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/480-13

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

19 - 30 November 1990

Nonlinear Inversion of Surface Wave Data

D. Lokchtanov

International Inst. of Earthquake Prediction & Math. Geophysics of the USSR Moscow, U.S.S.R.

NONLINEAR INVERSION OF SURFACE WAVE DATA

by D. Lokchtanov

The conventional approaches for surface wave data analysis, aimed to find media model parameters, consist mostly of the following three steps. First of all the raw seismograms are filtered to suppress nonstationary unwanted waves, that are not accomodated by the useful signal model. After this the filtered records are used to extract from them such wave characteristics, that are determined by the investigated region only. Commonly these characteristics are surface wave group and phase velocity (or slowness) dispersion curves. And, at the third step, the derived dispersion curves are used for finding the investigated region model parameters. Namely this last step will be considered here.

1. Determination of media parameters from surface wave dispersion curves.

Now we consider a task of finding the media model parameters from Rayleigh wave phase or group velocity (slowness) dispersion curves. The model is represented by a stack of horizontal homogeneous layers, bounded from above by the free surface and by the homogeneous halfspace from below. The sense of the model depends on a way of dispersion curves extracting. For example, if they are obtained as a result of tomography procedure, then the inversion of each reconstructed "local" dispersion curve would give us parameters of "local" vertical structure model. Otherwise, this model can

describe the horizontally averaged structure between source and receiver or between several receirvers.

So, let us have observations of Rayleigh wave fundamental mode phase slowness $p^{\circ}(\omega_i)$ ($p = 1/c$, c - phase velocity), available at frequencies ω_i, $i = 1, \ldots, N$. The errors of observations will be denoted by ξ_i, while its standard deviation by δ_i. Now, the task of dispersion curve inversion can be formulated as finding such values $\vec{m}$ of media model parameters, that provide the best fittness between theoretical slownesses $p^{T}(\omega_i, \vec{m})$ and observed ones. If this fittness is determined in the least-square sense, then the parameters estimates $\hat{\vec{m}}$ are found by minimizing with respect to $\vec{m}$ the following functional

$$\phi_p(\vec{m}) = \sum_{i=1}^{N} \frac{1}{\delta_i^2} \left[p^{T}(\omega_i, \vec{m}) - p^{\circ}(\omega_i) \right]^2. \qquad (1)$$

Here vector $\vec{m}$ consists of parameters of the layers, namely their thicknesses, densities, compressional and shear wave velocities. When observational errors have Gaussian distribution and uncorrelated at different frequences, the functional (1) corresponds to the functional of maximum likelyhood, that provides optimal statistical properties to the estimates $\vec{m}$.

By the similar way, if instead of phase slowness observations group slowness observations $g^{\circ}(\omega_i)$ are available, the parameters of the structure model can be found by minimizing the functional

$$\phi_g(\vec{m}) = \sum_{i=1}^{N} \frac{1}{\delta_i^2} \left[g^{T}(\omega_i, \vec{m}) - g^{\circ}(\omega_i, \vec{m}) \right]. \qquad (2)$$

2. Determination of medium parameters from spectra of wave records.

Up to now we assumed, that only one mode observational dispersion curve is available. In the most cases it is really so and this mode is a fundamental one. But in some cases the main portion of recorded energy represented not only by fundamental mode, but by sum of fundamental and higher modes. In this case, if dispersion curves for different modes can be obtained separately, then instead of functionals (1) and (2) we use the next ones

$$\phi_p(\vec{m}) = \sum_{j=1}^{K} \sum_{i=1}^{N_j} \frac{1}{\delta_{ij}^2} \left[p_j^{T}(\omega_{ij}, \vec{m}) - p_j^{\circ}(\omega_{ij}) \right]^2, \qquad (3)$$

$$\phi_g(\vec{m}) = \sum_{j=1}^{K} \sum_{i=1}^{N_j} \frac{1}{\delta_{ij}^2} \left[g_j^{T}(\omega_{ij}, \vec{m}) - g_j^{\circ}(\omega_{ij}) \right]^2. \qquad (4)$$

In this formulas K - the number of considered modes, ω_{ij} - the frequency value for i-th observation for j-th mode.

Now let us have the case, when the signals for different modes interfere with each other in the time-frequency domain and it is impossible to extract each of these signals separately without their significant distortion. In such situation the media model parameters can be inverted directly from the initial records or their spectra.

Consider, for example, the task of determining the horizontally layered structure model parameters $\vec{m}$ between two seismic stations, lying approximately on the same great circle with the source. We assume, that at each station the seismograms of radial and vertical displacement component are available and the main part of the recorded energy is carried by Rayleigh wave fundamental and first modes. In that case, for observations in the frequency domain at both stations, we have the next relation

$$\vec{U}(\omega_k) = T(\omega_k, \vec{m})\,\vec{C}(\omega_k) + \vec{\zeta}(\omega_k), \qquad (5)$$

where

$\vec{U}(\omega_k)$ - four dimensional vector, containing the results of discrete Fourier transform of the observed (filtered) seismograms, taken at the frequency ω_K :

$$\vec{U}(\omega_k) = \left(\vec{U}_1^T(\omega_k), \vec{U}_2^T(\omega_k)\right)^T, \qquad (6)$$

$$\vec{U}_\ell(\omega_k) = \left(u_{r\ell}(\omega_k), i u_{z\ell}(\omega_k)\right)^T, \qquad (7)$$

ℓ - the station index, $\ell = 1,2$; $u_{r\ell}$, $u_{z\ell}$ - the spectrum of radial and vertical displacement componenet; $i = (-1)^{1/2}$; $\omega_k = 2\pi k/\mathcal{D}$, $\mathcal{D}$ - duration of the analyzed part of the seismograms; T - symbol of transposition. The matrix $T(\omega_k, \vec{m})$ of the size (4x2) depends on media model parameters $\vec{m}$: phase slownesses $p_0(\omega_k, \vec{m})$, $p_1(\omega_k, \vec{m})$ of the Rayleigh wave fundamental and first mode; the values of the eigenfunctions at the free surface $r_{0r}(\omega_k, \vec{m})$, $r_{0z}(\omega_k, \vec{m})$, $r_{1r}(\omega_k, \vec{m})$, $r_{1z}(\omega_k, \vec{m})$.; the distance

ΔR between the stations and the factor q , describing the geometrical spreding :

$$T(\omega_k, \vec{m}) = \left[\frac{R(\omega_k, \vec{m})}{q\,R(\omega_k, \vec{m})\,L(\omega_k, \vec{m})}\right], \qquad (8)$$

and here R is the matrix of eigenfunctions values at the free surface

$$R(\omega_k, \vec{m}) = \begin{bmatrix} r_{0r}(\omega_k, \vec{m}) & r_{1r}(\omega_k, \vec{m}) \\ r_{0z}(\omega_k, \vec{m}) & r_{1z}(\omega_k, \vec{m}) \end{bmatrix} \qquad (9)$$

while the diagonal matrix L controls phase shifts in spectrum of fundamental and first mode due to propagation between stations 1 and 2

$$L(\omega_k, \vec{m}) = \begin{bmatrix} \exp\{i\omega_k p_0(\omega_k, \vec{m})\Delta R\} & 0 \\ 0 & \exp\{i\omega_k p_1(\omega_k, \vec{m})\Delta R\} \end{bmatrix} \qquad (10)$$

Also in the formula (5), $\vec{C}(\omega_k)$ - the two dimensional vector of unknown "input" signals of the fundamental and first mode at the first station (if eigenfunctions are normalised in a such way, that at the free surface $r_{0r}^2 + r_{0z}^2 = 1$, $r_{1r}^2 + r_{1z}^2 = 1$, then vector $\vec{C} = (C_0, C_1)^T$ consists of complex amplitudes of fundamental and first mode displacement vector in the plane RZ). In the same

formula $\vec{\zeta}(\omega_k) = (\vec{\zeta}_1^T(\omega_k), \vec{\zeta}_2^T(\omega_k))^T$ - four dimensional noise vector. We suppose, that $M[\vec{\zeta}(\omega_k)] = \vec{0}$, $M[\vec{\zeta}(\omega_k)\vec{\zeta}^*(\omega_k)] = \delta_{kj} F(\omega_k)$ M - mathematical expectation, $*$ - Hermitian conjugation symbol, δ_{kj} - Kronecker delta symbol.

Our observation model (5) is really a model of nonlinear regression analysis, and for estimating of $\vec{m}$. we can use the optimal statistical methods, as the method of maximum likelyhood (ML). If suppose, that noise $\vec{\zeta}(\omega_k)$ is Gaussian, then the functional of maximum likelyhood for observations (5) will have the following form

$$L(\vec{m}) = -\frac{1}{2}\sum_{K=1}^{N} \ln\det F(\omega_k) - \frac{1}{2}\sum_{K=1}^{N} \{\vec{U}(\omega_K) - \quad (11)$$

$$- T(\omega_K, \vec{m})\vec{C}(\omega_K, \vec{m})\}^* F^{-1}(\omega_K)\{\vec{U}(\omega_K) - T(\omega_K, \vec{m})\vec{C}(\omega_K)\},$$

and it depends not only on "useful" parameters $\vec{m}$, but also on unknown regressors $\vec{C}(\omega_K)$, $K = 1, \ldots, N$. When the number of nuisance parameters is fixed and the sample size is increasing, asimptotically optimal estimates of informational parameters are ML estimates, derived by maximization of ML functional both with respect to nuisance and informational parameters. In the considered task the number of nuisance parameters (unknown regressors) is equal to $2N$ and proportional to the sample size $4N$. Nevertheless, the ML estimates of $\vec{m}$., found by maximization of (11) with respect to $\vec{m}$ and $\vec{C}(\omega_K)$ are $\sqrt{N}$ consistent (Kushnir & Lokchtanov, 1988).

When $\vec{m}$ is fixed, the maximum of L (11) with respect to $\vec{C}(\omega_K)$ is achieved at the point

$$\hat{\vec{C}}(\omega_k) = \left[T^*(\omega_k)T(\omega_k)\right]^{-1} T^*(\omega_k)\vec{U}(\omega_k) \quad (12)$$

If we substitute (12) into (11), we will find, that estimates of $\vec{m}$ can be found by minimizing of the functional

$$\phi(\vec{m}) = \sum_{K=1}^{N} \| \vec{U}(\omega_K) - T(\omega_K, \vec{m})\left[T^*(\omega_K, \vec{m})T(\omega_K, \vec{m})\right]^{-1} T(\omega_K, \vec{m})\vec{U}(\omega_K)\|^2. \quad (13)$$

In the formulas (12) and (13) we suggested, that $F(\omega_k) = I$, where I is a unit matrix. The common case of arbitrary F can be reduced to the considering one by the following changing of variables $\vec{U} \to F^{-1/2}\vec{U}$, $T \to F^{-1/2}T$.

The functional (13) depends only on observations $\vec{U}(\omega_K)$ and informational parameters $\vec{m}$. So, we do not need any assumptions about the source parameters and media between the source and investigated region. In the same procedure we exploit not only the dispersion properties of Rayleigh waves, but also their polarization features. The described scheme was used for inverting Rayleigh wave seismograms, recorded at the NARS's stations (Kushnir et. al., 1988; Lokchtanov & Trusov, 1990).

3. Algorithm of nonlinear minimization.

As it was derived in the previous sections, the suggested procedures for inversion of surface wave data require minimization of functional ((1)-(4) or (13)) with essentially nonlinear dependence on the media model parameters $\vec{m}$. So we have a problem of choosing the most suitable scheme for nonlinear optimization. We suggest to implement one of the best of such schemes, based on the method of conjugate gradients (Polak, 1985). According to this scheme minimization of the functional $\phi(\vec{m})$ is realized by successive search for the points of the minimum in so-called conjugate directions in the space of parameters R^n . Determination of these directions requires calculation of the gradient $\nabla\phi(\vec{m})$ at each step of the iterations. If $\phi(\vec{m})$ is a positive definite quadratic form in space R^n , then the point of the minimum is reached during n iterations.

The conjugate gradient method leads to the following iterative scheme:

1. The initial point $\vec{m}_0$ is assigned and

$$\vec{d}_0 = -\nabla\phi(\vec{m}_0)$$

is calculated.

2. For $K = 0, \ldots, n-1$ the point $\vec{m}_{K+1}$ is calculated according to the iterative procedure

$$\vec{m}_{K+1} = \vec{m}_K + \alpha_K \vec{d}_K$$

where α_K is determined by one-dimensional minimization, i.e.

$$\phi(\vec{m}_K + \alpha_K \vec{d}_K) = \min_{\alpha} \phi(\vec{m}_K + \alpha \vec{d}_K).$$

If $K \geq 1$, vector $\vec{d}_K$ is calculated by formula

$$\vec{d}_K = -\nabla\phi(\vec{m}_K) + \beta_K \vec{d}_{K-1}, \qquad (14)$$

where

$$\beta_K = \frac{\nabla\phi^T(\vec{m}_K)\nabla\phi(\vec{m}_K)}{\nabla\phi^T(\vec{m}_K)\nabla\phi(\vec{m}_{K-1})} \qquad (15)$$

3. The process ends when $\|\nabla\phi(m_{K+1})\|$ or $\|\vec{m}_{K+1} - \vec{m}_K\|$ is sufficiently small. In opposite case we go to step 2.

4. In the case of $\phi(\vec{m})$ not quadratic with respect to $\vec{m}$ the procedure of searching for the points of the minimum usually does not end during n steps. In this case, if local quadratic approximation rapidly vary from iteration to iteration, the method begins to generate in several steps such directions of "movement", that are not effective for minimization. Taking this into account, the iteration procedure 1-3 is constructed usually by cycles with n iterations each of them, and every new cycle begins from the steepest descent iteration 1. Introduction of

such iteration of steepest descent also provides the global convergence to the algorithm.

Often, the value of β_K is calculated not by formula (15), but with the following expression

$$\beta_K = \frac{\nabla\phi^T(\vec{m}_K)\left(\nabla\phi(\vec{m}_K) - \nabla\phi(\vec{m}_{K-1})\right)}{\nabla\phi^T(\vec{m}_{K-1})\nabla\phi(\vec{m}_{K-1})} \qquad (16)$$

For quadratic forms formulas (15) and (16) are equalent (as $\nabla\phi^T(\vec{m}_K)\nabla\phi(\vec{m}_K) = 0$). But for nonquadratic function it is much more preferable to use (16). The reason is that due to nonquadratic effects and inexectitude of one-dimensional minimization, the procedure can temporally "stick", as it begins to generate directions of minimization, that are almost perpendicular to $\nabla\phi(\vec{m}_K)$. In a such case $\nabla\phi(\vec{m}_{K+1}) \approx \nabla\phi(\vec{m}_K)$ and $\beta_{K+1} \approx 0$ Then, according to (14), the next direction of minimization will be close to $-\nabla\phi(\vec{m}_{K+1})$ and the "stick" will be got over. The formula (15) does not posseses the same ability.

4. The forward problem solution

According to the implementing scheme of minimization at each step of iteration procedure we must solve the forward problem - calculate eigenvalues, eigenfunctions and their derivatives with respect to the media model parameters.

For solving the dispersion equation

$$F(\omega, p, \vec{m}) = 0 \qquad (17)$$

(ω - frequency, p - horizontal slowness, $\vec{m}$ - media model parameters) we use the method, developed by Kennett (1979,1981) and Kerry (1980). According to their approach, the dispersion function $F(\omega, p, \vec{m})$ is represented through matrixes R_D (2x2), T_D (2x2) of reflection and transmission coefficients of $P\text{-}SV$ waves for

a layered structure under the free surface and also the matrixes n_D (2x2), n_U (2x2), depending on parameters of the upper layer:

$$F(\omega, p, \vec{m}) = \det(n_U R_D + n_D)/\det T_D = 0 \qquad (18)$$

The matrixes R_D , T_D are evaluated by the scheme, suggested by Kennett (1974). According to this scheme at first we calculate the matrixes r_D, r_U, t_D, t_U, of reflection and transmission of $P\text{-}SV$ waves for individual boundaries and then use a recurcive procedure for evaluating R_D and T_D. Solving the dispersion equation, we fix slowness p and vary frequency ω . In elastic media matrixes R_D, T_D don't depend on a frequency ω . So, when slowness p . is fixed and frequency ω is varied, the main part of calculation of R_D, T_D is realized just once. Therefore, we can fastly evaluate the dispersion function $F(\omega, p, \vec{m})$ for many different values of ω and by this way to look for the roots of the dispersion equation. The roots $\omega_i(p)$ (i - the number of mode) are found by changing the sign of dispersion function. In corresponding interval of ω the value of root is defined more precisely by the bisection method.

For evaluating the derivatives of phase slowness p with respect to model parameters, we use the formulas, obtained from variational principles for surface waves (Aki & Richards, 1980; Keilis-Borok, 1989). If we introduce in the media model small perturbations of density $\delta\rho(z)$ and Lame parameters $\delta\lambda(z)$, $\delta\mu(z)$, then the perturbation of Rayleigh wave phase velocity C ($C = 1/p$) will be determined by the following formula

$$\delta C = \frac{1}{4UI_1K^2}\left\{\int_0^\infty \left(K r_z + \frac{dr_x}{dz}\right)^2 \delta\lambda\, dz + \right.$$
$$+ \int_0^\infty \left[2K^2 r_z^2 + r\left(\frac{dr_x}{dz}\right)^2 + \left(K r_x - \frac{dr_z}{dz}\right)^2 \delta\mu\, dz - \right.$$
$$\left. \left. - \int_0^\infty \omega^2 (r_x^2 + r_z^2)\,\delta\rho\, dz \right\}, \qquad (19)$$

where

U - group velocity; $K = \omega/c$; $I_1 = \int_0^\infty \rho(r_x^2 + r_z^2)\, dz$;

r_x, r_z - Rayleigh wave eigenfunctions. For horizontally layered structure all

underintegral expressions in (19) can be represented by sum of exponential terms (upgoing and downgoing P and SV waves), so analytical evaluation of these integrals is straightforward.

The same variational approach was used also for obtaining the derivative formula of phase velocity with respect to boundary level positions (Kushnir et. al., 1988; Bukchin, 1990)

$$\frac{\delta c}{\delta z_n} = \frac{1}{4UI_1K^2}\left\{(K^2\delta[\lambda+2\mu]-\omega^2\delta[\rho])r_r^2 + \right.$$
$$+(K^2\delta[\mu]-\omega^2\delta[\rho])r_z^2 - \delta\left[\mu\left(\frac{dr_r}{dz}\right)^2 - \right. \qquad (20)$$
$$\left.\left.-\delta\left[(\lambda+2\mu)\left(\frac{dr_z}{dz}\right)^2\right]\right\}.$$

In this formula $\delta[\ ]$ means the difference of inside bracket terms, evaluated just over and below the n-th boundary:

$$\delta[f] = f(z_n-) - f(z_n+).$$

When using the functional (2) or (4) for inverting the group slowness observations, we need also the formulas of $\partial g/\partial m_K$ ($g = 1/U$, U - group velocity) group slowness derivatives with respect to model parameters. Taking into account, that

$$g = \frac{\partial K}{\partial \omega} = \frac{\partial(\omega p)}{\partial \omega}$$

we get

$$\frac{\partial g}{\partial m_K} = \frac{\partial p}{\partial m_K} + \omega \frac{\partial}{\partial \omega}\left(\frac{\partial p}{\partial m_K}\right)$$

The terms $\partial p/\partial m_K$ are evaluated by the formulas (19) and (20), while derivative with respect to frequency by the finite difference approximation.

Concluding remarks.

Now we have all required "bricks" for inversion of surface wave data by means of nonlinear minimization. Let us also mention, that inversion of dispersion curves from computational point of view is much more simple task than inversion of records spectrum, requiring the minimization of a quite complex functional with a lot of local extremums. So, practically, it is preferable first to invert just dispersion curves and then, may be, to improve the model parameters estimates by minimization of (13), where polarization features of Rayleigh waves are also exploited.

REFERENCES

Aki K. and Richards P. (1980) Quantitative seismology: Theory and Methods, W.H.Freeman and Company, San Francisco, 932 pp.

Bukchin B.G. (1990) The dependence of surface waves phase velocity on boundaries depths, Computational Seismology, vol.22.

Kennett B.L.N. (1974) Reflections, rays and reverberations, Bull. Seism. Soc. Am., v.28, p.97-110.

Kennett B.L.N. (1981) Elastic wave propagation in stratified media, Adv. in Appl. Mech., v.21, p.80-167.

Kerry N.J. (1981) The synthesis of seismic surface waves, Geoph. J. R. Astr. Soc., v.64, p.425-446.

Keilis-Borok V.I. (ed.) (1989) Seismic surface waves in a laterally inhomogeneous Earth, D. Reidel Publishing Company, 278 pp.

Kushnir A.F., Levshin A.L. and Lokchtanov D.E. (1988) Determination of velocity structure from surface wave spectra by methods of nonlinear optimization, Comput. Seismol., Vol.21, p.137-152.

Kushnir A.F., Lokchtanov D.E. (1988) The statistical approach to inversion of surface wave records, Comput. Seismol., Vol.21., p. 124-136.

Kushnir A.F., Levshin A.L. and Lokchtanov D.E. (1988) Determination of a regional velocity structure from surface wave seismograms recorded at a set of stations, In A.Vogel, R.Gorenflo, B.Kummer, C.Ofoegbu (Eds.) Inverse Modeling in Exploration Geophysics, p.489-498, Friedr. Vieweg & Sohn.

Lokchtanov D.E., Trusov A.V. (1989) Determination of the crustal velocity structure between the NARS-s stations, Comput. Seismol., Vol. 22, p.153-160.

Polak E. (1985) Computational methods in optimization, New York: Academic Press, 56 pp.