PEAT8002 - SEISMOLOGY
Lecture 8: Ray tracing in practice

Nick Rawlinson

Research School of Earth Sciences
Australian National University

- Although 1-D whole Earth models are an acceptable approximation in some applications, lateral heterogeneity is significant in many regions of the Earth (e.g. subduction zones) and therefore needs to be accounted for.
- Ray tracing in laterally heterogeneous media is non-trivial, and many different schemes have been devised in the last few decades.
- The following techniques will be discussed in this lecture:
  - Shooting and bending methods of ray tracing
  - Finite difference solution of the eikonal equation
  - Shortest Path Ray tracing (SPR)

- Shooting methods of ray tracing are conceptually simple; they formulate the kinematic ray equation as an initial value problem, which allows rays 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.

- The appropriate form of the equation required to solve the initial value problem depends largely on the choice of parameterisation used to represent velocity variations.
- 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
- Analytic ray tracing can also be applied to other parameterisations; for example, triangular or tetrahedral meshes in which the velocity gradient is constant.

- The expression for ray trajectory in a medium with a constant velocity gradient can be expressed in various ways, including parametrically as

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

  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 traveltime is then given by

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

  where T_0 is the traveltime 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 nth power of slowness $1/v^n$.
- Although analytic ray tracing is possible for a few special cases, in general one needs to solve the kinematic ray tracing equation using numerical methods.
- This usually requires the ray equation to be reduced to a convenient first order initial value system of equations, which can be done in a variety of ways.

- In this case, we will derive initial value ray equations by considering the following unit vector in the direction of the ray
$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s} = [\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta],$$
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 $+$ve x-axis in xy plane).
- Substitution of this expression into $\frac{\mathrm{d}}{\mathrm{d}s}\left[U\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}\right] = \nabla U$ and application of the product rule yields:

$$\frac{\partial U}{\partial x} = U\cos\theta\cos\phi\frac{\mathrm{d}\theta}{\mathrm{d}s} - U\sin\theta\sin\phi\frac{\mathrm{d}\phi}{\mathrm{d}s} + \sin\theta\cos\phi\frac{\mathrm{d}U}{\mathrm{d}s}$$
$$\frac{\partial U}{\partial y} = U\cos\theta\sin\phi\frac{\mathrm{d}\theta}{\mathrm{d}s} + U\sin\theta\cos\phi\frac{\mathrm{d}\phi}{\mathrm{d}s} + \sin\theta\sin\phi\frac{\mathrm{d}U}{\mathrm{d}s}$$
$$\frac{\partial U}{\partial z} = -U\sin\theta\frac{\mathrm{d}\theta}{\mathrm{d}s} + \cos\theta\frac{\mathrm{d}U}{\mathrm{d}s}$$

- These three equations can be rearranged to remove the $\mathrm{d}U/\mathrm{d}s$ term and produce expressions for $\mathrm{d}\theta/\mathrm{d}s$ and $\mathrm{d}\phi/\mathrm{d}s$, which produces

$$\frac{\mathrm{d}x}{\mathrm{d}s} = \sin\theta\cos\phi$$

$$\frac{\mathrm{d}y}{\mathrm{d}s} = \sin\theta\sin\phi$$

$$\frac{\mathrm{d}z}{\mathrm{d}s} = \cos\theta$$

$$\frac{\mathrm{d}\theta}{\mathrm{d}s} = \frac{\cos\theta}{U}\left[\cos\phi\frac{\partial U}{\partial x} + \sin\phi\frac{\partial U}{\partial y}\right] - \frac{\sin\theta}{U}\frac{\partial U}{\partial z}$$

$$\frac{\mathrm{d}\phi}{\mathrm{d}s} = \frac{1}{U\sin\theta}\left[\cos\phi\frac{\partial U}{\partial y} - \sin\phi\frac{\partial U}{\partial x}\right]$$

- 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.
- The above initial value formulation of the kinematic ray tracing equations uses path length s 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.
- This can be achieved by using the following simple set of relationships

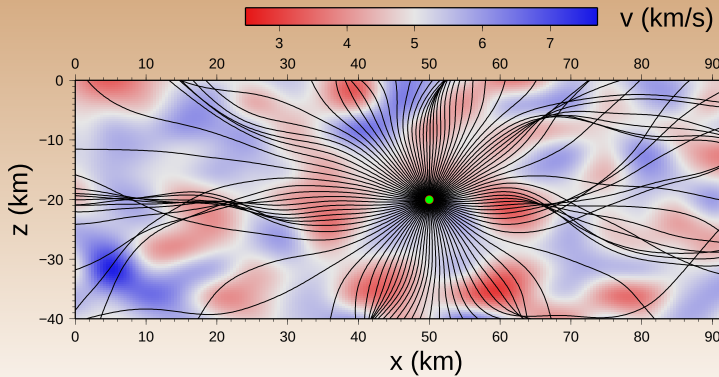
$$\frac{\mathrm{d}\theta}{\mathrm{d}s} = U\frac{\mathrm{d}\theta}{\mathrm{d}t}, \quad \frac{\mathrm{d}\phi}{\mathrm{d}s} = U\frac{\mathrm{d}\phi}{\mathrm{d}t}, \quad \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}s} = U\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}, \quad \frac{\partial U}{\partial \mathbf{x}} = -\frac{1}{v^2}\frac{\partial v}{\partial \mathbf{x}},$$

Shooting method
The initial value problem

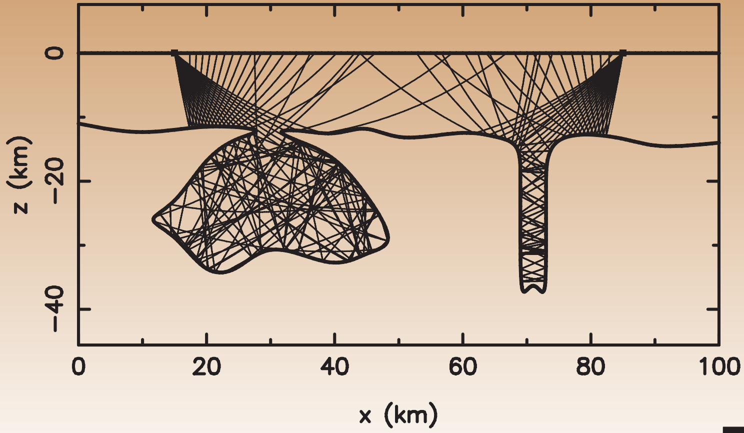
- Thus, the following system of equations results:

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

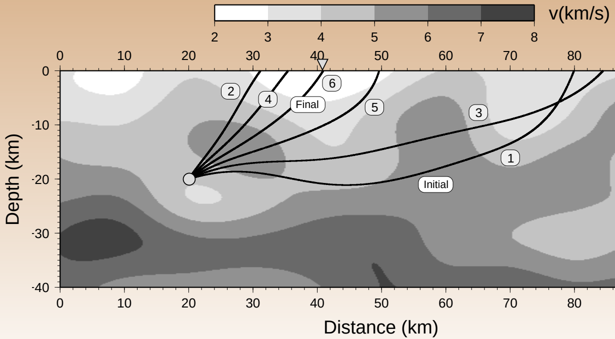
- The example below was computed by solving the above equation with a 4th order Runge Kutta method.



- In the presence of interfaces, Snell's law can be applied.



- 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



THE AUSTRALIAN NATIONAL UNIVERSITY

- If a ray 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 endpoint (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

$$x_e(\theta_o, \phi_o) = X_r$$
$$y_e(\theta_o, \phi_o) = Y_r$$

- 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.

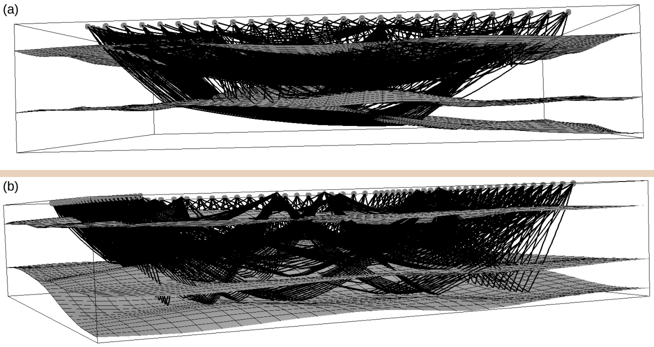
- Since the optimisation problem is non-linear, a range of iterative non-linear and fully non-linear schemes can be applied.
- A common iterative non-linear scheme is Newton's method, which amounts to iterative application of the following system of equations:

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

- Thus, given some starting initial trajectory θ_o^0, ϕ_o^0, solution of this equation provides an updated initial trajectory θ_o^1, ϕ_o^1, and the process is repeated until an appropriate tolerance criterion is met.

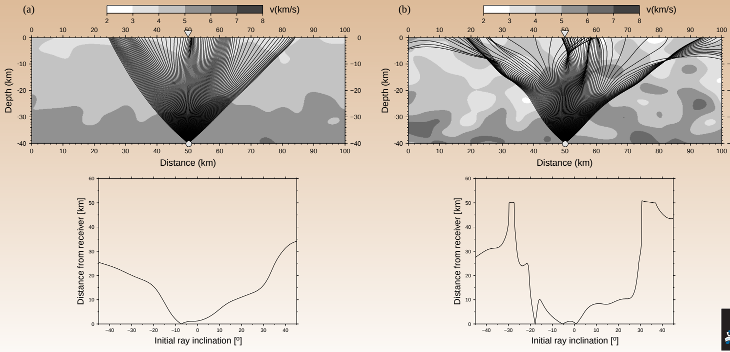
- The success of this scheme depends largely on two factors: (1) accurate calculation of the partial derivative matrix, and (2) obtaining an initial guess ray that will converge to the correct minimum under the assumption of local linearity.
- Both of these requirements can be difficult to satisfy, particularly in complex media.
- One way of obtaining an accurate initial guess ray is to shoot 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 until a suitably accurate initial ray is obtained.
- The partial derivative matrix can be directly computed by applying paraxial ray theory.

- The example below shows two-point ray paths computed through a 3-D layered model using an iterative non-linear shooting scheme.



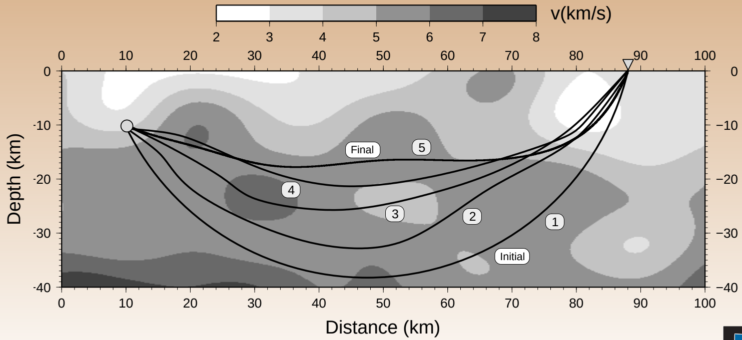
Shooting method
The boundary value problem

- The example below demonstrates the non-linear relationship between initial ray trajectory and distance from ray end point to receiver. Fully non-linear optimisation schemes can be successfully applied in complex media, but tend to be time consuming.



THE AUSTRALIAN NATIONAL UNIVERSITY

- 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).



- 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.
- The traveltime T of a ray path between source S and receiver R can in general be expressed by the integral

$$T = \int_S^R U ds,$$

where U is slowness and s is pathlength.
- The ray path can be described parametrically by a monotonic function λ, the normalised path length ($\lambda = s/L$, 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

$$\frac{\mathrm{d}s}{\mathrm{d}\lambda} = \sqrt{\dot{\mathbf{r}} \cdot \dot{\mathbf{r}}} = \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} = F,$$

  where $(\dot{})$ denotes differentiation with respect to λ, and the use of normalised path length means that $F = L$ and $\mathrm{d}F/\mathrm{d}\lambda = 0$.
- The traveltime can therefore be written:

$$T = \int_S^R UF\mathrm{d}\lambda.$$

- The ray tracing equations can be obtained by extremizing this integral using the calculus of variations.

- For any integrand $G(\lambda, \mathbf{r}(\lambda), \dot{\mathbf{r}}(\lambda))$ the Euler-Lagrange equations are
$$\frac{\partial G}{\partial \mathbf{r}} - \frac{\mathrm{d}}{\mathrm{d}\lambda}\left[\frac{\partial G}{\partial \dot{\mathbf{r}}}\right] = 0.$$
- In our case, the integrand is $G = UF = U\sqrt{\dot{\mathbf{r}} \cdot \dot{\mathbf{r}}}$, and substitution into the above equation yields
$$U\ddot{\mathbf{r}} + (\dot{\mathbf{r}} \cdot \nabla U)\dot{\mathbf{r}} - (\dot{\mathbf{r}} \cdot \dot{\mathbf{r}})\nabla U = 0.$$
- It is easy to show that only two of these equations are independent, 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 $\mathrm{d}F/\mathrm{d}\lambda = 0$ (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

$$U\ddot{x} + U_y\dot{y}\dot{x} + U_z\dot{z}\dot{x} - U_x(\dot{y}^2 + \dot{z}^2) = 0$$

$$U\ddot{y} + U_x\dot{x}\dot{y} + U_z\dot{z}\dot{y} - U_y(\dot{x}^2 + \dot{z}^2) = 0$$

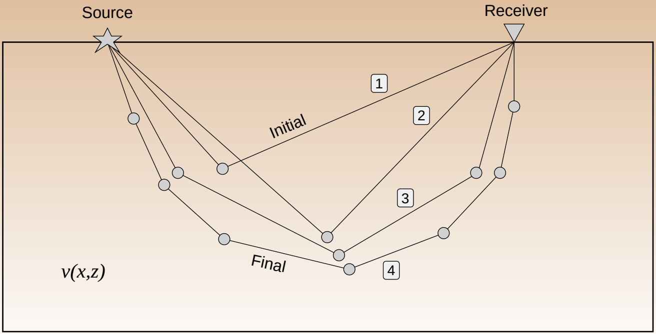
$$\dot{x}\ddot{x} + \dot{y}\ddot{y} + \dot{z}\ddot{z} = 0$$

  where $\nabla U = (U_x, U_y, U_z)$.
- 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 $\mathbf{r}(\lambda)^{n+1} = \mathbf{r}(\lambda)^n + \delta\mathbf{r}(\lambda)^n$.

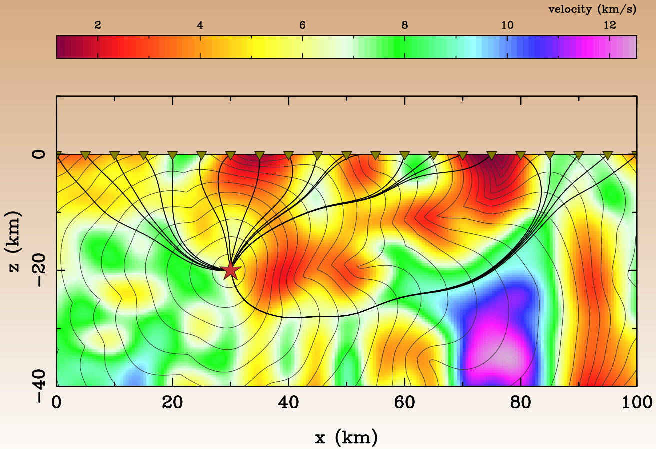
- Substitution of this expression into the previous equation and linearising the resulting equations for $\delta \mathbf{r}(\lambda)^n$ allows solutions to be obtained using, for example, second order finite difference techniques.
- The iterative process can be continued until some suitable convergence criterion, based on the path perturbation integrated along the ray, is satisfied.
- Bending methods can also be applied in the presence of interfaces, 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.

- Pseudo-bending methods are similar in principle to the bending scheme described previously, but avoid direct solution of the ray equations.
- One of the most common pseudo-bending schemes 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.

- Despite the relatively crude approximations made in pseudo-bending, it is often more computationally efficient than conventional bending schemes, and has become quite popular for problems which require large traveltime datasets to be predicted, e.g. 3-D local earthquake tomography



- The non-linearity of the two-point problem is the main limitation of both shooting and bending methods of ray tracing.



- 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, grid based traveltime schemes have a number of clear advantages:
  - Most are capable of computing traveltimes to all points of a medium, and will locate diffractions in ray shadow zones.

- other advantages include
  - 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.
  - 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.
  - 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:
  - 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.
  - Most practical schemes compute first-arrivals only - thus, features such as wavefront triplications cannot be predicted.
  - Quantities other than traveltime (such as amplitude) are difficult to compute accurately without first extracting path geometry and applying ray based techniques
- Two commonly used grid based schemes will now be discussed.

- Eikonal solvers seek finite difference solutions to the eikonal equation throughout a gridded velocity field.
- One of the first schemes (published in 1988) progressively computes the traveltime field outwards from the source along an expanding square in 2-D.
- The partial derivative terms in the eikonal equation are approximated by:

$$\frac{\partial T}{\partial x} = \frac{T_{i,j} + T_{i,j+1} - T_{i+1,j} - T_{i+1,j+1}}{2\delta x}$$

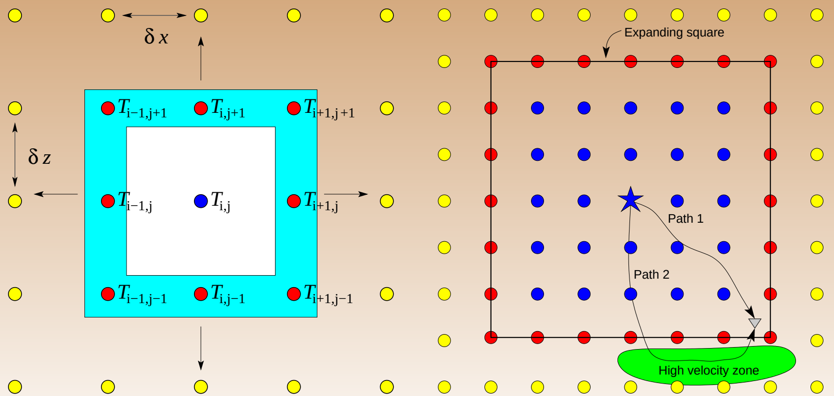
$$\frac{\partial T}{\partial z} = \frac{T_{i,j} + T_{i+1,j} - T_{i,j+1} - T_{i+1,j+1}}{2\delta z}$$

- Substitution into the eikonal equation yields:

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

  which is a quadratic equation for the traveltime at the new point $T_{i+1,j+1}$. $\bar{s}$ is the the average slowness of all four points defining a cell.
- The resultant scheme is fast and accurate (2nd order), but the expanding square formalism for the propagation of the computational front means that stability issues may arise.

- Schematic illustration of the expanding square scheme, and how it can breach causality.



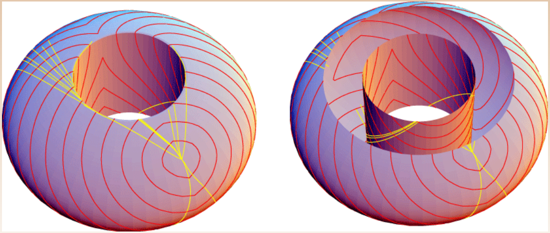
- One of the more recently developed grid based eikonal solvers which is both highly robust and computationally efficient is the so called Fast Marching Method (FMM).
- It was originally developed in the field of computational mathematics for solving various types of interface evolution problems, and to date has been applied in numerous areas of the physical sciences including optimal path planning, medical imaging, geodesics, and photolithographic development.
- Compared to the expanding square scheme, FMM is much more robust, as it overcomes the causality issue by using the evolving wavefront as the computational front.
- FMM also overcomes the problem of ∇T not being spatially differentiable at every point for a first-arrival traveltime field.

Grid based schemes

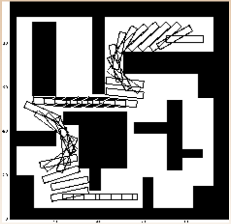
Fast Marching Method

Medical imaging

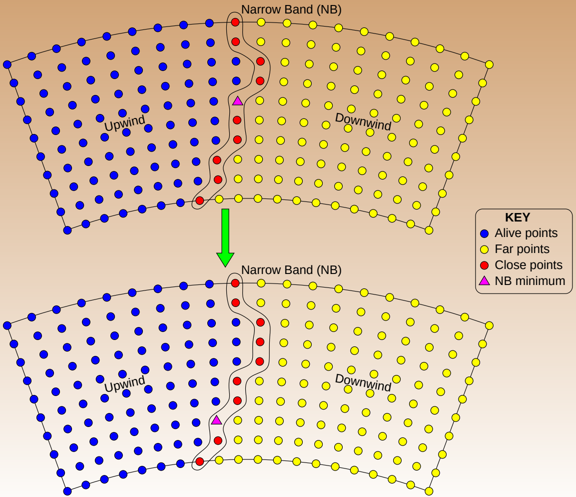
Geodesics



Robotic navigation



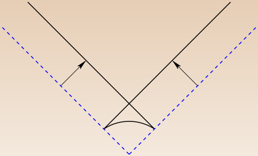
- Narrow band evolution scheme used by FMM for the ordered update of grid points. The narrow band advances from the *close* point with minimum traveltime.



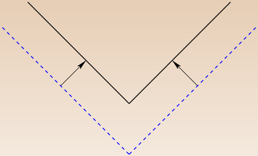
Grid based schemes

Fast Marching Method

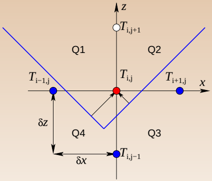
- Discontinuities in first-arrival wavefronts usually arise from discarding later-arriving information.
- FMM overcomes this problem by using an upwind entropy-satisfying finite difference stencil which properly respects the flow of information by considering solutions from each quadrant.



Swallowtail formation



First−arrival wavefront



Update scheme

- The entropy-satisfying upwind scheme that is often used can be written:

$$\begin{bmatrix} \max(D_a^{-x}T, -D_b^{+x}T, 0)^2 + \\ \max(D_c^{-y}T, -D_d^{+y}T, 0)^2 + \\ \max(D_e^{-z}T, -D_f^{+z}T, 0)^2 \end{bmatrix}_{ijk}^{\frac{1}{2}} = s_{i,j,k},$$

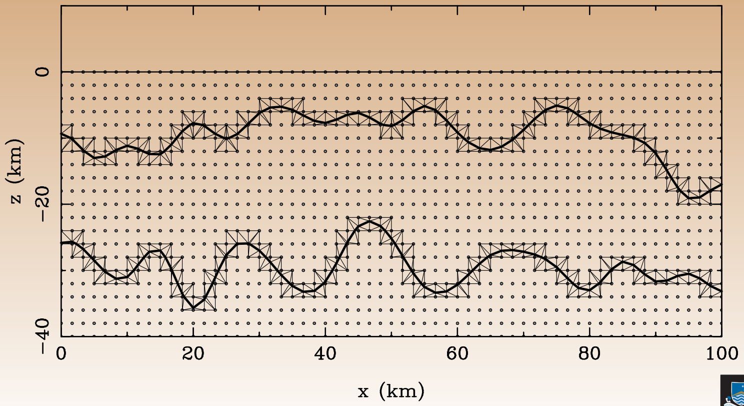
  where the integer variables a, b, c, d, e, f define the order of accuracy of the upwind finite difference operator used in each of the six cases.

- In a Cartesian coordinate system, the first and second order operators for $D^{-x}T_i$ are

$$D_1^{-x}T_{i,j,k} = \frac{T_{i,j,k} - T_{i-1,j,k}}{\delta x}$$

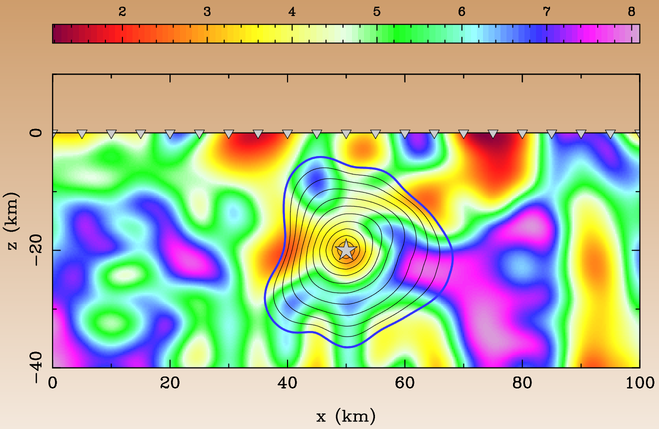
$$D_2^{-x}T_{i,j,k} = \frac{3T_{i,j,k} - 4T_{i-1,j,k} + T_{i-2,j,k}}{2\delta x}$$

- In layered media, an adaptive triangular mesh can be used to locally suture the irregular interface nodes to the regular nodes of the velocity grid.



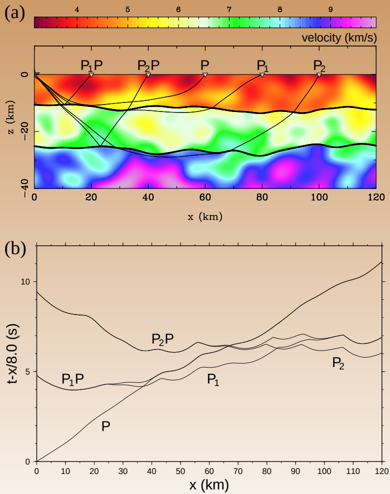
Grid based schemes

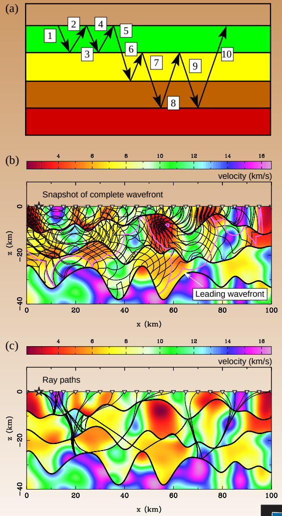
FMM examples



Grid based schemes

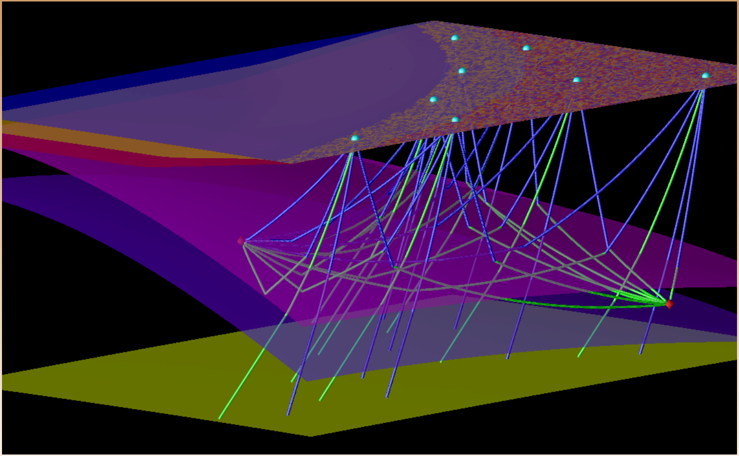
FMM examples



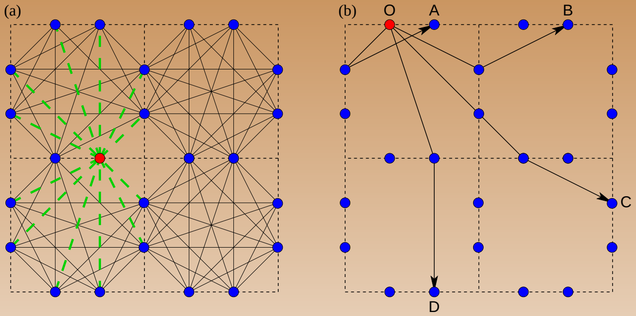


Grid based schemes

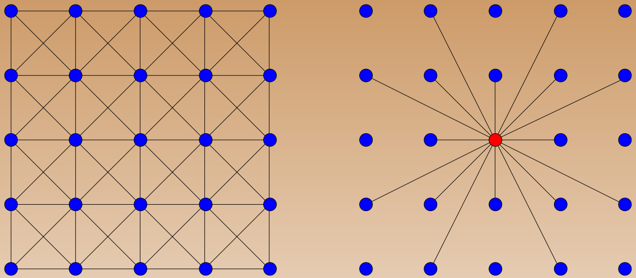
FMM examples



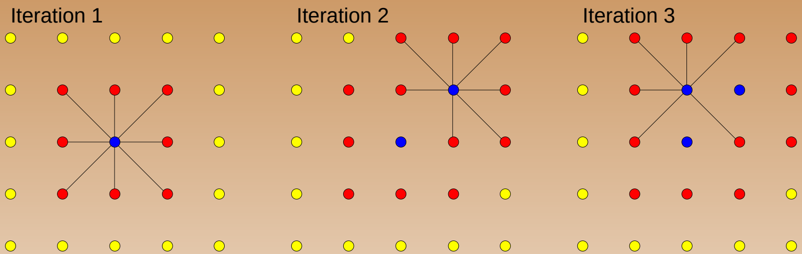
- Shortest path ray tracing or SPR is another popular method for determining first-arrival traveltimes at all points of a gridded velocity field.
- Rather than solve a differential equation, a network or graph is formed by connecting neighbouring nodes with traveltime path segments.
- Dijkstra-like algorithms 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.



- Example of a shortest path network built on a grid of constant velocity cells. Two network nodes are placed on the edge of each cell boundary. The green dashed lines highlight the connections from a single node. The plot on the right show the shortest paths between node O and a selection of surrounding nodes.



- Shortest path network built on a grid of velocity nodes. The plot on the left shows the use of a stencil with at most 8 connections per node. The plot on the right shows a 16-node stencil, which allows smaller path deviations to be more accurately represented.

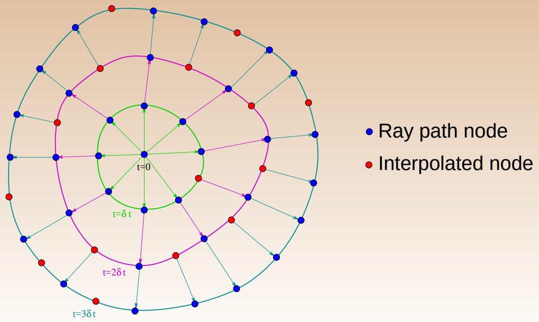


- Three iterations of a simple shortest path scheme using a forward star with 8 connections. Blue dots have known traveltimes, red dots have trial traveltimes, and yellow dots are yet to have traveltime computed.
- Note the similarity of this update scheme with that of FMM.

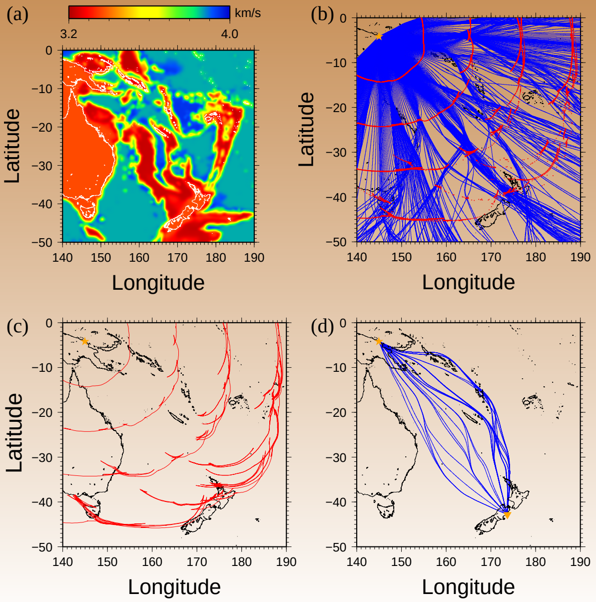
- Errors in SPR are due to finite node spacing and the angular distribution of node connectors. A coarse grid of nodes may poorly approximate the true velocity variations, while a limited range of angles between adjacent connectors may not allow for an accurate representation of the path.
- The accuracy 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.
- 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.

- 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.
- The ray tracing schemes discussed so far cannot be used to compute all arrivals in heterogeneous media.
- A technique knows as wavefront construction, which can find all arrivals will now briefly be discussed.

- Wavefront construction techniques represent the wavefront by a set of points, and generates a new wavefront at time $t_i + \delta t$ from a previous wavefront at time t_i by using initial value ray tracing and interpolation.
- It can be readily extended to 3-D, applied to anisotropic media, and allows amplitudes and synthetic seismograms to be calculated.



Ray tracing in practice
Wavefront construction



From Rawlinson et al. (2006)