

SMR: 1059/21

***FOURTH WORKSHOP ON THREE DIMENSIONAL MODELLING
OF SEISMIC WAVES GENERATION,
PROPAGATION AND THEIR INVERSION***

28 September - 9 October 1998

Body Wave Tomography and Ray Tracing

presented by:

Tatyana YANOVSKAYA

St. Petersburg State University

Institute of Physics

Ulyanovskaya St. 1

Petrodvoretz

198904 St. Petersburg

Russia

These are preliminary lecture notes, intended only for distribution to participants.

Body wave tomography and ray tracing

Contents

1. Basic principles of wave propagation in inhomogeneous media
 - 1.1 Equation of motion
 - 1.2 Eikonal equation
 - 1.3 Equation of ray
2. Computational methods for two-point ray tracing
 - 2.1 Shooting method
 - 2.2 Bending method
 - 2.3 Calculation of the time field
3. Principles of the ray tomography
 - 3.1 Formulation of the problem
 - 3.2 Linearization of the problem
 - 3.3 Main approaches for solving the tomography problems
4. Resolving power
 - 4.1 Resolution matrix
 - 4.2 Estimation of resolution by mean size of the averaging area
 - 4.3 Checkerboard test

1. Basic principles of wave propagation in inhomogeneous media

1.1. Equation of motion

If the elastic moduli λ , μ as well as density ρ depend on spatial coordinates $\mathbf{x}$, then from the general equation of motion

$$\nabla \mathbf{T} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2}$$

it follows

$$(\lambda + 2\mu)\nabla \text{div} \mathbf{u} - \mu \text{rot} \text{rot} \mathbf{u} + \nabla \lambda \text{div} \mathbf{u} + \nabla \mu \times \text{rot} \mathbf{u} + 2(\nabla \mu, \nabla) \mathbf{u} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \quad (1)$$

An exact solution of (1) cannot be determined. Therefore approximate methods are applied to estimate the solution. Widely used is the *ray method*. It is valid for high frequencies, as well as in the vicinity of the *wave fronts*.

The concept of the wave front can be understood from the following consideration. Assume that in the elastic medium a source begins to act in some area at $t=0$. At $t<0$ the medium was motionless, at $t>0$ a disturbance begins to propagate, and at a moment t it achieves a surface Σ , which bounds a disturbed area. This surface is called the wave front. Its location depends on the moment t . Equation for Σ may be written as $\tau(\mathbf{x})=t$. Therefore a solution of (1) may be written in the following general form

$$\mathbf{u}(\mathbf{x}, t) = \mathbf{A}(\mathbf{x}, t) f(t - \tau(\mathbf{x}))$$

where $\mathbf{A}(\mathbf{x}, t)$ is an analytical function, and

$$f(t) = \begin{cases} 0 & t < 0 \\ \neq 0 & t > 0 \end{cases}$$

According to the ray theory a solution of the equation (1) is determined in the form of the *ray series*

$$\mathbf{u}(\mathbf{x}, t) = \sum_{k=0}^{\infty} \mathbf{u}_k(\mathbf{x}) f_k(t - \tau(\mathbf{x})) \quad (2)$$

where the functions $f_k(t)$ satisfy the following recurrent relationship

$$f'_k(t) = f_{k-1}(t) \quad (3)$$

The highest order of discontinuity on the wave front is determined by the first function $f_0(t)$.

Substituting (2) to (1) we obtain a series in the left-hand side. Taking into account (3) and collecting the terms with the functions $f_k(t)$ with one and the same subscript k , we obtain

$$\sum_{k=0}^{\infty} \{ \mathbf{L}(\mathbf{u}_{k-2}) - \mathbf{M}(\mathbf{u}_{k-1}) + \mathbf{N}(\mathbf{u}_k) \} f_{k-2}(t - \tau) = 0 \quad (4)$$

where $\mathbf{L}(\mathbf{u})$ is differential operator in the left-hand side of the equation (1), $\mathbf{M}(\mathbf{u})$ is a differential operator of the first order, and

$$\mathbf{N}(\mathbf{u}) = (\lambda + \mu)(\mathbf{u}, \nabla \tau) \nabla \tau + (\mu(\nabla \tau)^2 - \rho) \mathbf{u}$$

The coefficients by all the functions $f_k(t)$ should vanish.

Since $\mathbf{u}_{k-2} = \mathbf{u}_{k-1} = 0$, the first equation is

$$\mathbf{N}(\mathbf{u}_0) = 0$$

or

$$(\lambda + \mu)(\mathbf{u}_0, \nabla \tau) \nabla \tau + (\mu(\nabla \tau)^2 - \rho)\mathbf{u}_0 = 0 \quad (5)$$

Equate to zero the components of (5) in the direction of $\nabla \tau$ in perpendicular direction (l):

$$N_r = ((\lambda + 2\mu)(\nabla \tau)^2 - \rho)\mathbf{u}_{0r} = 0$$

$$N_l = (\mu(\nabla \tau)^2 - \rho)\mathbf{u}_{0l} = 0$$

This system has two solutions:

$$\begin{aligned} (1) \quad u_{0l} &= 0, \quad (\nabla \tau)^2 = \frac{\rho}{\lambda + 2\mu} = \frac{1}{a^2} \\ (2) \quad u_{0r} &= 0, \quad (\nabla \tau)^2 = \frac{\rho}{\mu} = \frac{1}{b^2} \end{aligned} \quad (6)$$

The first of the solutions (6) corresponds to P-wave, the second one – to S-wave. The equation

$$(\nabla \tau)^2 = \frac{1}{c^2}$$

is called *the eikonal equation*.

1.2. Eikonal equation

Below we show that the solution of the eikonal equation is Fermat functional

$$\tau(M, M_0) = \int_{M_0}^M \frac{ds}{c(\mathbf{x})}$$

calculated along the extremal.

At first we derive the equation for the extremal of the Fermat functional. Represent the coordinates x, y, z on the extremal in parametric form: $x=x(\sigma)$, $y=y(\sigma)$, $z=z(\sigma)$, then $ds = \sqrt{x_\sigma^2 + y_\sigma^2 + z_\sigma^2} d\sigma$, so the integrand can be represented as

$$F(x, y, z, x_\sigma, y_\sigma, z_\sigma) d\sigma$$

where

$$F = \frac{\sqrt{x_\sigma^2 + y_\sigma^2 + z_\sigma^2}}{c(x, y, z)}$$

The equation for the extremal is obtained from the Euler equations:

$$\frac{\partial F}{\partial x} - \frac{d}{d\sigma} \left(\frac{\partial F}{\partial x_\sigma} \right) = 0, \quad \frac{\partial F}{\partial y} - \frac{d}{d\sigma} \left(\frac{\partial F}{\partial y_\sigma} \right) = 0, \quad \frac{\partial F}{\partial z} - \frac{d}{d\sigma} \left(\frac{\partial F}{\partial z_\sigma} \right) = 0, \quad (7)$$

Denote

$$\sqrt{x_\sigma^2 + y_\sigma^2 + z_\sigma^2} = |\mathbf{r}_\sigma|$$

Then the equations (7) may be written as

$$|\mathbf{r}_\sigma| \nabla \left(\frac{1}{c} \right) - \frac{d}{d\sigma} \left(\frac{\mathbf{r}_\sigma}{c |\mathbf{r}_\sigma|} \right) = 0$$

Denote the unit vector tangent to the extremal $\mathbf{t} = \frac{\mathbf{r}_\sigma}{|\mathbf{r}_\sigma|}$. Then, taking into account that

$ds = |\mathbf{r}_\sigma| d\sigma$, we can write the equation for the extremal as

$$\frac{d}{ds}\left(\frac{\mathbf{t}}{c}\right) = \nabla\left(\frac{1}{c}\right) \quad (9)$$

Extremal of the Fermat functional is a *ray*. Thus (9) is the *ray equation*.

Now we show that the value of τ calculated along the extremal is the solution of the eikonal equation.

From the eikonal equation follows $\nabla\tau = \frac{\mathbf{t}}{c}$, where $\mathbf{t}$ is unit vector orthogonal to a surface $\tau(\mathbf{x})=\text{const}$. Transform the derivative $\frac{d}{ds}\left(\frac{\mathbf{t}}{c}\right)$, where differentiation is along the direction $\mathbf{t}$:

$$\frac{d}{ds}\left(\frac{\mathbf{t}}{c}\right) = \frac{d\nabla\tau}{ds} = (\mathbf{t}, \nabla)\nabla\tau = c(\nabla\tau, \nabla)\nabla\tau = \frac{c}{2}\nabla(\nabla\tau)^2 = \frac{1}{2}c\nabla\left(\frac{1}{c^2}\right) = \nabla\left(\frac{1}{c}\right) \quad (10)$$

So the equation of a line orthogonal to the surface $\tau=\text{const}$ coincide with the equation for the extremal of the Fermat functional. Taking into account that $d\tau = ds/c$, where ds is an element of this line, we obtain that

$$\tau = \tau_0 + \int \frac{ds}{c}, \text{ where the integral is calculated along the extremal.}$$

1.3. Equation of ray

It follows from (9) that the equation for the slowness vector $\mathbf{p} = \mathbf{t}/c$ is

$$\frac{d\mathbf{p}}{ds} = \nabla\left(\frac{1}{c}\right) \quad (11)$$

Because $\mathbf{t}$ is unit vector tangent to the ray, $\mathbf{t} = \frac{d\mathbf{x}}{ds}$, or

$$\frac{d\mathbf{x}}{ds} = c\mathbf{p} \quad (12)$$

The equations (11),(12) determine the ray – they are so-called *equations of ray*.

They are the differential equations of the first order, and to solve them we have to know the initial conditions. It can be shown that for determination of the rays field in space it is sufficient to know the values of τ at some surface S . In fact, if we know $\tau(S)$, we can calculate the derivatives along two any orthogonal directions on S , i.e. x_1 and x_2 . But since $|\nabla\tau| = \frac{1}{c}$, we can also calculate

$$\frac{\partial\tau}{\partial n} = \sqrt{\frac{1}{c^2} - \left(\frac{\partial\tau}{\partial x_1}\right)^2 - \left(\frac{\partial\tau}{\partial x_2}\right)^2}$$

and determine the vector $\nabla\tau=\mathbf{p}$ on the whole surface S . So from any point of the surface we can construct the rays, because at the surface we know both $\mathbf{x}$ and $\mathbf{p}$.

The vector equations (11) and (12) can be represented as 6 scalar equations for the components of $\mathbf{x}$ and $\mathbf{p}$. However they are not independent, because

$$p_x^2 + p_y^2 + p_z^2 = \frac{1}{c^2}. \text{ Therefore 6 equations can be reduced to 5.}$$

If the solution is determined in the Cartesian coordinates x, y, z , then instead of the components of the slowness vector it is convenient to use the angles θ and φ specifying the unit vector $\mathbf{t}$. The equations for x, y, z , θ and φ are:

$$\begin{aligned}\frac{dx}{ds} &= \sin \theta \cos \varphi \\ \frac{dy}{ds} &= \sin \theta \sin \varphi \\ \frac{dz}{ds} &= \cos \theta \\ \frac{d\theta}{ds} &= -\frac{\cos \theta}{c} (c_x \cos \varphi + c_y \sin \varphi) + \frac{\sin \theta}{c} c_z \\ \frac{d\varphi}{ds} &= \frac{1}{c \sin \theta} (c_x \sin \varphi - c_y \cos \varphi)\end{aligned}\tag{13}$$

Thus, if a direction of the ray is known in a point x_0, y_0, z_0 , the ray may be calculated by numerical integration of the system (13).

2. Computational methods for two-point ray tracing

2.1. Shooting method

In different problems it is required to calculate a ray between two fixed points – source and receiver. But if we integrate the system (13) with the initial data in the source, the ray cannot hit the receiver point. To hit the receiver we have to vary the direction of the ray in the source using one or another method.

Assume that that we can calculate a ray sufficiently close to that, which crosses the receiver. The problem may be linearized, and is formulated as follows:

Let the receiver coordinates are x_r, y_r, z_r (without loss of generality we may assume $z_r = 0$). For given angles in the source θ_0, φ_0 the ray crosses a plane $z=0$ at the points x_q, y_q . It is necessary to determine the corrections to the angles $\delta\theta, \delta\varphi$, so as the ray leaving the source x_0, y_0, z_0 at the angles $\theta_0 + \delta\theta, \varphi_0 + \delta\varphi$ would hit the receiver point.

If the both rays are calculated for one and the same velocity distribution $c(\mathbf{x})$, this problem can be solved by interpolation using some numerical scheme.

Example. We calculate two additional rays corresponding to the initial angles

(1) $\theta_0 + \delta\theta_0, \varphi_0$; (2) $\theta_0, \varphi_0 + \delta\varphi_0$; let the rays cross the plane $z=0$ at the points

x'_q, y'_q and x''_q, y''_q respectively. If $x = x(\theta, \varphi), y = y(\theta, \varphi)$, then

$$\begin{aligned}x'_q - x_q &\approx \frac{\partial x}{\partial \theta_0} \delta\theta_0, & y'_q - y_q &\approx \frac{\partial y}{\partial \theta_0} \delta\theta_0 \\ x''_q - x_q &\approx \frac{\partial x}{\partial \varphi_0} \delta\varphi_0, & y''_q - y_q &\approx \frac{\partial y}{\partial \varphi_0} \delta\varphi_0\end{aligned}$$

These formulas allow the derivatives to be estimated. Using these derivatives we can estimate the corrections $\delta\theta, \delta\varphi$ from the linear system:

$$x_r - x_q = \left(\frac{\partial x}{\partial \theta_0} \right) \delta \theta + \left(\frac{\partial x}{\partial \varphi_0} \right) \delta \varphi$$

$$y_r - y_q = \left(\frac{\partial y}{\partial \theta_0} \right) \delta \theta + \left(\frac{\partial y}{\partial \varphi_0} \right) \delta \varphi$$

Another approach for interpolation of the ray into the given receiver point is based on the use of the *dynamic ray tracing*.

We introduce the so-called *ray coordinates*: τ, γ_1, γ_2 : τ is time on the ray, γ_1, γ_2 are the parameters specifying the ray (in the particular case they may be θ_0, φ_0). Also we introduce a coordinate system in the vicinity of the ray on the plane orthogonal to the ray: q_1, q_2 . The related components of the slowness vector define as p_1, p_2 . The ray equation in these coordinates takes the form:

$$\frac{\partial q_i}{\partial s} = c p_i$$

$$\frac{\partial p_i}{\partial s} = - \frac{\partial}{\partial q_i} \left(\frac{1}{c} \right)$$

Now consider the derivatives of q_1, q_2 and p_1, p_2 in respect to the ray parameters γ_1, γ_2 :

$$Q_y = \frac{\partial q_i}{\partial \gamma_j}, \quad P_y = \frac{\partial p_i}{\partial \gamma_j}$$

Then we obtain the following differential equations for Q_y, P_y :

$$\frac{dQ_y}{ds} = c P_y, \quad \frac{dP_y}{ds} = -c^2 \left(\frac{\partial^2 c}{\partial q_i \partial q_1} Q_{1j} + \frac{\partial^2 c}{\partial q_i \partial q_2} Q_{2j} \right) \quad (14)$$

The equation (14) may be integrated jointly with the ray equation, so that we can obtain the derivatives of the coordinates in respect to the parameters of the ray, and use them for interpolation of the coordinates of the ray. However this approach is valid only in the cases when c is continuous with its first derivatives.

Drawbacks of this approach:

- it is time consumable, because the procedure should be repeated until a required precision is achieved, and at each step it needs numerical integration;
- non-linearity in case of low-velocity zones.

2.2. Bending method

Whereas the shooting method is based on the equation of ray, the bending method proceeds from the Fermat's principle: travel time between the two points should be minimum. This method is used for cases, when the ray between the two points is known for an initial velocity model, and the problem is reduced to calculation of the ray for a perturbed model. The idea of the method – to perturb the initial ray so that to achieve minimum travel time. The initial ray is divided to some segments, and the end points of the segments are varied while the minimum of the following functional is achieved:

$$\Phi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n-1}) = \sum_n \int_{\mathbf{x}_k}^{\mathbf{x}_{k+1}} \frac{ds}{c(s)} \quad (15)$$

The segments are usually assumed to be straight lines, and if their lengths are sufficiently small, the integrals may be estimated as

$$\int_{\mathbf{x}_k}^{\mathbf{x}_{k+1}} \frac{ds}{c(s)} = \frac{|\mathbf{x}_{k+1} - \mathbf{x}_k|}{2} \left(\frac{1}{c_k} + \frac{1}{c_{k+1}} \right)$$

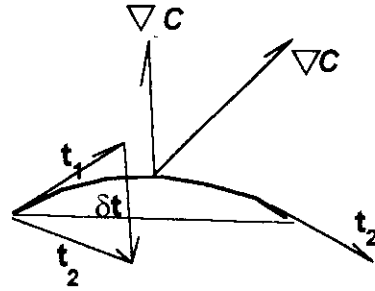
The most general approach is to minimize the functional (15) directly, using a proper computational scheme for minimization, e.g. the conjugate gradient method or the method of steepest descent. To simplify the search the properties of the ray resulting from the ray equation can be used.

Since $\mathbf{p} = \frac{\mathbf{t}}{c}$ we obtain from the ray equation that

$$\frac{d\mathbf{p}}{ds} = \mathbf{t} \frac{d}{ds} \left(\frac{1}{c} \right) + \frac{1}{c} \frac{d\mathbf{t}}{ds} = \nabla \left(\frac{1}{c} \right) \Rightarrow \frac{d\mathbf{t}}{ds} = c \nabla \left(\frac{1}{c} \right) - c \mathbf{t} \frac{d}{ds} \left(\frac{1}{c} \right) = c \nabla_{\perp} \left(\frac{1}{c} \right)$$

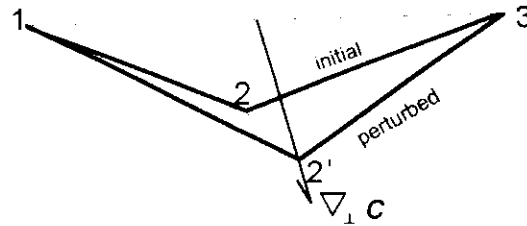
This means that the unit vector tangent to the ray is changing in the direction of the component of the velocity gradient perpendicular to the ray.

Thus, the ray is bending in the direction parallel to $\nabla_{\perp} c$. Therefore is easy to find in which direction the ray should be bend.



If in the initial ray consists of rectilinear segments, the successive end points of the segments could be changed using three-point perturbation scheme:

The point 2 is moved to a new position 2' along the gradient so that the time along the segments 12' + 2'3 becomes minimum. Then the same procedure is repeated for the segments 2'34, etc.



Drawbacks of this approach:

- it is time consuming;
- there is no guarantee that this procedure leads to the global minimum, especially in complex structures.

2.3. Calculation of the time field

This method is based on the Fermat's principle and on the Huygens' principle.

The method consists of two steps:

- preliminary calculation of the *time field*, i.e. travel times in each point, which are minimum of the times corresponding to all paths from a given source;
- calculation of rays from the receivers proceeding from the time field: the points on the ray are chosen so as to minimize the travel time.

If the times on the wave fronts are t_i and it is necessary to calculate the time in the point O , it is clear that it should be

$$t_3 + \min \{ \text{times from } O \text{ to the front } t_3 \}$$

It is evident that it is

$$t_3 + T(OA)$$

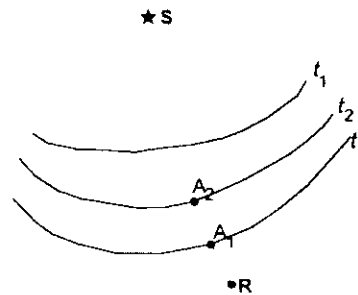
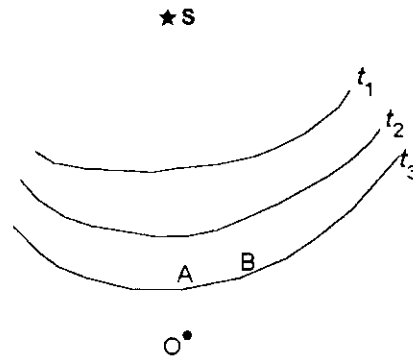
Thus we can calculate the times in all points of the medium.

The same principle is used for calculation of the rays between two given points: a source and a receiver.

If the time field from a given source has already been calculated, we begin to calculate a ray from the given receiver R . At first we find the point A_1 satisfying the condition

$$T(RA_1) + t_3 = \min ,$$

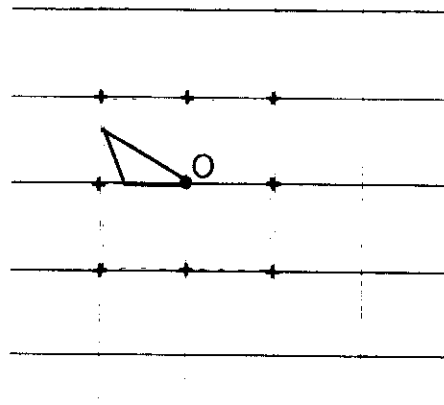
then find a point A_2 , etc.



This method is convenient when several receivers correspond to one and the same source: in this case it is sufficient to calculate the time field only once, and it can be used for all receivers.

The scheme for calculation of the time field in 2D case when the velocity distribution is approximated by constant values in rectangular cells is outlined below.

The times are calculated in the nodes, a method for interpolation between the neighbour nodes (along the cell's boundaries) being assumed.



To calculate the time in the point O , we must know the times in all points around as well as on the boundary indicated by dash line. A time in O is calculated as the time on the contour + travel time from this point to O along one of two possible paths. All points along the contour are considered, and the minimum time is chosen. Usually, when we begin to calculate the time in O , we know the times not in all the points around. In this case the point O remains as an 'open' point, and the calculations are repeated; the times in other points are calculated, and are used for calculation of the time in the point O . If the time in O calculated by the use of some other points on the contour becomes less than that calculated at the previous step, it replaces the previous time. This procedure is repeated until the time in O cannot be decreased. Then this point is accepted as a 'closed' one, and the time is assumed as the 'true' time. The calculations are finished when all points (nodes) belong to the closed class.

This method is convenient in the tomographic problems, if the velocity structure is approximated by constant values in the rectangular cells. It can be easily put into practice for 2D case, but much more difficult in 3D case.

3. Principles of the ray tomography

3.1. Formulation of the problem

The ray tomography is one of the inverse problems of seismology. Its specific character is that the initial data are functionals over lines (rays) in the space. But the rays do not cover the space entirely, so that the data seem to contain no information on the parts not crossed by the rays. However it is not quite. The rays are infinitely narrow lines only in case of infinitely high frequencies. But since the wavelengths are always finite, the rays may be regarded as if thick, the thickness being dependent on frequency: it is estimated by a dimension of the Fresnel zone. Therefore if a number of the rays is sufficiently large, the "thickened" rays may cover the whole space.

However the data in ray tomography – travel times, or attenuation factor, are usually expressed in a form of the functional along the *lines*:

$$t_i = \int_{L_i} \frac{ds}{V(\mathbf{r})} \quad \text{или} \quad t_i^* = \int_{L_i} \frac{ds}{Q(\mathbf{r})V(\mathbf{r})} \quad (16)$$

For solving the tomography problem it is convenient to express them in a form of the functionals over a spatial area Ω :

$$t_i = \int_{\Omega} G_i(\mathbf{r}) V^{-1}(\mathbf{r}) d\mathbf{r} \quad (17)$$

where $G_i(\mathbf{r})$ is singular on the ray L_i and equal to zero elsewhere. Extending this definition to the finite wavelengths, we may assume $G_i(\mathbf{r})$ to be non-zero not only on an infinitely thin ray, but within a finite part of the space. It achieves maximum on the ray, noticeably differs from zero within the Fresnel zone, and practically vanishes outside. Let us use local orthogonal coordinates related to the ray so that s specifies a point on the ray, and ξ_1, ξ_2 are orthogonal to the ray. If a size of the Fresnel zone is not too large (which is equivalent to validity of the ray theory), (17) may be written as

$$t_i = \int_{\Omega} G_i(s, \xi_1, \xi_2) V^{-1} ds d\xi_1 d\xi_2 = \int_{L_i} \bar{V}^{-1}(s) ds \quad (18)$$

where

$$\bar{V}^{-1}(s) = \iint V^{-1}(\mathbf{r}) G_i(s, \xi_1, \xi_2) d\xi_1 d\xi_2, \quad (19)$$

is an averaged velocity value in the vicinity of the ray.

Consequently, the functional (18) may be represented in the form (17), where $G_i(\mathbf{r})$, as before, is singular on the ray, though V^{-1} is not local slowness, but averaged over a spatial area in accordance with (19).

So we should keep in mind that because of finite size of the Fresnel zone the data are related not to local velocity values, but to averaged over a certain region. Therefore it follows that the unknown function should be sufficiently smooth: at least it is senseless to seek for spatial heterogeneity of size less than the Fresnel zone.

3.2. Linearization of the problem

Since the shape of ray L_i in (16) depends on the unknown velocity distribution, the tomography problem is non-linear. To linearize the problem we shall determine a small correction to a starting velocity distribution $V_0(\mathbf{r})$ rather than the velocity $V(\mathbf{r})$, properly:

$$\delta V(\mathbf{r}) = V(\mathbf{r}) - V_0(\mathbf{r})$$

For the starting model $V_0(\mathbf{r})$ we can calculate travel times between sources and receivers

$$t_{0i} = \int_{L_{0i}} \frac{ds}{V_0(\mathbf{r})} \quad (20)$$

A difference between (20) and (16)

$$\delta t_i = t_i - t_{0i} = \int_{L_i} \frac{ds}{V(\mathbf{r})} - \int_{L_{0i}} \frac{ds}{V_0(\mathbf{r})}$$

may be approximately written as the integral along the *starting ray* L_{0i} , because a contribution to δt_i due to variation of the ray is of the second order:

$$\delta t_i \cong \int_{L_{0i}} (V^{-1} - V_0^{-1}) ds = \int_{\Omega} G_i(\mathbf{r}) m(\mathbf{r}) d\mathbf{r} \quad (21)$$

where

$$m(\mathbf{r}) = (V^{-1} - V_0^{-1}) V_0$$

and $G_i(\mathbf{r})$ is a function singular on the ray L_{0i} , independent on the unknown velocity distribution and satisfying the condition

$$\int_{\Omega} G_i(\mathbf{r}) d\mathbf{r} = t_{0i} \quad (22)$$

This statement can be proved proceeding from the eikonal equation:

$$\begin{aligned} (\nabla \tau_0)^2 &= \frac{1}{V_0^2} \\ (\nabla(\tau_0 + \delta\tau))^2 &= \left(\frac{1}{V_0} + \delta \left(\frac{1}{V} \right) \right)^2 \\ 2(\nabla \tau_0, \nabla \delta\tau) &\approx 2 \frac{1}{V_0} \delta \left(\frac{1}{V} \right) \end{aligned}$$

or

$$\frac{d\delta\tau}{ds} = \delta \left(\frac{1}{V} \right)$$

Thus the linearized problem of seismic time delay tomography is formulated as follows:

Given are the values of time delays δt_i for different rays $i=1,2,...,N$, which are related to the unknown function $m(\mathbf{r})$ with the use of the functional (21), whose kernel satisfies the condition (22)

It is required to estimate $m(\mathbf{r})$.

The initial non-linear problem (17) may be solved by iterative procedure, the problem being linearized at each step.

3.3. Main approaches for solving the tomography problems

In general any problem of the ray tomography may be formulated as follows:

Known are the data γ_i ($i=1,2,\dots,N$) (time delays, values of t_i^* , polarization anomalies) related linearly with the unknown *model* (velocity corrections, attenuation coefficients) $m(\mathbf{r})$ with the functionals

$$\gamma_i = \int_{\Omega} G_i(\mathbf{r})m(\mathbf{r})d\mathbf{r} \quad (23)$$

The spatial region Ω may be 2D or 3D. The data γ_i may contain the observational errors described by a covariance matrix $\mathbf{R}_\gamma$. The number of the data is finite, and the coverage of the region Ω by the rays is not dense and non-uniform. The latter is specific for seismic tomography unlike the medical tomography.

It is evident that a solution of such problem is not unique. Therefore some *a priori* assumptions about $m(\mathbf{r})$ should be added. The approaches for solving the tomography problem are distinguished by a character of these assumptions.

A general approach for constructing the solution of (23) is to represent the unknown function as a series in some basis functions $\psi_j(\mathbf{r})$:

$$m(\mathbf{r}) = \sum_{j=1}^M \lambda_j \psi_j(\mathbf{r}) \quad (24)$$

Then the problem (23) is reduced to a system of linear equations respectively to the coefficients λ_j . If the data are exact, the system is obtained by substitution of (24) into (23):

$$\mathbf{S}\boldsymbol{\lambda}=\boldsymbol{\gamma}, \quad (25)$$

where $\mathbf{S}$ is the matrix with elements

$$S_{ij} = \int_{\Omega} G_i(\mathbf{r})\psi_j(\mathbf{r})d\mathbf{r},$$

$\boldsymbol{\gamma}^T=(\gamma_1, \gamma_2, \dots, \gamma_N)$ – vector of the data, and $\boldsymbol{\lambda}^T=(\lambda_1, \lambda_2, \dots, \lambda_M)$ – vector of the unknown coefficients.

If the data contain errors, the system is obtained from minimization of the functional

$$\sum_i \left(\gamma_i - \int_{\Omega} G_i(\mathbf{r})m(\mathbf{r})d\mathbf{r} \right) R_{ii}^{-1} \left(\gamma_i - \int_{\Omega} G_i(\mathbf{r})m(\mathbf{r})d\mathbf{r} \right) \quad (26)$$

where R_{ii}^{-1} is element of the matrix $\mathbf{R}_\gamma^{-1}$. In matrix notation this system is written as:

$$\mathbf{S}^T \mathbf{R}_\gamma^{-1} \mathbf{S} \boldsymbol{\lambda} = \mathbf{S}^T \mathbf{R}_\gamma^{-1} \boldsymbol{\gamma} \quad (27)$$

A choice of the basis functions depends on the choice of the *a priori* assumptions. The *a priori* assumptions also result in the constraints to the solution of (25) or (27) in case when a solution of the system is not unique.

The basis functions may be chosen either *a priori*, or proceeding from some requirements to the solution.

The widespread approach for a choice of the basis functions is representation of the region by blocks, the velocity corrections being assumed constant in the blocks. In this case

$$\psi_j(\mathbf{r}) = \begin{cases} 1 & \text{if } \mathbf{r} \in j\text{-th block} \\ 0 & \text{otherwise} \end{cases}$$

Then $S_{ij} = t_{ij}$, t_{ij} – is travel time along the i -th ray across the j -th block.

But in this case a question arises: what is the optimum number of blocks? If the number of blocks is much less than the number of rays, the least square method may be applied to obtain a solution. But in this case we cannot get a reasonable resolution.

Therefore a number of blocks should be assumed to be large enough, but in this case we cannot get a unique solution. Therefore some additional assumptions must be involved. Most of the algorithms of such kind are based (explicitly or implicitly) on the assumption that the norm of the solution must be minimum. Such is ACH-method, as well as some of ART-algorithms. An advantage of this approach is that for solving the large linear systems we may apply some iterative algorithms of the linear algebra, which allow not to keep the whole system in the computer memory. A disadvantage is that it is difficult to bring in correspondence the pattern of blocks with the data set.

Another approach is to approximate the unknown $m(\mathbf{r})$ by a fixed function with some unknown parameters. For example, in plane case it is convenient to represent the unknown function as a product of the orthogonal polynomials, in spherical case as a series in spherical functions. The unknown parameters are the coefficients in the series.

An alternative approach for choosing the basis functions is to construct them proceeding from some requirements to the solution. Tarantola and Nersessian proposed to apply the Bayesian approach, i.e to assume an *a priori* covariance function of the model. The simplest *a priori* covariance is the Gaussian function:

$$C(\mathbf{r}, \mathbf{r}') = \sigma^2 \exp\left(-\frac{|\mathbf{r} - \mathbf{r}'|^2}{2L^2}\right)$$

where L is *a priori* correlation radius, and σ is a mean amplitude of inhomogeneities. Then $m(\mathbf{r})$ should satisfy the condition:

$$\int_{\Omega} \int_{\Omega} m(\mathbf{r}) C^{-1}(\mathbf{r}, \mathbf{r}') m(\mathbf{r}') d\mathbf{r} d\mathbf{r}' + (\gamma - \mathbf{Gm})^T \mathbf{R}_\gamma^{-1} (\gamma - \mathbf{Gm}) = \min \quad (28)$$

where

$$(\mathbf{Gm})_i = \int_{\Omega} G_i(\mathbf{r}) m(\mathbf{r}) d\mathbf{r}$$

It follows from (28) that the basis functions are

$$\psi_j(\mathbf{r}) = \int_{\Omega} C(\mathbf{r}, \mathbf{r}') G_j(\mathbf{r}') d\mathbf{r}', \quad (29)$$

and the coefficients by these functions are determined from the equations

$$(\mathbf{R}_\gamma + \mathbf{GCG}^T) \lambda = \gamma, \quad (30)$$

where the following notation is used.

$$(\mathbf{GCG}^T)_{ij} = \iint G_i(\mathbf{r})C(\mathbf{r},\mathbf{r}')G_j(\mathbf{r}')d\mathbf{r}d\mathbf{r}' \quad (31)$$

In matrix notation a solution of (28) may be written in the form:

$$\mathbf{m} = \mathbf{CG}^T(\mathbf{R}_\gamma + \mathbf{GCG}^T)^{-1}\gamma \quad (32)$$

The function $m(\mathbf{r})$ is represented in (32) as a vector $\mathbf{m}$ in Hilbert space. This formula is also valid in the case, when $\mathbf{m}$ is a vector in finite-dimensional space, i.e. is a set of velocity corrections in blocks, and the elements G_{ij} of the matrix $\mathbf{G}$ are travel times of the i -th ray in j -th block. In this case $\mathbf{C}$ is *a priori* covariance matrix of the parameters (velocity corrections in blocks). Since division of the area into the blocks is usually made under the assumption that the velocities in the neighbouring blocks are not correlated, we may take $\mathbf{C} = \sigma_m^2 \mathbf{I}$, where $\mathbf{I}$ is unit matrix. Also, if we assume that all the data are not correlated and have equal precision, so that $\mathbf{R}_\gamma = \sigma_\gamma^2 \mathbf{I}$, then (32) may be written in the form:

$$\mathbf{m} = \mathbf{G}^T(\alpha \mathbf{I} + \mathbf{GG}^T)^{-1}\gamma \quad (33)$$

where $\alpha = \sigma_\gamma^2 / \sigma_m^2$ is the regularization parameter. Since usually the value of σ_m is unknown, the regularization parameter is selected arbitrarily. The resolution and standard error of the solution depend on the choice of this parameter.

Another method for constructing the basis functions based on a smoothness criterion was proposed by Ditmar and Yanovskaya for 2D case. If the data are exact, this criterion is following:

$$\int |\nabla m|^2 d\mathbf{r} = \min \quad (34a)$$

$$m = \text{const} \quad \text{при } |\mathbf{r}| \rightarrow \infty \quad (34b)$$

Condition (34a) ensures continuity of $m(\mathbf{r})$, though the derivatives of this function are discontinuous on the rays. The basis function are

$$\psi_i(\mathbf{r}) = \begin{cases} \int \ln|\mathbf{r} - \mathbf{r}'| G_i(\mathbf{r}') d\mathbf{r}' & i = 1, 2, \dots, N \\ 1 & i = N + 1 \end{cases}$$

The velocity obtained outside the rays may be interpreted as interpolated between the values on the rays. Higher rate of smoothness may be obtained by minimization of the functional

$$\int \underbrace{\|\nabla \nabla \dots \nabla m\|}_n^2 d\mathbf{r} \quad (35)$$

Minimization of the functional (35), or (34) as a particular case, is analogous to the spline approximation of a function determined by the values in discrete points. The condition (34) is analogous to the piece-linear approximation.

It should be noted that the condition (34) is valid only for 2D case: if we use it for 3D case, we obtain the solution concentrated on the rays and equal to zero outside the rays. In 3D case we must assume at least

$$\int_{\Omega} \|\nabla \nabla m\|^2 d\mathbf{r} = \min \quad (36)$$

(This is a phenomenon of dimension: in 1D case it is sufficient to minimize a norm of the solution: $\int_{\Omega} |m(\mathbf{r})|^2 d\mathbf{r} = \min$).

However, reduction of the initial tomography problem to determination of a finite number of parameters cannot ensure uniqueness and stability of the solution. For example, in case of the block approximation the rays may not cross some blocks, or one ray crosses several blocks. In such case it is necessary to impose some additional constraints to the solution. In some cases such constraints are imposed implicitly, by a choice of a method for solving the linear system. For example, if the SVD method is applied, the solution for the parameters fits to minimum norm:

$$\sum_i \lambda_i^2 = \min \quad (37)$$

However, if the number of blocks is much more than the number of rays this condition results in zero values in the blocks not crossed by the rays. To overcome this drawback we may combine two approaches: to divide the model to blocks, and to use a criterion analogous to (34) instead of (37) to get a 'smooth' solution. Whereas a solution for the block approximation is determined assuming $\sum_i m_i^2$ to be minimum, where m_i are the values in the

blocks, to determine a 'smooth' solution we should assume

$$\sum_{k,i} \xi_{ik} (m_k - m_i)^2 = \min$$

where

$$\xi_{ik} = \begin{cases} 1 & \text{if } i\text{-th and } k\text{-th blocks are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

Unlike the methods based on the minimum norm, for which it is important to divide the model into the optimum number of the blocks, this method is free of this limitation: the number of the blocks may be taken sufficiently large - much larger than the number of rays. This is valid both for 2D and 3D cases, though for 3D case it is more correct to use the criterion analogous to (36).

4. Resolving power

4.1. Resolution matrix

As in case of any inverse problems, where the number of unknowns is larger than the number of data, a solution of the tomography problem corresponds not to *local* values, but to averaged over an area, dimension of the area depending on the data set. Averaging of the solution takes place implicitly if the model is divided to blocks: the velocity value in a block may be regarded as an averaged over all points of the block. If the number of the blocks is too large, and the system of equation for the parameters is overdetermined, a value in each block is an average over some adjacent blocks. If the solution is determined as a continuous function, the velocity values in each point may be regarded as a result of interpolation, or as averaged over a more or less extended area. To interpret the solution it is always important to know a rate of averaging.

In the theory of linear inverse problems the resolving power of the data is estimated by a *resolution matrix* or by an *averaging kernel*. If a solution is determined by a linear operator $\mathbf{H}$, i.e. if a solution of the problem $\mathbf{Gm}=\gamma$ is $\hat{\mathbf{m}}=\mathbf{H}\gamma$, then $\hat{\mathbf{m}}=\mathbf{H}\mathbf{Gm}=\mathbf{Rm}$, where $\mathbf{R}$ is the resolution matrix.

A row of $\mathbf{R}$ may be interpreted as weights which average a solution for j -th block m_j over all other blocks:

$$m_j = \sum_k R_{jk} m_k$$

The closer R_{jj} to δ_{jj} , the better is the resolution in the j -th block.

If the operator $\mathbf{H}$ may be calculated explicitly, the resolution matrix is easily determined by multiplication: $\mathbf{R}=\mathbf{H}\mathbf{G}$. However, if the solution is determined by an iterative procedure, the operator $\mathbf{H}$ is implicit. But if the resolution matrix is symmetric, we may calculate the columns of $\mathbf{R}$ using the same iterative procedure and substituting the corresponding column of the matrix $\mathbf{G}$ instead of γ . In the case when the solution is determined from the equation (33), so that the operator $\mathbf{H}$ is $\mathbf{G}^T(\alpha\mathbf{I} + \mathbf{G}\mathbf{G}^T)^{-1}$, the resolution matrix is

$$\mathbf{R} = \mathbf{G}^T(\alpha\mathbf{I} + \mathbf{G}\mathbf{G}^T)^{-1}\mathbf{G}$$

It is easy to show that in this matrix is symmetric:

$$\begin{aligned} \mathbf{R} &= \mathbf{G}^T(\alpha\mathbf{I} + \mathbf{G}\mathbf{G}^T)^{-1}\mathbf{G} = (\alpha\mathbf{I} + \mathbf{G}^T\mathbf{G})^{-1}(\alpha\mathbf{I} + \mathbf{G}^T\mathbf{G})\mathbf{G}^T(\alpha\mathbf{I} + \mathbf{G}\mathbf{G}^T)^{-1}\mathbf{G} = \\ &= (\alpha\mathbf{I} + \mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{G} = \mathbf{I} - \alpha(\alpha\mathbf{I} + \mathbf{G}^T\mathbf{G})^{-1} \end{aligned}$$

This formula also shows that the smaller the regularization parameter, the better is the resolution (the closer is the matrix $\mathbf{R}$ to the unit matrix $\mathbf{I}$).

The columns (=row) of the matrix $\mathbf{R}$ can be determined by the use of the same iterative procedure, as is used for determining the solution, replacing the data γ by the corresponding column of the matrix $\mathbf{G}$.

A row of the resolution matrix describes a resolution in one separate point (or block). To draw conclusions about the resolution in all blocks we must calculate the total matrix rather than a single row, but this procedure requires is time consuming. Also it is difficult to interpret the result: each block (i) is specified by the values R_{ij} , i.e. $F_i(x_{1j}, x_{2j}, x_{3j})$ so we must analyze a lot of such matrices.

4.2. Estimation of resolution by mean size of the averaging area

To represent the resolution more clear, it would be desirable to estimate the resolution in each point by a single number rather than by a matrix $F_i(x_{1j}, x_{2j}, x_{3j})$. In classical Backus-Gilbert method for 1D inverse problems such characteristics of the resolution is the so-called 'averaging length'. In 2D or 3D tomography problems a 2D or 3D averaging area should replace the averaging length. In 2D tomography problem in case when the solution is constructed from the smoothness criterion (34), the real averaging area is replaced by an equivalent circular area, and its radius is accepted as a characteristics of the resolution in the center of this area.

This approach is not suitable in cases when coverage of the region by rays is not uniform in all directions. Usually we have a predominant direction of the rays because of concentration of the sources (and even stations) in some parts of the region under investigation. For this case we may assume the equivalent averaging area of elliptic (or ellipsoidal) shape, and estimate the main axes of the ellipsoid. The ellipsoid (ellipsoid) should be stretched in the predominant direction of the rays (the resolution in this direction is the poorest).

4.3. Checkerboard test

At present time to estimate the resolution widely used is the so-called *checkerboard test*. Its goal is to conclude about a possibility to resolve the velocity anomalies of a given size by the use of a given data set. For this purpose a synthetic model is constructed in a form of a checkerboard: positive and negative anomalies alternate with each other in chess-board fashion with the amplitudes of the order approximately the same as expected in the real medium. For such model travel time residuals are calculated along the same rays which are presented in the given data set. Random errors are added to the residuals, and such synthetic data are used as the input for the tomography problem. In the parts of the region, where the assumed anomalies are revealed in the solution, the blocks of the same size may be resolved. Sometimes synthetic models contain the anomalies of the same configuration as expected in the real structure.

A drawback of this approach is that it allows only to answer the question: can we resolve the anomalies of a *given* size, or not? Of course it is more important to estimate the sizes of the anomalies in different parts of the region, which are different. For such purpose we have to repeat the checkerboard test several times with different sizes of the blocks, though it is rather time consuming.

SCIENTIFIC COMMUNICATIONS

Resolution Estimation in the Problems of Seismic Ray Tomography

T. B. Yanovskaya

Saint Petersburg State University

Received November 13, 1996

The solution of the seismic tomography problem is nonunique, because initial data (functionals along rays) do not constrain seismic velocities at all points of a medium. For this reason, the data resolvability estimation is vital to seismic tomography. The knowledge of resolvability allows one to reliably estimate the inhomogeneity size adequate to a given data sample and, most importantly, decide, whether or not solution properties are artifacts specific to the method of solution.

In the application of discretization methods, an obvious procedure based on the calculation of the resolvability matrix is awkward, requiring an enormous amount of computations, which is as many times larger than that needed for the construction of the solution as the number of blocks comprising the medium. The graphical (particularly 3D) representation of resolvability is also cumbersome in this case.

Some authors estimate the resolvability of data in terms of specific synthetic models based on a fixed set of rays. Such a model commonly consists of checkerboarded anomalous zones of higher and lower velocity [1, 2]. In those areas where such zones are resolved as a result of solution, the data set in use ensures the resolution of inhomogeneities of the given size. In some cases, the synthetic model is similar to the actually employed one [3].

The drawbacks of this approach are obvious: due to the irregular ray coverage, the resolution is not the same in various parts of an area studied, and one should either resort to the above procedure involving model anomalous zones of various sizes and constraining the size and position of resolvable inhomogeneities or restrict oneself to a single model (as is commonly done) constraining the possible position of inhomogeneities of a fixed size.

An alternative approach to the estimation of data resolvability in the 2D tomography problem [4] generalizes the method proposed by Backus and Gilbert [5] for the 1D data inversion. This approach is applicable, if the solution is sought for as an expansion in continuous basis functions under certain assumptions on the smoothness of the solution [6, 7]. In this case, the resolution at each point of the area is determined by the smoothing kernel, whereas the kernel concentration in the vicinity of a given point is defined as the linear

dimension of the area, in which the kernel is essentially nonzero. This dimension is determined by equating the values of a preset concentration criterion applied to the kernel under consideration with a reference kernel constant within a circle with the center at a given point. The radius of this circle is adopted as the estimate of the smoothing area size.

However, a disadvantage of this procedure is that it is not capable of estimating the direction dependence of resolvability. In fact, if there is a preferred direction of rays (which is nearly always the case with fixed stations and sources in one region), the resolvability along this direction should be worse than across it.

This work offers a method for estimating the azimuth-dependent resolvability. By way of illustration, the method is applied to a 2D problem, but it is easily extended to the 3D case.

FORMULATION OF THE PROBLEM

Let a plane 2D domain Σ be covered with rays. Travel time residuals δt_i are the initial data. The solution of the tomography problem at a certain point with the coordinates $0 < x < X$, $0 < y < Y$ (or in a block containing the given point, given the blockwise approximation of the medium) is represented as a linear combination of initial data, regardless of the solution method:

$$m(x_0, y_0) = \sum_i a_i(x_0, y_0) \delta t_i, \quad (1)$$

where $m(x_0, y_0)$ is the relative correction to the velocity $V_0(x_0, y_0)$ chosen as an initial approximation,

$$m(x_0, y_0) = \frac{\delta V(x_0, y_0)}{V_0(x_0, y_0)}.$$

The coefficients $a_i(x_0, y_0)$ are assumed to be expressed in an explicit form.

The smoothing kernel at a given point is defined as

$$A(x, y; x_0, y_0) = \sum_i a_i(x_0, y_0) G_i(x, y), \quad (2)$$

where $G_i(x, y)$ are data kernels singular on rays and subject to the condition

$$\iint_{\Sigma} G_i(x, y) dx dy = t_{0i} \quad (3)$$

and t_{0i} is the i th-ray travel time in the model of initial approximation.

The sought-for model and smoothing kernel represented by (2) and (3) are assumed here to be continuous functions of coordinates, but the generalization to the case of blockwise approximation of the medium is immediate: the smoothing kernel should be replaced by the resolution matrix

$$R_{ij} = \sum_k a_{ik} t_{kj},$$

where t_{kj} is the j th-ray travel time in a k th block, and a_{ik} are the matrix elements relating the vector of parameters $\mathbf{m}$ with the data vector $\delta \mathbf{t}$.

ESTIMATION PROCEDURE OF THE DATA RESOLVABILITY

Now we define the criterion for estimating the resolvability at a point (x_0, y_0) along a fixed direction. Without loss of generality, the latter may be assumed to coincide with the x axis.

Integrating (2) in the y direction over the interval $(0, Y)$ corresponding to the area covered with rays yields a function of x and point (x_0, y_0) :

$$F(x; x_0, y_0) = \sum_i a_i(x_0, y_0) \int_0^Y G_i(x, y) dy. \quad (4)$$

This function measures the smoothness of the solution in the x direction, and its divergence with the delta function $\delta(x - x_0)$ may be taken as the resolvability estimate along this direction. However, both functions being singular, this divergence should be estimated as the difference between the integral

$$\Phi(x; x_0, y_0) = \int_0^x F(x'; x_0, y_0) dx' \quad (5)$$

and Heaviside function $\varepsilon(x - x_0)$ [5]. The substitution of (4) into (5) yields

$$\Phi(x; x_0, y_0) = \sum_i a_i(x_0, y_0) \tau_i(x), \quad (6)$$

where $\tau_i(x)$ is the travel time along the i th ray to the point with coordinate x . For the ray with the end coordinates x_i^1 and x_i^2 , this function is plotted in Fig. 1.

The divergence between (6) and Heaviside function is naturally expressed by the squared deviation

$$s(x_0, y_0) = \int_0^x (\Phi(x; x_0, y_0) - \varepsilon(x - x_0))^2 dx. \quad (7)$$

If the velocity of initial approximation is independent of coordinates, the function $\tau_i(x)$ is linear in the interval

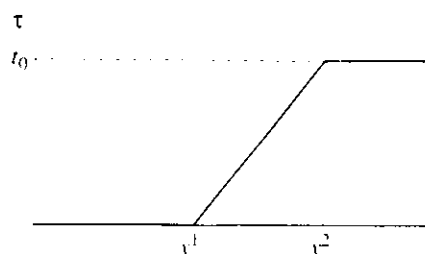


Fig. 1. The function $\tau(x)$ corresponding to the ray with the end coordinates x_i^1 and x_i^2 . The travel time on ray is t_0 .

(x_i^1, x_i^2) , and the function $\Phi(x; x_0, y_0)$ is, correspondingly, piecewise linear in the interval $(0, X)$. In this case, the expression for $s(x_0, y_0)$ can readily be written in an analytical form.

The smaller the value of s , the closer $\Phi(x; x_0, y_0)$ is to the Heaviside function and the better the resolution along the x axis. The value s provides constraints on the effective width of the smoothing interval just as this is done in the Backus–Gilbert method in the 1D problems.

Let the smoothing kernel be a nonzero constant in the rectangle $x_0 - \frac{a}{2} < x < x_0 + \frac{a}{2}$, $y_0 - \frac{b}{2} < y < y_0 + \frac{b}{2}$.

Assuming the kernel to be normalized, this constant should be equal to $1/ab$. For such a kernel, $s(x_0, y_0)$ is easily shown to be $a/12$. Thus, the value $12s$ may be taken as an estimate of the smoothing area width in the x direction.

As is mentioned above, the direction of the x axis may be chosen arbitrarily. Consequently, the smoothing area width a characterizing the data resolvability can be estimated along any direction through the proper rotation of the x axis. In other words, rotating the x axis, the azimuth dependence of resolvability, represented by $a(\vartheta)$, can be estimated at any point (x_0, y_0) .

The resolvability estimation by the smoothing area radius [4] actually is equivalent to the assumption that the linear size of this area is direction-independent and can be evaluated through the averaging of $a(\vartheta)/2$ over the whole range of ϑ . The data resolvability estimate can be further improved if $a(\vartheta)$ is approximated by the function of the type $a + b \cos(2\vartheta - 2\phi)$. In this case, the resolution is worst (the linear dimension of the smoothing area L is $a + b$) at the azimuths $\vartheta = \phi$ or $\vartheta = \phi + \pi$ and best ($L = a - b$) at $\vartheta = \phi + \pi/2$ or $\vartheta = \phi + 3\pi/2$. The resolution is nearly the same in all directions at b close to zero. A large value of b/a at some point indicates that the directions of most traces are similar and the original data are insufficient for the tomographic reconstruction.

The resolvability estimation by the minimum and maximum values of linear dimensions of the smoothing area, obtained along two mutually perpendicular directions, is equivalent to the assumption that the effective smoothing area may be represented by an ellipse with

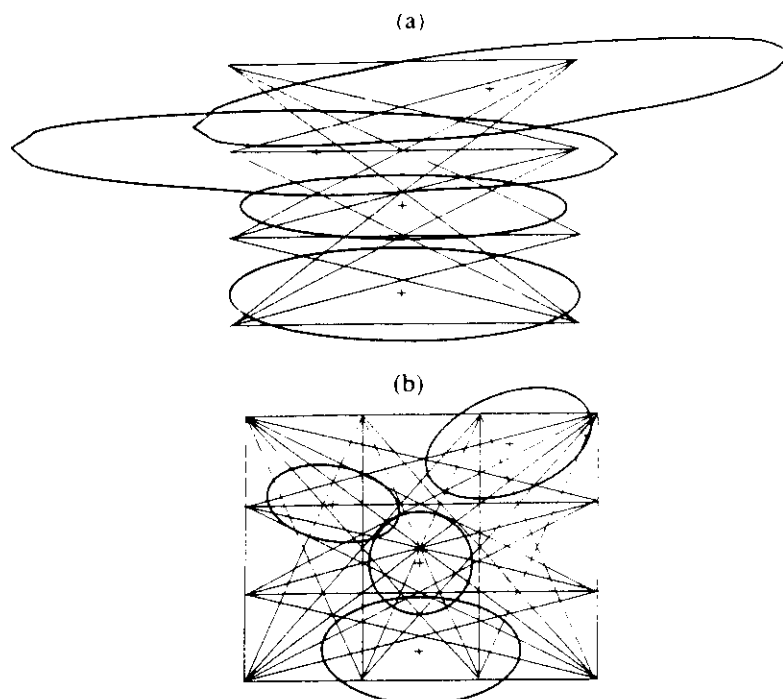


Fig. 2. Equivalent smoothing areas in the form of ellipses, centered at the points marked with the crosses. The thin lines are rays. The seismic tomography solution at these points may be considered as a result of smoothing over the respective areas.

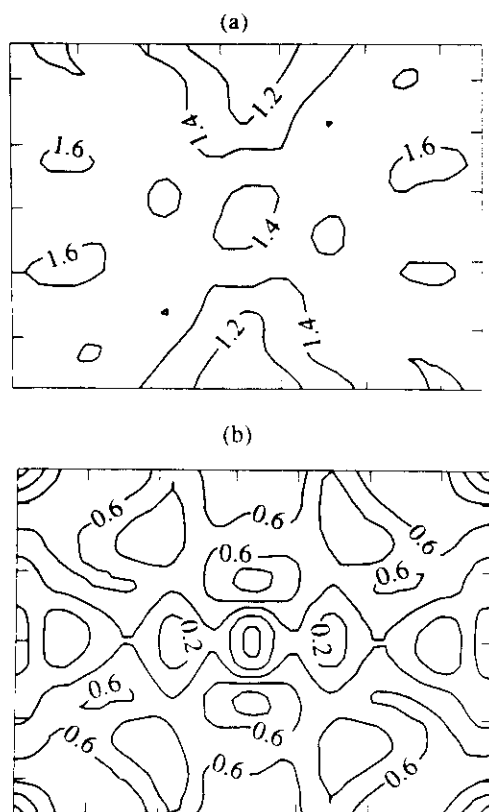


Fig. 3. Distribution of ϵ for the two ray systems shown in Fig. 2.

the axes equal to $a + b$ and $a - b$, with the azimuth of the major axis being ϕ .

A NUMERICAL MODEL TEST

The resolvability effect of path orientation is illustrated in Fig. 2, where effective smoothing areas are shown for various sets of traces. In Fig. 2a, traces are mostly oriented from right to left, and traces intersecting the area studied from top to bottom are absent. Thus, traces oriented in this way are typical of a system of the crosshole survey. The smoothing areas centered at four points marked by crosses are well elongated in a horizontal direction, so that the resolvability in this direction is nearly zero. The introduction of top-to-bottom traces (Fig. 2b) essentially changes the whole picture: now, the sizes of the smoothing areas are nearly the same in all directions. In this case, the resolvability may be estimated by the radius of the effective smoothing area, as is suggested in [4].

The elongation of the smoothing area can be estimated by the ratio of the difference between the maximum and minimum sizes of the area to its mean size $\epsilon = \frac{2b}{a}$.

Small values of ϵ imply that the traces are oriented more or less uniformly along all directions, and the resolvability at any point may be characterized by a single number, that is, the mean linear size of the effective smoothing area. On the contrary, if the ϵ values are large (at least larger than unity), the traces have a pre-

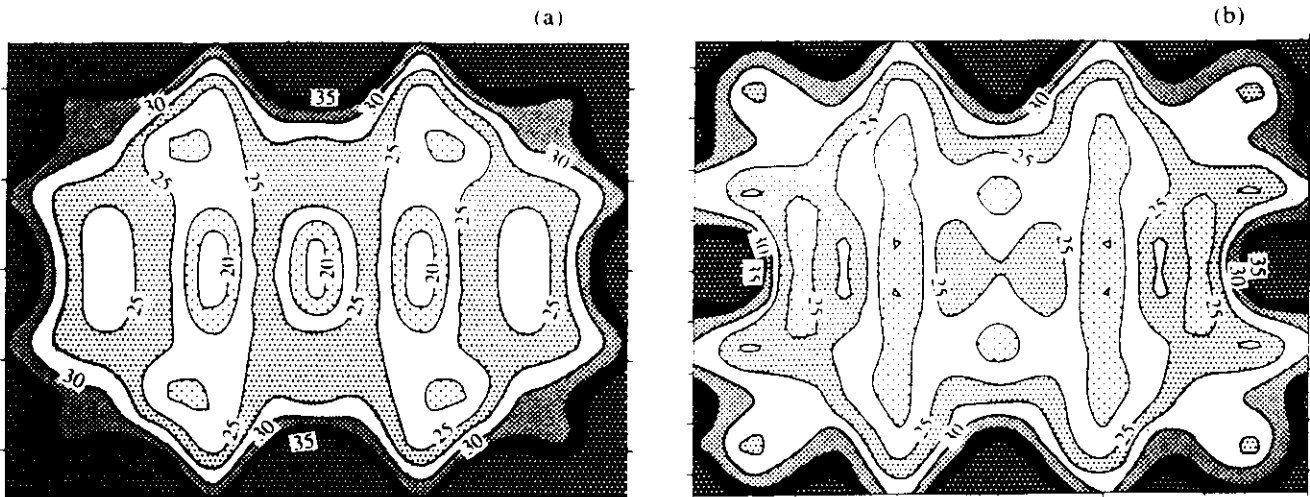


Fig. 4. Distribution of the smoothing area radius for the ray system of Fig. 1b, obtained by (a) the method offered in this work and (b) circular approximation of smoothing area [4].

ferred orientation, and the resolution in this direction is likely to be quite small.

The isolines of ϵ are shown in Figs. 3a and 3b for two ray systems considered above (see Figs. 2a and 2b). As is expected, the ϵ values average 1.4–1.6 for the first ray system characterized by weak resolution in the horizontal direction, and they are 0.2–0.6 for the second system. Correspondingly, in the second case, the resolution at any point can be characterized by the mean values of the smoothing area size ($a/2$) or “radius” ($a/4$). The distribution of this value over the area covered by the rays is shown in Fig. 4a for the second ray system. For comparison, Fig. 4b shows the distribution of the effective radius of the smoothing area, calculated by the method proposed in [4]. As might be anticipated, these distributions, differing in detail because of different approaches to their calculation, are similar on the whole and consistent with the ray configuration: in both cases, the smallest size of the smoothing area correspond to the zones most densely covered with variously oriented rays.

CONCLUSIONS

The approach proposed for estimating the resolvability of ray tomography problems using travel time data allows one to estimate the completeness of the ray system in use. The parameter ϵ may serve as a criterion of the completeness: the smallness of ϵ (~ 0.5) implies that the obtained solution is locally smoothed over an area of nearly the same size in all directions. At $\epsilon > 1$, the solution is the result of averaging over an elongated area and, in some cases, does not even provide any information additional to the original data (path-averaged velocities).

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 96-05-64091, and by the International Association for the Promotion of Cooperation with Scientists from the Independent States of the Former Soviet Union, grant no. 93-92.

REFERENCES

1. Humphreys, E. and Clayton, R.W., Adaptation of Back-Projection Tomography to Seismic Travel Time Problems, *J. Geophys. Res.*, 1988, vol. 93, pp. 1073–1085.
2. Spakman, W., van der Lee, S., and van der Hilst, R., Travel Time Tomography of the European-Mediterranean Mantle down to 1400 km, *Phys. Earth Planet. Inter.*, 1993, vol. 79, pp. 3–74.
3. Spakman, W., Stein, S., van der Hilst, R.D., and Wortel, M.J.R., Resolution Experiments for NW Pacific Subduction Zone Tomography, *Geophys. Res. Lett.*, 1989, vol. 16, pp. 1097–1100.
4. Ditmar, P.G. and Yanovskaya, T.B., Extension of the Backus–Gilbert Method to the Estimation of Lateral Variations in the Surface Wave Velocity, *Izv. Akad. Nauk SSSR, Fiz. Zemli*, 1987, no. 6, pp. 30–40.
5. Backus, G. and Gilbert, F., The Resolving Power of Gross Earth Data, *Geophys. J. R. Astron. Soc.*, 1968, vol. 16, pp. 169–205.
6. Yanovskaya, T.B. and Ditmar, P.G., Smoothness Criteria in Surface Wave Tomography, *Geophys. J. (Oxford)*, 1990, vol. 102, pp. 63–72.
7. Tarantola, A. and Nersessian, A., Three-Dimensional Inversion without Blocks, *Geophys. J. R. Astron. Soc.*, 1984, vol. 76, pp. 299–306.

SMR: 1059

**FOURTH WORKSHOP ON THREE DIMENSIONAL MODELLING
OF SEISMIC WAVES GENERATION.
PROPAGATION AND THEIR INVERSION**

28 September - 9 October 1998

*Additional pages to
T. Yanovskaya's*

lecture (SMR1059/21)

These are preliminary lecture notes, intended only for distribution to participants.

Alternatives for choice of the basis functions

1. *A priori choice of the basis functions.*

In *global* studies the model $m(\phi, \lambda)$ is usually represented as a series in spherical harmonics, so that the basis functions are just spherical harmonics.

In *regional* studies -the area under investigation is divided into some regions R_j ($j=1,2,\dots,M$), which are assumed to be laterally homogeneous on the basis of geomorphological, geotectonical or some other geophysical data.

The basis functions are defined as follows:

$$\psi_j(\mathbf{r}) = \begin{cases} 1 & \text{if } \mathbf{r} \in R_j \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

2) Bayesian approach (Tarantola & Nersessian, 1984). If the *a priori* covariance function of the model $C(\mathbf{r}, \mathbf{r}')$ is assumed, $m(\mathbf{r})$ is determined by minimizing the functional:

$$\int_S \int_S m(\mathbf{r}) C^{-1}(\mathbf{r}, \mathbf{r}') m(\mathbf{r}') d\mathbf{r} d\mathbf{r}' + (\mathbf{d} - \mathbf{Gm})^T \mathbf{R}_d^{-1} (\mathbf{d} - \mathbf{Gm}) \quad (7)$$

where $\mathbf{Gm}$ is a vector with components $\int_S G_i(\mathbf{r}) m(\mathbf{r}) d\mathbf{r}$.

The solution is $\mathbf{m} = \mathbf{CG}^T (\mathbf{R}_d + \mathbf{GCG}^T)^{-1} \mathbf{d}$, where the following notation is implied:

$$(\mathbf{GCG}^T)_{ij} = \iint G_i(\mathbf{r}) C(\mathbf{r}, \mathbf{r}') G_j(\mathbf{r}') d\mathbf{r} d\mathbf{r}'$$

The basis functions are

$$\psi_j(\mathbf{r}) = \int_S C(\mathbf{r}, \mathbf{r}') G_j(\mathbf{r}') d\mathbf{r}' \quad (8)$$

The *a priori* covariance function of the model is assumed usually as Gaussian

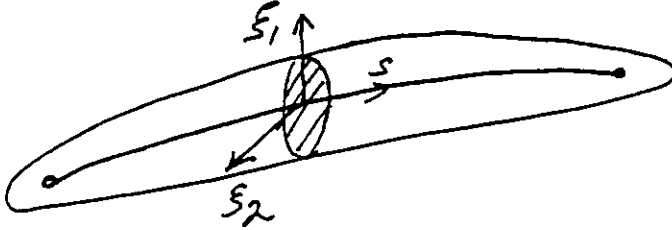
$$C(\mathbf{r}, \mathbf{r}') = \sigma^2 \exp\left(-\frac{|\mathbf{r} - \mathbf{r}'|^2}{2L^2}\right)$$

where L is a correlation length.

$$t_i = \int_{L_i} \frac{ds}{V(\mathbf{r})}, \quad L_i \text{ depends on } V(\mathbf{r})!$$

$\Downarrow$

$$t_i = \iiint_{\Omega} G_i(\mathbf{r}) V^{-1}(\mathbf{r}) d\mathbf{r} = \iiint_{\Omega} G_i(\xi_1, \xi_2, s) V^{-1}(\mathbf{r}) d\xi_1 d\xi_2 ds$$



For $\omega \rightarrow \infty$

$$G_i(\xi_1, \xi_2, s) = \delta(\xi_1) \delta(\xi_2) (H(s) - H(S_i - s)) \quad (*)$$

$$\iiint G_i(\xi_1, \xi_2, s) d\xi_1 d\xi_2 ds = S_i$$

For finite ω

$$G_i(\xi_1, \xi_2, s) \approx \frac{1}{2\pi R_F^2} \exp\left(-\frac{\xi_1^2 + \xi_2^2}{2R_F^2}\right) (H(s) - H(S_i - s))$$

Define 'average' slowness at the ray:

$$\bar{V}^{-1}(s) = \iint G_i(\xi_1, \xi_2, s) V^{-1}(\mathbf{r}) d\xi_1 d\xi_2$$

then

$$t_i = \int_{L_i} \bar{V}^{-1}(s) ds = \iiint_{\Omega} \bar{V}^{-1}(s) G_i(\xi_1, \xi_2, s) d\xi_1 d\xi_2 ds$$

where $G_i(\xi_1, \xi_2, s)$ is defined by (*)