INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

H4.SMR/650-9

Workshop on Three-Dimensional Modelling
of Seismic Waves Generation
Propagation and their Inversion

30 November - 11 December 1992

Some Problems of Seismic Tomography

T.B. Yanovskaya

St. Petersburg State University
Institute of Physics
St. Petersburg
Russian Federation

SOME PROBLEMS OF SEISMIC TOMOGRAPHY

T.B.Yanovskaya

Notes of the lectures at the Workshop
"Three-Dimensional Modelling of Seismic Wave
Generation, Propagation and their Inversion"

ITCP, 1992

Content:

- Basic principles of seismic tomography
- Resolving power of tomography data
- Joint determination of lateral velocity variations and azimuthal anisotropy
- Computational facilities

1. Basic principles of seismic tomography

1.1. Classification of tomography problems

In general, *tomography* means a reconstruction of image of an "object" from its *projections*. Projections are some functionals of unknown physical characteristics of the object. In seismic tomography "chsracteristics" are usually velocities of seismic waves, and "projections" are either travel times or waveforms obtained from observations.

Dependently on the data seismic tomography is subdivided to the *ray* and *diffractional* tomography. The ray tomography is based on the data of the waves, whose lengths are much less than dimensions of heterogeneities under investigation, so that for calculations of the wave fields the ray theory may be applied. The diffractional tomography is based on a long wavelength approximation (dimensions of heterogeneities are less than wavelength). An advantage of the diffrational tomography is the use of the data on waveforms rather than on travel times only, since the waveforms contain enlarged information about characteristics of the medium. But solution of the problems of diffrational tomography involvs some computational difficulties as well as difficulties in calculation of synthetic seismograms, especially for complex structures. Therefore the methods of the ray tomography are more developed by the present time, and these methods are widely used for determination of 3D distribution of seismic velocities in the Earth.

1.2 Formulation of the ray seismic tomography problem

The ray tomography is one of the inverse problems of seismology. Its peculiarity is that the initial data are functionals over lines (rays) in the space. But the rays do not cover the space entirely, so that such data seem to contain no information on the parts not crossed by the rays. However it is not quite so. 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 thick, their thickness being dependent on frequency: it is equal to a dimension of the Fresnel zone. Therefore if a number of the rays if sufficiently large, the "thickened" rays (let us call them *ray tubes*) may cover the whole space.

However we cannot express the "data" (travel times) as some functionals over 3D parts of the space, they are represented as the functionals over the lines:

$$t_i = \int_{L_i} \frac{ds}{V(r)} \qquad (1)$$

or in the form more suitable for solving the tomography problem:

$$t_i = \int_{\Omega} G_i(r) V^{-1}(r)\, dr \qquad (2)$$

where $G_i(r)$ is singular on L_i and equal to 0 elsewhere. Extending this definition to finite wavelengths we must assume $V(r)$ in (1) as an average value over a cross-section of the ray tube rather than a local value. Correspondingly a definition of the function $G_i(r)$ is changing: it differs from zero outside the ray tubes, and inside it may be roughly assumed to be equal to W^{-1}, where W is the area of cross-section of the ray tube.

This consideration allows on one hand to understand a difference between the approaches for solving the tomography problem, on the other hand to realize what is a meaning of the solution.

1.3. Linearization of the problem

The general formulation of the problem is given by (1) or (2): the observational data are t_i $(i=1,2,..,N)$, and the velocity function $V(r)$ has to be determined. This problem is obviously non-linear, since a shape of the ray depends on the unknown velocity distribution. But in most cases a rough approximation of the velocity distribution $V_0(r)$ (a starting model) is known from former geophysical studies. So instead of $V(r)$ we may determine a velocity correction assuming that the ray in the starting model differs slightly from the ray in the starting model. Usually instead of the velocity correction we determine the relative slowness correction:

$$m(r) = (V^{-1}(r) - V_0^{-1}(r))/V_0^{-1}(r) \qquad (3)$$

If we define travel time in the starting model as

$$t_{0i} = \int_{L_{0i}} \frac{ds}{V_0(r)} \qquad (4)$$

then the problem is linearized:

$$\delta t_i = \int_{L_{0i}} m(r) \frac{ds}{V_0(r)} \qquad (5a)$$

or

$$\delta t_i = \int_{\Omega} G_i(r) m(r)\, dr \qquad (5b)$$

where $G_i(r)$ does not depend on the unknown velocity function $V(r)$.

1.4. The approaches for solving the problem.

Since a solution of (5) is not unique, some *a priori* assumptions e to be made about the solution. Differences between the isting algorithms for solving the problem ultimately follow from fferences in the *a priori* assumptions. The assumptions may be volved into the algor thm in explicit or implicit form.

By the present time many different algorithms have been veloped for solving the tomography problem. This follows from ie fac. that any algorithm must satisfy different requirements. ie main requirements are following:
- the *a priori* assumptions must be well founded from physical int of view;
- a solution has not to be affected strongly by the values of the rameters of the algorithm, which are arbirtary to some exent;
- the algorithm must be as fast as possible and not require too ich computer memory.

In the most of algorithms the medium is divided into a number of lls, m(r) being assumed constant (m_i) in each cell. It is vident that in this case the problem is reduced to solving a inear system of equations. But in this case arises a question: iat is the optimum number of the cells? If the number of cells is ich less than the number of rays, the least square method may be plied to obtain a solution. But in this case we cannot get a asonable resolution.

Therefore a number of cells should be assumed to be large ough, but in this case we cannot get a unique solution, so some ditional assumtions must be involved. Most of the algorithms of ich kind are based (explicitely or implicitely) on the assumption iat the norm of the solution must be minimum. Such is ACH-method s well as some of ART-algorithms. An advantage of this approach s that for solving the large linear systems we may apply some terative algorithms of linear algebla, which allow not to keep in ie computer memory the whole system. A disadvantage is that it is ifficult to bring in correspondence the pattern of cells with the ata set.

To avoid this difficulty some authors proposed the algorithms, n which the medium is not divided into cells: a solution is pposed to be a continious function of coordinates, but some ditional *a priori* constrains are imposed to the unknown nction. In the algorithm proposed by Tarantola & Nersessian (984) the Bayesian approach is applied, and an *a priori* ovariance function of the model is assumed. Yanovskaya & Ditmar (990) proposed the algorithm based on the assumption of oothness of the unknown function, which is expressed as inimization of the functional

$$\int \|\nabla m(r)\|^2 dr \qquad (6)$$

for 2D case under the constrains (5a). For 3D case the functional

$$\int \|\nabla\nabla m(r)\|^2 dr \qquad (7)$$

has to be minimized.

For the 2D case the solution takes form:

$$m(r) = \sum_i a_i H_i(r) + C \qquad (8)$$

where

$$H_i(r) = \int_{L_i} g(r,r') \frac{ds'}{V_0(r')} \qquad (9)$$

and $g(r,r')$ is the Green's function of the Laplace equation

$$\Delta g(r,r') = \delta(r-r')$$

The coefficients a_i are determined after substitution of (8) into (5), so that a system of equations, number of which is equal to the number of the data, is to be solved.

The advantage of these approaches is that a resolution of the solution is in correspondence with the dataset: in the parts of the area covered densily by the rays the solution is detailed, whereas in the other parts it is more or less smoothed.

The disadvantage is that the methods of this sort require too much computer time comparatively with the former methods. The method based on the smoothness criterion is practically suitable only for 2D case if the velocity in the starting model V_0 is constant (i.e.for determination of lateral distribution of surface waves), whereas the approaches, in which the problem is reduced to solving a linear system of equations

$$\mathbb{G}m = y \qquad (10)$$

and the matrix $\mathbb{G}$ is sparse, allow the iterative algorithms to be applied, as mentioned before.

The simplest (though not widely used) algorithm is so-called *Back Projection Technique* (BPT). According to this algorithm the solution is determined by the formula:

$$m_k = \frac{\sum_{j=1}^{N} G_{kj} \frac{y_j}{\sum_i G_{ij}}}{\sum_j G_{kj}} \qquad (11)$$

This gives the exact solution of (10) in the only case, when each

cell is crossed not more than by one ray, but it is not clear to which solution this algorithm converges otherwise.

Some other iterative algorithms converge to the solution with minimum norm, i.e. they lead to the solution, which mimimize the mean square residual

$$(\mathbb{G}m - y)^T(\mathbb{G}m-y) \qquad (12)$$

and the norm $\|m\|$, or m^Tm.

In general case, when the random errors in the data are described by the covariance matrix $\mathbb{R}_y$, and an *a priori* covariance matrix of the vector of the unknowns is $\mathbb{R}_m$, it is reasonable to look for the solution which minimizes the functional

$$(\mathbb{G}m - y)^T\mathbb{R}_y^{-1}(\mathbb{G}m-y)+m^T\mathbb{R}_m^{-1}m \qquad (13)$$

If $\mathbb{R}_m=\sigma_m^2\mathbb{I}$, and $\mathbb{R}_y=\sigma_y^2\mathbb{I}$, then the functional (13) reduces to:

$$(\mathbb{G}m - y)^T(\mathbb{G}m-y)+\alpha m^Tm \qquad (13a)$$

and the solution is obtained from the linear system

$$(\mathbb{G}^T\mathbb{G}+\alpha\mathbb{I})m = \mathbb{G}^Ty \qquad (14)$$

which also may be solved with the use of one or another iterative algorithms. Here $\alpha=\sigma_y^2/\sigma_m^2$ may be regarded as a regularization parameter in Tikhonov's approach.

In the approach based on smoothness criterion in case of erroneous data the solution is determined by minimization of the functional

$$(\mathbb{G}m - \delta t)^T(\mathbb{G}m-\delta t)+ \alpha\iint|\nabla m|^2dr \qquad (15)$$

where α is again a regularization parameter. A solution for the coefficients a_i and th constant C are determined from the system

$$\begin{cases} (\mathbb{S}+\alpha\mathbb{I})a + Ct_0 = \delta t \\ a^Tt_0 = 0 \end{cases} \qquad (16)$$

where

$$S_{ij} = \iint g(r_i,r_j)\frac{ds_i}{V_0}\frac{ds_j}{V_0} \qquad (17)$$

1.3. Combination of two approaches.

It is possible to combine the two approaches so that to keep the advantages of them both. It can be done if the method based on the smoothness criterion would be applied to a discretized model. The medium is divided into cells, the number of cells being arbitrary and larger than the number of rays, and to determine the values of m_j we minimize the discrete analogue of (15). The functional (15) transfomrs into the following:

$$(\mathbb{G}m - \delta t)^T(\mathbb{G}m-\delta t) + \alpha m^T\mathbb{C}m \qquad (18)$$

where the matrix $\mathbb{C}$ is defined from the relationship:

$$m\mathbb{C}m = \frac{1}{2}\sum_i\zeta_{ik}(m_i-m_k)^2 \qquad (19)$$

where

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

The solution is obtained then from the linear system

$$(\alpha\mathbb{C} + \mathbb{G}^T\mathbb{G})m = \mathbb{G}^T\delta t \qquad (20)$$

When the solution is constructed proceeding from the minimization of the functional (13a), the system of the same form is to be solved, but in this case $\mathbb{C}=\mathbb{I}$ (unit matrix).

In 3D case the matrix $\mathbb{C}$ is determined as a discrete analogue of the matrix of second derivatives.

Returning to the system (20) we notice that the matrix $\mathbb{C}$ is sparse, the total matrix $\alpha\mathbb{C}+\mathbb{G}^T\mathbb{G}$ is positive-definite, so the solution may be obtained by the use of some iterative methods, in particular with the use of Conjugate Gradient Method (CGM). Computer time needed for calculations turns out to be much less than for the continious case.

The program developed in S-Petersburg University is based on solving the system (20) by the CGM algorithm with $\mathbb{C}=\mathbb{I}$ and $\mathbb{C}$ defined by (19). So the solutions corresponding to two different criteria can be easily compared.

Calculations made for some model examples allows to draw the following conclusions:

- When the number of cells increases, the solution obtained by the discrete analogue of the approach based on the smoothness criterion converges to the true model, whereas the method based on minimization of the norm leads to destruction of the solution.
- The former conclusion is kept even if the real model is not smooth.

2. Resolving power of tomography data

We can never obtain local values of the unknown velocity: the olution of the tomography problem is always smoothed. Not to ention that discretization implies estimation of averaged values ithin each cell, even these values may be interpreted as averaged ver some cells. If the solution is determined as a continious unction, the values, which are obtained in each point may be egarded as a result of interpolation. Therefore, to judge on eality of some features of the solution it is important to alculate not only the solution, but also a *resolution*.

In the theory of inverse problems the resolving power is haracterized by the *resolution matrix* or *averaging kernel*.
If a solution is determined by some linear operator $\mathbb{H}$ (for the ystem (20) $\mathbb{H}=(\alpha\mathbb{C}+\mathbb{G}^T\mathbb{G})^{-1}\mathbb{G}^T$), i.e. the solution $\bar{m}=\mathbb{H}\delta t$, and since $t=\mathbb{G}m$, we derive that $\bar{m} = \mathbb{H}\mathbb{G}m = \mathbb{R}m$, where $\mathbb{R}$ is the resolution atrix.

A row of the $\mathbb{R}$ may be interpreted as weights which average solution for j-th cell $\bar{m}_j$ over all other cells:

$$\bar{m}_j = \sum R_{jk} m_k$$

he closer R_{jk} to δ_{jk}, the better the resolution in the j-th cell.

If the operator $\mathbb{H}$ may be calculated explicitely, then the esolution matrix is easily determined by multiplication: $\mathbb{R}=\mathbb{H}\mathbb{G}$. owever, when the solution is determined by an iterative rocedure, the operator $\mathbb{H}$ is implicit. But if the resolution atrix is symmetric, we may calculate the columns of $\mathbb{R}$ using the ame iterative procedure and substituting the corresponding column f the matrix $\mathbb{G}$ instead of δt. In the case $\mathbb{C}=\mathbb{I}$ this condition is atisfied:

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

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

$$\mathbb{I}-\alpha(\mathbb{G}^T\mathbb{G}+\alpha\mathbb{I})^{-1}$$

ut if $\mathbb{C}$ is determined by (19) the resolution matrix turns out to e non-symmetric. Applying the similar transformation one can get:

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

nd though the two matrices $(\mathbb{G}^T\mathbb{G}+\alpha\mathbb{C})^{-1}$ and $\mathbb{C}$ are both symmetric, heir product is not symmetric. Thus, for calculation of the esolution matrix we can apply the same procedure, which is used or getting the solution, only for the solution with minimum norm. his is a disadvantage of the solution based on the smoothness riterion.

A row of the resolution matrix describes a resolution in one eparate point (or cell). To draw conclusions about resolution in ll cells we must calculate the total matrix rather than one row, but this equires too much time. Therefore it is desirable to be able to estimate resolution at least roughly in all points of the space. In classical Bakus-Gilbert method for 1D inverse problems such characteristics of the resolution is so-called 'averaging length'. If the 2D tomography problem is solved using the approach based on the smoothness criterion

$$\int |\nabla m(r)|^2 dr = \min,$$

the analogous characteristics is a linear dimension (or radius) of 'averaging area'. This radius is determined by the formula

$$R = \exp(0.75 - a^T\mathbb{S}a - 2H^Ta)$$

where the matrix $\mathbb{S}$ is determined by (17) and the vector H by (9).

Application of this criterion for estimating the resolution in the problems of surface wave tomography shows that the value of R in fact correlates with density of the wave paths and consequently may be taken as an estimate of the resolving power of the data.

3. Joint determination of lateral velocity variations and azimuthal anisotropy

Travel time residuals contain information not only on velocity variation but also about anisotropy. There are many papers, in which azimuthal anisotropy is estimated jointly with lateral velocity variations (Nakanishi & Anderson, 1983; Tanimoto & Anderson, 1985; Montagner & Tanimoto, 1990). In all such studies the velocity is assumed to be a function not only coordinates but also a direction of wave propagation. It is well known that for weak anisotropy in 2D case the velocity may be represented as

$$V = V_0 + A\cos 2\varphi + B\sin 2\varphi + C\cos 4\varphi + D\sin 4\varphi$$

where φ is azimuth.

Usually a contribution of the terms with 4φ is much less than those with 2φ, and they may be omitted. Assumung also lateral velocity variation to be small, we may represent the unknown velocity in the form:

$$V(x,y,\varphi) = V_0 + \delta V(x,y) + A(x,y)\cos 2\varphi + B(x,y)\sin 2\varphi$$

So now three unknown functions $\delta V(x,y)$, $A(x,y)$ and $B(x,y)$ are to be determined instead of one function $\delta V(x,y)$. These functions are usually represented in series of some basis functions with unknown coefficients (in spherical case - in series of spherical functions), and the coefficients are obtained from a linear system of equations.

But the question arises: is the resulting anisotropy is a real property of the Earth's material, or it is a consequence of a lateral heterogeneity?

To answer this question we should estimate a resolution for anisotropy , i.e. a dimension of averaging area, and compare it with correlation length of lateral velocity variations.

For estimation of the azimuthal anisotropy, i.e. the functions A(x,y) and B(x,y) we may apply the same technique as for estimation of lateral velocity variations in continious case.

As before, we substitute the unknown functions $\delta V(r)$, A(r) and B(r) with the following: $m(r)=-\delta V(r)/V_0$, $a(r)=-A(r)/V_0$, $b(r)=-B(r)/V_0$. Then the travel time residuals may be written in the form:

$$\delta t_i = \iint G_i(r)m(r)dr + \cos 2\varphi_i \iint G_i(r)a(r)dr + \sin 2\varphi_i \iint G_i(r)b(r)dr$$

The functions m(r), a(r) and b(r) may be determined by minimization of the functional analogous to (15):

$$(\mathbb{G}m - \delta t)^T(\mathbb{G}m-\delta t) + \alpha\iint|\nabla m|^2dr + \beta\iint(|\nabla a|^2+|\nabla b|^2)dr$$

Then it is easy to show that as before (see (8)):

$$m(r) = \sum \lambda_i H_i(r)+C_1$$

$$a(r) = \sum k\lambda_i \cos 2\varphi_i H_i(r)+C_2$$

$$b(r) = \sum k\lambda_i \sin 2\varphi_i H_i(r)+C_2$$

where $k=\alpha/\beta$.

The coefficients λ_i and C_1,C_2,C_3 are determined from the system similar to (16)

$$(\tilde{S}+\alpha\mathbb{I})a + C_1 t_0 + C_2 t_{0x} + C_3 t_{0y} = \delta t$$
$$\Lambda^T t_0 = 0$$
$$\Lambda^T t_{0x} = 0$$
$$\Lambda^T t_{0y} = 0$$

where

$$\tilde{S}_{ij} = S_{ij}(1+k\cos 2(\varphi_i-\varphi_j))$$

$$t_{0ix} = t_{0i}\cos 2\varphi_i$$

$$t_{0iy} = t_{0i}\sin 2\varphi_i$$

So the computational procedure developed for estimation of velocity variations with minor changes may be applied for joint estimation of lateral variations and azimuthal anizotropy. Azimuth of the fastest direction of wave propagation is then determined as

$$\varphi = \operatorname{arctg}(a/b)/2$$

and the anisotropy as

$$\alpha = 2\sqrt{a^2+b^2}$$

Analogously to the isotropic case the radii of averaging areas separately for a(x,y) and b(x,y) can be estimated.

Test examples show that even in case of exact data there is an influence of velocity variations on anisotropy and vice versa. This influence is strong in the cases when whithin the averaging area the averaged velocity is different along different direcrions.

Application of the described approach to the data on group velocities of Rayleigh waves in the region of South-eastern Europe shows a distinct anisotropy, which correlates with tectonic structures. However, the radii of averaging areas are found to be larger than dimensions of these structures, so that the anisotropy revealed in this study may be regarded as apparent anisotropy due to lateral heterogeneity similar to transverse anisotropy due to layering.

4. Computational facilities

For solving tomography problems two computer programs have been developed in S-Petersburg University:

- for 2D tomography problem based on continiuos solution, including determination of anisotropy;
- for 2D and 3D problems based on discretization.

The first program is designed for calculation of the solution (velocity distribution over a spherical surface), standard error of the solution and radius of averaging area over the region under investigation. Velocity in the starting model is assumed to be constant, only one step in improving the velocity being admitted.

In the second program a solution may be obtained by one of the three methods: BPT, MinNOrm (minimization of the functional (18)) and MaxSmooth (minimization of functional (20)). A column of resolution matrix may be calculated for any cell (in case of MinNorm solution it reflects resolution in the corresponding cell). Non-linearity is taken into account for 2D case, i.e. the solution obtained at the former step is assumed as a starting model at the next step. The program is supplied with graphic facilities for plotting pattern of rays and the solution at eacl step.