A strategy for nonlinear elastic inversion
of seismic reflection data

Albert Tarantola*

ABSTRACT

The problem of interpretation of seismic reflection data can be posed with sufficient generality using the concepts of inverse theory. In its roughest formulation, the inverse problem consists of obtaining the Earth model for which the predicted data best fit the observed data. If an adequate forward model is used, this best model will give the best images of the Earth's interior.

Three parameters are needed for describing a perfectly elastic, isotropic, Earth: the density $\rho(x)$ and the Lamé parameters $\lambda(x)$ and $\mu(x)$, or the density $\rho(x)$ and the P-wave and S-wave velocities $\alpha(x)$ and $\beta(x)$. The choice of parameters is not neutral, in the sense that although theoretically equivalent, if they are not adequately chosen the numerical algorithms in the inversion can be inefficient. In the long (spatial) wavelengths of the model, adequate parameters are the P-wave and S-wave velocities, while in the short (spatial) wavelengths, P-wave impedance, S-wave impedance, and density are adequate. The problem of inversion of waveforms is highly nonlinear for the long wavelengths of the velocities, while it is reasonably linear for the short wavelengths of the impedances and density. Furthermore, this parameterization defines a highly hierarchical problem: the long wavelengths of the P-wave velocity and short wavelengths of the P-wave impedance are much more important parameters than their counterparts for S-waves (in terms of interpreting observed amplitudes), and the latter are much more important than the density. This suggests solving the general inverse problem (which must involve all the parameters) by first optimizing for the P-wave velocity and impedance, then optimizing for the S-wave velocity and impedance, and finally optimizing for density.

The first part of the problem of obtaining the long wavelengths of the P-wave velocity and the short wavelengths of the P-wave impedance is similar to the problem solved by present industrial practice (for accurate data interpretation through velocity analysis and "prestack migration"). In fact, the method proposed here produces (as a byproduct) a generalization to the elastic case of the equations of "prestack acoustic migration."

Once an adequate model of the long wavelengths of the P-wave velocity and of the short wavelengths of the P-wave impedance has been obtained, the data residuals should essentially contain information on S-waves (essentially P-S and S-P converted waves). Once the corresponding model of S-wave velocity (long wavelengths) and S-wave impedance (short wavelengths) has been obtained, and if the remaining residuals still contain information, an optimization for density should be performed (the short wavelengths of impedances do not give independent information on density and velocity independently).

Because the problem is nonlinear, the whole process should be iterated to convergence; however, the information from each parameter should be independent enough for an interesting first solution.

INTRODUCTION

Modeling of seismic waveforms can be performed using numerical methods (e.g., finite differencing). Large-scale parallelism in the computers of the near future will make modeling accurate and inexpensive. Interpretation of industrial seismic data using inverse methods will probably become routine, and the stack-plus-migration method will become an ancient technique. Imaging will not be based on "principles," but on well-posed questions about the properties of the Earth's interior.

At present, it is urgent to develop the theory, to check approximations made in forward modeling and to study strategies for inversion.

Seismic reflection data are nonlinearly related to the Earth's model parameters. Classical velocity analysis as usually performed in the petroleum industry suggests that it is often pos-

sible to obtain good enough initial models to reduce the remaining nonlinearities.

There has been a considerable amount of recent work in linearized inversion of multioffset seismic reflection data (Clayton and Stolt, 1981; Berkhout, 1984; Tarantola, 1984a; Bleistein et al., 1985; Ikelle et al., 1986).

Some work exists in nonlinear inversion. Using a generalized least-squares method, Tarantola (1984b) develops a theory for the nonlinear, multidimensional inversion of multioffset seismic reflection data in the acoustic approximation. Tarantola (1984c) extends the theory to the elastic (isotropic) case. A coherent inverse theory including attenuation is missing.

Passing from theory to numerical experimentation, one-dimensional (1-D) problems (i.e., problems where the medium parameters depend only on one dimension) must be distinguished from two-dimensional (2-D) and three-dimensional (3-D) problems. Although 3-D forward modeling is within the capabilities of today's computers (Edwards et al., 1985), no attempt has yet been made to solve the 3-D nonlinear inverse problem with multioffset data numerically.

For 1-D nonlinear inversion with multioffset data, Mora (1984) uses ray theory to obtain models of density and P-wave and S-wave velocities. His forward modeling technique has the advantage of being inexpensive, but unfortunately it does not take into account multiply reflected energy. He shows that by using only the vertical displacement at the surface, it is possible to obtain the three parameters describing the medium, at least with perfect (synthetic) data. Kolb et al. (1986) solve the same problem using a more powerful (and much more expensive) finite-difference forward modeling technique. In particular, they discuss the problem of extracting the long wavelengths of the medium; unfortunately, they limit their scope to an acoustic medium and only invert for a single parameter (velocity).

Gauthier et al. (1986) give the first realistic example of multidimensional, nonlinear inversion of multioffset seismic reflection data. Their work is also limited to a single parameter in an acoustic medium, but they demonstrate the feasibility of such an inversion (with present-day software and hardware). Chen (1985) performs a 2-D elastic inversion of multioffset seismic data, but with so few parameters ($\simeq 100$) and for such unusual low-frequency "seismograms" [$\simeq \exp(-t)$] that I am unable to ascertain that their method is applicable to realistic data. Large-scale inverse problems are qualitatively different from the discrete problems obtained using rough parameterizations.

The aims of this paper are to use recent results to enumerate the remaining difficulties, and to propose a reasonable strategy for nonlinear inversion of real seismic reflection data with modern computers. Part of the mathematics needed for a full understanding of this paper has been developed by Tarantola (1984c). The elastodynamic wave equation is well-known and described in many texts (e.g., Morse and Feshbach, 1953; Aki and Richards, 1980; Ben-Menahem and Singh, 1981), so it is not presented here. The particular equations needed for the inverse theory are available in Tarantola (1984c, 1986).

CHOOSING AN ADEQUATE PARAMETERIZATION

To describe an isotropic elastic Earth, in addition to the density $\rho(x)$ two more parameters are needed. The simplest

choice is the Lamé parameters $\lambda(x)$ and $\mu(x)$, which appear in the wave equation. The compressional (P)-wave velocity

$$\alpha(x) = \sqrt{\frac{\lambda(x) + 2\mu(x)}{\rho(x)}} \qquad (1a)$$

and the shear (S) wave velocity

$$\beta(x) = \sqrt{\frac{\mu(x)}{\rho(x)}} \qquad (1b)$$

have more direct physical meaning for problems involving wave propagation.

For inversion, all choices are not equivalent. For instance, to accelerate iterative inversion algorithms, it is important to select parameters whose a posteriori uncertainties will be as uncorrelated as possible. Assume, for instance, that $\rho(x)$, $q_1(x)$, and $q_2(x)$ are chosen for describing the Earth, and imagine an Earth model that is homogeneous but for three point diffractors. The first point diffractor is a perturbation of density only, the second is a perturbation of q_1 only, and the third is a perturbation of q_2 only. A wave sent into the Earth will be diffracted by these three points. A good choice of parameters will give three diffraction patterns which are as different as possible, to allow easy identification of the parameters.

From numerical experiments, I conclude that for a seismic reflection experiment, an adequate parameter choice is density $\rho(x)$, P-wave impedance

$$IP(x) = \rho(x)\alpha(x) = \sqrt{\rho(x)\left[\lambda(x) + 2\mu(x)\right]}, \qquad (1c)$$

and S-wave impedance

$$IS(x) = \rho(x)\beta(x) = \sqrt{\rho(x)\mu(x)}. \qquad (1d)$$

The corresponding diffraction patterns are shown in Figures 1 and 2. For comparison, the diffraction patterns correspond-

FIG. 1. For each type of diffractor considered in Figures 2, 3, and 4, the diffracted waves are classified as indicated here. The columns correspond to incident P-, SV-, and SH-waves, respectively, while the rows correspond to diffracted P-, SV-, and SH-waves. Arrows indicate the sense of first displacement of the incident wave.
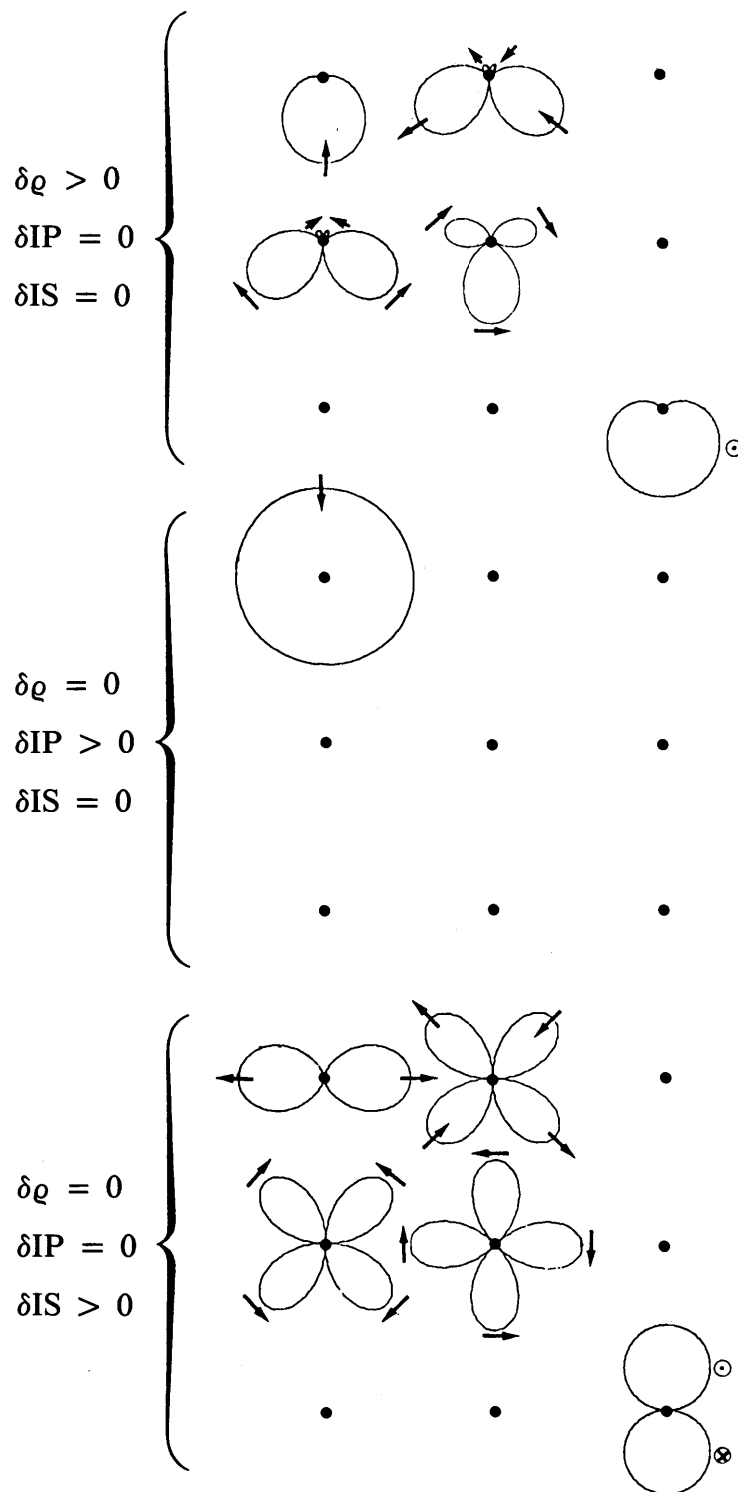
FIG. 2. The top section shows the amplitude diffraction patterns of a diffractor with a perturbation of density ρ but constant P-wave impedance IP and constant S-wave impedance IS. The middle of section shows the diffraction patterns of a diffractor with a perturbation of P-wave impedance IP but constant density ρ and S-wave impedance IS. The bottom section shows the diffraction patterns of a diffractor with a perturbation of S-wave impedance IS but constant ρ and IP. The waves are classified as in Figure 1. The arrows indicate the sense of first displacement of the diffracted wave. As is well-known, there are only P-P, P-SV, SV-P, SV-SV, and SH-SH conversions. The diffraction patterns for P-wave impedance diffractors and S-wave impedance diffractors (with constant density) are very different. A density diffractor (with constant impedances) only diffracts waves forward: this sort of diffractor will be hardly visible using surface seismic reflection data with moderate offsets.
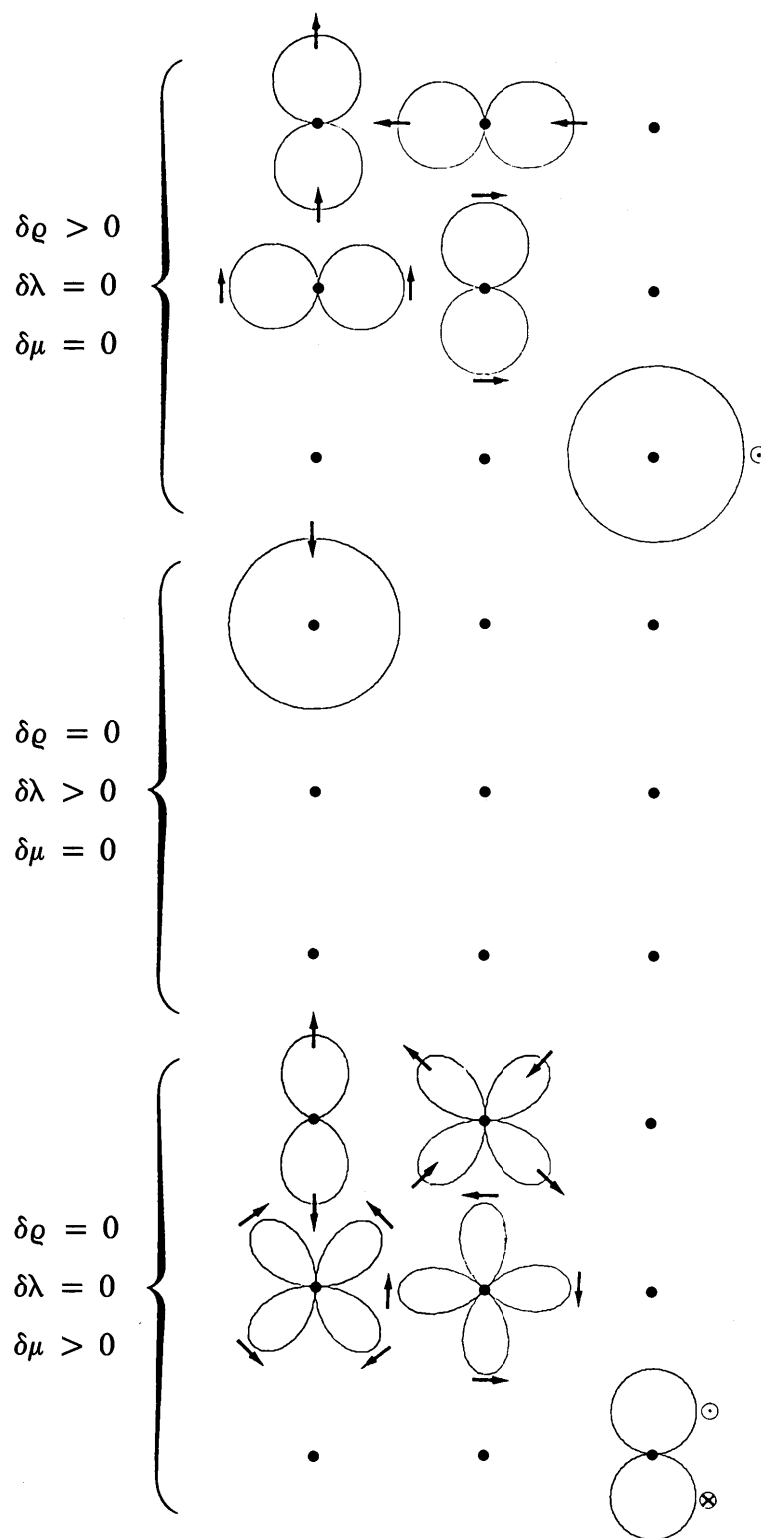
FIG. 3. Same as Figure 2 but for parameters ρ, λ, and μ (density and Lamé parameters). Different diffractors give similar diffraction patterns for moderate offsets. For instance, using only P-P waves and small offsets, the responses of the three types of diffractors considered are almost identical.
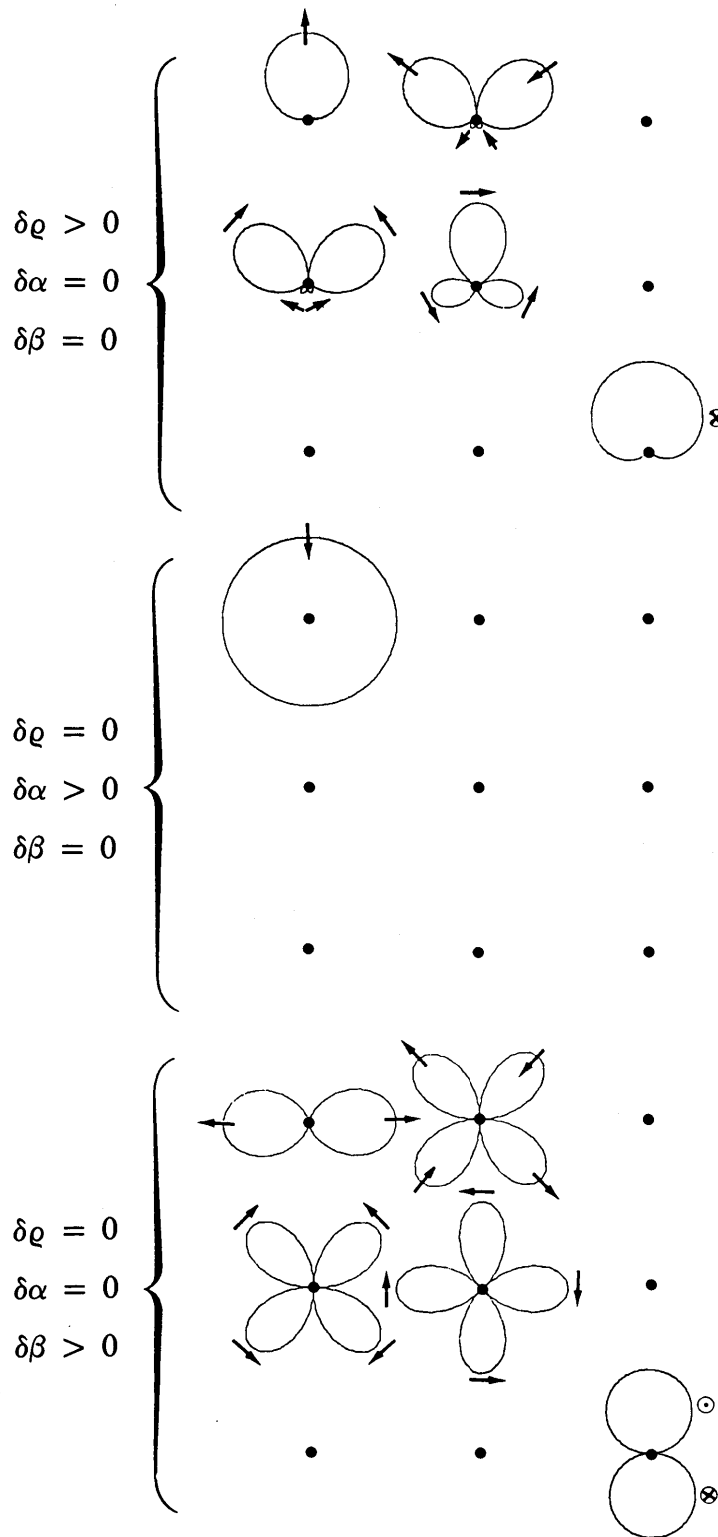
FIG. 4. Same as Figure 2 but for parameters ρ, α, and β (density and velocities of *P*- and *S*-waves). The situation is not as bad as in Figure 3, but using only *P-P* waves and small offsets, it is still difficult to distinguish between a density diffractor (with constant α and β) and a *P*-wave velocity diffractor (with constant ρ and β).

ing to a parameterization in density $\rho(\mathbf{x})$ and velocities $\alpha(\mathbf{x})$, $\beta(\mathbf{x})$, or to a parameterization in density $\rho(\mathbf{x})$ and Lamé parameters $\lambda(\mathbf{x})$ and $\mu(\mathbf{x})$, are shown in Figures 3 and 4. It can be seen that the backward diffraction patterns for small offsets are quite different for the choice of parameters in Figure 2, and they are more similar for the choice of parameters in Figures 3 and 4. For obtaining the diffraction patterns, I used a linearized solution of the wave equation with respect to a homogeneous medium, for parameter perturbations of the form $K\delta(\mathbf{x} - \mathbf{x}_0)$. Some (different) diffraction patterns are also shown in Sato (1984), Wu and Aki (1985a, b), and Wu and Ben-Menahem (1985).

As discussed below, only the short wavelengths of the impedances and density will be adequately resolved using a typical seismic reflection data set. This means that the parameters that will be truly resolved are the (approximately vertical) *gradients* of impedances and density. However, it is simpler to use impedances and density as parameters than their gradients; since the relationship between a parameter and its gradient is linear, the difference in resolution is not important.

NONLINEAR INVERSION OF MULTIOFFSET SEISMIC REFLECTION DATA

Notation

Consider a typical seismic reflection experiment. In what follows, $\mathbf{x}_r$ denotes a generic receiver position and $\mathbf{x}_s$ denotes a generic "shot" position. A typical seismic reflection data set may be represented by

$$\left[u^i(\mathbf{x}_r, t; \mathbf{x}_s) \right], \tag{2}$$

where $i = 1, 2, 3, r = 1, 2, \ldots, 0 \leq t \leq T$, and $s = 1, 2, \ldots$. The time variable t is reset to zero at each shot.

The least-squares criterion of goodness of fit

Define the best Earth's model $[\mathrm{IP}(\mathbf{x}), \mathrm{IS}(\mathbf{x}), \rho(\mathbf{x})]$ as the model that minimizes the least-squares expression

$$S(\mathrm{IP}, \mathrm{IS}, \rho)$$

$$= \frac{1}{2} \left(\| \mathbf{u}_{\mathrm{cal}} - \mathbf{u}_{\mathrm{obs}} \|^2 + \| \mathrm{IP} - \mathrm{IP}_{\mathrm{prior}} \|^2 \right.$$

$$\left. + \| \mathrm{IS} - \mathrm{IS}_{\mathrm{prior}} \|^2 + \| \rho - \rho_{\mathrm{prior}} \|^2 \right), \tag{3}$$

(the factor 1/2 allows subsequent simplifications), where

$$\| \mathbf{u}_{\mathrm{cal}} - \mathbf{u}_{\mathrm{obs}} \|^2$$

$$= \sum_s \sum_r \int_0^T dt \int_0^T dt' \left[u^i(\mathbf{x}_r, t; \mathbf{x}_s)_{\mathrm{cal}} - u^i(\mathbf{x}_r, t; \mathbf{x}_s)_{\mathrm{obs}} \right]$$

$$\times W^{ij}(t, t'; \mathbf{x}_r, \mathbf{x}_s) \left[u^j(\mathbf{x}_r, t'; \mathbf{x}_s)_{\mathrm{cal}} - u^j(\mathbf{x}_r, t'; \mathbf{x}_s)_{\mathrm{obs}} \right], \tag{4a}$$

$$\| \mathrm{IP} - \mathrm{IP}_{\mathrm{prior}} \|^2$$

$$= \int_V dV(\mathbf{x}) \int_V dV(\mathbf{x}') \left[\mathrm{IP}(\mathbf{x}) - \mathrm{IP}(\mathbf{x})_{\mathrm{prior}} \right]$$

$$\times W_p(\mathbf{x}, \mathbf{x}') \left[\mathrm{IP}(\mathbf{x}') - \mathrm{IP}(\mathbf{x}')_{\mathrm{prior}} \right], \tag{4b}$$

$$\| \mathrm{IS} - \mathrm{IS}_{\mathrm{prior}} \|^2$$

$$= \int_V dV(\mathbf{x}) \int_V dV(\mathbf{x}') \left[\mathrm{IS}(\mathbf{x}) - \mathrm{IS}(\mathbf{x})_{\mathrm{prior}} \right]$$

$$\times W_s(\mathbf{x}, \mathbf{x}') \left[\mathrm{IS}(\mathbf{x}') - \mathrm{IS}(\mathbf{x}')_{\mathrm{prior}} \right], \tag{4c}$$

and

$$\| \rho - \rho_{\mathrm{prior}} \|^2$$

$$\int_V dV(\mathbf{x}) \int_V dV(\mathbf{x}') \left[\rho(\mathbf{x}) - \rho(\mathbf{x})_{\mathrm{prior}} \right]$$

$$\times W_\rho(\mathbf{x}, \mathbf{x}') \left[\rho(\mathbf{x}') - \rho(\mathbf{x}')_{\mathrm{prior}} \right]. \tag{4d}$$

Here $u^i(\mathbf{x}_r, t; \mathbf{x}_s)_{\mathrm{cal}}$ represents the data predicted by the model $[\mathrm{IP}(\mathbf{x}), \mathrm{IS}(\mathbf{x}), \rho(\mathbf{x})]$ and $W^{ij}(t, t'; \mathbf{x}_r, \mathbf{x}_s)$, $W_p(\mathbf{x}, \mathbf{x}')$, $W_s(\mathbf{x}, \mathbf{x}')$, and $W_\rho(\mathbf{x}, \mathbf{x}')$ represent weighting functions.

The physical interpretation of criterion (3) is simple: one wishes a model $[\mathrm{IP}(\mathbf{x}), \mathrm{IS}(\mathbf{x}), \rho(\mathbf{x})]$ not too far from the a priori model $[\mathrm{IP}(\mathbf{x})_{\mathrm{prior}}, \mathrm{IS}(\mathbf{x})_{\mathrm{prior}}, \rho(\mathbf{x})_{\mathrm{prior}}]$ such that the predicted data are not too far from the observed data. The constraint that the final model cannot be too far from some a priori model is necessary for avoiding a possibly ill-posed problem: if very different models give approximately the same seismograms, I prefer the simplest model.

When so defined, the problem is fully nonlinear (the best model is defined without invoking any linear approximation of the basic equations). In particular, I do not use the Born approximation. Note that because the computed seismograms are *nonlinear* functions of the model parameters, the functional (3) is a *nonquadratic* function of the parameters.

Choice of the weighting functions

In the context of least-squares, a weighting function is the integral kernel of the inverse of a covariance operator. For instance, if $C(\mathbf{x}, \mathbf{x}')$ is a covariance function, the associated weighting function $W(\mathbf{x}, \mathbf{x}')$ verifies

$$\int_V dV(\mathbf{x}')C(\mathbf{x}, \mathbf{x}')W(\mathbf{x}', \mathbf{x}'') = \delta(\mathbf{x} - \mathbf{x}''). \tag{5}$$

Taking the covariance function

$$C(\mathbf{x}, \mathbf{x}') = C(x, y, z, x', y', z')$$

$$= K\delta(x - x')\delta(y - y') \min(z, z') \tag{6}$$

gives (Tarantola, 1986)

$$\| \phi - \phi_{prior} \|^2$$

$$= \int_V dV(\mathbf{x}) \int_V dV(\mathbf{x'}) \left[\phi(\mathbf{x}) - \phi(\mathbf{x})_{prior} \right]$$

$$\times W(\mathbf{x}, \mathbf{x'}) \left[\phi(\mathbf{x'}) - \phi(\mathbf{x'})_{prior} \right],$$

$$= \frac{1}{K} \int_V dV(\mathbf{x}) \left[\frac{\partial \phi}{\partial z}(x, y, z) - \frac{\partial \phi_{prior}}{\partial z}(x, y, z) \right]^2. \quad (7)$$

which is an adequate norm to impose for impedances or density in the Earth. The final model should not be close to the initial model, but the vertical gradient of the model should be close to the vertical gradient of the a priori model. For instance, taking a homogeneous a priori model in impedances and density, the norm (7) will require the final model to have small vertical gradients. The random process with covariance function (6) corresponds to a "random walk" in the z direction. Godfrey et al. (1980) suggest use of more general Markov chains. In the x-y directions, the random process with covariance function (6) is white noise. More general a priori constraints can be imposed by choosing adequate covariance functions; in particular, it is easy to impose lateral smoothness on the model by taking

$$C(\mathbf{x}, \mathbf{x'}) = C(x, y, z, x', y', z')$$

$$= K \exp \left[- \frac{\sqrt{(x - x')^2 + (y - y')^2}}{L} \right] \min(z, z'), \quad (8)$$

where L represents the length over which the model has to be smooth.

If there are uncorrelated errors in the data set, depending upon time or source and receiver positions, then

$$C^{ij}(t, t'; \mathbf{x}_r, \mathbf{x}_s) = \sigma^2(\mathbf{x}, t; \mathbf{x}_s)\delta^{ij}\delta(t - t'), \quad (9)$$

and

$$\| \mathbf{u}_{cal} - \mathbf{u}_{obs} \|^2 = \sum_s \sum_r \int_0^T dt$$

$$\times \sum_{i=1}^3 \frac{\left[u^i(\mathbf{x}_r, t; \mathbf{x}_s)_{cal} - u^i(\mathbf{x}_r, t; \mathbf{x}_s)_{obs} \right]^2}{\sigma^2(\mathbf{x}_r, t; \mathbf{x}_s)}. \quad (10)$$

Usually, only the vertical component u^3 is recorded. The sum over i then disappears from equation (10).

Methods of resolution

There are two classes of methods for minimizing a functional. The first has methods which extensively explore the model space. The exploration can be systematic or random (Monte Carlo). The inverse seismic problem has too many degrees of freedom for these methods to be useful with present-day computers. The second class contains iterative methods which, using the local properties (at a given point of the model space) of the functional to be minimized, define a "descent direction" along which a new and better point of the model space will be obtained. The more efficient methods are generally gradient methods. There are a number of good books on nonquadratic optimization (e.g., Céa, 1971; Fletcher, 1980; Scales, 1985). Some methods are reviewed in Tarantola (1984c, 1986). My

personal experience in the present problem suggests the following strategy.

First, as for all nonlinear problems, it is important to start iterating at a point as close as possible to the final solution, to minimize the nonlinearity of the problem. In the present context, this means starting from a model for which the long wavelengths of the P-wave and S-wave velocities are reasonably correct.

The three parameters IP(x), IS(x), and ρ(x) have been chosen to be as independent as possible. Furthermore, these parameters have very different importance. Most of the data features can be explained with P-waves alone, suggesting that the iteration starts with a gradient method for the P-wave impedance alone (i.e., maintaining fixed S-wave impedance and density). This requires a reasonably good model of the long wavelengths of the P velocity. Later I show that each iteration for the gradient strongly resembles a migration (for unstacked data).

Once a good model IP(x) has been obtained, the remaining data residuals will contain S-waves. If a reasonably good model for the long wavelengths of the S velocity can be obtained, then some gradient iterations to model S-wave impedance should be performed. The remaining residuals may contain some information on the short wavelengths of the density. Some gradient iterations for the density will end the process.

Because the total problem is nonlinear, the entire process should in principle be iterated until convergence. However, since the chosen parameters are acceptably independent, I hope the model obtained after a single loop will be good enough, if the long wavelengths of the P-wave and S-wave velocities in the starting model are right.

The long wavelengths

Claerbout (1985) suggests that normal seismic data sets contain information on the long wavelengths of the model which is totally independent of the information on the short wavelengths.

Gauthier et al. (1986) give an example of nonlinear inversion of multioffset seismic reflection data for a 2-D model, using the acoustic approximation. However, one of their negative conclusions is that the gradient methods have an extremely poor convergence rate if the starting model does not contain the long wavelengths of the true model.

Kolb et al. (1986) solve a nonlinear acoustic inverse problem with multioffset data, for a depth-dependent medium; they also use a gradient method. They show that if the first iterations are performed with a data set that has been severely low-pass filtered (i.e., using only the very low frequencies), then an adequate model for the long wavelengths may be obtained. It is not clear if such a method will work well with real data. If so, it allows use of the gradient methods from the beginning of the interpretation. If it does not work well on real data, the starting models of the long wavelengths of the P velocity and S velocity have to be obtained using an independent method, such as a τ-p method (Phinney et al., 1981; Carrion and Kuo, 1984) or a traveltime inversion (Nercessian et al., 1984). Stork and Clayton (1985) suggest a kind of iterative tomographic migration method of reconstruction. The use of methods based on Rytov's approximation (e.g., Devaney, 1984) is also promising.

In what follows, it is assumed that an Earth model $[\rho_0(x),$ $IP_0(x), IS_0(x)]$ that contains the long-wavelength component of the P and S velocities is given.

Optimization of the *P*-wave impedance

Denote by $\rho(x)_n$, $IP(x)_n$, and $IS(x)_n$ the model obtained at the nth iteration, which needs to be further optimized for the impedance $IP(x)$. I use a gradient iterative method which gives models $IP(x)_{n+1}, IP(x)_{n+2}, \ldots$.

Using the method from Tarantola (1984c, 1986), an iteration of the steepest-descent method is performed through the following equations (interpreted below):

$$\delta\hat{u}^i(x, t; x_s)_n = \frac{u^i(x_r, t; x_s)_n - u^i(x_r, t; x_s)_{obs}}{\sigma^2(x_r, t; x_s)}, \tag{11a}$$

$$\delta\Psi^i(x, t; x_s)_n = \sum_r \Gamma^{ij}(x, 0; x_r, t)_n * \delta\hat{u}^j(x_r, t; x_s)_n, \tag{11b}$$

$$\delta I\hat{P}(x)_n = -2\alpha(x)_n \sum_s \int_0^T dt\, u^{ii}(x, t; x_s)_n \times \delta\Psi^{jj}(x, t; x_s)_n, \tag{11c}$$

$$\delta IP(x)_n = \int dx'\, C_{Ip}(x, x')\delta I\hat{P}(x')_n, \tag{11d}$$

$$\Delta IP(x)_n = \text{Precond}\left[\delta IP(x)_n\right], \tag{11e}$$

and

$$IP(x)_{n+1} = IP(x)_n - a_n\left[\Delta IP(x)_n + IP(x)_n - IP(x)_{prior}\right], \tag{11f}$$

where a_n is the real constant which makes $S(IP_{n+1}, IS_n, \rho_n)$ minimum, and which can be analytically estimated (see Tarantola 1984c) or simply obtained by trial and error.

I now turn to the physical interpretation of this result.

Equation (11a).—$u^i(x_r, t; x_s)_n$ are the data predicted for the model IP_n, IS_n, ρ_n. $\lambda(x)_n$ and $\mu(x)_n$ are the Lamé parameters corresponding to the model IP_n, IS_n, ρ_n [through equations (1c) and (1d)]. By definition, $u^i(x_r, t; x_s)_n$ is the solution of the set of differential equations

$$\rho(x)_n \frac{\partial^2 u^i}{\partial t^2}(x, t; x_s)_n$$

$$-\frac{\partial}{\partial x^i}\left[\lambda(x)_n u^{kk}(x, t; x_s)_n\right]$$

$$-2\frac{\partial}{\partial x^j}\left[\mu(x)_n u^{ij}(x, t; x_s)_n\right] = 0, \tag{12a}$$

$$\lambda(x)_n u^{kk}(x, t; x_s)_n n^i(x)$$

$$+2\mu(x)_n u^{ij}(x, t; x_s)_n n^j(x)$$

$$= T^i(x, t; x_s), \qquad x \in S, \tag{12b}$$

$$u^i(x, 0; x_s) = 0, \tag{12c}$$

and

$$\hat{u}^i(x, 0; x_s) = 0, \tag{12d}$$

where it is assumed that the source of seismic waves is a vibrator whose action is completely described by the surface traction $T^i(x, t; x_s)$. For instance, a simplistic description could be $T^i(x, t; x_s) = \delta^{i3}S(t)\delta(x - x_s)$, thus assuming a vertical point source located at $x = x_s$ with time function $S(t)$. The actual computation of the field $u^i(x, t; x_s)_n$ may be made using any numerical method, such as finite differencing (Gauthier et al., 1986).

$\sigma^2(x_r, t; x_s)$ represents the estimated (squared) error at time t of the displacement measured at x_r for the source at x_s. $\delta\hat{u}^i(x_r, t; x_s)_n$ then clearly represents the *weighted residuals*.

Equation (11b).—Define the field $\delta\Psi^i(x, t; x_s)_n$ by the equations

$$\rho(x)_n \frac{\partial^2 \delta\Psi^i}{\partial t^2}(x, t; x_s)_n$$

$$-\frac{\partial}{\partial x^i}\left[\lambda(x)_n\delta\Psi^{kk}(x, t; x_s)_n\right]$$

$$-2\frac{\partial}{\partial x^j}\left[\mu(x)_n\delta\Psi^{ij}(x, t; x_s)_n\right] = 0, \tag{13a}$$

$$\lambda(x)_n\delta\Psi^{kk}(x, t; x_s)_n n^i(x)$$

$$+2\mu(x)_n\delta\Psi^{ij}(x, t; x_s)_n n^j(x)$$

$$= \sum_r \delta(x - x_r)\delta\hat{u}^i(x_r, t; x_s)_n, \qquad x \in S, \tag{13b}$$

$$\delta\Psi^i(x, T; x_s)_n = 0, \tag{13c}$$

and

$$\delta\Psi^i(x, T; x_s)_n = 0, \tag{13d}$$

where $\delta\Psi^i(x, t; x_s)_n$ satisfies homogeneous *final* conditions (13c) and (13d), instead of initial conditions. The "sources" of the field $\delta\Psi^i(x, t, x_s)$ are the weighted residuals $\delta\hat{u}^i(x_r, t; x_s)_n$, acting as if they were tractions (13b), all radiating in-phase.

Using the representation theorem [Aki and Richards, 1980, eq. (27)] with reversed time allows the following compact representation for $\delta\Psi^i(x, t; x_s)$:

$$\delta\Psi^i(x, t; x_s)_n = \sum_r \Gamma^{ij}(x, 0; x_r, t)_n * \delta\hat{u}^i(x_r, t; x_s)_n, \tag{27}$$

which corresponds to equation (11b). $\Gamma^{ij}(x, 0; x_r, t)_n$ represents the Green's function in the current model, satisfying homogeneous initial and (traction) boundary conditions. (The Green's function is useful for analytical developments or to simplify notation, but it never has to be used explicitly in numerical computations.)

The field $\delta\Psi^i(x, t; x_s)_n$ can, for instance, be numerically obtained using a finite-difference code, with time running backward from $t = T$ to $t = 0$, and where for a given shotpoint x_s, I consider virtual sources, one at each receiver, radiating the weighted residuals backward in time. See Gauthier et al. (1986) for a numerical implementation in an acoustic example.

Note that the physical dimension of $\delta\Psi^i(x, t; x_s)_n$ is not a displacement, because the right side of equation (13b) is not a traction. Because the surface "tractions" corresponding to

$\delta\Psi^i(\mathbf{x}, t; \mathbf{x}_s)_n$ equal the weighted residuals, this field is closely related to the *missing field* at iteration n, i.e., a field whose surface displacements equal the residuals at the receiver locations. I liberally name $\delta\Psi^i(\mathbf{x}, t; \mathbf{x}_s)_n$ the "current missing field."

Equation (11c).—This is the most important of the equations, because this is the inversion. After some corrections [equations (11d)–(11f)] $\delta I\hat{P}(\mathbf{x})_n$ will essentially be the correction to be applied to $\mathbf{IP}_n$ for obtaining $\mathbf{IP}_{n+1}$ [as shown by equation (11f)]. Equation (11c) shows that this correction at a given point $\mathbf{x}$, for given shot $\mathbf{x}_s$, equals the time correlation of the dilatation $u^{ii}(\mathbf{x}, t; \mathbf{x}_s)_n$ of the current predicted field with the dilatation $\delta\Psi^{jj}(\mathbf{x}, t; \mathbf{x}_s)_n$ of the current missing field. The physical interpretation is as follows. If for a given source point $\mathbf{x}_s$ and at a given point $\mathbf{x}$, the dilatation of the current predicted field is time-correlated with the dilatation of the missing field, this missing field should be created by adding a P impedance diffractor at point $\mathbf{x}$. This interpretation is strikingly similar to the imaging principle of Claerbout (1971), but here it is in an elastic context and results from a general optimization criterion.

Equation (11d).—The "migrated" field $\delta I\hat{P}(\mathbf{x})_n$ is operated with the covariance operator incorporating a priori information. If a covariance function such as equation (8) is used, this equation corresponds to a convolution over the x and y coordinates and a sum over the z coordinate which in fact corresponds to taking twice the primitive of $\delta I\hat{P}(\mathbf{x})_n$ with respect to z. The parameter K in equation (8) controls the tradeoff between the importance of the a priori information and the information obtained from the data set.

Equation (11e).—As in all gradient methods, some preconditioning may greatly speed convergence. At least, a preconditioning operator has to simulate the action of the inverse Hessian appearing in the Newton-like methods of optimization. Using physical intuition, operators can be defined that may be better than the inverse Hessian. The simplest operator in this example corresponds to a correction for spherical divergence of waves (Gauthier et al., 1986), i.e., to a multiplication by z^n, where $n \simeq 1$ for a 2-D problem and $n \simeq 2$ for a 3-D problem.

Equation (11f).—The new model $IP(\mathbf{x})_n$ is obtained here. An optimum value of a_n (for which the cost function is a minimum) is obtained by trial and error.

Each iteration corresponds to a sort of generalized elastic "prestack" migration. A few iterations should suffice (if it is not necessary to improve the long wavelengths). Readers not interested in inversion, but only in prestack migration, may consider these equations as a serious candidate for replacing acoustic migration equations.

Optimization of the S-wave impedance

Turning now to S-wave impedance, denote by $\rho(\mathbf{x})_n$, $IP(\mathbf{x})_n$, and $IS(\mathbf{x})_n$ the model already obtained, which must be further optimized for the impedance $IS(\mathbf{x})$. The gradient iterative method gives models $IS(\mathbf{x})_{n+1}$, $IS(\mathbf{x})_{n+2}, \ldots$.

An iteration of the steepest-descent method is performed through the following equations:

$$\delta\hat{u}^i(\mathbf{x}_r, t; \mathbf{x}_s)_n = \frac{u^i(\mathbf{x}_r, t; \mathbf{x}_s)_n - u^i(\mathbf{x}_r, t; \mathbf{x}_s)_{\text{obs}}}{\sigma^2(\mathbf{x}_r, t; \mathbf{x}_s)}, \tag{14a}$$

$$\delta\Psi^i(\mathbf{x}, t; \mathbf{x}_s)_n = \sum_r \Gamma^{ij}(\mathbf{x}, 0; \mathbf{x}_r, t)_n * \delta\hat{u}^j(\mathbf{x}_r, t; \mathbf{x}_s)_n, \tag{14b}$$

$$\delta I\hat{S}(\mathbf{x})_n = -4\beta(\mathbf{x})_n \sum_s \int_0^T dt$$
$$\times \left[u^{km}(\mathbf{x}, t; \mathbf{x}_s)_n \Psi^{km}(\mathbf{x}, t; \mathbf{x}_s)_n \right.$$
$$\left. - u^{ii}(\mathbf{x}, t; \mathbf{x}_s)_n \Psi^{jj}(\mathbf{x}, t; \mathbf{x}_s)_n \right], \tag{14c}$$

$$\delta IS(\mathbf{x})_n = \int d\mathbf{x}' C_{Is}(\mathbf{x}, \mathbf{x}') \delta I\hat{S}(\mathbf{x}')_n, \tag{14d}$$

$$\Delta IS(\mathbf{x})_n = \text{Precond}\left[\delta IS(\mathbf{x})_n \right], \tag{14e}$$

and

$$IS(\mathbf{x})_{n+1} = IS(\mathbf{x})_n - a_n \left[\Delta IS(\mathbf{x})_n + IS(\mathbf{x})_n - IS(\mathbf{x})_{\text{prior}} \right], \tag{14f}$$

where a_n is the real constant which makes $S(\mathbf{IP}_n, \mathbf{IS}_{n+1}, \boldsymbol{\rho}_n)$ minimum. The physical interpretation is as for the P-wave impedance.

Optimization of the density

Finally for the density, denote by $\rho(\mathbf{x})_n$, $IP(\mathbf{x})_n$, and $IS(\mathbf{x})_n$ the model already obtained, which must be further optimized for the density $\rho(\mathbf{x})$. The gradient iterative method gives models $\rho(\mathbf{x})_{n+1}$, $\rho(\mathbf{x})_{n+2}, \ldots$.

An iteration of the steepest-descent method is performed through the following equations:

$$\delta\hat{u}^i(\mathbf{x}_r, t; \mathbf{x}_s)_n = \frac{u^i(\mathbf{x}_r, t; \mathbf{x}_s) - u^i(\mathbf{x}_r, t; \mathbf{x}_s)_{\text{obs}}}{\sigma^2(\mathbf{x}_r, t; \mathbf{x}_s)}, \tag{15a}$$

$$\delta\Psi^i(\mathbf{x}, t; \mathbf{x}_s)_n = \sum_r \Gamma^{ij}(\mathbf{x}, 0; \mathbf{x}_r, t)_n * \delta\hat{u}^j(\mathbf{x}_r, t; \mathbf{x}_s)_n, \tag{15b}$$

$$\delta\hat{\rho}(\mathbf{x})_n = \sum_s \int_0^T dt \left\{ \dot{u}^i(\mathbf{x}, t; \mathbf{x}_s)_n \delta\dot{\Psi}^i(\mathbf{x}, t; \mathbf{x}_s)_n \right.$$
$$+ \left[\alpha^2(\mathbf{x})_n - 2\beta^2(\mathbf{x})_n \right] u^{ii}(\mathbf{x}, t; \mathbf{x}_s)_n \delta\Psi^{jj}(\mathbf{x}, t; \mathbf{x}_s)_n$$
$$\left. + 2\beta^2(\mathbf{x})_n u^{km}(\mathbf{x}, t; \mathbf{x}_s)_n \delta\Psi^{km}(\mathbf{x}, t; \mathbf{x}_s)_n \right\}, \tag{15c}$$

$$\delta\rho(\mathbf{x}) = \int d\mathbf{x}' C_\rho(\mathbf{x}, \mathbf{x}') \delta\hat{\rho}(\mathbf{x}')_n, \tag{15d}$$

$$\Delta\rho(\mathbf{x})_n = \text{Precond}\left[\delta\rho(\mathbf{x})_n \right], \tag{15e}$$

and

$$\rho(\mathbf{x})_{n+1} = \rho(\mathbf{x})_n - a_n \left[\Delta\rho(\mathbf{x})_n + \rho(\mathbf{x})_n - \rho(\mathbf{x})_{\text{prior}} \right], \tag{15f}$$

where a_n is the real constant which makes $S(\mathbf{IP}_n, \mathbf{IS}_n, \rho_{n+1})$ minimum. The physical interpretation is the same as for the P-wave impedance.

DISCUSSION

With the previous algorithms, nonlinear inversion of seismic reflection data requires computations which are quite similar to those in use today for accurate prestack migration. These algorithms are more expensive because they use a full elastodynamic wave equation instead of an approximate acoustic wave equation.

Although the equations given here are for three-component data, the restriction to vertical component data is obvious.

It is difficult to analyze errors in the solution of such an inverse problem. There are two sources of error, the simplest of which is due to noise in the data set. Generalized least-squares using a priori information on model parameters are known to be stable (Jackson, 1979; Tarantola and Valette, 1982), and the estimated errors as given by the a posteriori covariance operator in the model space seem to be reliable (as far as Gaussian statistics are acceptable for errors in the data set). Unfortunately, in a problem as large as the present one there is no computable method for estimating these errors (Tarantola, 1984c, 1986).

Errors more difficult to analyze are those due to discretization. All the main computations in the present method consist of wave propagation (both for solving the forward problem and for iterating in the inverse resolution), and these errors are relatively well-known (Alterman and Karal, 1968; Virieux, 1986).

Waves propagate in a 3-D Earth, and, even if the geology is invariant in the direction perpendicular to the survey line, accurate modeling of amplitudes can be obtained only by using 3-D models. Finite-difference approximations to the 3-D wave equation are possible (Edwards et al., 1985) but expensive with modern vector computers. The newest generation of parallel computers will probably make the modeling task feasible. For now, some adhoc corrections in the observed amplitudes may account for part of the effect. However, correct inversion is highly dependent upon forward modeling, and this could be a major limitation.

The source time function is not known exactly, and this is itself a full inverse problem. However, the relationship between observed amplitudes and the source time function is strictly linear, so the problem is not difficult.

Here I have used a simple, preconditioned steepest-descent method that is probably too rough to be efficient. More sophisticated methods using not only the gradient of the functional S, but also the Hessian (second derivative) have to be envisaged (see Tarantola, 1984c). Unfortunately no progress has been made in that direction. The Hessian is difficult to interpret and too complicated to use. The numerical experiments of Gauthier et al. (1986) suggest conjugate gradients to be more efficient than steepest descent, but good preconditioning is more important.

Here, I have used a least-squares (L_2-norm) criterion of goodness of fit. It is well-known (e.g., Claerbout and Muir, 1973) that this criterion is not as robust as the L_1-norm criterion for adequately handling blunders in a data set, but the

theory of large-scale inversion using an L_1-norm has not yet been developed.

CONCLUSION

Some numerical experiments suggest separation of the parameters describing the Earth in long spatial wavelengths and short spatial wavelengths. In the long wavelengths, the P-wave velocity $\alpha(\mathbf{x})$ and the S-wave velocity $\beta(\mathbf{x})$ are adequate parameters. In the short wavelengths, the P-wave impedance, the S-wave impedance, and the density are adequate.

The nonlinear inverse problem is set as the problem of minimizing a (nonquadratic) functional of the model parameters, essentially measuring the distance between observed and predicted data. Use of a generalized least-squares criterion allows introduction of a priori constraints in which the vertical gradients of the model have to be small so that the resulting equations are very simple. The minimization problem is solved iteratively, first optimizing the P-wave impedance, second optimizing the S-wave impedance, and then optimizing the density. Each partial optimization problem corresponds to a type of "generalized iterative elastic migration of unstacked data." Each iteration requires resolution of two forward problems per source.

ACKNOWLEDGMENTS

REFERENCES

Aki, K., and Richards, P., 1980, Quantitative seismology, 1: W. H. Freeman and Co.
Alterman, Z. S., and Karal, F. C., Jr., 1968, Propagation of elastic waves in layered media by finite difference methods: Bull., Seis. Soc. Am., 64, 367–398.
Ben-Menahem, A., and Singh, S. J., 1981, Seismic waves and sources: Springer-Verlag.
Berkhout, A. J., 1984, Multidimensional linearized inversion and seismic migration: Geophysics, 49, 1881–1895.
Bleistein, N., Cohen, J. K., and Hagin, F. G., 1985, Computational and asymptotic aspects of velocity inversion: Geophysics, 50, 1253–1265.
Carrion, P. M., and Kuo, J. T., 1984, A method for computation of velocity profiles by inversion of large offset records: Geophysics, 49, 1249–1258.
Céa, J., 1971, Optimisation, théorie et algorithmes: Dunod.
Chen, Y. M., 1985, Generalized pulse spectrum technique: Geophysics, 50, 1664–1675.
Claerbout, J. F., 1971, Toward a unified theory of reflector mapping: Geophysics, 36, 467–481.
———— 1985, Imaging the Earth's interior: Blackwell Scientific Publ.
Claerbout, J. F., and Muir, F., 1973, Robust modeling with erratic data: Geophysics, 38, 826–844.
Clayton, R. W., and Stolt, R. H., 1981, A Born-WKBJ inversion method for acoustic reflection data: Geophysics, 46, 1559–1567.
Devaney, A. J., 1984, Geophysical diffraction tomography: Inst. Electr. Electron. Eng., geosi. and remote sensing, Ge-22, 3–13.
Edwards, M., Hsiung, C., Kosloff, D., and Reshef, M., 1985, Elastic 3-D forward modeling by the Fourier method: Presented at the

55th Ann. Internat. Mtg., Soc. Explor. Geophys., Washington, D. C.

Fletcher, R., 1980, Practical methods of optimization, 1: Unconstrained optimization: John Wiley and Sons, Inc.

Gauthier, O., Virieux, J., and Tarantola, A., 1986, Nonlinear inversion of multioffset seismic reflection data: Geophysics, **51**, 1387–1403.

Godfrey, R., Muir, F., and Rocca, F., 1980, Modeling seismic impedance with Markov chains: Geophysics, **45**, 1351–1372.

Ikelle, L. T., Diet, J. P., and Tarantola, A., 1986, Linearized inversion of multioffset seismic reflection data in the ω-k domain: Geophysics, **51**, 1266–1276.

Jackson, D. D., 1979, The use of a priori data to resolve nonuniqueness in linear inversion: Geophys. J. Roy. Astr. Soc., **57**, 137–157.

Kolb, P., Collino, F., and Lailly, P., 1986, Prestack inversion of a 1-D medium: Special issue on seismic inversion, Proc., Inst. Electr. Electron. Eng., **64**, 498–508.

Mora, P., 1984, Elastic inversion using ray theory: Stanford Exploration Project, rep. no. 41, Stanford Univ.

Morse, P. M., and Feshbach, H., 1953, Methods of theoretical physics: McGraw-Hill Book Co.

Nercessian, A., Hirn, A., and Tarantola, A., 1984, Three-dimensional seismic transmission prospecting of the Mont-Dore volcano: Geophys. J. Roy. Astr. Soc., **76**, 307–315.

Phinney, R. A., Chowdhury, R. K., and Frazer, N. L., 1981, Transformation and analysis of record sections: J. Geophys. Res., **86**, 359–377.

Sato, H., 1984, Attenuation and envelope formation of three-component seismograms of small local earthquakes in randomly inhomogeneous lithosphere: J. Geophys. Res., **89**, 1221–1241.

Scales, L. E., 1985, Introduction to nonlinear optimization: Macmillan Publ. Co.

Stork, C., and Clayton, R. W., 1985, Iterative tomographic and migration reconstruction of seismic images: presented at the 55th Ann. Internat. Mtg., Soc. Explor. Geophys., Washington, D. C.

Tarantola, A., 1984a, Linearized inversion of seismic reflection data: Geophys. Prosp., **32**, 998–1015.

———— 1984b, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

———— 1984c, The seismic reflection inverse problem, in Santosa, F., Pao, Y.-H., Symes, W. W., and Holland, C., Eds., Inverse problems of acoustic and elastic waves: Soc. Industr. Appl. Math., 104–181.

———— 1986, Inverse problem theory. Methods for data fitting and model parameter estimation: Elsevier Science Publ. Co., in press.

Tarantola, A., and Valette, B., 1982, Generalized nonlinear inverse problems solved using the least squares criterion: Rev. Geophys. and Space Phys., **20**, 219–232.

Virieux, J., 1986, P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, **51**, 889–901.

Wu, R. S., and Aki, K., 1985a, Scattering characteristics of elastic waves by an elastic heterogeneity: Geophysics, **50**, 582–595.

———— 1985b, Elastic wave scattering by a random medium and the small scale inhomogeneities in the lithosphere: J. Geophys. Res., **90**, 10261–10273.

Wu, R. S., and Ben-Menahem, A., 1985, The elastodynamic near-field: Geophys. J. Roy. Astr. Soc., **81**, 609–621.