Flexible seismic traveltime tomography applied to diving waves

William S. Harlan [1]

ABSTRACT

To prepare for 3D anisotropic applications of traveltime tomography I found it necessary to reformulate the estimation of raypaths and the parameterization of velocities. Raypaths should be efficient to calculate and store in memory without recalculation. Velocities should be able to change with angle when necessary.

Ray relaxation can optimize ray paths through an anisotropic medium described only by group velocities. Three parameters adequately describe transversely isotropic group velocities with a vertical axis of symmetry. One parameter changes most arbitrarily, one changes only along the vertical axis of symmetry, and one remains a constant.

3D raypaths are described as sums of smooth curves with a small number of coefficients. A generic Gauss-Newton algorithm perturbs these coefficients to minimize traveltimes between endpoints. A small number of coefficients are saved in memory to describe paths efficiently.

Diving wave tomography can use conventional traveltime tomographic algorithms to model and invert the traveltimes of direct arrivals. Early synthetic tests minimized the complexity of velocity anomalies necessary to explain the data. Raypaths robustly converge to their appropriate distribution.

INTRODUCTION

In the past few years, I have found my previous methods of traveltime tomography too restrictive for many important current applications. Most importantly, the parameterization of velocities did not allow the introduction of anisotropy, and the methods of constructing paths and traveltimes were impractical for large sets of three dimensional picks. See Harlan et al (1991a; 1991b; 1990; 1992).

[1]**email:** harlan@sep.stanford.edu

Crosswell data frequently appear impossible to explain without anisotropy, and depth imaging often finds it impossible to explain both moveouts and well-ties with a single isotropic velocity. In this paper, I propose an algorithm using group velocities that can change with angle. The computational difficulty is only slightly greater, and the flexibility will be available when necessary.

Previously, I found explicit methods of extrapolating traveltime tables convenient, but such tables take too much memory to be stored for a large 3D dataset. Recomputation of traveltime tables is likely to be prohibitively expensive. Even sampled raypaths also can require too much memory to be stored. Instead I optimize the coefficients of smooth basis functions according to Fermat's principle, to minimize the ray traveltime. Only a few coefficients need be stored to describe each raypath.

The first obvious applications of this method are to crosswell data and to surface diving-wave tomography for near surface velocities. In later work, reflection points will be included. Because less work has been done on large 3D datasets for diving wave tomography and because there is less understanding of the non-uniqueness, I address diving waves first. Diving waves are generalizations of surface refractions. Conventional refraction analysis describes the near surface as layers of slabs, whose velocities change very smoothly, if at all, in the interior, and very sharply at the boundaries between slabs. The first arrival times are assumed to increase linearly with offset for a given refraction. An interpreter must distinguish individual refractions on the first arrival. Unfortunately, first arrivals rarely appear to be composed of a series of distinct linear segments.

Diving waves allow the interpreter to model all direct arrivals, without identifying individual refractions. Velocities are allowed to vary arbitrarily but are expected to increase overall with depth. When sources and receivers on the surface are relatively close (hundreds of meters), the wave energy that arrives first is confined to the very near surface. When the separation reaches several kilometers, then the earliest arrivals contain energy that has passed a kilometer or so in depth. For example, if the velocity at the surface of the earth were 2 km/s, and this velocity increased 1 km/s with every 1 km of depth (reaching 4 km/s at 2 km depth), then a source and receiver at an offset of 4.47 km would produce a circular arc that reached a maximum depth of 1 km. An offset of 6.93 km would reach 2 km depth.

Recent papers have used tomographic methods to reconstruct near surface velocities from surface measurements of first arrival times. Simmons, Bernitsas, and Backus (1988; 1992; 1994) were able to reconstruct impressive images with simple assumptions of semicircular raypaths. Zhu, Sixta, and Angstman (1992a; 1992b) allowed more flexible raypaths for improved 2D images. Stefani (1993) prepared an excellent 2D case study which demonstrated the accuracy of estimated velocities for depth conversion. Bell, Lara, and Gray (1994) showed that targeted tomography was appropriate in certain areas. Zhang and McMechan (1994) and Qin, Cai, and Schuster, (Qin et al., 1993; Cai and Qin, 1994) solved very general isotropic 3D geometries. J.A. Hole et al (1992; 1992) used explicit traveltime extrapolations for good inversions of 3D diving traveltimes with limited surface coverage. Many other

papers consider diving waves without tomography (Laski, 1978; Levin, 1994). The literature on the inversion of refracted waves also contains many good ideas which could be generalized for diving waves (Clayton and McMechan, 1981; Landa et al., 1994; Zanzi and Carlini, 1991; Palmer, 1981; Hagedoorn, 1964; Hawkins, 1961).

Diving wave tomography has only recently incorporated anisotropy, which appears more in crosswell tomographic applications (Michelena, 1994; Michelena et al., 1993; Vassiliou et al., 1994; Saito, 1991; Pratt and McGaughey, 1991). I will follow many of the suggestions of Grechka and McMechan (1995a; 1995b) for diving wave tomography. They use Chebyshev polynomials to describe raypaths and have already performed SVD analysis to show the non-uniqueness introduced by anisotropy into diving-wave tomography.

In this paper, a continuous anisotropic velocity model is parameterized with a minimal number of discrete parameters. Most of the cited publications prefer to work with discrete bins and may even discretize the physical modeling. Instead, I assume a continuous velocity model with a finite number of coefficients to scale continuous basis functions.

Most cited diving wave papers construct sampled rays from shooting methods or from traveltime tables. Shooting methods can be fast, but are efficient only if one sorts through the data in a particular order. Rays must be stored in their sampled form, which may take too much memory for 3D datasets. Traveltime tables cannot be saved even for smallish datasets. This paper constructs rays as a scaled sum of smooth basis functions. Only a small number of coefficients need to be stored to describe an entire continuous raypath. A generic Gauss-Newton algorithm optimizes raypaths and velocities without any features specific to this application.

ANISOTROPIC VELOCITY PARAMETERS

In this section I will propose a simple parameterization of angle-dependent velocity appropriate for relatively weak anisotropy with a vertical axis of symmetry. Anisotropy will be assumed to derive from layered media that may be isotropic on a fine scale but which will appear to be anisotropic on the scale of larger seismic wavelengths. The anisotropic parameters will be described as a sum of smooth basis functions, with no more spatial variations than necessary to explain the data. The form will allow easy linearization of slowness with respect to the model coefficients.

The parameterization of velocities should have enough degrees of freedom to describe all plausible models; the particular numerical formulation of this parameterization is less important. Our methods should be able to request the velocity at any point and in any direction from an essentially continuous model.

The model should also be differentiable, to allow the efficient optimization of raypaths. If discontinuities are included in the model, then it should be possible to smooth these values numerically to calculate gradients on different scales. I will also require that the resolution of the model should be adjustable at any time, even during optimization of the model. Adjustable resolution will allow the optimization

to converge first on the smoothest, most reliable components of the velocity model. As the accuracy of estimated raypaths improves, then more detail will be allowed in velocity models.

No convenient explicit equation exists to describe group velocity as a function of angle. Instead, I will use an approximation with enough degrees of freedom to explain the data well and still adequately span the same range of functions allowed by the exact theory. Tomography will have a limited ability to estimate arbitrary changes in velocity with angle. An approximate equation will have a form that is easy to optimize and yet describe the most important variations in velocity with angle. The errors introduced by analytic approximation are intended to be much smaller than errors introduced by inaccurate traveltimes. We might fit data first with more approximate curves and introduce refinements only when sensitivity improves.

The following equation parameterizes group velocities as a function of group angle ϕ, measured from the vertical axis of symmetry:

$$\begin{aligned} V(\phi)^{-1} &= V_x^{-1}\sqrt{1 + 2\eta\cos^2(\phi)\sin^2(\phi) + 2\epsilon\cos^2(\phi)} \\ &\approx V_x^{-1}[1 + \eta\cos^2(\phi)\sin^2(\phi) + \epsilon\cos^2(\phi)]. \end{aligned} \tag{1}$$

See the appendix for a fuller justification. This equation will have several advantages for tomography. The equation describes reciprocal velocity, or slowness, as an easily linearized function of three variables V_x^{-1}, η, ϵ. The parameters of anisotropy η and ϵ have small magnitudes, on the order of 0.05. The parameter ϵ controls an elliptical stretch, and η controls the bulge of anellipticity.

Slowness is integrated as a function of distance to give traveltimes. The horizontal velocity V_x is well defined by a physical experiment and is measured accurately from surface or crosswell experiments; the velocity V_z along the vertical axis of symmetry is much less well determined. The first parameter η in the square root is determined second best by surface experiments (Tsvankin and Thomsen, 1994; Alkhalifah and Tsvankin, 1994) because it specifies the difference between a normal moveout (NMO) velocity and V_x according to equation (A-3). Finally, the parameter ϵ expresses the third and least well determined part of anisotropy, giving the fractional change of vertical velocity according to (A-1).

The magnitude V_x will be allowed to change most arbitrarily, in many dimensions. Because this anisotropy is assumed to be a layered phenomenon, η will only be allowed to change vertically (or perpendicular to layering). Because ϵ is so poorly determined by surface data, I will assume that it has a single global value which can be chosen to fit well ties when available. Or one may attempt to predict ϵ from the other two parameters by observing the correlation in values that are produced by equivalent media calculations from well logs (Backus, 1962).

We can express the continuous velocity functions as a scaled sum of smooth basis functions. For example, if $\underset{\sim}{\mathbf{x}}$ is an arbitrary Cartesian coordinate, then we can express the slowness $V_x^{-1}(\underset{\sim}{\mathbf{x}})$ as a linear function of discrete parameters $\underset{\sim}{\mathbf{s}} = [s_i]$, using smooth

basis functions $S_i(\underset{\sim}{\mathbf{x}})$:

$$V_x^{-1}(\underset{\sim}{\mathbf{x}}) = \sum_i s_i S_i(\underset{\sim}{\mathbf{x}}) = \underset{\sim}{\mathbf{s}} \cdot \underset{\sim}{\mathbf{S}}(\underset{\sim}{\mathbf{x}}). \tag{2}$$

If $\hat{\underset{\sim}{\mathbf{z}}}$ is the unit vector in the direction of the vertical axis of symmetry, then we can express the function $\eta(\underset{\sim}{\mathbf{x}})$ as a linear function of scale factors $\underset{\sim}{\boldsymbol{\eta}} = [\eta_i]$ and smooth one-dimensional basis functions $E_i(z)$:

$$\eta(\underset{\sim}{\mathbf{x}}) = \sum_i \eta_i E_i(\underset{\sim}{\mathbf{x}} \cdot \hat{\underset{\sim}{\mathbf{z}}}) = \underset{\sim}{\boldsymbol{\eta}} \cdot \underset{\sim}{\mathbf{E}}(\underset{\sim}{\mathbf{x}} \cdot \hat{\underset{\sim}{\mathbf{z}}}). \tag{3}$$

The third parameter ϵ is already a single constant.

Let us designate this discrete set of velocity parameters as a single vector $\underset{\sim}{\mathbf{v}} = [\underset{\sim}{\mathbf{s}}, \underset{\sim}{\boldsymbol{\eta}}, \epsilon]$. The different elements of $\underset{\sim}{\mathbf{v}}$ are understood to have different scales, and will assume different variances during optimization. The approximate magnitude of V_x in (2), and thereby of $\underset{\sim}{\mathbf{s}}$, is easily anticipated from a quick glance at traveltimes over certain short distances. Theory (Backus, 1962) can easily anticipate reasonable magnitudes for η and ϵ for equivalent layered media.

Now we can write the continuous group velocity explicitly as a function of these discrete parameters

$$V(\underset{\sim}{\mathbf{x}}, \phi)^{-1} = [\underset{\sim}{\mathbf{s}} \cdot \underset{\sim}{\mathbf{S}}(\underset{\sim}{\mathbf{x}})][1 + \underset{\sim}{\boldsymbol{\eta}} \cdot \underset{\sim}{\mathbf{E}}(\underset{\sim}{\mathbf{x}} \cdot \hat{\underset{\sim}{\mathbf{z}}}) \cos^2(\phi) \sin^2(\phi) + \epsilon \cos^2(\phi)]. \tag{4}$$

A finite perturbation is easily linearized as

$$\begin{aligned} \Delta[V(\underset{\sim}{\mathbf{x}}, \phi)^{-1}] = \quad & [\Delta\underset{\sim}{\mathbf{s}} \cdot \underset{\sim}{\mathbf{S}}(\underset{\sim}{\mathbf{x}})][1 + \underset{\sim}{\boldsymbol{\eta}} \cdot \underset{\sim}{\mathbf{E}}(\underset{\sim}{\mathbf{x}} \cdot \hat{\underset{\sim}{\mathbf{z}}}) \cos^2(\phi) \sin^2(\phi) + \epsilon \cos^2(\phi)] \\ + \quad & [\underset{\sim}{\mathbf{s}} \cdot \underset{\sim}{\mathbf{S}}(\underset{\sim}{\mathbf{x}})][1 + \Delta\underset{\sim}{\boldsymbol{\eta}} \cdot \underset{\sim}{\mathbf{E}}(\underset{\sim}{\mathbf{x}} \cdot \hat{\underset{\sim}{\mathbf{z}}}) \cos^2(\phi) \sin^2(\phi) + \Delta\epsilon \cos^2(\phi)]. \end{aligned} \tag{5}$$

Unperturbed parameters in this equation are understood to remain at their reference values. Reference velocities will be iteratively updated and relinearized by a Gauss-Newton optimization algorithm described later.

PARAMETERIZATION OF PATHS

Raypaths will be parameterized as a sum of a small number of Chebyshev polynomials. Fermat's principle allows us to optimize these raypaths with only group velocities, and not phase velocities. The coefficients of the raypaths can be saved in little computer memory.

In an anisotropic medium, traveltime tomography needs to construct raypaths from group velocities. A raypath traces the path of wave energy and by definition always parallels the energy-flow vector, or Poynting vector. (See standard texts on electromagnetic theory.) Let us identify the Poynting vector with the group velocity vector, whose more restrictive definition applies only to a narrow band of frequencies. When the high-frequency ray approximation is valid, we can calculate the traveltime of a wave between two points by integrating the group velocity along the ray between the points.

Most shooting methods of dynamic ray tracing require alternate use of both phase and group velocities. Using both velocities would be more inconvenient for traveltime tomography, which needs only group velocities to calculate traveltime from a valid raypath. A perturbation of group velocities would have to be reconciled with perturbations of phase velocities.

By contrast a relaxation method of ray tracing needs only the group velocity. A raypath designates a stationary path according to Fermat's principle: perturbing any portion of the path should increase the integrated traveltime. A relaxation method initializes the raypath with parameters that allow few degrees of freedom. The initial path should be stiff and unable to make any sudden changes in direction because early iterations will also have large errors in the reference velocity model. Some stiffness in a raypath is also necessary to model accurately the path of wave energy at finite frequencies. Parameters of the path are perturbed and adjusted until the integrated traveltime is minimized. Because errors in the path are reduced, the path is allowed more degrees of freedom, and the new parameters are perturbed and adjusted further. Optimization ends when no more detail in the raypath can be justified for a measurable sensitivity to the resulting perturbations of traveltime.

Describe a ray then as a sum of smooth basis functions. Prothero et al (1988) described raypaths as sums of sine functions and optimized these functions with a simplex search, which avoids the need to calculate Frechet derivatives. Berryman (1990) used a similar raypath optimization for crosswell applications. However, Frechet derivatives do allow faster convergence.

More recently Grechka and McMechan (1995a; 1995b) have used Chebyshev polynomials. I find that these polynomials optimize shorter traveltimes for diving waves from fixed surface points than do sums of sines and semicircles. Sines have much flatter derivatives at the endpoints and cannot begin at such a steep angle as the polynomials. A semicircle appears to encourage *too* steep an angle at the endpoints.

Chebyshev polynomials $T_n(x)$ are defined over the range $-1 \leq x \leq 1$ by the recursion: $T_0(x) = 1$, $T_1(x) = x$, and $T_{n+2}(x) = 2xT_{n+1}(x) - T_n(x)$. As a minor modification, I prefer to shift the Chebyshev polynomials so that endpoints are fixed, and coefficients describe perturbations away from a straight line between endpoints. Define functions $C_n(x)$ over the range $0 \leq x \leq 1$ by

$$\begin{aligned} C_{2n}(x) &\equiv [1 - T_{2n+2}(2x-1)]/2, \text{and} \\ C_{2n+1}(x) &\equiv [2x - 1 - T_{2n+3}(2x-1)]/2, \end{aligned} \tag{6}$$

so that $C_0(x) = -4x^2 + 4x$ and $C_1(x) = -16x^3 + 24x^2 - 8x$, etc. Figure 1 shows the first 5 of these shifted Chebyshev polynomials.

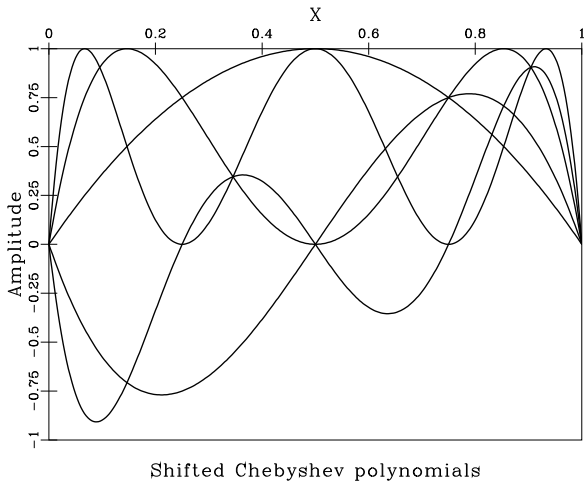

Figure 1: Shifted Chebyshev polynomials are used as basis functions for three-dimensional raypaths. fig1 [NR]

Let $\underset{\sim}{\mathbf{a}}$ and $\underset{\sim}{\mathbf{b}}$ be the Cartesian coordinates of the endpoints of the ray. I assume that these points lie along a surface that is roughly horizontal. Let $\hat{\underset{\sim}{\mathbf{z}}}$ be the unit vector pointing down the vertical axis of anisotropic symmetry. I specify the location $\underset{\sim}{\mathbf{x}}(r)$ along a raypath as a function of the variable r, which ranges from 0 to 1 between the endpoints:

$$\underset{\sim}{\mathbf{x}}(r) = \underset{\sim}{\mathbf{a}} + r(\underset{\sim}{\mathbf{b}} - \underset{\sim}{\mathbf{a}}) + [\sum_{i=0}^{n} \alpha_i C_i(r)]\hat{\underset{\sim}{\mathbf{z}}} + [\sum_{j=0}^{n} \beta_i C_j(r)][\hat{\underset{\sim}{\mathbf{z}}} \times (\underset{\sim}{\mathbf{b}} - \underset{\sim}{\mathbf{a}})]. \tag{7}$$

The parameters $[\alpha_i, \beta_j]$ scale a limited number of smooth curves that perturb the ray from a straight line between the endpoints. The $[\alpha_i]$ scale increasing orders of periods of polynomials in the vertical direction, and the $[\beta_j]$ scale in a perpendicular horizontal direction. The parameter r should be sampled more densely near the endpoints. I prefer the sampling $r_k = [1 - \cos(\pi k/N)]/2$ for $k = 0$ to N, which gives even sampling in distance along a semicircular path, to guarantee adequate sampling near possibly steep endpoints. This sampling also matches the density of zeros of the Chebyshev polynomials, which should be sampled more densely where they become more oscillatory.

A small number of coefficients may be adequate to describe a 3D raypath. Such a raypath is much easier to store in memory than an ordered series of samples from the raypath. The stiffness of the raypath may also be controlled to describe more accurately the path of energy with finite bandwidth. Detailed complications of a raypath should be avoided on a scale finer than the spatial wavelengths being modeled.

VELOCITY PERTURBATIONS

Here I show that traveltime data are easily linearized as a function of the parameters describing the velocity basis functions.

The functional form of an anisotropic group velocity must depend on the location and on the direction of a raypath. The vector $\partial \underset{\sim}{\mathbf{x}}(r)/\partial r$ points tangent to a given point on a raypath. Since ϕ is the angle of a raypath from the vertical axis of symmetry $\hat{\underset{\sim}{\mathbf{z}}}$, we can calculate the angle in the group velocity (4) as a function of the position along the raypath:

$$\phi(r) = \arccos\{[\partial \underset{\sim}{\mathbf{x}}(r)/\partial r \cdot \hat{\underset{\sim}{\mathbf{z}}}]/\|\partial \underset{\sim}{\mathbf{x}}(r)/\partial r\|_x\} \text{ with } \|\underset{\sim}{\mathbf{x}}\|_x^2 \equiv \underset{\sim}{\mathbf{x}} \cdot \underset{\sim}{\mathbf{x}}. \tag{8}$$

For this reason, we can write the parameterized group velocity $V[\underset{\sim}{\mathbf{x}}(r), \phi]$ in (4) as a function of the location and a tangent vector (with arbitrary magnitude).

A given raypath $\underset{\sim}{\mathbf{x}}(r)$, for r from 0 to 1, integrates for the traveltime

$$t = \int_0^1 V[\underset{\sim}{\mathbf{x}}(r), \phi(r)]^{-1} \|\partial \underset{\sim}{\mathbf{x}}(r)/\partial r\|_x dr. \tag{9}$$

Because the raypath represents a stationary minimum, tomography recognizes that perturbation of a valid raypath affects traveltime only to second order. To perturb traveltimes linearly with finite perturbations of slowness along a path, we need only integrate the slowness perturbations along the original path:

$$\Delta t = \int_0^1 \Delta\{V[\underset{\sim}{\mathbf{x}}(r), \phi(r)]^{-1}\} \|\partial \underset{\sim}{\mathbf{x}}(r)/\partial r\|_x dr. \tag{10}$$

The perturbation of slowness is given by the parameterization in (5). This formulation provides the linearized perturbation of the traveltime data as a function of perturbed velocity parameters. The adjoint uses the same weights for a backprojection of traveltime perturbations upon the velocity parameters.

OPTIMIZATION

For this application, I found it advantageous to write a generic "Gauss-Newton" optimization routine that minimizes a least-squares objective function with a nonlinear forward model. Both ray tracing and tomographic inversion of velocities are optimized with this algorithm. The ray parameters in equation (7) are perturbed until the traveltime is minimized. (The traveltime is a nonlinear function of local velocities.) The error between picked and modeled traveltimes (9) is minimized by perturbations of velocity parameters (4). In each case, a model damping term is included for numerical stability.

Let the vector $\underset{\sim}{\mathbf{m}}$ describe a model, and a vector $\underset{\sim}{\mathbf{d}}$ contain the data whose errors will be minimized. Define also a scalar product for each of these vectors: $< \underset{\sim}{\mathbf{m}}_1, \underset{\sim}{\mathbf{m}}_2 >_m$ and $< \underset{\sim}{\mathbf{d}}_1, \underset{\sim}{\mathbf{d}}_2 >_d$. The squared magnitude of each is defined by

$$\|\underset{\sim}{\mathbf{m}}\|_m^2 \equiv < \underset{\sim}{\mathbf{m}}, \underset{\sim}{\mathbf{m}} >_m \text{ and } \|\underset{\sim}{\mathbf{d}}\|_d^2 \equiv < \underset{\sim}{\mathbf{d}}, \underset{\sim}{\mathbf{d}} >_d . \tag{11}$$

These scalar products incorporate any non-stationary variances or covariances that can be assumed for the problem. For example, I will be assuming smaller variances for higher-order polynomials used to describe raypaths. The velocity parameters η and ϵ

will have small variances on the order of 0.05, and the velocity V_x will depend on the physical units of the survey. Rather than introduce correlations between samples into the scalar product, I prefer to encourage such correlations with the choice of basis functions. By scaling basis functions correctly, we can make the model norms become the trivial Cartesian norm, simply summing the squares of model parameters. I do not assume any correlation in the errors of traveltime data.

Assume we wish to fit the data $\mathbf{d}$ with a non-linear forward model $\mathbf{f}(\mathbf{m})$. We also must apply a linearized forward transform $\mathbf{F}(\mathbf{m}_0)$ for a given reference model $\mathbf{m}_0$, so that

$$\mathbf{f}(\mathbf{m}_0 + \Delta\mathbf{m}) \approx \mathbf{f}(\mathbf{m}_0) + \mathbf{F}(\mathbf{m}_0) \cdot \Delta\mathbf{m}. \tag{12}$$

We must be able to apply the transforms $\mathbf{f}(\mathbf{m}_0)$ and $\mathbf{F}(\mathbf{m}_0)$ when necessary and apply the adjoint $\mathbf{F}^*(\mathbf{m}_0)$ of the linear transform, defined by

$$< \mathbf{d}, \mathbf{F}(\mathbf{m}_0) \cdot \Delta\mathbf{m} >_d = < \mathbf{F}^*(\mathbf{m}_0) \cdot \mathbf{d}, \Delta\mathbf{m} >_m \ . \tag{13}$$

Let us assume that all optimum models $\mathbf{m}$ can then be specified to minimize an objective function of the form

$$\min_{\mathbf{m}} J_1(\mathbf{m}) = \|\mathbf{d} - \mathbf{f}(\mathbf{m})\|_d^2 + \|\mathbf{m} - \bar{\mathbf{m}}\|_m^2. \tag{14}$$

where $\bar{\mathbf{m}}$ contains the expected mean of the model. The relative weighting of the two terms ideally should be equal when covariances are included properly in the dot products. To optimize a raypath I minimize the traveltime. To optimize velocities I minimize the differences between measured and modeled traveltimes.

The objective function is iteratively approximated by a quadratic objective function, using the linearized forward model

$$\min_{\Delta\mathbf{m}} J_2(\Delta\mathbf{m}) = \|\mathbf{d} - \mathbf{f}(\mathbf{m}_0) - \mathbf{F}(\mathbf{m}_0) \cdot \Delta\mathbf{m}\|_d^2 + \|\mathbf{m}_0 + \Delta\mathbf{m} - \bar{\mathbf{m}}\|_m^2. \tag{15}$$

This quadratic objective function is easily optimized by a gradient method such as conjugate gradients. The gradient

$$\nabla_{\Delta\mathbf{m}} J_2(\Delta\mathbf{m}) = -\mathbf{F}^*(\mathbf{m}_0) \cdot [\mathbf{d} - \mathbf{f}(\mathbf{m}_0) - \mathbf{F}(\mathbf{m}_0) \cdot \Delta\mathbf{m}] + (\mathbf{m}_0 + \Delta\mathbf{m} - \bar{\mathbf{m}}) \tag{16}$$

requires application of the adjoint linearized transform. The resulting linearized perturbation is added to the reference model, after optimizing a scale factor λ by a line search:

$$\min_{\lambda} J_3(\lambda) = \|\mathbf{d} - \mathbf{f}(\mathbf{m}_0 + \lambda\Delta\mathbf{m})\|_d^2 + \|\mathbf{m}_0 + \lambda\Delta\mathbf{m} - \bar{\mathbf{m}}\|_m^2. \tag{17}$$

The reference model is updated by the scaled perturbation, the transform is relinearized, and the new quadratic (15) is optimized again, until convergence.

SYNTHETIC EXAMPLE

Scheduled tests begin with a simple 2D isotropic example, to test the convergence of the ray and velocity optimization. Non-uniqueness here will most likely remain with the additional complications of 3D and anisotropic models.

I define a 2D velocity model (in km/s) over the ranges $0 \leq x \leq 10$ (km) and $0 \leq z \leq 2$ (km), with

$$\begin{aligned} v(x,z) = 2 + z - 0.6 \quad & \exp \quad \{-\pi[(x-3)^2 + (z-1)^2]\} \\ + \quad & \exp \quad \{-\pi[(x-7)^2 + (z-1)^2]\}. \end{aligned} \tag{18}$$

This velocity model (Figure 2) has a background that increases linearly from 2 km/s to 4 km/s from $z = 0$ km to $z = 2$ km. Two smooth velocity anomalies, Gaussian with a width of 1, one slow and one fast, are placed at depths of $z = 1$ km, at $x = 3$ km and $x = 7$ km.

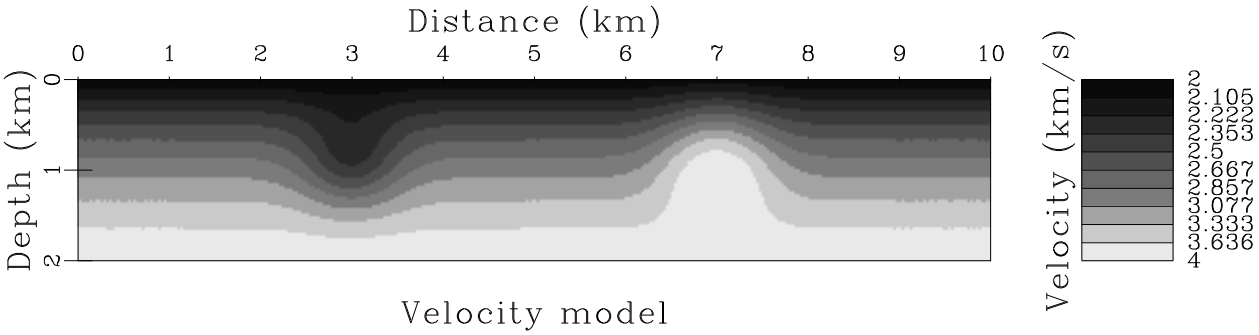

Figure 2: This is the assumed velocity model. fig2 [NR]

Sources and receivers are placed at 0.5 km intervals of x within $0 \leq x \leq 10$, with a minimum separation (offset) of 0.5 km and a maximum of 7 km, for a total of 189 data pairs. Figure 3 shows the resulting coverage of the velocity model by raypaths. If not for the velocity anomalies, a maximum offset of 7 km would produce a ray reaching 2 km depth. Rays through the faster part of the model cross at shallower depths. Rays also attempt to pass around the slow anomaly, leaving a small hole in coverage.

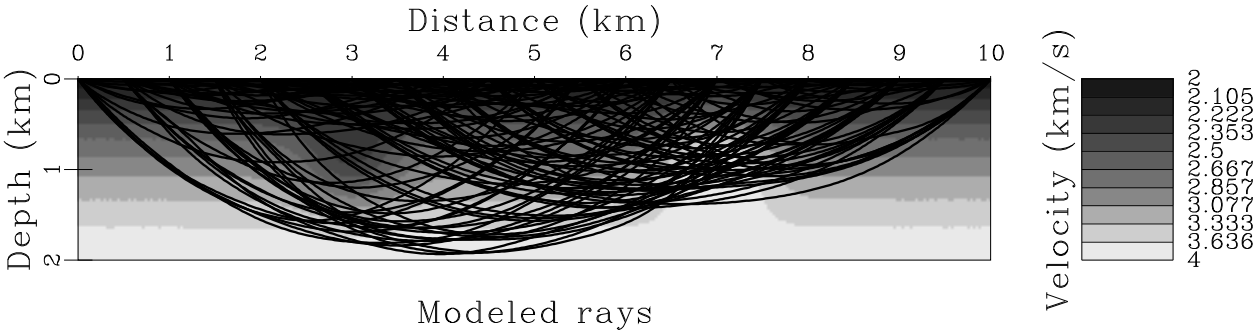

Figure 3: Rays are modeled from sources and receivers placed every 0.5 km, with a maximum offset of 7 km. fig3 [NR]

The most robust initialization of this velocity model would first optimize a velocity that changed only with depth. Only two parameters are necessary to describe a

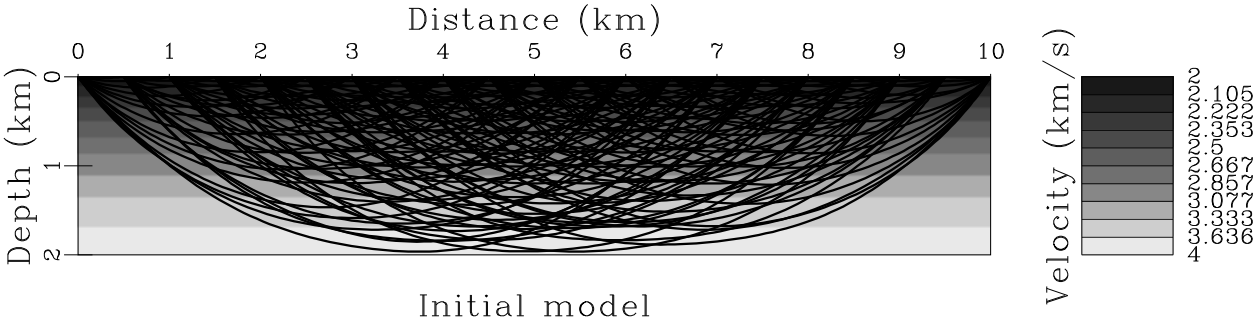

Figure 4: An initial model assumes velocity increases linearly from 2 km/s at the surface to 4 km/s at 2 km depth. fig4 [NR]

velocity that changes linearly with depth, so I initialize with the model in Figure 4: the velocity increases from 2 km/s at the surface to 4 km/s at 2 km depth. The rays are laterally invariant for a given offset. The ray coverage does not agree with that of the original model in Figure 2.

Eight large iterations were performed. In the first iteration isotropic velocities were described with overlapping Gaussian basis functions with a width of 1 km vertically and 4 km horizontally. During each large iteration, the raypaths were estimated once, then a linearized perturbation of velocities was estimated for these reference raypaths with a conjugate gradient algorithm. The perturbation was added to the reference velocities only after a line-search to find the proper scale factor. Rays were then reëstimated for the next large iteration and velocity perturbations were allowed to change more rapidly, with geometrically decreasing widths, until the final two iterations, where basis functions have widths of 0.1 km.

To control the rate of convergence for a general dataset, I always begin with velocity function that is smooth over most of the vertical and horizontal span. The final maximum resolution is limited by the density of rays and the spatial wavelengths. The widths of intermediate basis functions are reduced by a constant factor for each iteration. The final two iterations are at maximum resolution. Users are allowed to increase the total number of iterations for a more robust convergence.

After four iterations, the very smooth velocity model in Figure 5 has begun to show the velocity anomalies in outline. After two more iterations, the shape of the estimated anomalies in Figure 6 has approached the scale of the true anomalies, but still with flattened magnitudes. After eight large iterations, the velocity model in Figure 7 shows minor unreliable detail on the scale of the sampling of raypaths, but the anomalies do not much better resemble the original model. The raypaths in Figure 7 have converged to roughly the same coverage as in the original model in Figure 3.

The traveltimes plotted in Figure 9 were integrated from the raypaths in the original model in Figure 3. Traveltimes from the estimated raypaths in Figure 8 were subtracted from the correct traveltimes for the errors in Figure 10. These errors ranged from –6 ms to +10 ms, with a root-mean-square average of 2.3 ms. Those

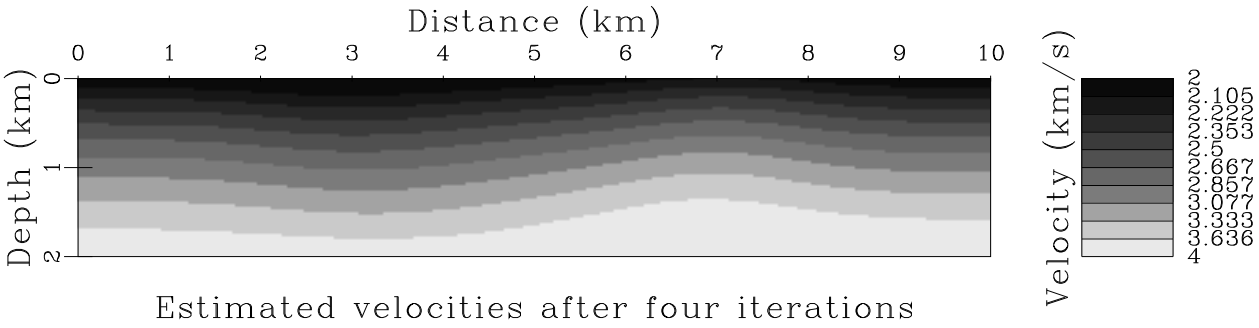

Figure 5: After four iterations, very smooth variations in interval velocity have been optimized. fig5 [NR]

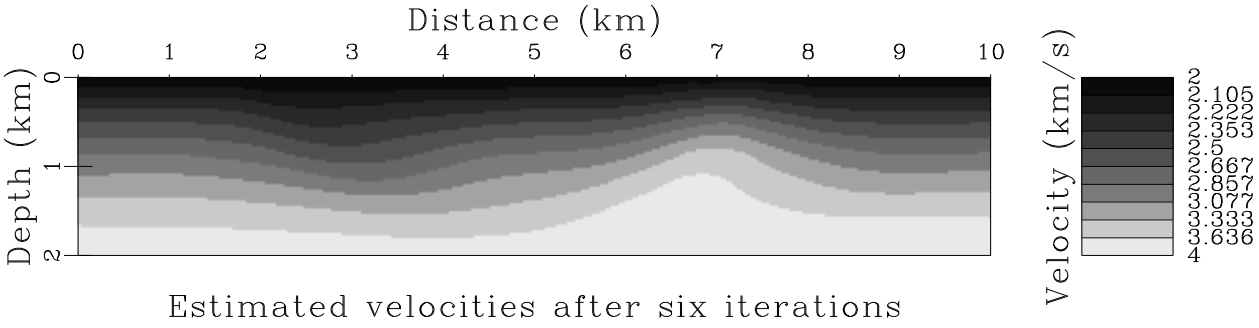

Figure 6: After six iterations, velocities are allowed to change more rapidly. fig6 [NR]

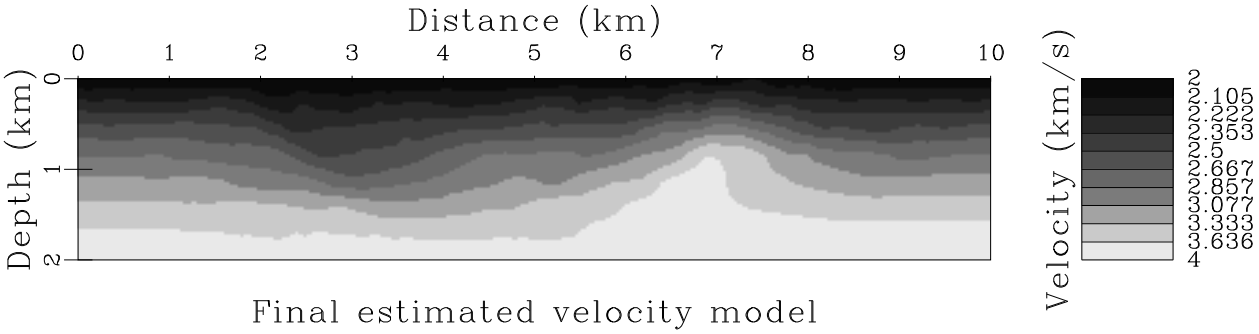

Figure 7: After eight iterations, velocities are allowed to change most arbitrarily. fig7 [NR]

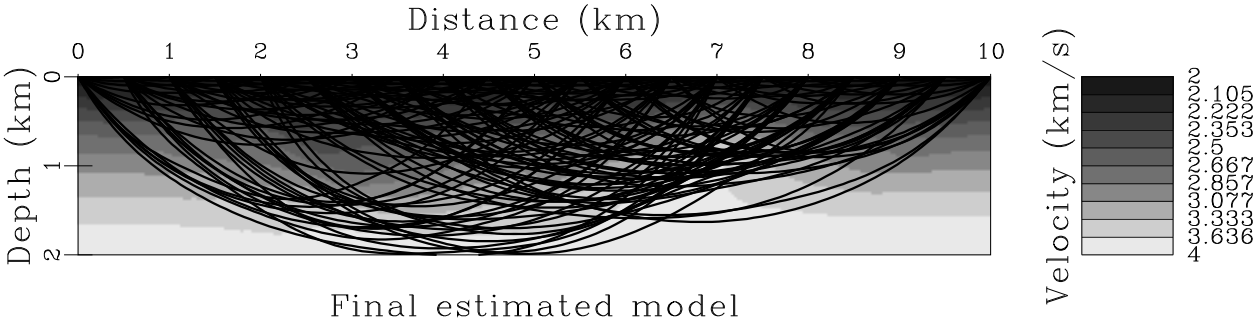

Figure 8: The final estimation of raypaths for the model in Figure 7 has converged to a coverage similar to that in Figure 3. fig8 [NR]

errors which remain do not appear to have any systematic correlation. Such time errors are much less than would be found at these scales in times picked from recorded surface seismic data, where a seismic temporal wavelength often exceeds 20 ms. Thus, the original and estimated velocity models in Figures 2 and 7 produce effectively indistinguishable traveltimes for this survey.

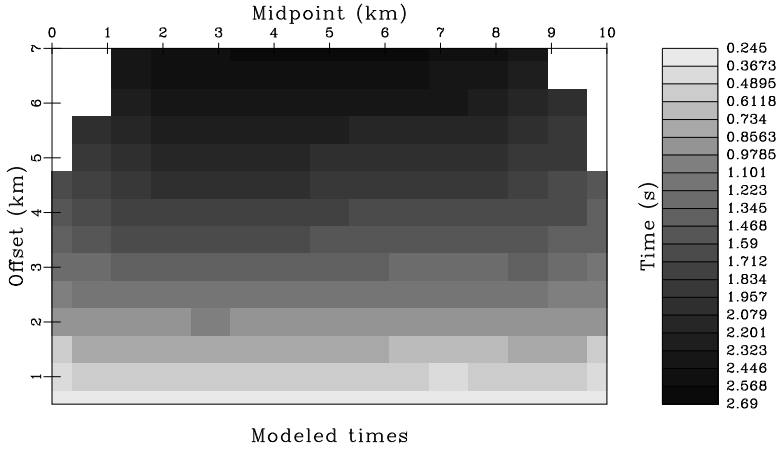

Figure 9: These traveltimes were modeled from the source and receiver positions in Figure 3. fig9 [NR]

Figure 11 shows the difference of the original velocities in Figure 2 minus the estimated velocities in Figure 7. The magnitudes of both anomalies has been underestimated. Moreover, sidelobes appear about the anomalies. The center of the estimated slow anomaly is too fast and the edges of the slow anomaly are too slow. The center of the estimated fast anomaly is too slow and the edges of the fast anomaly are too fast. Both anomalies have lost resolution in the reconstruction. The deeper portion of the fast anomaly was missed altogether because the original rays did not reach this depth.

The most important application of diving wave tomography is the correction of structural distortions in underlying reflections. To see the effect of errors in the velocity model, I calculate the vertical two-way traveltime from the surface to a

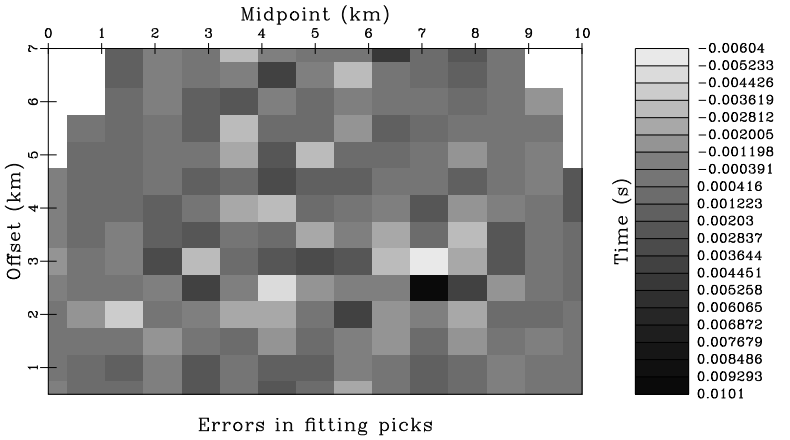

Figure 10: Traveltimes from the original model in Figure 2 minus traveltimes from the estimated model in Figures 7 and 8. The root-mean-square average of these errors was 2.3 ms, well below any possible accuracy in picking. fig10 [NR]

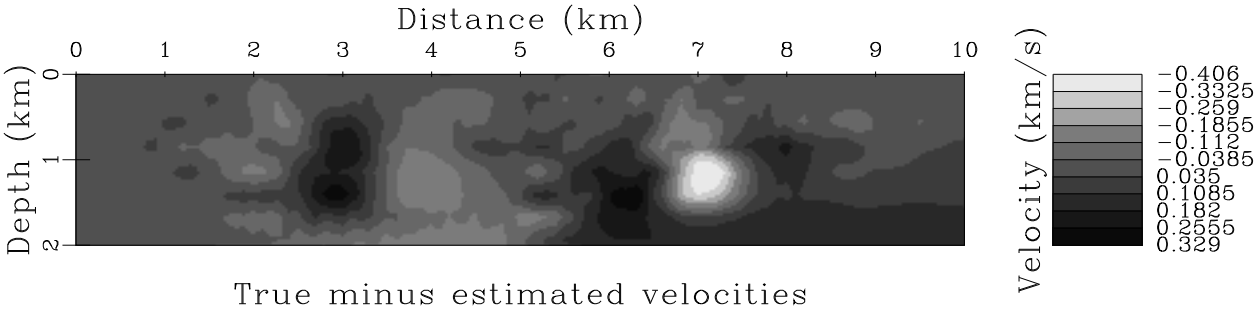

Figure 11: The original velocities in Figure 2 minus the estimated velocities in figure 7. fig11 [NR]

horizontal reflector at 2 km depth, for all surface positions. The true velocity model gives two distinct peaks in traveltime in Figure 12, but the estimated velocity model gives much smoother peaks. The positioning of peaks is correct, but lost resolution has the effect of reducing the magnitude of the relative changes in vertical time.

Smearing of the velocity anomalies clearly occurs parallel to the raypaths. Resolution is greatest perpendicular to the raypaths. The angular coverage of a particular anomaly is limited. Broader velocities anomalies are easier to invert than this example, but we must acknowledge this loss of resolution will affect the accuracy of our near-surface velocities. Other information, such as known shallow faults, might allow us to identify and introduce sharper edges on lateral velocity changes.

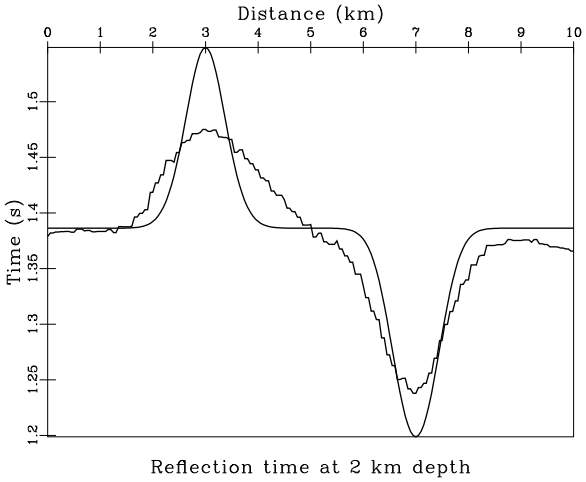

Figure 12: Vertical two-way times integrated from the surface down to a flat reflector at 2 km depth, for the original model in Figure 2 and for the estimated model in Figure 7. The estimated curve has broader and shallower peaks (time sags) than the actual curve. fig12 [NR]

FUTURE WORK

The description of anisotropy velocities and raypaths as sums of smooth basis was relatively simple to code, and a generic Gauss-Newton optimization algorithm greatly simplified their optimization. The non-uniqueness of the first synthetic inversion is not surprising and will not prevent us from making use of well-determined components of velocities.

The 2D isotropic synthetic converges to the smoothest and most stratified velocity model that will adequately model the picked traveltimes. 3D and anisotropic models could easily increase the non-uniqueness of the inversion. Next, I must test 3D isotropic models to discover what patterns of 3D coverage can isolate velocity anomalies. It would be very useful to know if a typical 3D marine acquisition is adequate, with offsets extending only along one azimuth.

I must test 2D and 3D anisotropic data for non-uniqueness. An isotropic inversion of anisotropically calculated times will demonstrate whether such data can

be explained isotropically. With crosswell surveys many have seen that a simple anisotropic model explains the same traveltimes as an isotropic model with much more complicated spatial variations in velocity. Possibly an analogous situation will occur with surface diving waves. I must then make choices on whether to discourage spatial variations or anisotropy more, while allowing for non-uniqueness in both components.

ACKNOWLEDGMENTS

Greg Lazear and Phil Anno of Conoco helped considerably with a year of discussions about anisotropy. Sergey Fomel introduced me to the work of Vladimir Grechka (1995a; 1995b), who has graciously allowed me to see his papers before publication. Thanks to Sean Crawley for his continuing interest.

REFERENCES

Alkhalifah, T. A., and Tsvankin, I., 1994, Velocity analysis for transversely isotropic media: 64th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1000–1003.

Backus, G. E., 1962, Long-wave elastic anisotropy produced by horizontal layering: Journal of Geophysical Research, **67**, no. 11, 4427–4440.

Bell, M. L., Lara, R., and Gray, W. C., 1994, Application of turning-ray tomography to the offshore Mississippi Delta: 64th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1509–1512.

Berryman, J., 1990, Stable iterative reconstruction algorithm for nonlinear traveltime tomography: Inverse Problems, **6**, 21–42.

Cai, W., and Qin, F., 1994, Three-dimensional refraction imaging: 64th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 329–332.

Clayton, R. W., and McMechan, G. A., 1981, Inversion of refraction data by wavefield continuation: Geophysics, **46**, no. 6, 860–868.

Grechka, V., and McMechan, G., 1995a, Fast 3d two-point ray tracing for anisotropic media: submitted to Geophysics.

Grechka, V., and McMechan, G., 1995b, Singular value decomposition for 3d anisotropic traveltime tomography: submitted to Geophysics.

Hagedoorn, J. G., 1964, The elusive first arrival: Geophysics, **29**, no. 5, 806–813.

Harlan, W. S., Hanson, D. W., and Boyd, M., 1991a, Common-offset depth migrations and traveltime tomography: 61st Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 971–973.

Harlan, W. S., Hanson, D. W., and Boyd, M., 1991b, Traveltime tomography and multioffset common-reflecton points: 61st Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 974–976.

Harlan, W. S., 1990, Tomographic estimation of shear velocities from shallow cross-well seismic data: 60th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 86–89.

Harlan, W. S., 1992, Robust strategies for reflection and cross-well traveltime tomography: 54th Mtg. Eur. Assoc. Expl Geophys., Abstracts, 250–251.

Hawkins, L. V., 1961, The reciprocal method of routine shallow seismic refraction investigations: Geophysics, **26**, no. 6, 806–819.

Hole, J., Clowes, R., and Ellis, R., 1992, Interface inversion using broadside seismic refraction data and three-dimensional travel time calculations: J. Geophys. Res., **97**, no. B3, 3417–3429.

Hole, J., 1992, Nonlinear high-resolution three-dimensional seismic trave ltime tomography: J. Geophys. Res., **97**, no. B5, 6553–6562.

Landa, E., Keydar, S., and Kravtsov, A., 1994, Determination of shallow velocity-depth model from seismic refraction data by coherency inversion: 56th Mtg., Eur. Assoc. Expl. Geophys., Extended Abstracts, Session:H023.

Laski, J. D., 1978, Transformation of time-distance curves of diving waves into a multilayered model: Geophys. Prosp., **26**, no. 1, 48–81.

Levin, F. K., 1994, The anatomy of diving waves: 64th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1561–1563.

Michelena, R. J., Muir, F., and Harris, J. M., 1993, Anisotropic traveltime tomography: Geophys. Prosp., **41**, no. 4, 381–412.

Michelena, R. J., 1994, Elastic constants of transversely isotropic media from constrained aperture traveltimes: Geophysics, **59**, no. 4, 658–667.

Palmer, D., 1981, An introduction to the generalized reciprocal method of seismic refraction interpretation: Geophysics, **46**, no. 11, 1508–1518.

Pratt, R. G., and McGaughey, W. J., 1991, Velocity and anisotropy tomograms of a crown pillar: 61st Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1616–1621.

Prothero, W., Taylor, W., and Eickemeyer, J., 1988, A fast, two-point, three-dimensional raytracing algorithm using a simple step search method: Bull. Seismol. Soc. Am., **78**, no. 3, 1190–1198.

Qin, F., Cai, W., and Schuster, G. T., 1993, Inversion and imaging of refraction data: 63rd Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 613–615.

Saito, H., 1991, Anisotropic traveltime tomography at the Buckhorn Test Facility in Illinois: 61st Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 123–126.

Schoenberg, M., and Muir, F., 1989, A calculus for finely layered anisotropic media: Geophysics, **54**, no. 5, 581–589.

Simmons, Jr., J. L., and Backus, M. M., 1988, First arrival times and diving-wave tomography: 58th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, Session:S21.5.

Simmons, Jr., J. L., and Backus, M. M., 1992, Linearized tomographic inversion of first-arrival times: Geophysics, **57**, no. 11, 1482–1492.

Simmons, Jr., J. L., and Bernitsas, N., 1994, Nonlinear inversion of first-arrival times: 64th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 992–995.

Stefani, J. P., 1993, Possibilities and limitations of turning ray tomography—a synthetics study: 63rd Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 610–612.

Thomsen, L., 1986, Weak elastic anisotropy: Geophysics, **51**, no. 10, 1954–1966.

Tsvankin, I., and Thomsen, L., 1994, Nonhyperbolic reflection moveout in anisotropic media: Geophysics, **59**, no. 8, 1290–1304.

Vassiliou, A. A., Savage, C. W., Liner, C. L., Bozhurt, G., and Lines, L. R., 1994, Glenn pool project: Initial tomographic results: 64th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 302.

Zanzi, L., and Carlini, A., 1991, Refraction statics in the wavenumber domain: Geophysics, **56**, no. 10, 1661–1670.

Zhang, J., and McMechan, G. A., 1994, 3-D transmission tomography using wide aperture data for velocity estimation of irregular salt bodies: Geophysics, **59**, no. 10, 1620–1630.

Zhu, X., Sixta, D. P., and Angstman, B. G., 1992a, Tomostatics: Turning-ray tomography + static corrections: 62nd Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1108–1111.

Zhu, X., Sixta, D. P., and Angstman, B. G., 1992b, Tomostatics: Turning-ray tomography + static corrections: The Leading Edge, **11**, no. 12, 15–23.

APPENDIX A

PARAMETERIZATION OF ANISOTROPIC GROUP VELOCITIES

Assume that anisotropic velocities have a vertical axis of symmetry, like the transversely isotropic (TI) media described in Thomsen (1986). Although that paper describes "weak" anisotropy, the same equations can be applied to very strong anisotropy (Tsvankin and Thomsen, 1994).

Three of Thomsen's parameters, V_z, δ, and ϵ, are defined by the elastic constants of a general TI medium. These constants can be used to specify three different effective velocities at a single point in the model. V_z is the velocity of a wave traveling vertically along the axis of symmetry. The velocity in any horizontal direction is V_x determined by

$$\epsilon = V_x^2(V_z^{-2} - V_x^{-2})/2 > 0 \qquad (A-1)$$

and a "normal moveout velocity" (NMO) velocity V_n defined by

$$\delta \equiv V_n^2(V_z^{-2} - V_n^{-2})/2 < 0. \qquad (A-2)$$

Phil Anno of Conoco has shown that if the TI properties represent the equivalent medium of many isotropic layers (Backus, 1962; Schoenberg and Muir, 1989), then the above inequalities can be expected to hold. (One additional assumption is that the V_s/V_p ratio and V_s have a positive correlation.) Notice that $V_x \approx (1+\epsilon)V_z$ and $V_n \approx (1+\delta)V_z$, so that $V_n \leq V_z \leq V_x$.

For convenience, researchers at the Colorado School of Mines (Tsvankin and Thomsen, 1994; Alkhalifah and Tsvankin, 1994) have also defined a constant

$$\eta \equiv (\epsilon - \delta)/(1+2\delta) = V_x^2(V_n^{-2} - V_x^{-2})/2 > 0. \qquad (A-3)$$

Many combinations of three of these parameters can be used to describe a TI medium. An approximation has already dropped a fourth constant to which compressional P waves are very insensitive. The exact equations for TI phase velocity as a function of angle are rather clumsy, and *no explicit form* is available for group velocity. Alternative approximate equations can used which fit almost the same family of curves as the original correct equations (Michelena et al., 1993). I use an approximate equation for group velocity which appears to emulate closely the exact curves for large ranges of positive ϵ and negative δ. Since I aim to estimate these anisotropic velocities from noisy measurements, I expect our estimated curves to have larger errors than introduced by these approximations.

I choose approximate curves with the three velocities defined above. Let ϕ be the group angle of a raypath from the vertical. Then the group velocity $V(\phi)$ can be expressed as

$$V(\phi)^{-2} = V_z^{-2}\cos^2(\phi) + (V_n^{-2} - V_x^{-2})\cos^2(\phi)\sin^2(\phi) + V_x^{-2}\sin^2(\phi). \qquad (A-4)$$

Greg Lazear of Conoco found that a good approximation of the phase velocity $v(\theta)$ as a function of the phase angle θ takes a similar form, but with reciprocals of velocities:

$$v(\theta)^2 = V_z^2 \cos^2(\theta) + (V_n^2 - V_x^2) \cos^2(\theta) \sin^2(\theta) + V_x^2 \sin^2(\theta). \qquad (A-5)$$

The NMO velocity also turns out to have a physical interpretation. Imagine an experiment on a homogeneous and anisotropic medium (or imagine a small scale experiment on a smooth model). Measure the traveltime t_0 between two points placed on a vertical line, separated by a vertical distance $V_z t_0$. Now displace the upper point a distance h along a horizontal line and measure the new traveltime t_h.

Then according to equation (A-4) the traveltime t_h as a function of offset h is exactly

$$t_h^2 = t_0^2 + \left[V_n^{-2} + (V_x^{-2} - V_n^{-2}) \frac{h^2}{h^2 + V_z^2 t_0^2} \right] h^2. \qquad (A-6)$$

When $h \ll V_z t_0$ then the value of t_h in this "moveout equation" is controlled by the NMO velocity V_n rather than V_x. In the other case $h \gg V_z t_0$, the raypath is almost horizontal and V_x dominates.

I find it convenient to define a stacking velocity $V_h(h)$ as a function of the offset h for a fixed vertical distance:

$$V_h(h)^{-2} \equiv (t_h^2 - t_0^2)/h^2 = V_n^{-2} + (V_x^{-2} - V_n^{-2}) \frac{h^2}{h^2 + V_z^2 t_0^2}. \qquad (A-7)$$

Thus, we can construct a equation which describes the best fitting hyperbola to traveltimes at zero offset and at a single finite offset h:

$$t^2 = t_0^2 + h^2 / V_h(h)^2. \qquad (A-8)$$

Note that this stacking velocity covers the range $V_n \le V_h(h) \le V_x$, increasing in value as h increases. (To use two-way reflection times in A-8 we need only replace the half offset h by the full offset.)

Theoretically, three measurements of traveltimes at three different offsets h should uniquely determine the three velocity constants V_z, V_x, V_n. However, the traveltimes are much more sensitive to V_n, which determines moveouts at small offsets, and to V_x, which determines moveout at larger offsets. The vertical velocity V_z affects only the rate at which the stacking velocity changes from one limit to the other. As long as V_z has roughly the correct magnitude, then we can fit all measured traveltimes very well. Remember that we expect $V_n \le V_z \le V_x$ for equivalent layered media.

For imaging data in time, we can set $V_z = V_n$ and simplify our equations even further. To image surface data in depth, we can focus images very well with good values for V_x and V_n, then adjust imaged depths to tie wells with V_z, holding the other two velocities constant.

Stacking velocity analysis can be optimized with V_n because it is relatively close to V_h. If the maximum offset equals the depth, and if $V_x = 1.1 V_n$, then $V_h = 1.02 V_n$ for a flat reflection, which is small enough difference for such a large anisotropy.

This anisotropy model, although certainly not the most general, describes the most important properties of transversely isotropic velocities. The very simple form allows for easy optimization and inversion. A tomographic algorithm which builds on such a model will include the necessary dependence of velocities on angle. Any refinements in the anisotropic behavior will be easy to introduce without major alterations of the computer program.