathbf{x}+\mathbf{h};\mathbf{x}_r)G(\tau,\mathbf{x}-\mathbf{h};\mathbf{x}_s)\frac{\partial^3 d}{\partial t^3}(t,\mathbf{x}_r;\mathbf{x}_s),$$

which of course is exactly (10). That is,

> *the depth oriented image volume (output of survey sinking or "wave equation migration") is the image of the data under the adjoint tangent extended modeling operator, for the depth oriented offset extension:*

$$I_{\rm SS} = D\bar{F}_{\rm DO}[v]^* d,\ I = E^*_{\rm DO} I_{\rm SS}.$$

As noted above, in a much wider range of cases, including all in which rays carrying significant energy propagate without turning horizontal, the tangent extended operator for the depth oriented extension is near-unitary: its pseudoinverse differs from its adjoint by dip-dependent scaling and filterin. Otherwise put, the normal operator for tangent extended depth oriented extended modeling is pseudodifferential, absent turning rays. This variant of PSDM does produce image volumes near the range of the extension map E for data-consistent velocity models, even in the presence of significant multipathing.

OBJECTIVE MIGRATION VELOCITY ANALYSIS

Migration velocity analysis has a very simple description in the context of tangent extended modeling:

MVA $\equiv$ adjust the velocity model v to bring the pseudoinverse solution of the linearized inverse problem $D\bar{F}[v]^{\dagger}d$, or its PSDM approximation $D\bar{F}[v]^{*}d$, as near as possible to the range of the extension map E (the physical models).

The pseudoinverse solution is only computationally feasible when it is closely related (by dip-dependent scaling and filtering) to the PSDM operator. Therefore this procedure really only makes sense for the surface oriented extension when energy moves from data to image volume along essentially unique ray paths. When multiple ray paths carry any significant amount of energy, the depth oriented extension provides a more robust framework for velocity analysis.

Many software tools exist to carry out this task interactively. However it is also possible to view the aim of MVA as inversion, and to automate its accomplishment. In fact,

Provided that the underlying extemsion is invertible, MVA is an approximate solution method for the partially linearized inverse problem: given data d, find velocity model v **and** *short-scale perturbation δv so that $DF[v]\delta v \simeq d$ in the least squares sense.*

To see this, suppose that MVA produces a pseudoinverse solution $D\bar{F}[v]^{\dagger}d$ that lies near the range of E, i.e.

$$D\bar{F}[v]^{\dagger}d \simeq E\delta v$$

for a perturbational model δv. Invertibility of the extension implies that

$$DF[v]\delta v \simeq D\bar{F}[v]D\bar{F}[v]^{\dagger}d \simeq d$$

That is, v and δv solve the partially linearized inverse problem.

For those cases in which the pseudoinverse is closely related to the adjoint, i.e. in the absence of kinematic artifacts, the criteria for membership in the range of E may be verified by examination of $D\bar{F}[v]^{*}d$ instead, as already noted.

Annihilators

Since MVA actually (implicitly) solves a version of the waveform inversion problem, it is natural to seek objectives whose extrema represent the solution. Since the range of E is a linear subspace of $\bar{M}$, any linear operator vanishing on this subspace gives rise to a quadratic form which can serve as such an objective. A (possibly) v-dependent map $A[v]$ from extended model space $\bar{M}$ to some other Hilbert space H is an *annihilator* of the range of E if

$$r = E\delta v \Leftrightarrow A[v]r = 0$$

.

If $A[v]$ is an annihilator in the sense just defined, then the function

$$J_A[v,d] \equiv \frac{1}{2}\|A[v]D\bar{F}[v]^{\dagger}d\|^2 \tag{14}$$

attains its global minimum at velocities which produce physical short-scale models, i.e. image volumes in the range of the extension operator. Any such function could potentially serve as the objective for an optimization approach to MVA. Computational advantage can be gained if the pseudoinverse $D\bar{F}[v]^\dagger$ can be replaced by the adjoint $D\bar{F}[v]^*$ (i.e. PSDM instead of linearized inversion) without seriously perturbing the locations of minimizing models. Such replacement is possible with some choices of $A[v]$.

Of course, the linear space of operators $A[v]$ annihilating the range of E typically has high dimension. Somewhat more surprisingly, the behaviour of the corresponding functions J_A varies dramatically with the choice of $A[v]$. The following three types of annihilator have been studied extensively:

- $A[v] = (DF[v]DF[v]^\dagger - I)D\bar{F}[v]$.Under the standing assumption on invertibility of the extension,

$$J_A[v,d] = \frac{1}{2}\|(DF[v]DF[v]^\dagger - I)D\bar{F}[v]D\bar{F}[v]^\dagger d\|^2$$

$$\simeq \frac{1}{2}\|DF[v]DF[v]^\dagger d - d\|^2 = \min_{\delta v}\frac{1}{2}\|DF[v]\delta v - d\|^2.$$

  Thus minimizing J_A over v is equivalent to OLS solution of the partially linearized problem. So even OLS fits in this "annihilator" framework.

- $A = EE^\dagger - I$. Note that

$$\|(EE^\dagger - I)r\|^2 = \|EE^\dagger r\|^2 - 2\langle r, EE^\dagger r\rangle + \|r\|^2.$$

  Recalling that E may be taken to be an isometry, i.e. $E^* = E^\dagger$. Using this fact and the pseudoinverse property $EE^\dagger E = E$, the last line simplifies to

$$= \|r\|^2 - \|E^\dagger r\|^2.$$

  Thus minimizing J_A in this case is the same as maximizing $\|E^\dagger D\bar{F}[v]^\dagger d\|^2$. Provided that $D\bar{F}[v]$ is almost unitary, same optima occur using PSDM rather than pseudoinversion, i.e. minimizing J_A yields the same velocity model as does maximizing $\|E^* D\bar{F}[v]^* d\|^2 = \|DF[v]^* d\|^2$. For the surface oriented extension, the resulting functional is the simplest version of *(total)stack power*, and was studied in the context of convolutional modeling by Toldi (Toldi, 1989). For either type of extension described above, it amounts to the *image power* (Soubaras and Gratacos, 2006; Shen and Calandra, 2005).

- $A = (1 - \nabla^2_{\mathbf{x},\mathbf{h}})^{-\frac{1}{2}}\partial_{\mathbf{h}}$ for the surface oriented extension, and $\mathbf{h}$ (i.e. multiplication by offset) for the depth oriented extension, and various inessential modifications of these operators. Introduced by the author (Symes, 1986b), this type of annihilator has come to be called *differential semblance* ("DS").

We have already discussed the behaviour of the OLS function at some length, and this partially linearized version of it shares the same propensity for multimodality. PSDM cannot in general be used as a straightforward substitute for linearized inversion in this partially linearized OLS formulation of objective velocity analysis: the mismatches of amplitudes can actually generate yet more spurious stationary points.

The second choice (image power) is much less affected by the amplitude differences between (pseudo)inversion and PSDM, but also exhibits a pronounced tendency to multimodality: with prototypical data bandwidth and acquisition geometry, many critical points typically appear, most having nothing to do with a physically meaningful solution. Chauris has carefully illustrated this phenomenon, and noted that it is caused by isolated tangencies between unrelated events in predicted and target data (Chauris, 2000; Chauris and Noble, 2001). As is the case with OLS, a sufficiently good initial guess, made easier to supply by downfiltering the data in early stages, can yield convergence for image power optimization. Soubaras has recently provided an excellent example of this strategy (Soubaras and Gratacos, 2006), and it has been used to good effect by Shen and colleagues as well (Shen and Calandra, 2005; Shen et al., 2005a).

Differential Semblance

The differential semblance (DS) class has many variants, some employing PSDM (i.e. $D\bar{F}[v]^*$) rather than (pseudo)inversion to construct the function J_A. DS variants based on surface oriented extension have used

- convolutional model simulation of plane wave data (Symes, 1986b; Symes, 1990; Symes and Carazzone, 1991; Symes, 1991a; Symes and Carazzone, 1992; Minkoff and Symes, 1997),
- convolutional model simulation of CMP data (Symes, 1993; Symes and Gockenbach, 1995; Symes, 1998; Li and Symes, 2005; Dussaud and Symes, 2005; Verm and Symes, 2006),
- two way wave equation modeling (two way reverse time migration) (Symes, 1991c; Symes and Versteeg, 1993; Kern and Symes, 1994),
- generalized Radon transform simulation of acoustic scattering (Kirchhoff migration) (Chauris and Noble, 2001; Mulder and ten Kroode, 2002; de Hoop et al., 2003), and
- generalized Radon transform simulation of anisotropic elastic scattering (de Hoop et al., 2005).

Variants based on depth oriented extension have employed

- one way wave equation migration of shot profiles (Shen et al., 2005b; Albertin et al., 2006), and

- one way wave equation migration via the DSR equation (Shen et al., 2003; Khoury et al., 2006).

The same concept has been applied to cross-well tomography (Plessix, 2000).

Differential semblance has a precise mathematical characterization, which partly explains its significance:

> *Amongst all possible quadratic forms in the data, parametrized by velocity, of the form given in equation (14), only differential semblance (and inessential variations thereof) is smooth jointly as a function of smooth perturbations in velocity and finite energy perturbations in data (Stolk and Symes, 2003).*

This statement is a paraphrase of the precise mathematical fact; the reader may consult the reference for the "fine print". The significance of this statement lies in the requirement of smoothness under variation of both velocity and *finite energy* data, i.e. data perturbations of unlimited bandwidth. A related statement is that the *shape* of the differential semblance objective function is *stable* as data sampling is refined and/or upper bandlimit increased, and *no other form* of J_A has this property - not image power, not OLS, nor any other yet-to-be-invented quadratic forms. Since any smooth objective in this sense necessarily has a partial Hessian operator in d, of the form given in equation (14), this result also shows that any function of v, d must be identical to DS to second order in d, if it is to be smooth in the sense specified.

All of the works cited above suggest in one way or another than not only is the DS objective stable against high-frequency data perturbation, but it is also essentially *monomodal*: the only stationary points are physically significant solutions of the waveform inversion problem (and, in particular, velocities kinematically consistent with data). In one special case, this has even been proven with mathematical rigor: for the DS variant for CMP data, based on hyperbolic NMO (Symes and Gockenbach, 1995; Symes, 1998; Li and Symes, 2005; Verm and Symes, 2006), *all stationary points are global minima, up to an error proportional to a dominant wavelength* (Symes, 1999; Symes, 2001). The essential idea of the proof is the relation between the DS objective and fitting of apparent velocities in the data, an approach to velocity estimation also known as *stereotomography* (Billette and Lambaré, 1998). That is, DS is essentially data-weighted stereotomography, implicitly computed without event picking (see also (Chauris, 2000; Chauris and Noble, 2001)).

Influence of Coherent Noise

The linearized description of seismic scattering models single or primary scattering. It neglects multiple scattering, so that any multiply scattered energy in the data appears as coherent noise, with apparent velocity differing from that of the primary events. Since evidence of nonlinear seismic field - earth structure coupling is commonplace in reflections from the upper crust, the influence of this signal type on objective MVA algorithms must necessarily play a role in determining their practical impact.

The contrasting behaviors of the various MVA objectives carry over to their response to the presence of coherent noise such as multiple reflections. As noted above, OLS and image power optimization via Newton-like algorithms require good initial estimates of velocity structure in order to converge to a physically relevant final velocity estimate. However, OLS and image power optima are not much influenced by presence of events with differing apparent velocity, which are effectively orthogonal signal components hence thrown into the residual (OLS) or stacked out (image power). DS optima, on the other hand, are quite stable against changes in initial estimate but are strongly influenced by coherent noise, tending to yield averages of apparent velocities weighted by the corresponding event strengths (Gockenbach and Symes, 1999; Mulder and ten Kroode, 2002; Li and Symes, 2005; Verm and Symes, 2006).

Several "signal processing" approaches have been suggested to retain the global convergence properties of DS optimization while reducing its sensitivity to coherent noise. An obvious approach is application of one or more multiple suppression algorithms to the data, to render it more similar to the primaries-only data modeled by the Born approximation (Gockenbach and Symes, 1999; Li and Symes, 2005). Since the tendency of DS is to average the apparent velocities of events, the lower velocities of multiple reflections in near-layered structures will tend to appear undercorrected in image gathers, whereas the primary events appear overcorrected. Therefore, dip-filtering the undercorrected events and remodeling will have the effect of suppressing multiple reflections, under these assumptions (Mulder and ten Kroode, 2002; Verm and Symes, 2006). Finally, the *extremal regularization* algorithm (Gockenbach and Symes, 1999) perturbs data as well as velocity to find the well-fit data closest to given data. In the event that the primary reflections form the strongest consistent family of reflections in the data, the extremal regularization algorithm will find them (Gockenbach and Symes, 1999).

Another approach to desensitization of DS optimization to the presence of coherent noise relies on continuation or homotopy, using the DS velocity estimate as a starting guess, then introducing either a blend of DS and image power or OLS objectives, or simply switching to image power or OLS. Gockenbach analyzed a homotopy between DS and OLS in his thesis, and showed that with some reasonable assumptions one could guarantee that the path beginning at the DS solution terminates as the OLS solution (Gockenbach et al., 1995). Chauris (Chauris, 2000; Chauris and Noble, 2001) employed a similar approach using image power, as did Shen and collaborators (Shen et al., 2005b; Shen and Calandra, 2005). Celis and Larner proposed a similar homotopy, computing image power or an equivalent crosscorrelation over offset windows of varying width; in the limit of small width, the crosscorrelation is very close to DS in its behviour (Celis and Larner, 2002). All of these homotopies share an implicit constraint on the overall noise energy with extremal regularization: if the noise is sufficiently large, the homotopy begins at a highly corrupted DS estimate and terminates at a spurious stack power peak or OLS local minimum.

All of these ideas have been shown to be effective in limited circumstances. However, all are somewhat *ad hoc*. Since multiple scattering is a physical phenomenon predicted by modeling, a different and more fundamental approach to presence of multiply scattered energy is development of an MVA method based on full nonlinear modeling.

TOWARDS EFFECTIVE NONLINEAR WAVEFORM INVERSION

Each of the extensions underlying the two common variants of PSDM is the linearization of a full waveform modeling extension. In all cases examined in this paper, the extension map E is linear, so the very same annihilators A may be used to characterize its range and hence the physically significant solutions of the extended inverse problem.

Accordingly, suppose that $\bar{F} : \bar{M} \rightarrow D$ and $E : M \rightarrow \bar{M}$ form an extension of $F : M \rightarrow D$, and that A is an annihilator of the range of E. For the present discussion, in contrast to that of the last section and without compelling reason to do otherwise, suppose that A is a fixed, linear map. Given $d \in D$ and a fit tolerance $\epsilon > 0$, define the *feasible model* set $S_\epsilon[d]$ by

$$S_\epsilon[d] = \{\bar{m} \in \bar{M} : \|\bar{F}[\bar{m}] - d\| \leq \epsilon \|d\|\},$$

and $J_A : \bar{M} \rightarrow \mathbf{R}$ by

$$J_A[\bar{m}] = \frac{1}{2}\|A\bar{m}\|^2.$$

A solution of the constrained optimization problem

$$\text{minimize}\, J_A[\bar{m}] \text{ subject to}\, \bar{m} \in S_\epsilon[d] \tag{15}$$

is then a candidate for solution of the waveform inversion problem. If the objective value is near zero, then the solution $\bar{m}$ predicts the data to within the specified tolerance, and is close to the range of E in the sense that its image under A is small. Therefore there exists a model $m \in M$ with $Em \simeq \bar{m}$, hence $F[m] \simeq d$. Conversely, if the data $d \in D$ is consistent with the model to tolerance ϵ, that is, if $m \in M$ exists for which $\|F[m]-d\| < \epsilon$, then $\bar{m} = Em$ is *a priori* a solution of problem (15).

Of course the constrained optimization problem (15) can be rewritten as an unconstrained problem by use of a penalty (Lagrange multiplier) formulation. The author studied such a penalty reforumulation of (15) for plane wave solutions of a layered acoustic model with surface-oriented extension in (Symes, 1991b). I used a natural version of DS for A. I showed that the Hessian of the penalty function at a consistent solution is positive definite, which implies the uniqueness of the stationary point at least locally, i.e. absence of spurious stationary points. To obtain positivity of the Hessian, I imposed a *scale dichotomy* property on the model, which is reminiscent of the scale dichotomy used to justify use of perturbation theory. This study did not address the determination of appropriate values for the penalty parameter, an essential step in any practical implementation.

In general, two major issues arise in formulating any approach to the solution of problem (15):

- representation of the extended model $\bar{M}$, and

- parametrization of the feasible model set $S_\epsilon[d]$.

The first issue presents few difficulties for surface-oriented extensions, a circumstance evident in the theory developed in (Symes, 1991b). The extended model simply amounts to permitting the coefficients in the wave equation to depend on a surface acquisition parameter, and simulation is independent for each value of the parameter. Thus the computational complexity of the extended modeling operator is no greater than that of the basic modeling operator. The memory complexity is equivalent to that of image volume formation in PSDM.

For depth oriented extensions, the coefficients in the wave equation (vectors in the extended model space) are positive definite symmetric operators. The kernels or matrices of these operators occupy storage equivalent to that of image volumes in PSDM. However the computational complexity of modeling via timestepping for generalized wave equations such as (12) is potentially enormous, apparently involving a matrix multiplication at each time step. In special cases, such as the layered model discussed below, the cost of modeling can be reduced dramatically by judicious choice of basis in model space. I will come back to this issue, which is currently very much open, in the concluding discussion.

The force of the second issue arises from the very irregular geometry of the set $S_\epsilon[d]$. [For an explicit discussion of a closely related phenomenon and a careful quantification, see (Chavent, 1991).] This puts the constrained optimization problem (15), in its original form, beyond the reach of contemporary numerical optimization (Nocedal and Wright, 1999). A reparametrization is essential, so that the problem becomes a smooth one to which once can apply the Lagrangian theory behind successful large scale methods.

Note that MVA, viewed as a route to partially linearized inversion via tangent extended modeling, solves the reparametrization problem via scale dichotomy and linearization. The analogue of S_ϵ for MVA consists of the set of pairs $v, D\bar{F}[v]^\dagger d$, parametrized by the macromodel velocity v and the data d. This set is also highly curved, but the evaluation of J_A (as defined in this section) on this set, viewed as a function of v, is smooth provided that an appropriate choice of A is made, as noted above.

An effective reparametrization of $S_\epsilon[d]$ for layered modeling uses the well-known slowness-dependent traveltime parametrization of depth (Symes, 1991b). One can equally well view this step as parametrization of the model via the data, via the solvability of the one-dimensional *impulsive* inverse problem (i.e. with data at all frequencies including 0 Hz). Once again, the restriction of J_A is smooth as a function of the reparametrized model, for the DS choice of A.

Some numerical evidence (but little theoretical backup) exists that the impulsive several dimensional problem has unique solutions, in several of its settings. For example, the previously mentioned work of Bunks, Chavent, and collaborators (Bunks et al., 1995), also that of Shin and collaborators (Shin et al., 2001), may be interpreted as suggesting that the 2D common source point inverse problem has a well-defined and computationally tractable solution provided that essentially impulsive data is available. This evidence supports the hypothesis that solution of impulsive source problems may provide appropriate reparametrization for $S_\epsilon[d]$, at least for surface oriented extension.

No evidence, numerical or otherwise, seems to exist concerning constructive reparametrization of the depth oriented extension in its nonlinear form. It is possible however to analyze a special case, that of layered models.

Depth oriented extension for layered models

Assume that the Earth is horizontally layered, and that the depth oriented extended model (12) should also model laterally homogeneous action-at-a-distance. That is, shifting the source location $\mathbf{x}_s$ horizontally should produce a corresponding shift in the wavefield $\bar{u}$ in the extended model (12). Assuming that the source and receiver positions lie at the same depth, and that the offsets are the same from shot to shot, it follows that the data is also a function of offset only. Moreover, it follows that the operator $\bar{v}^{-2}$ is necessarily a *convolution operator in the horizontal vaiables*, For computational purposes, it is convenient to represent $\bar{v}^2$ as convolution with a kernel σ, and for convenience I will assume a 2D Earth (or equally well a line source in the y direction). For any function $g(z,x)$ even in x,

$$\bar{v}^2 g(z,x) = \frac{1}{\pi}\int dk \cos kx \hat{\sigma}(z.k)\hat{g}(z,k),$$

where $\hat{g}$ is the cosine transform of g. Assume, also for convenience, that the velocity field is known near the surface, and in particular at the source depth, and is physical: for z near z_s, $\bar{v}(z,x) = v_0\delta(z-z_s)\delta(x-x_s)$. Locating the source for convenience at $x_s = 0$, the wavefield u is perforce even in x. The extended modeling equation (12) then implies for the cosine transform $\hat{u}$

$$\left[\frac{\partial^2}{\partial t^2} - \hat{\sigma}(z,k)\left(\frac{\partial^2}{\partial z^2} - k^2\right)\right]\hat{u}(t,z,k) = v_0^2 w(t)\delta(z-z_s). \tag{16}$$

Representing the operator $\bar{v}^2$ in this way has made its action *sparse* (diagonal) in an appropriate basis, so that its computational cost is reduced to the same level as that of a physical velocity model (i.e. multiplication operator). Accordingly, the wave equation (16) may be solved as efficiently as the basic modeling equation (2) (actually, more efficiently, in this case).

The extended forward modeling operator has the representation

$$\bar{F}[\bar{v}](t,z_r,x_r) = \frac{1}{\pi}\int dk \cos kx_r \frac{\partial \hat{u}}{\partial t}(t,z_r,k) \tag{17}$$

in terms of the cosine transformed field $\hat{u}$, which suggests an efficient implementation of $\bar{F}$.

A simple finite difference scheme of order 2 in time and 4 in space is amply accurate to approximate the solution of the extended modeling equation (16). Discrete cosine transform then provides a natural and efficient implementation of the extended modeling operator via (17). To obtain points in the feasible model set $S_\epsilon[d]$, I approximated the solution of the extended inversion problem

$$\text{minimize } \|\bar{F}[\bar{v}] - d\| \text{ over } \bar{v} \tag{18}$$

using a quasi-Newton method (limited memory BFGS, see (Nocedal and Wright, 1999)). This technique required computation of the gradient of the objective appearing in (18). I computed the gradient using the adjoint state method (Tarantola, 1987; Plessix, 2006). A useful by-product of these constructions was a quasi-Newton method for solving the OLS problem (1) for the basic forward modeling operator F of layered acoustics.

The examples to follow all pertain to the same simple three-layer model, depicted in Figure 1. The surface $z = 0$ of this model is pressure-free, modeled by a Dirichlet boundary condition. The reflectivities are chosen so that the free surface multiple reflection from the first layer is roughly the same strength as the primary reflection from the second layer, see Figure 2.

Iterative OLS inversion for a layered velocity model yields results which depend strongly on the initial estimate, consistent with the accumulated experience of other studies as reviewed above. Three such results are depicted in Figure 3. Only the inversion from the initial model with correct velocity in the second layer (blue curves in Figure 3) places the second reflector at the right depth. Inversion cannot of course correct the velocity below the second interface, as no reflections return from deeper in the model. Each iteration was halted when the gradient decreased in length by two orders of magnitude from its initial value.

Iterative *extended* OLS inversion also produced results depending on the initial estimate. For the same initial estimates as in used in the OLS trials (Figure 3), or rather their extensions (images under E), the extended inversions are depicted in Figures 4, 5, and 6. In all three cases, BFGS iteration was terminated when the RMS fit error declined to 1 %. The predicted seismograms, that is, $\bar{F}[\bar{v}]$, are shown for the three cases in Figures 7, 8, and 9. All are visually indistinguishable from the data depicted in Figure 2.

These results suggest several conjectures about this particular extended inverse problem:

- The range of $\bar{F}$ appears to be stable (analogue of surjectivity for linearized extention), and fairly precise fits to data are possible via local optimization (limited memory BFGS) starting at a wide variety of initial model estimates;
- Only the extended OLS estimate produced by a kinematically accurate initial model (Figure 6) is reasonably close to the range of the extension map E, i.e. concentrated near zero offset. Moreover, the degree of concentration appears to change smoothly as the initial velocity becomes kinematically more accurate;
- The treatment of the extended OLS problem could benefit from some regularization (not applied in these experiments): all estimates produced in this series are relatively noisy.

Thus the nonlinear analogue of the surjectivity property important in the MVA analysis, appears to hold at least to the limited extent investigated in these examples. Furthermore, the behavior of these extended OLS inversions is roughly consistent with the expectation that the function J_A, for A = multiplication by offset, may in some sense change smoothly with position in the feasible model set $S_\epsilon[d]$ (in which Figures 4 through 6 represent three points).

DISCUSSION AND CONCLUSION

The foregoing discussion has reviewed the failure of waveform (OLS) inversion to update macromodels and the contrasting ability of MVA to do so. MVA is based on linearized or Born modeling, and has at least two major variants corresponding to two approaches to image volume formation. The extension concept formalizes image volume formation as application to data of an adjoint (or pseudoinverse) tangent modeling operator. This identification clarifies the meaning of the tendency of surface oriented image volumes to exhibit kinematic artifacts in the presence of multipathing: the corresponding extended adjoint tangent extended modeling operator is not a good approximation to the pseudoinverse. The depth oriented extension on the other hand maintains a close relationship (modulo dip-dependent amplitude scaling and filtering) between adjoint and pseudoinverse tangent extended modeling operators under a much wider range of circumstances. In both cases, MVA may be cast as an optimization problem in a number of ways. Only one of these, differential semblance, appears to yield optima that are robust against large errors in initial model estimates.

The extension concept permits a unified view of OLS and MVA approaches to Earth model estimation: MVA can be regarded as a solution method for a partially linearized OLS inverse problem. The extension framework also suggests an approach to nonlinear waveform inversion incorporating elements of MVA. Preliminary numerical experiments hint that this nonlinear extended waveform inversion may exhibit the global convergence attribute of MVA while preserving the ability of waveform inversion to model nonlinear material-wave interaction (and, in more general settings, other physics of waves).

Two remaining obstacles must be overcome before this nonlinear MVA approach to waveform inversion becomes in any sense practical. The first concerns the depth oriented extension: in this formulation, nonphysical models are operators, and must be represented as sparse matrices in order that timestepping simulation is computationally feasible. The simple solution sketched above for layered models, via convolution and the Fourier transform, is not available in general. A possible resolution may lie in a *phase space representation*: the kernels of these operators are analogous to space shift image volumes in survey-sinking PSDM, and these data tend to be very sparse in phase space (one or at most a few dips at any one point in space). Fast algorithms have been developed in the last few years for representation of functions possessing the typical phase space sparsity properties of seismic images (Candes and Demanet, 2005; Chauris, 2006). These may lead to a resolution of the depth oriented extension representation problem.

The second obstacle is the problem of extreme geometric irregularity of the feasible model set $S_\epsilon[d]$, and the consequent poor performance of standard constrained optimization tools such as sequential quadratic programming, which typically fail to solve the problem (15) when applied directly. Earlier work on the layered acoustic problem (Symes, 1991b) suggested that reparametrizing the feasible model set via impulse response inversion may circumvent this difficulty. Some numerical evidence indicates that impulse response inversion for surface oriented extensions may indeed be possible, for nonlayered problems as well. Whether a similar statement can be made for depth oriented extensions remains to be seen. This is a matter of some import, as the nonlinear scattering effects which motivate the nonlinear MVA approach sketched here tend to occur in conjunction with complex ray geometry. A notorous example of this phenomenon occurs in the vicinity of salt bodies embedded in sedimentary sequences, giving rise to rugose high contrast interfaces generating complex ray paths and strong multiple reflections. In the presence of such complexity (multipathing), the structure of the inverse problem for depth oriented extensions is simpler (more straightforward relation between adjoint and pseudoinverse for tangent operators) than is the case for surface oriented extensions. This reasoning suggests that the impulse response inverse problem for depth oriented extension is worth some study.

ACKNOWLEDGEMENTS

I am indebted to many colleagues for their insights into velocity analysis and seismic inversion. The work reported here was supported in part by the sponsors of The Rice Inversion Project.

REFERENCES

Albertin, U., P. Sava, J. Etgen, and M. Maharramov, 2006, Adjoint wave equation velocity analysis: 76th Annual International Meeting, Expanded Abstracts, TOM2.1, Society of Exploration Geophysicists.

Backus, G. and F. Gilbert, 1968, The resolving power of gross earth data: Geophysical Journal of the Royal Astronomical Society, **16**, 169–205.

———, 1970, Uniqueness in the inversion of inaccurate gross earth data: Philosophical Transactions of the Royal Society of London, Series A, **266**, 123–192.

Bamberger, A., G. Chavent, and P. Lailly, 1977, Etude mathématique et numérique d'un problem inverse pour l'Équation des ondes a une dimension: Rapport Interne 14, Centre de Mathématiques Appliqués, École Polytechnique, Paris.

———, 1979, About the stability of the inverse problem in 1-d wave equation — application to the interpretation of seismic profiles: Appl. Math. Opt., **5**, 1–47.

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

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

Biondi, B., 2003, Equivalence of source-receiver migration and shot-profile migration: Geophysics, **68**, 1340–1347.

Biondi, B. and G. Palacharla, 1996, 3-d prestack migration of common-azimuth data: Geophysics, **61**, 1533–1543.

Biondi, B. and G. Shan, 2002, Prestack imaging of overturned reflections by reverse time migration: 72nd Annual International Meeting, Expanded Abstracts, 1284–1287, Society of Exploration Geophysicists.

Biondi, B. and W. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, **69**, 1283–1298.

Bleistein, N., 1987, On the imaging of reflectors in the earth: Geophysics, **52**, 931–942.

Brandsberg-Dahl, S., M. De Hoop, and B. Ursin, 2003, Focusing in dip and AVA compensation on scattering angle/azimuth common image gathers: Geophysics, **68**, 232–254.

Brenders, A. and G. Pratt, 2006a, Efficient waveform tomography for lithospheric imaging: implications for realistic, 2-d acquisition geometries and low frequency data: Geophysical Journal International. (in press).

———, 2006b, Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model: Geophysical Journal International. (in press).

Bube, K. and R. Burridge, 1983, The one dimensional inverse problem of reflection seismology: SIAM Review, **25**, 497–559.

Bunks, C., F. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Burridge, R., M. de Hoop, D. Miller, and C. Spencer, 1998, Multiparameter inversion in anisotropic elastic media: Geophysical Journal International, **134**, 757–777.

Candes, E. and L. Demanet, 2005, The curvelet representation of wave propagatorss is optimally sparse: Communications on Pure and Applied Mathematics, **58**, 1472–1528.

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

Carazzone, J. and L. Srnka, 1989, Elastic inversion of Gulf of Mexico data: Proc. 59th Annual International Meeting, 956–958. Expanded abstract.

Celis, V. and K. Larner, 2002, Selective-correlation velocity analysis: 72nd Annual International Meeting, Expanded Abstracts, 2289–2292, Society of Exploration Geophysicists.

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

———, 2006, Seismic imaging in the curvelet domain and its implications for the curvelet design: 76th Annual International Meeting, Expanded Abstracts, SPMI3.2, Society of Exploration Geophysicists.

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

Chavent, G., 1991, New size times curvature conditions for strict quasiconvexity of sets: SIAM Journal on Control and Optimization, **29**, 1348–1372.

Chavent, G. and P. Lemmonier, 1974, Identification de la non-linéarité d'une équation parabolique quasilinéaire: Applied Mathematics and Optimization, **1**, 121–162.

Claerbout, J., 1985, Imaging the earth's interior: Blackwell Scientific Publishers.

Claerbout, J. and S. Doherty, 1972, Downward continuation of moveout corrected seismograms: Geophysics, **37**, 741–768.

Claerbout, J. F., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481.

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

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

de Hoop, M., S. Brandsberg-Dahl, and B. Ursin, 2003, Seismic velocity analysis in the scattering-angle/azimuth domain: Geophysical Prospecting, **51**, 295–314.

de Hoop, M., S.-K. Foss, and B. Ursin, 2005, Depth-consistent reflection tomography using PP and PS seismic data: Geophysics, **70**, U51–U65.

de Hoop, M. and C. Stolk, 2002, Microlocal analysis of seismic inverse scattering in anisotropic, elastic media: Communications on Pure and Applied Mathematics, **55**, 261=301.

———, 2005, Modeling of seismic data in the downward continuation approach: SIAM Journal on Applied Mathematics, **65**, 1388–1406.

———, 2006, Seismic inverse scattering in the downward continuation approach: Wave Motion, **43**, 579–598.

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

Duquet, B. and P. Lailly, 2006, Efficient 3-d wave-equation migration using virtual planar sources: Geophysics, **71**, S185–S197.

Dussaud, E. and W. Symes, 2005, Velocity analysis from interferometric data: 75th Annual International Meeting, Expanded Abstracts, SPVA1.1, Society of Exploration Geophysicists.

Fernandez-Berdaguer, E. M., J. E. Santos, and D. Sheen, 1996, An iterative procedure for estimation of variable coefficients in a hyperbolic system: Applied Mathematics and Computation, **76**, 213–250.

Gardner, G., ed. 1985, Migration of seismic data. Geophysics Reprint Series No. 4: Society of Exploration Geophysicists,.

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

Gockenbach, M. and W. Symes, 1999, Coherent noise suppression in velocity inversion: 69th Annual International Meeting, Expanded Abstracts, 1719–1723, Society of Exploration Geophysicists.

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

Goupillaud, P., 1961, An approach to inverse filtering of near-surface layers from seismic records: Geophysics, **26**, 754–760.

Gray, S. and W. Symes, 1985, Stability considerations for one-dimensional inverse problems: Geophys. J. Roy. Astr. Soc., **80**, 149–163.

Jackson, D. D., 1972, Interpretation of inaccurate, insufficient, and inconsistent data: Geophysical Journal of the Royal Astronomical Society, **28**, 97–110.

———, 1976, Most squares inversion: J. Geophys. Research, **81**, 1027–1030.

———, 1979, The use of *a priori* data to resolve nonuniqueness in linear inversion: Geophysical Journal of the Royal Astronomical Society, **57**, 137–157.

Jin, S. and Madariga, 1994, Nonlinear velocity inversion by a two-step Monte-Carlo method: Geophysics, **59**, 577–590.

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

Khoury, A., W. W. Symes, P. Williamson, and P. Shen, 2006, DSR migration velocity analysis by differential semblance optimization: 76th Annual International Meeting, Expanded Abstracts, SPMI3.4, Society of Exploration Geophysicists.

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

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

———, 1984, Migration methods: partial but efficient solutions to the seismic inverse problem, *in* Santosa, F., ed., Inverse Problems of Acoustic and Elastic Waves. SIAM.

Lewis, M. and W. Symes, 1991, On the relation between the velocity coefficient and boundary values for solutions of the one-dimensional wave equation: Inverse Problems, **7**, 597–632.

Li, J. and W. Symes, 2005, Fast interval velocity estimation via nmo-based differential semblance: 75th Annual International Meeting, Expanded Abstracts, SPVA1.8, Society of Exploration Geophysicists.

Lines, L. and S. Treitel, 1984, A review of least-squares inversion and its application to geophysical problems: Geophysical Prospecting, **32**, 159–186.

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

Mora, P., 1988, Elastic wavefield inversion of reflection and transmission data: Geophysics, **53**, 750–759.

Mosegard, K. and A. Tarantola, 1991, Monte Carlo analysis of geophysical inverse problems: Presented at the Proc. 61st Annual International Meeting and Exposition.

Mulder, W. and A. ten Kroode, 2002, Automatic velocity analysis by differential semblance optimization: Geophysics, **67**, 1184–1191.

Nocedal, J. and S. Wright, 1999, Numerical optimization: Springer Verlag.

Nolan, C. and W. Symes, 1997, Global solution of a linearized inverse problem for the wave equation: Communications on Partial Differential Equations, **22**, 919–952.

Pao, Y.-H., F. Santosa, and W. Symes, 1984, Inverse problems of acoustic and elastic waves, *in* Santosa, F., ed., Inverse Problems of Acoustic and Elastic Waves. SIAM.

Parker, R., 1977, Understanding inverse theory: Annual Review of Earth and Planetary Science, **5**, 35–64.

Plessix, R.-E., 2000, Automatic cross-well tomography: an application of the differential semblance optimization to two real examples: Geophysical Prospecting, **48**, 937–951.

———, 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophys. J. Int., **167**, 495–503.

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

———, 1999, Waveform inversion of reflection seismic data for kinematic parameters by local optimization: SIAM Journal on Scientific Computation, **20**, 1033–1052.

Pratt, R., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory, and verification in a physical scale model: Geophysics, **64**, 888–901.

Pratt, R. and R. Shipp, 1999, Seismic waveform inversion in the frequency domain, part 2: Fault delineation in sediments using crosshole data: Geophysics, **64**, 902–914.

Prucha, M., B. Biondi, and W. Symes, 1999, Angle-domain common image gathers by wave-equation migration: 69th Annual International Meeting, Expanded Abstracts, 824–827, Society of Exploration Geophysicists.

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

Rickett, J. and P. Sava, 2002, Offset and angle-domain common image-point gathers for shot profile migration: Geophysics, **67**, 883–889.

Robinson, E. and S. Treitel, 1980, Geophysical signal analysis: Prentice-Hall.

Sacks, P., 1988, The inverse problem for a weakly inhomogeneous two-dimensional medium: SIAM Journal on Applied Mathematics, **48**, 1167–1193.

Sacks, P. and F. Santosa, 1987, A simple computational scheme for determining the sound speed of an acoustic medium from the surface values of its impulse response: SIAM J. Sci. and Stat. Comp., **3**, 501–520.

Santosa, F. and H. Schwetlick, 1982, The inversion of acoustic impedance profile by methods of characteristics: Wave Motion, **4**, 99–110.

Santosa, F. and W. Symes, 1989, An analysis of least-squares velocity inversion, volume **4** *of* Geophysical Monographs: Soc. of Expl. Geophys.

Sava, P. and S. Fomel, 2003, Angle-domain common-image gathers by wavefield continuation methods: Geophysics, **68**, 1065–1074.

———, 2005, Time-shift imaging condition: 75nd Annual International Meeting, Expanded Abstracts, SPMI2–2, Society of Exploration Geophysicists.

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

———, 1991b, Nonlinear one-dimensional seismic waveform inversion using simulated annealing: Geophysics, **56**, 1624–1636.

Shen, P. and H. Calandra, 2005, One-way waveform inversion within the framework of adjoint state differential migration: 75th Annual International Meeting, Expanded Abstracts, SI3.5, Society of Exploration Geophysicists.

Shen, P., E. Menyoli, and H. Calandra, 2005a, Automatic sub-salt velocity analysis: an integrated strategy of ray-based tomography and wave-equation migration velocity inversion: 75th Annual International Meeting, Expanded Abstracts, SPMI6.3, Society of Exploration Geophysicists.

Shen, P., W. Symes, S. Morton, and H. Calandra, 2005b, Differential semblance velocity analysis via shot profile migration: 75th Annual International Meeting, Expanded Abstracts, SPVA1.4, Society of Exploration Geophysicists.

Shen, P., W. Symes, and C. Stolk, 2003, Differential semblance velocity analysis by wave-equation migration: 73rd Annual International Meeting, Expanded Abstracts, 2135–2139, Society of Exploration Geophysicists.

Shin, C., S. Jang, and D.-J. Min, 2001, Improved amplitude preservation for prestack depth migration by invoerse scattering theory: Geophysical Prospecting, **49**, 592–606.

Sirgue, L. and G. Pratt, 2004, Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.

Sommerfeld, A., 1964, Lectures on theoretical physics: Academic Press.

Soubaras, R. and B. Gratacos, 2006, Velocity model building by semblance maximization of modulated-shot gathers: 76th Annual International Meeting, Expanded Abstracts, SVIP1.3, Society of Exploration Geophysicists.

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

———, 2000b, On the modeling and inversion of seismic data: PhD thesis, Universiteit Utrecht.

Stolk, C. C. and W. W. Symes, 2003, Smooth objective functionals for seismic velocity inversion: Inverse Problems, **19**, 73–89.

———, 2004, Kinematic artifacts in prestack depth migration: Geophysics, **69**, 562–575.

Symes, W., 1980, Numerical stability in an inverse scattering problem: SIAM J. Num. Analysis, **17**. MRC TSR 1990, (August 1979).

———, 1983, Impedance profile inversion via the first transport equation: J. Math. Anal. Appl., **94**, 435–453.

———, 1986a, On the relation between coefficient and boundary values for solutions of webster's horn equation: SIAM J. Math. Anal., **17**, 1400–1420.

———, 1986b, Stability and instability results for inverse problems in several-dimensional wave propagation, *in* Glowinski, R. and J. Lions, eds., Proc. 7th International Conference on Computing Methods in Applied Science and Engineering. North-Holland.

———, 1990, Velocity inversion: A case study in infinite-dimensional optimization: Mathematical Programming, **48**, 71–102.

———, 1991a, A differential semblance algorithm for the inverse problem of reflection seismology: Computers and Mathematics with Applications, **22**, 147–178.

———, 1991b, Layered velocity inversion: a model problem from reflection seismology: SIAM Journal on Mathematical Analysis, **22**, 680–716.

———, 1991c, Non-interactive estimation of the Marmousi velocity model by differential semblance optimization: Initial trials: Presented at the The Marmousi Experience: Proceedings of the EAEG Workshop on Practical Aspects of Inversion.

———, 1993, A differential semblance criterion for inversion of multioffset seismic reflection data: Journal of Geophysical Research, **98**, 2061–2073.

———, 1995, Mathematical Foundations of Reflection Seismology: Technical report, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA. `www.trip.caam.rice.edu`.

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

———, 1992, Velocity inversion by coherency optimization, *in* Bednar, J., ed., Geophysical Inversion, 59–89. Society for Industrial and Applied Mathematics.

Symes, W. and M. Gockenbach, 1995, Waveform inversion for velocity: Where have all the minima gone?: 65th Annual International Meeting of the Society of Exploration geophysicists, Expanded Abstracts, 1235–1239.

Symes, W. and R. Versteeg, 1993, Velocity model determination using differential semblance optimization: 63rd Annual International Meeting, Expanded Abstracts, 696–699, Society of Exploration Geophysicists.

Symes, W. W., 1998, High frequency asymptotics, differential semblance, and velocity analysis: 68th Annual International Meeting, Expanded Abstracts, 1616–1619, Society of Exploration Geophysicists.

———, 1999, All stationary points of differential semblance are asymptotic global minimizers: layered acoustics: Technical Report 99-29, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.

———, 2001, Asymptotic analysis of differential semblance for layered acoustics: Technical Report 01-21, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.

———, 2002, Kinematics of reverse time shot-geophone migration: The Rice Inversion Project, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA: http://www.trip.caam.rice.edu.

Taner, M. and F. Koehler, 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.

———, 1987, Inverse problem theory: Elsevier.

Tarantola, A., E. Crase, M. Jervis, Z. Konen, J. Lindgren, K. Mosegard, and M. Noble, 1990, Nonlinear inversion of seismograms: State of the art: Presented at the Proc. 60th Annual International Meeting and Exposition. Expanded abstract S13.7.

Ten Kroode, A. P. E., D. J. Smit, and A. R. Verdel, 1998, A microlocal analysis of migration: Wave Motion, **28**, 149–172.

Toldi, J., 1989, Velocity analysis without picking: Geophysics, **54**, 191–199.

Treitel, S., P. Gutowski, and D. Wagner, 1982, Plane-wave decomposition of seismograms: Geophysics, **47**, 1375–1401.

Verm, R. and W. Symes, 2006, Practice and pitfalls in nmo-based differential semblance velocity analysis: 75th Annual International Meeting, Expanded Abstracts, SI3.7, Society of Exploration Geophysicists.

Xu, S., H. Chauris, G. Lambaré, and M. Noble, 2001, Common angle migration: A strategy for imaging complex media: Geophysics, **66**, 1877–1894.

Yilmaz, O., 2001, Seismic data processing, *in* Investigations in Geophysics No. 10. Society of Exploration Geophysicists.

FIGURE CAPTIONS

Figure 1. A simple three layer velocity model.

Figure 2. Common shot gather for velocity model of Figure 1. Source wavelet is (4,10,30,40) Hz trapezoidal zero-phase bandpass filter.

Figure 3. Results of three trials of OLS inversion for the data of Figure 2. In each case (red, green, blue), the dashed line is the initial velocity estimate, the wiggly line is the final estimate. LBFGS iteration halted when the gradient norm dropped to 1% of its initial value. Note that (a) only when the initial velocity is correct up to the second interface (blue lines) is the estimate of the second interface position correct; (b) the surface multiple from the second interface shows up as an inhomogeneity in the estimated velocity for both the constant velocity initial estimate (red lines) and the estimate with a second layer at 2.0 km/s (green lines), but disappears when the kinematics are correct (blue lines); (c) the iteration fails in both of the first two cases to update the second layer mean velocity; (d) in all cases the velocity below the second interface remains at its initial value, because no constraint is present in the data.

Figure 4. Inversion for extended model, constant initial velocity ($\equiv$ 1.5 km/s, red dashed line in Figure 3). Note the large amount of energy spread across the entire offset range.

Figure 5. Inversion for extended model, initial velocity includes jump to 2 km/s at depth of 0.2 km (green dashed line in Figure 3). Energy still present at nonzero offsets, but less so than in Figure 4.

Figure 6. Inversion for extended model, initial velocity includes jump to 2.4 km/s at depth of 0.2 km (blue dashed line in Figure 3), which is kinematically correct down to the second interface. Energy mostly focused at zero offset, except for inversion noise which should be suppressed by proper regularization. Note that second reflector now appears at correct depth, and multiple reflection (visible in Figures 4 and 5) has disappeared.

Figure 7. Resimulated data from inverted model of Figure 4.

Figure 8. Resimulated data from inverted model of Figure 5.

Figure 9. Resimulated data from inverted model of Figure 6.

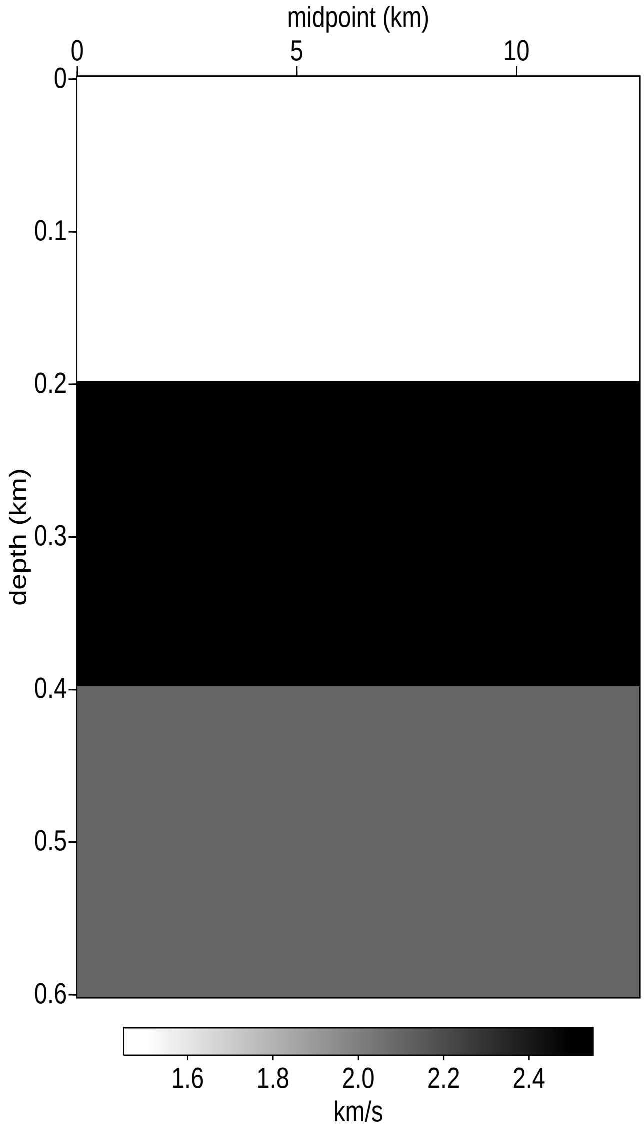

FIG. 1.

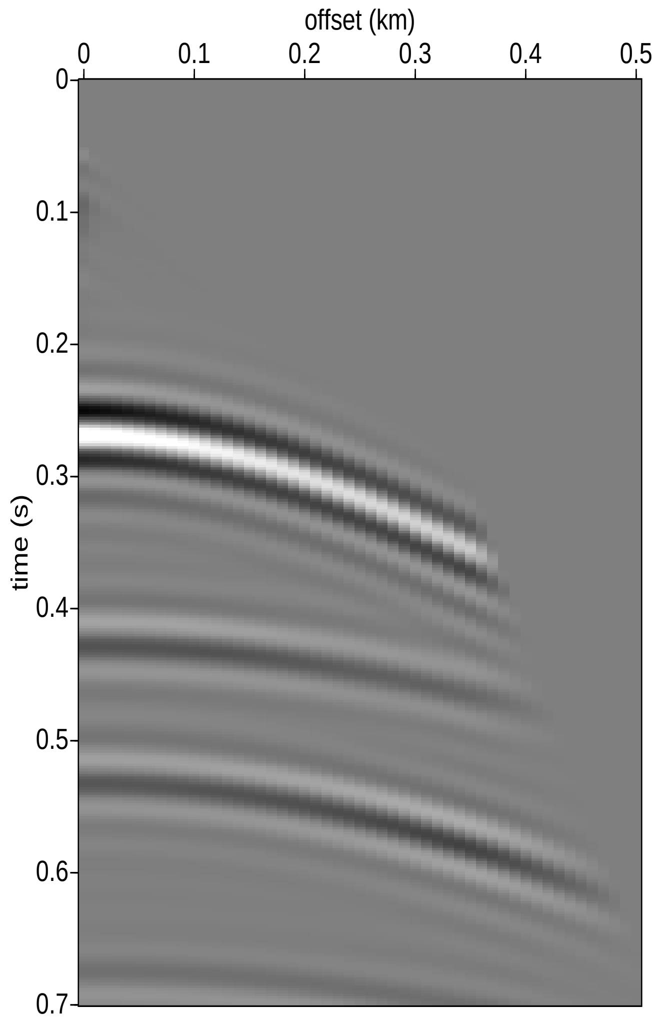

FIG. 2.

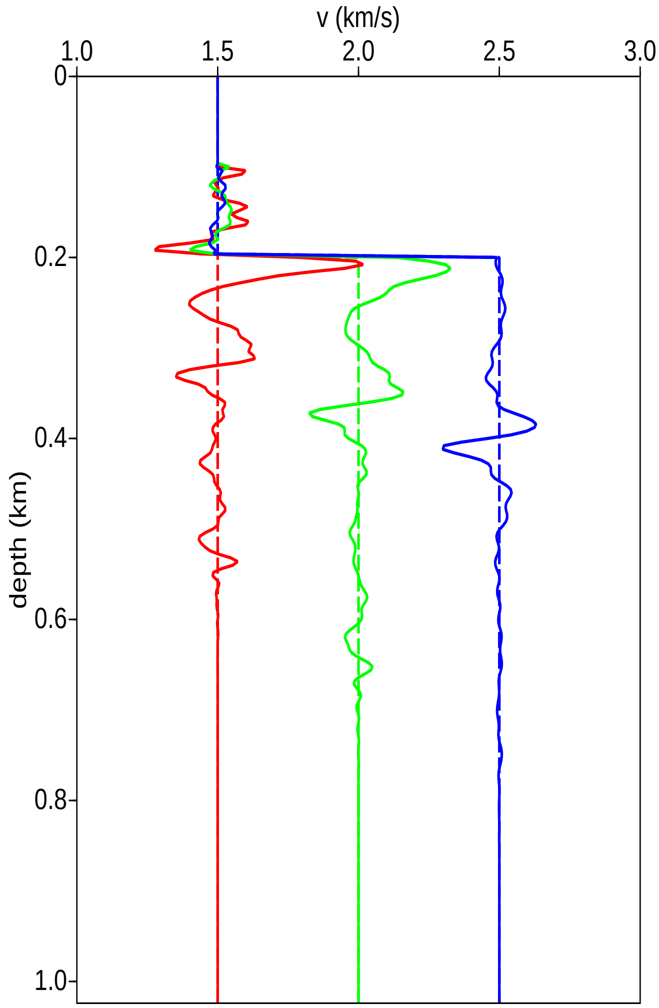

FIG. 3.

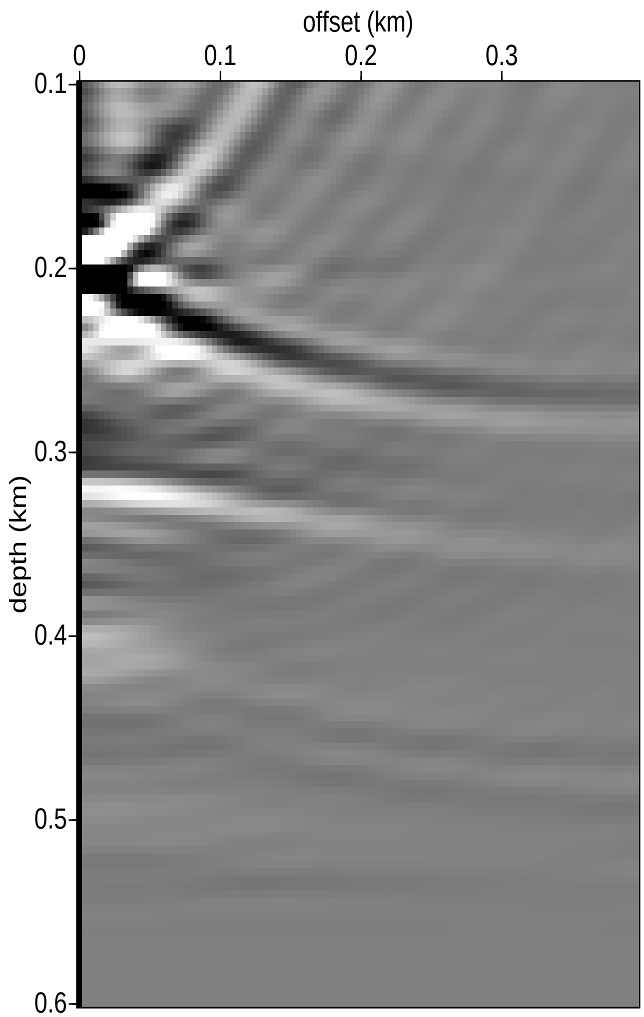

Fig. 4.

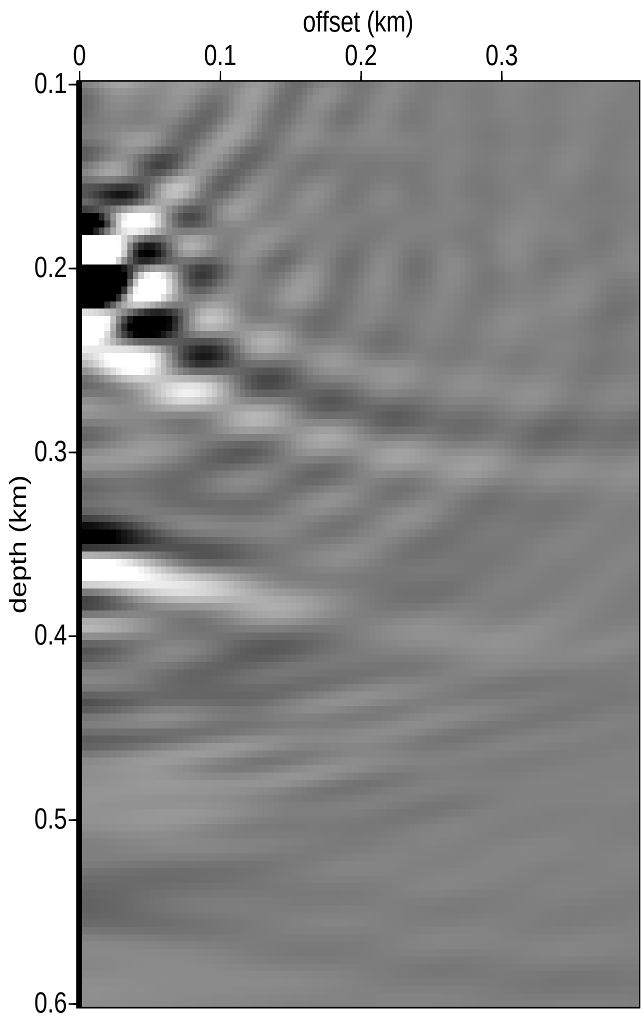

FIG. 5.

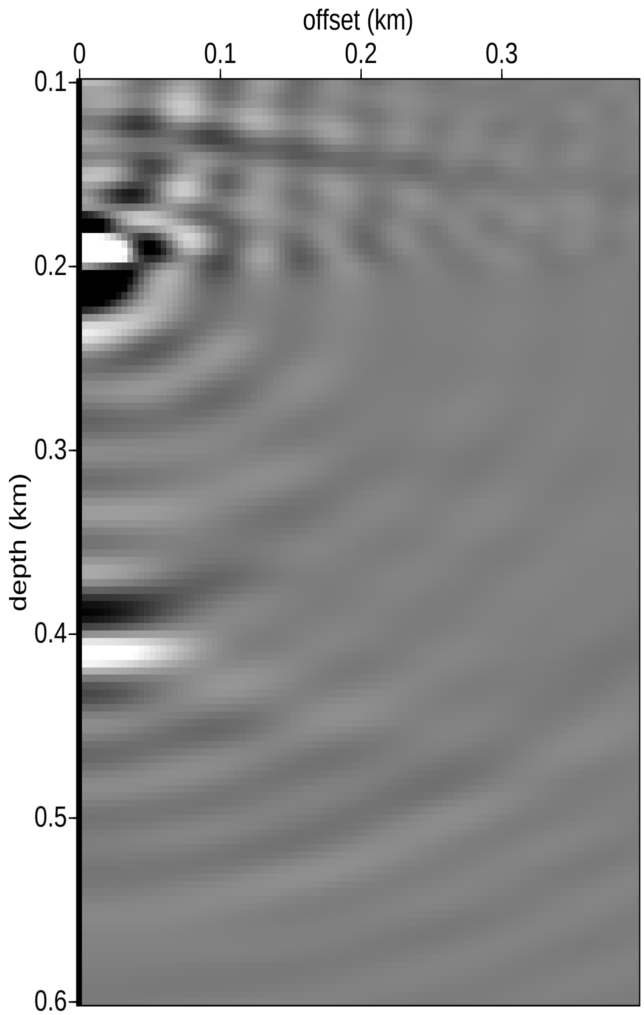

FIG. 6.

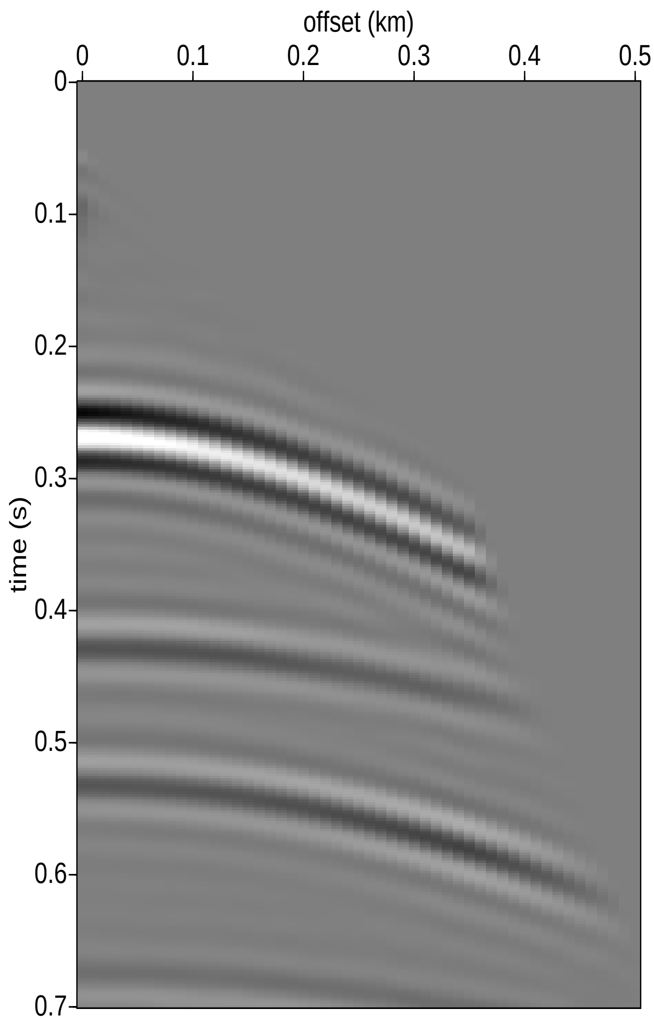

Fig. 7.

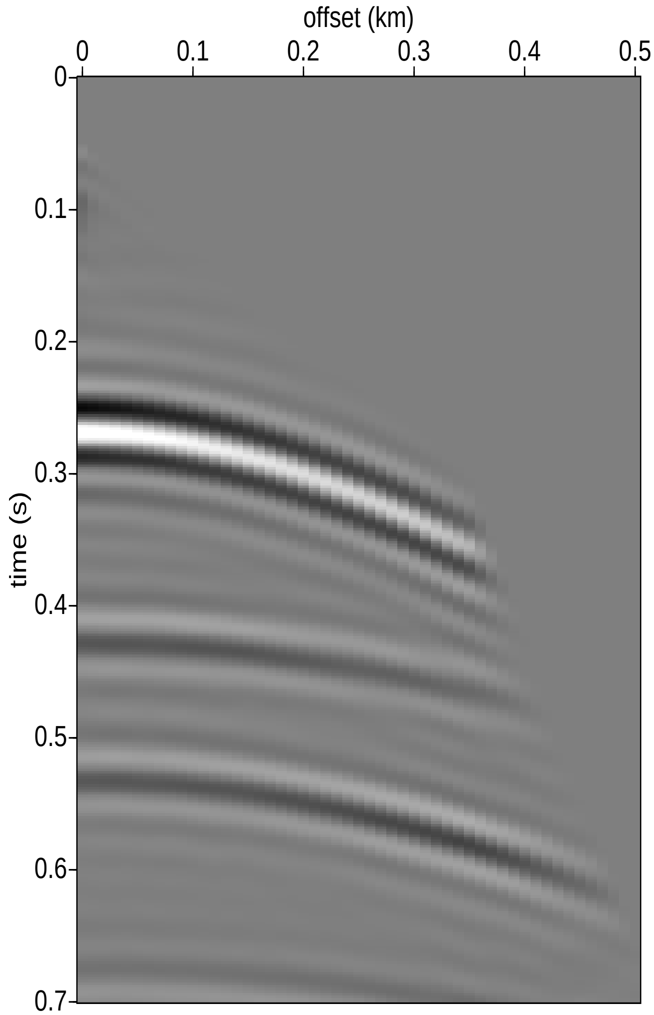

Fig. 8.

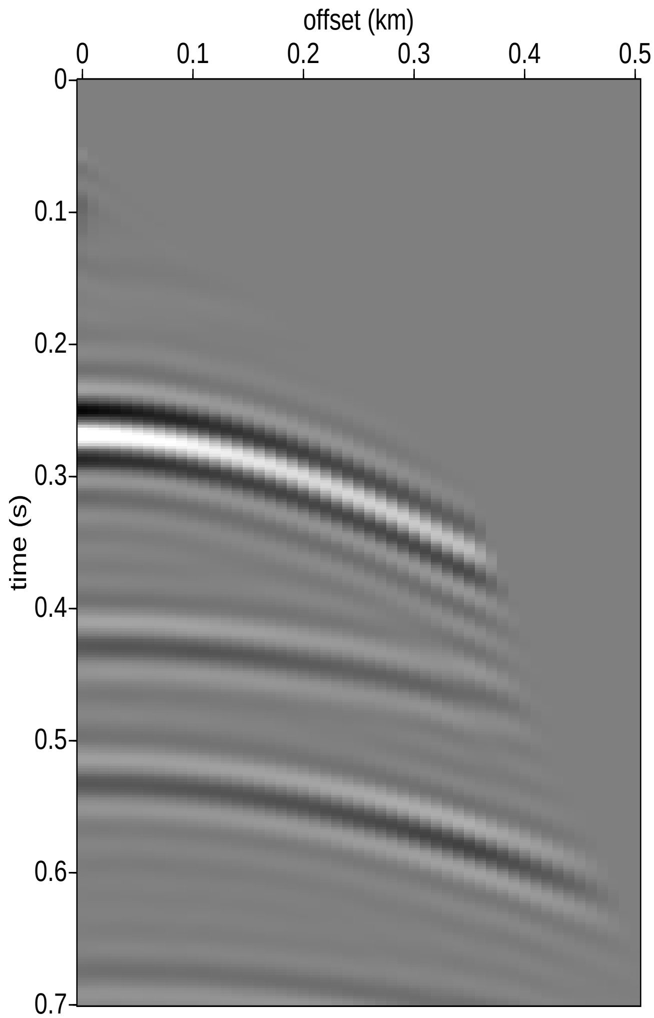

FIG. 9.