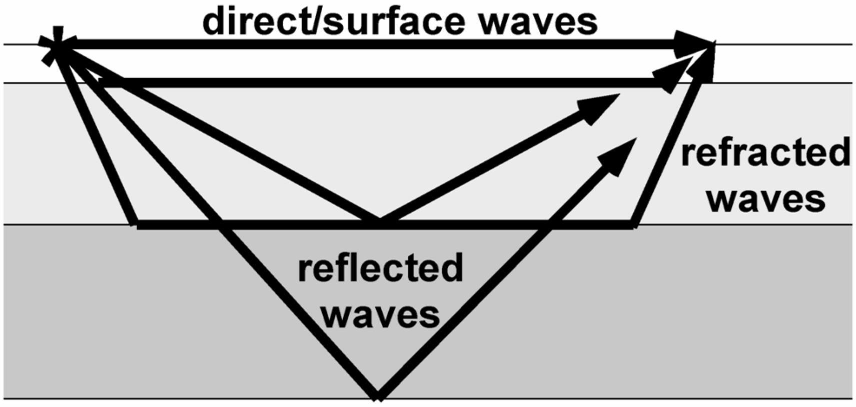
Fundamentals of Seismic Tomography and Full Waveform Inversion

Satish Singh
Institut de Physique du Globe de Paris
University of Paris City, France

Seismic methods are most powerful methods to characterise the nature of sub-surface. A seismic source is used to generate seismic energy. This seismic energy travels through the earth and get either reflected (scattered) and transmitted (propagation) through the earth and is recorded by set of receivers. Depending upon the source-receiver distance, we either record reflected energy and transmitted (refracted) energy. When the source-receiver distance is small as compared to the depth of the target, reflections are observed and when the source-receiver distance is large, refractions are observed. Seismic reflection method is used to obtain seismic image, commonly used for oil and gas exploration, whereas seismic refraction methods are used to determine large-scale velocity structures, mainly used by academic community to determine crustal and upper mantle velocity structures. Initially, these methods developed independently, but with the advent of modern acquisition technology, and the development of new theoretical methodology, the boundary between these methods have become less rigid. This is because, the physics of the seismic wave propagation and scattering is the same. I this course, I will take the advantage of these developments, and treat the seismic reflection and refraction problems together as a one problem. As industry professionals have extensive experience in seismic reflection imaging, I will not discuss seismic reflection imaging but mainly focus on refraction problems, while highlighting the link between reflections and refractions. This fits well with the development of tomography, based on arrival times, that first developed for refraction problems. On the other hand, the initial development in seismic full waveform started with the application to seismic reflection data, but its full potential was realised when it was applied to refraction data. So in this course I will cover the full offset range, from zero offset reflection to post-critical reflections and refraction or turning rays. In order to establish this link, I will first introduce seismic reflection and refractions using rays, their subtility under different conditions, such as the presence of low velocity layers, high velocity gradient layers, etc for simple 1D case, then introduce link between waves and rays, introduce different methods to compute travel times, develop concept of travel time tomography, present different methods, their strength and weaknesses, how to obtain uncertainty in velocity models, and then introduce seismic full waveform inversion, basic theory, different methods of waveform inversion, basic requirements, and finally show some applications of these two methods to sub-basalt imaging and reservoir characterisation problems, and suggest how the method could be used for monitoring. As I come from academia, where I am used to teach students on black board, I will follow the same approach using slide, but the text below will form the backbone of the course, which students can use afterwards.

Figure 1: Ray diagram for direct, refracted and reflected rays.



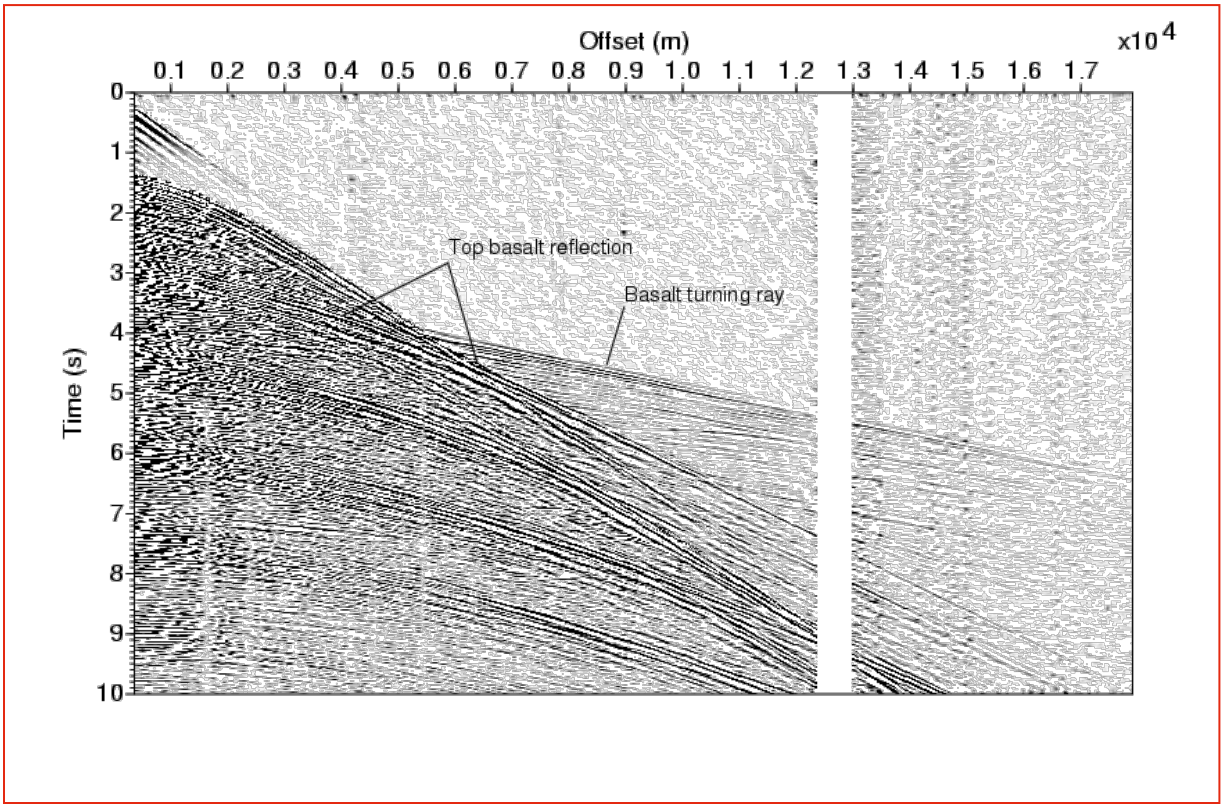


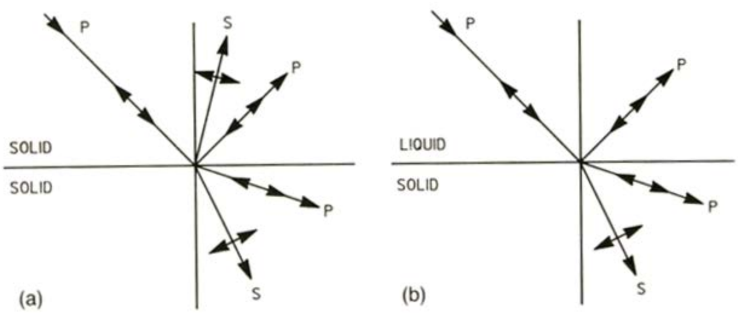
Figure 2: An example combined seismic reflection and refraction data as a function of distance (horizontal) and time (vertical). Turning rays corresponds to refraction. Reflection arrivals are also marked. Two ships, one carrying a 12 km streamer and another 6 km, were used to acquire 18 km offset data.

Although I will come back to wave theory later, here I would like to introduce the acoustic wave equation, which is fundamental for seismic methods:

$$\left[\frac{1}{k(r)}\frac{\partial^2}{\partial t^2} - \nabla\cdot\left(\frac{1}{\rho(r)}\nabla\right)\right]P(r,t;r_s) = s(r,t;r_s)$$

where *P* is wavefield that propagates in the earth generated by source *s,* which could a dynamite or aur gun source. ρ, κ are density and compressibility in the media. This equation can be solved using numerical methods or using ray theory.

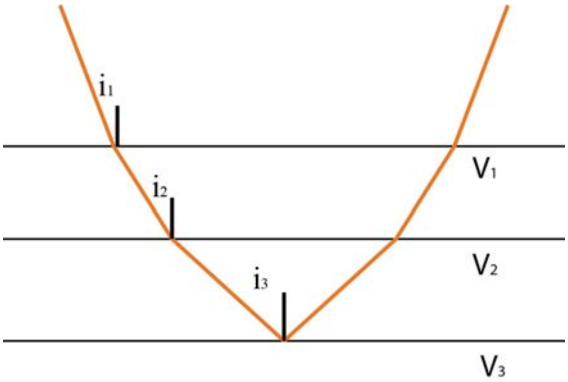
Basic Concept of Reflection and transmission



Snell's Law (fundamental for seismic waves)

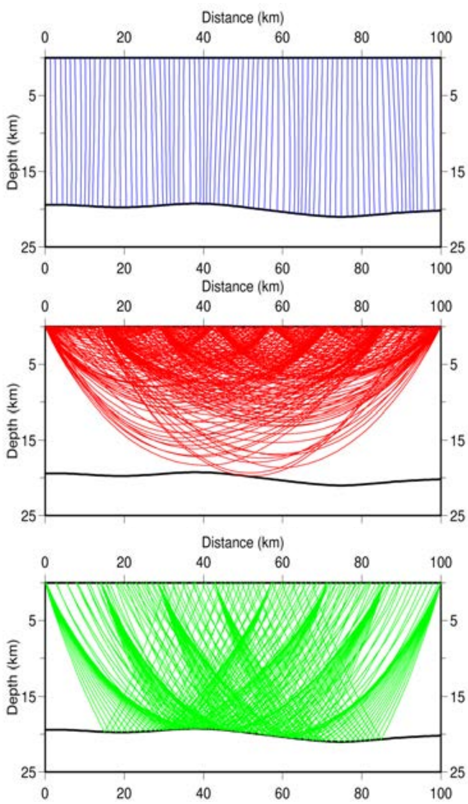
$$\frac{\sin i_1}{v_1} = \frac{\sin i_2}{v_2} = \frac{\sin i_3}{v_3} = \text{constant} = p \text{ (horizontal slowness)}$$

Horizontal slowness is constant along a ray path.



Vertical Slowness

$$q = \sqrt{\frac{1}{v^2} - p^2}$$



© Michael

Different rays:

Vertically propagating rays

Turning rays

Wide-angle reflection

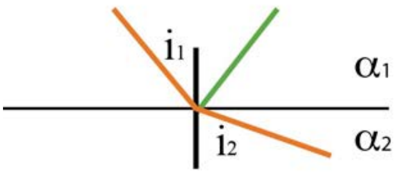
Basic Theory of Refraction Method

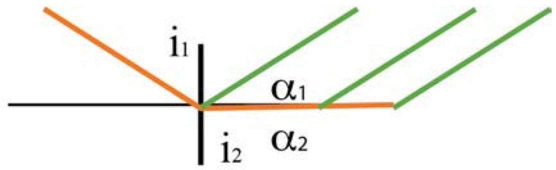
The travel time of first arrivals are generally used in refraction study. In order to understand the tomography, it is important to start with simple two layers example. Here, I provide detailed mathematical description, so that students can understand the basic concept if they wish to, but in the class I will simply the discussion, and focus on the concepts and the main point.

Two Layer Case: Let us assume that the crust beneath a refraction profile consists of two horizontal layers, with distinct and constant P-wave velocities α_1 and α_2 such that α_2>α_1. Energy from the source can reach the receiver directly through the top layer (direct wave), by reflection from the interface between the two layers, or travelling along the interface as a critically refracted wave or head wave. The head wave has a travel time corresponding to a ray, which has travelled down to the interface act the critical angle i_c with the velocity of layer 1, α_1, and then along the interface with the velocity of the lower layer, α_2, then back to the receiver, again at the critical angle with velocity, α_1.

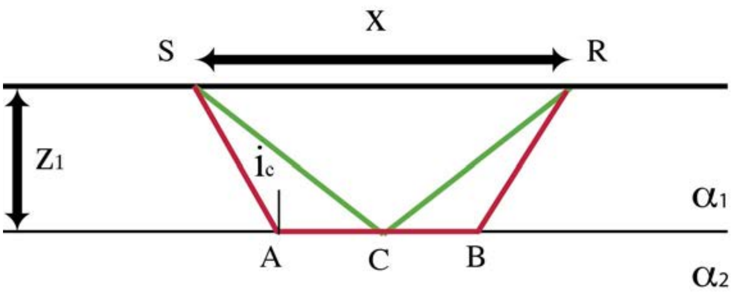
Recalling Snell's law for two layers

$$\frac{sin\, i_1}{\alpha_1} = \frac{sin\, i_2}{\alpha_2} = constant$$

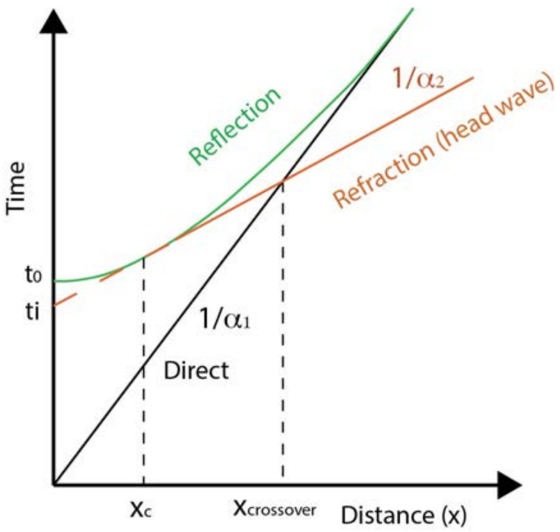




When refracted angle i_2=90°, the ray travels along the interface, and the angle is called critical angle, i_c. It is called head wave or interface wave. Although the wave travel at the interface with velocity in the lower medium, the part of energy is transmitted in the upper layer, and travel at the critical angle, and arrive at the receiver.



Ray diagram for reflected and refracted arrivals



Travel time for reflected and refracted arrivals

Direct waves:
$t = \frac{x}{\alpha_1}$ is straight line in the x-t plot.

Reflected wave: Travel time

$$t = \frac{SC}{\alpha_1} + \frac{CR}{\alpha_1}$$

$$SC = CR = \sqrt{z_1^2 + \frac{x^2}{4}}$$

Travel time

$$t = \frac{2}{\alpha_1}\sqrt{z_1^2 + \frac{x^2}{4}}$$

or

$$\alpha_1^2 t^2 = 4z_1^2 + x^2$$

is a parabola. This equation forms the basis for velocity analysis in the seismic reflection imaging.

Refracted or Head wave

$$t = \frac{SA}{\alpha_1} + \frac{AB}{\alpha_2} + \frac{BR}{\alpha_1}$$

Since i_2=90°, $\sin i_c = \frac{\alpha_1}{\alpha_2}$ from Snell's law, and
$SA = BR = \frac{z_1}{\cos i_c}$ (symmetry)
$AB = x - 2z_1 \tan i_c$

Travel time

$$t = \frac{2z_1}{\alpha_1 \cos i_c} + \frac{x}{\alpha_2} - \frac{2z_1 \tan i_c}{\alpha_2}$$

$$= \frac{2z_1}{\alpha_1 \cos i_c}[1 - \alpha_1/\alpha_2 \sin i_c] + \frac{x}{\alpha_2}$$

$$= \frac{2z_1}{\alpha_1}\cos i_c + \frac{x}{\alpha_2}$$
$$= \frac{2z_1}{\alpha_1}\sqrt{1-\frac{\alpha_1^2}{\alpha_2^2}} + \frac{x}{\alpha_2}$$
This is a equation of a straight line with slope=$1/\alpha_2$ and intercept time
$$t_i = \frac{2z_1}{\alpha_1}\sqrt{1-\frac{\alpha_1^2}{\alpha_2^2}}$$
at x=0.

Critical Distance: The shortest distance at which the head wave can be recorded is the critical distance, corresponding to critical angle, i_c, when i_2=90°, and AB=0,
$$x_c = 2z_1 \tan i_c$$
$$= \frac{2z_1\alpha_1}{\sqrt{\alpha_2^2-\alpha_1^2}}$$
The slope of the reflection parabola (tangent) can be defined as
$$\frac{dt}{dx} = \frac{1}{dx}\left[\frac{2}{\alpha_1}\sqrt{z_1^2+\frac{x^2}{4}}\right]$$
$$= \frac{x}{\alpha_1}\frac{1}{\sqrt{4z_1^2+x^2}}$$
At the critical distance, i_1=i_c, this slope is
$$\left.\frac{dt}{dx}\right]_{x=x_c} = \frac{2z_1 \tan i_c}{\alpha_1}\frac{1}{\sqrt{4z_1^2+4z_1^2\tan^2 i_c}}$$
$$= \frac{\sin i_c}{\alpha_1}$$
$$= \frac{1}{\alpha_2}$$
At critical distance, the head wave is tangent to the reflection parabola (hyperbola).

Crossover distance is the range at which the direct and head waves have the same time

t_{direct}=$t_{headwave}$
$$\frac{x_{cross}}{\alpha_1} = \frac{2z_1}{\alpha_1}\sqrt{1-\frac{\alpha_1^2}{\alpha_2^2}} + \frac{x_{cross}}{\alpha_2}$$
$$x_{cross} = 2z_1\sqrt{\frac{\alpha_2+\alpha_1}{\alpha_2-\alpha_1}}$$
So based on the above information, we could determine the velocity in the first layer, second layer and the depth of the interface as follows:

α_1 = inverse of the slope of the direct wave for $x \le x_{cross}$

α_2 = inverse of the slope of the head wave for $x \ge x_{cross}$

$$z_1 = \frac{t_i\alpha_1}{2}\frac{1}{\sqrt{1-\frac{\alpha_1^2}{\alpha_2^2}}}$$
t_i is the intercept time. Once we know the intercept time, we could compute the depth z_1.

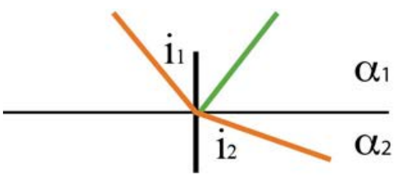
Multi-layer case: The travel times for a model consisting of n uniform horizontal layers of thickness z_j and P-wave velocity α_j are determined in the exactly the same way as the two-layer model. The only extra matter to be remembered is that the rays bend according to Snell's law as they cross each interface (i.e. horizontal slowness p is constant along each ray). The travel time for a wave refracted along the top of the mth layer is

$$t = \sum_{j=1}^{m-1} \left(\frac{2z_j}{\alpha_j} \sqrt{1 - \frac{\alpha_j^2}{\alpha_m^2}} \right) + \frac{x}{\alpha_m}$$

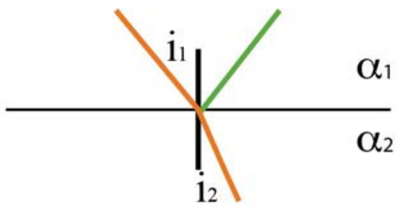
for $\alpha_j < \alpha_{j+1}$.

We cannot have two refraction arrivals at the same offset under normal circumstances.

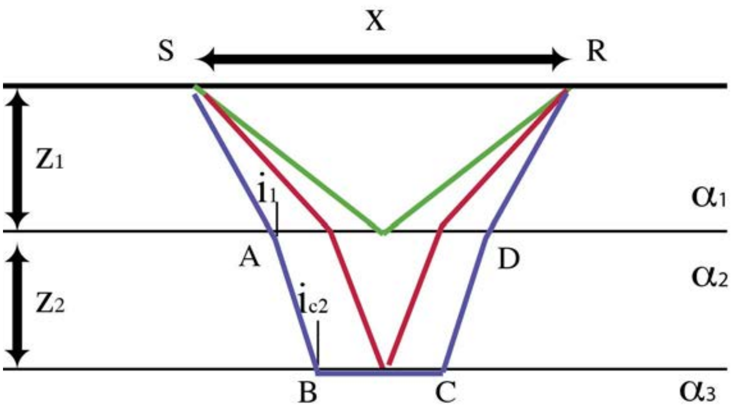
A low velocity sediment layer beneath high velocity basalt layer: A low-velocity layer cannot give rise to any head wave at its interface because the transmitted ray bends towards the normal. Furthermore, there are no critical angle reflections from an interface where the velocity contrast is negative.



for $\alpha_1 < \alpha_2$

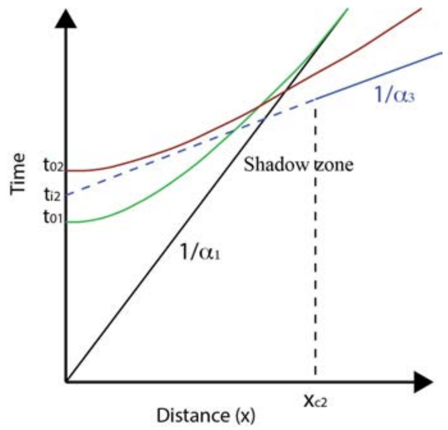


for $\alpha_1 > \alpha_2$



Ray path when
$\alpha_1 > \alpha_2 < \alpha_3$

Therefore, the presence or absence of critical angle arrivals can be used to get some idea about the presence of a high velocity layer or low velocity layer. Note that the critical distance depends on the velocity contrast at the interface following the Snell's law.

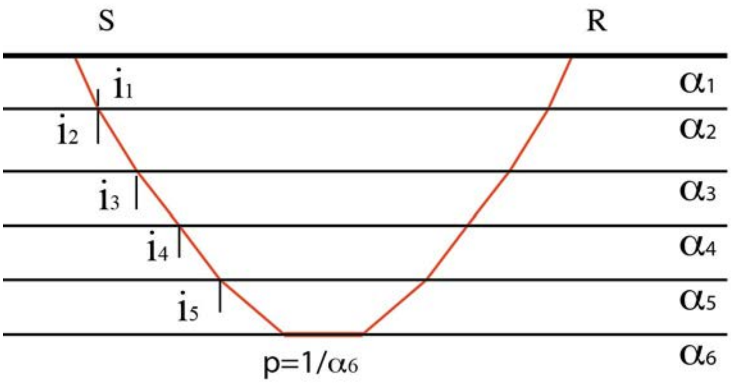


Travel time curve when
$\alpha_1 > \alpha_2 < \alpha_3$

Turning Rays: Another level of complexity arises when the earth is not layered cake and does not consist of thick constant velocity layers, but of very thin layers or velocity varied continuously. We will first examine a ray travelling downwards through a series thing layers, each of which is faster than the layers above. As discussed above, the horizontal slowness, also called ray parameter, p will constant along the ray

$$p = \frac{sin\, i_1}{\alpha_1} = \frac{sin\, i_2}{\alpha_2} = \frac{\sin i_3}{\alpha_3} = \frac{sin\, i_4}{\alpha_4} = \frac{sin\, i_5}{\alpha_5} = \frac{1}{\alpha_6}$$

If the velocity continues to increase, the angle *i* will eventually equal to 90°, and the ray will turn back:



Recall

$$t = \sum_{j=1}^{m-1} \left(\frac{2z_j}{\alpha_j} \sqrt{1 - \frac{\alpha_j^2}{\alpha_m^2}} \right) + \frac{x}{\alpha_m}$$

For turning ray $p = \frac{1}{\alpha_6}$

$$t = px + \sum_{j=1}^{m-1} \left(2z_j \sqrt{\frac{1}{\alpha_j^2} - p^2} \right)$$

as for the layered case. If the layer thickness is infinitely small, or medium is continuous, we can obtain the travel time equation for turning ray

$$t_{turn} = xp + 2\int_0^z \sqrt{\frac{1}{\alpha^2(z)} - p^2}\ dz$$

The horizontal slowness can also be defined as

$$\frac{dt}{dx} = p$$

Which is the derivative of the time with respect to distance, slope of the arrival time. As we have mentioned that slowness (p) along the ray is constant, and hence the arrival time for a particular ray with slowness p would be constant:

$$\frac{dt}{dp} = 0 = X(p) + 2\frac{d}{dp}\left[\int_0^z \sqrt{\frac{1}{\alpha^2(z)} - p^2}\ dz\right]$$

This gives the distance of arrival for a particular p

$$X(p) = 2p\int_0^z \frac{1}{\sqrt{\frac{1}{\alpha^2(z)} - p^2}} dz$$

And corresponding time for turning ray at p will be

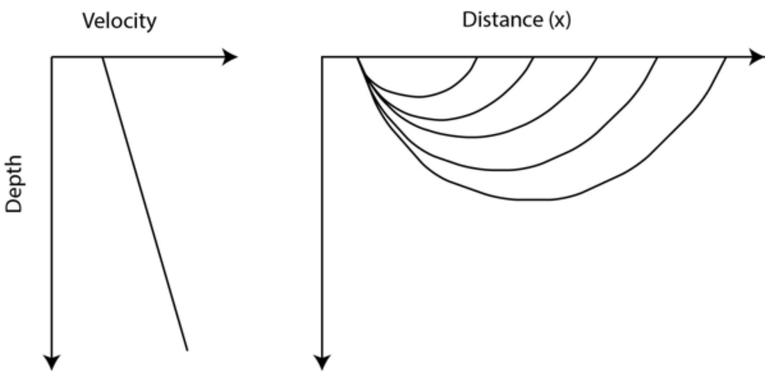
$$T(p) = 2\int_0^z \frac{1}{\alpha^2(z)} \frac{1}{\sqrt{\frac{1}{\alpha^2(z)} - p^2}} dz$$

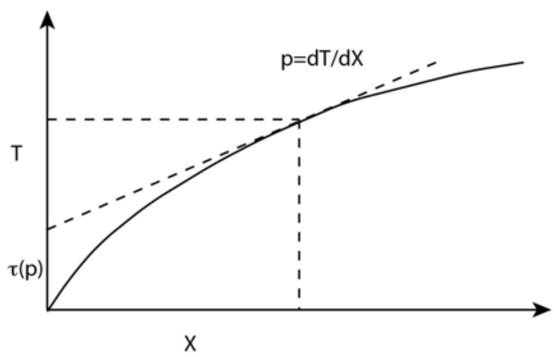
If we define the intercept time as

$$\tau(p) = \int_0^z \sqrt{\frac{1}{\alpha^2(z)} - p^2}\ dz$$

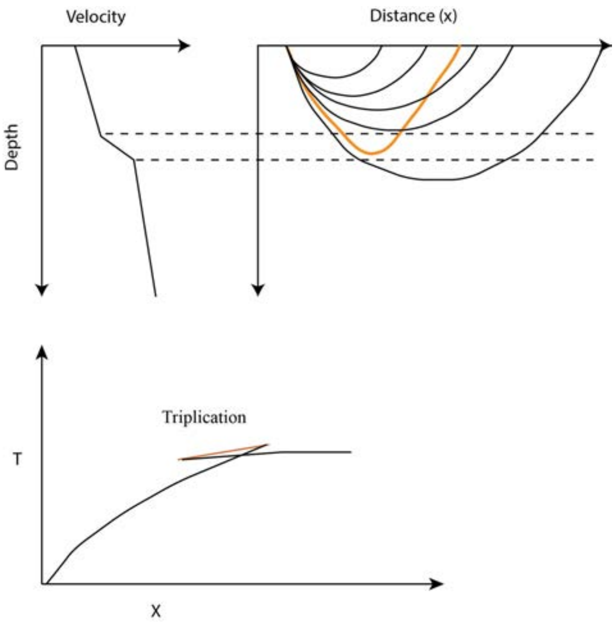
Then the time can be written as

$$T(p) = pX(p) + \tau(p)$$

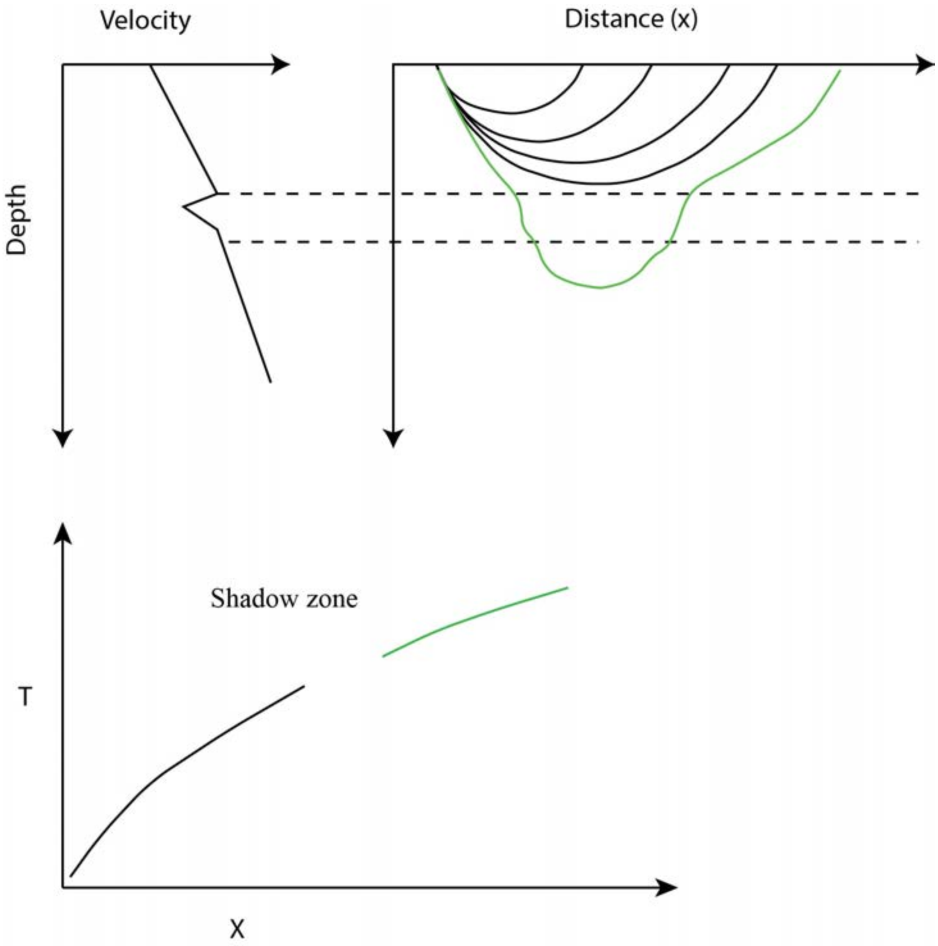


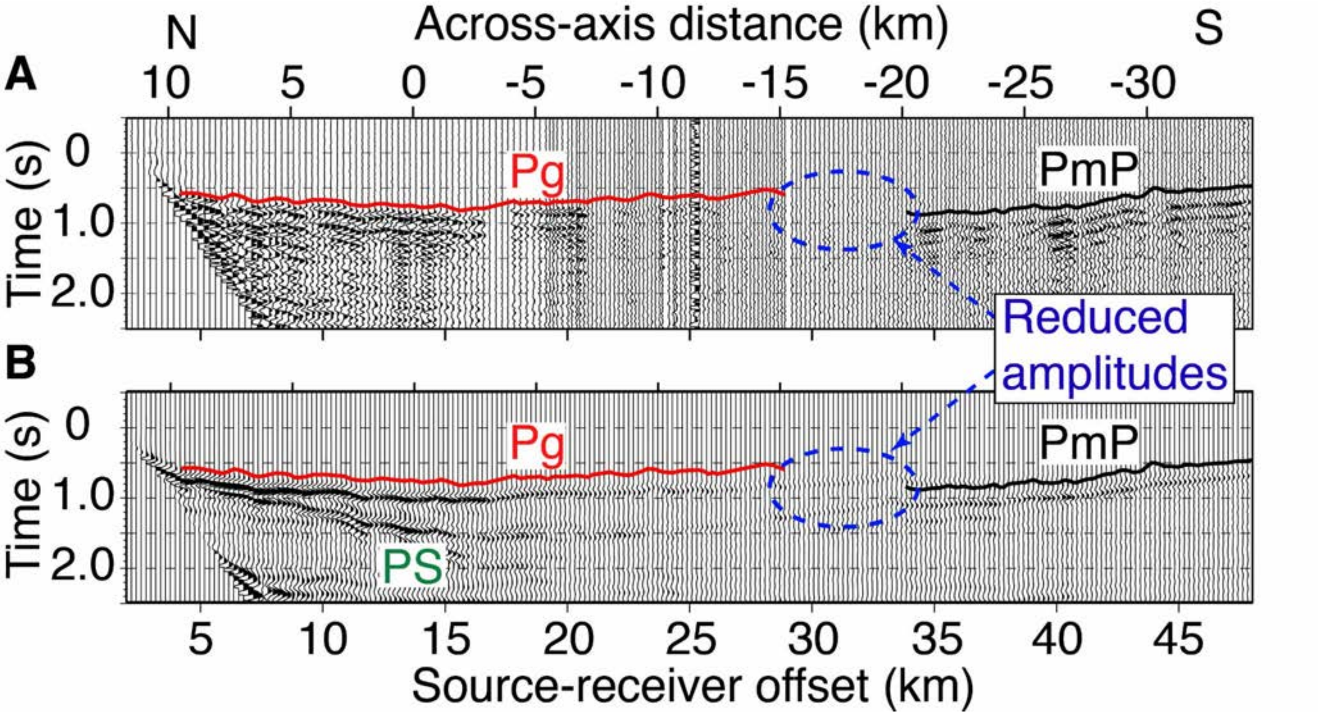


Triplication:



Shadow Zone:

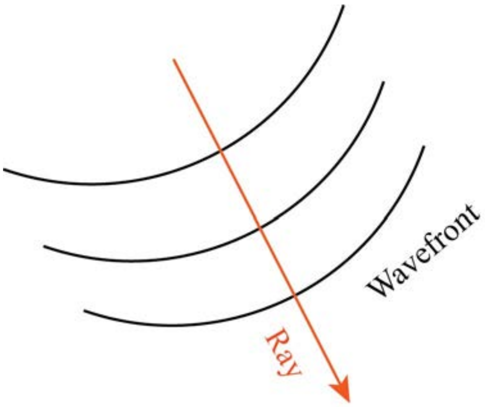




N
Across-axis distance (km)
S
A
10
5
0
-5
-10
-15
-20
-25
-30
Time (s)
0
1.0
2.0
Pg
PmP
Reduced amplitudes
B
Time (s)
0
1.0
2.0
Pg
PmP
PS
5
10
15
20
25
30
35
40
45
Source-receiver offset (km)

Ray Theory

Until now, we have studied the velocity variation with depth only, but the Earth is three dimensional, and therefore, we need to solve the problem in 3D. Unfortunately, there are no simple equations for 3D. The simplest approximation of the wave theory is ray theory. In the ray theory, we assume that the wave propagation can be treated as set of rays of infinite thickness that are orthogonal to the wave fronts.



Here we will first establish the link between wave theory and the ray theory. Therefore, we will start with a scalar wave equation

$$\nabla^2 \phi(\mathrm{x},t) - \frac{1}{v^2}\frac{\partial^2 \phi(\mathrm{x},t)}{\partial t^2} = 0$$

Fourier Transform

$$f(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)e^{-i\omega t}\,dt$$

$$\frac{d}{dt} = i\omega$$
$$\frac{d^2}{dt^2} = (i\omega)^2$$

The wave equation becomes

$$\nabla^2 \hat{\phi}(\mathrm{x},\omega) + \frac{1}{v^2}\omega^2 \hat{\phi}(\mathrm{x},\omega) = 0$$

Asymptotic ray theory (geometrical ray theory)

$$\hat{\phi}(\mathrm{x},\omega) = A(\mathrm{x},\omega)e^{i\omega \mathrm{T}(\mathrm{x})}$$

This equation suggests the wavefield could be decomposed of two parts, the amplitude term $A(\mathrm{x},\omega)$, which defines the amplitude of the wavefield and a phase term $\mathrm{T}(\mathrm{x})$, determine the arrival time. The amplitude term can be written in Ansatz form

$$A(\mathrm{x},\omega) = \sum_{k=0}^{\infty}\frac{A_k(\mathrm{x})}{(-i\omega)^k}$$

which suggests that the amplitude term $A_k(\mathrm{x})$ is independent of frequency and the contribution of the later terms in the Ansatz deceases rapidly for higher frequency, hence this is a high-frequency approximation. This means that this approximation is valid for high frequencies. The first term will be $A(\mathrm{x},\omega) \approx A_0(\mathrm{x})$ or

$$\hat{\phi}(\mathrm{x},\omega) = A_0(\mathrm{x})e^{i\omega \mathrm{T}(\mathrm{x})}$$

which is the zeroth order solution of the wave equation. Substituting this equation in the above wave equation, we get

$$A_0(\mathrm{x})(\nabla \mathrm{T}(\mathrm{x}))^2(i\omega)^2 e^{i\omega \mathrm{T}(\mathrm{x})} + A_0(\mathrm{x})\nabla^2 \mathrm{T}(\mathrm{x})\ (i\omega)\ e^{i\omega \mathrm{T}(\mathrm{x})} + \nabla A_0(\mathrm{x})\nabla \mathrm{T}(\mathrm{x})(i\omega)e^{i\omega \mathrm{T}(\mathrm{x})}$$
$$+ \nabla A_0(\mathrm{x})\nabla \mathrm{T}(\mathrm{x})(i\omega)e^{i\omega \mathrm{T}(\mathrm{x})} + \nabla^2 A_0(\mathrm{x})e^{i\omega \mathrm{T}(\mathrm{x})} - \frac{(i\omega)^2}{v^2(\mathrm{x})} A_0(\mathrm{x})e^{i\omega \mathrm{T}(\mathrm{x})} \cong 0$$

Re-arranging this equation

$$(i\omega)^2 \left[(\nabla \mathrm{T}(\mathrm{x}))^2 - \frac{1}{v^2(\mathrm{x})}\right] A_0(\mathrm{x})e^{i\omega \mathrm{T}(\mathrm{x})} + (i\omega)[A_0(\mathrm{x})\nabla^2 \mathrm{T}(\mathrm{x}) + 2\nabla A_0(\mathrm{x})\nabla \mathrm{T}(\mathrm{x})]e^{i\omega \mathrm{T}(\mathrm{x})}$$

$$+\nabla^2 A_0(\mathrm{x})e^{i\omega \mathrm{T}(\mathrm{x})} \cong 0$$

For the left hand side of the terms equal to zero, each term in the bracket must be equal to zero. The term multiplied by $(i\omega)^2$ will give:

$$\left[(\nabla \mathrm{T}(\mathrm{x}))^2 - \frac{1}{v^2(\mathrm{x})}\right] = 0$$

This equation is the eikonal equation, relating the time with velocity and is the equation of propagation of wavefront.

In tomography, we mainly use the travel time, and hence here we will focus on the eikonal equation.

$$\left[(\nabla \mathrm{T}(\mathrm{x}))^2 - \frac{1}{v^2(\mathrm{x})}\right] = 0$$
$$(\nabla \mathrm{T}(\mathrm{x}))^2 = \frac{1}{v^2(\mathrm{x})}$$

This equation could be solved using three different methods:

1. Finite difference
2. Shooting method
3. Bending method
4. Graph binning method

Finite difference method:

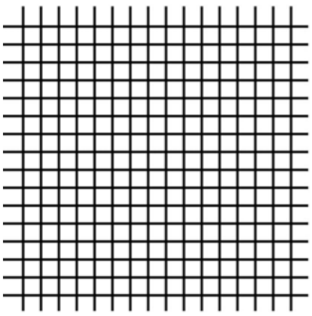
$$(\nabla \mathrm{T}(\mathrm{x}))^2 = \frac{1}{v^2(\mathrm{x})}$$

This equation has two solutions

$$\nabla \mathrm{T}(\mathrm{x}) = \pm \frac{1}{v(\mathrm{x})}$$

In this computation, we only take the positive solution, which means the wave only propagates in the forward direction

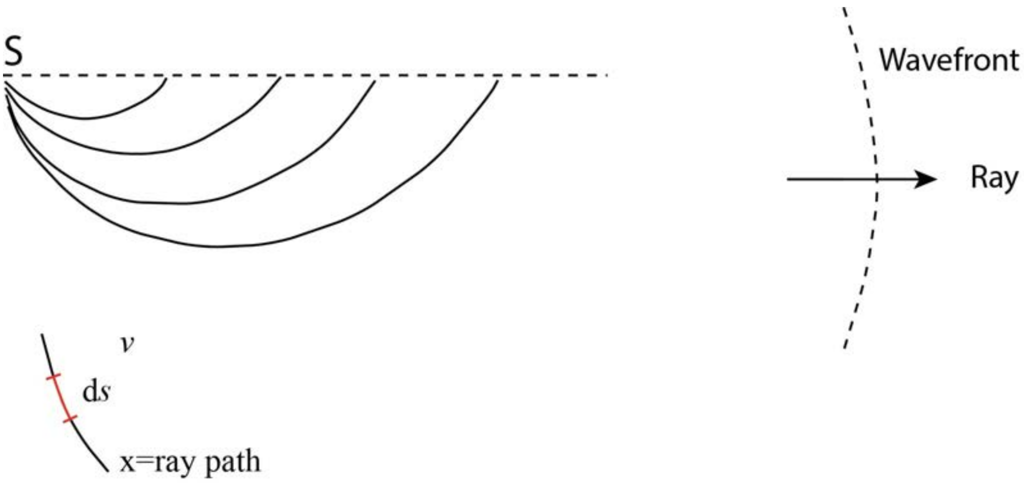
$$\frac{\partial \mathrm{T}(x,y,z)}{\partial x} + \frac{\partial \mathrm{T}(x,y,z)}{\partial y} + \frac{\partial \mathrm{T}(x,y,z)}{\partial z} = \frac{1}{v(\mathrm{x})}$$



Vidale, Geophysics, 1988.

The eikonal solver is very efficient for computing travel time in 3D media, but the main problem is that it can mainly handle the first arrival, i.e., no reflections or triplications. Furthermore, since it is based on forward marching algorithms, it computes arrival times in shadow zones, that do not exist.

Shooting Method:



$$\mathbf{T(x)} = \int_{ray} \frac{\mathbf{1}}{\boldsymbol{v}(\mathbf{x})} \boldsymbol{d}\mathbf{x}$$

We use analytical solution to solve the equation efficiently, and update using interpolation. The integration is carried out along the x=ray path. If d*s* is a small segment along the ray in a medium with velocity *v*, we can write the equation for ray tracing:

$$\frac{d\mathrm{x}(s)}{ds} = v\,\nabla\mathrm{T}$$

Differentiating the eikonal equation with respect to *s*

$$\frac{d(\nabla\mathrm{T})}{ds} = \nabla\frac{1}{v}$$

We know

$$\mathrm{p} = \nabla\mathrm{T}$$

we get

$\frac{d\mathrm{x}(s)}{ds} = v\,\mathrm{p}$ and $\frac{d(\mathrm{p})}{ds} = \nabla\frac{1}{v}$, the ray tracing equation.

In numerical sense

$\mathrm{X}_{j+1} = \mathrm{X}_j + (v\mathrm{p})\Delta s$ along the ray

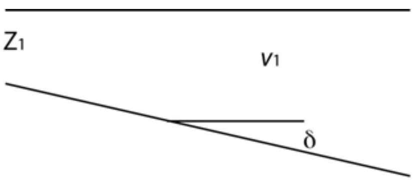
$\mathrm{p}_{j+1} = \left(\frac{1}{v_{j+1}} - \frac{1}{v_j}\right)\Delta s$ perpendicular to the ray.

These are the two common methods to compute the travel time in 2D and 3D earth.

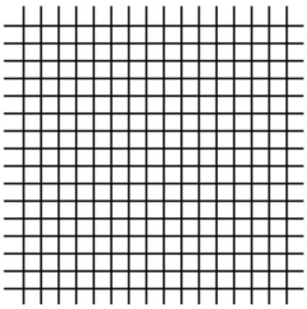
Tomography

Once we can compute the travel time from a source to receiver, we can use the modern tomographic methods to estimate velocity models. While the computation of travel time is termed as a forward problem, the tomography is termed as an inverse problem. Before we talk about the inverse problem, we should first discuss the parameterization.

Parameterization: The ideal case of solving any problem is to have number of unknowns (parameters) equal to the number of equations, then we can obtain all the unknown parameters from the given equations. The earth is not like that; there are cases we have only a few observations and a large number of unknown parameters, and in other cases, there are a large number of observations for only a few unknowns.



Under parameterization



Over parameterization

Under parameterization when $n(m) \ll n(d)$ (e.g. Zelt and Smith, GJI, 1992)
Over parameterization when $n(m) \gg n(n(d)$ (McCaughey and Singh, GJI, 1997; Zelt and Barton, JGR, 1998; Hobbro et al., GJI, 2002).

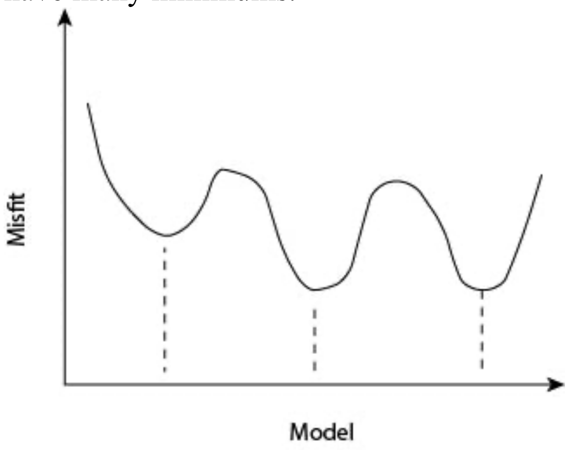
Other thing to consider is how to define model, using velocity (v), slowness $u(=1/v)$, square of the slowness $u^2(=1/v^2)$ depending up on modeling algorithm used.

Inverse Problem:

The objective of any inverse problem is to minimize the difference between the observed data (d_{obs}) and calculated data (d_{cal})

$$S(\mathrm{m}) = \|d_{obs} - d_{cal}\|^m$$

where m is a norm. For least squares case m=2. The misfit function S(m) could be non-linear and may have many minimums.



The non-linear problems could be solved using global search algorithms, such as Monte Carlo, Genetic Algorithm, Simulated Annealing etc. However, for large-scale inverse problems, these techniques are very expensive, and hence we use local search techniques starting from an initial model, mc, which is close to the true model, i.e. lies in the global minimum.

Weakly non-linear problem: Re-writing the travel time equation in terms of slowness

$$t(\mathrm{x}) = \int_{ray} \boldsymbol{u}(\mathrm{x})\, \boldsymbol{d}\mathbf{x}$$

Where $\boldsymbol{u}$(x) is the inverse of velocity or slowness. Let us assume that we have some idea of the slowness, $\mathbf{u}_0$(x), which is close to the true slowness by a small amount $\delta\boldsymbol{u}$(x), the we can write

$$\boldsymbol{u}(\mathrm{x}) = \boldsymbol{u}_0(x) + \delta\boldsymbol{u}(\mathrm{x})$$

Then the travel time in the initial model can be written as

$$\boldsymbol{t_0}(\mathbf{x}) = \int_{ray} \boldsymbol{u}_0(\mathrm{x})\, \boldsymbol{d}\mathbf{x}$$

The travel time in the perturbed medium will be

$$t(\mathrm{x}) = \int_{ray} \boldsymbol{(}\boldsymbol{u}_0(x) + \delta\boldsymbol{u}(\mathrm{x})\boldsymbol{)}\, \boldsymbol{d}\mathbf{x}$$

or

$$t(\mathrm{x}) - \boldsymbol{t_0}(\mathbf{x}) = \int_{ray} \boldsymbol{(}\delta\boldsymbol{u}(\mathrm{x})\boldsymbol{)}\, \boldsymbol{d}\mathbf{x}$$

$$\boldsymbol{\delta t}_0(\mathbf{x}) = \mathbf{A}\, \delta\boldsymbol{u}(\mathrm{x}) = \mathbf{A}\, \delta\mathrm{m}$$

which is the data misfit. We can generalize this equation

$$\delta\boldsymbol{m}(\mathrm{x}) = \boldsymbol{m}(\mathrm{x}) - \boldsymbol{m}_c(x)$$

and the travel time residual will be

$$\boldsymbol{\delta t}(\mathbf{x}) = \boldsymbol{t_{cal}}(\mathbf{x}) - \boldsymbol{t_{obs}}(\mathbf{x})$$

$$A_{ij} = \frac{\partial t_j}{m_i}$$

which the derivative of the data with respect to the model. It is very difficult to minimize the data alone, as it would lead to a very unstable solution, particularly we have a large number of model parameters. In this case, we minimize the misfit function along with the model misfit function

$$\psi(\delta m) = \delta t^T . C_D^{-1} . \delta t + \lambda(m - m_0) . C_M^{-1}(m - m_{c)}$$

where the second term in the equation is the model misfit. λ is a constant, used to give different weight for model misfit function. We can write $m - m_0 \rightarrow m_0 + \delta m$ and expand the misfit function near the current mode m_0

$$\psi(m) = \psi(0) + l(0) . \delta m + \frac{1}{2} \delta m^T . H . \delta m$$

The first term is the data misfit

$$\psi(0) = \delta t_0^T . C_D^{-1} . \delta t_D$$

The second term is gradient (first derivative with respect to model parameter) of the misfit function, defined as

$$l(\delta m) = -2(\delta t_0 - A\delta m)^T . C_D^{-1} A + 2\lambda(m_0 + \delta m) . C_M^{-1}$$

The second derivate of the misfit function, called Hessian, can be written as

$$H = 2(A^T C_D^{-1} A + \lambda C_M^{-1})$$

The minimum occurs when $\psi(m) = \psi(0)$, or the last terms equal to zero

$$l(0) . \delta m + \frac{1}{2} \delta m^T . H . \delta m = 0$$

hence the solution of the inverse problem lies at

$$\delta\widehat{m} = -l(0) . H^{-1}$$

Substituting the value of l and H, we find the solution of the inverse problem

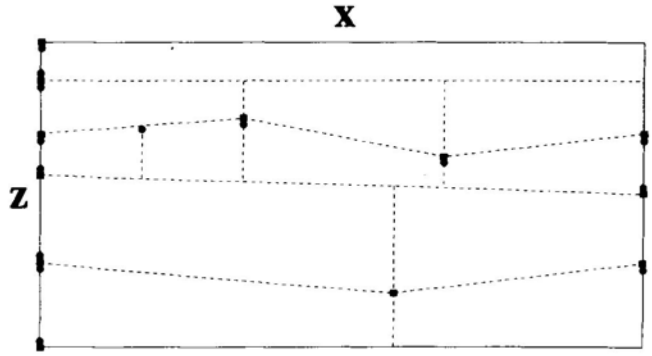
$$\delta\widehat{m} = (A^T C_D^{-1} A + \lambda C_M^{-1})^{-1}(\delta t_0^T . C_D^{-1} A - \lambda m_0^T C_M^{-1})$$

The model parameter could be estimated using single step solution or an iterative method. There are different iterative methods, such as Newton method, steepest decent method, conjugate gradient method. In each method, we would require local gradient, which is also called Frechet derivative, at model m_j and Hessian H, the inverse of which is difficult to compute.

The above equation is a general equation for any inverse problem, e.g., seismic reflection, refraction, waveform inversion, and will appear time and again, in different forms.

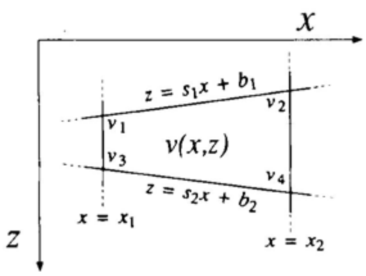
Under Parameterization Method (Zelt and Smith, GJI, 1992)

Model is parameterized by a limited number of nodes, much less than the data.
It is layered cake model. Each layer is divided into trapezoids.



A trapezoid that has four boundaries in the z-x plan is defined by four velocities (v_1, v_2, v_3, v_4) at its corner where the distance between the nodes is

$$x = x_1,\;\; x = x_2,\;\; z = s_1 x + b_1, \quad z = s_2 x + b_2$$



which means only depth of the nodes vary, not the x-location. The constants, s and b, are predefined.

The velocity in the trapezoid is

$$v(x,z) = \frac{(c_1 x + c_2 x^2 + c_3 z + c_4 xz + c_5)}{(c_6 x + c_7)}$$

where the coefficient c_i are linear combinations of corner velocities

$$c_1 = s_2(x_2 v_1 - x_1 v_2) + b_2(v_2 - v_1) - s_1(x_2 v_3 - x_1 v_4) - b_1(v_4 - v_3)$$
$$c_2 = s_2(v_2 - v_1) - s_1(v_4 - v_3)$$
$$c_3 = x_1 v_2 - x_2 v_1 + x_2 v_3 - x_1 v_4$$
$$c_4 = v_1 - v_2 + v_4 - v_3$$
$$c_5 = b_2(x_2 v_1 - x_1 v_2) - b_1(x_2 v_3 - x_1 v_4)$$
$$c_6 = (s_2 - s_1)(x_2 - x_1)$$
$$c_7 = (b_2 - b_1)(x_2 - x_1)$$

You could have layer boundary with different velocities or same layer by having velocity at the interface same. The main advantage of this parameter is that ray can be traced efficiently.

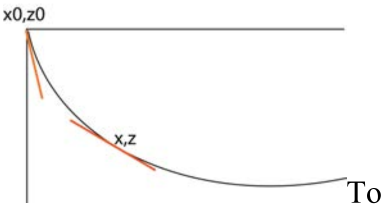
Ray Tracing

$$\frac{dz}{dx} = \mathrm{cotan}(\theta), \quad \frac{d\theta}{dx} = \frac{(v_z - v_x \mathrm{cotan}(\theta))}{v}$$

for ray path near horizontal

$$\frac{dx}{dz} = \tan(\theta), \quad \frac{d\theta}{dx} = \frac{(v_z \tan(\theta) - v_x)}{v}$$

for ray path near vertical,
with initial conditions, $x = x_0,\ \ z = z_0,\ \ \theta = \theta_0$. v_x and v_z are partial derivatives of velocity with respect x and z, θ_0 is ray take off angle from the vertical.

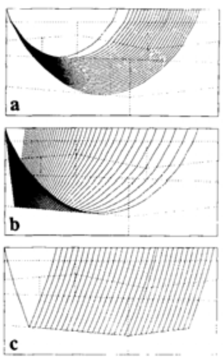


Step length for the ray can be calculated as

$$\delta l = \frac{\alpha v}{|v_x| + |v_z|}$$

where α is a constant.

Ray take-off angle need to be assigned for different types of rays:



The rays nearest to the receivers are used to interpolate the time at the receiver.

Inversion

Rewriting the travel time as

$$t = \int_L \frac{1}{v(x,z)} dl \ \text{ or } \ t = \sum_{i=1}^{n} \frac{l_i}{v_i}$$

After linearizing $\mathbf{m} = \mathrm{m}_0 + \Delta\mathbf{m}$ we can write

$\mathbf{A}\Delta\mathbf{m} = \Delta\mathbf{t}$

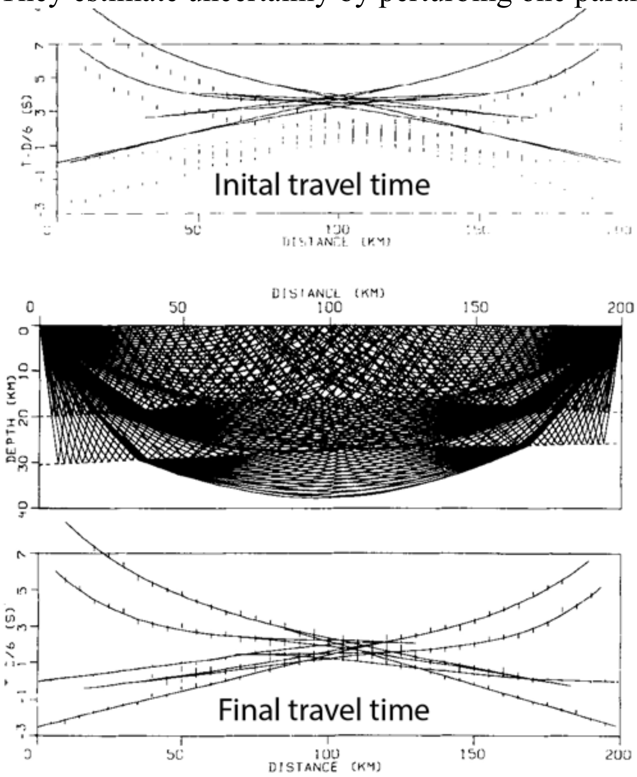
A is the partial derivative matrix, and can be calculated using

$$\frac{\partial t}{\partial v_j} = \int_L -\frac{1}{v^2}\frac{\partial v}{\partial v_j} dl$$

analytically using the definition of the velocity.

They use a damped-least squares inversion, by adding an extra damping term that depends on the initial model, similar to the regulation term discussed above.

They estimate uncertainly by perturbing one parameter at a time and performing the inversion.



The advantage of this approach is that it is efficient, but the user has to define the nodes, give the initial velocity and gradient, and inversion only perturbs a limited number of parameters. So the method is very user based, and can be biased, i.e. no objectivity.

In order to obtain unbiased inversion results, we would need to use a large number of parameters and regularized inversion.

Regularised Tomographic Inversion for velocity and interface
(McGaughey and Singh, GJI, 1997; Hobro et al., 2002)

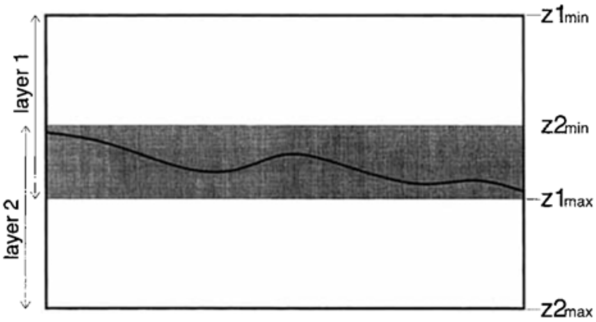
Parameterization: Fine grid parameterization both for velocity and interface

Slowness squared and depth parameters are interpolated using a B-spline

$$u^2(x,z) = \sum_{i=1}^{r} \beta_i^V\,(x,z) m_i$$

$$D^l(x,z) = \sum_{i=r+1}^{s} \beta_i^D\,(x,z) m_i$$

where β_i^V and β_i^D are appropriate B-spines, and m_i are the slowness squared and interface depth parameters, respectively.
To allow the interfaces to move during the inversion, the velocities in the layers are defined so that the layers overlap (Figure). The velocity across the interface could have a contrast.



Recall the linearization

$$t(\mathrm{x}) = \int_{ray} (\delta \boldsymbol{u}(\mathrm{x}))\, d\mathrm{x}$$

Recall the linearization $u(x) = u_0(x) + \delta u(x)$

$$\boldsymbol{\delta t} = \int_{ray} (\delta \boldsymbol{u}(\mathrm{x}))\, \boldsymbol{dx}$$

rewriting
$u d\tau = dx$, we can get

$$\boldsymbol{\delta t} = \frac{\mathbf{1}}{\mathbf{2}} \int_{ray} (\delta \boldsymbol{u}^2(\tau))\, \boldsymbol{d\tau}$$

The time due to small perturbation in the interface δD^l will be

$$\delta t = \Delta p_z^k \delta D^l(x^k)$$

where $\Delta p_z^k = (p_z - \widehat{p_z})$, and p_z, $\widehat{p_z}$ are the vertical component of slowness vectors for the incident and transmitted/reflected rays at the kth interface. So the total perturbed travel time can be defined as

$$\boldsymbol{\delta t} = \frac{\mathbf{1}}{\mathbf{2}} \int_{ray} (\delta \boldsymbol{u}^2(\tau))\, \boldsymbol{d\tau} + \sum_k \Delta p_z^k \delta D^l(x^k)$$

due to velocity and interface perturbations. $\mathbf{A\delta m} = \boldsymbol{\delta}\mathbf{t}$, where A_{ij}=$\partial t_i/\partial m_j$ is the partial derivative. The travel time and the partial derviatives are computed analytically using the shooting method discussed above. A fan of rays are traced as discussed above at neighbouring points and then are interpolated to obtain the time as well as the derivatives at the receiver.

Velocity and depth parameter equalization:

The slowness squared and interface parameters have different units, and the partial derivatives for each parameters will be significantly different. Furthermore, a ray will cross many cells, whereas it would interact with interface only a few times. Thus the total of the partial derivatives with respect to the slowness squared will be much greater than those due to interface depths.

To address the problem of this difference in units and sensitivity, the units of the parameter sets are changed so that the equal emphasis is placed on each set by the inversion. A correction factor, *w,* is calculated and all the partial derivatives with respect to depth parameters are weighted by it. W is calculated by summing the absolute values of all the elements of A which concern the slowness squared parameters, and divided by the sum of the absolute values of all the elements which concern the depth parameters,

$$\boldsymbol{w} = \sqrt{\frac{\sum_i \sum_{j=1}^{r} |\boldsymbol{A}_{ij}|}{\sum_i \sum_{j=r+1}^{s} |\boldsymbol{A}_{ij}|}}$$

Regularization

As mentioned above, instead of minimizing the travel time misfit function, we include a regularization term that depends on the model

$$\psi(\delta m) = \delta t^T . C_D^{-1} . \delta t + \lambda (m_0 + \delta m) . C_M^{-1} (m_0 + \delta m)$$

When λ is large, the inversion will be influenced the initial model and when λ=0, the inversion will mainly depend on the data misfit.

Damped least-squares

The second term can be replaced by

$$\psi_{DLS} = \lambda \sum_i \delta m_i^2$$

At each iteration, it will produce solution, which is close to the previous solution in a least squares sense. This is simple to implement and is useful for course parameterization, but it is very sensitive to the starting model. Secondly, the inversion can become unstable as this tends to produce rough model. One can mitigate the problem by smoothing the gradient at each iteration, but this would be ad hoc and not very satisfactory.

Regularized least-squires

An alternative to the second term of the objective function is

$$\psi_{RLS} = \lambda \| m + \delta m \|_M^2$$

where the norm $\|.\|_M$ a differential property of the model (flatness or smoothness), and where this property is measured for the perturbed (new) model and is therefore independent of the starting model. This approach allows the model smoothness to be measured and regulated during the inversion (using λ) in a controlled and objective manner.

There are two approaches: flatness and smoothness. To measure the extent of non-flatness, a measurement of first spatial derivatives across the model is used, giving

$$\|m\|_F^2 = (x_2 - x_1) \int_{x_1}^{x_2} \left(\frac{\partial m}{\partial x} \right)^2 dx$$

For smoothness criteria, the second derivatives across the model is used,

$$\|m\|_S^2 = (x_2 - x_1)^3 \int_{x_1}^{x_2} \left(\frac{\partial^2 m}{\partial x^2} \right)^2 dx$$

Von Avendonk et al (1998) uses

$$\|m\|_F^2 = \int_A \left\{ L_H^2 \left(\frac{\partial u}{\partial x} \right)^2 + L_V^2 \left(\frac{\partial u}{\partial z} \right)^2 \right\} dA$$

where L_H and L_V are horizontal and vertical smoothing length.

McCaughey and Singh (1997) and Hobro et al. (2003) use the smoothness criteria

$$\|m\|_{su}^2 \approx \int_A \left\{ \left(\frac{\partial^2 u}{\partial x^2} \right)^2 + \left(\frac{\partial^2 u}{\partial z^2} \right)^2 + 2 \left(\frac{\partial^2 u}{\partial x \partial z} \right)^2 \right\} dA$$

Similarly, one can have regularisation for the interface

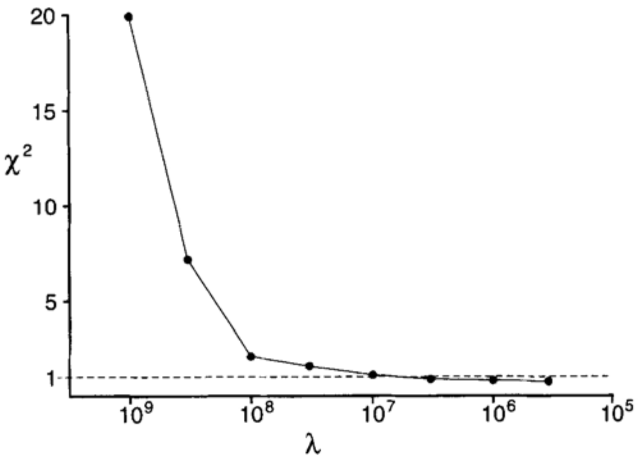
$$\|m\|_{sI}^2 \approx \int_A \left\{ \left(\frac{\partial^2 z}{\partial x^2} \right)^2 \right\} dA$$

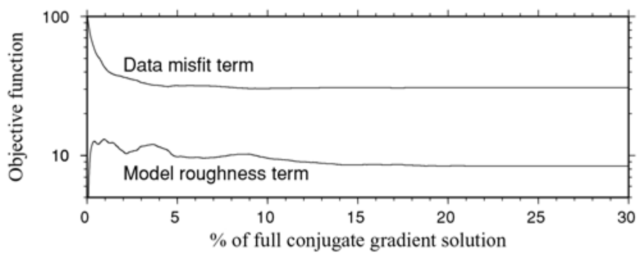
Variable smoothing approach

The strength of the regularization is controlled by the variable parameter λ. A high value will produce a smooth solution but higher travel time misfit whereas a low value will produce rough model, which if the inversion remained stable, will give a lower residual. The fit of the calculated traveltimes to the data traveltimes relative to the noise levels can be quantified by

$$\chi^2 = \frac{1}{n}\, \delta t^T . C_D^{-1} . \delta t$$

where n is number of traveltimes used in the inversion. The optimum value of λ is that which results in $\chi^2 = 1$, producing the smoothest model, which fits the data to the level of the noise. The variable smoothing consists of starting with a high value of λ where the model is smooth and decrease slowly until the $\chi^2 = 1$.





Data Error

We must have some idea about the data error. You could have picking error, depending upon the frequency of the signal and noise in the data. Near offset, far offset, first arrival, secondary arrivals etc.

Optimization

As mentioned above, we start from the starting model m_0, which is close enough to the global model but far enough to provide unbiased solution. The model is updated $m \rightarrow m + \delta m$. Initially δm is set to zero. The updated model is obtained using different optimization methods, such as Gauss-Newton method, steepest decent method or conjugate gradient method. The basic rule is to update the model using a step length (α) along a gradient direction (d_i)

$$\delta m_i = \delta m_{i-1} + \alpha_i d_i$$

The steepest descent direction will be the gradient direction

$$g_i = \begin{bmatrix} -l(\delta m_{i-1}) \quad for\ i > 1, \\ -l(0) \quad for\ i = 1 \end{bmatrix}$$

The conjugate gradient direction vector will be

$$d_i = \begin{bmatrix} g_i + \left(\frac{(g_i - g_{i-1})}{g_i \cdot g_i}\right) d_{i-1} \quad for\ i > 1, \\ g_1 \quad for\ i = 1 \end{bmatrix}$$

The step length along the conjugate gradient direction will

$$\alpha_i = \frac{g_i \cdot d_i}{d_i^T H d_i}$$

As you will notice that conjugate direction depends not only the current gradient but also on the previous gradient, and the convergence is faster, and the solution is more stable.

Resolution and Error Analysis

As we saw earlier with Zelt and Smith method, we need to be sure what we invert makes sense and we have some idea about the uncertainty in our results. If the misfit function has a Gaussian probability distribution, the posteriori co-variance matrix can be defined as an inverse of the Hessian matrix

$$C_M = H^{-1} = \frac{1}{2}(-A^T C_D^{-1} A + \lambda C_M^{-1})^{-1}$$

The diagonal term of this matrix will give the variance

$$\sigma_i = \sqrt{C_M^{ii}}$$

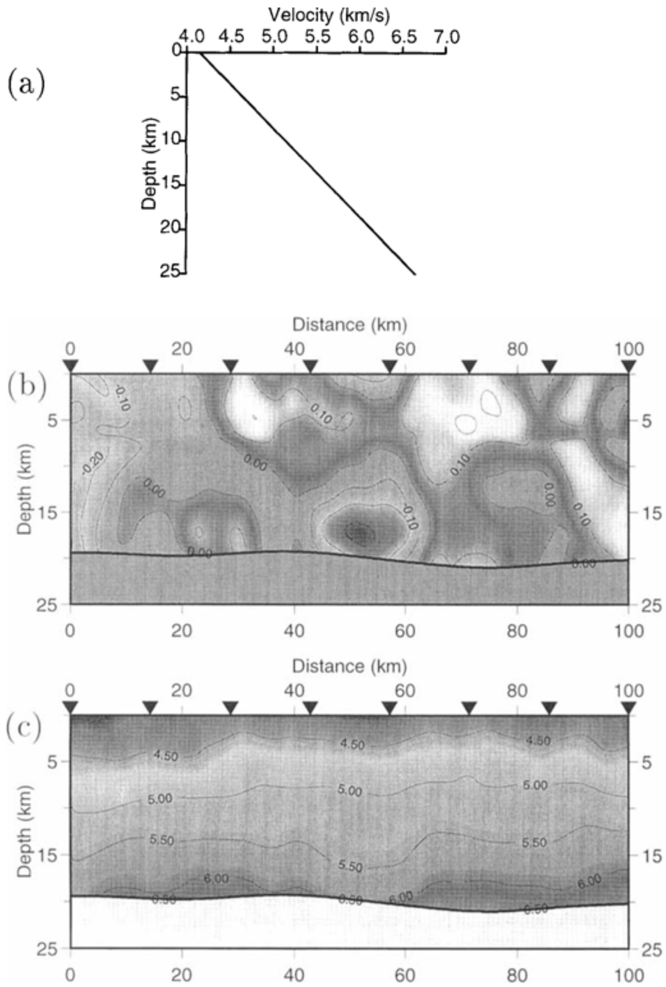
Off-diagonal term will give correlation

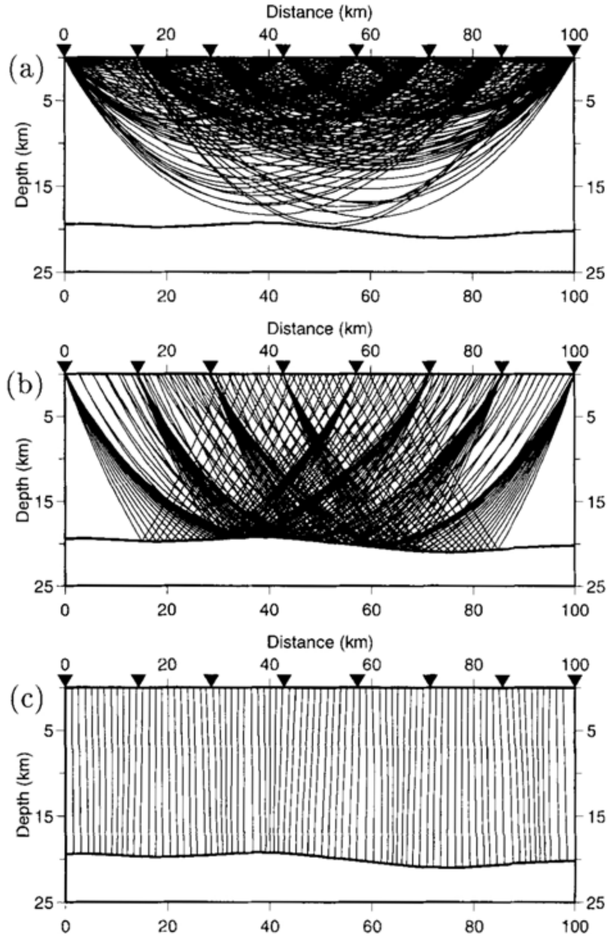
$$\varrho_{ij} = \frac{C_M^{ij}}{\sqrt{C_M^{ii} C_M^{jj}}}$$

$$-1 \leq \varrho_{ij} \leq 1$$

Zero correlation means parameters are not correlated and ± 1 means they are highly correlated.

The inverse of the Hessian Matrix can be very expensive, but it is sparse matrix and hence can inverted using standard sparse matrix inversion, such as Gauss-Jordan method.





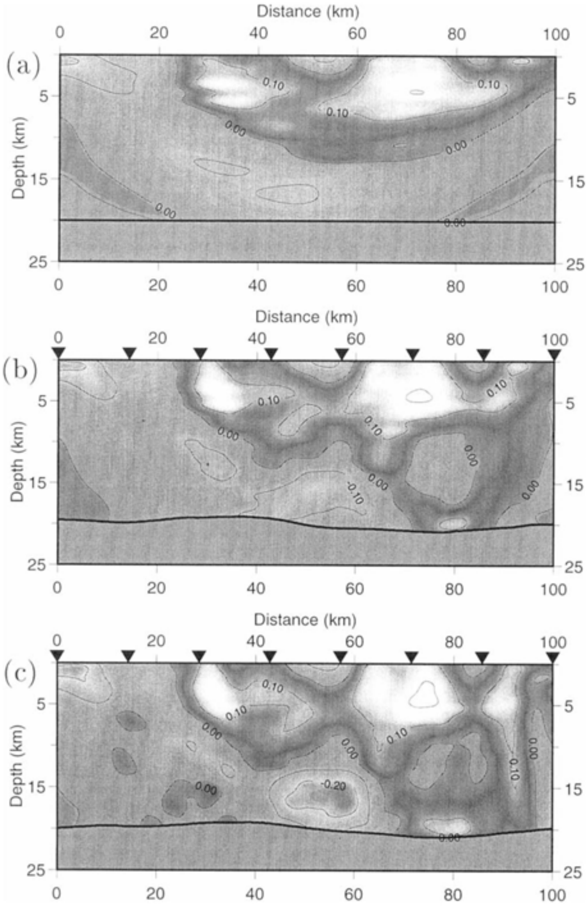
Defining the rays

Normal incidence data have very good lateral resolution, not vertical resolution.

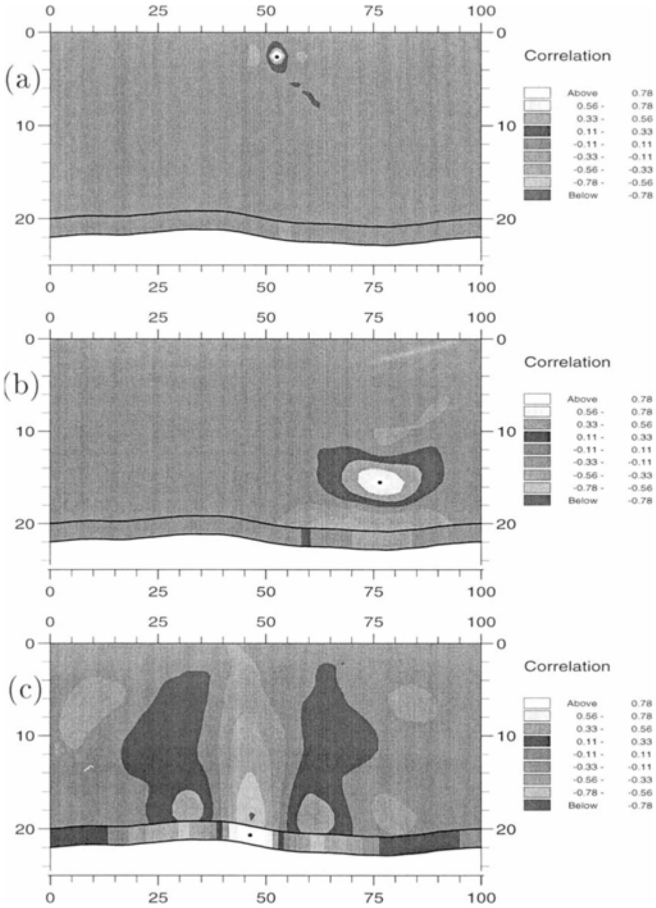
Turning rays have very good vertical resolution but very poor lateral resolution

Wide-angle reflection has resolution both in vertical and horizontal directions.

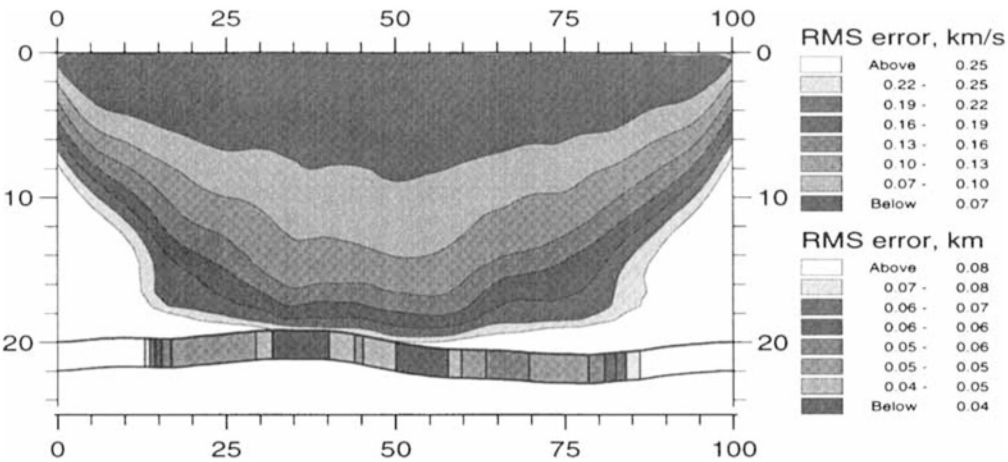
Pre-critical reflections



Final model



Correlation



Uncertainty

Uncertainty using random models:

Korenaga et al (2000, JGR) proposes to start from 50-100 random initial models and then perform linearized tomographic inversion. Assuming all the models converge to a solution, you will have a set of inverted models that will fit the data equally well and others may converge but may have larger traveltime residual. The average model can be defined as

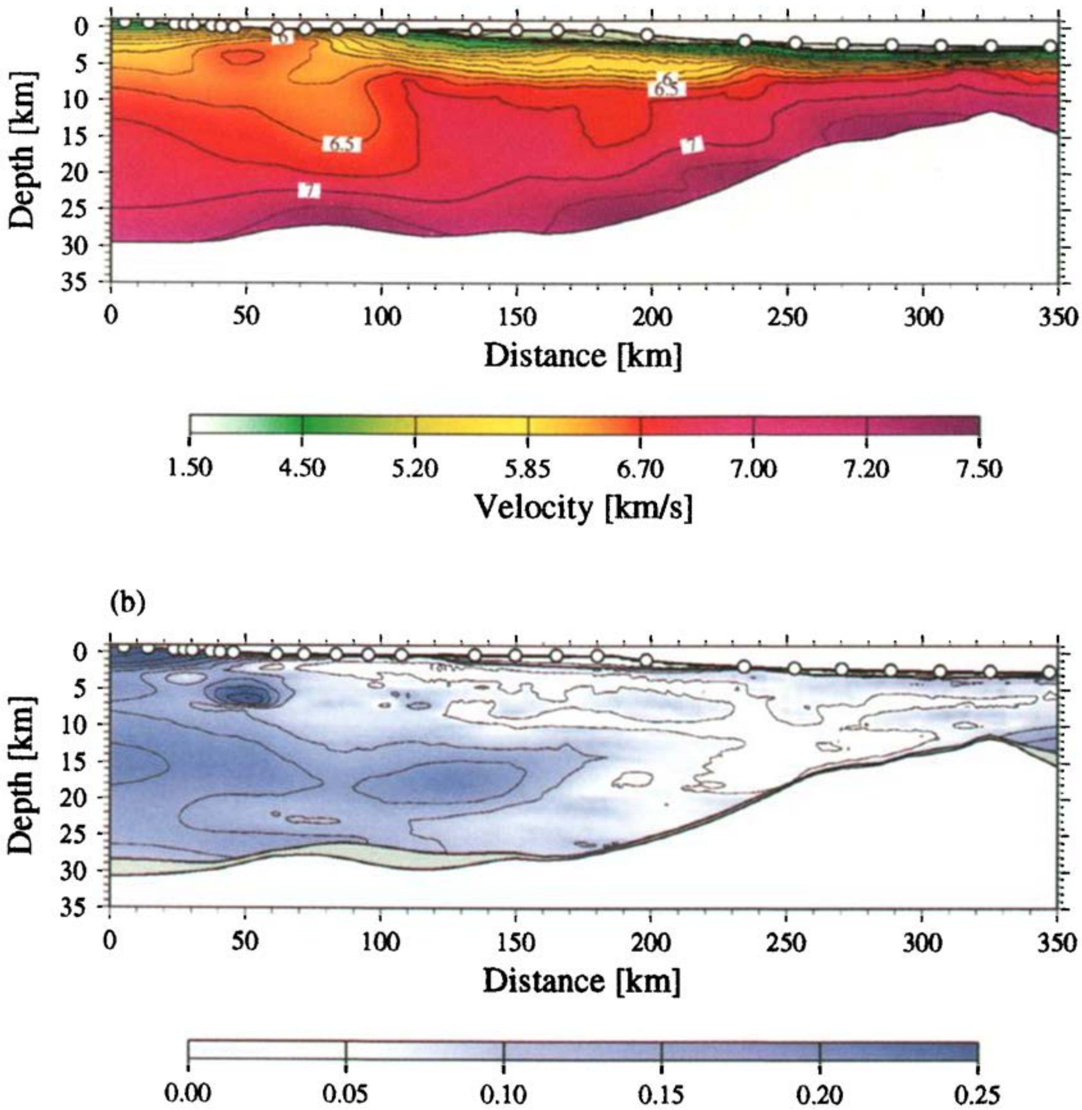
$$m_{av} = \frac{1}{n}\sum m_i \pm \sigma$$

where n is number of different starting models, m_i is the different final models and σ is the uncertainty obtained from the posteriori co-variance matrix defined as

$$C = \frac{1}{n}\sum_{i=1}^{n}(m_i - m_{av}).(m_i - m_{av})^T$$

and $\sigma_i = \sqrt{C_{ii}}$

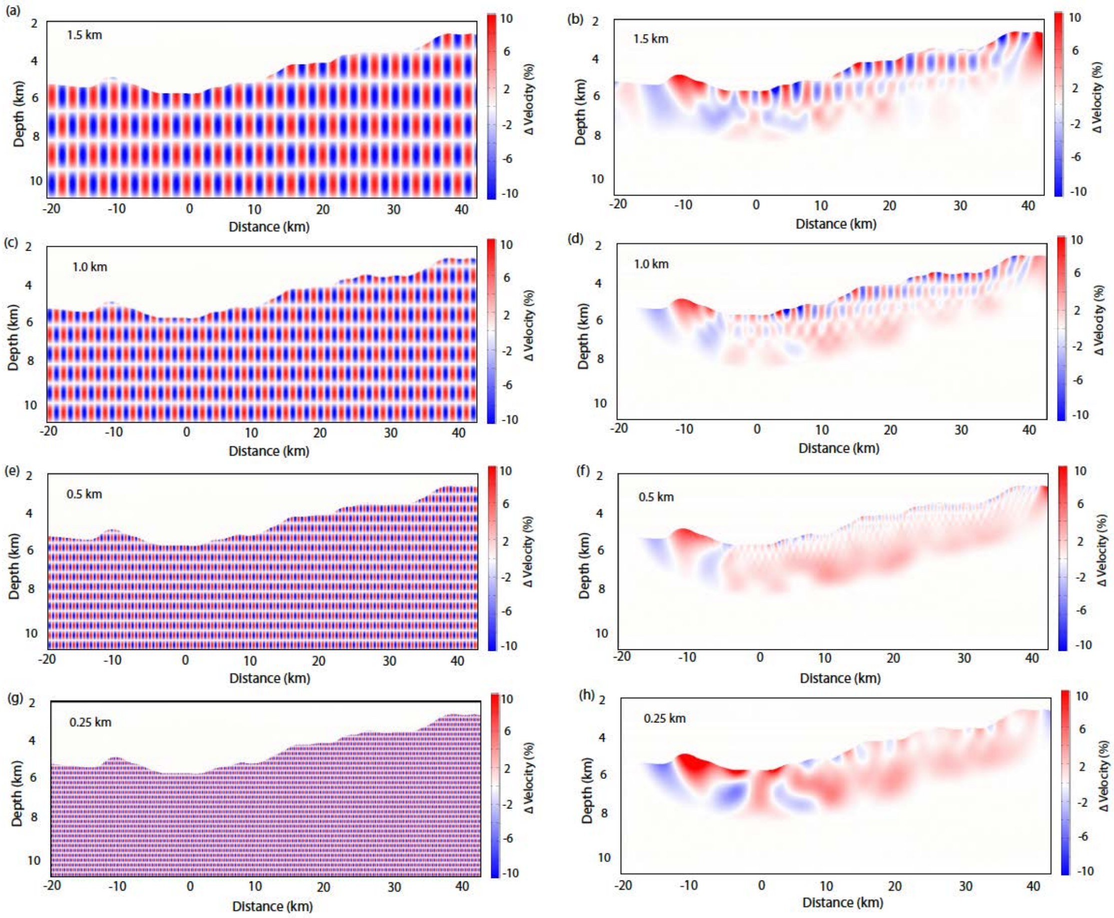
This approach assumes that the final solution of all the inversion results have a Gaussian distribution.



Checkerboard Test:

Other way to assess the resolution is to perform checkerboard test:

Take the final inverted velocity model, add checkerboard with positive and negative velocity anomaly having small perturbation (2-3%) with the same size, compute travel time and perform the inversion. Then change the size of the anomalies and perform the inversion. This will allow you to get some idea of size of the anomaly that you can resolved as a function of depth.



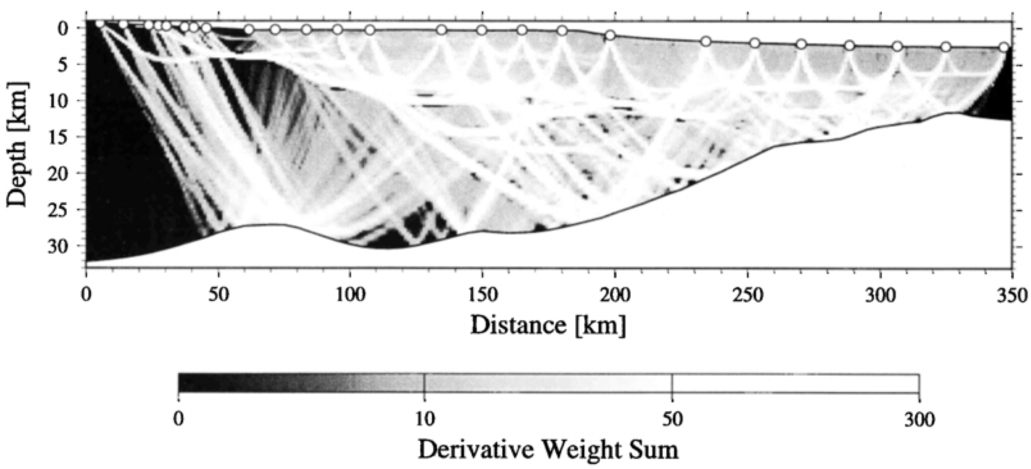
Ray Diagram

You must plot a ray diagram to show in which part of the model rays have travelled.

Hit Count: Number of times a ray travels through a cell. It does not contain any information on the direction or length of the ray etc. It allows to get a digital number telling something about rays and cells.

Derivative Weight Sum (DWS)

It is total weighted ray length in a cell (weighted sum of Frechet derivatives)
The length of the ray is included, not the direction. It measures the ray density.



Wave Equation Tomography

So far we have picked the arrival times, leading to human error in picking. We do not take into account the type of wavelet, frequency of the data etc. A new method has been developed recently called wave equation tomography. It is based cross-correlation of real data with synthetically calculated data (Wang et al, 2014, GJI). Synthetic data could be computed using either ray theory or by solving the wave equation.

Cross-correlation

The cross-relation measures the similarly between two functions and can be defined as

$$\emptyset_{fg}(\tau) = \int_{-\infty}^{\infty} f(t)g(\tau+t)dt \quad \text{or } \emptyset_{fg}(\tau = n\Delta t) = \sum_i f_i\, g_{n+i}$$

Shift the second time series by τ, multiply the two time series, and then add for all possible shifts. The cross-correlation is non-cumulative, i.e.

$$\emptyset_{fg}(\tau) = \emptyset_{fg}(-\tau)$$

The cross-correlation can also be used to estimate the time difference between two signals. The maximum cross-relation will occur when synthetic data and observe data match perfectly:

$$CC_{max}(\Delta\tau_{sr}) = \int W(t)u_{sr}^{obs}(t)u_{sr}^{syn}(t+\Delta\tau_{sr})\, dt$$

$\Delta\tau$ is the time delay between the synthetic data and observe data, and W(t) is window in which the data are cross-related. The synthetic data, not just the time, but full wave field is computed solving a full wave equation using a numerical method (e.g. finite difference).

In this case, you could define your misfit function

$$S(m) = \|1 - \mathrm{CC}_{max}(\Delta\tau)\|^2$$

which is equivalent to

$$S(m) = \Delta\tau^T . C_D^{-1} . \Delta\tau$$

minimizing the time difference between the observed and synthetic data.
Interestingly, the gradient of the misfit will contain the term

$$S_\tau = \sum_r \frac{1}{N_r}\dot{u}(x_r)\Delta\tau(x_r)\delta(x_r)$$

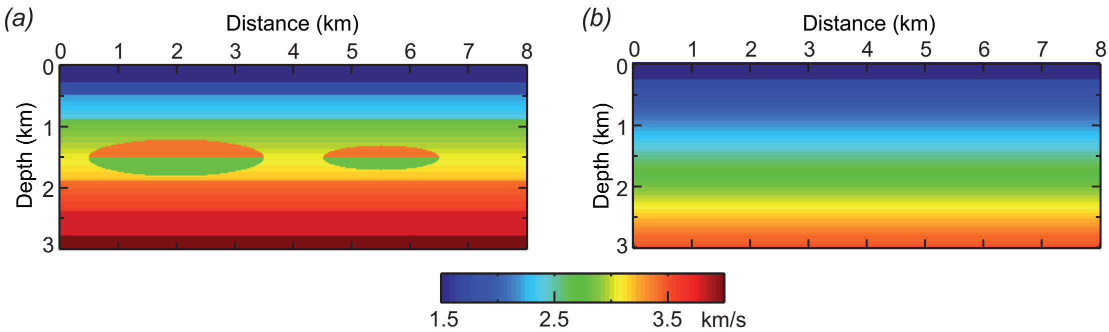
where the normalization factor N_r is

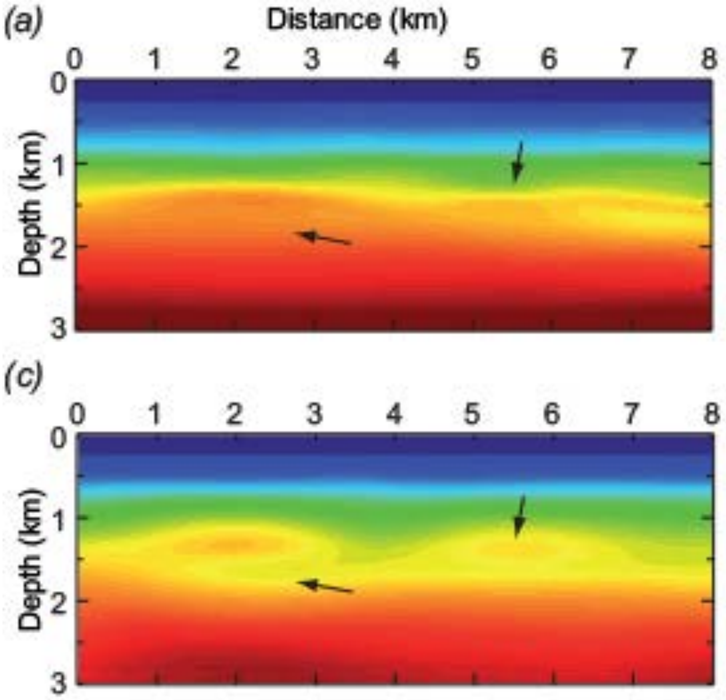
$$N_r = \int_0^W u(x_r)\,\ddot{u}(x_r)dt$$

which means the waveform will also contribute to the inversion. The dots indicate the first and second order time derivative.
The gradient will be of the form

$$g = -2\rho v \sum_s \int_0^W \nabla u * S_\tau\, dt$$

Which again depends on the full wave field.



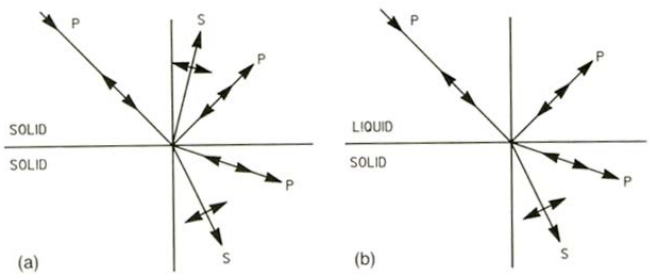


Travel time tomography of reflection data:

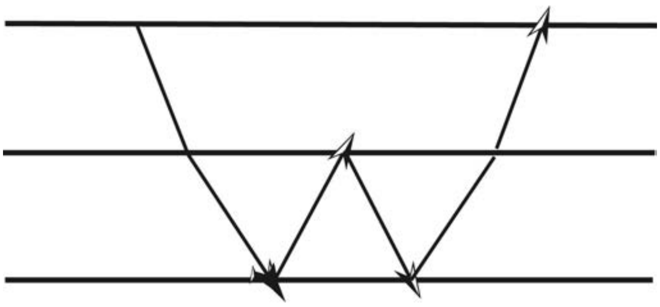
Although we have focused on refraction or turning rays, all the basics presented above is valid for just inverting the seismic reflection data only. The main difference is that we normally pick zero offsets two-way travel time, and the corresponding moveout of the reflection arrivals.

Full waveform inversion of seismic data

Until now we have learned about tomography that has been mainly based on travel time information, and hence provide low-resolution velocity structures of the sub-surface. Secondly, it is based on ray theory, assuming infinite frequency, but the real data have band-limited frequency. Thirdly, it is based on assuming only one type of wave, P-wave, but there are S-wave in the data. In order to shed light on the lithology of the subsurface (rock types, presence of fluids, melt, oil, gas etc) we need the quantitative estimation of P and S-wave velocities, anisotropy, attenuation etc.



Reflection and transmission at an interface



Inter-bed multiples

Waveform inversion can address these problems as well provide quantitative image of the subsurface. Waveform inversion has been developed for refraction data (Chapman and Orcutt, 1985, Carry and Chapman, 19988) for one dimension earth model. Tarantola started the application of the full waveform inversion to reflection data in eighties at IPG Paris and made a significant progress (Tarantola, 1984, 1988) and set up the foundation of waveform inversion. The idea is based on minimising the misfit between data and synthetic in a least-squares sense. The idea has been applied to 1D (Dietrich and Kormendi, 1989; Singh et al, 1991) and 2D problems (Crase et al., 1990). Here we discuss the basic theory of waveform inversion and show some inversion results and point out the strength and weakness of the waveform inversion.

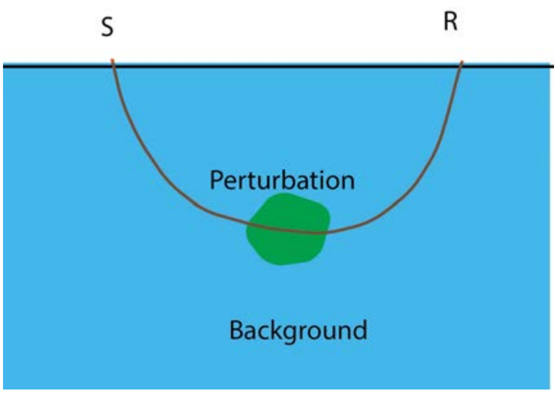
There are two main parts in inversion: Forward problem and inverse problem. In forward problem, there are two points to keep into consideration: model parameters and computation of synthetic data. In the inverse problem, we need to think about optimisation of a misfit between the observe data and synthetic data.

Let us first setup the formalism for acoustic medium

$$\left[\frac{1}{k(r)}\frac{\partial^2}{\partial t^2} - \nabla\cdot\left(\frac{1}{\rho(r)}\nabla\right)\right]P(r,t;r_s) = s(r,t;r_s)$$

$k(r)$ and $\rho\ (r)$ are compressibility and density in the medium, $P(r,t;r_s)$ is the pressure field and $s(r,t;r_s)$ is the source. This equation can be solve using numerical methods, such as finite difference method either in the time domain or in the frequency domain, using finite element or spectral element method. Show 2D model.

Let us assume that the parameters are perturbed by a small amount



$k(r) = k(r) + \delta k(r), \quad \rho(r) = \rho(r) + \delta\rho(r)$
and effecting the wave field such that

$$P(r,t;r_s) = P(r,t;r_s) + \delta P(r,t;r_s)$$

In this case the fields will follow

$$\left[\frac{1}{k(r)}\frac{\partial^2}{\partial t^2} - \nabla\cdot\left(\frac{1}{\rho(r)}\nabla\right)\right]P(r,t;r_s) = s(r,t;r_s)$$

$$\left[\frac{1}{k(r)}\frac{\partial^2}{\partial t^2} - \nabla\cdot\left(\frac{1}{\rho(r)}\nabla\right)\right]\delta P(r,t;r_s) = \Delta s(r,t;r_s)$$

$$\Delta s(r,t;r_s) = s(r,t;r_s) + \frac{\delta k(r)}{k^2(r)}\frac{\partial^2 P(r,t;r_s)}{\partial t^2} + \nabla\cdot\left[\frac{\delta\rho(r)}{\rho(r)}\nabla\right]P(r,t;r_s)$$

The solution of these equations can be written as function of the Green's function (g)

$$P(r,t;r_s) = \int dr' g(r,t;r',0) * s(r',t;r_s)$$

$$\delta P(r,t;r_s) = \int dr' g(r,t;r',0) * \Delta s(r',t;r_s)$$

The perturbed field can be expressed as

$$\delta P(r_g,t;r_s) = \frac{\partial P(r_g,t;r_s)}{\partial k(r)}\delta k(r) + \frac{\partial P(r_g,t;r_s)}{\partial \rho(r)}\delta\rho(r)$$

Let us take only the term related to the compressibility

$$\frac{\partial P(r_g,t;r_s)}{\partial k(r)}\delta k(r) = U(r_g,t;r_s)\delta k(r)$$

Including the source term for the compressibility only

$$U(r_g,t;r_s)\delta k(r) = \int dr\, g(r,t;r_g,0) * \frac{\delta k(r)}{k^2(r)}\frac{\partial^2 P(r,t;r_s)}{\partial t^2}$$

or

$$U(r_g,t;r_s)\delta k(r) = \int dr\, \frac{1}{k^2(r)}\frac{\partial g(r,t;r_g,0)}{\partial t} * \frac{\partial P(r,t;r_s)}{\partial t}\delta k(r)$$

The gradient can be written as

$$U(r_g,t;r_s) = \int dr \frac{1}{k^2(r)}\frac{\partial g(r,t;r_g,0)}{\partial t} * \frac{\partial P(r,t;r_s)}{\partial t}$$

At the nth iteration, the perturbed wavefield can be written as

$$\delta P_n(r,t;r_s) = P_0(r,t;r_s) + P_n(r,t;r_s)$$

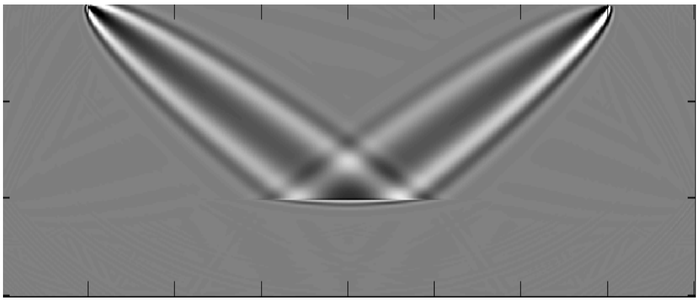
$$U(r_g,t;r_s)\delta k_n(r) = \delta P_n(r,t;r_s)$$

$$\delta \bar{k}_n = U^* \delta P_n$$

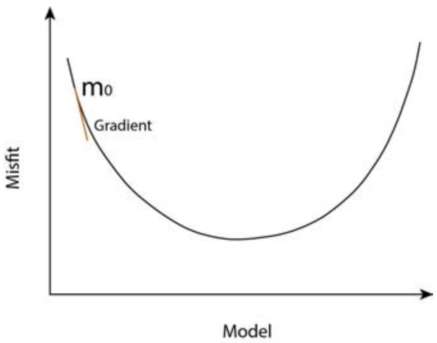
$$\delta \bar{k}_n(r) = \frac{1}{k_n^2(r)} \int dt \sum_s \frac{\partial \delta P_n(r,t;r_g,0)}{\partial t} * \frac{\partial P_n(r,t;r_s)}{\partial t}$$

The first term in the above equation perturbed wavefield propagated backwards and the second term is the forward propagated wave field

$$\delta \bar{k}_n(r) = \frac{1}{k_n^2(r)} \int dt \sum_s \frac{\partial \overleftarrow{\delta P_n(r,t;r_g,0)}}{\partial t} * \frac{\partial \overrightarrow{P_n(r,t;r_s)}}{\partial t}$$



There are different methods to converge to the local minima:



Newton, Quasi-Newton, Gauss-Newton, Steepest Descent, Conjugate Gradient methods. Here, we will discuss the conjugate gradient method. Starting from the initial model, $\mathbf{m_0}$, current model at iteration I can be written

$$\boldsymbol{m_{i+1}} = \boldsymbol{m_i} + \boldsymbol{\alpha_i g^c(m_i)}$$

where $\boldsymbol{\alpha}$ step length in the direction of conjugate gradient $\boldsymbol{g^c(m_i)}$. The first step of the gradient is

$$g_1^c = g_1$$

is steepest descent at the starting model. The conjugate gradient can be computed as

$$g_i^c = g_i + \frac{(g_i - g_{i-1})}{g_i^T g_i} g_{i-1}^c$$

where g_i^c and g_{i-1}^c are conjugate, and more stable and efficient. The step length can be computed as

$$\alpha_i = \frac{g_i . g_i^c}{g_i^T H g_i^c}$$

Parameters for inversion:

Reflection coefficient depends on the impedance contrast across an interface, which can be defined as $IP = \rho V_p$, *and* $IS = \rho V_s$. Therefore, one can parameterise the model either using (λ, μ, ρ), (V_p, V_s, ρ) or (IP, IS, ρ). For long wavelength spatial inversion, (V_p, V_s, ρ) are the best parameters because they influence the travel time arrivals along with amplitude and phase. This is the case for wide-angle or refraction data. For short wavelength spatial variations, (IP, IS, ρ) are the best parameters for the inversion. Waveform of seismic reflection data contains only short wavelength structural information. However, if one wishes to invert both seismic reflection and refraction data simultaneously, one should invert for (V_p, V_s, ρ).

These values can be replaced by the P and S-wave velocities as

$$\delta \overline{V_p} = 2\rho V_P \overline{\delta\lambda}$$
$$\delta \overline{V_S} = -4\rho V_S \overline{\delta\lambda} + 2\rho V_S \overline{\delta\mu}$$

Starting model: As mentioned before, the waveform inversion requires that the starting model gives travel time within a half of the period of the final model. Results from tomography could be used as starting model. Since waveform contains information on fine details (short wavelengths), the starting model should be smooth. For near offset data reflection or for very far offset refraction data, an acoustic approximation could be used, which means only P-wave velocity model is required. For elastic inversion, one requires information on P-wave and S-wave velocities and density. Since long wavelength information on density and S-wave velocity is generally not known, one uses some relationship between P-wave velocity and these parameters. For example, we have used the following relationship based on empirical formula (Gardener,1974; Hamilton, 1978, Shipp and Singh, 2002):
S-wave velocity from P-wave velocity

$$\beta = \begin{cases} 0.0 & \text{for } \alpha \le 1.5 \\ (\alpha - 1.36)/1.16 & \text{for } 1.5 < \alpha < 3.5 \\ 0.53\alpha & \text{for } \alpha \ge 3.5 \end{cases}$$

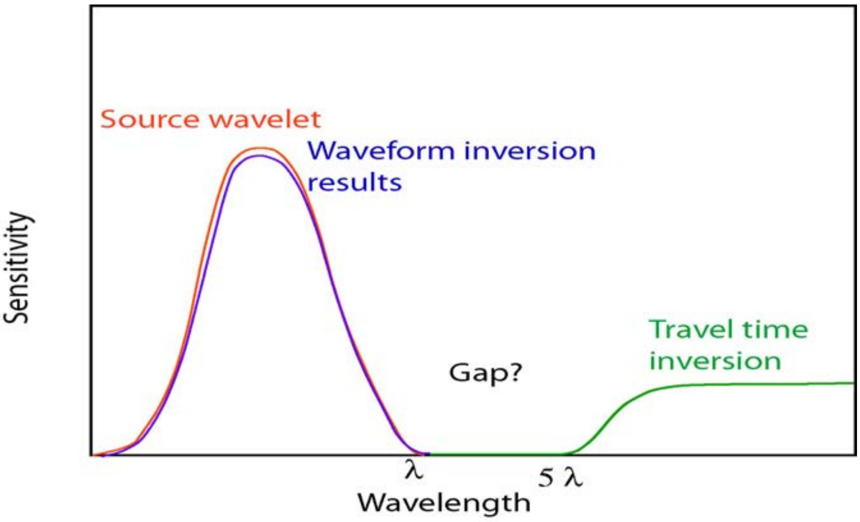
and density from P-wave velocity

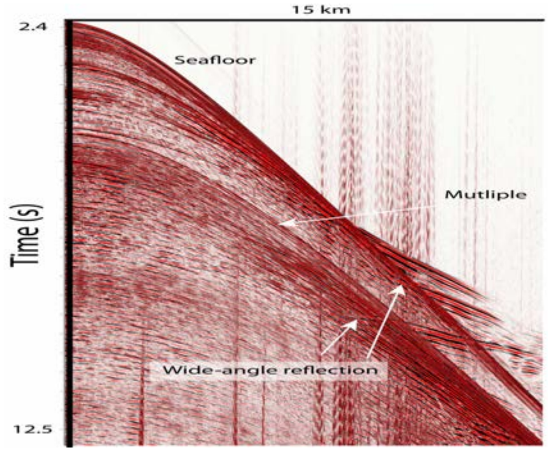
$$\rho = \begin{cases} 1000 & \text{for } \alpha \le 1.5 \\ 2351 - 7497\alpha^{-4.656} & \text{for } 1.5 < \alpha < 3.5 \\ 1740\alpha^{0.25} & \text{for } \alpha \ge 3.5 \end{cases}$$

where density in kg m^{-3} .

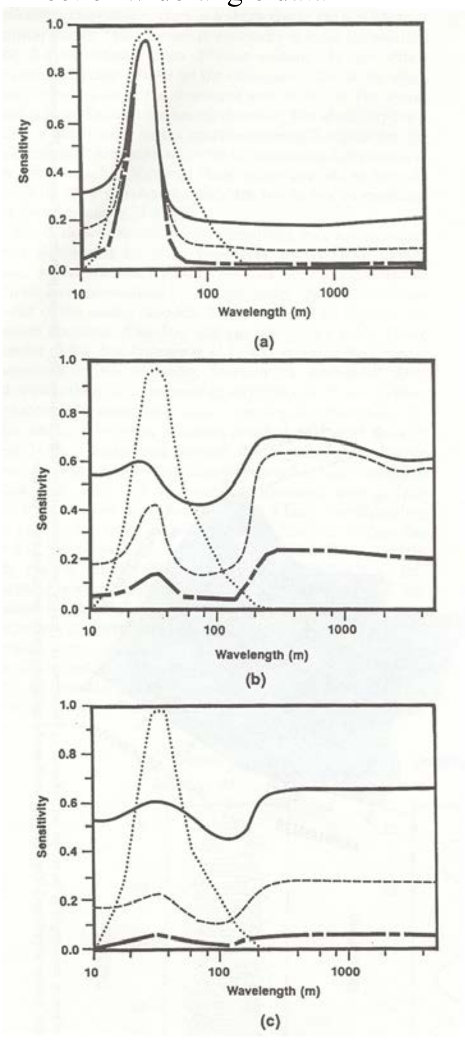
In nineties, the Tarantola's group at IPG Paris carried out extensive work on waveform inversion, but the results were not very promising. The main reasons were the short offset and the wavelength gap.

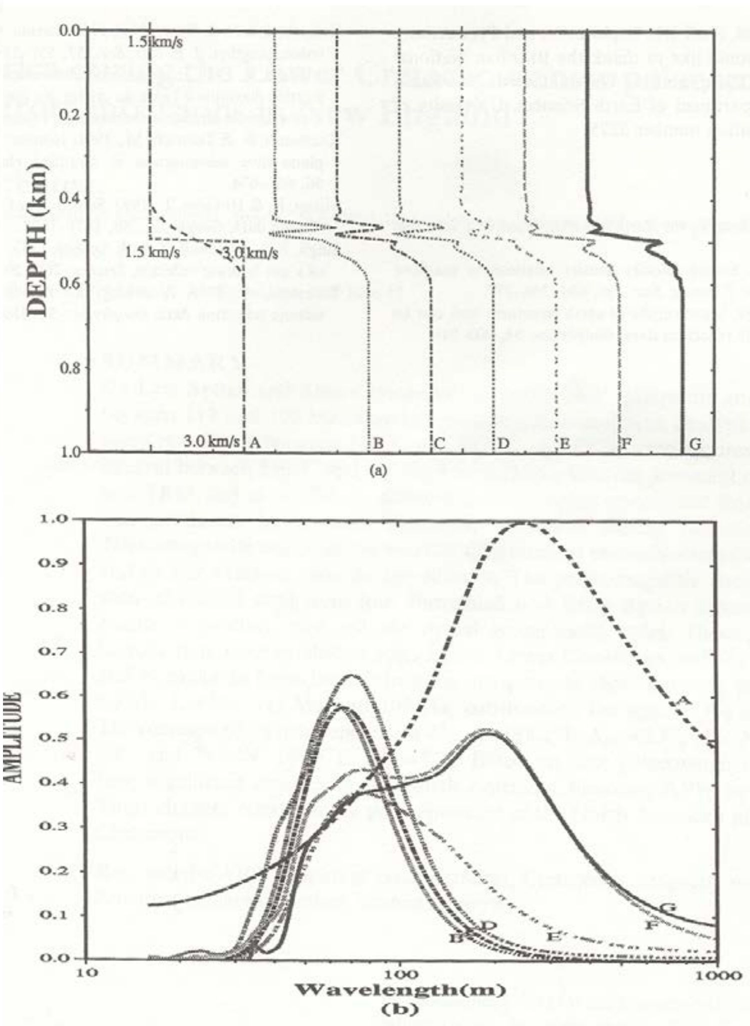
Sensitivity





Effect of wide-angle data





Synthetic example:

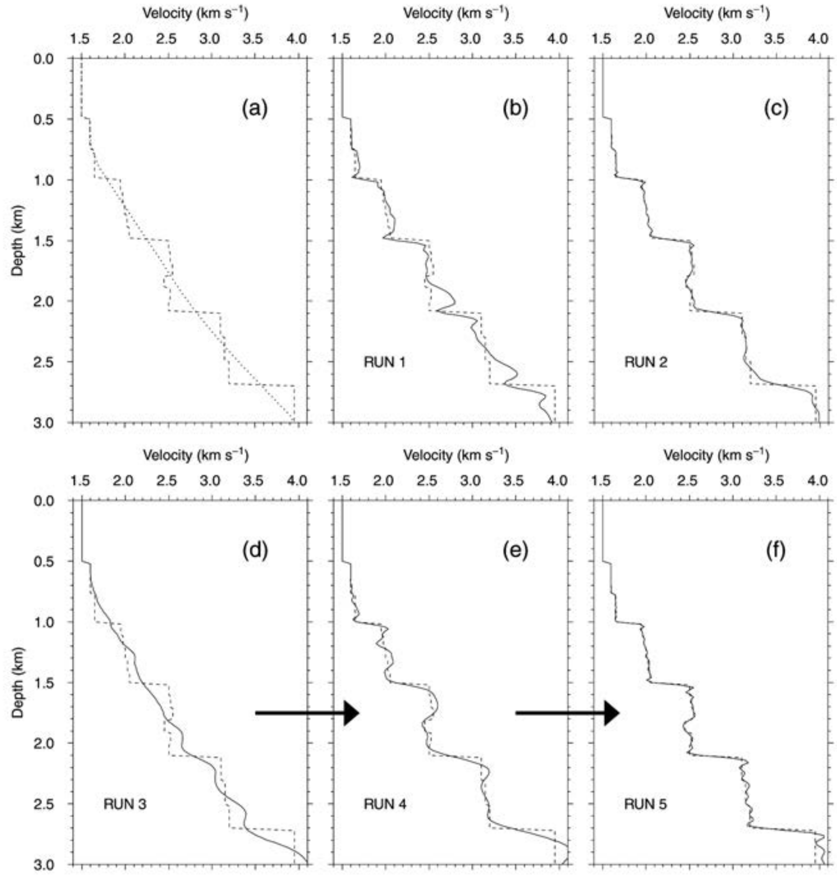


Figure: Synthetic inversion results. (a) True model (dashed) and starting velocity model (dotted) obtained after tomographic inversion. (b) Near offset (0-4 km) inversion results (solid line). (c) Inversion of 0-12 offset data at same time. Long offset to near offset inversion strategy: (d) Inversion of long offset data (8-12 km), (e) 4-8 km and (f) then 0-4 km offset.

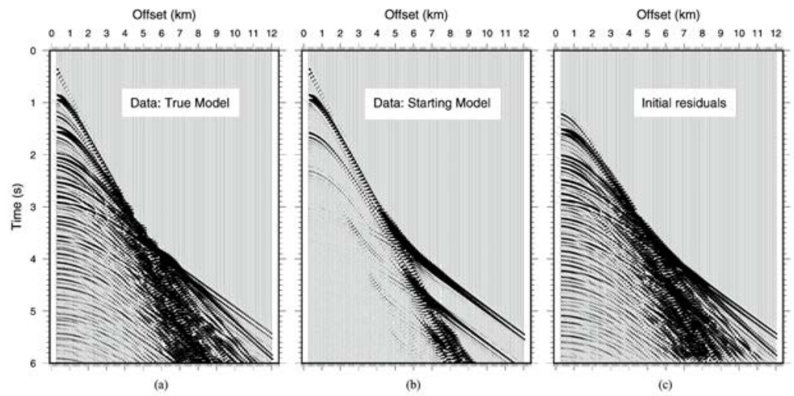


Figure 2. Results from example 1: Data from the wide-aperture wavefield inversion of Fig. 1(c). Synthesized observed data, initial synthetic data and initial residuals plotted with the same scaling.

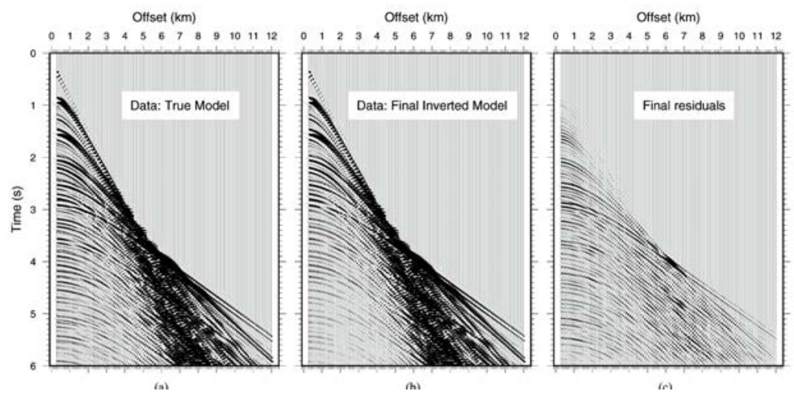
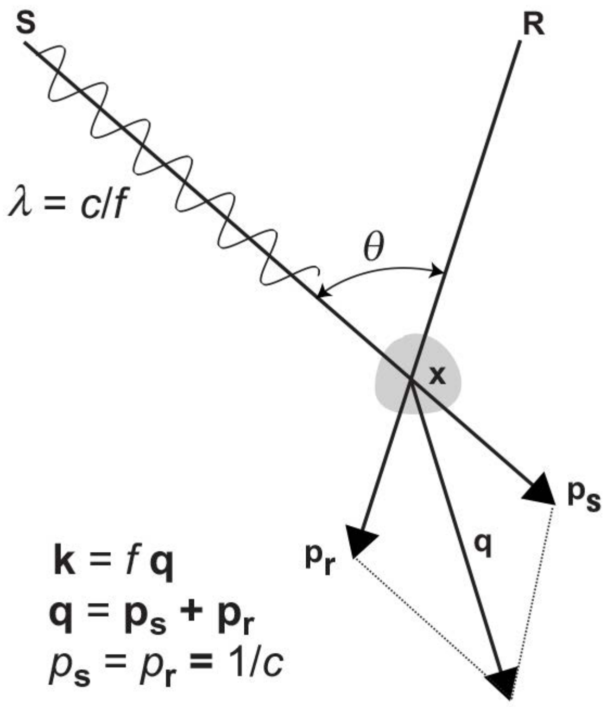
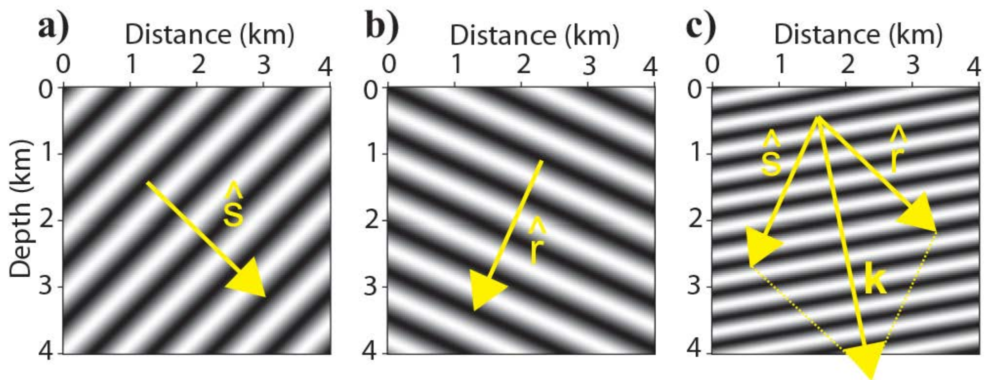


Figure: Real, synthetic and residual data before and after the inversion.

Data redundancy:

The data could be decimated either in the offset domain or in the frequency domain, which brings out the idea of performing the inversion in the frequency domain. Before we discuss the data redundancy, let us look at the relationship between the frequency and wave number.





Using the Born scattering theory, we could easily derive a relationship between wavenumber that can be retrieved from a single frequency inversion (Virieux and Operto, 2009)

$$\mathbf{k} = \frac{2f}{c_0}\cos\left(\frac{\theta}{2}\right)\mathbf{n}$$

where **n** is a unit vector in the direction of slowness vector. This equation suggests that for one frequency and one aperture in the data space map to one wavenumber in the model space. And therefore, frequency and aperture have redundant control on the wavenumber coverage, and this redundancy increases with aperture coverage. Pratt et al (1996) used this idea to propose decimating this wavenumber-coverage redundancy in the frequency domain and developed a FWI of single discrete frequencies. This allows model seismic data one frequency at a time and perform single frequency at a time starting from low frequencies and increasing to higher frequencies.

The wave acoustic equation can be re-written

$$\left[\frac{1}{k(r)}\frac{\partial^2}{\partial t^2} - \nabla\cdot\left(\frac{1}{\rho(r)}\nabla\right)\right]P(r,t;r_s) = s(r,t;r_s)$$

which in frequency domain can be written as

$$\left\{-\frac{1}{k(r)}\omega^2 - \nabla.\left(\frac{1}{\rho(r)}\nabla\right)\right\}P(r,\omega,r_s) = s(r,\omega,r_s)$$

$$B(r,\omega)P(r,\omega,r_s) = s(r,\omega,r_s)$$

You could solve this equation for single frequency for all sources simultaneously, which can be solved efficiently, and then performed single frequency inversion.

One should note that the misfit function:

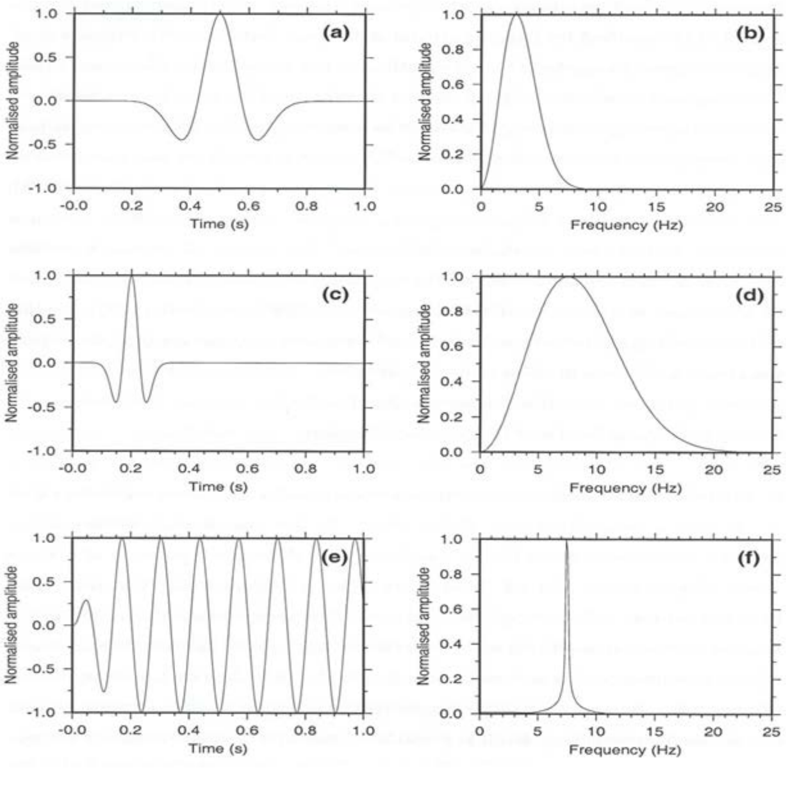
$$S(m) = \|d_{obs}(x,t) - d_{cal}(x,t)\|^2 = \|d_{obs}(x,\omega) - d_{cal}(x,\omega)\|^2$$

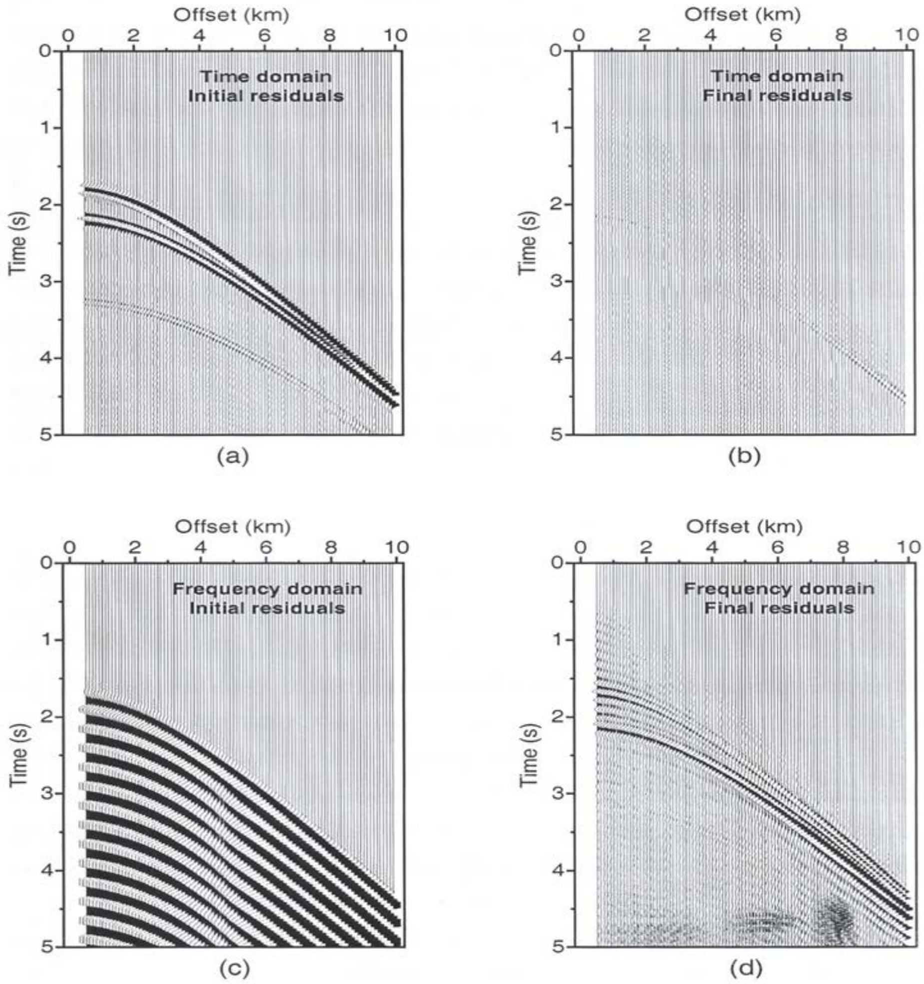
Because of the Parseval theorem

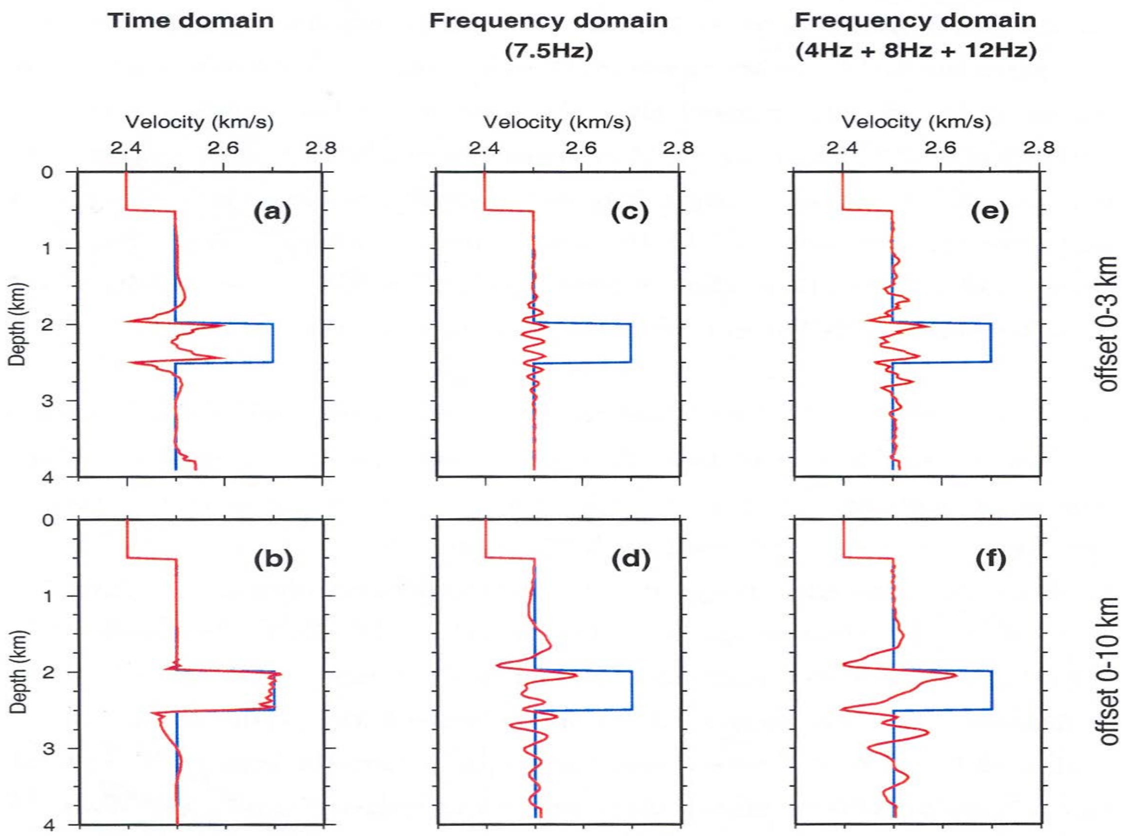
$$\int |u(t)|^2\, dt = \int |u(\omega)|^2\, d\omega$$

As mentioned above, we can perform one frequency at a time for different sources simultaneously, which would make the inversion very efficient.

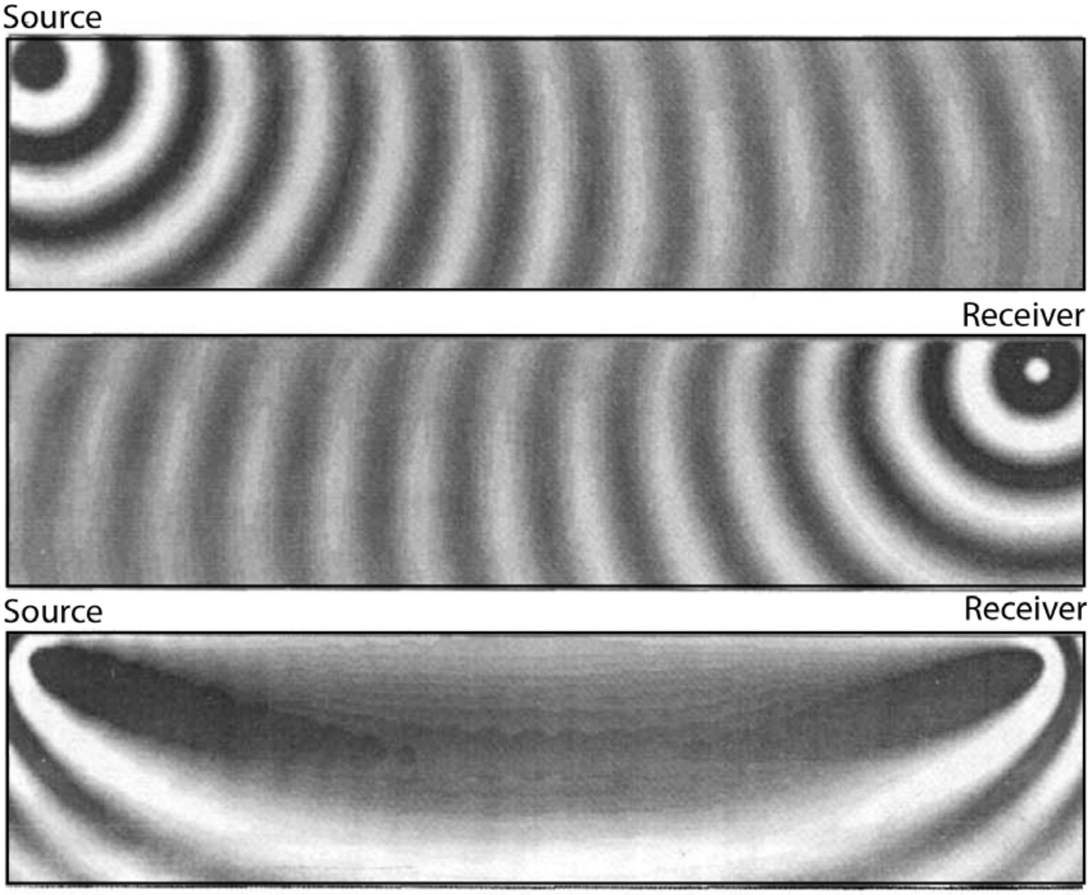
But remember one frequency and one aperture will map to a one wave number. Note a single frequency corresponds to sine wave. If you look at the correspondence between single frequency and inversion of pre-critical reflection data, you will find that we will get one wave number.




Offset (km)
Time (s)
Time domain
Initial residuals
Final residuals
Frequency domain
(a)
(b)
(c)
(d)


Time domain
Frequency domain
(7.5Hz)
Frequency domain
(4Hz + 8Hz + 12Hz)
Velocity (km/s)
Depth (km)
(a)
(b)
(c)
(d)
(e)
(f)
offset 0-3 km
offset 0-10 km

Transmitted waves in frequency domain



Pratt et al (1996)

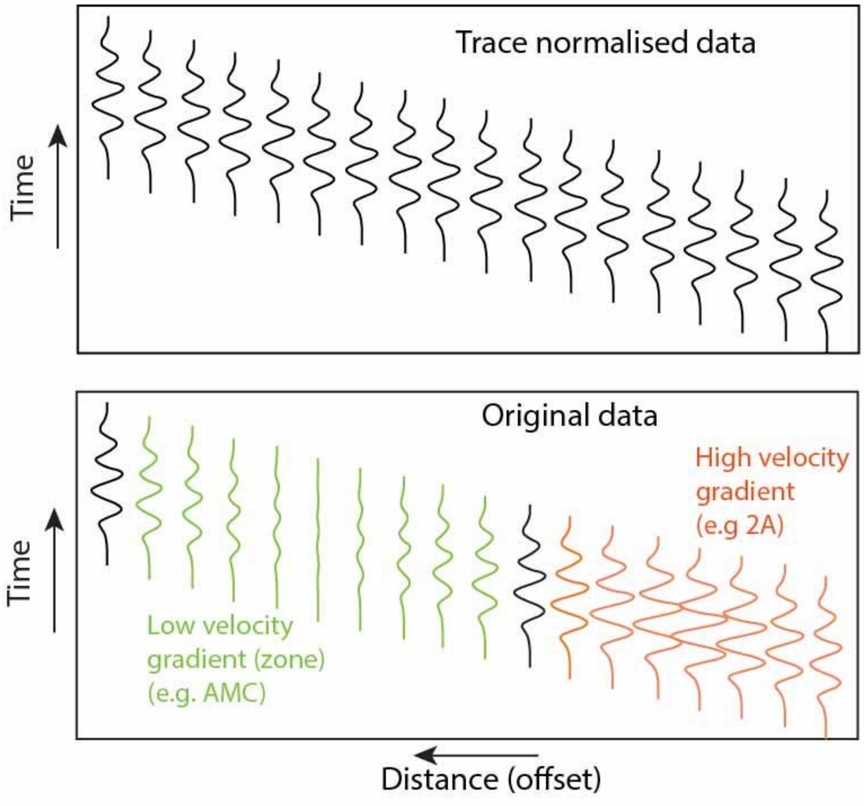
Note the lowest frequency is 0.5 Hz. The source wavelets used in active source seismology experiments have a bandwidth of 4-70 Hz. If we had zero frequency, we could invert long wavelength structures. So there is a trade-off between offsets and frequency content. Longer the offset, better the constraints we have on long wavelength structures.

Waveform tomography (Trace normalised waveform inversion)

Both amplitude and phase of seismic data change with offset. The solution of the acoustic wave equation can be written as

$$u(x,\omega) = A(x,\omega)e^{-i\omega\psi(t)}$$

Where A is amplitude and ψ is the phase or travel time. The amplitude would vary with offset.
For example, if you have shadow zone due to low velocity zone where amplitude goes to zero or high velocity gradient zone where one can have triplication, and the amplitude could very be large because of caustics. In the case of reflection, the reflection coefficient changes with offsets. However, when your amplitude is not reliable, e.g., different instruments may have different instruments response, local coupling etc., people normalise the trace, e.g., remove the effect of amplitudes. In this case, only waveform is fitted in the inversion, and it is called waveform tomography. Some people use the normalised cross-correlation function (similar to wave equation tomography), which contains come information about the wavefield. Others applying a pre-whitening in frequency domain, and invert one frequency at a time, hence fitting the phase term only.



Trace normalisation consists of normalisation a trace with the sum of the square of its amplitude, which is equivalent to energy term in a trace, defined as $E_{ij} = \sum_{t=0}^{t} u_{ij}^2(t) = \|u_{ij}\|^2$, where i is source number and j is the receiver number. We can define the normalisation factor as $|u_{ij}| = \sqrt{E_{ij}}$. If we define d data and u as synthetic, then misfit function can be written as

$$J_N(m) = \sum_{i,j} u_{ij}^T \left(\frac{u_{ij}}{|u_{ij}|} - \frac{d_{ij}}{|d_{ij}|} \right)$$

$$J_N(m) = \sum_{i,j} \left(|u_{ij}| - u_{ij}^T \frac{d_{ij}}{|d_{ij}|} \right)$$

The gradient of this misfit function as function of velocity will be

$$\frac{\partial J_M}{\partial m} = \sum_{i,j} \left(\frac{u_{ij}}{|u_{ij}|} - \frac{d_{ij}}{|d_{ij}|} \right) \left(\frac{\partial u_i}{\partial m} \right)$$

So the adjoint source for the trace will be $\left(\frac{u_{ij}}{|u_{ij}|} - \frac{d_{ij}}{|d_{ij}|} \right)$, which is the normalised residual.

Full waveform inversion

In full waveform inversion we minimise the difference between observed and synthetic data in a least square sense

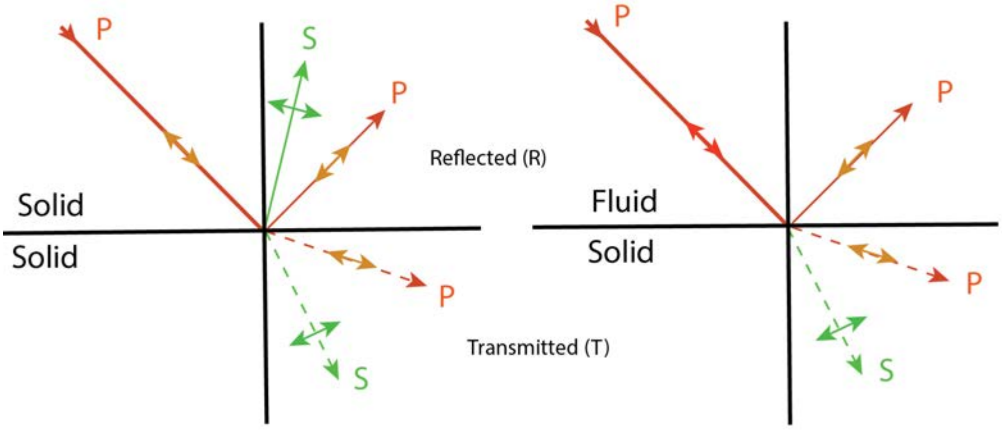
$$J(m) = \sum_{i.j} (u_{ij} - d_{ij})^T (u_{ij} - d_{ij})$$

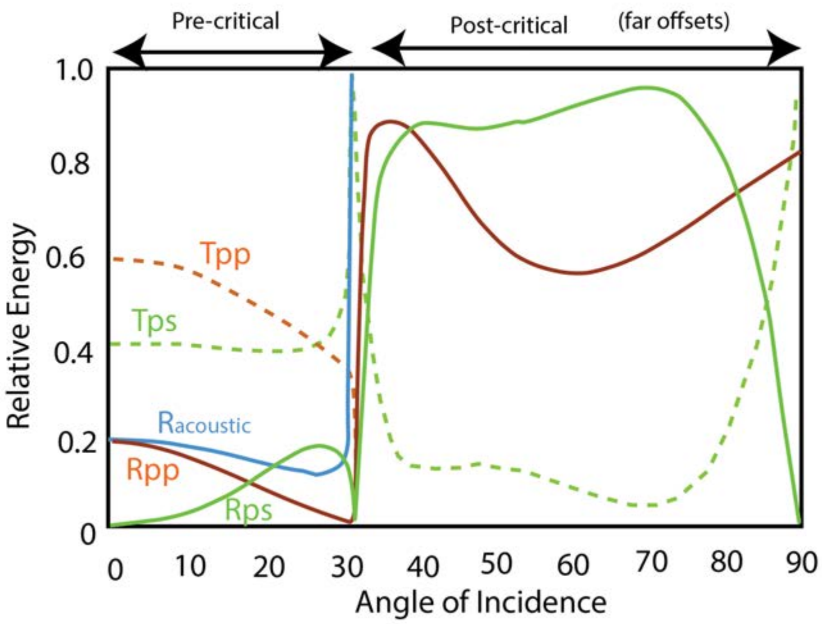
and gradient will be

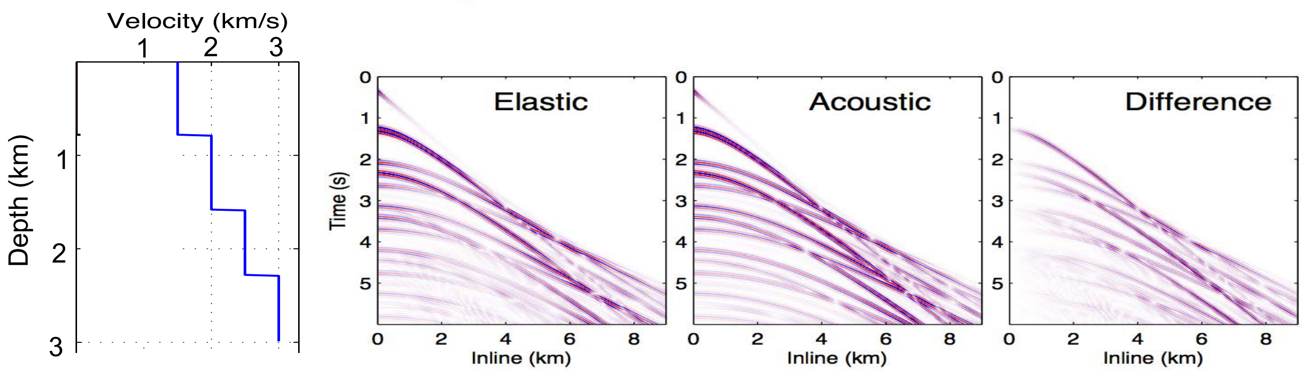
$$\frac{\partial J}{\partial m} = \sum_{i.j} (u_{ij} - d_{ij}) \left(\frac{\partial u_i}{\partial m}\right)$$

The data residual, $u_{ij} - d_{ij}$, will be the adjoint source.

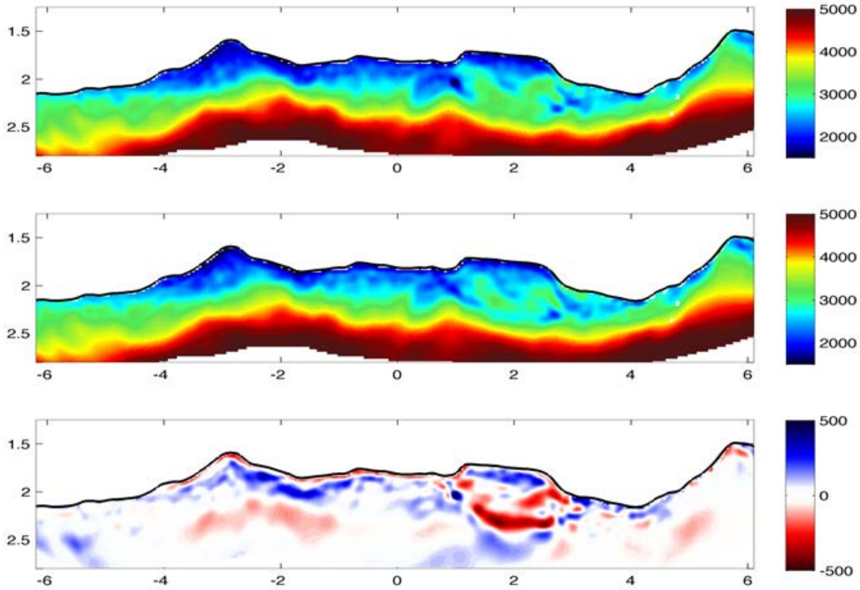
Acoustic Versus Elastic







Comparison between acoustic and elastic inversion:



Elastic FWI

Acoustic FWI

Difference

Arnulf et al GJI (2014)

Since full waveform inversion depends on the point-by-point difference between the observed data and synthetic data, anything that has not been modelled will map in the model. Here, we have mainly discussed elastic media, but the media could anisotropic, visco-elastic etc., and hence these effects should be taken into account to get accurate results, but this would increase the computation time, and therefore, most of the inversion are performed using acoustic approximation.

In any case, full waveform inversion is most powerful method available to determine fine-scale quantitative structures of the sub-surface and is being developed/used both at local and global scale.

3D localised full waveform inversion

Sub-basalt imaging problems: