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/480-4

WORKSHOP ON EARTHQUAKE SOURCES & REGIONAL LITHOSPHERIC STRUCTURES FROM SEISMIC WAVE DATA

19 - 30 November 1990

REGIONALIZATION OF SURFACE WAVES

T.B. YANOVSKAYA

Leningrad State University
Institute of Physics
Leningrad, U.S.S.R.

REGIONALIZATION OF SURFACE WAVES

T.B.Yanovskaya

1. Introduction.

In recent years dispersion curves of phase or group velocities are widely used for both regional and global-scale tomography. The 3D tomography problem using surface wave data is easily separated into two independent problems: (1) 2D tomography for estimating local surface wave velocities for a set of periods, or for constructing "local" dispersion curves, (2) determination of S-wave velocity distribution with depths in each point of the Earth's surface from the local dispersion curves. The first problem is sometimes referred as "regionalization" problem, and it is of independent importance, because it allows some inferences about lateral inhomogeneities in the upper part of the Earth to be drawn.

The initial data used for regionalization are dispersion curves obtained from observations along the paths across different tectonic structures, so that the data are the average velocities over the paths. Thus the problem is formulated as follows:

using the average velocities along different paths, crossing the region under investigation, to estimate local values of the velocities.

Since the data set is always limited, we can determine only "locally averaged " velocities, therefore a resolving power also must be determined. If the data are inaccurate, the standard error of the solution should be estimated as well.

2. Methods for solving 2D tomography problem

2.1. General approach

The data, which are obtained from observations, are averaged surface wave velocities, or travel times along the paths L_i (i=1,2,..N). At first we shall assume that the data over all the paths correspond to one and the same period of the wave. Let the unknown lateral distribution of the surface wave velocity corresponding to this period be $V(x,y)$ or $V(\theta,\varphi)$. Then the relationship between the travel-time data and the unknown function is as follows:

$$t_i = \int_{L_i} \frac{ds}{V(x,y)} \qquad (1)$$

The path L_i depends on the unknown velocity distribution $V(x,y)$. But if the lateral velocity variations are sufficiently small, as takes place for surface waves, the paths in the first approximation may be taken as straight lines or great circles.

Let the mean velocity (averaged over the whole area), corresponding to the period under consideration, be V_0 .The laterally homogeneous medium with this surface wave velocity will be referred as "starting model". Instead of evaluating the unknown velocity distribution $V(x,y)$ we shall estimate the relative slowness correction respectively the starting model

$$m(x,y) = (V^{-1}(x,y) - V_0^{-1})/V_0^{-1} \qquad (2)$$

The travel times in the starting model along the same paths L_i are

$$t_{oi} = \int_{L_i} \frac{ds}{V_0}$$

so that the travel time residuals (differences between the observed travel times and calculated for the starting model) are

$$\delta t_i = t_i - t_{oi} = \int_{L_i} m(x,y)\frac{ds}{V_0} \qquad (3)$$

where L_i is a segment of a straight line.

If the observations contain experimental errors, we must add a random error ε_i to the right-hand side of (3):

$$\delta t_i = \int_{L_i} m(x,y)\frac{ds}{V_0} + \varepsilon_i \qquad (4)$$

To determine the function $m(x,y)$ from a limited set of data it is necessary to impose some a priori constrains to the unknown function. The constrains may be of different kind, but finally the solution may be written in the following general form

$$m(x,y) = \sum_j a_j \varphi_j(x,y) \qquad (5)$$

where $\varphi_j(x,y)$ are some basis functions. The basis functions may be either assumed a priori, or constructed proceeding from the given data set. Now, if the basis functions $\varphi_j(x,y)$ are known, we may determine the coefficients a_j proceeding from the linear system of equations.

If the data contain the experimental errors ε_i and the covariance matrix of the errors is R_t , then the solution is obtained by minimizing the functional

$$(A\alpha - \delta t)^T R_t (A\alpha - \delta t) \qquad (6)$$

where

$$A_{ij} = \int_{L_i} \varphi_j(x,y)\frac{ds}{V_0}$$

2.2 The alternatives for constructing the basis functions.

2.2.1. A priori choice of the basis functions.

One approach is the so-called "regionalization": the area under investigation is divided into some regions R_j (j=1,2,..,M), which are assumed to be laterally homogeneous, proceeding from geomorphological, geotectonical or some other geophysical data. The velocities within each region are assumed to be constants (for the given period), so that the problem is reduced to estimation of a number of parameters, which are the unknown slowness corrections m_j . The basis functions are defined as follows:

$$\varphi_j(x,y) = \begin{cases} 1 & \text{if } (x,y)\in R_j \\ 0 & \text{otherwise} \end{cases} \qquad (7)$$

If the length of the i-th path in the j-th region is l_{ij}, then $A_{ij} = l_{ij}/V_0$, i.e A_{ij} is the travel time along the i-th path in the j-th region. If the errors ε_i are not correlated and all have the same variance, the parameters m_j are estimated by the least-square method from the system

$$\sum_j (l_{ij}/V_0) m_j = \delta t_i \qquad (8)$$

A drawback of this approach is that it requires á priori regionalization, which is usually made proceeding from differences in surface features, but they may not reflect the lateral heterogeneities at depth.

Another approach is to represent the unknown function $m(x,y)$ or $m(\theta,\varphi)$ as a series in some appropriate functions: polynomials in a plane case or in spherical harmonics in a spherical case:

$$m(\theta,\varphi) = \sum_{n=0}^{K} \sum_{m=0}^{n} (A_n^m \cos m\varphi + B_n^m \sin m\varphi) P_n^m(\cos\theta) \qquad (9)$$

Since the number of the unknown coefficients A_n^m, B_n^m must be limited, this series is always truncated. The number of the coefficients is taken to be less than the number of the data, so the coefficients are estimated by minimizing (6), and in the case $R_t = \sigma^2 I$ it is reduced to estimating by the least-square method.

A drawback of this approach is that the solution does not reflect adequately the information, which the data contain: in the parts of the area covered densily by wave paths the solution becomes too smooth, and on the other hand, in the parts where the number of paths is small, the solution may contain spurious anomalies connected with peculiarities of the basis functions. Therefore, it is more expedient to determine a set of the basis functions in concordance with the pattern of paths.

2.2.2. Determination of the basis functions proceeding from the given data set

Now we consider the methods for constructing the basis functions $\varphi_j(x,y)$ proceeding from assumptions about some extremal features of the unknown function $m(x,y)$. In other words, we must impose some constrains to the function $m(x,y)$, which would be in accordance with our a priori notions about a behaviour of $m(x,y)$. The simplest constrain seems to be minimization of the functional

$$\iint_S m^2(x,y)\,dx\,dy$$

where S is the area under study. It is of reasonable physical meaning, because according to our initial assumtion, the function $m(x,y)$ is a correction to the starting model, which must be small. But if we used this criterion for constructing the basis functions, we should get the solution concentrated along the rays and equal to zero elsewhere. Therefore we must assume some correlation between the values of m in nearby points. This can be expressed by some different criteria.

One criterion may be referred as "smoothness criterion": the function $m(x,y)$ is assumed to be smooth in the following sense:

$$\iint_S |\nabla m(x,y)|^2\,dx\,dy = \min \qquad (10)$$

If the data did not contain experimental errors, we should solve the well-known variational problem: to determine the function $m(x,y)$, which mininizes the functional (10) and fit the constrains (3). For convenience we shall write the equations (3) in the form

$$\iint_S G_i(\mathbf{r})\, m(\mathbf{r})\,d\mathbf{r} = \delta t_i \qquad (11)$$

where $\mathbf{r} = (x,y)$ and the data kernel $G_i(\mathbf{r})$ is defined by the identity

$$\iint_S G_i(\mathbf{r})\, m(\mathbf{r})\,d\mathbf{r} = \int_{L_i} m(x,y)\,\frac{ds}{V_0}$$

i.e $G_i(\mathbf{r})$ is a generalized function tending to infinity at the i-th ray and equal to zero elswhere. It is easy to see that

$$\iint_S G_i(\mathbf{r})\,d\mathbf{r} = t_{oi}$$

It appears however, that the condition (10) is insufficient for determining the function $m(x,y)$ uniquely: it is necessary to assume some boundary condition at the contour of S. One of possible boundary conditions is following:

$$\left(\frac{\partial m}{\partial n}\right)_{C_s} = 0 \qquad (12)$$

Solving this variational problem, we easily find that the function $m(x,y)$ must be a solution of the Poisson's equation

$$\Delta m(\mathbf{r}) = -2\pi \sum_i \lambda_i G_i(\mathbf{r}) \qquad (13)$$

where λ_i are the Lagrange multipliers to be determined from (11) The factor -2π is inserted for convenience.

We choose Green's function of the Laplace equation as

$$g(\mathbf{r}',\mathbf{r}) = -\frac{1}{2\pi} \ln|\mathbf{r}'-\mathbf{r}| + \psi(\mathbf{r}',\mathbf{r}) \qquad (14)$$

where $\psi(\mathbf{r}',\mathbf{r})$ is harmonic function within S. Using the Green's formula and taking into account (12) we get

$$m(\mathbf{r}) = 2\pi \sum_j \lambda_j \iint_S G_j(\mathbf{r}')\, g(\mathbf{r}',\mathbf{r})\,d\mathbf{r}' - \int_{C_s} m(\mathbf{r}')\,\frac{\partial g(\mathbf{r}',\mathbf{r})}{\partial n}\,dl' \qquad (15)$$

For simplicity we remove the contour C_s away to infinity. Substituting (14) into (15) we conclude that for the last term in the right-hand side of (15) to be finite it is necessary that

(1) $\psi(\mathbf{r}',\mathbf{r})$ would be constant (for instance, C);

(2) $m(\mathbf{r})$ would be finite at infinity. The latter condition leads to

$$\sum_i \lambda_i t_{oi} = 0 \qquad (16)$$

Thus the solution may be written in the form

$$m(\mathbf{r}) = -\sum_i \lambda_i \int_{L_i} \ln|\mathbf{r}'-\mathbf{r}|\,\frac{ds'}{V_0} + C \qquad (17)$$

To determine N Langrange multipliers λ_i and the unknown constant C we have N constrains (11) and the condition (16).

Using the matrix notation

$$\Lambda^T = (\lambda_1, \lambda_2, \ldots \lambda_N)$$

$$t_0^T = (t_{01}, t_{02}, \ldots, t_{0N})$$

$$\delta t^T = (\delta t_1, \delta t_2, \ldots, \delta t_N)$$

$$S = \{S_{ij}\}, \quad S_{ij} = -\iint_{L_i L_j} \ln|r_i - r_j| \frac{dl_i}{V_0} \frac{dl_j}{V_0} \qquad (18)$$

we may write these equations in the form

$$S\Lambda + C t_0 = \delta t$$
$$\Lambda^T t_0 = 0 \qquad (19)$$

From (19) we obtain

$$C = \frac{t_0^T S^{-1} \delta t}{t_0^T S^{-1} t_0} \qquad \Lambda = S^{-1}\delta t - \frac{t_0^T S^{-1}\delta t}{t_0^T S^{-1} t_0} S^{-1} t_0 \qquad (20)$$

Denoting

$$K_j(r) = -\int_{L_j} \ln|r' - r| \frac{dl'}{V_0} \qquad (21)$$

and assuming $K_j(r)$ to be components of the vector $K(r)$, we may write the solution in the form

$$m(r) = \Lambda^T K + C$$

So the system of the basis function is following:

$$\psi_j(r) = \begin{cases} K_j(r) & j=1,2,..,N \\ 1 & j=N+1 \end{cases} \qquad (22)$$

This approach can be easily expanded to inaccurate data in the framework of Tikhonov's 'regularization' method. According to this method we shall look for the solution proceeding from the following assumption on the function $m(r)$:

$$(Gm - \delta t)^T R_t^{-1} (Gm - \delta t) + \alpha \iint |\nabla m|^2 dr = \min \qquad (23)$$

where α is regularization parameter, the notation

$$(Gm)_i = \iint G_i(r) m(r) dr$$

is implied, and $G(r)$ is regarded as a vector with components $G_i(r)$. The unknown parameter α must be chosen so that the first term in (23) would be equal to the total number N of data.

It appears that in this case the system of the basis functions will be the same as in the case of exact data, i.e. they are defined by (22), therefore the solution may be also written in the form (22), but in the expressions for Λ and C the matrix S must be replaced by the matrix $S + \alpha R_t$.

Another criterion for constructing the basis functions was proposed by Tarantola & Nersessian (1984). The mainpoint in this approach is that the a priori covariance function of the model $R_{m_0}(r', r)$ is assumed to be known, and $m(r)$ is determining by minimizing the functional

$$m^T R_{m_0}^{-1} m + (Gm - \delta t)^T R_t^{-1} (Gm - \delta t) \qquad (24)$$

where $m = m(r)$ and $R_{m_0} = R_{m_0}(r', r)$ are regarded as vector and matrix in Hilbert space, so that

$$m^T R_{m_0}^{-1} m = \iint m(r') R_{m_0}^{-1}(r', r) m(r) dr\, dr' \qquad (25)$$

The solution of this problem is obtained easily:

$$m = (R_{m_0}^{-1} + G^T R_t^{-1} G)^{-1} G^T R_t^{-1} \delta t \qquad (26)$$

but this form of the solution is inconvenient, because (26) involves the inverse function $R_{m_0}^{-1}$. After some transformations of (26) we obtain

$$\begin{aligned} m &= (R_{m_0}^{-1} + G^T R_t^{-1} G)^{-1} G^T R_t^{-1} (R_t + G R_{m_0} G^T)(R_t + G R_{m_0} G^T)^{-1} \delta t = \\ &= (I + R_{m_0} G^T R_t^{-1} G)^{-1} R_{m_0} G^T (I + R_t^{-1} G R_{m_0} G^T)(R_t + G R_{m_0} G^T)^{-1} \delta t = \\ &= (I + R_{m_0} G^T R_t^{-1} G)^{-1} (I + R_{m_0} G^T R_t^{-1} G) R_{m_0} G^T (R_t + G R_{m_0} G^T)^{-1} \delta t = \\ &= R_{m_0} G^T (R_t + G R_{m_0} G^T)^{-1} \delta t \end{aligned} \qquad (27)$$

It is clear from the final formula that a vector of the basis functions is GR_{m_0}, or the basis functions are expressed as follows

$$\psi_j(r) = \iint R_{m_0}(r', r) G_j(r') dr' = \int_{L_j} R_{m_0}(r', r) \frac{ds'}{V_0} \qquad (28)$$

A question arises: how to choose the a priori covariance function R_{m_0} ? The simplest analytical form of the function is Gaussian:

$$R_{m_0}(r', r) = \sigma^2 \exp\left\{-\frac{|r' - r|^2}{2L^2}\right\} \qquad (29)$$

where L is a correlation length.

The correlation length is a kind of smoothing parameter: if it is too large, the solution will be very smooth, and if it is too small, the solution will be concentrated along rays. Proceeding from the theory of surface wave propagation we can conclude that L must be not smaller than the wavelength, and on the other side, it must be large enough that the paths of effective width ensure a good coverage of the area.

2.3. Resolving power and variance of the solution.

The solution, which is obtained from a finite set of data, is always an averaged solution. This concernes the both approaches for solving the tomography problem: a priori choice of the basis functions and determining the basis functions from the given set of data according to some criterion. Since the problem is linearized, so that the solution is a linear combination of the data, it is a linear average. Indeed, if the data are exact,

$$\hat{m} = Q \delta t$$

where Q is a matrix, the form of which depends on a chosen approach, $\hat{m}$ is the solution (either a set of parameters, or a function $\hat{m}(r)$). Taking into account that

$$\delta t = Gm$$

where m is a real velocity distribution, we obtain

$$\hat{m} = (QG) m = A m$$

where A is the so-called resolution matrix, characterizing the resolution. If the solution if a function $\hat{m}(r)$, then

$$\hat{m}(r) = \iint_S A(r',r)m(r')dr'$$

$A(r',r)$ is called resolving or averaging kernel: the closer is $A(r',r)$ to the delta-function $\delta(r'-r)$, the better is the resolution.

If the data are inaccurate, i.e.

$$\delta t = Gm + \varepsilon$$

then

$$\hat{m} = Q(Gm + \varepsilon) = Am + Q\varepsilon$$

If as usual $E(\varepsilon) = 0$, then $E(\hat{m}) = Am$, so that the resolution matrix is the same: $A = QG$.

Now since the solution has a random error, we can estimate its a posteriori covariance matrix (or covariance function):

$$R_m = E\{(\hat{m} - E(\hat{m}))(\hat{m} - E(\hat{m}))^T\}$$

Tarantola & Valette (1982) have shown that á posteriori covariance matrix R_m is related to á priori covariance matrix R_{mo} as follows:

$$R_m = (I - A)R_{mo}$$

Diagonal elements of the covariance matrix are variances of the parameters, or in the continuous case, for which $R_m = R_m(r',r)$, the variance of the solution in the given point r is

$$Var\{m(r)\} = R_m(r,r)$$

In another way we may write the solution in the form

$$m(r) = a^T \delta t$$

where the vector $a = a(r)$, then

$$Var\{m(r)\} = a^T R_t a$$

This formula is convenient for numerical estimation of the standard error of the solution.

It is convenient to estimate the resolving power by such a quantity, which characterizes a proximity of the resolving kernel to the delta-function. For one-dimensional case Backus & Gilbert introduced a conception of an averaging length, which is an estimate (according to some criterion) of a 'width' of the averaging kernel. The similar approach can be used also for 2D case.

For this purpose we chose some functional $S^2(r)$, which expresses a proximity of the averaging kernel $A(r',r)$ to the delta-function $\delta(r'-r)$: $S^2=0$ if $A(r',r)=\delta(r'-r)$ and increases with increase of the difference between $A(r',r)$ and $\delta(r'-r)$. But it is inconvenient for interpreting the resolving power: it is more convenient to chose some quantity related to S^2, but having a

certain physical meaning. If we now define a special averaging kernel $A(r',r)$, which differs from zero in $\mathcal{R}$-vicinity of the given point and vanishes outside, and determine $\mathcal{R}$ so that the chosen functional S^2, calculated for this averaging kernel would be equal to that determined for the given set of data, we may use this value of $\mathcal{R}$ as an estimate of linear dimension of the averaging area.

Examples. **(1)** For the solution satisfying the criteria (10), (12) it is convenient to take for S^2 the following functional:

$$S^2 = \iint |E(r',r) - H(r',r)|^2 dr'$$

where

$$div\, E(r',r) = A(r',r)$$
$$H(r',r) = (2\pi)^{-1} V_0' \nabla' \ln|r'-r|$$

The solution (17) alternatively can be obtained from the criterion $S^2 = \min$ under the condition

$$\iint A(r',r)\,dr' = 1$$

The 'equivalent' averaging kernel $A(r',r)$ in this case is following:

$$A(r',r) = \begin{cases} 1/\pi\mathcal{R}^2 & \text{if } |r'-r| < \mathcal{R} \\ 0 & \text{if } |r'-r| > \mathcal{R} \end{cases}$$

and $\mathcal{R}$ is related to S^2 as follows:

$$\mathcal{R} = \exp\{3/4 - a^T S a - 2K^T a\}$$

where a is vector of the coefficients in the relationship

$$m = a^T \delta t$$

$K(r)$ is a vector of the functions $K_j(r)$ defined by (21) and S is a matrix defined by (18).

(2) The solution (27) also may be obtained from a deltaness criterion of the averaging kernel. This criterion is written in the form

$$S^2(r) + Var\{m(r)\} = \min$$

where

$$S^2(r) = \iint [\iint A(r',r)f(r',\rho)dr' - f(r,\rho)]^2 d\rho$$

and the function $f(r',r)$ satisfies the condition

$$\iint f(r',r)f(r',\rho)dr' = R_{mo}(r,\rho)$$

If the a priori covariance function $R_{mo}(r,\rho)$ is determined by (29), then

$$f(r',r) = \sqrt{\frac{2}{\pi}}\frac{\sigma}{L}\exp\{-|r'-r|^2/L^2\}$$

In this case we may take $S^2(r)$ as a measure of deltaness of the averaging kernel $A(r',r)$ and for the 'equivalent' averaging kernel assume the Gaussian function

$$\frac{1}{2\pi R^2} \exp\left\{-\frac{|r'-r|^2}{2R^2}\right\}$$

It is easy to show that R is related to S^2 as follows:

$$S^2 = \frac{4R^4\sigma^2}{(L^2+2R^2)(L^2+R^2)}$$

2.4. *Surface wave tomography based on joint use of phase and group velocities for different periods.*

If the dispersion curves for phase and group velocities are available for the area under consideration, these data can be used jointly, since the phase and group velocities are related:

$$U^{-1}(\omega) = C^{-1}(\omega) + \omega \frac{dC^{-1}}{d\omega} \qquad (30)$$

In this case we must estimate a function of three variables, i.e. $C(x,y,\omega)$ rather than $C(x,y)$ or $U(x,y)$ for a given period, so that we must solve 3D inverse problem. In general we could apply the same approach as for solving 2D tomography problem. It is easy to see from the relationship (30) that if we define the unknown function as

$$\mu(x,y,\omega) = C^{-1}(x,y,\omega) - C_0^{-1}(\omega)$$

then the travel-time residuals δt^{ph} and δt^{gr} will be the linear functionals of $\mu(x,y,\omega)$ in 3D space:

$$\delta t^{ph}(\omega_{kj}) = \iiint \mu(x,y,\omega)[H_k(x,y)\delta(\omega-\omega_{kj})]\,dx\,dy\,d\omega$$

$$\delta t^{gr}(\omega_{kj}) = \iiint \mu(x,y,\omega)\left[\omega H_k(x,y)\frac{d\delta(\omega-\omega_{kj})}{d\omega}\right]dx\,dy\,d\omega$$

where $H_k(x,y)$ is the kernel similar to $G_k(x,y)$:

$$\iint H_k(x,y)\,dx\,dy = \int_{L_k} ds = l_{0k}$$

subscript k corresponds to the path, and j corresponds to a frequency for the given path.

But since $C(x,y,\omega)$ as a function of ω is monotonous and sufficiently smooth, we may approximate C^{-1} as a function of ω by a polynnomial:

$$C^{-1} = \sum_{q=0}^{Q} a_q(x,y)\,\omega^q$$

According to (30)

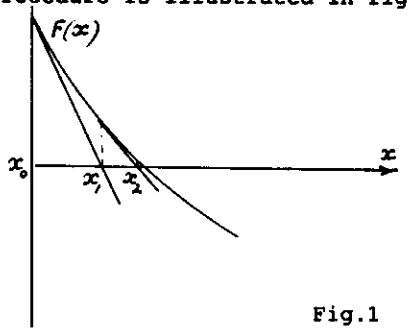
$$U^{-1} = \sum_{q=0}^{Q} (1+q)\,a_q(x,y)\,\omega^q$$

So now the 3D problem is reduced to 2D problem for estimation of a set of functions $a_q(x,y)$, and for this purpose we can use one of the approaches desribed above.

2.5. *A simple iterative procedure for solving a non-linear problem.*

We assumed the lateral velocity variations to be so small that the wave paths may be approximated by straight lines. But in some cases (especially in regional studies) this assumption is not valid, so that the paths L_k are markedly curvilinear and their shape depends on the unknown phase velocity distribution.

For solving a non-linear problem we can apply the usual iterative procedure: at the first iteration we assume the standard model as before, then the solution at the first step is used as the standard model at the second step, etc. This procedure is equivalent to the well-known Newton's method for solving transcendent algebraic equations. Geometrically this iterative procedure is illustrated in fig.1.



$$F(x) = 0$$

$$x_1 = x_0 - F(x_0)/F'(x_0),$$

$$x_2 = x_1 - F(x_1)/F'(x_1),$$

etc.

Fig.1

But for solving the tomography problem this procedure is inconvenient, because it results in tremendous complication of computing. The analogy to the derivative $F'(x)$ is the matrix A, elements of which are calculated by integration of a basis functions along the rays. When the standard model is a constant velocity, these integrals are calculated easily. But beginning from the second iteration the calculations become much more complicated, because these integrals should be calculated numerically along curvilinear rays. Especially this concerns the cases, when the basis functions are themselves calculated by integration along the rays (see formulas (21) or (28)).

To avoid this complication another iterative procedure may be proposed. It can be demonstrated by solving the equation $F(x)=0$. Let the first approximation be $x = x_0$. Then we may write the equation in the form:

$$F(x) = F(x_0) + F'(x_0)(x-x_0) + \Phi(x) = 0$$

where $\Phi(x)$ involves all non-linear terms. As before we assume $\Phi(x_0)=0$ at the first iteration. Then

$$x_1 = x_0 - \frac{F(x_0)}{F'(x_0)}$$

Then we use this solution to correct the non-linear term:

$$F(x_0) + F'(x_0)(x - x_0) + \Phi(x_1) = 0$$

or

$$x_2 = x_0 - \frac{F(x_0) + \Phi(x_1)}{F'(x_0)}$$ etc.

Geometrically this process is illustrated in fig.2a

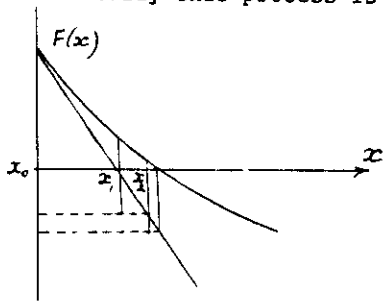


Fig.2a

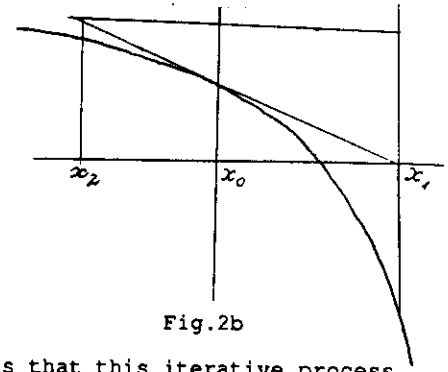


Fig.2b

A drawback of this approach is that this iterative process converges not always. It is clear from fig.2b. But the divergence of the process may be avoided by damping.

For solving the tomography problem this approach leads to the following iterative procedure:

1) at the first iteration we solve the linear problem and obtain the solution $C(x,y)$ (C is the phase velocity!)
2) then we calculate the differences between the travel times along the straigth lines in the model $C(x,y)$ and along the rays corresponding to the velocity distribution $C(x,y)$; let us denote them as $\Delta t_i^{(1)}$ (apparently they must be positive);
3) again solve the linear problem with the same matrix A as at the first iteration, but with the data $\delta t_i + \Delta t_i^{(1)}$;
4) repeat from (2)-(3) until $|\Delta t_i^{(n)} - \Delta t_i^{(n+1)}| < \varepsilon$.

3. Comparison of the methods.

The question arises: which approach for surface wave tomography is preferable? The answer depends on that how the real Earth's structure fits the a priori assumptions, which one or another method is based on. The two extreme assumptions are: (1) the lithosphere is consisted of some laterally homogeneous regions (blocks), the velocities are discontinious at the boundaries of the blocks; (2) surface wave velocities are continious everywhere and vary smoothly.

The former assumption seems to be more realistic, although of course the velocities vary within the blocks, and the blocks are divided some transition zones rather than the sharp boundaries. Applying the latter approach for studying such media we should obtain a too smooth solution: in high velocity regions we should get lower velocities and vice versa. But to obtain a reasonable solution by applying the former approach we must know location of the boundaries of the blocks as exact as possible. Otherwise we might obtain unrealistic velocities. This is demonstrated by some examples: surface