Introduction to Seismology 12.510 Sophie Michelet
March 28, 2005

Seismic tomography

We will study travel time tomography (body waves), which can be assimilated to imaging. But one has to be aware that *imaging* can be interpreted as reflection seismic, velocity reconstruction by the industry (SEG).
One can also look at surface wave tomography but these two types of tomography are not part of the same theory. There is no easy way to combine them; the assumptions made are different in each case.

The steps:

- We have observations, represented by dots on the figure below (picking of the different waves: P, S, PP…)

Image removed due to copyright considerations.

Please see:
Stein, S., and M. Wysession. Figure 3.5-3 in *An Introduction to Seismology, Earthquakes, and Earth Structure*. 1st ed. Malden, MA: Blackwell, September 2002, p. 163. ISBN: 0865420785.

- People generate linear regression to give the best fit of these observations
- We obtain models (ex: IASP91, PREM…)

[Wiechert and Herglotz (GER) found a 1D classical model from travel time curves.]

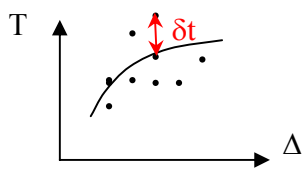
- We use an inversion technique to get the P & S wave velocity as a function of the depth (see the figure below).

Image removed due to copyright considerations.

Please see:
Stein, S., and M. Wysession. Figure 3.5-1 in *An Introduction to Seismology, Earthquakes, and Earth Structure*. 1st ed. Malden, MA: Blackwell, September 2002, p. 162. ISBN: 0865420785.

- From these models, we can produce travel time curves again.
- Go back to the observations again and look at the deviation from the model → the scattering around it.

$$\delta t = T_{obs} - T_{ref}$$



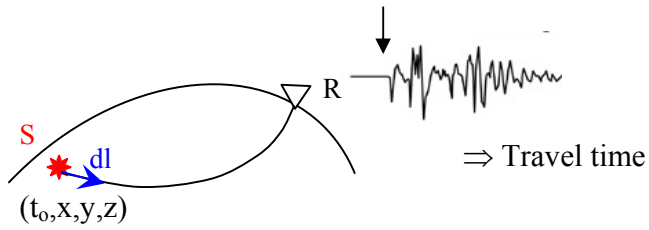
We look at the difference between the data and the model and NOT at the travel time itself!

OBJECTIVE: find a model (3D) that minimizes δt.

$$T_{obs} = \text{3D structure} + \underbrace{\text{error in source location}}_{\text{Hypocenter: to,x,y,z (or } \theta,\varphi,r)} + \text{noise (instrument, measurement)}$$

$$\Rightarrow \delta t = \delta t_{\mathbf{3D}} + \delta t_{\mathbf{mislocation}} + \delta t_{\mathbf{noise}}$$

But can we isolate $\delta t_{\mathbf{3D}}$?
What people do is combine the model ($\delta t_{\mathbf{3D}}$) and the source location ($\delta t_{\mathbf{mislocation}}$).
The noise is assumed to be white, with a Gaussian distribution.



If we know the 1D model, we can apply Snell's law to estimate the geometry:

$$T_{obs} = \int_{\substack{\text{3D}\\\text{ray}\\\text{path}}} \frac{1}{v(\underline{r})} dl$$

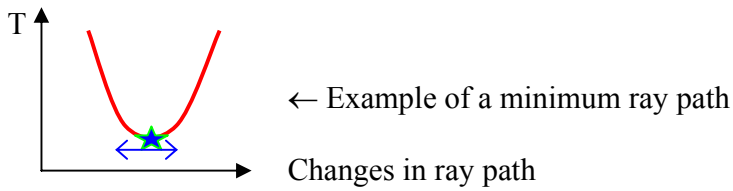
But we know that:

$$\nabla T = \frac{1}{c}\hat{k}$$

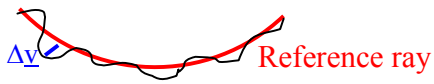
If c changes, the ray path changes. We end up with a **non-linear** problem. The ray path depends on the answer of the problem.

What do we need to do?
⇒ **linearize the inversion**

To do so, we use **the Fermat's principle**:



If we change a little bit the ray path around the optimum, we'll end up with a small change in travel time.



We have two kinds of "deviations" from the reference ray:

1. First contribution: effect of changes in velocity $\Delta\underline{v}$
2. Second contribution: effect of change in the ray path

Fermat's principle says that the 2nd effect is of 2nd order: we can ignore it!

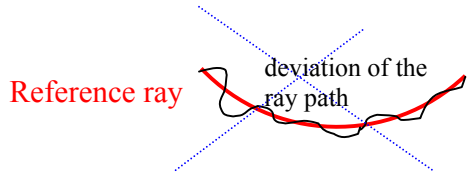
Linearization of the travel time

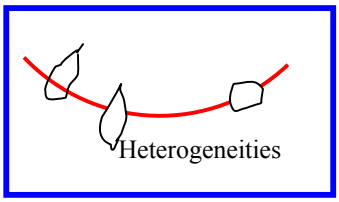
$$\delta t = T_{obs} - T_{ref} = T - T_o = \underbrace{\int \frac{1}{v(\underline{r})} dl}_{\text{3D ray (unknown)}} - \underbrace{\int \frac{1}{v_o(\underline{r})} dl'}_{\text{Reference ray (known), model}}$$

$$= \int s.dl - \int s_o.dl'$$

Because the deviation of the ray is ignored, we only look at the effect of the heterogeneities. It means that in linearizing the travel time, we obtain:

$$\underbrace{\int s.dl}_{\text{3D ray (unknown)}} \cong \underbrace{\int s_o.dl'}_{\text{Reference ray (known)}} + 2^{nd} \text{ order effect}$$





$$\delta t = \int (s - s_o).dl' = \int \Delta s.dl'$$

Reference ray (known)

In linearizing the problem, we get rid of the unknown ray. We can do our calculation in a reference earth model.

[Rayleigh 's principle is similar]

The travel time tomography is an *iterative process*:

- create 1D model
- ray tracing and get new rays in the model
- update ray geometry
- get the reference ray related to the 3D
  $\delta t = T_{obs} - T_{ref}(3D)$
  (The reference model does not have to be a 1D model.)

Linearization of the hypocenter mislocation

$$\delta t_{misloc} = \frac{\partial T}{\partial t_o}\delta t_o + \frac{\partial T}{\partial x}\delta x + \frac{\partial T}{\partial y}\delta y + \frac{\partial T}{\partial z}\delta z$$

$$\delta \otimes = \frac{\partial \otimes}{\partial \oplus}\delta \oplus$$

with t_o the origin time, (x,y,z) the location of the earthquake.

What we want to know is: Δs, δt_o, δx, δy, δz

From the travel time tomography, we get the wavespeed model and the estimation of the mislocation of the source.

- If we want to work with the **velocity**, not the slowness:

  Fermat's principle $-\int \frac{\Delta v}{v_o^2} dl'$ for small $\Delta v = v - v_o$

  $$\Delta s \cong -\frac{\Delta v}{v_o^2}$$ conversion

- **Inversion problem**: linear relationship between δt and the model.

observations $\delta t = T - T_o$

model perturbation ($m-m_o$)

$$d_i = \int_{volume} G_i(\underline{r})\mu(\underline{r})d\underline{r}$$

sensitivity kernel/ Frechet derivatives

$$d\otimes = \frac{\partial \otimes}{\partial \oplus}\delta \oplus$$

Partial derivatives / Frechet derivatives

  - The first thing to do next is to discretize the problem: do a **parametrization** (not work with a continuous integral).

$$\mu(\underline{r}) = \sum_{k=1}^{M} \gamma_k h_k(\underline{r})$$

γ_k: weight
h_k: basis function

For example,

1. *Fourier theory, harmonic functions:*
   $h_k(\underline{r}) = \cos(k\theta)$
   Just add a weight and we obtain a signal.

2. *Plane wave summation:*

$$\phi = \iiint \Phi(x,y,z,w) e^{i(kx-wt)} dk.dw$$

weight *basis function*

- We inject μ into the equation:

$$d_i = \sum_{k=1}^{M} \left\{ \underbrace{\int G_i(\underline{r}) h_k(\underline{r}) d\underline{r}} \right\} \gamma_k$$

kernel projected on the basis function h

$$d_i = \sum_{k=1}^{M} A_{ik} \gamma_k \quad \text{for i=1,2,…, N data}$$

sensitivity matrix
kernel projected on the basis functions

projector

$$\underline{d} = \underline{\underline{A}}\,\underline{m} \qquad m = (\gamma_1, \gamma_2, ..., \gamma_M)$$

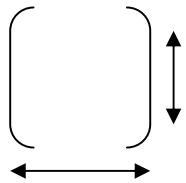
model (includes the values of the weight)

data

- <u>Green functions</u>: solution of a point source
  G_i is a Green function for a point perturbation. We need to do a convolution with a point perturbation in order to get the observations.

- M: number of model parameters
  N: number of observations (data)
  in general, N≠M.

  As a consequence, the matrix A is not square in general.

A = 

N rows

M columns

$\underline{m} = \underline{\underline{A}}^{-1} \underline{d}$ → we cannot do that!

We have to multiply by A^T:

$$A^T d = (A^T A)m$$

- $A^T A$ is a square matrix (M x M)

$\rightarrow \boxed{\hat{m} = (A^T A)^{-1} A^T \underline{d}}$ **Generalized Least Squares inversion**

$\| \hat{m} - (A^T A)^{-1} A^T \underline{d} \| = \varepsilon \rightarrow$ Goal is to **minimize ε**.

- Usually, we work with 10 million x 10 million matrix. So, we need to use numerical techniques to solve the problem: LSQR, SIRT…

- $\underline{d} = \underline{\underline{A}}\underline{m}$

  $\delta t = \int \Delta s.dl \rightarrow$ parametrize Δs

  $\Delta s = \sum_{k=1}^{M} \gamma_k h_k$

  <u>One way</u> is to take $\mathbf{h_k}$ as a series of **cells/blocks.**
  h_k will have a value for <u>r</u> inside the cell k and will be zero for <u>r</u> outside the cell k.

  <u>Another way:</u>
  - $\mathbf{h_k}$ as splines, wavelets
  - $\mathbf{h_k}$: **spherical harmonics** (in global seismology)

  $h_{klm} = f_k(\underline{r}) Y_l^m(\theta, \varphi)$

  surface spherical harmonic

  radial function

<u>model perturbation:</u>

$$\mu(\mathrm{r}) = \mu(\mathrm{r}, \theta, \varphi) \sum_k \sum_l \sum_m \left({}_k\alpha_l^m \cos(m\varphi) + {}_k\beta_l^m \sin(m\varphi) \right) f_k(r) P_l^m(\cos\theta)$$

for gravity $(1/\mathrm{r})^l$

Legendre polynomial

$$\sum \gamma_k \times h_k = \mu(\underline{r})$$

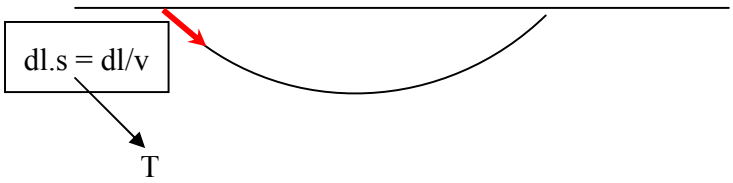
$$\sum weight \times basis\,function = \mathrm{model}$$

- $\delta t_i = \int \Delta s.dl = \sum_{k=1}^{Mcell} \Delta s_k (dl)_{ik}$

path length

in a block k

i: event-station pair



- d = A m

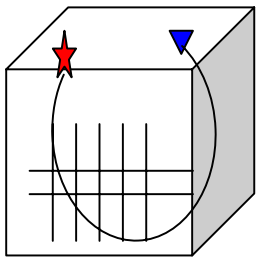
with A the sensitivity matrix containing the path length

$$Am = \begin{pmatrix} l_1\ l_2 \ldots\ldots\ldots l_m \\ \\ \\ \end{pmatrix} \begin{pmatrix} \Delta s_1 \\ \Delta s_2 \\ \ldots \\ \Delta s_m \end{pmatrix} = \begin{pmatrix} \delta t_1 \\ \delta t_2 \\ \ldots \\ \delta t_m \end{pmatrix}$$

model data vector

each ray gives a row

We have an average wavespeed along the ray. In order to construct a model vector, we need to get data from different rays crossing each other.
A is a sparse matrix. If we look at one ray:



M~300,000
~ 20 layers
the ray samples ~ 100 cells

A will have only ~ 100 elements non-zero.

The **good** thing about sparse matrix is that $A^T A$ is ~ diagonal.
The **problem** is that there are many singularities, which makes the inversion unstable (in that case, we need to add a damping factor or regularize the problem).

One possibility is to not use cells of the same size. Consequently, it reduces the number of cells; the inverse matrix is less singular. Nevertheless, the computation time increases.