PEAT8002 - SEISMOLOGY
Lecture 16: Seismic Tomography I

Nick Rawlinson

Research School of Earth Sciences
Australian National University

- Seismic data represent one of the most valuable resources for investigating the internal structure and composition of the Earth.
- One of the first people to deduce Earth structure from seismic records was Mohorovičić, a Serbian seismologist who, in 1909, observed two distinct traveltime curves from a regional earthquake.
- He determined that one curve corresponded to a direct crustal phase and the other to a wave refracted by a discontinuity in elastic properties between crust and upper mantle. This world-wide discontinuity is now known as the Mohorovičić discontinuity or Moho for short.
- On a larger scale, the method of Herglotz and Wiechart was first implemented in 1910 to construct a 1-D whole Earth model.

- Today, an abundance of methods exist for determining Earth structure from seismic waves. Different components of the seismic record may be used, including traveltimes, amplitudes, waveform spectra, full waveforms or the entire wavefield.
- Source-receiver configurations also differ - receiver arrays may be in-line or 3-D, sources may be close or distant to the receiver array, sources may be natural or artificial, and the scale of the study may be from tens of meters to the whole Earth.
- In this lecture, we will review a particular class of method for extracting information on Earth structure from seismic data, known as seismic tomography.

- Seismic tomography combines data prediction with inversion in order to constrain 2-D and 3-D models of the Earth represented by a significant number of parameters.
- If we represent some elastic property of the subsurface (e.g. velocity) by a set of model parameters $\mathbf{m}$, then a set of data (e.g. traveltimes) $\mathbf{d}$ can be predicted for a given source-receiver array by line integration through the model.
- The relationship between data and model parameters, $\mathbf{d} = \mathbf{g}(\mathbf{m})$, forms the basis of any tomographic method.
- For an observed dataset $\mathbf{d}_{obs}$ and an initial model $\mathbf{m}_0$, the difference $\mathbf{d}_{obs} - \mathbf{g}(\mathbf{m}_0)$ gives an indication of how well the current model predictions satisfy the data.

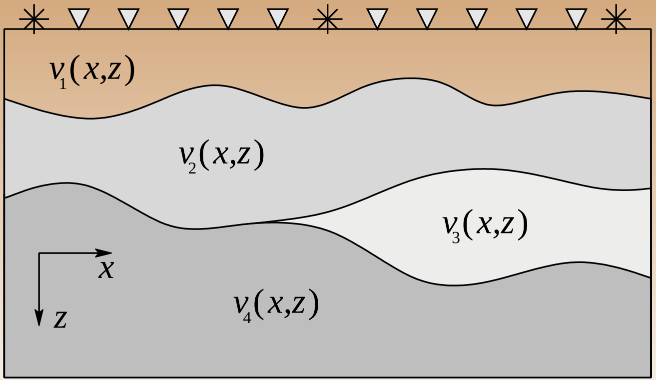
- The inverse problem in tomography is then to manipulate $\mathbf{m}$ in order to minimise the difference between observed and predicted data subject to any regularisation that may be imposed.
- The reliability of the final model will depend on a number of factors including:
  - how well the observed data are satisfied by the model predictions
  - assumptions made in parameterising the model
  - errors in the observed data
  - accuracy of the method for determining model predictions $\mathbf{g}(\mathbf{m})$
  - the extent to which the data constrain the model parameters
- The tomographic method therefore depends implicitly on the general principles of inverse theory.

Seismic tomography in a nutshell

- Model parameterisation The seismic structure of the target region is represented by set of model parameters.
- Forward calculation A procedure is defined for the prediction of model data (e.g. traveltimes) given a set of values for the model parameters.
- Inversion Automated adjustment of the model parameter values with the object of better matching the model data to the observed data subject to any regularisation that may be imposed.
- Solution non-uniqueness Investigate solution robustness (e.g. estimates of covariance and resolution, synthetic reconstructions).

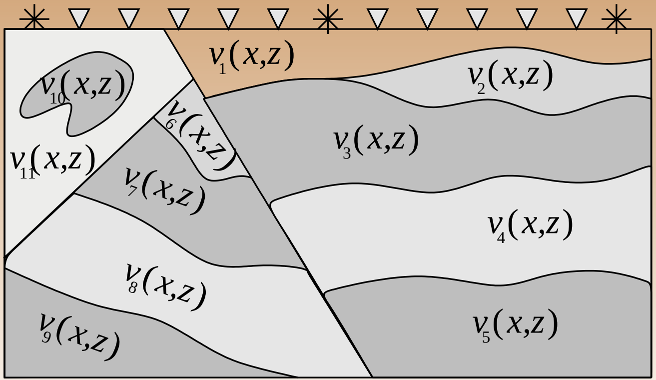
- Given our fundamental relationship between data and model parameters $\mathbf{d} = \mathbf{g}(\mathbf{m})$, we need to be able to represent variations in structure such that we can compute $\mathbf{g}(\mathbf{m})$, and satisfy the data observations.
- In laterally heterogeneous media, the most general type of parameterisation needs to allow for both continuous and discontinuous variations in seismic properties.
- 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 assumptions are usually made.
- In seismic tomography, it is common practice, when interfaces are required, to represent the medium by layers, within which properties vary continuously, separated by sub-horizontal interfaces which vary in depth.

- Schematic illustration of a layered parameterisation.



- The relative simplicity of this representation makes it amenable to fast and robust data prediction, and 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 within which material property varies smoothly, but is discontinuous across element boundaries
- However, in the presence of such complexity, the data prediction problem and inverse problem are much more difficult to solve.

- Schematic illustration of a block-model parameterisation.



- Common parameterisations used to describe wave speed variations (or other seismic properties) in a continuum include constant velocity (or slowness) blocks, triangular/tetrahedral meshes within which velocity is constant or constant in gradient, and grids of velocity nodes which are interpolated using a predefined function.
- 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 unrealisticly 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.

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

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

  where $V(x_i, y_j, z_k)$ are the velocity (or some other seismic property) values at eight grid points surrounding (x, y, z).

- In the above equation v is continuous, but its gradient ∇v is not (i.e. C^0 continuity). Despite this limitation, pseudo linear interpolation between nodes is a popular choice in seismic tomography.

- 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, B-splines and splines under tension.
- Cubic B-splines are particularly useful, 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

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

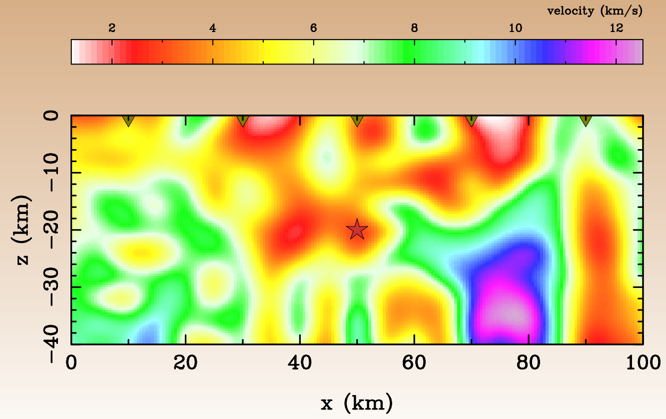
  where $\mathbf{q}_{i,j,k} = (V_{i,j,k}, \mathbf{p}_{i,j,k})$. Thus, the three independent variables $0 \leq u, v, w \leq 1$ define the velocity distribution in each volume element.
- The weighting factors $\{b_i\}$ are the uniform cubic B-spline functions.

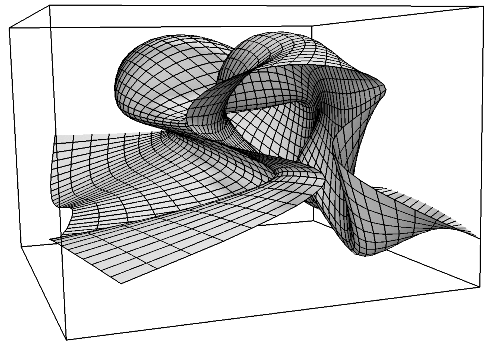
- Rather than use velocity grids in the spatial domain to describe smooth media, one could also exploit the wavenumber domain by employing a spectral parameterisation.
- These are often popular for global applications e.g. spherical harmonics, but can also be used for problems on a local or regional scale.
- One approach (in 2-D) is to use a truncated Fourier series expressed as:

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

- In the above equation, $\mathbf{r} = x\mathbf{i} + z\mathbf{j}$ 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 the above equation 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 tomographic studies, particularly those involving global datasets.

- Example of a continuously varying velocity medium defined using a regular mesh of grid points to control the values of a cubic B-spline continuum.



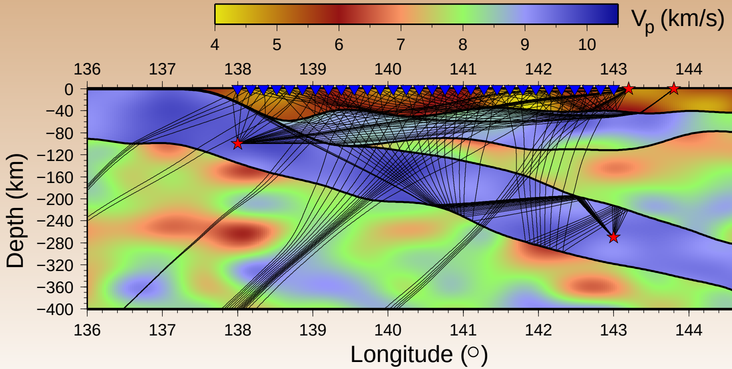


- Interfaces are often described using equivalent parameterisations to those used for the continuum.
- For example, linear segments (2-D volume), triangular meshes (3-D volume) or nodes with a specified interpolant are common.

- Finally, it is worth noting that irregular parameterisations are occasionally used.
- 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
- Completely unstructured meshes, such as those that use Delaunay tetrahedra or Voronoi polyhedra, offer high levels of adaptability, but have special book-keeping requirements when solving the forward problem of data prediction.

- The forward problem in seismic tomography requires the calculation of model data, given a set of values for the model parameters.
- In the case of traveltime tomography, the aim is to compute source-receiver traveltimes for a given velocity model.
- For surface wave tomography, on the other hand, the aim is usually to try and compute a long period synthetic waveform.
- In the majority of cases, some form of ray tracing or wavefront tracking is used. These techniques were discussed in detail in Lecture 8.

- The example below shows a variety of paths tracked through a complex subduction zone using FMM. This type of approach can be used to solve the forward problem in seismic tomography.



- The inversion step, which involves the adjustment of the model parameters **m** to better satisfy the observed data $\mathbf{d}_{obs}$ through the known relationship $\mathbf{d} = \mathbf{g}(\mathbf{m})$, can be performed in a number of ways.
- When the model unknowns **m** are velocity or slowness, then the functional **g** is non-linear because the ray path and hence traveltime depends on the velocity structure.
- Ideally, an inversion scheme should account for this non-linearity if it is present.
- The three approaches to solving the inversion step that will be considered below are backprojection, gradient methods and global optimisation techniques.

- Backprojection methods have traditionally been quite popular in traveltime tomography, but it is probably true to say that their popularity is on the wane.
- In an earlier lecture, we showed that the perturbation of a ray path only has a second order effect on traveltime (Fermat's principle). In terms of slowness, this is written:

$$\delta t = \int_{L_0} \delta s(\mathbf{x}) \mathrm{d}l + O(\delta s(\mathbf{x})^2)$$

- If a continuum is described by M constant slowness blocks, then the discrete form of the above integral equation for N rays can be written:

$$\mathbf{d} = \mathbf{G}\mathbf{m}$$

- In the above equation, **d** are the traveltime residuals, **m** the slowness perturbations and **G** an $N \times M$ matrix of ray lengths l_{ij} corresponding to the distance traversed by each ray in each block.
- Note that for the general case **m** (e.g. velocity nodes, interface depths etc.) $\mathbf{G} = \partial \mathbf{g} / \partial \mathbf{m}$ where $\mathbf{g}(\mathbf{m})$ is the model prediction.
- Many of the elements of **G** will be zero since each ray path will usually only traverse a small subset of the M blocks.
- Backprojection methods can be used to solve $\mathbf{d} = \mathbf{G}\mathbf{m}$ for the slowness perturbations **m** by iteratively mapping traveltime anomalies into slowness perturbations along the ray paths until the data are satisfied.

- Two well known backprojection techniques for solving $\mathbf{d} = \mathbf{G}\mathbf{m}$ are the Algebraic Reconstruction Technique (ART) and the Simultaneous Iterative Reconstruction Technique (SIRT), both of which originate from medical imaging.
- In ART, the model is updated on a ray by ray basis. The residual d_n for the n^{th} ray path is distributed along the path by adjusting each component of $\mathbf{m}$ in proportion to the length l_{nj} of the ray segment in the j^{th} cell:

$$m_j^{k+1} = m_j^k + \frac{t_n^{k+1} l_{nj}}{\sum_{m=1}^{M} l_{nm}^2}$$

- In the above equation, $t_n^{k+1} = d_n - t_n^k$ is the difference between the residuals at the 0^{th} and k^{th} iteration, m_j^k is the approximation to the j^{th} model parameter at the k^{th} iteration, $m_j^1 = 0$ and $t_n^1 = 0$.

- The main problem with ART is that it suffers from poor convergence properties, but it has been used in cross-hole and local earthquake tomography in the past.
- SIRT addresses some of the convergence problems associated with ART by averaging the perturbations applied to each parameter from all the rays that are influenced by the parameter.
- Thus, the SIRT algorithm may be written:

$$m_j^{k+1} = m_j^k + \frac{1}{R_j^k} \sum_{n=1}^{R_j^k} \left[\frac{t_n^{k+1} l_{nj}}{\sum_{m=1}^{M} l_{nm}^2} \right]$$

where R_j^k is the number of rays that the j^{th} model parameter influences for the k^{th} iteration.

- The SIRT method has been used in the inversion of teleseismic traveltime residuals, and in the inversion of reflection traveltimes for both velocity structure and interface depth.
- Inversion using backprojection tends to be computationally more rapid at each iteration compared to other techniques, but often converges more slowly and with less stability.
- This is at least partly due to the use of more *ad hoc* regularisation (like spatial averaging) compared to, for example, the formal inclusion of such constraints in the inversion permitted by gradient methods.

- The inverse problem in seismic tomography can be formulated as one of minimising an objective function consisting of a data residual term and one or more regularisation terms.
- As before, let $\mathbf{d}$ denote a data vector of length N which is dependent on a model vector $\mathbf{m}$ of length M as $\mathbf{d} = \mathbf{g}(\mathbf{m})$.
- For an initial estimate $\mathbf{m}_0$ of the model parameters, comparing $\mathbf{d} = \mathbf{g}(\mathbf{m}_0)$ with the observed traveltimes $\mathbf{d}_{obs}$ gives an indication of the accuracy of the model.
- The misfit can be quantified by constructing an objective function $S(\mathbf{m})$, consisting of a weighted sum of data misfit and regularisation terms, that is to be minimised.

- An essential component of the objective function is a term $\Psi(\mathbf{m})$ which measures the difference between the observed and predicted data.
- If it is assumed that the error in the relationship $\mathbf{d}_{obs} \approx \mathbf{g}(\mathbf{m}_{true})$ is Gaussian, then a least squares or L_2 measure of this difference is suitable:

$$\Psi(\mathbf{m}) = \|\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs}\|^2$$

- If uncertainty estimates have been made for the observed data (usually based on picking error), then more accurate data are given a greater weight in the objective function by writing $\Psi(\mathbf{m})$ as:

$$\Psi(\mathbf{m}) = (\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs})^T \mathbf{C}_d^{-1} (\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs})$$

where $\mathbf{C}_d$ is a data covariance matrix.

- If the errors are assumed to be uncorrelated, then $\mathbf{C}_d = [\delta_{ij}(\sigma_d^j)^2]$ where σ_d^j is the uncertainty of the j^{th} data point.
- A common problem with tomographic inversion is that not all model parameters will be well constrained by the data alone (i.e. the problem may be under-determined or mixed-determined).
- A regularisation term $\Phi(\mathbf{m})$ is often included in the objective function to provide additional constraints on the model parameters, thereby reducing the non-uniqueness of the solution.
- The regularisation term is typically defined as:

$$\Phi(\mathbf{m}) = (\mathbf{m} - \mathbf{m}_0)^T \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0)$$

where $\mathbf{C}_m$ is an *a priori* model covariance matrix.

- If uncertainties in the initial model are assumed to be uncorrelated, then $\mathbf{C}_m = [\delta_{ij}(\sigma_m^j)^2]$ where σ_m^j is the uncertainty associated with the j^{th} model parameter of the initial model.
- The effect of $\Phi(\mathbf{m})$ is to encourage solution models $\mathbf{m}$ that are near a reference model $\mathbf{m}_0$. The values used in $\mathbf{C}_m$ are usually based on prior information.
- Another approach to regularisation is the minimum structure solution which attempts to find an acceptable trade-off between satisfying the data and finding a model with the minimum amount of structural variation.
- One way of including this requirement in the objective function is to use the term:

$$\Omega(\mathbf{m}) = \mathbf{m}^T\mathbf{D}^T\mathbf{D}\mathbf{m}$$

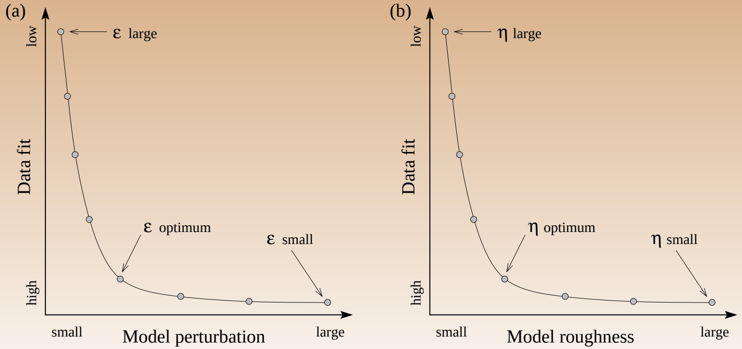
where $\mathbf{Dm}$ is a finite difference estimate of a specified spatial derivative.

- An explicit smoothing term in the objective function may be necessary if crude parameterisations such as constant velocity blocks are used to simulate a continuously varying velocity field.
- However, if an implicitly smooth parameterisation like cubic splines is used, then an explicit smoothing term may be unnecessary.
- Using the L_2 terms described above, the objective function $S(\mathbf{m})$ can be written in full as:

$$S(\mathbf{m}) = \frac{1}{2}\left[\Psi(\mathbf{m}) + \epsilon\Phi(\mathbf{m}) + \eta\Omega(\mathbf{m})\right]$$

where ϵ is referred to as the *damping factor* and η as the *smoothing factor* (when **D** is the second derivative operator, which is usually the case).

- ϵ and η govern the trade-off between how well the solution $\mathbf{m}_{est}$ will satisfy the data, how closely $\mathbf{m}_{est}$ is to $\mathbf{m}_0$, and the smoothness of $\mathbf{m}_{est}$.



- Gradient-based inversion methods make use of the derivatives of $S(\mathbf{m})$ at a specified point in model space.
- A basic assumption that is shared by all practical gradient methods is that $S(\mathbf{m})$ is sufficiently smooth to allow a local quadratic approximation about some current model:

$$S(\mathbf{m} + \delta\mathbf{m}) \approx S(\mathbf{m}) + \hat{\gamma}\delta\mathbf{m} + \tfrac{1}{2}\delta\mathbf{m}^T\hat{\mathbf{H}}\delta\mathbf{m}$$

- In the above equation, $\delta\mathbf{m}$ is a perturbation to the current model and $\hat{\gamma} = \partial S/\partial\mathbf{m}$ and $\hat{\mathbf{H}} = \partial^2 S/\partial\mathbf{m}^2$ are the gradient vector and Hessian matrix respectively.

- For our chosen form of objective function,

$$\hat{\gamma} = \mathbf{G}^T \mathbf{C}_d^{-1}[\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs}] + \epsilon \mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{m}_0) + \eta \mathbf{D}^T \mathbf{D}\mathbf{m}$$

$$\hat{\mathbf{H}} = \mathbf{G}^T \mathbf{C}_d^{-1}\mathbf{G} + \nabla_{\mathbf{m}} \mathbf{G}^T \mathbf{C}_d^{-1}[\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs}] + \epsilon \mathbf{C}_m^{-1} + \eta \mathbf{D}^T \mathbf{D}$$

  where $\mathbf{G} = \partial \mathbf{g}/\partial \mathbf{m}$ is the Fréchet matrix of partial derivatives calculated during the solution of the forward problem.
- Since $\mathbf{g}$ is generally non-linear, the minimisation of $S(\mathbf{m})$ requires an iterative approach:

$$\mathbf{m}_{n+1} = \mathbf{m}_n + \delta \mathbf{m}_n$$

  where $\mathbf{m_0}$ is the initial model.

- The objective function is minimised for the current ray path estimate at each step to produce $\mathbf{m}_{n+1}$, after which new ray paths are computed for the next iteration.
- The iterations cease either when the observed traveltimes are satisfied or when the change in $S(\mathbf{m})$ with iteration gets sufficiently small.
- One approach to estimating $\delta\mathbf{m}_n$ is the Gauss-Newton method. It locates the minimum of the tangent paraboloid to $S(\mathbf{m})$ at $\mathbf{m}_n$.
- At the minimum of S, the gradient will vanish, so $\mathbf{m}$ is required such that:

$$\mathbf{F}(\mathbf{m}) = \mathbf{G}^T\mathbf{C}_d^{-1}(\mathbf{g}(\mathbf{m})-\mathbf{d}_{obs})+\epsilon\mathbf{C}_m^{-1}(\mathbf{m}-\mathbf{m}_0)+\eta\mathbf{D}^T\mathbf{D}\mathbf{m} = 0$$

where $\mathbf{F}(\mathbf{m}) = \hat{\gamma}$.

- If we are at some point $\mathbf{m}_n$, then a more accurate estimate $\mathbf{m}_{n+1}$ can be obtained using a Taylor series expansion:

$$F_i(m^1_{n+1}, \ldots, m^M_{n+1}) = F_i(m^1_n, \ldots, m^M_n) + \sum_{j=1}^{M}(m^j_{n+1} - m^j_n) \left.\frac{\partial F_i}{\partial m^j}\right|_{\mathbf{m}_n} = 0$$

- This may be rewritten as

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \left[\frac{\partial \mathbf{F}}{\partial \mathbf{m}}\right]^{-1}_n [\mathbf{F}_n] = \mathbf{m}_n - \left[\frac{\partial^2 S}{\partial \mathbf{m}^2}\right]^{-1}_n \left[\frac{\partial S}{\partial \mathbf{m}}\right]_n$$

where $(\partial S/\partial \mathbf{m})_n$ is the gradient vector and $(\partial^2 S/\partial \mathbf{m}^2)_n$ is the Hessian matrix.

- Substitution of the expressions given earlier for $(\partial S/\partial \mathbf{m})_n$ and $(\partial^2 S/\partial \mathbf{m}^2)_n$ yields the Gauss-Newton solution:

$$\begin{aligned}\delta \mathbf{m}_n &= -[\mathbf{G}_n^T \mathbf{C}_d^{-1} \mathbf{G}_n + \nabla_{\mathbf{m}} \mathbf{G}_n^T \mathbf{C}_d^{-1} (\mathbf{g}(\mathbf{m}_n) - \mathbf{d}_{obs}) + \epsilon \mathbf{C}_m^{-1} + \eta \mathbf{D}^T \mathbf{D}]^{-1} \\ &\quad \times [\mathbf{G}_n^T \mathbf{C}_d^{-1} [\mathbf{g}(\mathbf{m}_n) - \mathbf{d}_{obs}] + \epsilon \mathbf{C}_m^{-1} (\mathbf{m}_n - \mathbf{m}_0) + \eta \mathbf{D}^T \mathbf{D} \mathbf{m}_n]\end{aligned}$$

- If instead we assume that the relationship $\mathbf{d} = \mathbf{g}(\mathbf{m})$ is linearisable then

$$\mathbf{d}_{obs} \approx \mathbf{g}(\mathbf{m}_0) + \mathbf{G}(\mathbf{m} - \mathbf{m}_0)$$

or $\delta \mathbf{d} = \mathbf{G} \delta \mathbf{m}$ with $\delta \mathbf{d} = \mathbf{d}_{obs} - \mathbf{g}(\mathbf{m}_0)$ and $\delta \mathbf{m} = \mathbf{m} - \mathbf{m}_0$.

- Because a one-step solution is possible in the linear case, the objective function is sometimes written:

$$S(\mathbf{m}) = \frac{1}{2}\Big[(\mathbf{G}\delta\mathbf{m} - \delta\mathbf{d})^T \mathbf{C}_d^{-1}(\mathbf{G}\delta\mathbf{m} - \delta\mathbf{d}) + \epsilon\delta\mathbf{m}^T\mathbf{C}_m^{-1}\delta\mathbf{m} + \eta\delta\mathbf{m}^T\mathbf{D}^T\mathbf{D}\delta\mathbf{m}\Big]$$

  where last term on the RHS smooths the perturbations to the prior model.

- The functional in this case is:

$$\mathbf{F}(\mathbf{m}) = \mathbf{G}^T\mathbf{C}_d^{-1}(\mathbf{G}\delta\mathbf{m} - \delta\mathbf{d}) + \epsilon\mathbf{C}_m^{-1}\delta\mathbf{m} + \eta\mathbf{D}^T\mathbf{D}\delta\mathbf{m} = 0$$

- The solution can therefore be written as:

$$\delta \mathbf{m} = [\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \epsilon \mathbf{C}_m^{-1} + \eta \mathbf{D}^T \mathbf{D}]^{-1} \mathbf{G}^T \mathbf{C}_d^{-1} \delta \mathbf{d}$$

- When no smoothing is used ($\eta = 0$), then:

$$\delta \mathbf{m} = [\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \mathbf{C}_m^{-1}]^{-1} \mathbf{G}^T \mathbf{C}_d^{-1} \delta \mathbf{d}$$

  which is the maximum likelihood solution to the inverse problem or the stochastic inverse.

- The above two expressions are sometimes referred to as the Damped Least Squares (DLS) solutions to the inverse problem (particularly when $\eta = 0$).

- Solutions to the above equations can be obtained using standard methods for solving large linear systems of equations.
- Techniques such as LU decomposition or SVD can be used to solve small to moderate sized problems.
- For large problems, conjugate gradients and LSQR are more appropriate, particular for sparse matrices.
- Methods for minimisation of the objective function that don’t require the solution of a large linear system of equations include steepest descent, conjugate gradient and subspace methods.

- The inversion methods described above are local in that they exploit information in regions of model space near an initial model estimate and thus avoid an extensive search of model space.
- Consequently, they cannot guarantee convergence to a global minimum solution. Local methods are prone to entrapment in local minima, especially if the subsurface velocity structure is complex and the starting model is not close to the true model.
- In many realistic applications, particularly at regional and global scales, the need for global optimisation techniques is hard to justify, because the *a priori* model information is relatively accurate and lateral heterogeneities are not very large

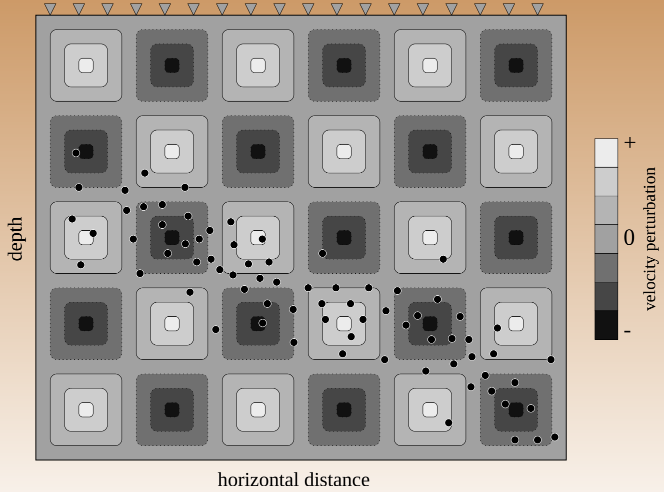
- However, the crust and lithosphere are generally less well constrained by *a priori* information and are also much more heterogeneous. This means that the initial model is likely to be more distant from the global minimum solution, and entrapment in a local minimum becomes more of a concern.
- Two non-linear schemes that have been used in seismic tomography are genetic algorithms and simulated annealing.
- Genetic algorithms use an analogue to biological evolution to develop new models from an initial pool of randomly picked models.
- Simulated annealing is based on an analogy with physical annealing in thermodynamic systems to guide variations to the model parameters.

- Global optimisation using stochastic methods is a rapidly developing field of science. However, current applications to seismic tomography problems have been limited due to computational expense.
- Hybrid approaches are sometimes implemented, which involve using a non-linear technique to find the global minimum solution of a coarse model (i.e. few model parameters). Iterative non-linear schemes are then applied to refine the solution.
- Simulated annealing has been used in the inversion of reflection traveltimes for 2-D velocity structure and interface geometry.
- Genetic algorithms have been used in the inversion of refraction traveltimes for 2-D velocity structure.

- The process of producing a solution to an inverse problem using the above methods is not complete until some estimate of solution robustness or quality is made.
- Simply producing a single solution that minimises an objective function (i.e. best satisfies the data and *a priori* constraints) without knowledge of resolution or non-uniqueness is inadequate.
- In seismic tomography, one common approach to assessing solution robustness is to apply synthetic resolution tests.
- Alternatively, if the assumption of local linearity is justifiable, formal estimates of model covariance and resolution can be made.

- Parameterisations that describe continuous variations in seismic properties often opt for resolution tests that attempt to reconstruct a synthetic model using the same source-receiver geometry as the real experiment.
- The rationale behind this approach is that if a known structure with similar length scales to the solution model can be recovered using the same (for linearised solutions) or similar (for iterative non-linear solutions) ray paths, then the solution model should be reliable
- The quality criterion is the similarity between the recovered model and the synthetic model.
- The so-called “checkerboard test", in which the synthetic model is divided into alternating regions of high and low velocity with a length scale equal (or greater) to the smallest wavelength structure recovered in the solution model, is a common test model.

Seismic tomography I
Solution non-uniqueness



- For an objective function with no smoothing term, the resolution matrix can be written:

$$\mathbf{R} = \mathbf{I} - \mathbf{C}_M \mathbf{C}_m^{-1}$$

- The corresponding a posteriori covariance matrix can be written

$$\mathbf{C}_M = \epsilon[\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \epsilon \mathbf{C}_m^{-1}]^{-1}$$

- These quantities are derived from linear theory. The diagonal elements of $\mathbf{C}_M$ indicate the posterior uncertainty associated with each model parameter.
- The diagonal elements of $\mathbf{R}$ range between zero and 1; in theory, when $\mathbf{R} = \mathbf{I}$, the model is perfectly resolved.