

Seismic ray tracing and wavefront tracking in laterally heterogeneous media

N. Rawlinson^{*}, J. Hauser, M. Sambridge

Research School of Earth Sciences, Australian National University, Canberra ACT

0200, Australia

1 Introduction

1.1 Motivation

One of the most common and challenging problems in seismology is the prediction of source-receiver paths taken by seismic energy in the presence of lateral variations in wavespeed. The solution to this problem is required in many applications that exploit the high frequency component of seismic records, such as body wave tomography, migration of reflection data and earthquake relocation. The process of tracking the kinematic evolution of seismic energy also brings with it the possibility of computing various other wave-related quantities such as traveltime, amplitude, attenuation, or even the high frequency waveform, which can then be compared to observations.

The difficulties associated with locating a two point path arise from the non-linear relationship between velocity and path geometry. Fig. 1, which shows a fan of ray paths

^{*} Corresponding author.

Email addresses: `Nicholas.Rawlinson@anu.edu.au` (N. Rawlinson),
`jurg.hauser@anu.edu.au` (J. Hauser), `Malcolm.Sambridge@anu.edu.au` (M. Sambridge).

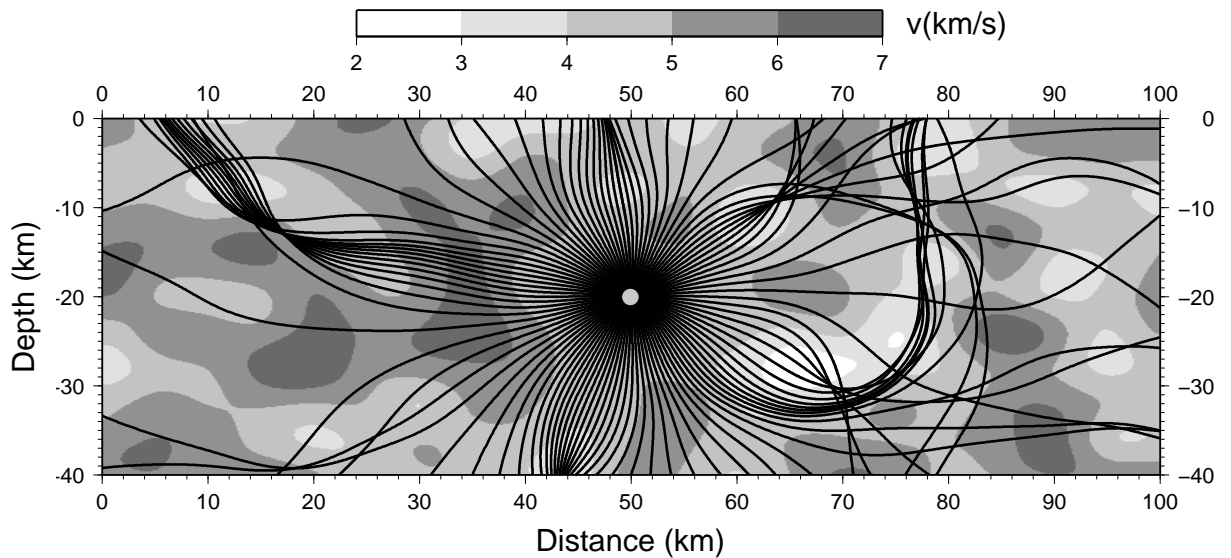


Fig. 1. Trajectories followed by a uniform fan of 100 rays emitted by a source point (grey dot) in a smoothly varying heterogeneous medium. Although the angular distance between all adjacent paths at the source point is identical, this relationship is not preserved as the rays track through the medium.

propagate from a point source in a strongly heterogeneous medium, provides useful insight into this non-linearity. If the medium had been homogeneous, then the paths would simply have been straight lines emitting at uniform angular distance from the source. However, the focusing and defocusing effects of the velocity heterogeneity have imposed strong and varying curvature to the paths. In addition to the extremely non-linear distribution of rays along the boundaries of the medium, the phenomenon of multi-pathing, which equates to wavefront self-intersection, can also be observed. Thus, the already difficult problem of locating a valid path which connects two points has been further complicated by the fact that there may well be more than one path.

Over the past few decades, the growing need for fast and accurate prediction of high frequency wave properties (most commonly traveltimes) in complex 2-D and 3-D media has spawned a prolific number of grid and ray based solvers. Traditionally, the method of choice has been ray tracing (Julian and Gubbins, 1977; Červený, 1987; Virieux and Farra, 1991; Červený, 2001), in which the trajectory of paths corresponding to wavefront normals are computed between two points. This approach is often highly accurate and

efficient, and naturally lends itself to the prediction of various seismic wave properties. However, it is non-robust, and may fail to converge to a true two-point path even in mildly heterogeneous media. In addition, it usually provides no guarantee as to whether a located path corresponds to a first or later arrival.

Grid based schemes, which usually involve the calculation of traveltimes to all points of a regular grid that spans the velocity medium, have become increasingly popular in recent times. They are often based on finite difference solution of the eikonal equation (Vidale, 1988; Qin et al., 1992; Hole and Zelt, 1995; Kim and Cook, 1999; Popovici and Sethian, 2002; Rawlinson and Sambridge, 2004a) or shortest path (network) methods (Nakanishi and Yamaguchi, 1986; Moser, 1991; Cheng and House, 1996), both of which tend to be computationally efficient and highly robust, a combination which makes them viable alternatives to ray tracing. Wavefronts and rays can be obtained *a posteriori* if required by either contouring the traveltime field or following the traveltime gradient from receiver to source, respectively. Disadvantages of these schemes include that in most cases they only compute the first-arrival traveltime, and their accuracy is generally not as high as ray tracing. In addition, it is difficult to compute quantities other than traveltime without first extracting ray paths.

The aim of this review paper is to describe a variety of ray and grid based solvers for locating ray paths, or implicitly or explicitly tracking wavefronts, in laterally heterogeneous media. The profusion of different schemes that can be found in the literature means that it is not possible to provide a comprehensive review; instead, we focus on methods that have been successful in practical applications. In addition, we only look at the kinematic component of the problem i.e. ray trajectory and wavefront evolution, rather than the dynamic component, which is required for computing various quantities such as amplitudes or waveforms. After providing some introductory background material on asymptotic ray theory and geometric optics, the review proceeds with a description of various shooting and bending methods of ray tracing. This is followed by eikonal solvers and shortest path

methods.

In recent years, there has been an increase in the development of new methods aimed at tracking all arrivals in heterogeneous media. These schemes can be used to predict more of the seismic wavefield when multi-pathing of seismic energy results in more complex wave-trains. One possible use of these schemes is in seismic tomography, where the exploitation of multi-arrivals may result in improved images. A detailed description of several recently developed grid and ray based schemes for predicting multi-arrivals is included in Section 4.

1.2 The eikonal equation

All ray and grid based methods we consider in this review are subject to the so-called “high frequency approximation”; that is, the wavelength of the propagating wave is substantially shorter than the seismic heterogeneities that characterise the medium through which they pass. Under this assumption, the full elastic wave equation can be greatly simplified, and the problem of computing the seismic wavefield made much more tractable. For a seismic P-wave in an isotropic medium, the elastic wave equation can be written (Chapman, 2004)

$$\nabla^2\Phi - \frac{1}{\alpha^2} \frac{\partial^2\Phi}{\partial t^2} = 0, \quad (1)$$

where Φ represents the scalar potential of a P-wave, α is P-wavespeed and t is time. If we assume that the solution of Eq. 1 has the general form

$$\Phi = A \exp[-i\omega(T(\mathbf{x}) + t)],$$

where $A = A(\mathbf{x})$ is amplitude, ω is angular frequency and T is a surface of constant phase, then the Laplacian of the scalar potential is

$$\begin{aligned} \nabla^2\Phi = & \nabla^2 A \exp[-i\omega(T + t)] - i\omega \nabla T \cdot \nabla A \exp[-i\omega(T + t)] \\ & - i\omega \nabla A \cdot \nabla T \exp[-i\omega(T + t)] - i\omega A \nabla^2 T \exp[-i\omega(T + t)] \\ & - \omega^2 A \nabla T \cdot \nabla T \exp[-i\omega(T + t)] \end{aligned}$$

and the second derivative of Φ with respect to time is

$$\frac{\partial^2 \Phi}{\partial t^2} = -\omega^2 A \exp[-i\omega(T + t)].$$

Substitution of the above two expressions into Eq. 1 yields

$$\nabla^2 A - \omega^2 A |\nabla T|^2 - i[2\omega \nabla A \cdot \nabla T + \omega A \nabla^2 T] = \frac{-A\omega^2}{\alpha^2}, \quad (2)$$

which can be divided into real and imaginary parts. If we take the real part and divide through by $A\omega^2$, then

$$\frac{\nabla^2 A}{A\omega^2} - |\nabla T|^2 = \frac{-1}{\alpha^2}.$$

Application of the high frequency approximation ($\omega \rightarrow \infty$) then yields the **eikonal equation**

$$|\nabla T| = s, \quad (3)$$

where $s = 1/\alpha$ is slowness. $T(\mathbf{x})$ is a time function (the eikonal) which describes surfaces of constant phase (wavefronts) when T is constant. If we now take the imaginary part of Eq. 2 and divide through by ω , we obtain the **transport equation**

$$2\nabla A \cdot \nabla T + A \nabla^2 T = 0 \quad (4)$$

which can be used to compute the amplitude of the propagating wave. Substitution of the appropriate general S-wave vector potential into the elastic wave equation for an S-wave leads to identical expressions for the eikonal and transport equations; thus, Eq. 3 and 4 are valid for any high frequency body wave with slowness s . In fully anisotropic media (Červený, 2001), the eikonal and transport equations have a slightly more complex form due to the presence of the elastic tensor \mathbf{c} .

1.3 The kinematic ray tracing equations

Rather than directly solve the eikonal equation, one can instead consider its characteristics, which are the trajectories orthogonal (in isotropic media) to the wavefront. If \mathbf{r}

represents the position vector of a point on a wavefront, and l the pathlength of the curve traced out by this point as the wavefront evolves (see Fig. 2), then

$$\frac{d\mathbf{r}}{dl} = \frac{\nabla T}{s} \quad (5)$$

since both $d\mathbf{r}/dl$ and $\nabla T/s$ are unit vectors parallel to the path. The rate of change of traveltime along the path is simply the slowness, so

$$\frac{dT}{dl} = s \quad (6)$$

and by taking the gradient of both sides (noting the commutation of d/dl and ∇)

$$\frac{d\nabla T}{dl} = \nabla s. \quad (7)$$

Eq. 5 and 7 can be combined in order to remove ∇T which gives

$$\frac{d}{dl} \left[s \frac{d\mathbf{r}}{dl} \right] = \nabla s. \quad (8)$$

Eq. 8 is the **kinematic ray equation** and describes the trajectory of ray paths in smoothly varying isotropic media. It will be shown later how Eq. 8 can be reduced to forms suitable for initial and boundary value ray tracing. The kinematic ray equation can also be derived using the calculus of variations, because Fermat's principle of stationary time states that ray paths correspond to extremal curves of the integral

$$T = \int_L s dl \quad (9)$$

where L represents the path. In this case, Eq. 8 turns out to be the corresponding Euler-Lagrange equation.

In the presence of wavespeed discontinuities, Eq. 8 cannot be used because ∇s is not defined. Instead, Snell's law can be applied, which in its simplest form can be expressed:

$$\frac{\sin \theta_i}{v_i} = \frac{\sin \theta_o}{v_o} \quad (10)$$

where θ_i and v_i are the angle and wavespeed of the incoming ray, and θ_o and v_o are the angle and wavespeed of the outgoing ray. For reflected rays, $v_1 = v_2$ so $\theta_1 = \theta_2$, which

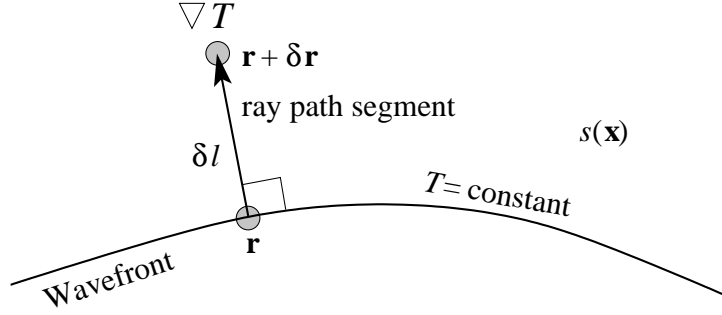


Fig. 2. Variables used to describe wavefronts and rays. T is traveltime, \mathbf{r} is the position vector of a point on the wavefront, l is ray path length and $s(\mathbf{x})$ is slowness.

in general will not be the case for transmitted rays provided a wavespeed discontinuity exists.

The derivation of Eq. 8 is relatively straightforward, due largely to the assumption of isotropic media. However, if we also wish to include anisotropy, then a more general approach is required. A commonly used treatment in this case is the Hamiltonian formalism of classical mechanics (Chapman, 2004; Červený, 2001). In this case, rays are equivalent to the characteristic curves of the Hamiltonian, which may be expressed in various ways. In isotropic media, the Hamiltonian is often written (Chapman, 2004)

$$H(\mathbf{x}, \mathbf{p}) = \frac{\mathbf{p}^2}{2s^2} \quad (11)$$

or (Virieux and Farra, 1991)

$$H(\mathbf{x}, \mathbf{p}) = \frac{1}{2}[\mathbf{p}^2 - s^2], \quad (12)$$

where $\mathbf{p} = \nabla T$. Setting $H = 1/2$ in Eq. 11 or $H = 0$ in Eq. 12 results in characteristic curves which satisfy the eikonal equation. The Hamilton equations, which describe the characteristic curves of the Hamiltonian, can be written (Chapman, 2004)

$$\frac{d\mathbf{x}}{dt} = \frac{\partial H}{\partial \mathbf{p}} \quad \text{and} \quad \frac{d\mathbf{p}}{dt} = -\nabla H \quad (13)$$

which is a coupled system of six ordinary differential equations. These equations can be integrated forward in time from given initial conditions using standard numerical solvers such as the Runge Kutta method (e.g. Kreyszig, 1993). The Hamilton equations need not be written in the form of Eq. 13 with time as the independent variable; for example, one

could also use path length (see Červený, 2001, for more details).

In anisotropic media, Eq. 13 remains valid, but an alternative form of the Hamiltonian, which takes into account the 21 independent elastic parameters $c_{i,j,k,l}$ (where $i, j, k, l = 1, \dots, 3$) and density ρ , is required. The presence of anisotropy means that we can no longer treat P- and S-waves as equivalent, with only their propagation speeds being different. Instead, there will be three distinct wave types, a quasi-compressional wave qP, and two quasi shear waves qS1 and qS2. It turns out that the behaviour of these waves can be described by finding the eigenvectors and eigenvalues corresponding to the solution of the equation (see Červený and Firbas, 1984):

$$(\mathbf{\Gamma} - \mathbf{I})\mathbf{w} = \mathbf{0} \quad (14)$$

where the 3×3 matrix $\mathbf{\Gamma} = \{\Gamma_{jk}\} = p_i p_l \mathbf{c}_{il} / \rho$ (note implied summation over i and l) is the so-called Christoffel matrix and \mathbf{w} is the three component displacement vector. Eq. 14 can be derived by seeking the asymptotic solution to the full elastic wave equation. The eigenvalues G_m ($m = 1, \dots, 3$) satisfy:

$$\det[\mathbf{\Gamma} - \mathbf{I}G_m] = 0 \quad (15)$$

and the eigenvectors \mathbf{g}_m satisfy:

$$[\mathbf{\Gamma} - \mathbf{I}G_m]\mathbf{g}_m = \mathbf{0} \quad (16)$$

noting that m is not used as an implied summation variable. Eq. 14 is satisfied provided any of the eigenvalues $G_m = 1$ ($m = 1, \dots, 3$). Each G_m corresponds to the eikonal equation for a different wave type, and the associated eigenvectors \mathbf{g}_m describe the direction of particle motion imposed by the wave. It turns out (Červený and Firbas, 1984) that the eigenvalues can be written in terms of the eigenvectors as

$$G_m = \mathbf{g}_m^T \mathbf{\Gamma} \mathbf{g}_m = \frac{p_i p_l \mathbf{g}_m^T \mathbf{c}_{il} \mathbf{g}_m}{\rho}. \quad (17)$$

In isotropic media, $G_1 = \alpha^2 \mathbf{p} \cdot \mathbf{p}$ and $G_2 = \beta^2 \mathbf{p} \cdot \mathbf{p}$, where α and β are the P and S

wavespeeds respectively.

For general anisotropic inhomogeneous media, the Hamiltonian can therefore be written as

$$H_m(\mathbf{x}, \mathbf{p}) = \frac{1}{2}G_m(\mathbf{x}, \mathbf{p}) = \frac{1}{2}\mathbf{g}_m^T \mathbf{\Gamma} \mathbf{g}_m = 1. \quad (18)$$

Substitution of these expressions into the Hamilton equations (Eq. 13) allows ray paths for any of the three different wave types to be traced using an appropriate numerical solver. For mildly anisotropic media, Červený and Firbas (1984) derive a linearisation procedure which allows anisotropic traveltimes to be computed using paths provided by an isotropic ray tracer. If an interface is located within an anisotropic media, then up to three reflected and three transmitted waves can be generated. Although Eq. 10 is still valid, it strictly applies to phase angle and phase velocity; thus, in order to correctly propagate rays in the presence of interfaces, additional constraints are required (see Slawinski et al., 2000).

In this section, we have only touched on kinematic ray theory; later, we will explore the more practical aspects of implementation. For further details on the underlying theory, the interested reader is referred to the comprehensive texts of Červený (2001) and Chapman (2004), which cover various aspects of kinematic and dynamic ray theory, ray amplitudes and synthetic seismograms.

1.4 *Common model parameterisations*

When rays or wavefronts are tracked through 2-D or 3-D laterally heterogeneous media, a formal description or parameterisation of the spatial variations in seismic properties is required. There are many ways that this can be done, and the final choice can depend on a variety of factors including the types of seismic structures that are present, the prediction scheme that is to be applied, and the dataset that is being simulated. For example, if the dataset was very large, then a fast and robust prediction scheme would be desirable; if the dataset contained refraction and reflection phases, then a layered medium would

be required and the prediction scheme would need to account for discontinuities. On the other hand, if the preferred prediction scheme was based purely on numerically solving Eq. 8, then only isotropic structures described by continuous variations in wavespeed would be permitted. However, if one was trying to predict phases in the presence of complex structures such as salt domes, subduction zones or heavily faulted sedimentary basins, then this class of prediction scheme would not be appropriate.

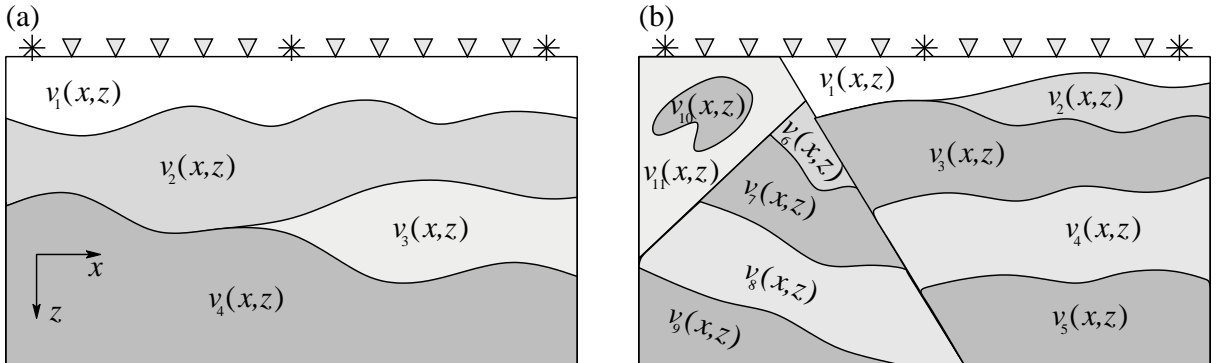


Fig. 3. General schemes for representing structure. (a) Laterally continuous interfaces separating layers within which wavespeed $v_i(x, z)$ varies smoothly, (b) more flexible framework based on an aggregate of irregular blocks within which wavespeed $v_i(x, z)$ varies smoothly.

In laterally heterogeneous media, the most general type of parameterisation needs to allow for both velocity and interface variation. This variation would need to be almost arbitrary if one wanted to represent all possible types of Earth structure, but in practice, a number of acceptable assumptions can usually be made. In seismic tomography (Nolet, 1987; Iyer and Hirahara, 1993; Rawlinson and Sambridge, 2003a), for example, it is common practice, when interfaces are required, to represent the medium by layers, within which wavespeed varies continuously, separated by sub-horizontal interfaces which vary in depth (e.g. Chiu et al., 1986; Farra and Madariaga, 1988; Williamson, 1990; Sambridge, 1990; Wang and Houseman, 1994; Zelt, 1999; Rawlinson et al., 2001a) as shown in Fig. 3a. The relative simplicity of this representation makes it amenable to fast and robust data prediction, and it also allows a variety of later arriving phases to be computed. However, in exploration seismology, where data coverage is usually dense, and near surface complexities (particularly faults) often need to be accurately represented, this class of parameterisation

can be too restrictive.

An alternative approach is to divide the model region up into an aggregate of irregularly shaped volume elements (see Fig. 3b), within which material property varies smoothly, but is discontinuous across element boundaries (e.g. Pereyra, 1996; Bulant, 1999). This allows most geological features such as faults, folds, lenses, overthrusts etc. to be faithfully represented. However, in the presence of such complexity, the data prediction problem becomes much more difficult to resolve, and reconciling data observations with these predictions (e.g. via seismic tomography) would be extremely challenging in the absence of accurate *a priori* information.

Common parameterisations used to describe wavespeed variations (or other seismic properties) in a continuum include constant velocity (or slowness) blocks (e.g. Aki et al., 1977; Nakanishi, 1985; Williamson, 1990; Saltzer and Humphreys, 1997), triangular/tetrahedral meshes within which velocity is constant or constant in gradient (e.g. White, 1989; Sambridge and Faletic, 2003), and grids of velocity nodes which are interpolated using a predefined function (e.g. Thomson and Gubbins, 1982; Thurber, 1983; Červený et al., 1984; Virieux and Farra, 1991; Zhao et al., 1992; Neele et al., 1993, etc.). Constant velocity blocks are conceptually simple, but require a fine discretisation in order to subdue the undesirable artifact of block boundaries. These discontinuities also have the potential to unrealistically distort the wavefield and make the two-point ray tracing problem more unstable. Triangular/tetrahedral meshes are flexible and allow analytic ray tracing when velocity is constant or constant in gradient within a cell; however, like constant velocity blocks, they usually require a fine discretisation, and can also destabilise the data prediction problem.

Velocity grids which describe a continuum using an interpolant offer the possibility of smooth variations with relatively few parameters, but are generally more computationally intensive to evaluate. In addition, analytic solution of the ray tracing equations is usually not possible. However, for many practical applications, the benefits of smooth-

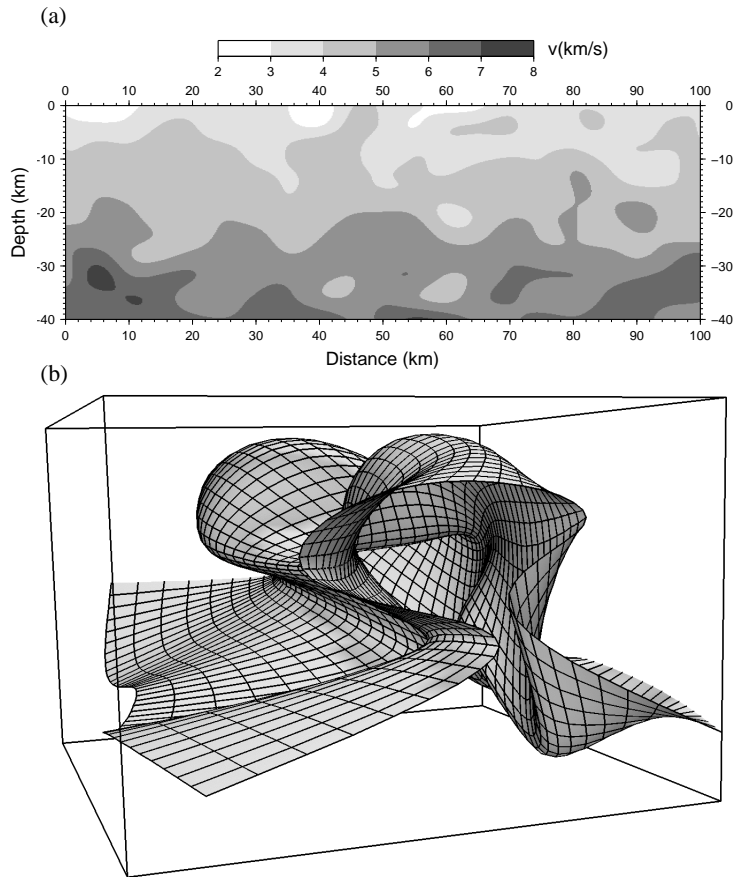


Fig. 4. Examples of cubic B-spline parameterisation for (a) velocity structure (b) interface structure.

ness outweigh these considerations. One of the simplest and most popular interpolants is pseudo-linear interpolation, which in 3-D Cartesian coordinates is:

$$v(x, y, z) = \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^2 V(x_i, y_j, z_k) \left(1 - \left| \frac{x - x_i}{x_2 - x_1} \right| \right) \left(1 - \left| \frac{y - y_j}{y_2 - y_1} \right| \right) \left(1 - \left| \frac{z - z_k}{z_2 - z_1} \right| \right) \quad (19)$$

where $V(x_i, y_j, z_k)$ are the velocity (or slowness) values at eight grid points surrounding (x, y, z) . For Eq. 19, velocity is continuous, but the velocity gradient is not (i.e. C^0 continuity). Despite this feature, pseudo linear interpolation has been frequently used in problems which require traveltime prediction (Eberhart-Phillips, 1986; Zhao et al., 1992; Scott et al., 1994; Steck et al., 1998).

Higher order interpolation functions are required if the velocity field is to have continuous first and second derivatives, which is usually desirable for schemes which numerically solve the ray tracing or eikonal equations. There are many types of spline functions that can

be used for interpolation, including Cardinal (Thomson and Gubbins, 1982; Sambridge, 1990), Bezier (Bartels et al., 1987), B-splines (Farra and Madariaga, 1988; Virieux and Farra, 1991; Rawlinson et al., 2001a) and splines under tension (Červený et al., 1984; Smith and Wessel, 1990; VanDecar et al., 1995). Cubic B-splines are particularly useful (Virieux and Farra, 1991; Rawlinson et al., 2001a) as they offer C^2 continuity, local control and the potential for an irregular distribution of nodes. For a set of velocity values $V_{i,j,k}$ on a 3-D grid of points $\mathbf{p}_{i,j,k} = (x_{i,j,k}, y_{i,j,k}, z_{i,j,k})$, the B-spline for the ijk th volume element is

$$\mathbf{B}_{i,j,k}(u, v, w) = \sum_{l=-1}^2 \sum_{m=-1}^2 \sum_{n=-1}^2 b_l(u)b_m(v)b_n(w)\mathbf{q}_{i+l,j+m,k+n}, \quad (20)$$

where $\mathbf{q}_{i,j,k} = (V_{i,j,k}, \mathbf{p}_{i,j,k})$. Thus, the three independent variables $0 \leq u, v, w \leq 1$ define the velocity distribution in each volume element. The weighting factors $\{b_i\}$ are the uniform cubic B-spline functions (Bartels et al., 1987). Fig. 4a shows a 2-D velocity field described by a mosaic of cubic B-spline elements.

Rather than use velocity grids in the spatial domain to describe smooth media, one could also exploit the wavenumber domain by employing a spectral parameterisation. These are often popular for global applications e.g. spherical harmonics (Dziewonski et al., 1977), but can also be used for problems on a local or regional scale. For example, Wang and Pratt (1997) use the following Fourier series to describe a 2-D slowness distribution in their inversion of reflection amplitude and traveltimes

$$s(\mathbf{r}) = a_{00} + \sum_{m=1}^N [a_{m0} \cos(\mathbf{k} \cdot \mathbf{r}) + b_{m0} \sin(\mathbf{k} \cdot \mathbf{r})] + \sum_{m=-N}^N \sum_{n=1}^N [a_{mn} \cos(\mathbf{k} \cdot \mathbf{r}) + b_{mn} \sin(\mathbf{k} \cdot \mathbf{r})], \quad (21)$$

where $\mathbf{r} = x\mathbf{i} + z\mathbf{j}$ and $\mathbf{k} = m\pi k_0\mathbf{i} + n\pi k_0\mathbf{j}$ are the position and wavenumber vector respectively, and a_{mn} and b_{mn} are the amplitude coefficients of the $(m, n)^{th}$ harmonic term. Although Eq. 21 is infinitely differentiable, it is globally supported in that adjustment of any amplitude coefficient influences the entire model. Spectral parameterisations have been used in a number of studies which require traveltime prediction (Hildebrand et al.,

1989; Hammer et al., 1994; Wiggins et al., 1996).

Interfaces are often described using equivalent parameterisations to those used for velocity. For example, linear segments (2-D volume), triangular meshes (3-D volume) or nodes with a specified interpolant are common. In 2-D, piecewise linear segments have been used in several studies (e.g. Zelt and Smith, 1992; Williamson, 1990), but the discontinuities in gradient between adjacent segments can have a destabilising effect on traveltime prediction schemes - for example, two ray paths with similar trajectories which impinge on an interface at either side of a discontinuity may reflect at very different angles. Zelt and Smith (1992) overcome this problem by applying a smoothing filter to the interface normals which are required by Snell’s law. Cubic B-splines in parametric form have been used by a number of authors (Farra and Madariaga, 1988; Virieux and Farra, 1991; Rawlinson et al., 2001a) to describe interface structure for the data prediction problem. For interface surfaces, Eq. 20 becomes

$$\mathbf{B}_{i,j}(u, v) = \sum_{k=-1}^2 \sum_{l=-1}^2 b_k(u)b_l(v)\mathbf{p}_{i+k,j+l}, \quad (22)$$

where $\mathbf{p}_{i,j} = (x_{i,j}, y_{i,j}, z_{i,j})$ is a set of control vertices on a topologically regular grid. Eq. 22 has the same desirable properties as its velocity counterpart (Eq. 20), and thanks to its parametric representation, allows multi-valued surfaces to be represented. Fig. 4b shows a complex interface surface that has been parameterised using cubic B-splines.

Finally, it is worth noting that irregular parameterisations have been used for both ray based and grid based schemes. In the case of grid based schemes which solve the eikonal equation, irregular grids have the potential to improve computational efficiency by varying grid resolution in response to wavefront curvature (e.g. Kimmel, 1998; Qian and Symes, 2002). Another motivation for adopting irregular grids comes from seismic tomography; for many large seismic datasets, path distribution can be highly heterogeneous, resulting in a spatial variability in resolving power. The ability to “tune” a parameterisation to these variations using some form of irregular mesh has a range of potential benefits, including

increased computational efficiency (fewer unknowns), improved stability of the inverse problem, and improved extraction of structural information (Michelini, 1995; Curtis and Snieder, 1997; Vesnaver et al., 2000; Spakman and Bijwaard, 2001; Rawlinson and Sambridge, 2003a; Sambridge and Faletic, 2003; Sambridge and Rawlinson, 2005). Completely unstructured meshes, such as those that use Delaunay tetrahedra or Voronoi polyhedra (Sambridge et al., 1995), offer high levels of adaptability, but have special book-keeping requirements when solving the forward problem of data prediction. Sambridge et al. (1995) and Sambridge and Gudmundsson (1998) describe techniques for locating points within these meshes, which allows ray tracing to be performed efficiently.

2 Ray tracing schemes

In this section, we describe a number of practical ray tracing schemes for solving the boundary value problem of locating source-receiver ray paths in various classes of media (2-D, 3-D, with and without discontinuity). There are two broad categories - shooting and bending - which exploit different formulations of the ray tracing equations (Eq. 8).

2.1 Shooting methods

Shooting methods of ray tracing are conceptually simple; they formulate Eq. 8 as an initial value problem which allows a complete ray path (with appropriate application of Snell's law in the presence of any interface) to be traced given an initial trajectory of the path. The two point problem of finding a source-receiver path then becomes an inverse problem in which the unknown is the initial direction vector of the ray, and the function to be minimised is a measure of the distance between the ray end point and receiver. The main challenge that faces this class of method is the non-linearity of the inverse problem, which tends to increase dramatically with the complexity of the medium (as Fig. 1 testifies).

2.1.1 The initial value problem

The appropriate form of the equation required to solve the initial value problem depends largely on the choice of parameterisation. In a medium described by constant velocity (slowness) blocks, the ray path is simply described by a piecewise set of straight line segments; all that is required to solve the initial value problem is repeated application of Snell's law at cell boundaries. This can be accomplished with high computational efficiency (e.g. Williamson, 1990). Analytic ray tracing can also be applied to other parameterisations; for example, triangular or tetrahedral meshes in which the velocity gradient is constant (e.g. White, 1989). The expression for ray trajectory in a medium with a constant velocity gradient can be expressed in various ways, including parametrically as (Rawlinson et al., 2001a)

$$\mathbf{x} = \frac{v(z_0)}{k} \left[\frac{a_0(c - c_0)}{1 - c_0^2}, \frac{b_0(c - c_0)}{1 - c_0^2}, 1 - \sqrt{\frac{1 - c^2}{1 - c_0^2}} \right] + \mathbf{x}_0, \quad (23)$$

where \mathbf{x}_0 is the origin of the ray segment, $[a, b, c]$ is a unit vector tangent to the ray path, $[a_0, b_0, c_0]$ is a unit vector tangent to the ray path at x_0 , k is the velocity gradient, and $v(z_0)$ is the velocity at z_0 . The associated traveltimes is then given by

$$T = \frac{1}{2k} \ln \left[\left(\frac{1 + c}{1 - c} \right) \left(\frac{1 - c_0}{1 + c_0} \right) \right] + T_0, \quad (24)$$

where T_0 is the traveltimes from the source to \mathbf{x}_0 . For application to tetrahedra (or triangles in 2-D), it is simply a matter of rotating the coordinate system so that the velocity gradient is in the direction of the z -axis. A number of other velocity functions yield analytic ray tracing solutions, such as the constant gradient of $\ln v$, and the constant gradient of the n th power of slowness $1/v^n$ (Červený, 2001).

Although analytic ray tracing is possible for a few special cases, in general one needs to solve Eq. 8 using numerical methods. This usually requires Eq. 8 to be reduced to a convenient first order initial value system of equations, which can be done in a variety of

ways. For example, by considering the following unit vector in the direction of the ray

$$\frac{d\mathbf{r}}{dl} = [\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta], \quad (25)$$

where θ is the inclination of the ray with the vertical (z -axis), and ϕ is the azimuth of the ray (angle between ray and positive x -axis in xy plane). Substitution of this expression into Eq. 8 and application of the product rule yields:

$$\left. \begin{aligned} \frac{\partial s}{\partial x} &= s \cos \theta \cos \phi \frac{d\theta}{dl} - s \sin \theta \sin \phi \frac{d\phi}{dl} + \sin \theta \cos \phi \frac{ds}{dl} \\ \frac{\partial s}{\partial y} &= s \cos \theta \sin \phi \frac{d\theta}{dl} + s \sin \theta \cos \phi \frac{d\phi}{dl} + \sin \theta \sin \phi \frac{ds}{dl} \\ \frac{\partial s}{\partial z} &= -s \sin \theta \frac{d\theta}{dl} + \cos \theta \frac{ds}{dl} \end{aligned} \right\} \quad (26)$$

These three equations can be rearranged to remove the ds/dl term and produce expressions for $d\theta/dl$ and $d\phi/dl$, which together with Eq. 25 produce the following system of equations

$$\left. \begin{aligned} \frac{dx}{dl} &= \sin \theta \cos \phi \\ \frac{dy}{dl} &= \sin \theta \sin \phi \\ \frac{dz}{dl} &= \cos \theta \\ \frac{d\theta}{dl} &= \frac{\cos \theta}{s} \left[\cos \phi \frac{\partial s}{\partial x} + \sin \phi \frac{\partial s}{\partial y} \right] - \frac{\sin \theta}{s} \frac{\partial s}{\partial z} \\ \frac{d\phi}{dl} &= \frac{1}{s \sin \theta} \left[\cos \phi \frac{\partial s}{\partial y} - \sin \phi \frac{\partial s}{\partial x} \right] \end{aligned} \right\} \quad (27)$$

Thus, given some initial position and trajectory, a ray path can be obtained by solving this coupled system of equations e.g. using a fourth order Runge-Kutta scheme (e.g. Kreyszig, 1993).

The initial value formulation of the kinematic ray tracing equations given by Eq. 27

uses path length l as the independent variable. However, it is often more convenient to use traveltime t , since this parameter is usually required in addition to path geometry. Conversion of Eq. 27 into a form that uses t as the independent variable and velocity v instead of slowness s (the use of slowness or velocity is often a matter of convention, but it can have practical implications for certain classes of problem, e.g. Rawlinson and Sambridge, 2003b) can be achieved using the following simple set of relationships

$$\frac{d\theta}{dl} = s \frac{d\theta}{dt}, \quad \frac{d\phi}{dl} = s \frac{d\phi}{dt}, \quad \frac{d\mathbf{x}}{dl} = s \frac{d\mathbf{x}}{dt}, \quad \frac{\partial s}{\partial \mathbf{x}} = -\frac{1}{v^2} \frac{\partial v}{\partial \mathbf{x}}, \quad (28)$$

which result in the following system of equations

$$\left. \begin{aligned} \frac{dx}{dt} &= v \sin \theta \cos \phi \\ \frac{dy}{dt} &= v \sin \theta \sin \phi \\ \frac{dz}{dt} &= v \cos \theta \\ \frac{d\theta}{dt} &= -\cos \theta \left[\cos \phi \frac{\partial v}{\partial x} + \sin \phi \frac{\partial v}{\partial y} \right] - \sin \theta \frac{\partial v}{\partial z} \\ \frac{d\phi}{dt} &= \frac{1}{\sin \theta} \left[\sin \phi \frac{\partial v}{\partial x} - \cos \phi \frac{\partial v}{\partial y} \right] \end{aligned} \right\}. \quad (29)$$

This system of equations has a similar form to Eq. 27, and therefore can be solved using the same class of technique. The nine ray paths shown in Fig. 5 were computed by solving Eq. 29 using a 4th order Runge Kutta method (with constant ϕ). This procedure is very efficient; for example, tracing 10,000 ray paths using a time step of 0.1 s from the same source point to the edge of the medium only takes 6 s on a 1.6 GHz Opteron PC running Linux.

Rather than use ray inclination and azimuth to describe ray trajectory, one could also use the components of the ray direction (or *slowness*) vector $\mathbf{p} = \nabla T$. A system of first-order equations can then be simply derived from Eq. 8 by using this vector as the substitution

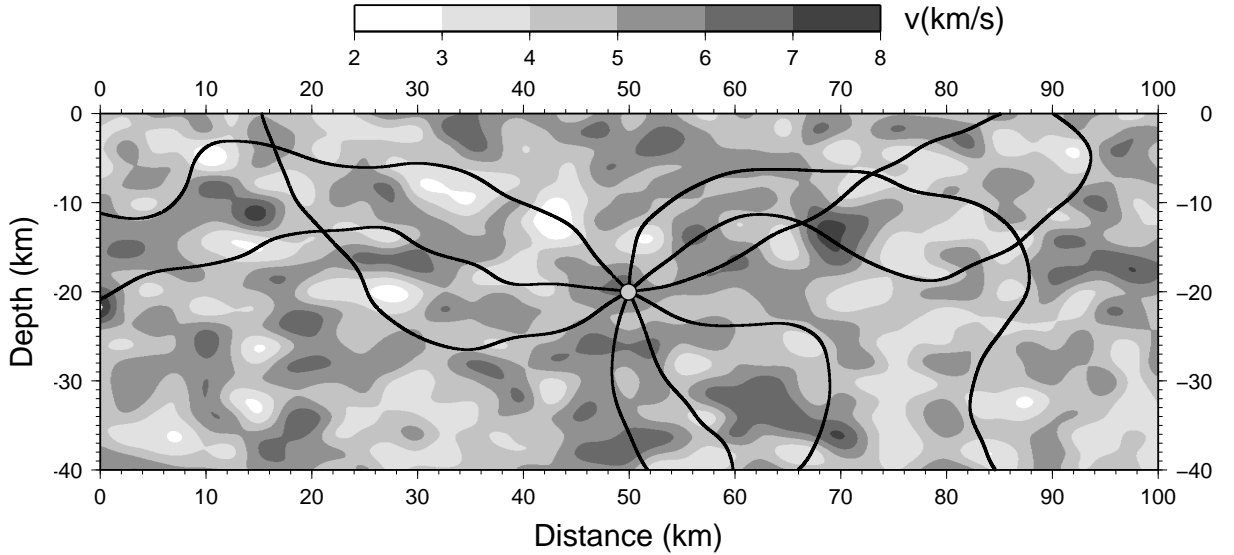


Fig. 5. A fan of nine ray paths traced from a source point in a complex 2-D medium described by a mesh of cubic B-spline functions. A fourth order Runge Kutta scheme is used to solve Eq. 29 in this case.

variable i.e. $\nabla T = s \mathbf{dr}/dl$. This results in

$$\left. \begin{aligned} \frac{d\mathbf{r}}{dl} &= v\mathbf{p} \\ \frac{d\mathbf{p}}{dl} &= \nabla \frac{1}{v} \end{aligned} \right\}, \quad (30)$$

which is a system of six equations. The independent variable s can be replaced by travelttime t in the same way as before to produce

$$\left. \begin{aligned} \frac{d\mathbf{r}}{dt} &= v^2\mathbf{p} \\ \frac{d\mathbf{p}}{dt} &= -\frac{1}{v}\nabla v \end{aligned} \right\}. \quad (31)$$

This is the same set of equations that would result from substituting Eq. 11 into the Hamilton equations 13. Although Eq. 29 has one less dependent variable to compute than Eq. 31, it contains trigonometric functions which, from a computational point of view, are less desirable. The 6D vector (\mathbf{r}, \mathbf{p}) , which uniquely describes the position and trajectory of a ray in 3-D space, is sometimes referred to as the *phase space* vector (Chapman, 2004), while its 5-D counterpart $(\mathbf{r}, \theta, \phi)$ is sometimes referred to as the *reduced phase*

space vector (e.g. Osher et al., 2002).

It is worth emphasising that a variety of other first order formulations of the kinematic ray tracing equations in isotropic media can be derived - for more details, see Červený (1987, 2001). In general anisotropic media, a system of six first-order equations can be obtained by substituting the Hamiltonian defined by Eq. 13 into the Hamilton equations, which results in (Červený and Firbas, 1984; Chapman, 2004)

$$\left. \begin{aligned} \frac{dx_i}{dt} &= \frac{\partial G_m}{\partial p_i} = \frac{c_{ijkl} p_k g_m^j g_m^l}{\rho} \\ \frac{dp_i}{dt} &= -\frac{\partial G_m}{\partial x_i} = -\frac{1}{2} \frac{\partial [c_{ijkl}/\rho]}{\partial x_i} p_j p_m g_m^k g_m^l \end{aligned} \right\}. \quad (32)$$

Although much more computationally expensive to solve than its isotropic equivalent, initial value anisotropic ray tracing in 2-D or 3-D media does not pose a significant challenge for modern computers.

In the presence of smooth velocity variations, the kinematic ray tracing equations provide the required solution, but as soon as velocity discontinuities are introduced, two additional problems need to be solved: (1) locate ray-interface intersection point; (2) find trajectory of departing ray path. The first problem can be difficult to solve, particularly if both ray paths and interfaces are described by non-linear functions. If a medium is divided up into cells, then a first level of refinement can be achieved by simply knowing which cells contain an interface; thus when a ray enters a cell, one knows whether or not an intersection is possible. Sambridge and Kennett (1990) devise a boundary value ray tracing scheme for 3-D media which contain discontinuities; rather than exactly locate the ray-interface intersection point, the time step length used in Eq. 29 is iteratively adjusted in the vicinity of an interface based on a linear update procedure. The iteration process ceases when the distance between the ray end point and interface satisfies a tolerance criteria.

In an alternative approach, Virieux and Farra (1991) exploit the convex hull and subdivision properties of B-splines. The first property states that the B-spline surface must lie

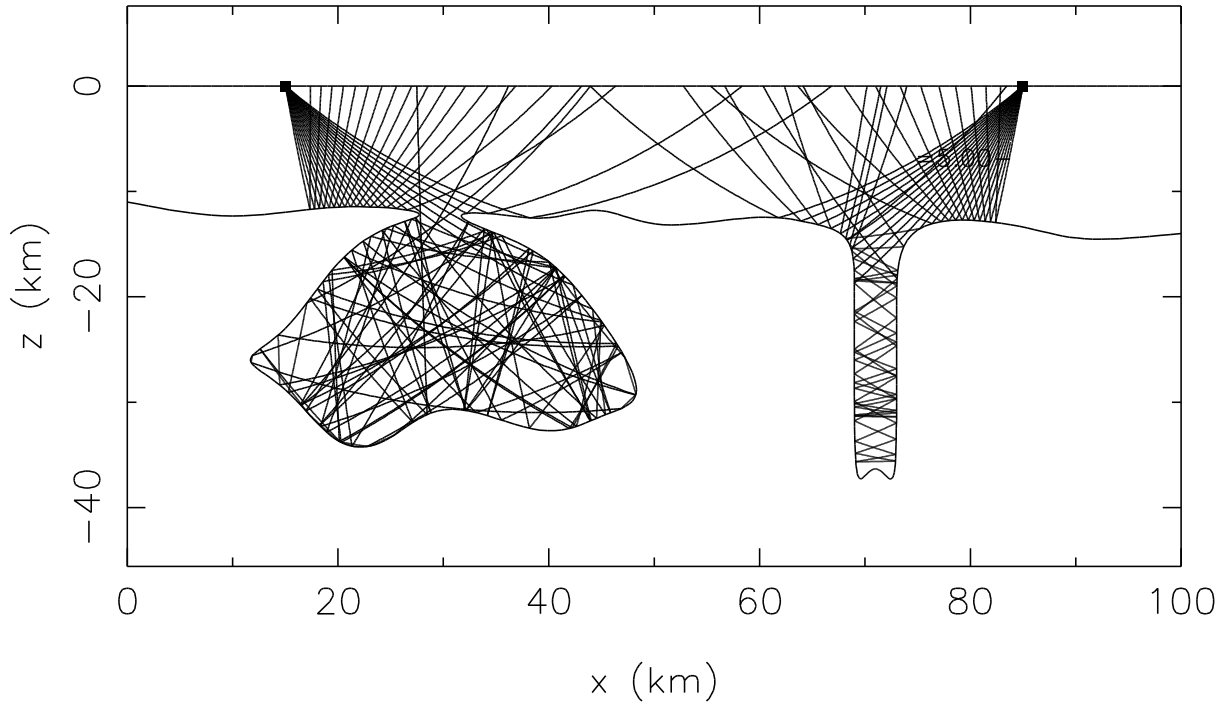


Fig. 6. Reflection paths from two source points in the presence of a complex multi-valued interface described by cubic B-splines. Path trajectories are obtained by solving an initial value problem until the ray impinges on the surface $z = 0$. Note that it is decided *a priori* that all rays impinging on the interface will reflect. Thus, rays trapped inside complex features of the model keep reflecting until they finally emerge and intersect with the surface.

within the convex volume defined by the control points; the second property allows the same surface to be described by larger numbers of control points. Thus, the approximate location of a ray-interface intersection point can be found by determining whether the ray (locally linear in this case) intersects a parallelepiped containing the convex volume; if it does, then the convex volume can be subdivided and the simple ray-parallelepiped problem can be resolved for each of the new, smaller, parallelepipeds. Repetition of this process allows all possible intersection points to be found. This scheme is very robust, but can be computationally expensive; as a result, Virieux and Farra (1991) switch to a more efficient local search method once it is considered a unique intersection point is being targeted by the subdivision scheme.

Like Virieux and Farra (1991), Rawlinson et al. (2001a) also use cubic B-splines in para-

metric form to represent interface surfaces, but adopt a different approach to finding ray-interface intersection points. In this case, rays are defined by circular arc segments. Rather than exploit the convex hull property of B-splines, an initial intersection point is found by approximating the surface by a mosaic of triangular patches, for which analytic solution of the ray-interface intersection problem is possible. Once a preliminary intersection point is found, an iterative non-linear scheme based on the generalised Newton method is used to target the true intersection point. Although computationally efficient, the procedure may fail to locate some intersection points, particularly if the trajectory of the ray is nearly parallel to the interface when intersection occurs. Another possible problem is that the local method converges to the incorrect intersection point. However, in practice the scheme is effective, as demonstrated in 2-D by Fig. 6, which shows reflection paths propagating through a highly complex interface structure. As well as demonstrating the robustness of the intersection scheme, this example also illustrates the power of initial value ray tracing: given virtually any structure, it is possible to trace ray paths of almost any specified initial trajectory.

The second problem that needs to be solved when a ray impinges on an interface is to find the trajectory of the departing ray path. In 2-D, a simple application of Snell's law (Eq. 10) can produce up to four departing paths (transmission, reflection, converted transmission, converted reflection) for isotropic media. In 3-D, the same class of paths can be generated, but in addition to Snell's law, the continuity of projection of the incident ray must also be considered. This is equivalent to requiring that the incident path, departing path and normal to the interface at the intersection point, all lie in the same plane. By combining this constraint with Snell's law, the slowness vector \mathbf{p}_r of the reflected or refracted ray path can be defined by (Červený, 1987; Sambridge and Kennett, 1990):

$$\mathbf{p}_r = \mathbf{p}_i + \left\{ \kappa \left[\frac{1}{v_r^2} - \frac{1}{v_i^2} + (\mathbf{p}_i \cdot \mathbf{n})^2 \right]^{1/2} - \mathbf{p}_i \cdot \mathbf{n} \right\} \mathbf{n} \quad (33)$$

where \mathbf{p}_i is the slowness vector of the incident path, \mathbf{n} is a normal vector to the interface at the intersection point, and v_i and v_r are the wavespeeds of the incident and departing

rays. $\kappa = \text{sign}(\mathbf{p}_i \cdot \mathbf{n})$ and equals $+1$ if \mathbf{p}_i makes an acute angle with \mathbf{n} and -1 otherwise. When unconverted reflected waves are required ($v_i = v_r$), Eq. 33 reduces to:

$$\mathbf{p}_r = \mathbf{p}_i - 2(\mathbf{p}_i \cdot \mathbf{n})\mathbf{n} \quad (34)$$

For general anisotropic media, up to three reflected and three transmitted waves can be generated for a single incident path - see Slawinski et al. (2000) for details on how they can be calculated.

2.1.2 The boundary value problem

Shooting methods of ray tracing usually solve the boundary value problem by probing the medium with initial value ray paths and then exploiting information from the computed paths to better target the receiver. Fig. 7 illustrates this basic concept in 2-D. If a ray

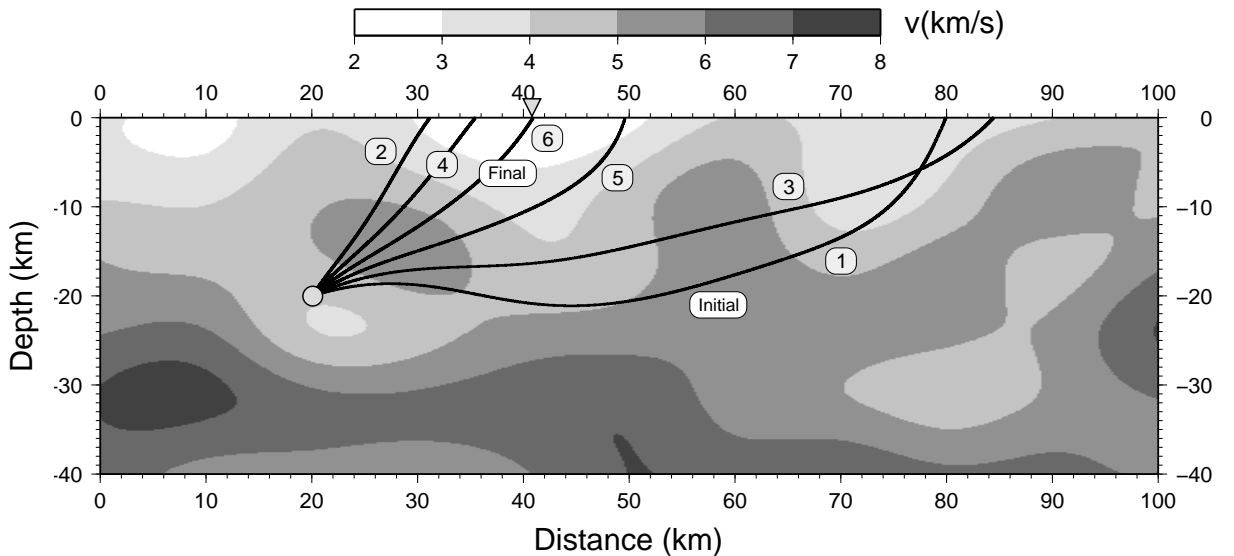


Fig. 7. Principle of the shooting method. In this case, an initial path trajectory is updated until it converges at the receiver.

emanates from a source point in a 3-D medium with take off angles θ_o and ϕ_o , and the aim is for the ray end point (x_e, y_e) on the receiver plane ($z = \text{constant}$) to coincide with the receiver location (X_r, Y_r) , then the boundary value problem amounts to finding the

(θ_o, ϕ_o) that solve the two non-linear simultaneous equations

$$\left. \begin{aligned} x_e(\theta_o, \phi_o) &= X_r \\ y_e(\theta_o, \phi_o) &= Y_r \end{aligned} \right\}. \quad (35)$$

Given that (x_e, y_e) cannot be expressed explicitly as a function of (θ_o, ϕ_o) for most velocity fields, it is usually the case that the boundary value problem is posed as an optimisation problem, with the misfit function to be minimised expressed as some measure of the distance between the ray end point and its intended target. Since the optimisation problem is non-linear, a range of iterative non-linear and fully non-linear schemes can be applied.

Julian and Gubbins (1977) propose two iterative non-linear schemes for solving Eq. 35.

The first of these is Newton's method, which amounts to a simple linearisation:

$$\begin{bmatrix} \frac{\partial x_e}{\partial \theta_o} & \frac{\partial x_e}{\partial \phi_o} \\ \frac{\partial y_e}{\partial \theta_o} & \frac{\partial y_e}{\partial \phi_o} \end{bmatrix} \begin{bmatrix} \theta_o^{n+1} - \theta_o^n \\ \phi_o^{n+1} - \phi_o^n \end{bmatrix} = \begin{bmatrix} X_r - x_e(\theta_o^n, \phi_o^n) \\ Y_r - y_e(\theta_o^n, \phi_o^n) \end{bmatrix}. \quad (36)$$

Thus, given some starting initial trajectory θ_o^0, ϕ_o^0 , solution of Eq. 36 provides an updated initial trajectory θ_o^1, ϕ_o^1 , and the process is repeated until an appropriate tolerance criterion is met. The success of this scheme depends largely on two factors: (1) accurate calculation of the partial derivative matrix, and (2) obtaining an initial guess ray that will converge to the correct minimum under the assumption of local linearity. Both of these requirements can be difficult to satisfy, particularly in complex media. One approach to estimating the partial derivatives involves fitting two planes through the end points of a cluster of three ray paths with different initial projection angles (i.e. one plane through all three $x_o(\theta_o, \phi_o)$ and the other plane through all three $y_o(\theta_o, \phi_o)$). The gradient of these planes in the θ_o and ϕ_o directions provide estimates of the four partial derivatives in Eq. 36. This approach is actually equivalent to the method of false position, which is the second iterative non-linear scheme proposed by Julian and Gubbins (1977). The method of false position is unlikely

to be first-order accurate like a true Newton method, and therefore will converge more slowly and possibly with less stability. However, it will be faster at each iteration, and Rawlinson et al. (2001a,b) found it to be sufficiently robust to use in solving the forward step of large 3-D tomographic inverse problems. Fig. 8 shows two-point paths through a 3-D laterally heterogeneous structure that were computed using this scheme.

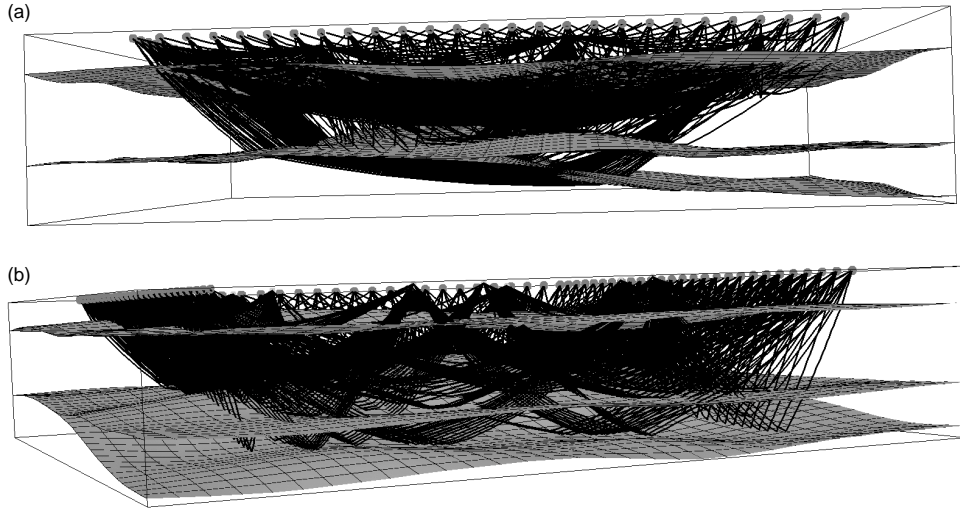


Fig. 8. Two-point paths computed through a 3-D layered model using the shooting scheme of Rawlinson et al. (2001a). (a) Refracted paths, (b) reflected paths.

Sambridge and Kennett (1990) directly compute the partial derivatives in Eq. 36 by exploiting wavefront curvature information at the end point of the ray. This can be done by differentiating Eq. 29 with respect to the initial take-off angles (θ_o, ϕ_o) and reversing the order of differentiation, resulting in an additional set of 10 first-order equations which must be solved together with the original set of five. The computed variables are closely related to the derivative terms in Eq. 36, and allow the Newton scheme to be applied with first order accuracy. Note that the additional equations now contain second-order derivatives of velocity, which means that the velocity field must have C_2 continuity.

The second requirement of a successful iterative non-linear shooting method is a sufficiently accurate initial guess ray. This can be obtained in various ways, including shooting a broad fan of rays in the general direction of the receiver array, and then (if necessary) shooting out increasingly targeted clusters of rays towards zones containing receivers un-

til a suitably accurate initial ray is obtained (Virieux and Farra, 1991; Rawlinson et al., 2001a). Another approach is to use the correct two-point ray for a laterally averaged version of the model as the initial guess ray (Thurber and Ellsworth, 1980; Sambridge, 1990). It is worth noting that the two-point problem in ray tracing is not the only type of boundary value problem; for example, one may wish to compute paths from an incident teleseismic wavefront below the crust or lithosphere to a receiver array on the surface. In this case, rays that end at a receiver begin at specific points along the wavefront surface. Fig. 9 shows a solution to this class of boundary value problem obtained using the iterative non-linear shooting scheme of Rawlinson and Houseman (1998).

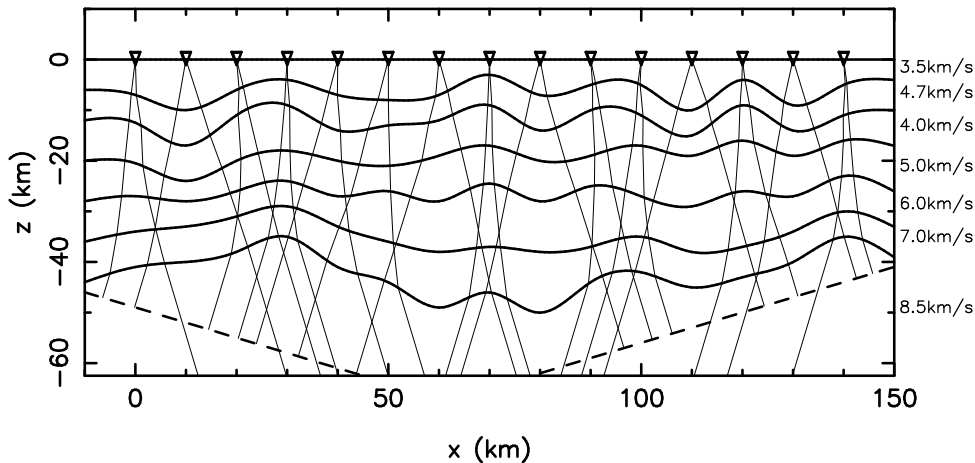


Fig. 9. A boundary value problem involving teleseismic wavefronts rather than point sources can also be solved using shooting methods of ray tracing.

Shooting methods of ray tracing are widely used in seismology due to their conceptual simplicity, and potential for high accuracy and efficiency. One area in which they enjoy frequent application is seismic tomography, where 2-D or 3-D variations in seismic properties are imaged by matching data observations with data predictions using inversion techniques (e.g. Cassell, 1982; Benz and Smith, 1984; Langan et al., 1985; Farra and Madariaga, 1988; White, 1989; Sambridge, 1990; Zelt and Smith, 1992; VanDecar et al., 1995; McCaughey and Singh, 1997; Rawlinson et al., 2001b).

2.1.3 Paraxial Ray Tracing

An important field of ray theory that is yet to be mentioned is the paraxial ray approximation (Červený and Pšenčík, 1983; Červený and Firbas, 1984; Červený, 1987; Farra and Madariaga, 1988; Virieux and Farra, 1991; Červený, 2001; Červený et al., 2006), which is widely employed by the seismology community for various aspects of data prediction. It essentially involves using first order perturbation theory to deduce characteristics of the wavefield in the neighbourhood of a reference ray. Thus, given some reference path with position and slowness vector $\mathbf{y}_0(t) = [\mathbf{r}_0(t), \mathbf{p}_0(t)]^T$, a paraxial ray can be defined by the first-order approximation

$$\mathbf{y}(t) = \mathbf{y}_0(t) + \delta\mathbf{y}(t) = [\mathbf{r}_0(t) + \delta\mathbf{x}(t), \mathbf{p}_0(t) + \delta\mathbf{p}(t)]. \quad (37)$$

To obtain the paraxial ray tracing equations, consider the linearisation of the kinematic ray tracing equations $\dot{\mathbf{r}} = \nabla_{\mathbf{p}}H$ and $\dot{\mathbf{p}} = -\nabla_{\mathbf{r}}H$ (equivalent to Eq. 13 with $()$ denoting differentiation with respect to t), which can be written as

$$\left. \begin{aligned} \delta\dot{\mathbf{r}} &= \frac{\partial\dot{\mathbf{r}}}{\partial\mathbf{r}}\delta\mathbf{r} + \frac{\partial\dot{\mathbf{r}}}{\partial\mathbf{p}}\delta\mathbf{p} \\ \delta\dot{\mathbf{p}} &= \frac{\partial\dot{\mathbf{p}}}{\partial\mathbf{r}}\delta\mathbf{r} + \frac{\partial\dot{\mathbf{p}}}{\partial\mathbf{p}}\delta\mathbf{p} \end{aligned} \right\}, \quad (38)$$

or in more compact form as

$$\delta\dot{\mathbf{y}} = A\delta\mathbf{y} \quad \text{where} \quad A = \begin{bmatrix} \nabla_{\mathbf{r}}\nabla_{\mathbf{p}}H & \nabla_{\mathbf{p}}\nabla_{\mathbf{p}}H \\ -\nabla_{\mathbf{r}}\nabla_{\mathbf{r}}H & -\nabla_{\mathbf{p}}\nabla_{\mathbf{r}}H \end{bmatrix}. \quad (39)$$

The paraxial ray tracing equation has six independent solutions in 3-D and four in 2-D. Since we can write that

$$\mathbf{y}(t) = \mathbf{y}(t_0) + \frac{d\mathbf{y}(t)}{dt}\delta t = \mathbf{y}(t_0) + \frac{\partial\mathbf{y}(t)}{\partial\mathbf{y}(t_0)} \frac{d\mathbf{y}(t_0)}{dt}\delta t, \quad (40)$$

it is common to express the solution to Eq. 39 in terms of a ray propagator matrix $P(t, t_0)$ (see Červený, 1987; Virieux and Farra, 1991)

$$\delta\mathbf{y}(t) = P(t, t_0)\delta\mathbf{y}(t_0) \quad \text{where} \quad P(t, t_0) = \begin{bmatrix} \frac{\partial\mathbf{x}(t)}{\partial\mathbf{x}(t_0)} & \frac{\partial\mathbf{x}(t)}{\partial\mathbf{p}(t_0)} \\ \frac{\partial\mathbf{p}(t)}{\partial\mathbf{x}(t_0)} & \frac{\partial\mathbf{p}(t)}{\partial\mathbf{p}(t_0)} \end{bmatrix}, \quad (41)$$

which has the initial condition that $P(t = t_0, t_0) = I$, the identity matrix. The power of paraxial rays is that they allow information about the wavefield in the vicinity of a reference ray to be used, for example to detect caustics in two-point ray tracing. In fact, the shooting method of Sambridge and Kennett (1990) described in the previous section, which makes use of wavefront curvature information in the vicinity of a ray to compute partial derivatives for an iterative non-linear update scheme, is an example of using paraxial ray theory in two-point ray tracing. Virieux and Farra (1991) describe a similar scheme, and in fact there are many shooting schemes of ray tracing which exploit the paraxial ray approximation (e.g. Červený and Firbas, 1984; Farra and Madariaga, 1988; Bulant, 1996). For more details on paraxial ray theory and its many other potential applications, refer to Červený and Pšenčík (1983); Červený and Firbas (1984); Červený et al. (1984); Červený (1987); Farra and Madariaga (1987); Klimeš (1989); Bulant (1996); Červený (2001); Červený et al. (2006).

2.1.4 Fully non-linear shooting methods

Shooting methods which use an iterative non-linear approach to solve the boundary value problem are often very efficient in mildly heterogeneous media, but generally become less robust as the complexity of the medium increases. Fig. 10 illustrates why this is the case by showing the relationship between ray inclination angle and distance from a receiver for two velocity models of different complexity. In the first model (Fig. 10a), velocity increases linearly with depth but has random velocity fluctuations with a standard deviation of 0.4 km/s superimposed. These velocity perturbations are sufficient to cause some mild

focusing and defocusing of the wavefront, as reflected by the variable density of ray paths along the surface. This effect is also manifest in the corresponding plot of ray end point to receiver distance vs. initial ray inclination; the plot is asymmetrical about 0° (vertical initial inclination) and contains significant variations in curvature. Despite these non-linearities, it is likely that any iterative non-linear shooting scheme will converge to the correct solution using a starting ray with initial angle in the range $-45^\circ \rightarrow 45^\circ$, although the initiation of a triplication several kilometres to the right of the receiver (Fig. 10a) may cause some difficulties (it appears as a small region of near zero gradient in the ray inclination vs. distance plot at about -4°).

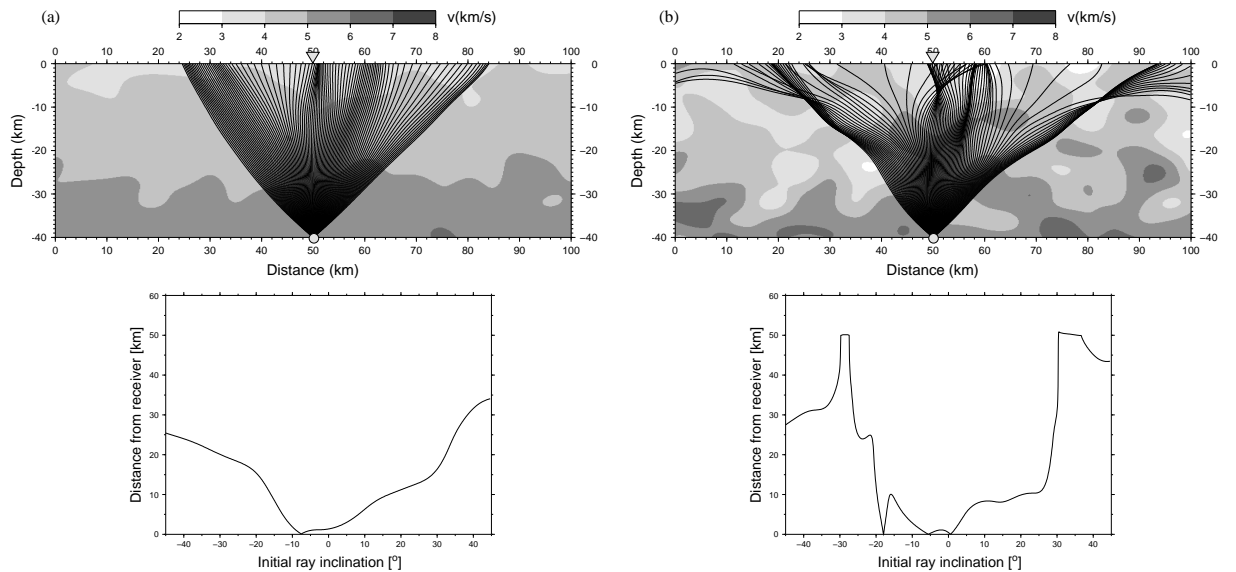


Fig. 10. Demonstration of the non-linear relationship that exists between initial ray trajectory and distance from ray end point to receiver. (a) Mildly heterogeneous medium with lateral velocity standard deviation of 0.4 km/s; (b) strongly heterogeneous medium with lateral velocity standard deviation of 1.4 km/s. The source is denoted by a grey circle and the receiver by a grey triangle. Note that while 100 ray paths are shown in each of the ray tracing plots (top), the corresponding ray end point vs. initial ray inclination diagrams (bottom) were generated using 1,000 rays.

The second model (Fig. 10b) is identical to Fig. 10a except that the random velocity fluctuations now have a standard deviation of 1.4 km/s, which results in significant lateral heterogeneity. Note that the pattern of anomalies is identical to the first model; it is

only their amplitude which has changed. The effect of the increased amplitude on the ray distribution is dramatic, with several triplications of the ray field now evident. The corresponding distance vs. ray inclination plot reveal these multi-pathing effects as extrema. Clearly, at least three arrivals are detected by the receiver. If one was to use an iterative non-linear shooting method in this situation with a starting ray in the initial inclination range of $-45^\circ \rightarrow 45^\circ$, then it is possible that a two point path will be located, but it is also possible that the scheme will become trapped in a local minimum. Shooting a broad fan of rays, and then increasingly more targeted clusters based on previous sampling of the wavefield, may provide sufficiently accurate initial rays (e.g. Rawlinson et al., 2001a) to locate one or more global minima. Ultimately, though, there will be a trade-off between computing time and the number of two-point paths found, and one must always make a decision as to which point along this trade-off curve is adequate for the problem at hand.

Given the potential pitfalls of using an iterative non-linear solver in two point shooting methods, it would appear that fully non-linear solvers would be at least worthy of investigation. However, there are relatively few examples in the literature, perhaps due to the recent proliferation of grid based and wavefront construction type schemes that are designed to overcome these limitations (these schemes will be discussed later in some detail). Velis and Ulrych (1996) describe a fully non-linear shooting method of ray tracing that uses simulated annealing to locate the global minimum path. Simulated annealing (e.g. Kirkpatrick et al., 1983) is based on an analogy with physical annealing in thermodynamic systems to guide variations to the model parameters, in this case the initial ray trajectories. Velis and Ulrych (2001) extend their method to heterogeneous 3-D velocity models which can include variable thickness layers, faults, and complex structures such as salt domes. The scheme exhibits several advantages compared to more conventional ray tracing schemes; in particular increased robustness for locating the global minimum solution. However, it does not appear to be practical for finding all multipaths, and tends to be more computationally expensive than iterative non-linear solvers.

2.2 Bending methods

The principle of the bending method of ray tracing is to iteratively adjust the geometry of an initial arbitrary path that joins source and receiver until it becomes a true ray path (i.e. it satisfies Fermat's principle of stationary time) - see Fig. 11. A common approach to implementing the bending method is to derive a boundary value formulation of the kinematic ray tracing equations which can then be solved iteratively. There are many ways that this can be done; here, we describe a scheme that was first devised by Julian and Gubbins (1977). The traveltime T of a ray path between source S and receiver R can in general be expressed by the integral

$$T = \int_S^R s dl, \quad (42)$$

where s is slowness and l is path length. The ray path can be described parametrically by a monotonic function λ , the normalised path length ($\lambda = l/L$, where L is the total path length of the ray), in which case $\mathbf{r} = \mathbf{r}(\lambda)$. A perturbation in path length can therefore be written as

$$\frac{dl}{d\lambda} = \sqrt{\dot{\mathbf{r}} \cdot \dot{\mathbf{r}}} = \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} = F, \quad (43)$$

where $(\dot{})$ denotes differentiation with respect to λ , and the use of normalised path length means that $F = L$ and $dF/d\lambda = 0$. Using this expression, Eq. 42 can be rewritten as

$$T = \int_S^R s F d\lambda. \quad (44)$$

The ray tracing equations can be obtained by extremizing this integral using the calculus of variations (e.g. Jeffreys and Swirles, 1966). For any integrand $G(\lambda, \mathbf{r}(\lambda), \dot{\mathbf{r}}(\lambda))$ the Euler-Lagrange equations are

$$\frac{\partial G}{\partial \mathbf{r}} - \frac{d}{d\lambda} \left[\frac{\partial G}{\partial \dot{\mathbf{r}}} \right] = 0. \quad (45)$$

In our case, the integrand is $G = sF = s\sqrt{\dot{\mathbf{r}} \cdot \dot{\mathbf{r}}}$, and substitution into Equation Eq. 45 yields

$$s\ddot{\mathbf{r}} + (\dot{\mathbf{r}} \cdot \nabla s)\dot{\mathbf{r}} - (\dot{\mathbf{r}} \cdot \dot{\mathbf{r}})\nabla s = 0. \quad (46)$$

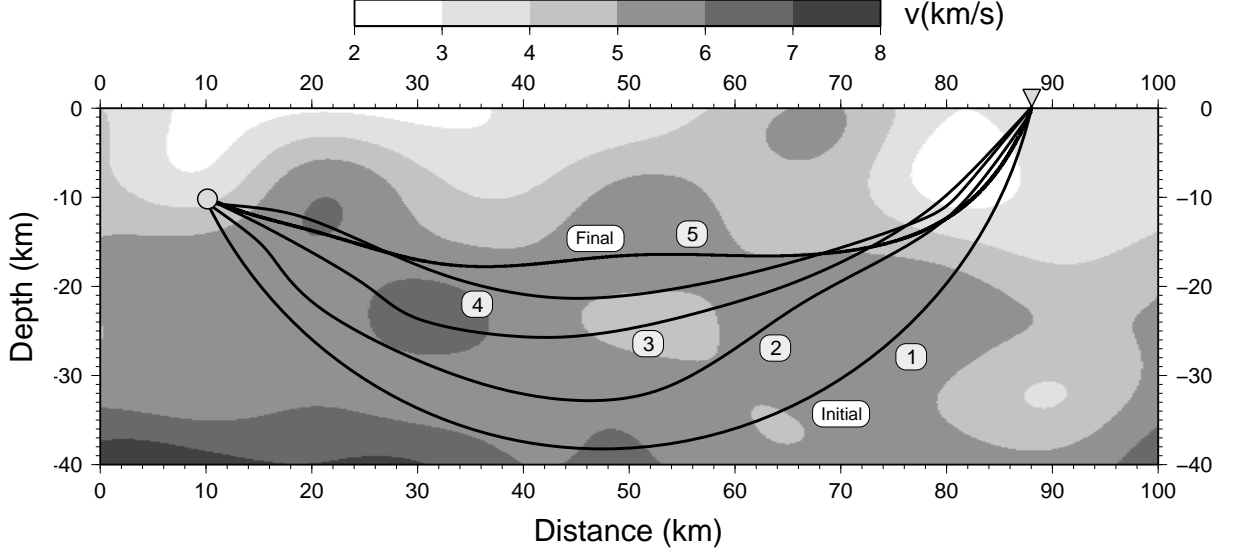


Fig. 11. Principle of the bending method. In this case, an initial two point path is perturbed until it satisfies Fermat's principle.

It is easy to show that only two of these equations are independent (Julian and Gubbins, 1977), and that one may be ignored without loss of generality. This leaves two equations with three unknowns $\mathbf{r} = (x, y, z)$; a final constraint comes from $dF/d\lambda = 0$ (so $\dot{\mathbf{r}} \cdot \ddot{\mathbf{r}} = 0$). Thus, a system of three independent non-linear second order differential equations can be explicitly written (Julian and Gubbins, 1977)

$$\left. \begin{aligned} s\ddot{x} + s_y\dot{y}\dot{x} + s_z\dot{z}\dot{x} - s_x(\dot{y}^2 + \dot{z}^2) &= 0 \\ s\ddot{y} + s_x\dot{x}\dot{y} + s_z\dot{z}\dot{y} - s_y(\dot{x}^2 + \dot{z}^2) &= 0 \\ \dot{x}\ddot{x} + \dot{y}\ddot{y} + \dot{z}\ddot{z} &= 0 \end{aligned} \right\}, \quad (47)$$

where $\nabla s = (s_x, s_y, s_z)$. The boundary conditions for this problem are $\mathbf{r}(0) = \mathbf{r}_S$ and $\mathbf{r}(1) = \mathbf{r}_R$, and an iterative non-linear solution approach is possible given some initial estimate of the path $\mathbf{r}(\lambda)^0$, so that in general $\mathbf{r}(\lambda)^{n+1} = \mathbf{r}(\lambda)^n + \delta\mathbf{r}(\lambda)^n$. Substitution of this expression into Eq. 47 and linearising the resulting equations for $\delta\mathbf{r}(\lambda)^n$ allows solutions to be obtained using, for example, second order finite difference techniques (Julian and

Gubbins, 1977). The iterative process can be continued until some suitable convergence criterion, based on the path perturbation integrated along the ray, is satisfied.

Pereyra et al. (1980) devise a similar technique to that described above for ray bending, but extend it to allow for the presence of interfaces. This can be achieved by considering a separate system of differential equations in each smooth region, and coupling them using the known discontinuity condition at each interface that is traversed by the ray path. The order in which the interfaces are intersected by the ray path needs to be known in advance, which may be a drawback in complex structures. This scheme is developed further in very complex 3-D models by Pereyra (1996), who also implements shooting to obtain an initial guess ray. This helps to overcome the problem of knowing *a priori* the correct path sequence.

2.2.1 Pseudo-bending methods

Pseudo-bending methods are similar in principle to the bending scheme described above, but avoid direct solution of the ray equations. One of the first of these schemes was developed by Um and Thurber (1987), and is based upon the ray path being represented by a set of linearly interpolated points. Given some initial arbitrary path, the aim is to sequentially adjust the location of each point so that the path better satisfies the ray equations. This can be accomplished quite efficiently by locating the direction of the ray path normal and then directly exploiting Fermat's principle of stationary time. If we denote \mathbf{m}_t and \mathbf{m}_n as vectors tangent and normal to the ray path respectively at some point \mathbf{r} (see Fig. 12a), then

$$\mathbf{m}_t = \frac{d\mathbf{r}}{dl} \quad \text{and} \quad \mathbf{m}_n = \frac{d\mathbf{m}_t}{dl}, \quad (48)$$

where \mathbf{m}_t is a unit vector. By using Eq. 5, \mathbf{m}_n can be written

$$\mathbf{m}_n = \frac{d}{dl} \left(\frac{\nabla T}{s} \right) = \frac{1}{s} \frac{d\nabla T}{dl} + s \nabla T \frac{d}{dl} \left(\frac{1}{s} \right). \quad (49)$$

\mathbf{m}_n can be expressed completely in terms of s and \mathbf{m}_t by making use of Eq. 7, so that

$$\mathbf{m}_n = \frac{\nabla s}{s} + \nabla T \frac{d}{dr} \left(\frac{1}{s} \right) \cdot \frac{d\mathbf{r}}{dl} = \frac{1}{s} (\nabla s - (\nabla s \cdot \mathbf{m}_t) \mathbf{m}_t). \quad (50)$$

If we now let \mathbf{n} represent the anti-normal unit vector to the ray path at \mathbf{r} , and substitute velocity v for slowness s , then

$$\mathbf{n} = \frac{\nabla v - (\nabla v \cdot \mathbf{m}_t) \mathbf{m}_t}{|\nabla v - (\nabla v \cdot \mathbf{m}_t) \mathbf{m}_t|} \quad (51)$$

which is equivalent to the expression derived by Um and Thurber (1987). The vector \mathbf{n} thus defines the direction of ray path curvature.

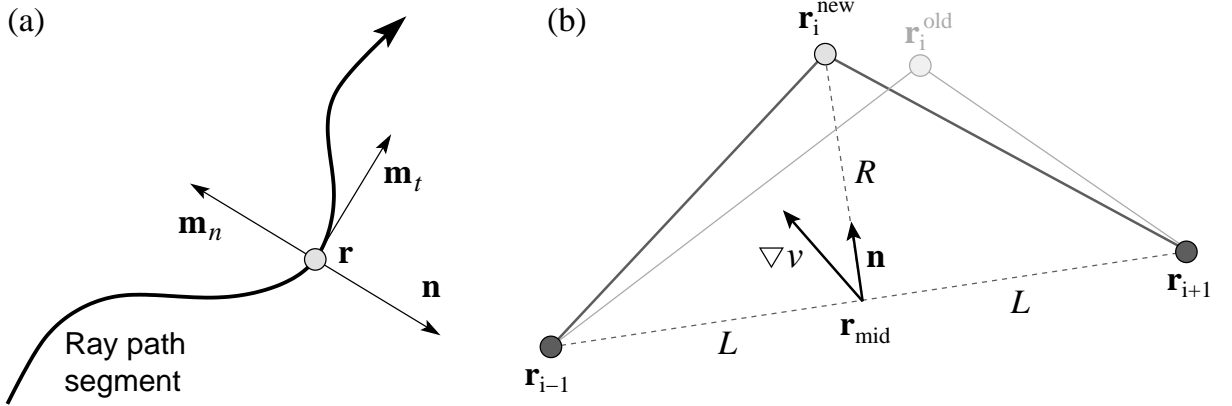


Fig. 12. Schematic representation of parameters used in the pseudo bending method. (a) Definition of ray tangent \mathbf{m}_t , ray normal \mathbf{m}_n and antinormal unit vector \mathbf{n} at a point \mathbf{r} along a ray path; (b) stencil for the three point perturbation scheme.

A three-point perturbation scheme is devised by Um and Thurber (1987) to sequentially update points along a path. Consider Fig. 12b, which shows three points \mathbf{r}_{i-1} , \mathbf{r}_i and \mathbf{r}_{i+1} . The aim is to replace the initial guess point $\mathbf{r}_i^{\text{old}}$ with an improved estimate $\mathbf{r}_i^{\text{new}}$. The improved estimate is obtained by considering a perturbation to the point \mathbf{r}_{mid} , which lies at the midpoint between \mathbf{r}_{i-1} and \mathbf{r}_{i+1} . The vector \mathbf{m}_t can then be simply approximated by

$$\mathbf{m}_t = \frac{\mathbf{r}_{i+1} - \mathbf{r}_{i-1}}{|\mathbf{r}_{i+1} - \mathbf{r}_{i-1}|} \quad (52)$$

and the anti-normal unit vector \mathbf{n} , which specifies the bending direction, is computed from Eq. 51. The next step is to find the distance R in the direction \mathbf{n} which results in

an improved estimate of the path. An approximate analytic expression for the traveltime $T(R)$ between \mathbf{r}_{i-1} and \mathbf{r}_{i+1} can be obtained using the trapezoidal rule

$$T(R) = \sqrt{L^2 + R^2} \left[\frac{1}{v_i^{\text{new}}} + \frac{1}{2} \left(\frac{1}{v_{i-1}} + \frac{1}{v_{i+1}} \right) \right]. \quad (53)$$

The appropriate value for R can be obtained by appealing directly to Fermat's principle of stationary time, which in this case equates to setting $dT/dR = 0$. Therefore,

$$\left. \begin{aligned} \frac{dT}{dR} &= \sqrt{L^2 + R^2} \frac{d}{dR} \left(\frac{1}{v_i^{\text{new}}} \right) + \left(\frac{1}{v_i^{\text{new}}} + c \right) \frac{R}{\sqrt{L^2 + R^2}} \\ &= (L^2 + R^2) \frac{dv_i^{\text{new}}}{dR} - Rv_i^{\text{new}}(cv_i^{\text{new}} + 1) = 0 \end{aligned} \right\}, \quad (54)$$

where $c = (1/v_{i-1} + 1/v_{i+1})/2$. The quantity v_i^{new} is unknown, but a first-order accurate (in R) estimate based on the velocity and velocity gradient at \mathbf{r}_{mid} can be made

$$v_i^{\text{new}} \approx v_{\text{mid}} + (\mathbf{n} \cdot \nabla v_{\text{mid}})R. \quad (55)$$

Substitution of this expression into Eq. 54 results in

$$[c(\mathbf{n} \cdot \nabla v_{\text{mid}})^2]R^3 - [2cv_{\text{mid}}(\mathbf{n} \cdot \nabla v_{\text{mid}})]R^2 - [v_{\text{mid}}(cv_{\text{mid}} + 1)]R + (\mathbf{n} \cdot \nabla v_{\text{mid}})L^2 = 0. \quad (56)$$

Um and Thurber (1987) ignore the cubic term of R , which results in a quadratic equation with two roots. The solution which produces real positive traveltimes is

$$R = \frac{-v_{\text{mid}}(cv_{\text{mid}} + 1) + \sqrt{v_{\text{mid}}^2(cv_{\text{mid}} + 1)^2 + 8cv_{\text{mid}}(\mathbf{n} \cdot \nabla v_{\text{mid}})^2L^2}}{4cv_{\text{mid}}(\mathbf{n} \cdot \nabla v_{\text{mid}})}. \quad (57)$$

The update procedure for pseudo bending as outlined above is very simple and computationally efficient, as it only requires two relatively simple equations to be solved (Eq. 51 and 57). In practice, Um and Thurber (1987) apply the update scheme simultaneously from both end points of the ray path to the central point. This process is repeated until a convergence criterion is met. An initial path can be approximated in various ways, but one simple option is to begin with a three point ray joining source and receiver. Once the central point has been perturbed, two new points are introduced that bisect each

line segment. The central three points can then be relocated, before four new points are introduced in the same way as before (see Fig. 13). This process can be continued until a suitably accurate path is obtained.

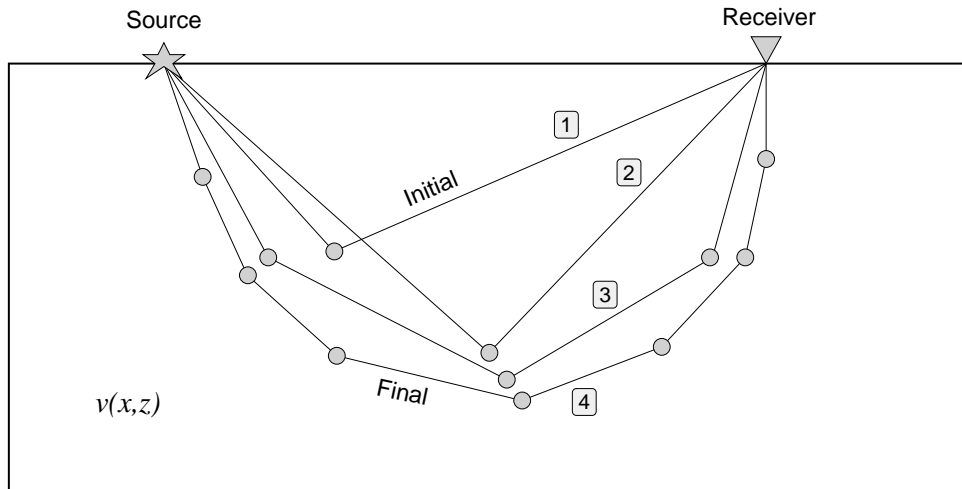


Fig. 13. Principle of the pseudo-bending method of Um and Thurber (1987). In this schematic example, an initial guess ray is defined by three points. The center point is perturbed to satisfy Fermat's principle of stationary time. The number of path segments is then doubled and the process repeated.

Despite the relatively crude approximations made in pseudo bending, Um and Thurber (1987) find it to be much more computationally efficient than conventional bending schemes; consequently, it has become quite popular for problems which require large traveltimes datasets to be predicted, e.g. 3-D local earthquake tomography (Eberhart-Phillips, 1990; Scott et al., 1994; Eberhart-Phillips and Reyners, 1997; Graeber and Asch, 1999). Zhao et al. (1992) modify the three point perturbation scheme of Um and Thurber (1987) to allow for the presence of interfaces. In this case, the sequence of points which discretely defines the path includes points which lie on each interface traversed by the ray. When the update scheme reaches these points, they are perturbed along the interface (with adjacent points on either side of the interface held fixed) until Snell's law is satisfied. Koketsu and Sekine (1998) devise a similar scheme in a 3-D spherical coordinate system.

2.2.2 Other bending schemes

Two classes of bending method have been described above. In the first, the ray equation is linearised and iteratively solved as a boundary value problem with fixed end points. The second approach - pseudo bending - applies a simple perturbation scheme based on a direct application of Fermat's principle of stationary time, to a ray path described by a sequence of points. Although these are the two principal methods, several other schemes have been developed which warrant a brief mention.

Prothero et al. (1988) develop a 3-D bending scheme based on the simplex method of function minimisation. An initial path is obtained by using an exhaustive search method to find the minimum-time circular path between source and receiver. Perturbations to this path, described by a sum of sine-wave harmonics, are then made using the simplex method, which searches for the amplitude coefficients that produce the path of least time. Although the method appears to be more robust than pseudo-bending, it is significantly slower.

Like ray shooting, ray bending can also be carried out using fully non-linear update schemes. Sadeghi et al. (1999) develop a method which uses genetic algorithms (GAs) to globally search for the minimum time ray path between two fixed points. The scheme is similar to pseudo-bending in that paths are described by a set of linearly interpolated points, which are perturbed until a convergence criterion is satisfied. However, in this case, a population of multiple two-point paths join source and receiver, and the GA drives the bending of paths until the traveltimes converges to a minimum. When this occurs, all two point rays within the population should follow almost identical paths.

Dębski and Ando (2004) develop a so-called "spectral ray tracer", which bears some resemblance to the scheme of Prothero et al. (1988), except that ray paths are parameterised as a series of Chebyshev polynomials. The bending problem can then be formulated as one of function minimisation, in which the decomposition coefficients of the Chebyshev poly-

nomials become the variables to be adjusted until the two point traveltime is minimised. Instead of adopting a linearised approach to the optimisation problem, Dębski and Ando (2004) use a genetic algorithm to generate and select the decomposition components.

Most schemes for solving the boundary value problem in ray tracing can usually be characterised as either shooting or bending. However, other schemes do exist, most notably those that are based on structural perturbation (Červený, 2001). In this type of scheme, a known two point path exists in a reference medium, and the aim is to locate the equivalent two point path in a medium that is slightly modified from the reference medium. Solution of this class of problem can be achieved using ray perturbation theory, which is described in various papers including Farra and Madariaga (1987); Snieder and Sambridge (1992); Pulliam and Snieder (1996). Snieder and Spencer (1993) show that ray bending and ray perturbation theories can in fact be combined into a single perturbation theory.

3 Grid based schemes

An alternative to tracing rays between source and receiver is to compute the traveltime of the evolving wavefront at all points of a grid which spans the medium. The complete traveltime field implicitly contains the wavefront location as a function of time (i.e. isochrons of $T(\mathbf{x})$) and all possible ray path trajectories (specified by ∇T). Compared to conventional shooting and bending methods of ray tracing (Julian and Gubbins, 1977; Cassell, 1982; Um and Thurber, 1987; Virieux and Farra, 1991; Koketsu and Sekine, 1998), grid based traveltime schemes have a number of clear advantages: (1) Most are capable of computing traveltimes to all points of a medium, and will locate diffractions in ray shadow zones; (2) the non-linearity of both ray shooting and bending means that they may fail to converge to a true two-point path, whereas most grid based schemes are highly stable and will find the correct solution even in strongly heterogeneous media; (3) grid based schemes can be very efficient in computing traveltime and path information to the level of

accuracy required by practical problems. Ray tracing schemes can be inefficient if solution non-linearity is significant; (4) most grid-based schemes consistently find first-arrivals in continuous media. It is often difficult to ascertain with ray tracing whether the located path is a first or later arrival.

Despite these advantages, grid based schemes have a number of limitations which should be considered prior to application. These include: (1) accuracy is a function of grid spacing - in 3-D halving the spacing of a grid will increase computation time by at least a factor of 8. Thus, computation time may become unacceptable if highly accurate traveltimes are required; (2) most practical schemes compute first-arrivals only - thus, features such as wavefront triplications cannot be predicted; (3) quantities other than traveltime (such as amplitude) are difficult to compute accurately without first extracting path geometry and applying ray based techniques. Two grid-based schemes - finite difference solution of the eikonal equation and shortest path methods - have emerged in the last few decades as popular alternatives to conventional ray tracing, and are described below.

3.1 Eikonal solvers

One of the first grid schemes based on finite difference solution of the eikonal equation was proposed by Vidale (1988). From a given source point, traveltimes are progressively computed outwards along an expanding square in 2-D (see Fig. 14). Traveltimes to points which lie in the initial square around the source point are computed as follows. The four points $(i \pm 1, j)$ and $(i, j \pm 1)$ have traveltimes estimated by the formulae $T_{i\pm 1, j} = \delta x / 2 (s_{i\pm 1, j} + s_{i, j})$ and $T_{i, j\pm 1} = \delta z / 2 (s_{i, j\pm 1} + s_{i, j})$, where s is slowness and δx and δz are grid spacing in x and z . The remaining four points are computed by appealing to the eikonal equation (Eq. 3); in particular, the ∇T term can be approximated in the cell defined by

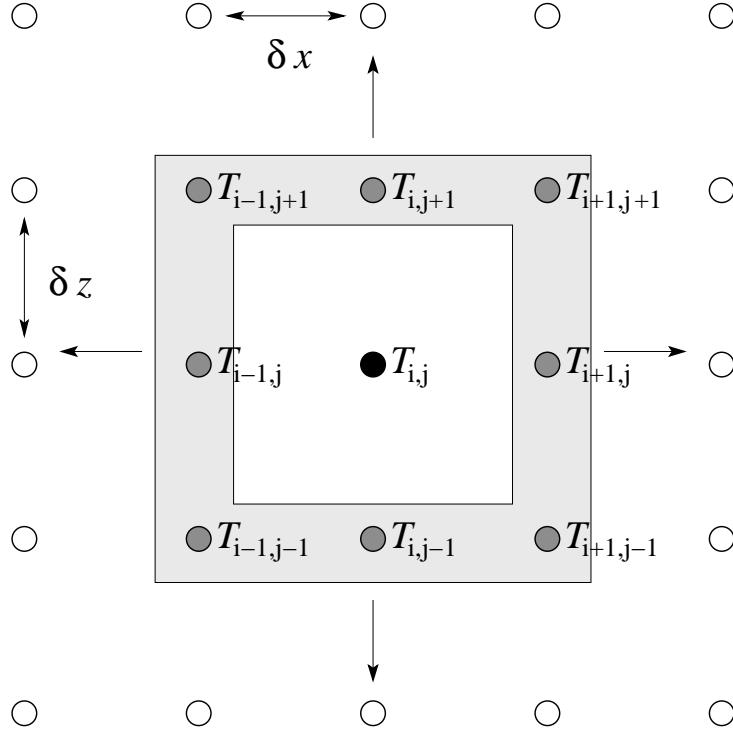


Fig. 14. The expanding square method for progressive calculation of traveltimes throughout a gridded velocity field.

points (i, j) , $(i + 1, j)$, $(i, j + 1)$ and $(i + 1, j + 1)$ by

$$\left. \begin{aligned} \frac{\partial T}{\partial x} &= \frac{T_{i,j} + T_{i,j+1} - T_{i+1,j} - T_{i+1,j+1}}{2\delta x} \\ \frac{\partial T}{\partial z} &= \frac{T_{i,j} + T_{i+1,j} - T_{i,j+1} - T_{i+1,j+1}}{2\delta z} \end{aligned} \right\}. \quad (58)$$

Substitution into Eq. 3 produces the following quadratic equation

$$\frac{(T_{i,j} + T_{i,j+1} - T_{i+1,j} - T_{i+1,j+1})^2}{\delta x^2} + \frac{(T_{i,j} + T_{i+1,j} - T_{i,j+1} - T_{i+1,j+1})^2}{\delta z^2} = 4\bar{s}^2, \quad (59)$$

where \bar{s} is the average slowness of all four points defining the cell. This equation can easily be solved for $T_{i+1,j+1}$. If $h = \delta x = \delta z$ (Vidale, 1988), then the solution to Eq. 59 has the simple form $T_{i+1,j+1} = T_{i,j} + \sqrt{2(h\bar{s})^2 - (T_{i,j+1} - T_{i+1,j})^2}$. Vidale (1988) also defines a solution stencil for a locally circular wavefront. This allows a mixed scheme to be devised, which uses the locally circular assumption in regions of high wavefront curvature (e.g. in the source neighbourhood), and Eq. 59 when the wavefront is more planar.

As the computational front evolves outward from the source, points within the square

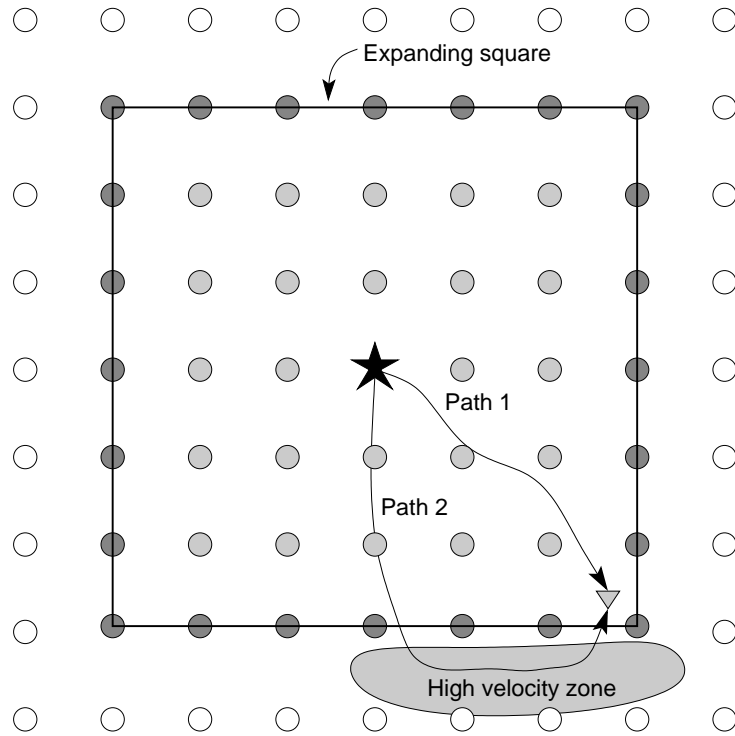


Fig. 15. Schematic illustration of how the expanding square method can fail. In this example, path 1 would be computed by the expanding square, but path 2 actually arrives first.

band cannot have traveltimes computed in arbitrary order; causality at least requires that new traveltimes be computed using only those traveltimes from surrounding points of lesser value. Vidale (1988) devises a scheme which sweeps through each of the four sides of the square in order to locate the minimum time solution at each point. The resulting computational scheme is both fast and accurate, with CPU time being approximately proportional to the number of points defining the grid. The method is readily extendable to 3-D, as demonstrated by Vidale (1990).

The use of an expanding square formalism to define the shape of the computational front cannot always respect the direction of flow of traveltime information through the medium. This is demonstrated in Fig. 15, which shows how the expanding square can fail if the first-arriving path to a point inside the square needs to sample structure outside the square. As a consequence, first-arrivals are not always guaranteed, which can ultimately lead to instability. Nevertheless, the basic scheme proposed by Vidale (1988) remains popular, and its stability has been improved upon by introducing new features to the

finite difference stencil and evolution of the computational front. Both Hole and Zelt (1995) and Afnimar and Koketsu (2000) introduce special headwave operators to better deal with the presence of strong velocity contrasts, and Hole and Zelt (1995) implement an iterative post-sweeping scheme to help account for the non-causal nature of the expanding square. van Trier and Symes (1991) use entropy-satisfying first-order upwind difference operators to improve the computational efficiency of the method and better deal with wavefront discontinuities. Comparable improvements are made by Podvin and Lecomte (1991), who systematically apply Huygen’s principle in the finite difference approximation in order to account for all possible modes of propagation.

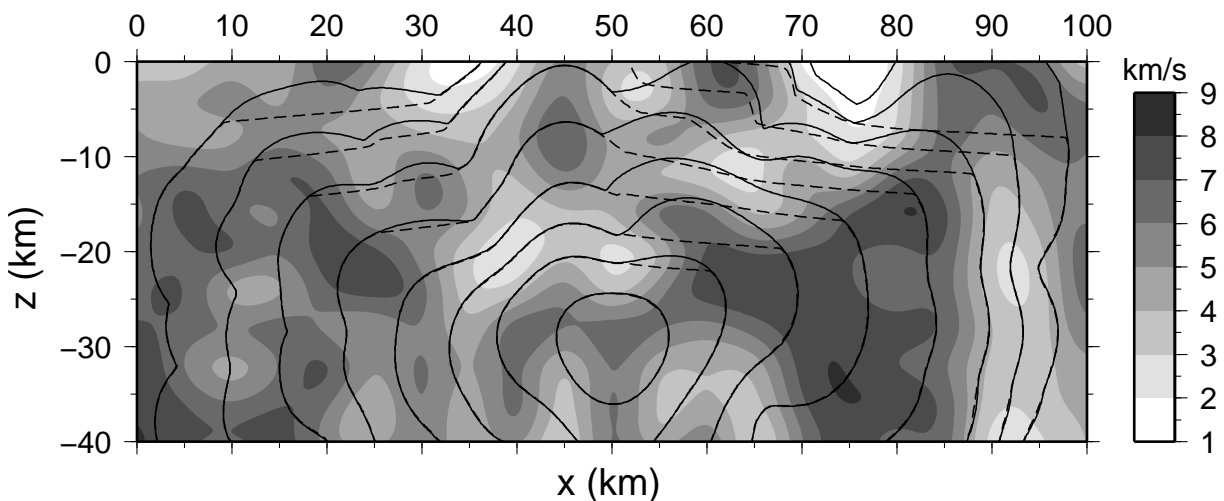


Fig. 16. Illustration of how post-sweeping can correct for errors due to the use of an expanding box strategy for computing traveltimes. Wavefronts for WENO-DNO denoted by dashed lines; wavefronts for WENO-DNO-PS denoted by solid lines.

More recently, Essentially Non-Oscillatory (ENO) finite difference schemes developed in the field of computational mathematics (Shu and Osher, 1988, 1989) have been used to solve the eikonal equation within an expanding square framework. The attraction of ENO schemes is that they can be readily extended to very high orders of accuracy, yet remain stable. Kim and Cook (1999) present a scheme, which they describe as ENO-DNO-PS, to efficiently compute first-arrival traveltime fields. DNO (or *down 'n' out*) refers to the expanding box scheme first introduced by Vidale (1988), and PS refers to post-sweeping, which is an iterative correction strategy applied to the traveltime field computed by

ENO-DNO in an attempt to deal with causality breaches introduced by the expanding box. This correction scheme is similar in principle to the one used by Hole and Zelt (1995). Fig. 16 demonstrates the importance of post-sweeping in complex models; the expanding box results in the overestimation of traveltimes in various regions of the model (dashed lines), which appear to be corrected by the post-sweeping step (solid line). A third-order WENO-DNO-PS solver is used in this example, which is slightly different to that used by Kim and Cook (1999). Weighted Essentially Non-Oscillatory or WENO schemes have a number of advantages over ENO including computation time and stability (Liu et al., 1994; Jiang and Shu, 1996; Jiang and Peng, 2000).

Other authors to use ENO-type schemes for seismic traveltimes include Qian and Symes (2002), who implement a WENO scheme with adaptive gridding, and Buske and Kästner (2004), who use an ENO scheme in polar coordinates; in both cases, the computed traveltimes are sufficiently accurate to solve the transport equation (Eq. 4) and obtain amplitudes.

Rather than use an expanding square as the computational front for traveltimes, it is also possible to use the expanding wavefront itself (Qin et al., 1992; Cao and Greenhalgh, 1994). In the case of Qin et al. (1992), the finite difference stencil proposed by Vidale (1988) is retained, and traveltimes to the first band of points surrounding the source are computed in the same way. However, thereafter the point of global minimum traveltime along the perimeter of points computed so far is used as the next source to locally expand the solution region. This approach ensures that the shape of the computational front conforms to that of the first-arriving wavefront, which minimises the possibility of computing arrivals other than the first. The need to locate the global minimum traveltime along the computational front requires an ordered storage of traveltimes (e.g. heap sorting), which is less efficient than the expanding box regime.

Eikonal solvers have been used in a variety of seismological applications, and are particularly useful for those classes of problems that require large traveltime datasets to be

predicted, such as seismic imaging. For example, they have been frequently used in 3-D refraction/reflection tomography (Hole, 1992; Zelt, 1996; Riahi et al., 1997; Zelt and Barton, 1998; Zelt, 1999; Zelt et al., 2001; Day et al., 2001) and in the migration of coincident reflection sections (Gray and May, 1994; Bevc, 1997; Buske, 1999a,b).

3.1.1 The Fast Marching Method

One of the more recently developed grid based eikonal solvers which is both highly robust and computationally efficient is the so called Fast Marching Method (Sethian, 1996, 1999; Popovici and Sethian, 2002) or FMM. It was originally developed in the field of computational mathematics for solving various types of interface evolution problems, and to date has been applied in numerous areas of the physical sciences including optimal path planning, medical imaging, geodesics, and photolithographic development (Sethian, 1999, 2001). In seismology, FMM has been used in the migration of coincident reflection profiles (Popovici and Sethian, 2002) and teleseismic tomography (Rawlinson et al., 2006a,b).

A common feature of first-arrival traveltimes fields, particularly in complex media, is that they are not spatially differentiable at every point (i.e. the ∇T term in Eq. 3 is not defined everywhere). This can lead to instability in schemes that do not implicitly or explicitly recognise this behaviour (e.g. Vidale, 1988; Qin et al., 1992). When a wavefront self-intersects (i.e. multipathing occurs), the first-arrival wavefront will contain a kink or discontinuity which would normally be spanned by the later arriving wavefront. However, since only first-arrival traveltimes are computed, this information is discarded, and to ensure stability, it becomes important not to use traveltimes from both sides of the discontinuity in the same finite difference stencil to compute new traveltimes. One way to overcome this problem is to solve the viscous version of the eikonal equation, which smooths out discontinuities; the limit of smooth solutions is a weak solution that corresponds to the first arriving wavefront. It turns out that the viscous limit solution can also be obtained by solving Eq. 3 using upwind entropy-satisfying operators which take into

account the direction of flow of information when evaluating ∇T .

A commonly used entropy-satisfying upwind scheme (e.g. Sethian and Popovici, 1999; Chopp, 2001; Popovici and Sethian, 2002) may be expressed as follows

$$\left[\begin{array}{l} \max(D_a^{-x}T, -D_b^{+x}T, 0)^2 + \\ \max(D_c^{-y}T, -D_d^{+y}T, 0)^2 + \\ \max(D_e^{-z}T, -D_f^{+z}T, 0)^2 \end{array} \right]_{ijk}^{\frac{1}{2}} = S_{i,j,k}, \quad (60)$$

where (i, j, k) are grid increment variables in any orthogonal coordinate system (x, y, z) , and the integer variables a, b, c, d, e, f define the order of accuracy of the upwind finite difference operator used in each of the six cases. For example, in a Cartesian coordinate system, the first and second order operators for $D^{-x}T_i$ are

$$\begin{aligned} D_1^{-x}T_{i,j,k} &= \frac{T_{i,j,k} - T_{i-1,j,k}}{\delta x} \\ D_2^{-x}T_{i,j,k} &= \frac{3T_{i,j,k} - 4T_{i-1,j,k} + T_{i-2,j,k}}{2\delta x} \end{aligned} \quad (61)$$

where δx is the grid spacing in x . First order accurate schemes only use D_1 operators and second order accurate schemes preferentially use D_2 operators. Strictly speaking, the second order method is really mixed order because it will use first order approximations when causality does not permit the use of the required operator. For example, if we implement a second order method and $T_{i-1} > T_{i-2}$, then the operator used would be $D_2^{-x}T_i$; while if $T_{i-1} < T_{i-2}$ we would have to resort to $D_1^{-x}T_i$. Mixed schemes which use D_3 or even higher order operators could also be devised, but in practice, the need to resort to D_1 operators on occasion means that improvements in overall accuracy are usually not substantial (Rawlinson and Sambridge, 2004b). It should also be noted that the first order FMM scheme has been proven to be unconditionally stable (Sethian, 2001), but no such proof exists for higher order schemes.

Fig. 17 schematically illustrates why the entropy-satisfying upwind finite difference stencil of Eq. 60 preserves stability in the presence of a wavefront discontinuity. In this case,

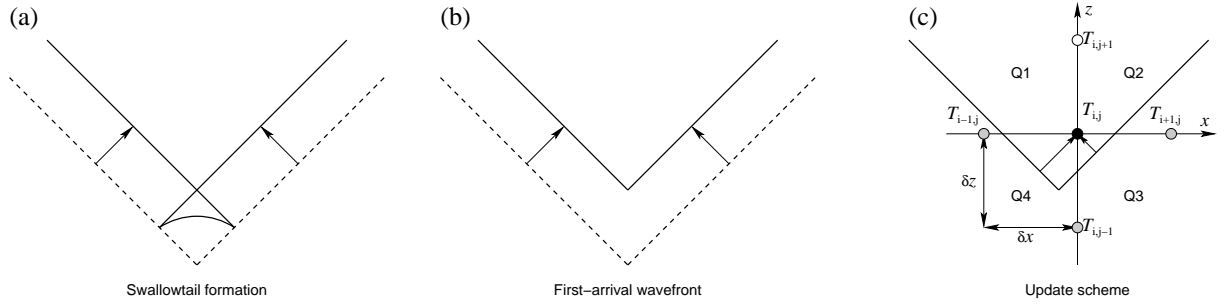


Fig. 17. Discontinuities in first-arrival wavefronts usually arise from discarding later-arriving information. (a) Formation of a swallowtail from an initial wavefront (dashed line); (b) the first-arrival wavefront contains a discontinuity; (c) the upwind entropy-satisfying finite difference stencil properly respects the flow of information by considering solutions from each quadrant.

the complete wavefront includes a swallowtail, but this is discarded and the first-arrival wavefront propagates with a discontinuity. The solution of Eq. 60, in 2-D media with first-order operators, can be obtained by considering each quadrant of Fig. 17c, where the aim is to compute $T_{i,j}$ from known traveltimes at surrounding points. In Q1 and Q2, the only approximations that can be made are $T_{i,j} = T_{i-1,j} + \delta x s_{i,j}$ and $T_{i,j} = T_{i+1,j} + \delta x s_{i,j}$ respectively, which are not even first-order accurate. However, in Q3 and Q4, which both have two known traveltimes, true first-order accurate traveltimes can be obtained. In the case of Q3, the quadratic equation is

$$\left(\frac{T_{i+1,j} - T_{i,j}}{\delta x}\right)^2 + \left(\frac{T_{i,j} - T_{i,j-1}}{\delta z}\right)^2 = (s_{i,j})^2. \quad (62)$$

Of the two possible solutions, the larger one corresponds to the correct local plane wave approximation. Similarly in quadrant Q4, the quadratic equation is

$$\left(\frac{T_{i,j} - T_{i-1,j}}{\delta x}\right)^2 + \left(\frac{T_{i,j} - T_{i,j-1}}{\delta z}\right)^2 = (s_{i,j})^2, \quad (63)$$

where again the larger solution corresponds to the correct local plane wave approximation. Of the four possible solutions that have been identified, the one with minimum traveltime is correct; in this case, it is the plane wave solution from Q3. The extension to 3-D and higher-order operators is straightforward, although the number of candidate solutions obviously increases.

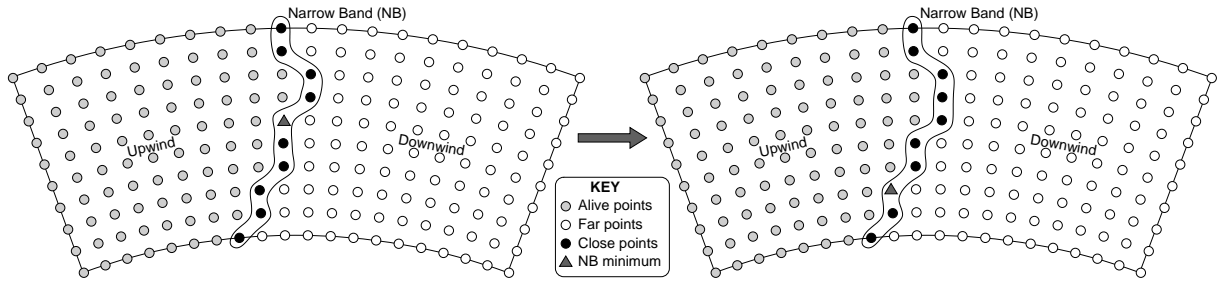


Fig. 18. Narrow band (NB) evolution scheme used for the ordered update of grid points. *Close* points have trial traveltimes computed using Eq. 60 (using appropriate operators for polar coordinates in this case). The narrow band locally advances from the *close* point with minimum traveltime.

The FMM stencil encapsulated in Eq. 60 describes how to calculate new traveltimes using known traveltimes from neighbouring grid points, but in order to populate a grid of nodes with traveltime values, the order in which nodes are updated must be consistent with the direction of flow of information; that is, from smaller values of T to large values of T . FMM achieves this by systematically constructing traveltimes in a downwind fashion from known values upwind by employing a *narrow band* approach. This is illustrated in Fig. 18. *Alive* points have their values correctly calculated, *close* points lie within the narrow band and have trial values computed using Eq. 60 with *alive* points only, and *far* points have no values calculated. The narrow band is evolved by identifying the *close* point with minimum traveltime, tagging it as *alive* and then updating any adjacent *close* or *far* point, the latter being re-tagged as *close*. Using this approach, the shape of the narrow band approximates the shape of the first-arrival wavefront, and the idea is to propagate the band through the grid until all points become alive. This basic approach is similar to that advocated by Qin et al. (1992). In fact, looking back at the literature produced by the seismology community on grid-based eikonal solvers, it is evident that all the ideas on which FMM is based have been considered at one time or another, but had not previously been combined in this particular way.

The use of a heap sort algorithm to identify the node with global minimum traveltime within the narrow band means that FMM has an operation count of $O(M \log M)$, where M is the total number of grid points. Thus, in order to achieve greatest efficiency in

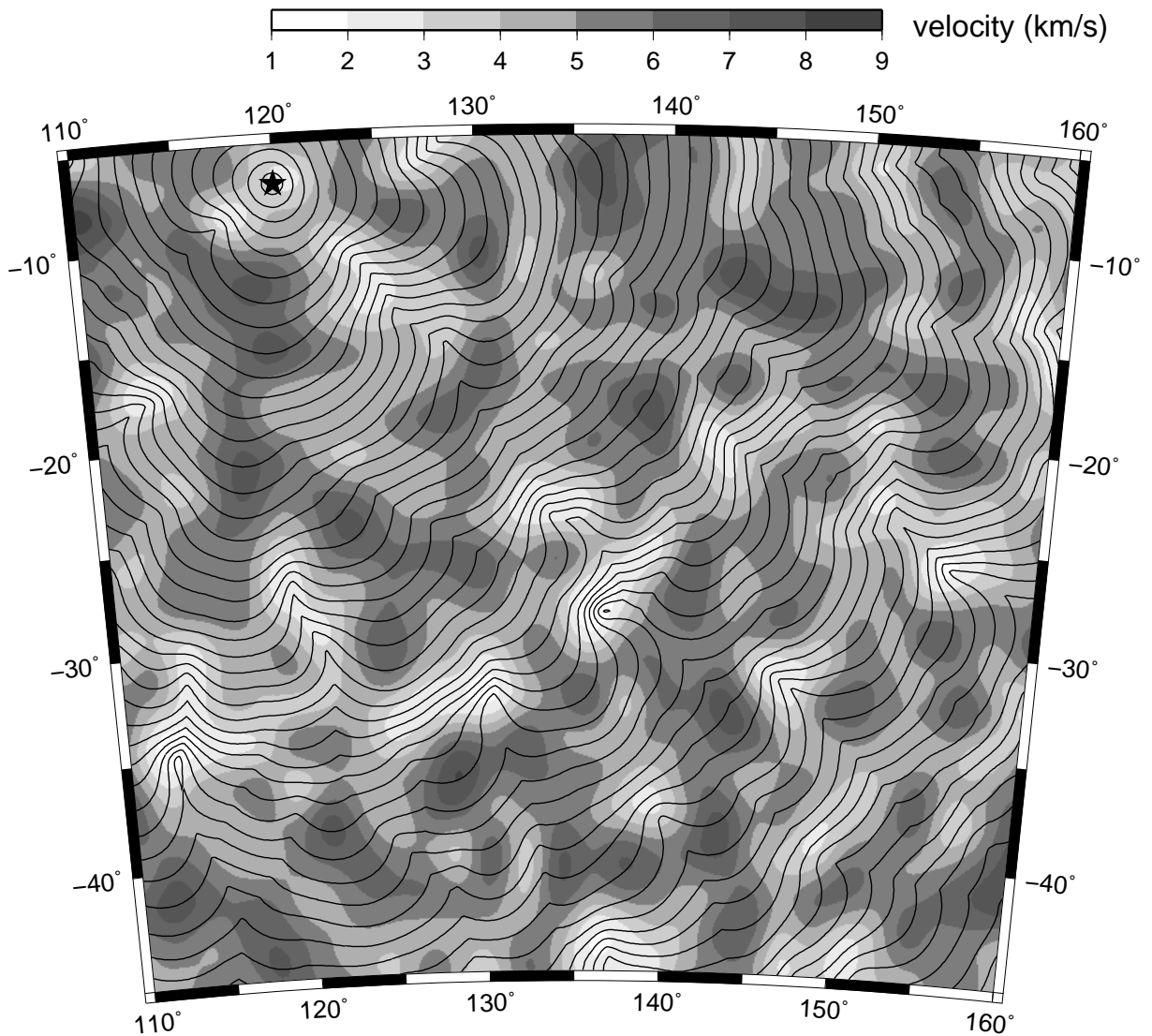


Fig. 19. Wavefronts implicitly computed by FMM in a complex velocity field. Note the formation and propagation of wavefront discontinuities. Wavefronts are contoured at 20 s intervals, and the synthetic model is in spherical shell coordinates.

practical applications (particularly in 3-D), it is important to choose the largest grid spacing permitted by the uncertainty in the data.

The stability of FMM is demonstrated by Fig. 19, which shows the evolution of a wavefront in a highly complex medium. In this case, FMM has been solved in spherical shell coordinates using the appropriate spherical form of ∇T in Eq. 3 with constant radius r . This particular implementation would be useful for tracking the evolution of higher frequency surface waves e.g. for surface wave tomography using ambient seismic noise (Shapiro et al., 2005; Yao et al., 2006). The velocity variations shown in Fig. 19 are somewhat pathologi-

cal, but the example seeks to demonstrate the robustness of FMM. Discontinuities in the first-arriving wavefront clearly develop, but are successfully evolved without the introduction of instabilities. Further FMM examples and detailed analysis of computational efficiency and accuracy can be found in Rawlinson and Sambridge (2004b,a); de Kool et al. (2006).

3.1.2 *Improving accuracy in the source neighbourhood*

A point source is an upwind singularity of the travelttime field and can be a major contributor to the overall error of finite difference eikonal solvers due to high wavefront curvature in the source neighbourhood. Schemes that use regular travelttime grids often poorly approximate this curvature, and the resulting error is carried forward in subsequent calculations. This problem has been recognised and addressed in a variety of ways. As alluded to earlier, Vidale (1988) formulates two finite difference stencils, one that is most accurate for locally plane wavefronts, and one that is most accurate for locally circular wavefronts. Using the circular finite difference scheme in the source vicinity would largely address this problem, although it is difficult to gauge the stability of this approach.

Other schemes for addressing near-source error include using spherical grids centered on the source point (Alkhalifah and Fomel, 2001), and local grid refinement in the source neighbourhood (Kim and Cook, 1999). In the latter case, grid spacing is progressively increased away from the source as wavefront curvature decreases. Adaptive gridding with grid refinement and coarsening based on *a posteriori* error estimation (Qian and Symes, 2002) has also been successfully used to minimise near source error. Another approach is to solve the eikonal equation in the so-called celerity domain. Celerity is defined as the distance from the source divided by the travelttime from the source, and transformation of the problem into this domain can significantly reduce source error (Zhang et al., 2005).

In the context of FMM, Rawlinson and Sambridge (2004a) develop a grid refinement

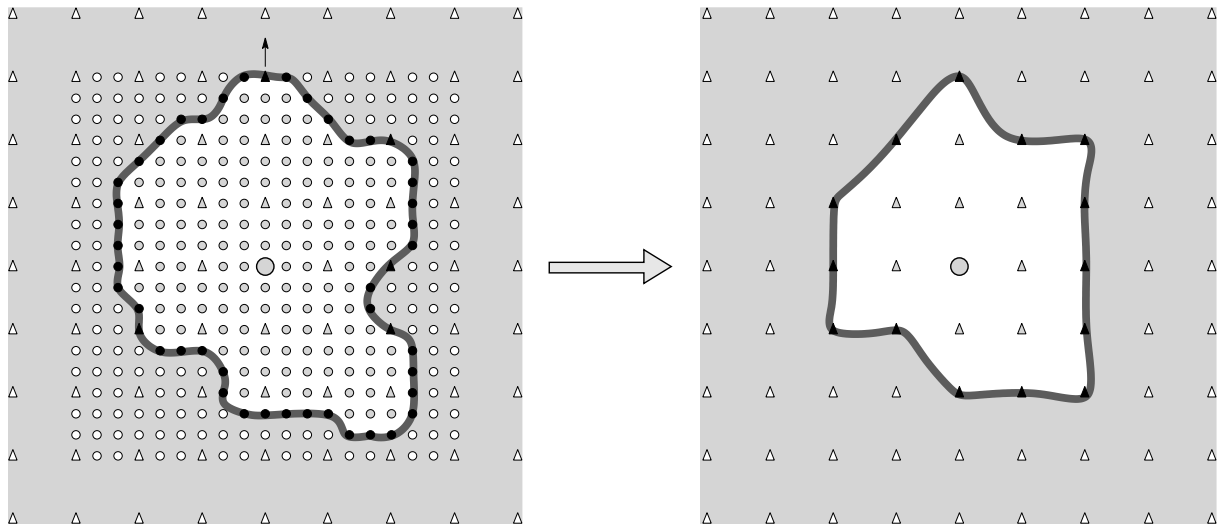


Fig. 20. Implementation of source-grid refinement when the computational front approximates the first-arrival wavefront. When the narrow band impinges on the edge of the refined grid, it is mapped onto the underlying coarse grid before continuing to evolve.

strategy that significantly improves the CPU time versus accuracy trade-off. The scheme is similar to that of Kim and Cook (1999), except that only a single level of refinement is used, and the edge of the refined grid ultimately conforms to the shape of the first-arrival wavefront. This latter property ensures that the stability of the scheme is not compromised. Fig. 20 schematically illustrates how the computational front evolves through the refined grid, then is mapped back onto the coarser grid as soon as the wavefront reaches the edge of the refined region. Rawlinson and Sambridge (2004a) show that in 2-D, accuracy can be improved by an order of magnitude with little increase in computing time.

3.1.3 Reflections and transmissions

Grid-based eikonal solvers have largely been used to compute first-arrival traveltimes in continuous media, and in fact, most methods cannot be naturally extended to track later arrivals, as they achieve their efficiency by only computing a single valued traveltime field. However, it is possible to compute reflection and transmission phases, thereby allowing much more of the complete wavefield to be simulated. Podvin and Lecomte (1991) and Riahi and Juhlin (1994) find reflections by tracking first-arrival traveltime fields from both source and receiver to the entire interface. Fermat's principle of stationary time can then

be applied to locate reflection points along the interface. Although this scheme has the advantage that multiple reflection paths can be found for a single interface, traveltime fields need to be computed for all sources and receivers. Hole and Zelt (1995) overcome this problem by explicitly applying Snell’s law in a local plane wave approximation in the neighbourhood of an interface.

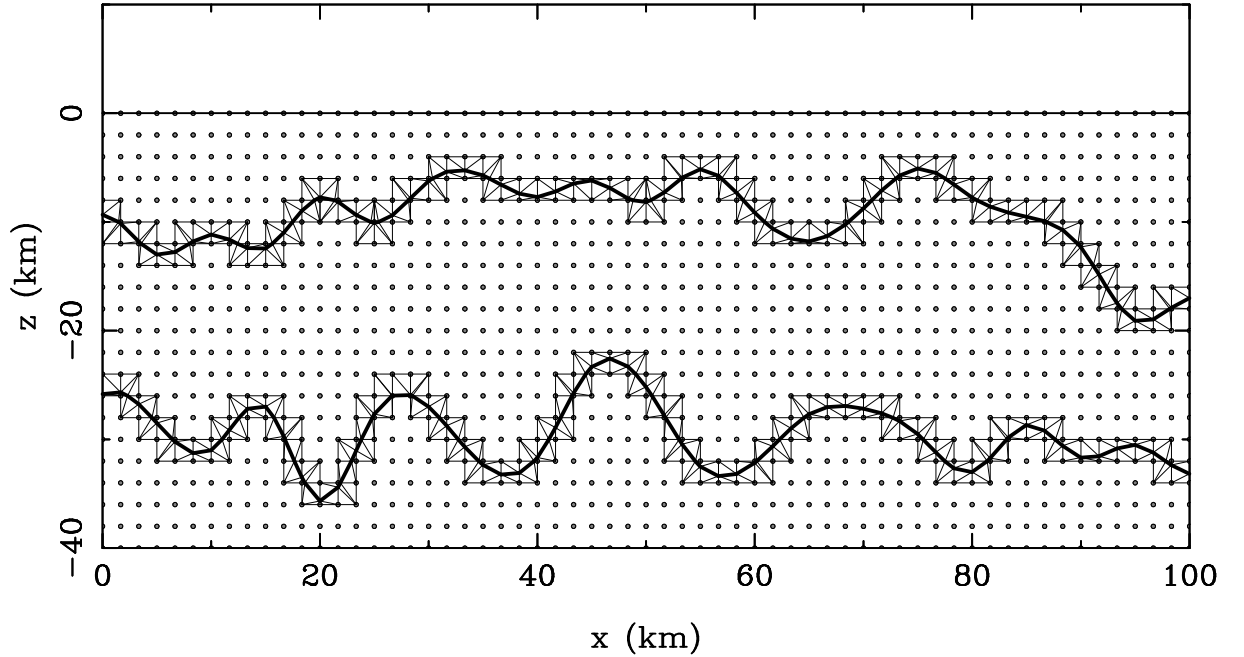


Fig. 21. An adaptive triangular mesh can be used to locally suture the irregular interface nodes to the regular nodes of the velocity grid (grey dots). To aid visualisation, the grid is coarser than would be used in practice. Refer to Rawlinson and Sambridge (2004b,a) for more details.

In an alternative approach, Li and Ulrych (1993) compute reflected and refracted traveltimes in two dimensions by using a local regridding technique to decompose a cell containing an interface into several rectangular and triangular cells so that the true interface shape is better represented by the computational grid. The incident traveltime field is computed using the scheme of Vidale (1988), and the reflected traveltime field is obtained by reinitialising the computational front from the point of minimum traveltime on the interface. The multi-stage FMM scheme of Rawlinson and Sambridge (2004b,a) is similar in principle to the Li and Ulrych (1993) approach, but has been successfully applied to much more complex velocity models, and allows paths composed of any number

of reflection and transmission segments to be tracked. Two main difficulties need to be overcome when reflection and refraction phases are computed: (1) accurate description of the embedded interface; (2) correct propagation of refracted and reflected wavefronts. In a heterogeneous layered media, interfaces vary with depth, and therefore do not conform to the regular distribution of grid nodes used by the velocity field (of course, completely irregular meshes could be used, but ease of implementation and computational cost can suffer). This can be overcome by using a locally adaptive mesh of triangles which sutures regular velocity nodes to irregular interface nodes, as illustrated in Fig. 21. Within the irregular mesh, a first-order accurate scheme for triangular elements is used to update traveltimes; elsewhere, the usual regular scheme (e.g. Eq. 60) can be applied.

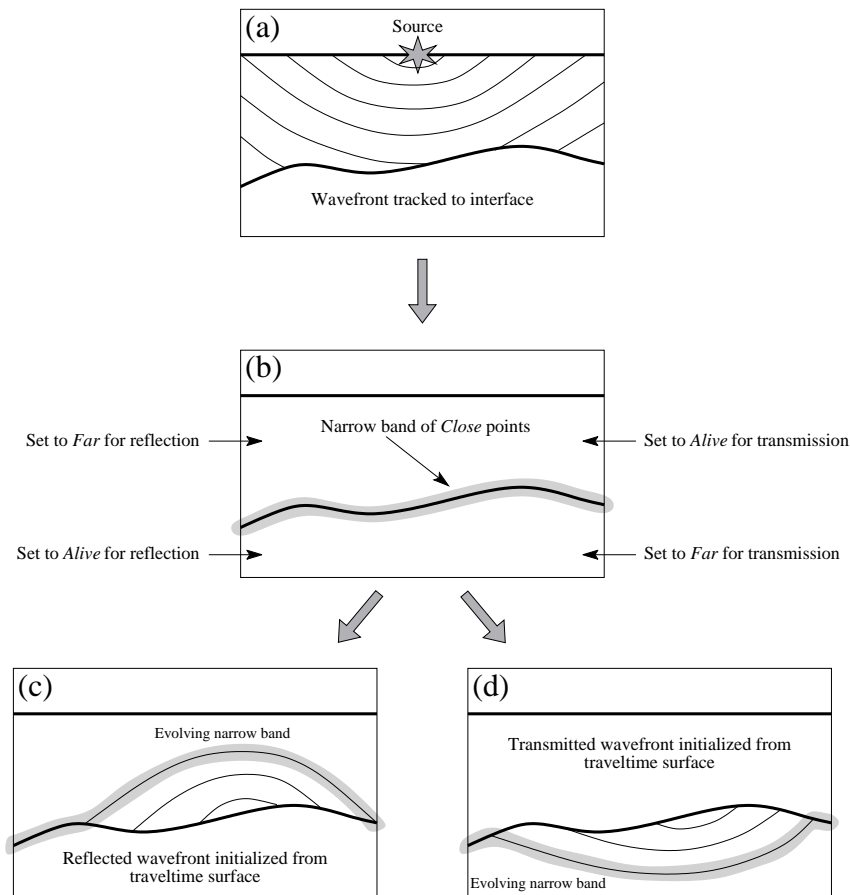


Fig. 22. Principle behind the multi-stage FMM scheme. Each layer represents a separate computational domain in which FMM is initialised either from a source or a layer boundary. (a) Incident wavefront generated from a point source; (b) narrow band defined by a set of interface nodes; (c) reflected wavefront tracked; (d) refracted wavefront tracked.

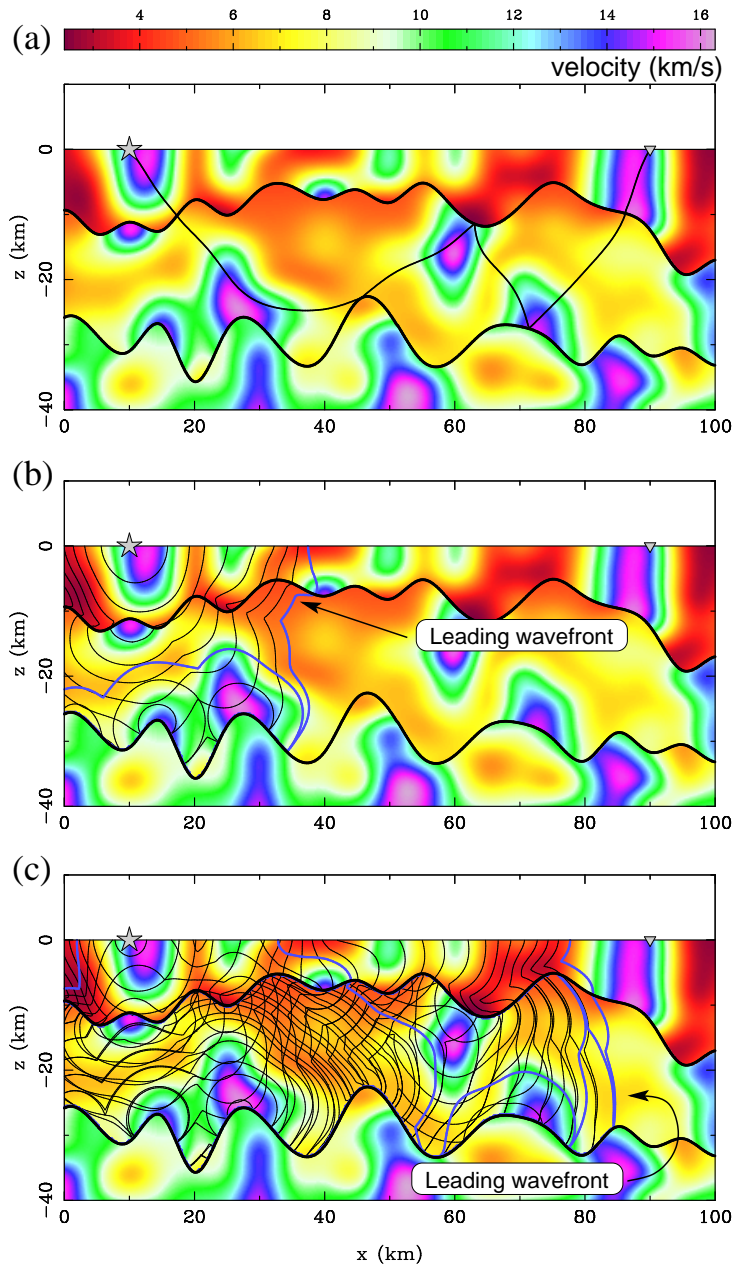


Fig. 23. A reflection multiple computed using the multi-stage FMM scheme of Rawlinson and Sambridge (2004a). (a) A two point ray path corresponding to the reflection multiple, computed after the completion of the multi-stage FMM; (b) snapshot of the evolving wavefront. Isochrons are contoured at 0.4 s intervals, and the leading wavefront is shown in blue; (c) same as (b) but at a later point in time.

To understand how more than one arrival can be tracked with FMM, consider Fig. 22a, which shows a wavefront propagating through a layer from a point source and impinging on an interface. Rather than continuing to evolve the narrow band, the wavefront is tracked only as far as the interface, which is treated as one of the four boundaries

of the computational domain. Once all points on the interface become *alive*, the FMM process is terminated. The next step is to track a reflected or transmitted wavefront into the top or bottom layer respectively; it turns out that this can be done with first-order accuracy by using only the traveltimes values of nodes which lie on the interface, with no explicit application of Snell's law. Thus, in order to track a reflection, all nodes in the top layer are set to *far* and all interface nodes are set to *close*. Therefore, the starting narrow band conforms to the shape of the interface, and FMM can be reinitialised. Using a starting narrow band that does not conform to the shape of the first-arriving wavefront is acceptable provided it is understood that the wavefront cannot reflect from or transmit through the interface more than once. A refracted wavefront can be tracked in the same way, although in this case all nodes in the bottom layer are set to *far*. This basic procedure can be applied iteratively to track phases comprising any number of reflected and transmitted phases in media of virtually any complexity. It should be noted, however, that the arrival found in each case is the absolute first-arrival of the specified phase, and is therefore likely to have small amplitude, particularly if it has experienced multiple interactions with complex interfaces and velocity gradients.

Fig. 23 shows an example of a reflection multiple tracked in the presence of a highly heterogeneous velocity model. As in Fig. 19, the velocity variations are pathological relative to real earth structure, but the purpose of the example is to demonstrate the robustness of the method. The two-point ray-path (Fig. 23a) favours fast regions of the model and preferentially reflects from convex regions of the interface; this is due to the fact that the first-arrival of this class of phase is tracked. The wavefront becomes more complex as it evolves (Fig. 23b,c) in response to strong velocity contrasts and interface geometry, but the multi-stage FMM remains stable. This stability is highly desirable in many applications including seismic tomography and non-linear earthquake relocation, which typically require many data prediction steps in order to obtain a solution.

In a recent paper, de Kool et al. (2006) extend the multi-stage FMM scheme to 3-D

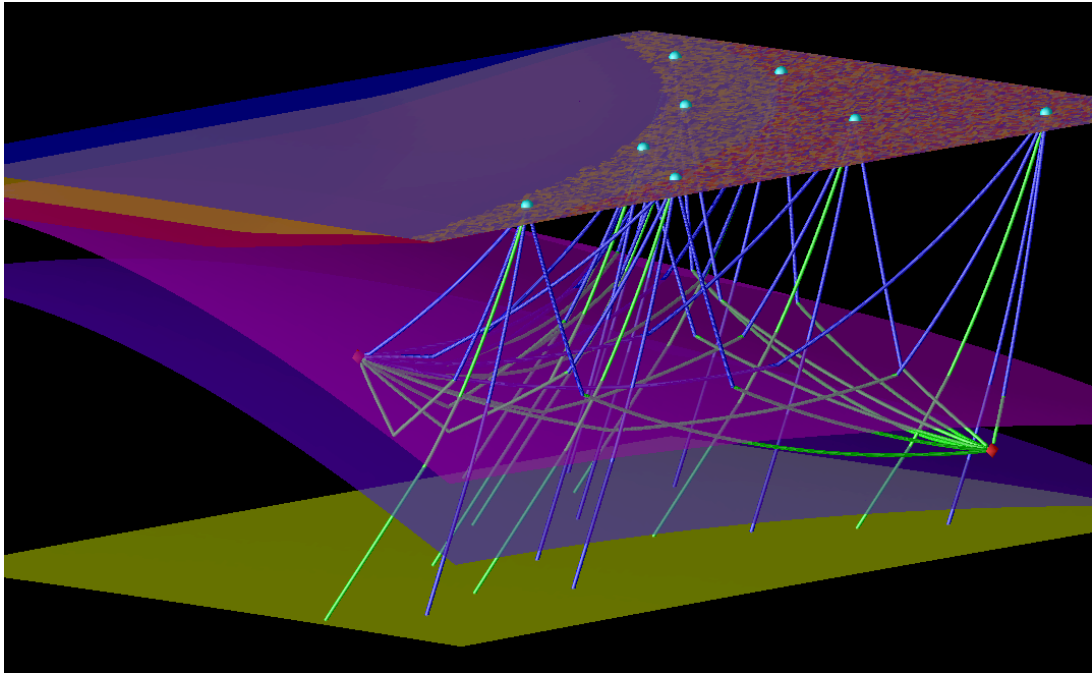


Fig. 24. Ray paths from local earthquake and teleseismic sources, computed using the multi-stage FMM scheme of de Kool et al. (2006), in the presence of a subduction zone. Receivers are denoted by blue spheres and local earthquake sources by red diamonds. P-wave paths are blue, and S-wave paths are green. S to P conversions occur at the upper interface of the dipping slab for the local earthquake sources.

spherical coordinates. The new implementation shares many of the basic principles used in the 2-D Cartesian version of Rawlinson and Sambridge (2004a), but several improvements have been implemented. These include discarding the explicit irregular mesh used to suture interface nodes to neighbouring velocity nodes; in 3-D one would need to use an irregular tetrahedral mesh, which can get complicated, particularly for interfaces with high curvature. Instead, a procedure based on testing all combinations of nodes (up to three at a time) adjacent to the node to be updated (interface node or velocity node adjacent to the interface) is used. Each combination of nodes allows a plane wave approximation to be made, and the one that yields the minimum traveltime is chosen. Another procedure built into the new 3-D method allows later-arriving reflections such as the global PP phase to be tracked. This is done by initiating FMM from both source and receiver and matching reflection points using Fermat's principle of stationary time. The flexibility of the new scheme is demonstrated in Fig. 24, which shows ray paths from teleseismic

and local earthquake sources (extracted from traveltimes fields computed using the multi-stage FMM) in a subduction zone setting. [The 3-D FMM software is freely available to download from <http://rses.anu.edu.au/seismology/fmmcode>.]

3.2 Shortest path ray tracing

Shortest path ray tracing or SPR is another popular method for determining first-arrival traveltimes at all points of a gridded velocity field (Nakanishi and Yamaguchi, 1986; Moser, 1991; Fischer and Lees, 1993; Cheng and House, 1996; Zhao et al., 2004; Zhou and Greenhalgh, 2005). Rather than solve a differential equation, a network or graph is formed by connecting neighbouring nodes with traveltimes path segments. Dijkstra-like algorithms can then be used to find the shortest path between a given point and all other points in the network. According to Fermat's principle of stationary time, the shortest time path between two points corresponds to a true ray path.

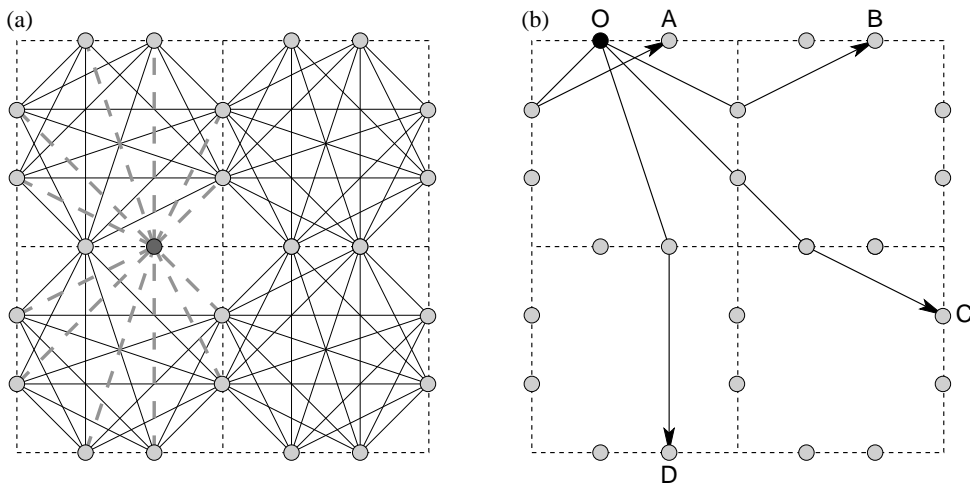


Fig. 25. Shortest path network built on a grid of constant velocity cells (Nakanishi and Yamaguchi, 1986; Moser, 1991). In this example, two network nodes are placed on the edge of each cell boundary. (a) All allowable path connections between nodes. Thick dashed grey lines highlight the connections from a single node; (b) shortest paths between node O and a selection of surrounding nodes in a homogeneous medium.

Shortest path networks are commonly defined in terms of either a cell or a grid centered

framework. In a seminal paper by Nakanishi and Yamaguchi (1986), the velocity field is defined by a set of constant velocity cells with network nodes placed on the interface between each cell (see Fig. 25a). The advantage of this approach is that traveltimes between each node pair can be easily evaluated as $t = ds$ where d is the distance between the two nodes and s is the slowness of the cell containing the ray segment. Accuracy may be increased by either reducing cell size or increasing the number of nodes on cell edges. Fig. 25b illustrates a selection of minimum time ray paths from one of the nodes to several other nodes in a constant velocity medium. This example provides some intuitive understanding of the sources of error inherent to this approach, as the true paths would be straight lines. Paths such as OC and OD are relatively straight, but OB and in particular OA are poor approximations due to the absence of connections between nodes on the same cell surface.

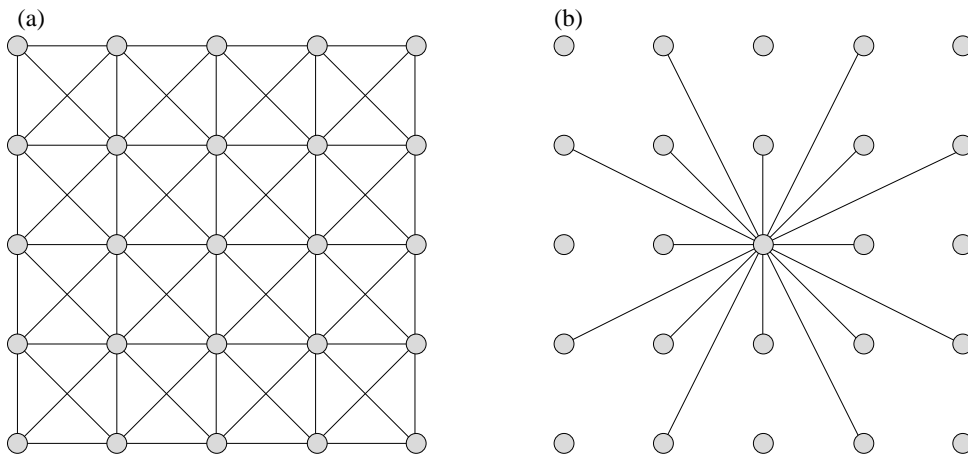


Fig. 26. Shortest path network built on a grid of velocity nodes (Moser, 1991). (a) A grid of 25 nodes with at most 8 connections per node; (b) increasing the number of connections allows smaller path deviations to be more accurately represented.

Another way of creating a network is to use a regular grid of velocity nodes and form linear connections between adjacent velocity nodes (Moser, 1991), as shown in Fig. 26a. The travelttime between two connected nodes A and B can be simply approximated by $t = d(s_A + s_B)/2$ where s_A and s_B are the slowness at nodes A and B respectively. The angular distance between branches emanating from a single node in Fig. 26b is 45° , which is relatively large. This can be reduced by increasing the number of node connections,

as demonstrated in Fig. 26b. The connection stencil for a particular node is sometimes referred to as the “forward star” (Klimeš and Kvasnička, 1994). In the case of Fig. 26a, the forward star has 8 connections, while in Fig. 26b the forward star has 16 connections. The advantage of a grid formalism for the shortest path network is that continuously varying velocity fields can be more accurately represented, and interfaces are more easily inserted.

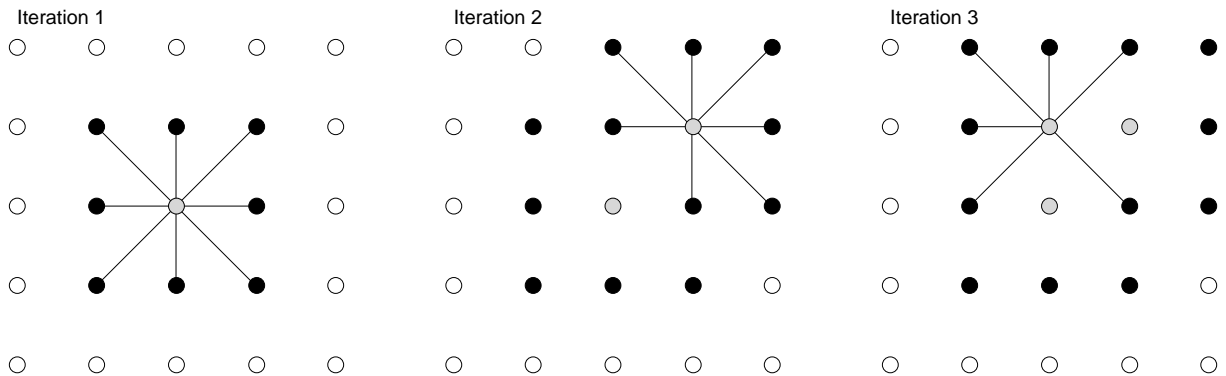


Fig. 27. Three iterations of a simple shortest path scheme using a forward star with 8 connections. Grey dots have known traveltimes, black dots have trial traveltimes, and white dots are yet to have traveltime computed.

Once a network structure and method of traveltime determination between node pairs has been chosen, the next step is to use a shortest path algorithm to compute the full traveltime field and associated ray paths. The original network theory algorithm was developed by Dijkstra (1959), for which computation time scales as $O(M^2)$, where M is the number of nodes. The basic approach is conceptually simple, with nodes divided into two groups: a set of P nodes with known traveltimes, and a set of Q nodes with unknown traveltimes (Moser, 1991); initially, Q contains M elements and P is empty, with the traveltimes of nodes in Q set to an arbitrarily large value. The scheme is initiated from a source node by adding it to P , and then calculating traveltimes to all neighbouring nodes as defined by the forward star (see Fig. 27). These constitute trial traveltime values, and the method proceeds by choosing the trial node with minimum traveltime, adding it to P and then computing trial times to all neighbouring nodes in Q . If the node in Q already has a trial value from a previous iteration, then the one with minimum traveltime is

selected. The complete traveltimes field is found in exactly M iterations, and ray paths are obtained by simply recording the update order of nodes.

The most time consuming part of the Dijkstra algorithm is locating the minimum traveltime node, because in its traditional usage (Moser, 1991), the entire set of nodes in Q is tested, even those with arbitrarily large values. Thus, the total operation count is $\sum_{i=1}^M i = \frac{1}{2}M(M+1) \rightarrow O(M^2)$. However, it is possible to use much more efficient schemes such as the heap sort, which has an operation count of $O(M \log M)$, due to the ordered storage of trial traveltimes in a binary tree. Much work has been done to increase the computational speed of shortest path algorithms by introducing increasingly efficient sorting algorithms (Moser, 1991; Klimeš and Kvasnička, 1994; Cheng and House, 1996; Zhang and Toksöz, 1998).

It is worth noting the similarity between SPR and eikonal solvers such as FMM, which use the shape of the first-arrival wavefront as the computational front (cf. Fig. 27 and Fig. 18). The only real difference is in the way traveltimes are updated to neighbouring nodes. Errors in SPR are due to finite node spacing and the angular distribution of node connectors. A coarse grid of nodes may poorly approximate the true velocity variations, while a limited range of angles between adjacent connectors may not allow for an accurate representation of the path. The accuracy of eikonal solvers is also a function of grid spacing, and parallels can be drawn between the complexity of the forward star used in SPR and the finite difference stencil used to solve the eikonal equation.

Although SPR methods can only be used to locate absolute first-arrivals in continuous media, they can be adapted to track refracted and reflected waves in much the same way as eikonal methods. Moser (1991) describes a constrained shortest path approach to computing this class of phase in layered media, which requires that paths visit a specified set of nodes that lie on an interface. In principle, the scheme is equivalent to the multi-stage FMM described earlier, which separates the medium into different computational domains. Thus, shortest paths are computed from the source node to all other nodes in

the layer; the traveltimes at interface nodes are recorded and ordered in a heap; and then the layer into which the wave propagates has all nodes placed in Q and P set to empty. The evolution of the new time field will therefore be initiated from the interface node with minimum traveltime. Like the multi-stage FMM, this scheme could be extended to compute paths comprising any number of reflection and refraction branches.

SPR has proven to be effective in a number of practical seismic applications that require large datasets to be predicted in the presence of significant lateral heterogeneity. In their original implementation of SPR, Nakanishi and Yamaguchi (1986) invert traveltimes from local earthquakes for 2-D velocity structure. Zhang and Toksöz (1998) use an updated version of the SPR scheme, which samples the wavefront with uniform angular coverage and removes unnecessary nodes from the network, in a 2-D refraction tomography algorithm. In 3-D, Toomey et al. (1994) invert refraction traveltimes for crustal structure using a similar SPR scheme to that proposed by Moser (1991). More recently, Bai (2005) use SPR in the inversion of local and regional earthquake traveltimes for the 3-D velocity structure of Rabaul volcano in Papua New Guinea.

4 Multi-arrival wavefront tracking

Most of the ray and grid-based schemes described previously are only suitable for tracking a single or limited number of arrivals between two points. In many cases, the presence of velocity heterogeneity results in an evolving wavefront self-intersecting, a phenomenon commonly referred to as *multipathing*, because it results in more than one ray path passing through a given point. This behaviour is readily observed in Fig. 1. Eikonal and shortest path methods will only provide the first-arrival in such cases; ray tracing schemes can potentially provide more than one arrival, but the severe non-linearity of the two-point problem makes reliability a major issue. The heterogeneity of Earth structure means that multipathing commonly contributes to the complexity of a recorded waveform, and having

a scheme that can efficiently predict all arrivals of significant amplitude has important implications in various areas of seismology, such as seismic imaging. In particular, it has already been noted that first-arrival traveltimes, which are commonly exploited in seismic tomography, tend to avoid regions of low wavespeed. Later arrivals, which do not necessarily exhibit such behaviour, could therefore provide better path coverage and consequently result in improved imaging.

To date, a number of methods have been developed to solve the multi-arrival problem. These include grid-based (Benamou, 1999; Steinhoff et al., 2000; Engquist et al., 2002; Fomel and Sethian, 2002; Osher et al., 2002; Symes and Qian, 2003), ray-based (Vinje et al., 1993; Lambaré et al., 1996; Vinje et al., 1996, 1999) and hybrid schemes (Benamou, 1996). Due to the relative infancy of this branch of seismology, many of the methods that have been proposed so far are not sufficiently developed for practical application. Here, several ray and grid based schemes are discussed.

4.1 Ray based schemes

One of the first ray-based schemes for computing multi-paths was proposed by Vinje et al. (1993), and is commonly referred to as wavefront construction. The basic principle underlying the method is that a wavefront can be evolved by repeated application of local ray tracing from a set of points lying on the wavefront surface. Thus, for some given time step δt , a new wavefront is defined by the end points of the rays traced from the old wavefront, as illustrated schematically in Fig. 28. In order to avoid undersampling of the wavefront as it evolves, new points are interpolated based on a distance criterion between adjacent rays. Thus, in order to apply this scheme, all that is required is some initial value formulation of the ray equation (e.g. Eq. 29), an interpolation criterion and an interpolation method. From a practical implementation point of view, an efficient indexing scheme for storing the path end points and their connectivity is also required.

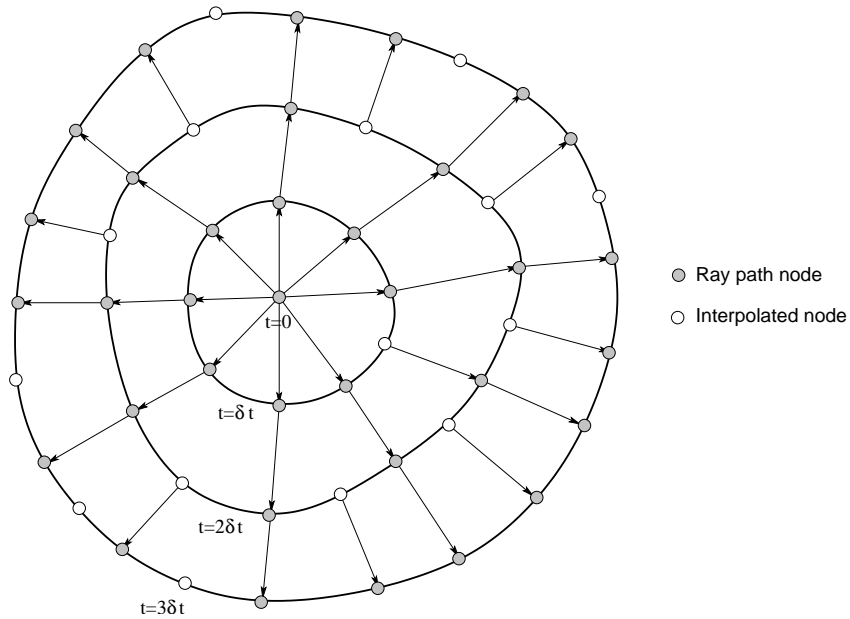


Fig. 28. Schematic diagram illustrating the principle behind the wavefront construction scheme proposed by Vinje et al. (1993). At each iteration, a new wavefront is constructed from the previous wavefront using local ray tracing and interpolation.

The basic 2-D scheme of Vinje et al. (1993) was subsequently extended to 3-D (Vinje et al., 1996), and modified for media involving complex interface structures (Vinje et al., 1999). The solution of the initial value ray tracing equations can be computed with high accuracy, so the main source of error in the wavefront construction method comes from interpolation. To decide whether a point needs to be interpolated, Vinje et al. (1993) uses a distance criterion between adjacent nodes i.e. $|\mathbf{r}_{i+1} - \mathbf{r}_i| \geq d_n$, where \mathbf{r}_{i+1} and \mathbf{r}_i are two adjacent points on the same wavefront and d_n is the distance tolerance above which a new point is interpolated. The problem with this interpolation criterion is that it does not account for variations in wavefront curvature, so regions of high detail are likely to be undersampled. This can be ameliorated to some extent by using a criterion based on angular distance as well, but difficulties can still arise due to the presence of sharp corners, which usually occur when a wavefront triplicates. Since most interpolation schemes are best applied to smooth, differentiable functions, they may not produce satisfactory results.

An alternative approach to interpolation in physical (or “normal”) space is to instead per-

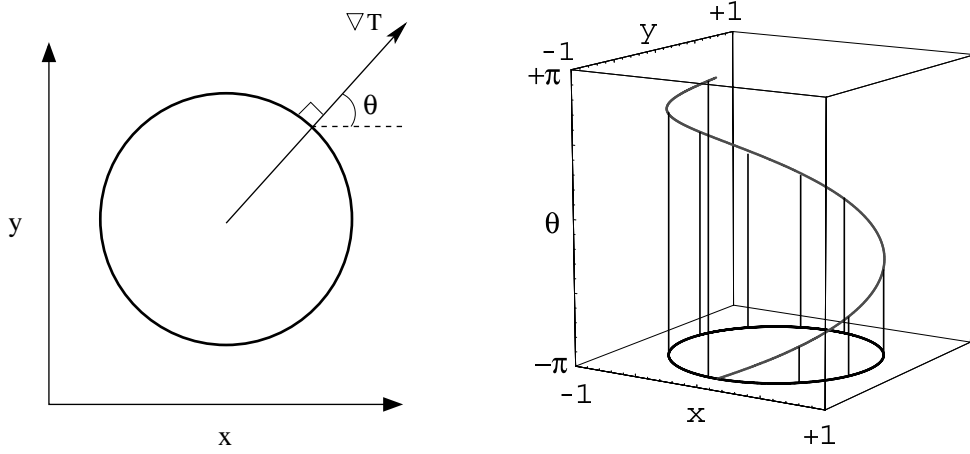


Fig. 29. Representation of a circular wavefront in 3-D reduced phase space; the ray path inclination θ is used as the third coordinate.

form it in phase space (Lambaré et al., 1996; Hauser et al., 2006). As noted earlier, phase space is defined as $(\mathbf{r}, \nabla T)$, where \mathbf{r} is a point on the wavefront and ∇T is the traveltime gradient, equivalent to the ray trajectory. If normal space is extended to higher dimensional phase space by using ∇T as the other coordinate directions, then the wavefront is “unfolded” into a smooth curve sometimes referred to as a “bicharacteristic strip” or curve (Osher et al., 2002). Fig. 29 shows a circular wavefront in 2-D real space unfolded into a spiral curve in 3-D reduced phase space.

Reduced phase space makes use of the relationship $\nabla T = f(\theta, \phi)$, where θ and ϕ are ray inclination and azimuth respectively (see Eq. 29); thus, it is possible to transform 6-D phase space into 5-D reduced phase space for problems in three dimensional normal space, and 4-D phase space into 3-D reduced phase space for problems in two dimensional normal space. The advantage of a reduced phase space representation is that when the wavefront self intersects and contains sharp corners, its corresponding bicharacteristic curve in phase space will be locally smooth and single valued, as demonstrated in Fig. 30.

In order to compute the distance between neighbouring points in reduced phase space, it is necessary to scale the normal space coordinates so that they lie in the interval $[-\pi, \pi]$; for example, in 2-D Cartesian coordinates, $-\pi \leq x, y \leq \pi$. This scaling correlates the metric distance with the angular distance. The criterion for adding a new point is then

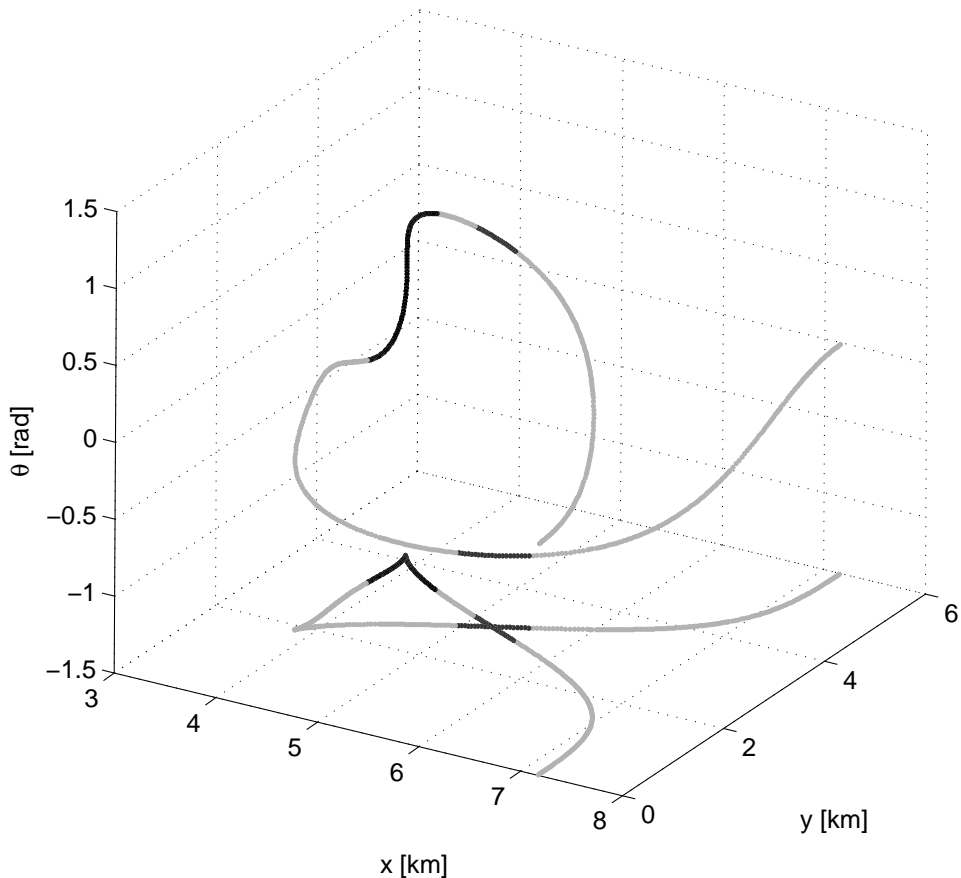


Fig. 30. A triplicating wavefront in normal space is represented by a smooth bicharacteristic curve in reduced phase space. Note that the intersecting segments in normal space no longer intersect in phase space, and the sharp corner is unfolded into a smooth segment of the bicharacteristic curve.

given by $|\mathbf{x}_{i+1} - \mathbf{x}_i| \geq d_p$, where \mathbf{x}_{i+1} and \mathbf{x}_i are adjacent points on the wavefront and d_p is the tolerance. The important difference compared to the criterion used by Vinje et al. (1993) is that here $\mathbf{x} = (\mathbf{r}, \boldsymbol{\theta})$, rather than $\mathbf{x} = \mathbf{r}$. The smoothness of the bicharacteristic curve in reduced phase space means that linear interpolation of \mathbf{x} (five variables in 3-D, three variables in 2-D) is often sufficient. However, it is relatively straightforward to use higher-order interpolators. For example, in 3-D phase space, one could insert a new point \mathbf{x}_m midway between \mathbf{x}_{i+1} and \mathbf{x}_i using a weighted average defined by (Dyn et al., 1990)

$$\mathbf{x}_m = \frac{-\mathbf{x}_{i-1} + 9\mathbf{x}_i + 9\mathbf{x}_{i+1} - \mathbf{x}_{i+2}}{16}, \quad (64)$$

which provides a smoother interpolation. As well as using interpolation to introduce new points to the wavefront, it also makes sense to have a minimum distance criterion to

allow the removal of points which become too close together as a result of wavefront focusing. Properly applied, this can significantly improve computational efficiency without degrading accuracy.

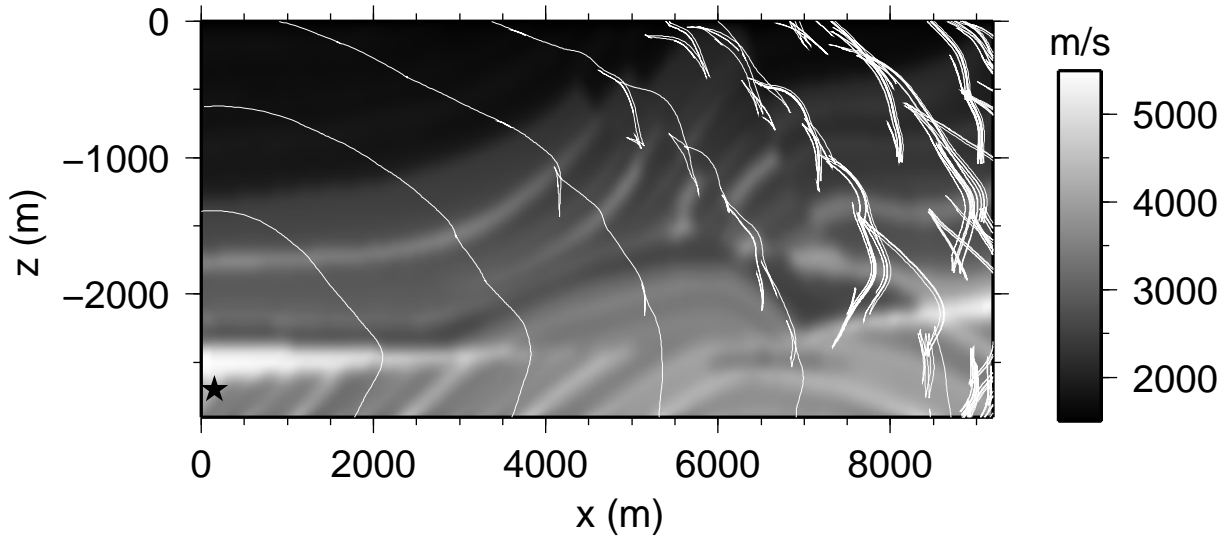


Fig. 31. Wavefronts tracked through the smoothed Marmousi model using the wavefront construction method of Hauser et al. (2006).

The robustness of wavefront construction in reduced phase space is demonstrated in Fig. 31, which shows several snapshots of an evolving wavefront in the smooth version of the Marmousi model. The Marmousi model is one of the most challenging benchmarks for ray tracing and grid-based traveltimes schemes, and is frequently used to test new methods (e.g. Lambaré et al., 1996; Alkhalifah and Fomel, 2001). In the case of Fig. 31, which uses the scheme of Hauser et al. (2006), the wavefront gradually develops multiple self-intersections as it propagates, which is testament to the complexity of the underlying model. However, the smoothness of the corresponding bicharacteristic curve in reduced phase space ensures that the scheme remains well behaved. In this example, over 60 arrivals from a single source are detected in some regions of the model.

A second wavefront construction example, which tracks a surface wavefront from an earthquake in northern Papua New Guinea through the southwest Pacific region, is shown in Fig. 32. The phase velocity model (Fig. 32a) was constructed on the basis of five dispersion curves for fundamental mode Rayleigh waves with a 15s period, with interpolation using

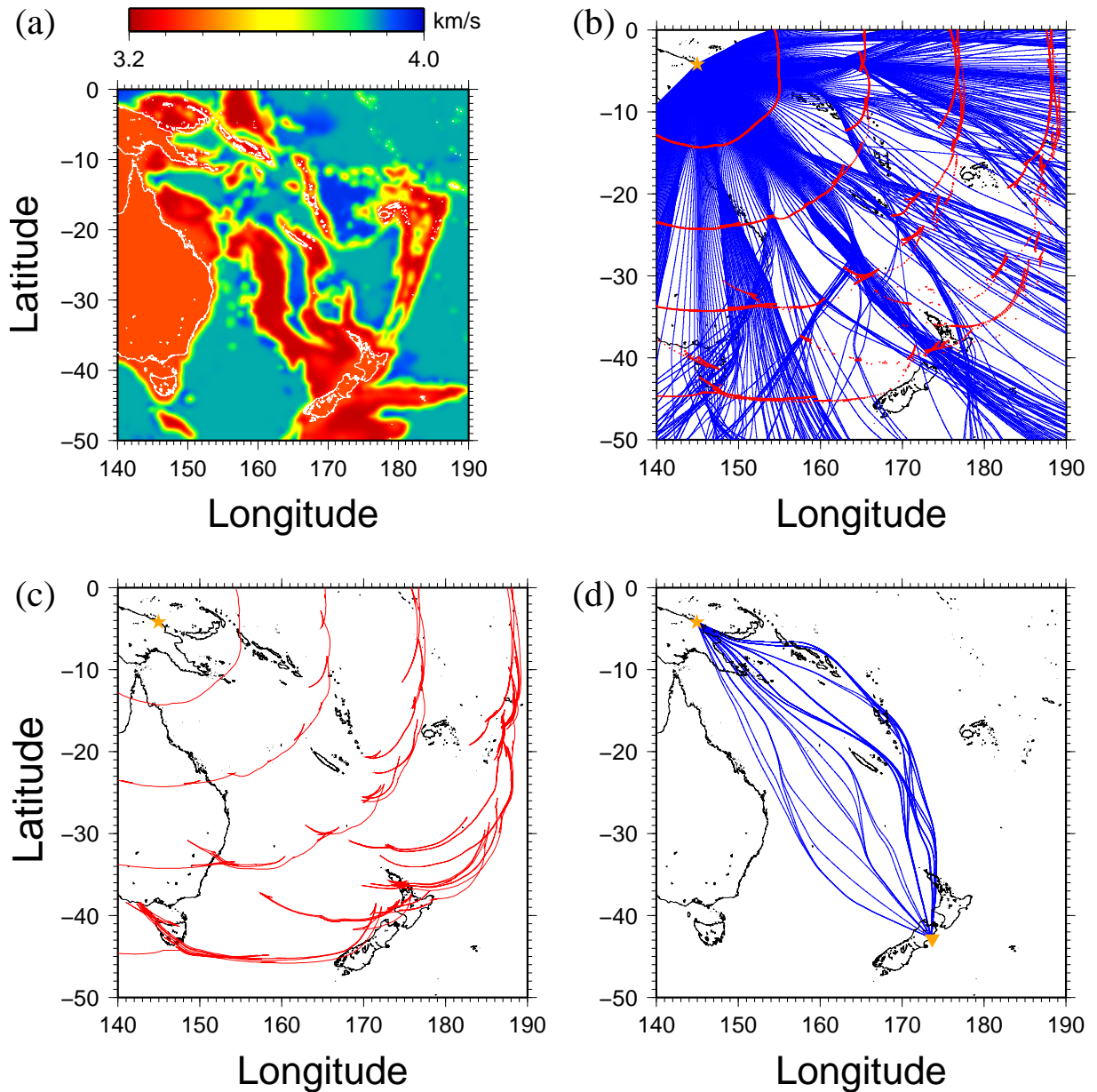


Fig. 32. Surface wavefront propagation in the southwest Pacific. (a) Phase velocity model for fundamental mode Rayleigh waves with a period of 15 s; (b) rays and wavefronts computed using a fan shooting scheme. Wavefronts are visualised by placing red dots along ray paths at predetermined intervals in time. Note that there are more red dots (10,000 per time step) than rays (500) because not all rays are plotted; (c) wavefronts plotted at the same time increments as (b), but computed using the wavefront construction scheme of Hauser et al. (2006); (d) ray paths between the source and a receiver located in New Zealand, obtained from the wavefront construction solution.

average local topography to create the complete 2-D velocity field. At very low frequencies, great circle path propagation is a reasonable approximation, but at higher frequencies,

the effects of velocity variations on path geometry should be considered. Fig. 32b shows the result of shooting a large fan of equally spaced rays (in azimuth) from the source. The strong velocity gradients present in the model leads to significant multipathing, and ray coverage is highly non-uniform. Using this approach, it is not possible to extract the shape of the evolving wavefront with much fidelity. In contrast, the phase space wavefront construction method of Hauser et al. (2006) appears to capture the full detail of the wavefront (Fig. 32c), with many levels of triplications evident. All ray paths (41 in total) from the source to a station located in the South Island of New Zealand are shown in Fig. 32d. Clearly, the energy received by the station will have sampled a broad region of the velocity model, which at the very least suggests that a great circle path approximation for higher frequency surface waves would not be valid in this case.

Extension of the phase space wavefront reconstruction scheme to 3-D is not trivial, as the connectivity between neighbouring points on a wavefront surface is not unique (unlike the 2-D case). Points could be connected using a triangular mesh, but a carefully considered interpolation strategy would be required to properly target regions of the wavefront that become undersampled. In addition, the use of higher order interpolation would be desirable to minimise the number of points that are required to accurately represent the evolving wavefront. One possibility is to use an edge splitting strategy that subdivides adjacent triangles that have a common edge which exceeds some distance tolerance. The butterfly subdivision scheme of Dyn et al. (1990) works along these lines, and allows for linear as well as higher order interpolation.

Several wavefront construction schemes have already been devised for 3-D media, most notably those of Vinje et al. (1996, 1999) and Lucio et al. (1996). Both use triangular meshes to define the wavefront surface, although they use different interpolation strategies. In the case of Vinje et al. (1999), wavefronts can also be tracked in so called “open models” which may contain interfaces with holes. This allows multiple reflection and refraction arrivals to be computed. The use of local ray tracing to evolve the propagating wavefront

means that other ray-related variables such as amplitude may also be computed (Vinje et al., 1993; Lambaré et al., 1996). Amplitude is a potentially important quantity when multiple paths are considered, because it offers a means of discriminating between the relative importance of each arrival.

Of all the multi-arrival schemes proposed so far, wavefront construction is by far the most developed, and is perhaps the only one that is sufficiently robust and efficient for practical problems. It has become a standard tool in the exploration industry via the NORSAR-3D modelling package, which is based on the scheme of Vinje et al. (1999), and has been used in various applications including coincident reflection migration (Xu and Lambaré, 2004; Xu et al., 2004).

Another scheme for tracking multiple arrivals is “big ray-tracing” (Benamou, 1996; Abgrall, 1999), which is actually a hybrid method that attempts to combine initial value ray tracing and finite difference solution of the eikonal equation in order to compute the multi-valued traveltime field. The first step of the procedure is to use initial value ray tracing to partition the traveltime field into single-valued regions via a fan shoot. The second step involves computing traveltimes on an irregular grid of points that span each adjacent pair of ray paths using an eikonal solver. This results in the complete traveltime field spanned by the ray paths being found, and allows wavefronts and intermediate ray paths to be extracted from the solution. However, the principal drawback of this scheme is that the fan shoot, even with a large number of initial rays, is unlikely to properly sample the evolving wavefront (as demonstrated previously in Fig. 32b); for example, the detail contained in an evolving triplication is unlikely to be retained. Rietdijk (1999) discusses the limitations of big ray-tracing in more detail.

4.2 Grid based schemes

Compared to a ray-based technique such as wavefront construction, grid based schemes are potentially attractive as they avoid the need for complex data structures to track irregular ray geometries, and the use of interpolation to retain a uniform sampling of the propagating wavefront. Over the last decade, a number of new grid-based methods for tracking multi-valued wavefronts have been proposed, and new developments appear regularly in the current literature.

One possibility for tracking multiple arrivals is to exploit the stability and efficiency of first-arrival eikonal solvers by partitioning the multi-valued solution into a series of single valued solutions in a way that is analogous to big ray-tracing. A detailed examination of this approach is given by Benamou (1999); other papers on domain decomposition methods include those by Symes and Qian (2003) and Bevc (1997). The main challenge faced by these schemes is devising a robust splitting strategy, as the structure of the multi-valued solution is not known *a priori*. Given that many levels of triplication can occur (e.g. Fig. 31 and 32), it is difficult to envisage this approach becoming generally applicable. Benamou (2003) provides an overview of a number of different splitting methods.

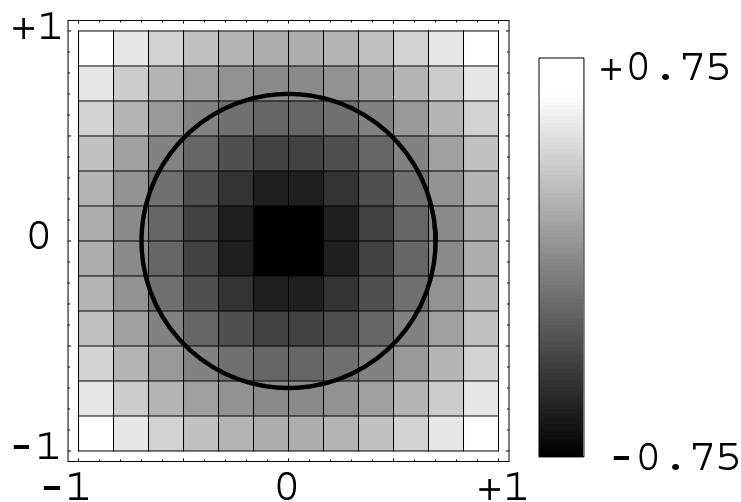


Fig. 33. Level set representation of an interface. In this example, the zero level set of the signed distance function is a circle.

Recently, a number of schemes have been proposed for tracking multi-valued wavefronts using the level set method. Level set methodology was pioneered by Osher and Sethian (1988) for the tracking of interfaces, and has been applied to various problems in the physical sciences (Sethian, 1999; Osher and Fedkiw, 2003), perhaps most notably in fluid dynamics for the evolution of interfaces between fluids and gasses (e.g. Mulder et al., 1992; Sussman et al., 1994; Chang et al., 1996). The basic idea behind a level set formulation is to represent an interface as the zero level set, or zero contour, of a function defined as the signed distance from each grid point to the closest point on the interface (see Fig. 33). This implicitly represented interface can then be evolved by finite difference solution of a set of partial differential equations that explicitly describe the behaviour of the signed distance function over time. In general, the level set method describes the evolution of an $m - 1$ dimensional manifold in m dimensional space; thus, a curve is tracked on a 2-D grid of points, and a surface is tracked on a 3-D grid of points. Note that while the FMM scheme presented earlier is also an interface tracking algorithm, it deals strictly with monotonically advancing interfaces, which allows for a boundary value formulation of the problem. Level set methods do not have this restriction, and consequently require an initial value formulation.

If we denote the signed distance function by $\phi(\mathbf{x})$, then at some given point $\mathbf{x} = \mathbf{x}_A$, $\phi(\mathbf{x}_A)$ is the distance from \mathbf{x}_A to the closest point on the interface, and is negative if \mathbf{x}_A is inside the interface, and positive if \mathbf{x}_A is outside the interface (Fig. 33). The propagation of the surface implicitly described by $\phi(\mathbf{x})$ can be defined in space and time by $\phi(\mathbf{r}(t), t) = 0$, where $\mathbf{r}(t)$ is the location of the zero level set in space at time t . The time derivative of this expression is given by the chain rule as

$$\phi_t + \nabla\phi(\mathbf{r}(t), t) \cdot \mathbf{r}_t(t) = 0, \quad (65)$$

where ϕ_t and \mathbf{r}_t are time derivatives of ϕ and \mathbf{r} respectively. If F is defined as the scalar speed function in the outward normal direction to the evolving surface, then $F = \mathbf{r}_t(t) \cdot \mathbf{n}$, where the unit normal direction $\mathbf{n} = \nabla\phi/|\nabla\phi|$. The evolution equation for the signed

distance function ϕ is therefore given by

$$\phi_t + F|\nabla\phi| = 0, \quad (66)$$

with some given initial condition $\phi(\mathbf{x}, t = 0)$. This is the level set equation that was originally formulated by Osher and Sethian (1988). The signed distance function ϕ is well behaved when $|\nabla\phi| = 1$ at every grid point in the computational domain. When this occurs, $\phi_t = -F$, so when $F > 0$ the interface moves in the outward normal direction, and when $F < 0$ the interface moves in the inward normal direction.

The signed distance function can become spatially discontinuous in gradient when points exist which are equidistant to at least two points on the front. Thus, numerical methods used for solving Eq. 66 must exhibit reasonable behaviour in the presence of gradient discontinuities; fortunately, a large number of schemes, such as ENO and WENO, have been developed for solving differential equations of this type (e.g. Osher and Shu, 1991; Jiang and Peng, 2000).

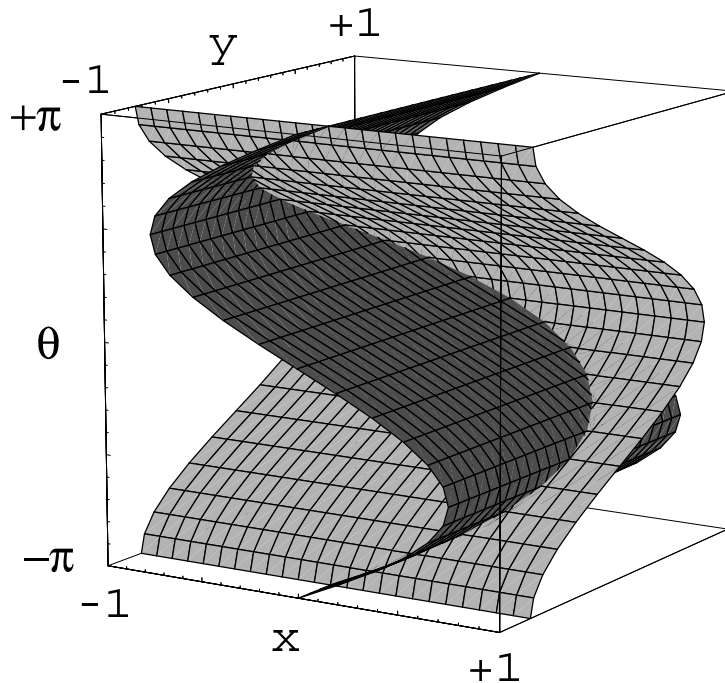


Fig. 34. In 3-D reduced phase space, the bicharacteristic curve can be described by the intersection of two zero level set surfaces. The projection of the bicharacteristic curve into normal space produces a circular wavefront in this example.

When the scalar speed function F is dependent on some vector velocity field $\mathbf{v}(\mathbf{x})$ defined at the grid points, then $F = \mathbf{v} \cdot \mathbf{n}$, so Eq. 66 becomes

$$\phi_t + \mathbf{v}(\mathbf{x}) \cdot \nabla \phi = 0. \quad (67)$$

The strength of the level set method lies in its implicit representation of an interface; topological changes like breaking and merging are handled naturally, and there is no need for resampling of the wavefront as it evolves. However, the interface cannot self-intersect, which at first appears to rule out using the method for multi-valued wavefronts. An approach devised by Osher et al. (2002), which overcomes this limitation, is to consider the problem in a reduced phase space, where the bicharacteristic curve is defined as the intersection of two zero level set surfaces (see Fig. 34); these surfaces are perpendicular along their line of intersection. The two surfaces can be separately evolved by solving the following pair of level set equations

$$\left. \begin{aligned} \phi_t + \mathbf{v} \cdot \nabla \phi &= 0 \\ \psi_t + \mathbf{v} \cdot \nabla \psi &= 0 \end{aligned} \right\}, \quad (68)$$

where the velocity vector \mathbf{v} is given by Eq. 29, and ϕ and ψ represent the two level set functions. The bicharacteristic curve is thus implicitly defined by the curve $\phi = \psi = 0$.

An example of a triplicating wavefront that is tracked using the level set approach described above is shown in Fig. 35. The two zero level set surfaces of ϕ and ψ are perpendicular where they intersect, and the line of intersection describes the bicharacteristic curve, which can be projected into normal space to obtain the wavefront. In this case, the level set equations defined by Eq. 68 are solved using a fifth order WENO scheme in space and a third order TVD Runge Kutta scheme in time (Hauser et al., 2006). Each signed distance function is updated at each time step for the entire computational domain. When computed on a $51 \times 31 \times 51$ grid using 70 time steps and a time increment of 0.1 s, the CPU time is approximately 45 s on a 3.0GHz Intel Pentium 4 PC running GNU/Linux; however, reducing the grid spacing by a factor of two ($101 \times 61 \times 101$ points)

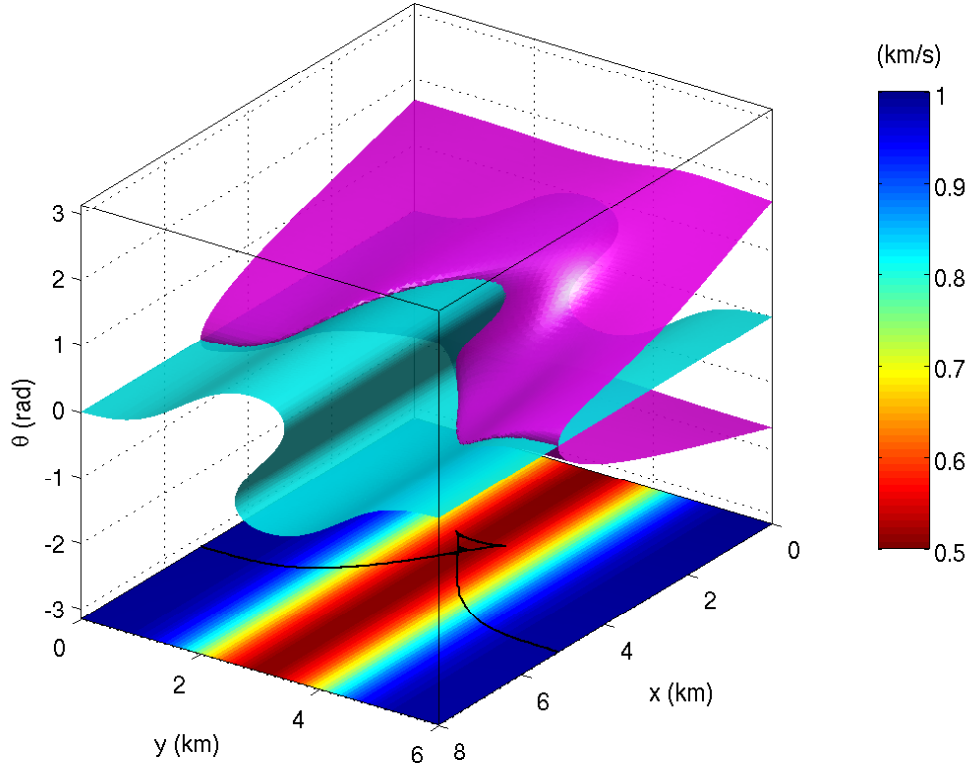


Fig. 35. The evolution of a triplicating wavefront computed using a level set scheme. The zero level set of ϕ is denoted by the magenta surface, and the cyan surface denotes the zero level set of ψ . The bicharacteristic curve is defined by the intersection of these two surfaces, and the projection of the line of intersection into (x, y) space yields the triplicating wavefront. The initial plane wavefront enters the box at $x = 0$.

results in CPU time increasing to approximately 717 s. This increase can be explained by two factors: first, halving the grid spacing in 3-D phase space will increase the number of points by a factor of 8; second, the CFL criterion used (Hauser et al., 2006) requires that a surface does not cross more than one grid cell per time step. Thus, halving the grid spacing means that it is now necessary to use 140 time steps with a time increment of 0.05s. Combined, these two factors result in a 16-fold increase in CPU time when grid spacing is halved.

The relationship between grid spacing and CPU time is the main limitation of the level set scheme of Osher et al. (2002). By comparison, wavefront construction schemes can increase the resolution of a wavefront by a factor of two (doubling the number of rays)

with only a factor of two increase in CPU time (Hauser et al., 2006). For strongly curved wavefronts (e.g. Fig. 35), it is often necessary to have high resolution grids, as the zero level set surfaces tend to become more tightly folded as the wavefront evolves; if a surface overturns within several grid cells or less, then information will be lost and the wavefront will be more poorly represented. Although this scheme can be extended to 3-D by tracking the evolution of three surfaces in five dimensions (Osher et al., 2002), the computational burden would be immense, because increasing grid resolution by a factor of two would increase CPU time by a factor of 64.

Despite the apparent limitations of the level set approach described above, research in this area is ongoing, with particular attention given to improving the efficiency of the scheme. This can be done by using a local level set scheme which only updates the signed distance functions within a narrow band about the bicharacteristic curve (Osher et al., 2002; Osher and Fedkiw, 2003; Qian and Leung, 2006), and adaptive gridding, which provides higher resolution only where it is required. Qian and Leung (2006) devise a local level set scheme which uses a paraxial formulation of the ray equations in which the z coordinate becomes the independent variable. For two dimensional problems, phase space can therefore be defined by (x, θ) rather than (x, z, θ) , which simplifies the problem, but imposes limitations on path geometry (in this case, wavefronts can only move in the positive z direction). Using a waveguide model similar to Fig. 35 and a 120×120 grid, Qian and Leung (2006) show that CPU time (on a Pentium IV desktop) is 72 minutes for a global level set method, but only 8 minutes for their local level set scheme, which is a significant improvement. Other papers on level set based methods for wavefront tracking in phase space include those by Leung et al. (2004), Qian and Leung (2004) and Cockburn et al. (2005).

An alternative approach to computing multi-arrivals with a grid-based scheme has been proposed by Fomel and Sethian (2002). Starting with the Liouville formulation of the ray tracing equations, a system of time-independent PDEs (referred to as “Escape Equations”)

are formulated which can be solved numerically on a grid in reduced phase space. The solutions correspond to the arrival times at the boundary from every point in the phase space domain. As a result, the computational efficiency of the scheme is $O(M \log M)$, where M is the total number of grid points. Multi-arrival information such as wavefront geometry and two-point traveltimes is extracted with post-processing. In comparison with the level set method described previously, Osher et al. (2002) claims that the Fomel and Sethian (2002) scheme is less computationally efficient at each update step (by a factor of $O(M^{1/2})$ in 2-D and $O(M)$ in 3-D), but given that the Escape Equations are boundary value PDEs, and therefore do not require solution at multiple time increments, it is debatable which scheme is ultimately more efficient.

It is interesting to note that a number of new schemes for computing multi-arrivals using a grid based (or “Eulerian”) framework in phase space appeared in the literature at the same time. Two of these have been discussed above, but another, referred to as the “segment projection” method (Engquist et al., 2002), which explicitly tracks a wavefront in phase space, will also be mentioned. In three dimensional reduced phase space, the bicharacteristic curve is represented by its projection onto the two-dimensional xy , $x\theta$ and $y\theta$ planes. Each of these curves is subdivided into a contiguous set of single valued line segments. The segments are then evolved by solving the relevant equations of motion, with dynamic creation and elimination of segments used to retain the minimum number of single-valued segments necessary to represent the wavefront. The method bears some resemblance to wavefront construction in phase space, but uses an Eulerian framework. It appears to be robust in the presence of strong velocity contrasts (such as a waveguide) which cause multiple foldings of a wavefront, and Engquist et al. (2002) claim that it has a similar level of computational complexity as ray tracing.

A final grid-based technique that is worth considering is Dynamic Surface Extension or DSE (Steinhoff et al., 2000). DSE shares some similarities with the level set scheme of Osher et al. (2002), except that the grid lies in normal space and the information

carried by each grid point is the coordinate of the closest point on the wavefront rather than the distance. The DSE algorithm proceeds in two steps; in the first, the location of the wavefront is updated throughout the grid using a knowledge of the the wavefront velocity; the second step updates the wavefront location information carried by the grid. Self-intersecting wavefronts can be tracked, but not all types of wavefront behaviour can be captured (Engquist et al., 2002). In addition, the method is not purely Eulerian as interpolation is needed to maintain an even sampling of the wavefront on the grid.

Although a number of grid and ray based schemes for tracking multi-arrival wavefronts has been described above, our coverage has not been exhaustive. For more information, the interested reader is directed to two relatively recent review papers on the subject (Benamou, 2003; Engquist and Runborg, 2003).

5 Concluding Remarks

In this review paper, we have described a variety of schemes for tracking the kinematic evolution of high frequency seismic waves in heterogeneous 2-D and 3-D structures. Where possible, we have focused on methods that have been used in practical applications; most of these can be characterised as either ray based or grid based. Ray tracing has traditionally been the method of choice in many seismic applications due to its high accuracy and potential for computational efficiency. Common ray tracing schemes include shooting, bending and pseudo bending. Shooting methods formulate the kinematic ray equation as an initial value problem, which allows a complete ray path to be computed given an initial projection vector. The boundary value problem of locating source-receiver trajectories is typically solved using iterative non-linear strategies which exploit information from nearby paths to update the projection parameters and better target the receiver. Bending methods perturb the geometry of an initial path joining source and receiver, until it becomes a true (stationary) path, by iteratively solving a boundary value formulation of

the ray equations. Pseudo-bending schemes usually represent a path as a series of points which are perturbed using a simple algorithm based directly on Fermat's principle of stationary time. All three methods are non-robust in the sense that the assumption of local linearity is made even though it is not generally valid.

Grid based methods include those that solve the eikonal equation using finite differences on a grid of points, and shortest path ray tracing, which uses Dijkstra-like algorithms to find traveltimes and paths through a network of nodes with predetermined connectivity. Both techniques are capable of computing traveltimes to all points of a medium in a stable and computationally efficient manner. It was shown that eikonal solvers which use the expanding wavefront as the computational front, such as the fast marching method, share many similarities with SPR. Compared to ray tracing, advantages of these grid based schemes include the ability to compute traveltimes to all points of the medium in one pass, stability, and in many cases, computational efficiency. However, the fact that they only compute the first arrival traveltime field in continuous media is a limitation, and excessive computing time can become an issue if highly accurate traveltimes are required. In addition, quantities other than traveltime are difficult to compute without first extracting ray paths from the traveltime field and then resorting to ray based techniques.

In Section 4, a variety of ray and grid based schemes for tracking multi-arrival wavefronts were presented. Although this area of research is arguably still in its infancy (particularly with regard to grid-based methods), there has been a proliferation of new schemes appearing in the literature over the past decade. The most developed and practical scheme to emerge so far has been wavefront construction, which uses local ray tracing and interpolation to advance a wavefront surface in discrete time steps. Of the grid-based solvers, segment projection perhaps shows the most promise, although research is still ongoing. The attraction of these techniques is that they are capable of computing all arrivals, which may allow more of an observed seismic wavetrain to be predicted. However, from an application point of view, there are still significant hurdles to be overcome. For instance, it

is almost impossible to pick arrivals other than the first from a seismogram unless they represent specific reflection or refraction phases. Thus, the issue of how to best exploit the multi-arrival capability of these schemes is yet to be fully explored.

In practice, the most appropriate scheme for predicting an observational dataset depends on a number of factors including computational efficiency, accuracy, robustness and the type of information that is required. For very large datasets, computational efficiency and robustness are crucial. In cases where there are many sources and few receivers or vice-versa, then eikonal or SPR solvers should be considered. When lateral heterogeneity is not particularly significant and true two-point paths do not deviate significantly from those computed through a laterally averaged model, then shooting or bending methods of ray tracing may be more appropriate. Global phases that penetrate deep into the earth tend to fall into this category, which perhaps explains why ray based schemes are still the preferred option with this class of dataset. Traditionally, ray tracing has always been preferred in models that include interfaces, but recent developments in grid-based eikonal solvers suggest that they can be just as effective for computing phases that experience multiple refractions and reflections. In cases where multi-arrival information is required, then a method like wavefront construction would be recommended.

In a paper by Leidenfrost et al. (1999), six different methods for calculating traveltimes are compared. These include three different varieties of eikonal solver, SPR, wavefront construction and a combined eikonal and Runge Kutta method. The first of the eikonal solvers is essentially the Vidale (1988) method; the second adapts the same scheme to polar coordinates with an expanding circle rather than square; and the third uses the method of Podvin and Lecomte (1991). The combined eikonal and Runge Kutta method solves the eikonal equation in polar coordinates by employing first-order finite differences in θ and a fourth-order Runge-Kutta scheme in radius. The SPR method is that of Klimeš and Kvasnička (1994) and the wavefront construction scheme is essentially that of Ettrich and Gajewski (1996). All solvers were benchmarked in 2-D using both a simple model

with analytic solutions and the Marmousi model. The tests show that the expanding square eikonal solver offers the best compromise between speed and accuracy for models that are not too complicated, while wavefront construction proved to be the most reliable for the more complex Marmousi model. It is interesting to note that despite computing multi-arrivals, wavefront construction was of the same order of accuracy as the expanding square method and had computation times better than the SPR scheme.

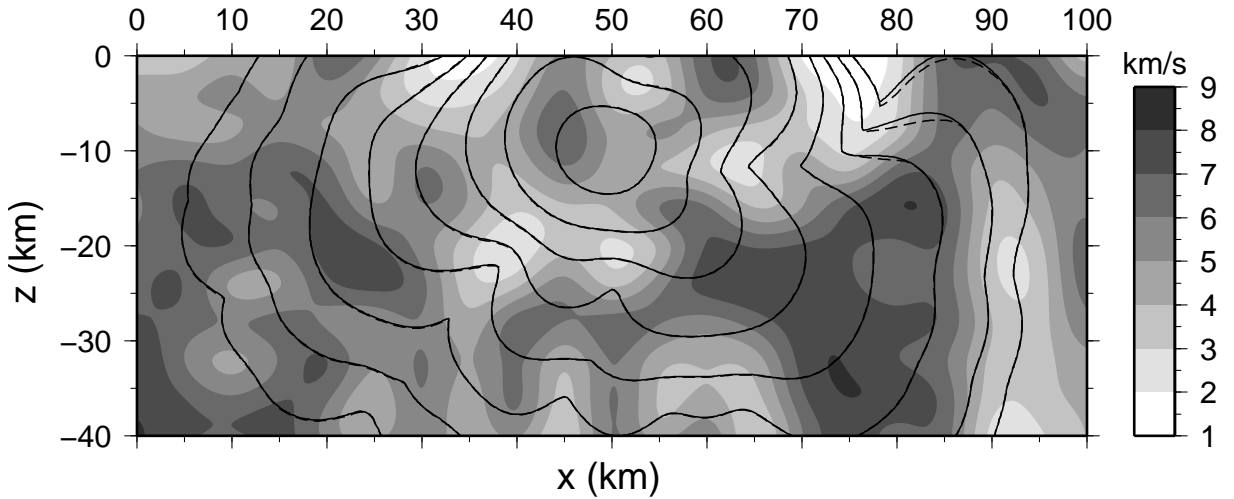


Fig. 36. Comparison of wavefronts computed using FMM (solid lines) and WENO-DNO-PS (dashed lines) for the complex velocity model of Fig. 16. The WENO-DNO-PS scheme clearly overestimates traveltimes in the neighbourhood of $(x, z) = (80, -5)$ km.

In a separate study, Kim (2002) carry out a comparison between FMM, GMM and a second order ENO-DNO-PS scheme using a 2-D smooth velocity model. GMM or the “Group Marching Method” (Kim, 2001), is a variant of FMM which avoids the need to sort narrow band traveltimes; hence computational cost scales as $O(M)$ rather than $O(M \log M)$, where M is the total number of grid points. As described previously, ENO-DNO-PS is an eikonal solver which uses an essentially non-oscillatory finite difference stencil, an expanding square framework, and post sweeping to repair causality breaches introduced by the expanding square. Results using the same grid spacing suggest that ENO-DNO-PS is approximately an order of magnitude more computationally efficient than the other two schemes, yet produces more accurate results. However, it should be pointed out that only first-order FMM and GMM schemes were used, and no source-grid

refinement was implemented. Another factor in favour of ENO-DNO-PS was the use of a relatively simple velocity field; in the presence of complex velocity structures, the use of *ad hoc* post-sweeping may not always succeed. Fig. 36 shows a comparison between the WENO-DNO-PS solution of Fig. 16 and a second-order FMM scheme with local grid refinement. Although the solutions are very similar, WENO-DNO-PS significantly overestimates the traveltimes in the region $(x, z) = (80, -5)$. It is possible that a judicious post-sweeping strategy may alleviate this problem, but in general, it is not clear that the correct solution can always be guaranteed.

ACKNOWLEDGMENTS

N. Rawlinson is supported by ARC Discovery Project DP0556282 and J. Hauser is supported by ARC Discovery Project DP0451133. Most figures were created using the freely distributed software packages GMT, xfig, Geomview and PGPLOT.

References

- Abgrall, R., 1999. Big ray-tracing and eikonal solver on unstructured grids: Application to the computation of a multivalued traveltime field in the Marmousi model. *Geophysics* 64, 230–239.
- Afnimar, Koketsu, K., 2000. Finite difference traveltime calculation for head waves travelling along an irregular interface. *Geophys. J. Int.* 143, 729–734.
- Aki, K., Christoffersson, A., Husebye, E. S., 1977. Determination of the three-dimensional seismic structure of the lithosphere. *J. Geophys. Res.* 82, 277–296.
- Alkhalifah, T., Fomel, S., 2001. Implementing the fast marching eikonal solver: Spherical versus Cartesian coordinates. *Geophysical Prospecting* 49, 165–178.
- Bai, C.-Y., 2005. 3-d multi-step travel time tomography: Imaging the local, deep velocity

- structure of Rabaul volcano, Papua New Guinea. *Phys. Earth Planet. Inter.* 151, 259–275.
- Bartels, R. H., Beatty, J. C., Barsky, B. A., 1987. An introduction to splines for use in computer graphics and geometric modelling. Morgan Kaufmann, Los Altos.
- Benamou, J. D., 1996. Big ray tracing: Multivalued travel time field computation using viscosity solutions of the Eikonal equation. *J. Comp. Phys.* 128, 463–474.
- Benamou, J. D., 1999. Direct computation of multivalued phase space solutions for Hamilton-Jacobi equations. *Comm. Pure Appl. Math.* 52, 1443–1475.
- Benamou, J. D., 2003. An introduction to Eulerian Geometrical Optics (1992-2002). *SIAM J. Sci. Comput.* 19, 63–93.
- Benz, H. M., Smith, R. B., 1984. Simultaneous inversion for lateral velocity variations and hypocenters in the Yellowstone region using earthquake and refraction data. *J. Geophys. Res.* 89, 1208–1220.
- Bevc, D., 1997. Imaging complex structures with semirecursive Kirchhoff migration. *Geophysics* 62, 577–588.
- Bulant, P., 1996. Two-point ray tracing in 3-D. *Pure Appl. Geophys.* 148, 421–447.
- Bulant, P., 1999. Two-point ray-tracing and controlled initial-value ray-tracing in 3-d heterogeneous block structures. *J. Seismic Explor.* 8, 57–75.
- Buske, S., 1999a. 3-D prestack Kirchhoff migration of the ISO89-3D data set. *Pure Appl. Geophys.* 156, 157–171.
- Buske, S., 1999b. Three-dimensional pre-stack Kirchhoff migration of deep seismic reflection data. *Geophys. J. Int.* 137, 1243–260.
- Buske, S., Kästner, U., 2004. Efficient and accurate computation of seismic traveltime and amplitudes. *Geophysical Prospecting* 52, 313–322.
- Cao, S., Greenhalgh, S., 1994. Finite-difference solution of the eikonal equation using an efficient, first-arrival, wavefront tracking scheme. *Geophysics* 59, 632–643.
- Cassell, B. R., 1982. A method for calculating synthetic seismograms in laterally varying media. *Geophys. J. Royal Astr. Soc.* 69, 339–354.

- Červený, V., 1987. Ray tracing algorithms in three-dimensional laterally varying layered structures. In: Nolet, G. (Ed.), *Seismic tomography: With applications in global seismology and exploration geophysics*. D. Reidel, Dordrecht, pp. 99–133.
- Červený, V., 2001. *Seismic Ray Theory*. Cambridge University Press, Cambridge.
- Červený, V., Firbas, P., 1984. Numerical modelling and inversion of travel times of seismic body waves in inhomogeneous anisotropic media. *Geophys. J. Royal Astr. Soc.* 76, 41–51.
- Červený, V., Klimes, L., Psencik, I., 2006. *Seismic ray method: Recent developments*. *Advances in Geophysics* 48, in press.
- Červený, V., Klimeš, L., Pšenčík, I., 1984. Paraxial ray approximations in the computation of seismic wavefields in inhomogeneous media. *Geophys. J. Royal Astr. Soc.* 79, 89–104.
- Červený, V., Pšenčík, I., 1983. Gaussian beams and paraxial ray approximation in three-dimensional elastic inhomogeneous media. *J. Geophys.* 53, 1–15.
- Chang, Y. C., Hou, T. Y., Merriman, B., Osher, S., 1996. A level set formulation of Eulerian interface capturing methods for incompressible fluid flows. *J. Comp. Phys.* 124, 449–464.
- Chapman, C., 2004. *Fundamentals of seismic wave propagation*. Cambridge University Press, Cambridge.
- Cheng, N., House, L., 1996. Minimum traveltimes calculations in 3-D graph theory. *Geophysics* 61, 1895–1898.
- Chiu, S. K. L., Kanasewich, E. R., Phadke, S., 1986. Three-dimensional determination of structure and velocity by seismic tomography. *Geophysics* 51, 1559–1571.
- Chopp, D. L., 2001. Some improvements of the fast marching method. *SIAM J. Sci. Comput.* 23, 230–244.
- Cockburn, B., Qian, J., Reitich, F., Wang, J., 2005. An accurate spectral/discontinuous finite-element formulation of a phase-space-based level set approach to geometrical optics. *J. Comp. Phys.* 208, 175–195.
- Curtis, A., Snieder, R., 1997. Reconditioning inverse problems using the genetic algorithm

- and revised parameterization. *Geophysics* 62, 1524–1532.
- Day, A. J., Peirce, C., Sinha, M. C., 2001. Three-dimensional crustal structure and magma chamber geometry at the intermediate-spreading, back-arc Valu Fa Ridge, Lau Basin - results of a wide-angle seismic tomographic inversion. *Geophys. J. Int.* 146, 31–52.
- Dębski, W., Ando, M., 2004. Spectral ray tracer: A class of accurate two-point ray tracers. *Acta Geophysica Polonica* 52, 1–14.
- de Kool, M., Rawlinson, N., Sambridge, M., 2006. A practical grid based method for tracking multiple refraction and reflection phases in 3d heterogeneous media. *Geophys. J. Int.* 167, 253–270.
- Dijkstra, E. W., 1959. A note on two problems in connection with graphs. *Numer. Math.* 1, 269–271.
- Dyn, N., Levin, D., Gregory, J. A., 1990. A butterfly subdivision scheme for surface interpolation with tension control. *ACM Transaction on Graphics* 9, 160–169.
- Dziewonski, A. M., Hager, B. H., O’Connell, R. J., 1977. Large-scale heterogeneities in the lower mantle. *J. Geophys. Res.* 82, 239–255.
- Eberhart-Phillips, D., 1986. Three-dimensional velocity structure in northern California coast ranges from inversion of local earthquake arrival times. *Bull. Seism. Soc. Am.* 76, 1025–1052.
- Eberhart-Phillips, D., 1990. Three-dimensional P and S velocity structure in the Coalinga Region, California. *J. Geophys. Res.* 95, 15,343–15,363.
- Eberhart-Phillips, D., Reyners, M., 1997. Continental subduction and three-dimensional crustal structure: The northern South Island, New Zealand. *J. Geophys. Res.* 102, 11,848–11,861.
- Engquist, B., Runborg, O., 2003. Computational high frequency wave propagation. *Acta Numerica* 12, 181–266.
- Engquist, B., Runborg, O., Tornberg, A.-K., 2002. High-frequency wave propagation by the segment projection method. *J. Comp. Phys.* 178, 373–390.
- Ettrich, N., Gajewski, D., 1996. Wave front construction in smooth media for prestack

- depth migration. *Pageoph* 148, 481–502.
- Farra, V., Madariaga, R., 1987. Seismic waveform modelling in heterogeneous media by ray perturbation theory. *J. Geophys. Res.* 92, 2697–2712.
- Farra, V., Madariaga, R., 1988. Non-linear reflection tomography. *Geophys. J.* 95, 135–147.
- Fischer, R., Lees, J. M., 1993. Shortest path ray tracing with sparse graphs. *Geophysics* 58, 987–996.
- Fomel, S., Sethian, J. A., 2002. Fast-phase space computation of multiple arrivals. *Proc. Nat. Acad. Sci.* 99, 7329–7334.
- Graeber, F. M., Asch, G., 1999. Three-dimensional models of *P* wave velocity and *P*-to-*S* velocity ratio in the southern central Andes by simultaneous inversion of local earthquake data. *J. Geophys. Res.* 104, 20,237–20,256.
- Gray, S. H., May, W. P., 1994. Kirchhoff migration using eikonal equation traveltimes. *Geophysics* 59, 810–817.
- Hammer, P. T. C., Dorman, L. M., Hildebrand, J. A., Cornuelle, B. D., 1994. Jasper Seamount structure: Seafloor seismic refraction tomography. *J. Geophys. Res.* 99, 6731–6752.
- Hauser, J., Sambridge, M., Rawlinson, N., 2006. Phase space methods for multi arrival wavefronts. *Exploration Geophysics* submitted.
- Hildebrand, J. A., Dorman, L. M., Hammer, P. T. C., Schreiner, A. E., Cornuelle, B. D., 1989. Seismic tomography of Jasper Seamount. *Geophys. Res. Lett.* 16, 1355–1358.
- Hole, J. A., 1992. Nonlinear high-resolution three-dimensional travel-time tomography. *J. Geophys. Res.* 97, 6553–6562.
- Hole, J. A., Zelt, B. C., 1995. 3-D finite-difference reflection travel times. *Geophys. J. Int.* 121, 427–434.
- Iyer, H., Hirahara, K., 1993. *Seismic Tomography: Theory and Practice*. Chapman & Hall, London.
- Jeffreys, H., Swirles, B., 1966. *Methods of mathematical physics*. Cambridge University

- Press, Cambridge.
- Jiang, G. S., Peng, D. P., 2000. Weighted ENO schemes for hamilton-jacobi equations. SIAM J. Sci. Comput. 21, 2126–2143.
- Jiang, G. S., Shu, C., 1996. Efficient implementation of weighted ENO schemes. J. Comp. Phys. 126, 202–228.
- Julian, B. R., Gubbins, D., 1977. Three-dimensional seismic ray tracing. J. Geophys. 43, 95–113.
- Kim, S., 2001. An $O(N)$ level set method for eikonal equations. SIAM J. Sci. Comput. 22, 2178–2193.
- Kim, S., 2002. 3-D eikonal solvers: First-arrival traveltimes. Geophysics 67, 1225–1231.
- Kim, S., Cook, R., 1999. 3D traveltime computation using second-order ENO scheme. Geophysics 64, 1867–1876.
- Kimmel, R. and Sethian, J. A., 1998. Fast marching methods on triangulated domains. Proc. Nat. Acad. Sci. 95, 8341–8435.
- Kirkpatrick, S., Gelatt, Jr, C. D., Vecchi, M. P., 1983. Optimization by simulated annealing. Science 220, 671–680.
- Klimeš, L., 1989. Gaussian packets in the computation of seismic wavefields. Geophys. J. Int. 99, 421–433.
- Klimeš, L., Kvasnička, M., 1994. 3-D network ray tracing. Geophys. J. Int. 116, 726–738.
- Koketsu, K., Sekine, S., 1998. Pseudo-bending method for three-dimensional seismic ray tracing in a spherical earth with discontinuities. Geophys. J. Int. 132, 339–346.
- Kreyszig, E., 1993. Advanced engineering mathematics. John Wiley & Sons, Inc., New York.
- Lambaré, G., Lucio, P. S., Hanyga, A., 1996. Two-dimensional multivalued traveltime and amplitude maps by uniform sampling of a ray field. Geophys. J. Int. 125, 584–598.
- Langan, R. T., Lerche, I., Cutler, R. T., 1985. Tracing of rays through heterogeneous media: An accurate and efficient procedure. Geophysics 50, 1456–1465.
- Leidenfrost, A., Ettrich, N., Gajewski, D., Kosloff, D., 1999. Comparison of six different

- methods for calculating traveltimes. *Geophysical Prospecting* 47, 269–297.
- Leung, S., Osher, S. J., Qian, J., 2004. A level set method for three-dimensional paraxial geometrical optics with multiple point sources. *Commun. Math. Sci.* 2, 643–672.
- Li, X. G., Ulrych, T. J., 1993. Traveltime computation in discrete heterogeneous layered media. *J. Seismic Explor.* 2, 305–318.
- Liu, X., Osher, S., Chan, T., 1994. Weighted essentially non-oscillatory schemes. *J. Comp. Phys.* 115, 200–212.
- Lucio, P. S., Lambaré, G., Hanyga, A., 1996. 3D multivalued travel time and amplitude maps. *Pageoph* 148, 449–479.
- McCaughey, M., Singh, S. C., 1997. Simultaneous velocity and interface tomography of normal-incidence and wide-aperture seismic traveltime data. *Geophys. J. Int.* 131, 87–99.
- Michelini, A., 1995. An adaptive-grid formalism for traveltime tomography. *Geophys. J. Int.* 121, 489–510.
- Moser, T. J., 1991. Shortest path calculation of seismic rays. *Geophysics* 56, 59–67.
- Mulder, W., Osher, S., Sethian, J. A., 1992. Computing interface motion in compressible gas dynamics. *J. Comp. Phys.* 100, 209–228.
- Nakanishi, I., 1985. Three-dimensional structure beneath the Hokkaido-Tohoku region as derived from a tomographic inversion of P -arrival times. *J. Phys. Earth* 33, 241–256.
- Nakanishi, I., Yamaguchi, K., 1986. A numerical experiment on nonlinear image reconstruction from first-arrival times for two-dimensional island arc structure. *J. Phys. Earth* 34, 195–201.
- Neele, F., VanDecar, J., Snieder, R., 1993. The use of P wave amplitude data in a joint inversion with travel times for upper mantle velocity structure. *J. Geophys. Res.* 98, 12,033–12,054.
- Nolet, G., 1987. Waveform tomography. In: Nolet, G. (Ed.), *Seismic tomography: With applications in global seismology and exploration geophysics*. D. Reidel, Dordrecht, pp. 301–322.

- Osher, S., Cheng, L.-T., Kang, M., Shim, H., Tsai, Y.-H., 2002. Geometric optics in a phase-space-based level set and Eulerian framework. *J. Comp. Phys.* 179, 622–648.
- Osher, S., Fedkiw, R., 2003. *Level Set Methods and Dynamic Implicit Surfaces*. Springer Verlag, 2nd edn.
- Osher, S., Sethian, J., 1988. Fronts propagating with curvature dependent speed: Algorithms based on hamilton-jacobi formulations. *J. Comp. Phys.* 79, 12–49.
- Osher, S., Shu, C. W., 1991. High-order essentially nonoscillatory schemes for hamilton-jacobi equations. *SIAM J. Numer. Anal.* 28, 907–922.
- Pereyra, V., 1996. Modelling, ray tracing, and block nonlinear travel-time inversion in 3d. *Pure Appl. Geophys.* 148, 345–386.
- Pereyra, V., Lee, W. H. K., Keller, H. B., 1980. Solving two-point seismic-ray tracing problems in a heterogeneous medium. *Bull. Seism. Soc. Am.* 70, 79–99.
- Podvin, P., Lecomte, I., 1991. Finite difference computation of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools. *Geophys. J. Int.* 105, 271–284.
- Popovici, A. M., Sethian, J. A., 2002. 3-D imaging using higher order fast marching traveltimes. *Geophysics* 67, 604–609.
- Prothero, W. A., Taylor, W. J., Eickemeyer, J. A., 1988. A fast, two-point, three-dimensional raytracing algorithm using a simple step search method. *Bull. Seism. Soc. Am.* 78, 1190–1198.
- Pulliam, J., Snieder, R., 1996. Fast, efficient calculation of rays and travel times with ray perturbation theory. *J. Acoust. Soc. Am.* 99, 383–391.
- Qian, J., Leung, S., 2004. A level set based eulerian method for paraxial multivalued traveltimes. *J. Comp. Phys.* 197, 711–736.
- Qian, J., Leung, S., 2006. A local level set method for paraxial geometrical optics. *SIAM J. Sci. Comput.* 28, 206–223.
- Qian, J., Symes, W. W., 2002. An adaptive finite-difference method for traveltimes and amplitudes. *Geophysics* 67, 167–176.

- Qin, F., Luo, Y., Olsen, K. B., Cai, W., Schuster, G. T., 1992. Finite-difference solution of the eikonal equation along expanding wavefronts. *Geophysics* 57, 478–487.
- Rawlinson, N., Houseman, G. A., 1998. Inversion for interface structure using teleseismic traveltimes residuals. *Geophys. J. Int.* 133, 756–772.
- Rawlinson, N., Houseman, G. A., Collins, C. D. N., 2001a. Inversion of seismic refraction and wide-angle reflection traveltimes for 3-D layered crustal structure. *Geophys. J. Int.* 145, 381–401.
- Rawlinson, N., Houseman, G. A., Collins, C. D. N., Drummond, B. J., 2001b. New evidence of Tasmania’s tectonic history from a novel seismic experiment. *Geophys. Res. Lett.* 28, 3337–3340.
- Rawlinson, N., Kennett, B. L. N., Heintz, M., 2006a. Insights into the structure of the upper mantle beneath the murray basin from 3-d teleseismic tomography. *Australian Journal of Earth Sciences* 53, 595–604.
- Rawlinson, N., Reading, A. M., Kennett, B. L. N., 2006b. Lithospheric structure of tasmania from a novel form of teleseismic tomography. *J. Geophys. Res.* 111, doi:10.1029/2005JB003803.
- Rawlinson, N., Sambridge, M., 2003a. Irregular interface parameterization in 3-D wide-angle seismic traveltimes tomography. *Geophys. J. Int.* 155, 79–92.
- Rawlinson, N., Sambridge, M., 2003b. Seismic traveltimes tomography of the crust and lithosphere. *Advances in Geophysics* 46, 81–198.
- Rawlinson, N., Sambridge, M., 2004a. Multiple reflection and transmission phases in complex layered media using a multistage fast marching method. *Geophysics* 69, 1338–1350.
- Rawlinson, N., Sambridge, M., 2004b. Wavefront evolution in strongly heterogeneous layered media using the Fast Marching Method. *Geophys. J. Int.* 156, 631–647.
- Riahi, M. A., Juhlin, C., 1994. 3-D interpretation of reflected arrival times by finite-difference techniques. *Geophysics* 59, 844–849.
- Riahi, M. A., Lund, C. E., Pederson, L. B., 1997. Three-dimensional image of the Moho undulations beneath the Gulf of Bothnia using wide-angle seismic data. *Geophys. J.*

- Int. 129, 461–471.
- Rietdijk, R. H., 1999. Notes on big ray tracing. *J. Comp. Phys.* 148, 149–168.
- Sadeghi, H., Suzuki, S., Takenaka, H., 1999. A two-point, three-dimensional seismic ray tracing using genetic algorithms. *Phys. Earth Planet. Inter.* 113, 355–365.
- Saltzer, R. L., Humphreys, E. D., 1997. Upper mantle *P* wave velocity structure of the eastern Snake River Plain and its relationship to geodynamic models of the region. *J. Geophys. Res.* 102, 11,829–11,841.
- Sambridge, M., Faletic, R., 2003. Adaptive whole Earth tomography. *Geochem. Geophys. Geosyst.* 4, doi: 10.1029/2001GC000213.
- Sambridge, M., Gudmundsson, O., 1998. Tomographic systems of equations with irregular cells. *J. Geophys. Res.* 103, 773–781.
- Sambridge, M., Rawlinson, N., 2005. Seismic tomography with irregular meshes. In: Levander, A., Nolet, G. (Eds.), *Seismic Earth: Array analysis of broadband seismograms*. Vol. 157. American Geophysical Union, pp. 49–65.
- Sambridge, M. S., 1990. Non-linear arrival time inversion: Constraining velocity anomalies by seeking smooth models in 3-D. *Geophys. J. Int.* 102, 653–677.
- Sambridge, M. S., Braun, J., McQueen, H., 1995. Geophysical parametrization and interpolation of irregular data using natural neighbours. *Geophys. J. Int.* 122, 837–857.
- Sambridge, M. S., Kennett, B. L. N., 1990. Boundary value ray tracing in a heterogeneous medium: A simple and versatile algorithm. *Geophys. J. Int.* 101, 157–168.
- Scott, J. S., Masters, T. G., Vernon, F. L., 1994. 3-D velocity structure of the San Jacinto fault zone near Anza, California-I. *P* waves. *Geophys. J. Int.* 119, 611–626.
- Sethian, J. A., 1996. A fast marching level set method for monotonically advancing fronts. *Proc. Nat. Acad. Sci.* 93, 1591–1595.
- Sethian, J. A., 1999. *Level Set Methods and Fast Marching Methods*. Cambridge University Press, Cambridge.
- Sethian, J. A., 2001. Evolution, implementation, and application of level set and fast marching methods for advancing fronts. *J. Comp. Phys.* 169, 503–555.

- Sethian, J. A., Popovici, A. M., 1999. 3-D traveltimes computation using the fast marching method. *Geophysics* 64, 516–523.
- Shapiro, N. M., Campillo, M., Stehly, L., Ritzwoller, M. H., 2005. High-resolution surface wave tomography from ambient seismic noise. *Science* 307, 1615–1618.
- Shu, C.-W., Osher, S., 1988. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *J. Comp. Phys.* 77, 439–471.
- Shu, C.-W., Osher, S., 1989. Efficient implementation of essentially non-oscillatory shock-capturing schemes, ii. *J. Comp. Phys.* 83, 32–78.
- Slawinski, M. A., Slawinski, R. A., Brown, R. J., Parkin, J. M., 2000. A generalized form of snell’s law in anisotropic media. *Geophysics* 65, 632–637.
- Smith, W. H. F., Wessel, P., 1990. Gridding with continuous curvature splines in tension. *Geophysics* 55, 293–305.
- Snieder, R., Sambridge, M., 1992. Ray perturbation theory for traveltimes and ray paths in 3-d heterogeneous media. *Geophys. J. Int.* 109, 294–322.
- Snieder, R., Spencer, C., 1993. A unified approach to ray bending, ray perturbation and paraxial ray theories. *Geophys. J. Int.* 115, 456–470.
- Spakman, W., Bijwaard, H., 2001. Optimization of cell parameterizations for tomographic inverse problems. *Pure Appl. Geophys.* 158, 1401–1423.
- Steck, L. K., Thurber, C. H., Fehler, M., Lutter, W. J., Roberts, P. M., Baldrige, W. S., Stafford, D. G., Sessions, R., 1998. Crust and upper mantle P wave velocity structure beneath Valles caldera, New Mexico: Results from the Jemez teleseismic tomography experiment. *J. Geophys. Res.* 103, 24,301–24,320.
- Steinhoff, J., Fan, M., Wang, L., 2000. A new Eulerian method for the computation of propagating short acoustic and electromagnetic pulses. *J. Comp. Phys.* 157, 683–706.
- Sussman, S., Smereka, P., Osher, S. J., 1994. A level set approach for computing solutions to incompressible two-phase flow. *J. Comp. Phys.* 114, 146–159.
- Symes, W. W., Qian, J., 2003. A slowness matching eulerian method for multivalued solutions of eikonal equations. *SIAM J. Sci. Comput.* 19, 501–526.

- Thomson, C. J., Gubbins, D., 1982. Three-dimensional lithospheric modelling at NOR-SAR: linearity of the method and amplitude variations from the anomalies. *Geophys. J. Royal Astr. Soc.* 71, 1–36.
- Thurber, C. H., 1983. Earthquake locations and three-dimensional crustal structure in the Coyote Lake area, central California. *J. Geophys. Res.* 88, 8226–8236.
- Thurber, C. H., Ellsworth, W. L., 1980. Rapid solution of ray tracing problems in heterogeneous media. *Bull. Seism. Soc. Am.* 70, 1137–1148.
- Toomey, D. R., Solomon, S. C., Purdy, G. M., 1994. Tomographic imaging of the shallow crustal structure of the East Pacific Rise at 9°30'N. *J. Geophys. Res.* 99, 24,135–24,157.
- Um, J., Thurber, C., 1987. A fast algorithm for two-point seismic ray tracing. *Bull. Seism. Soc. Am.* 77, 972–986.
- van Trier, J., Symes, W. W., 1991. Upwind finite-difference calculation of traveltimes. *Geophysics* 56, 812–821.
- VanDecar, J. C., James, D. E., Assumpção, M., 1995. Seismic evidence for a fossil mantle plume beneath South America and implications for plate driving forces. *Nature* 378, 25–31.
- Velis, D. R., Ulrych, T. J., 1996. Simulated annealing two-point ray tracing. *Geophys. Res. Lett.* 23, 201–204.
- Velis, D. R., Ulrych, T. J., 2001. Simulated annealing ray tracing in complex three-dimensional media. *Geophys. J. Int.* 145, 447–459.
- Vesnaver, A., Böhm, G., Madrussani, G., Rossi, G., Granser, H., 2000. Depth imaging and velocity calibration by 3D adaptive tomography. *First Break* 18, 303–312.
- Vidale, J. E., 1988. Finite-difference calculations of traveltimes. *Bull. Seism. Soc. Am.* 78, 2062–2076.
- Vidale, J. E., 1990. Finite-difference calculations of traveltimes in three dimensions. *Geophysics* 55, 521–526.
- Vinje, V., Åstebøl, K., Iversen, E., Gjøystdal, H., 1999. 3-D ray modelling by wavefront construction in open models. *Geophysical Prospecting* 64, 1912–1919.

- Vinje, V., Iversen, E., Åstebøl, K., Gjøystdal, H., 1996. Estimation of multivalued arrivals in 3D models using wavefront construction - Part I. *Geophysical Prospecting* 44, 819–842.
- Vinje, V., Iversen, E., Gjøystdal, H., 1993. Traveltime and amplitude estimation using wavefront construction. *Geophysics* 58, 1157–1166.
- Virieux, J., Farra, F., 1991. Ray tracing in 3-d complex isotropic media: An analysis of the problem. *Geophysics* 56, 2057–2069.
- Wang, Y., Houseman, G. A., 1994. Inversion of reflection seismic amplitude data for interface geometry. *Geophys. J. Int.* 117, 92–110.
- Wang, Y., Pratt, R. G., 1997. Sensitivities of seismic traveltimes and amplitudes in reflection tomography. *Geophys. J. Int.* 131, 618–642.
- White, D. J., 1989. Two-dimensional seismic refraction tomography. *Geophys. J.* 97, 223–245.
- Wiggins, S. M., Dorman, L. M., Cornuelle, B. D., Hildebrand, J. A., 1996. Hess deep rift valley structure from seismic tomography. *J. Geophys. Res.* 101, 22,335–22,353.
- Williamson, P. R., 1990. Tomographic inversion in reflection seismology. *Geophys. J. Int.* 100, 255–274.
- Xu, S., Lambaré, G., 2004. Fast migration/inversion with multivalued rayfields: Part 1 - Method, validation test, and application in 2D to Marmousi. *Geophysics* 69, 1311–1319.
- Xu, S., Lambaré, G., Calandra, H., 2004. Fast migration/inversion with multivalued rayfields: Part 1 - Applications to the 3d seg/eage salt model. *Geophysics* 69, 1320–1328.
- Yao, H., van der Hilst, R. D., de Hoop, M. V., 2006. Surface-wave array tomography in SE Tibet from ambient seismic noise and two-station analysis - I. Phase velocity maps. *Geophys. J. Int.* 166, 732–744.
- Zelt, B. C., Ellis, R. M., Zelt, C. A., Hyndman, R. D., Lowe, C., Spence, G. D., Fisher, M. A., 2001. Three-dimensional crustal velocity structure beneath the Strait of Georgia, British Columbia. *Geophys. J. Int.* 144, 695–712.
- Zelt, C. A., 1996. Seismic velocity structure of the central Chilean margin near the sub-

- ducting Juan Fernández ridge: effective inversion of travelttime data across complex, laterally varying structure. In: Abstracts. 7th Int. Symp., Deep Seismic Profiling of the Continents. Asilomar, CA.
- Zelt, C. A., 1999. Modelling strategies and model assessment for wide-angle seismic travelttime data. *Geophys. J. Int.* 139, 183–204.
- Zelt, C. A., Barton, P. J., 1998. Three-dimensional seismic refraction tomography: A comparison of two methods applied to data from the Faeroe Basin. *J. Geophys. Res.* 103, 7187–7210.
- Zelt, C. A., Smith, R. B., 1992. Seismic travelttime inversion for 2-D crustal velocity structure. *Geophys. J. Int.* 108, 16–34.
- Zhang, J., Rector III, J. W., Hoversten, G., 2005. Eikonal solver in the celerity domain. *Geophys. J. Int.* 162, 1–8.
- Zhang, J., Toksöz, M. N., 1998. Nonlinear refraction travelttime tomography. *Geophysics* 63, 1726–1737.
- Zhao, A., Zhongjie, Z., Teng, J., 2004. Minimum travel time tree algorithm for seismic ray tracing: improvement in efficiency. *J. Geophys. Eng.* 1, 245–251.
- Zhao, D., Hasegawa, A., Horiuchi, S., 1992. Tomographic imaging of *P* and *S* wave velocity structure beneath Northeastern Japan. *J. Geophys. Res.* 97, 19,909–19,928.
- Zhou, B., Greenhalgh, S. A., 2005. 'shortest path' ray tracing for most general 2d/3d anisotropic media. *J. Geophys. Eng.* 2, 54–63.