GEOPHYSICS, VOL. 49, NO. 8 (AUGUST 1984); P. 1259–1266.

Inversion of seismic reflection data in the acoustic approximation

Albert Tarantola*

ABSTRACT

The *nonlinear* inverse problem for seismic reflection data is solved in the acoustic approximation. The method is based on the generalized least-squares criterion, and it can handle errors in the data set and a priori information on the model. Multiply reflected energy is naturally taken into account, as well as refracted energy or surface waves. The inverse problem can be solved using an iterative algorithm which gives, at each iteration, updated values of bulk modulus, density, and time source function. Each step of the iterative algorithm essentially consists of a forward propagation of the actual sources in the current model and a forward propagation (backward in time) of the data residuals. The correlation at each point of the space of the two fields thus obtained yields the corrections of the bulk modulus and density models. This shows, in particular, that the general solution of the inverse problem can be attained by methods strongly related to the methods of migration of unstacked data, and commercially competitive with them.

INTRODUCTION

This is the second of a series of papers giving the solution of the inverse problem for seismic reflection data. In Tarantola (1984), herein referred as "paper I," I discussed the philosophy of inverse theory, compared to the philosophy of "migration" or "direct inversion." I demonstrated in that paper that the linearization of the forward problem leads to an inverse solution strongly related to the Kirchhoff migration method (French, 1974; Schneider, 1978). In this paper I attack the nonlinear acoustic problem, as a new step toward the solution of the nonlinear viscoelastic problem.

Although this paper is clearly a generalization of paper I, a special effort has been made to make it self-contained.

The objective of the paper is ambitious in the sense that I look for a method which is able to provide accurate models of the Earth starting with very crude (i.e., homogeneous) models. In addition, I want the method to handle waves other than the usual primary reflections. This means that the approach is necessarily nonlinear. I confess that the task appeared insurmountable at the beginning, but the computer time necessary for solving the problem has been decreasing by one order of magnitude per month, over many months. The current computing time is reasonable enough to justify this report.

THE FORWARD PROBLEM

I limit this paper to the acoustic approximation of the elastic wave equation. The generalization of the method to the elastic case will be developed later.

In the acoustic approximation, a medium is characterized by the density $\rho(\mathbf{r})$ and the bulk modulus $K(\mathbf{r})$. Given a source field $s(\mathbf{r}, t)$ and given initial and boundary conditions, the pressure field $p(\mathbf{r}, t)$ is uniquely defined by the acoustic wave equation

$$\left[\frac{1}{K(\mathbf{r})}\frac{\partial^2}{\partial t^2} - \operatorname{div}\left(\frac{1}{\rho(\mathbf{r})}\operatorname{grad}\right)\right]p(\mathbf{r}, t) = s(\mathbf{r}, t). \tag{1}$$

For short, one can denote by $\mathbf{K}$, $\boldsymbol{\rho}$, $\mathbf{s}$, and $\mathbf{p}$ the functions appearing in equation (1) considered as elements of a suitably chosen space. Formally, the solution of equation (1) can be written

$$\mathbf{p} = \mathbf{f}(\mathbf{K}, \boldsymbol{\rho}, \mathbf{s}), \tag{2}$$

where $\mathbf{f}$ represents a given, nonlinear operator. Throughout the paper I assume that one is able to solve equation (1). The results given in this paper will be valid for any method used for solving the forward problem, e.g., finite-differencing or ray-tracing methods.

Introducing the Green's function

$$\left[\frac{1}{K(\mathbf{r})}\frac{\partial^2}{\partial t^2} - \operatorname{div}\left(\frac{1}{\rho(\mathbf{r})}\operatorname{grad}\right)\right]g(\mathbf{r}, t; \mathbf{r}', t') = \delta(\mathbf{r} - \mathbf{r}')\delta(t - t'),$$

one can write

$$p(\mathbf{r}, t) = \int d\mathbf{r}' \int dt'\, g(\mathbf{r}, t; \mathbf{r}', t')s(\mathbf{r}', t'). \tag{3}$$

Since $\mathbf{K}$ and $\boldsymbol{\rho}$ are assumed independent of time, the Green's function will be invariant with respect to time-translation i.e.,

Manuscript received by the Editor February 25, 1983; revised manuscript received October 5, 1983.
*Laboratoire de Sismologie (LA CNRS 195), Institut de Physique au Globe de Paris, Univ. P. et M. Curie, 4 Place Jussieu, 75230, Paris Cedex 05, France.

$$g(\mathbf{r}, t; \mathbf{r}', t') = g(\mathbf{r}, t - t'; \mathbf{r}', 0) = g(\mathbf{r}, 0; \mathbf{r}', t' - t). \quad (4)$$

Equation (3) can then be rewritten as

$$p(\mathbf{r}, t) = \int d\mathbf{r}' \; g(\mathbf{r}, t; \mathbf{r}', 0) * s(\mathbf{r}', t), \quad (5)$$

where $*$ denotes time convolution.

Throughout this paper, equations (2) or (5) will represent the solution to the forward problem. Equation (2) has the advantage of allowing the solution of the forward problem to appear as the output of a *black box*, without reference to any numerical algorithm (finite-differencing or ray-tracing), while equation (5) allows an easy demonstration of the formulas used.

In the seismic reflection problem, actual sources can be considered as points in space, and if we assume that they are isotropic, they can simply be described using a source time function $S(t)$. Denoting by $\mathbf{r} = \mathbf{r}_s$ the source position,

$$s(\mathbf{r}, t) = \delta(\mathbf{r} - \mathbf{r}_s)S(t).$$

Denoting by $p(\mathbf{r}, t; \mathbf{r}_s)$ the pressure field due to a source located at $\mathbf{r} = \mathbf{r}_s$, equation (5) can be simplified to

$$p(\mathbf{r}, t; \mathbf{r}_s) = g(\mathbf{r}, t; \mathbf{r}_s, 0) * S(t).$$

Equation (2) will be rewritten

$$\mathbf{p} = \mathbf{f}(\mathbf{K}, \boldsymbol{\rho}, \mathbf{S}), \quad (6)$$

where I have replaced the source field $\mathbf{s}$ by the source time function $\mathbf{S}$.

THE DATA SET, THE A PRIORI INFORMATION ON THE MODEL

The solution of the forward problem allows the computation of the pressure field $p(\mathbf{r}, t; \mathbf{r}_s)$ for any value of $\mathbf{r}$ and t. Measurements of the actual pressure field are performed at discrete values of $\mathbf{r}$. Let $\mathbf{r} = \mathbf{r}_g$ represent a generic receiver position (g stands for "geophone"). The observations take then the form $p(\mathbf{r}_g, t; \mathbf{r}_s)$ where $\mathbf{r}_g$ and $\mathbf{r}_s$ belong to a *discrete* and *finite* set, while the variable t can either be considered continuous (analog recording) or discrete (digital recording). Actual measurements will give some definite values for $p(\mathbf{r}_g, t; \mathbf{r}_s)$ which will be named the "observed" values of the pressure field and which will be denoted by $p_0(\mathbf{r}_g, t; \mathbf{r}_s)$ or, for short, $\mathbf{p}_0$.

Experimental data are never perfect. A useful and rather general way for describing estimated uncertainties in a data set (due to noise) is the use of a covariance operator, which describes not only the estimated variance for each particular datum, but also the estimated correlation between errors. The most general covariance operator corresponding to this particular data set takes the form $C_p(\mathbf{r}_g, t; \mathbf{r}_s; \mathbf{r}'_g, t'; \mathbf{r}'_s)$. One particular example is

$$C_p(\mathbf{r}_g, t; \mathbf{r}_s \mid \mathbf{r}'_g, t'; \mathbf{r}'_s) = \sigma_{gs}^2 \delta_{gg'} \delta_{ss'} \delta_{tt'} \quad (7)$$

where σ_{gs}^2 represents the estimated error in the seismogram corresponding to the gth receiver for the sth source, and where errors are assumed to be uncorrelated. For short, the covariance operator will be denoted by C_p.

We will now try to introduce the a priori information about bulk modulus and density. By a priori I mean information which has been obtained independently of the observed values of the data set. The use of a priori information is useful to avoid instability in the inversion of data, which could otherwise arise in the present problem if, for instance, a given small region in the space was very poorly resolved by the data set, or if the data set could not resolve separately density and bulk modulus in a given region. However, one can expect that in the regions which are well covered by the seismic survey, the final solution will be practically independent of the a priori estimate. Nevertheless, I will try to give some rules for setting reasonable a priori models.

From an ideal point of view, one could collect a statistically significant collection of actual models for $\mathbf{K}$ and $\boldsymbol{\rho}$ from logging and seismic surveys in regions similar to the region presently under survey. In that case, and using the classical definitions of statistics, one could obtain the *mean model* $K_0(\mathbf{r})$ and $\rho_0(\mathbf{r})$ and the covariance functions $C_{KK}(\mathbf{r}, \mathbf{r}')$, $C_{\rho\rho}(\mathbf{r}, \mathbf{r}')$, and $C_{K\rho}(\mathbf{r}, \mathbf{r}')$. As I have not yet performed such a numerical experiment, I must, for the moment, be overconservative in the statement of the a priori information. That means that the a priori variances will be taken as very large, and that the covariances will be neglected.

An example of a priori information is the following:

$$K_0(\mathbf{r}) = K_0 = \text{const},$$

$$\rho_0(\mathbf{r}) = \rho_0 = \text{const},$$

$$S_0(t) = \text{any estimate of the source pulse},$$

$$C_{KK}(\mathbf{r}, \mathbf{r}') = \sigma_K^2 \exp\left\{-\frac{1}{2}\left[\frac{(x-x')^2 + (y-y')^2}{L_H^2} + \frac{(z-z')^2}{L_V^2}\right]\right\},$$

$$C_{\rho\rho}(\mathbf{r}, \mathbf{r}') = \sigma_\rho^2 \exp\left\{-\frac{1}{2}\left[\frac{(x-x')^2 + (y-y')^2}{L_H^2} + \frac{(z-z')^2}{L_V^2}\right]\right\}, \quad (8)$$

$$C_{SS}(t, t') = \sigma_S^2 \exp\left\{-\frac{1}{2}\frac{(t-t')^2}{T_S^2}\right\},$$

$$C_{K\rho}(\mathbf{r}, \mathbf{r}') \equiv 0,$$

$$C_{\rho S}(\mathbf{r}, t') \equiv 0,$$

and

$$C_{SK}(t, \mathbf{r}') \equiv 0,$$

where σ_K, σ_ρ, and σ_S represent the estimated departures of the true values $K(\mathbf{r})$, $\rho(\mathbf{r})$, and $S(t)$ from the a priori values $K_0(\mathbf{r})$, $\rho_0(\mathbf{r})$, and $S_0(t)$, where L_H and L_V represent the horizontal and vertical distances along which one expects the models to be smooth, and where T_S is the expected correlation length of errors in our estimate of the source.

For making the notation more compact, I introduce the model vector

$$\mathbf{m} = \begin{pmatrix} \mathbf{K} \\ \boldsymbol{\rho} \\ \mathbf{S} \end{pmatrix}. \quad (9)$$

The a priori information on the model is then described by the a priori model

$$\mathbf{m}_0 = \begin{pmatrix} \mathbf{K}_0 \\ \boldsymbol{\rho}_0 \\ \mathbf{S}_0 \end{pmatrix}$$

and the a priori covariance operator

$$C_m = \begin{pmatrix} C_{KK} & C_{K\rho} & C_{KS} \\ C_{\rho K} & C_{\rho\rho} & C_{\rho S} \\ C_{SK} & C_{S\rho} & C_{SS} \end{pmatrix} \tag{10}$$

(most of the off-diagonal terms are null in general).

THE INVERSE PROBLEM

The solution of the forward problem was written [equation (6)]:

$$\mathbf{p} = \mathbf{f}(\mathbf{K}, \boldsymbol{\rho}, \mathbf{S})$$

or, using the notations of the preceding section,

$$\mathbf{p} = \mathbf{f}(\mathbf{m}). \tag{11}$$

The a priori model $\mathbf{m}_0$ will, of course, not in general predict for the pressure field the observed value $\mathbf{p}_0$, i.e., we will have, in general,

$$\mathbf{p}_0 \neq \mathbf{f}(\mathbf{m}_0).$$

The generalized nonlinear inverse problem can be stated as the problem of finding the pair $\mathbf{p}$ and $\mathbf{m}$ which is, between all pairs satisfying equation (11), the closest to the a priori pair $\mathbf{p}_0$, $\mathbf{m}_0$, i.e., such that $\mathbf{p} = \mathbf{f}(\mathbf{m})$, distance between $(\mathbf{p}, \mathbf{m})$ and $(\mathbf{p}_0, \mathbf{m}_0)$ minimum. We now introduce a precise definition of distance. I take the distance associated with the norm

$$\|(\mathbf{p}, \mathbf{m})\|^2 = \|\mathbf{p}\|^2 + \|\mathbf{m}\|^2 = \mathbf{p}^* C_p^{-1}\mathbf{p} + \mathbf{m}^* C_m^{-1}\mathbf{m},$$

which corresponds to the usual definition for least-squares problems (Tarantola and Valette, 1982). In this equation, if we define

$$\tilde{\mathbf{p}} = C_p^{-1}\mathbf{p},$$

and

$$\tilde{\mathbf{m}} = C_m^{-1}\mathbf{m},$$

then, by $\mathbf{p}^*\tilde{\mathbf{p}}$ and $\mathbf{m}^*\tilde{\mathbf{m}}$, I denote, respectively,

$$\mathbf{p}^*\tilde{\mathbf{p}} = \mathbf{p}^* C_p^{-1}\mathbf{p} = \sum_g \sum_t \sum_s p(\mathbf{r}_g, t; \mathbf{r}_s)\, \tilde{p}(\mathbf{r}_g, t; \mathbf{r}_s),$$

and

$$\mathbf{m}^*\tilde{\mathbf{m}} = \mathbf{m}^* C_m^{-1}\mathbf{m} = \int d\mathbf{r}\; K(\mathbf{r})\, \tilde{K}(\mathbf{r}) + \int d\mathbf{r}\; \rho(\mathbf{r})\, \tilde{\rho}(\mathbf{r}) + \int dt\; S(t)\, \tilde{S}(t).$$

Among all norms, the least-squares (L^2) norm is the one which allows the easiest computations. Other norms, such as the absolute-values (L^1) norm, which has some advantages for the inversion of geophysical data (see Claerbout and Muir, 1973) are not envisaged in this paper. I arrive then at the statement of the generalized nonlinear least-squares inverse problem: to find the pair $\mathbf{p}$ and $\mathbf{m}$ such that $\mathbf{p} = \mathbf{f}(\mathbf{m})$, $(\mathbf{p}_0 - \mathbf{p})^* C_p^{-1}(\mathbf{p}_0 - \mathbf{p}) + (\mathbf{m}_0 - \mathbf{m})^* C_m^{-1}(\mathbf{m}_0 - \mathbf{m})$ minimum. This is equivalent to finding the model $\mathbf{m}$ which minimizes the functional

$$2S(\mathbf{m}) = [\mathbf{p}_0 - \mathbf{f}(\mathbf{m})]^* C_p^{-1}[\mathbf{p}_0 - \mathbf{f}(\mathbf{m})] + (\mathbf{m}_0 - \mathbf{m})^* C_m^{-1}(\mathbf{m}_0 - \mathbf{m}), \tag{12}$$

where the factor 2 is introduced for subsequent simplifications.

Before turning to the problem of the effective resolution of the minimization problem (12), let us emphasize that with such a formulation one would not have troubles with the problem of data sufficiency; even in the worst case when the number of data points tends to zero, the solution of the minimization of (12) tends to the a priori model $\mathbf{m}_0$, and the inverse problem remains well posed. With more and more data, and particularly with a data set as redundant as the one collected in present day seismic reflection experiments, the influence of $\mathbf{m}_0$ on the final model will tend to vanish.

In order to solve the minimization problem, let us introduce the linear operator F as the derivative of $\mathbf{f}$ at the point $\mathbf{m}$:

$$\mathbf{f}(\mathbf{m} + \delta\mathbf{m}) = \mathbf{f}(\mathbf{m}) + F\delta\mathbf{m} + \mathbf{o}(\|\delta\mathbf{m}\|^2).$$

A formal differentiation of $S(\mathbf{m})$ with respect to $\mathbf{m}$ gives the gradient of the functional S:

$$-\text{grad } S = C_m F^* C_p^{-1}[\mathbf{p}_0 - \mathbf{f}(\mathbf{m})] + (\mathbf{m}_0 - \mathbf{m}), \tag{13}$$

where F^* denotes the transpose of F. At the minimum of S the gradient must vanish, so the solution $\mathbf{m}$ of our least-squares problem must verify

$$\mathbf{m} - \mathbf{m}_0 = C_m F^* C_p^{-1}[\mathbf{p}_0 - \mathbf{f}(\mathbf{m})]. \tag{14}$$

This equation shows that the difference $\mathbf{m} - \mathbf{m}_0$ belongs to the image of C_m. Adding the term $C_m F^* C_p^{-1} F(\mathbf{m} - \mathbf{m}_0)$ to both sides of equation (14) and after reordering,

$$(I + C_m F^* C_p^{-1} F)(\mathbf{m} - \mathbf{m}_0) = C_m F^* C_p^{-1}\{\mathbf{p}_0 - \mathbf{f}(\mathbf{m}) + F(\mathbf{m} - \mathbf{m}_0)\}.$$

Since covariance operators are positive definite (except for null variances or infinite correlations), the operator $F^* C_p^{-1} F$ is necessarily definite nonnegative. It follows that the operator $I + C_m F^* C_p^{-1} F$ is regular and one can write

$$\mathbf{m} - \mathbf{m}_0 = (I + C_m F^* C_p^{-1} F)^{-1} C_m F^* C_p^{-1} \cdot \{\mathbf{p}_0 - \mathbf{f}(\mathbf{m}) + F(\mathbf{m} - \mathbf{m}_0)\}. \tag{15}$$

Although equation (15) is equivalent to equation (14), it can be directly solved using a fixed point algorithm:

$$\mathbf{m}_{k+1} = \mathbf{m}_0 + (I + C_m F_k^* C_p^{-1} F_k)^{-1} C_m F_k^* C_p^{-1} \cdot \{\mathbf{p}_0 - \mathbf{f}(\mathbf{m}_k) + F_k(\mathbf{m}_k - \mathbf{m}_0)\}, \tag{16}$$

where F_k denotes the value of F at the point $\mathbf{m}_k$. For a linear problem [$f(\mathbf{m}) = F\mathbf{m}$] this algorithm converges in only one iteration, giving

$$\mathbf{m} = \mathbf{m}_0 + (I + C_m F^* C_p^{-1} F)^{-1} C_m F^* C_p^{-1}\{\mathbf{p}_0 - F\mathbf{m}_0\}.$$

The reader can easily verify that equation (16) can also be written

$$\mathbf{m}_{k+1} = \mathbf{m}_k + (I + C_m F_k^* C_p^{-1} F_k)^{-1} \cdot \cdot \{C_m F_k^* C_p^{-1}[\mathbf{p}_0 - \mathbf{f}(\mathbf{m}_k)] - (\mathbf{m}_k - \mathbf{m}_0)\}. \tag{17}$$

The algorithm thus obtained (or any of its equivalents) was named the algorithm of total inversion in Tarantola and Valette (1982). The operator $W = (I + C_m F_k^* C_p^{-1} F_k)$ has a kernel of the form $W(\mathbf{r}, \mathbf{r}')$. The only practical means of handling the inverse of W is by discretizing the kernel $W(\mathbf{r}, \mathbf{r}')$ over a grid of points and inverting the system thus obtained. This matrix is so big that the inversion cannot be effected by present day computers. This simply means that the beautiful equation (17) is useless for our problem.

Let us come back to equation (12). For obtaining the minimum of S, one could try a gradient (steepest descent) method:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha(\text{grad } S)_k, \tag{18}$$

where α is an arbitrary constant, sufficiently small for ensuring the convergence of the algorithm.

Using equation (13) then

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha\{C_m F_k^* C_p^{-1}[\mathbf{p}_0 - \mathbf{f}(\mathbf{m}_k)] - (\mathbf{m}_k - \mathbf{m}_0)\}. \tag{19}$$

One can now see that this gradient equation is very similar to the total inversion equation (17): the operator $(I + C_m F_k^* C_p^{-1} F_k)^{-1}$ has simply been replaced by the diagonal operator αI.

Equations (17) and (19) are extreme cases of the general equation:

$$\mathbf{m}_{k+1} = \mathbf{m}_k + W\{C_m F_k^* C_p^{-1}[\mathbf{p}_0 - \mathbf{f}(\mathbf{m}_k)] - (\mathbf{m}_k - \mathbf{m}_0)\}, \tag{20}$$

where W represents an arbitrary regular operator close enough to $(I + C_m F_k^* C_p^{-1} F_k)$ to allow convergence

$$W \approx (I + C_m F_k^* C_p^{-1} F_k)^{-1}. \tag{21}$$

It is clear that if the algorithm (20) converges, then equation (14) is satisfied, which means that it necessarily converges to the solution of the problem (disregarding possible secondary minima).

The choice

$$W = \alpha I \tag{22}$$

gives the gradient equation (19), while the choice

$$W = (I + C_m F_k^* C_p^{-1} F_k)^{-1} \tag{23}$$

gives the total inversion equation (17). The number of iterations needed will greatly depend upon the choice of W. It may range from only one iteration for a quasi-linear problem and W given by equation (23), to a great number of iterations for W given by equation (22). Other choices of W can be imagined, for instance

$$W = [\text{DIAGONAL OF } (I + C_m F_0^* C_p^{-1} F_0)]^{-1} \tag{24}$$

or

$$W = [\text{DIAGONAL BAND OF } (I + C_m F_0^* C_p^{-1} F_0)]^{-1}. \tag{25}$$

While equation (24) is clearly related with Jacobi methods (see Clayton, 1982) of discrete linear algebra (used here in a more general context), it seems, at the present state of experience, that W given by equation (25) is the best compromise (with present day computers) between the simplicity of computations and the number of iterations.

For a given a priori model $\mathbf{m}_0$, the value of the operator $F_0^* C_p^{-1} F_0$ in equation (25) only depends upon the geometry of the problem, i.e., number of sources, receivers, and their relative positions. I hope to be able to obtain a set of empirical rules giving the conditions for optimum convergence.

In what follows I will focus attention on the choice $W = \alpha I$, the modifications needed for more elaborate choices of W being straightforward. Now turn to the problem of choosing the numerical value for α. The simplest strategy is to take for α a constant value, the same through all the iterations. For the first iteration, using (19),

$$\mathbf{m}_1 = \mathbf{m}_0 + \alpha\delta\mathbf{m}_1,$$

where

$$\delta\mathbf{m}_1 = C_m F_0^* C_p^{-1}[\mathbf{p}_0 - \mathbf{f}(\mathbf{m}_0)].$$

Once $\delta\mathbf{m}_1$ has been obtained (see next section), the numerical value of α can be chosen such that the deviations $\mathbf{m}_1 - \mathbf{m}_0$ are smaller than the a priori standard deviations (in C_m).

Using the methods of optimization theory (see for instance Walsh, 1975), an approximation to the optimum value of α_k to be used at the kth iteration is given by

$$\alpha_k = 1/(1 + \lambda_k C_p^{-1} \lambda_k / \gamma_k C_m^{-1} \gamma_k),$$

where

$$\lambda_k = F_k \gamma_k,$$

and

$$\gamma_k = (\text{grad } S)_k.$$

I point out that, when using a gradient algorithm for solving nonlinear inverse problems, there is no warranty that the value chosen for the scalar α will give a point $\mathbf{m}_{k+1}$ such that $S(\mathbf{m}_{k+1})$ is smaller than $S(\mathbf{m}_k)$. But, as $-(\text{grad } S)_k$ is a direction of descent, there must exist a constant α' $(0 < \alpha' < \alpha)$ for which $\mathbf{m}_{k+1}$ is better than $\mathbf{m}_k$. It is usual to take $\alpha' = \alpha/2$ as a new test value, and so on, until the condition $S(\mathbf{m}_{k+1}) < S(\mathbf{m}_k)$ is realized. It is clear that, with such a procedure, a gradient algorithm will necessarily converge.

I have to point out that I have not discussed the problem of estimating error and resolution on the a posteriori model. It can be demonstrated (Tarantola and Valette, 1982) that a useful estimate of error and resolution is obtained, in the linear approximation, through the a posteriori covariance operator C_m' given by

$$C_m' = (I + C_m F^* C_p^{-1} F)^{-1} C_m.$$

It is not clear at present which kind of information one should be able to obtain on C_m' with realistic computer requirements.

I now give explicit formulas for the problem, introducing the functions $\mathbf{K}$, $\boldsymbol{\rho}$, and $\mathbf{S}$. By definition,

$$\mathbf{f}(K + \delta\mathbf{K}, \boldsymbol{\rho} + \delta\boldsymbol{\rho}, \mathbf{S} + \delta\mathbf{S}) = \mathbf{f}(\mathbf{K}, \boldsymbol{\rho}, \mathbf{S}) + U\delta\mathbf{K} + V\delta\boldsymbol{\rho} + T\delta\mathbf{S} + \mathbf{o}(\delta\mathbf{K}, \delta\boldsymbol{\rho}, \delta\mathbf{S})^2. \tag{26}$$

Using equation (9) it is clear that the operator F can be written, in partitioned form,

$$F = (U \quad V \quad T), \tag{27}$$

where U, V, and T represent, respectively, the derivatives of the pressure field with respect to $\mathbf{K}$, $\boldsymbol{\rho}$, and $\mathbf{S}$. The recurrence relation (19) then takes the form

$$\begin{pmatrix} \mathbf{K}_{n+1} \\ \boldsymbol{\rho}_{n+1} \\ \mathbf{S}_{n+1} \end{pmatrix} = \begin{pmatrix} \mathbf{K}_{n} \\ \boldsymbol{\rho}_{n} \\ \mathbf{S}_{n} \end{pmatrix} + \alpha \left\{ \begin{pmatrix} C_{KK} & C_{K\rho} & C_{KS} \\ C_{\rho K} & C_{\rho\rho} & C_{\rho S} \\ C_{SK} & C_{S\rho} & C_{SS} \end{pmatrix} \begin{pmatrix} \delta\tilde{\mathbf{K}}_n \\ \delta\tilde{\boldsymbol{\rho}}_n \\ \delta\tilde{\mathbf{S}}_n \end{pmatrix} - \begin{pmatrix} \mathbf{K}_n - \mathbf{K}_0 \\ \boldsymbol{\rho}_n - \boldsymbol{\rho}_0 \\ \mathbf{S}_n - \mathbf{S}_0 \end{pmatrix} \right\}, \tag{28}$$

where

$$\begin{aligned} \delta\tilde{\mathbf{K}}_n &= U_n^* \delta\tilde{\mathbf{p}}_n, \\ \delta\tilde{\boldsymbol{\rho}}_n &= V_n^* \delta\tilde{\mathbf{p}}_n, \\ \delta\tilde{\mathbf{S}}_n &= T_n^* \delta\tilde{\mathbf{p}}_n, \end{aligned} \tag{29}$$

and

$$\delta\tilde{\mathbf{p}}_n = C_p^{-1}[\mathbf{p}_0 - \mathbf{f}(\mathbf{K}_n, \boldsymbol{\rho}_n, \mathbf{S}_n)]. \tag{30}$$

In the particular case of independent a priori information ($C_{K\rho} \equiv 0$, $C_{\rho S} \equiv 0$, $C_{SK} \equiv 0$), equation (28) simplifies to

$$\begin{aligned} \mathbf{K}_{n+1} &= \mathbf{K}_n + \alpha[C_{KK}\delta\tilde{\mathbf{K}}_n - (\mathbf{K}_n - \mathbf{K}_0)], \\ \boldsymbol{\rho}_{n+1} &= \boldsymbol{\rho}_n + \alpha[C_{\rho\rho}\delta\tilde{\boldsymbol{\rho}}_n - (\boldsymbol{\rho}_n - \boldsymbol{\rho}_0)], \end{aligned} \tag{31}$$

and

$$\mathbf{S}_{n+1} = \mathbf{S}_n + \alpha[C_{SS}\delta\tilde{\mathbf{S}}_n - (\mathbf{S}_n - \mathbf{S}_0)],$$

while equations (29)–(30) remain the same.

In the next section I give a physical interpretation of these formulas.

PHYSICAL INTERPRETATION AND PRACTICAL RESOLUTION

At each iteration of the algorithm (29)–(31) we have the values $\mathbf{K}_n$, $\boldsymbol{\rho}_n$, and $\mathbf{S}_n$ resulting from the previous iteration.

The first step is the computation of equation (30)

$$\delta\tilde{\mathbf{p}}_n = C_p^{-1}[\mathbf{p}_0 - \mathbf{f}(\mathbf{K}_n, \boldsymbol{\rho}_n, \mathbf{S}_n)]. \tag{32}$$

One can see that $\delta\tilde{\mathbf{p}}_n$ can be interpreted as the "weighted residuals." Its computation involves solving the forward problem, for each of the shotpoints, to evaluate the predicted data for the nth model:

$$\mathbf{p}_n = \mathbf{f}(\mathbf{K}_n, \boldsymbol{\rho}_n, \mathbf{S}_n). \tag{33}$$

Using for C_p the example given in equation (7),

$$\delta\tilde{p}_n(\mathbf{r}_g, t; \mathbf{r}_s) = \frac{p_0(\mathbf{r}_g, t; \mathbf{r}_s) - p_n(\mathbf{r}_g, t; \mathbf{r}_s)}{\sigma_{gs}^2}.$$

The second step of the iterative loop is the computation of equation (29)

$$\begin{aligned} \delta\tilde{\mathbf{K}}_n &= U_n^* \delta\tilde{\mathbf{p}}_n, \\ \delta\tilde{\boldsymbol{\rho}}_n &= V_n^* \delta\tilde{\mathbf{p}}_n, \end{aligned} \tag{34}$$

and

$$\delta\tilde{\mathbf{S}}_n = T_n^* \delta\tilde{\mathbf{p}}_n.$$

It has just been shown that $\delta\tilde{\mathbf{p}}_n$ are the weighted data residuals. As defined in equations (26) and (27), U_n, V_n, and T_n are the derivatives of the pressure field with respect to $\mathbf{K}$, $\boldsymbol{\rho}$, and $\mathbf{S}$ (at the current point $\mathbf{K}_n$, $\boldsymbol{\rho}_n$, and $\mathbf{S}_n$), and U_n^*, V_n^*, and T_n^* are the transposes of the derivative operators. The effect of the action of these operators on the residuals was first demonstrated by Lailly (1984). I give in the Appendix an equivalent demonstration. It is shown that for obtaining the values of $\delta\tilde{\mathbf{K}}_n$, $\delta\tilde{\boldsymbol{\rho}}_n$, and $\delta\tilde{\mathbf{S}}_n$, as defined by equation (34) one has to perform the following two operations.

(1) Propagating the residuals backward in time. I show in the Appendix that the first operation to be done is the computation of the field $p'_n(\mathbf{r}, t; \mathbf{r}_s)$ defined by

$$p'_n(\mathbf{r}, t; \mathbf{r}_s) = \int d\mathbf{r}'\, g(\mathbf{r}, 0; \mathbf{r}', t) * \delta s(\mathbf{r}', t; \mathbf{r}_s), \tag{35}$$

where

$$\delta s(\mathbf{r}, t; \mathbf{r}_s) = \sum_g \delta(\mathbf{r} - \mathbf{r}_g)\delta\tilde{p}_n(\mathbf{r}_g, t; \mathbf{r}_s). \tag{36}$$

For given $\mathbf{r}_s$, the field $p'_n(\mathbf{r}, t; \mathbf{r}_s)$ defined by equation (35) clearly corresponds to the propagation of the source $\delta s(\mathbf{r}', t; \mathbf{r}_s)$ backward in time (because of the Green's function $g(\mathbf{r}, 0; \mathbf{r}', t)$ rather than $g(\mathbf{r}, t; \mathbf{r}', 0)$). Equation (36) shows that the source $\delta s(\mathbf{r}', t; \mathbf{r}_s)$ is, for given $\mathbf{r}_s$, a composite source with one elementary source at each geophone position having as source time function the weighted residual (observed minus computed pressure field) at that point. The residuals represent that part of the signal (data set) which is not taken into account by the current model (of bulk modulus, density, and source). That the field $p'_n(\mathbf{r}, t; \mathbf{r}_s)$ is obtained by propagation of the residuals backward in time means that $\mathbf{p}'_n$ is the field that, propagated in the natural sense (forward in time), gives at the observation points a pressure field simply related with the residuals: $p'_n(\mathbf{r}, t; \mathbf{r}_s)$ is a sort of *missing diffracted field* (not taken into account by the current model). From a computational point of view, the knowledge of $p'_n(\mathbf{r}, t; \mathbf{r}_s)$ requires the solution of one forward problem, with a composite source and with reversed time, per source point.

(2) Correlating the "upgoing" and "downgoing" wave fields at each point of the space. The wave field $p_n(\mathbf{r}, t; \mathbf{r}_s)$ represents the field which propagates in the current model (of bulk modulus and of density) for the current values of the source time function. I have just shown that $p'_n(\mathbf{r}, t; \mathbf{r}_s)$ is related with the missing diffracted field for the current model. As shown in the Appendix, the weighted model corrections $\delta\tilde{K}(\mathbf{r})$, $\delta\tilde{\rho}(\mathbf{r})$, and $\delta\tilde{S}(t)$, of equation (34) can be obtained using the following equations:

$$\delta\tilde{K}_n(\mathbf{r}) = \frac{1}{K_n^2(\mathbf{r})} \int dt \sum_s \frac{\partial p'_n}{\partial t}(\mathbf{r}, t; \mathbf{r}_s) \frac{\partial p_n}{\partial t}(\mathbf{r}, t; \mathbf{r}_s), \tag{37a}$$

$$\delta\tilde{\rho}_n(\mathbf{r}) = \frac{1}{\rho_n^2(\mathbf{r})} \int dt \sum_s \operatorname{grad} p'_n(\mathbf{r}, t; \mathbf{r}_s) \cdot \operatorname{grad} p_n(\mathbf{r}, t; \mathbf{r}_s), \tag{37b}$$

and

$$\delta\tilde{S}_n(t) = \sum_s p'_n(\mathbf{r}_s, t; \mathbf{r}_s). \tag{37c}$$

These equations can easily be understood. If at a point in space there is a diffractor (a point perturbation of bulk modulus or of density), there will be a diffracted pressure field which at that point will be correlated with the incident wave field. The sums over time in equations (37) are the appropriate measures of correlation between the incident field and the missing diffracted field. At points where these correlations do not vanish, the current values of bulk modulus and density have to be modified according to equations (37a)–(37b) and to equation (39) below. This resembles the "imaging principle" of Claerbout (1971). Here, rather than *defining* a "reflectivity field" by the ratio of upgoing and downgoing wave fields, I show that the correlations in (37) rigorously give weighted *models* of bulk modulus and density.

The last step for the implementation of the iterative loop is the computation of equation (31):

$$\mathbf{K}_{n+1} = \mathbf{K}_n + \alpha[C_{KK}\,\delta\tilde{\mathbf{K}}_n - (\mathbf{K}_n - K_0)],$$
$$\boldsymbol{\rho}_{n+1} = \boldsymbol{\rho}_n + \alpha[C_{\rho\rho}\,\delta\tilde{\boldsymbol{\rho}}_n - (\boldsymbol{\rho}_n - \boldsymbol{\rho}_0)], \tag{38}$$

and

$$\mathbf{S}_{n+1} = \mathbf{S}_n + \alpha[C_{SS}\,\delta\tilde{\mathbf{S}}_n - (\mathbf{S}_n - \mathbf{S}_0)].$$

Explicitly, we have

$$K_{n+1}(\mathbf{r}) = K_n(\mathbf{r}) + \alpha\{\delta K_n(\mathbf{r}) - [K_n(\mathbf{r}) - K_0(\mathbf{r})]\},$$
$$\rho_{n+1}(\mathbf{r}) = \rho_n(\mathbf{r}) + \alpha\{\delta\rho_n(\mathbf{r}) - [\rho_n(\mathbf{r}) - \rho_0(\mathbf{r})]\}, \tag{39}$$

and

$$S_{n+1}(t) = S_n(t) + \alpha\{\delta S_n(t) - [S_n(t) - S_0(t)]\},$$

where

$$\delta K_n(\mathbf{r}) = \int d\mathbf{r}'\, C_{KK}(\mathbf{r}, \mathbf{r}')\delta\tilde{K}_n(\mathbf{r}'),$$
$$\delta\rho_n(\mathbf{r}) = \int d\mathbf{r}'\, C_{\rho\rho}(\mathbf{r}, \mathbf{r}')\delta\tilde{\rho}_n(\mathbf{r}') \tag{40}$$

and

$$\delta S_n(t) = \int dt'\, C_{SS}(t, t')\,\delta\tilde{S}_n(t'),$$

or slightly more complicated formulas if one uses the recurrence relation (28) instead of relation (31). Since C_{KK}, $C_{\rho\rho}$, and C_{SS} are covariance operators, they will be in general smoothing operators, so $\delta\mathbf{K}_n$, $\delta\boldsymbol{\rho}_n$, and $\delta\mathbf{S}_n$ are simply smoothed versions of $\delta\tilde{\mathbf{K}}_n$, $\delta\tilde{\boldsymbol{\rho}}_n$, and $\delta\tilde{\mathbf{S}}_n$.

Starting with $K_n(\mathbf{r})$, $\rho_n(\mathbf{r})$, and $S_n(t)$, I have gone through an entire loop of the iterative sequence leading to $K_{n+1}(\mathbf{r})$, $\rho_{n+1}(\mathbf{r})$, and $S_{n+1}(t)$. I have shown that the operations required for updating the current model are essentially the solution of two forward problems per shotpoint.

CONCLUSION

In paper I, I demonstrated that the inverse solution of the linearized problem for a homogeneous reference model can be solved using a slightly modified version of the Kirchhoff migration. It is nice to see that when applied to the heterogeneous problem, the same inverse method leads naturally to an algorithm for a solution which strongly resembles the migration method based on the imaging principle of Claerbout (1971).

I have shown in the previous section that each iterative loop of the inversion requires the solution of two times as many forward problems as there are source locations. Although this is a big task, it is within the capabilities of present day vector computers.

The advantage of an inversion of the type presented here with respect to classical migration is of two orders. First, it can probably handle strong lateral variations much better than conventional methods, because the velocity model is elaborated *as the iterative sequence proceeds*, and it is not given independently, as in migration. Second, it is important to emphasize that the inversion gives *absolute values* of density and bulk modulus. If errors due to the neglect of attenuation and elastic (versus acoustic) effects in the forward modelization are not too severe, these absolute values of density and bulk modulus can be of great help for direct hydrocarbon detection.

I do not assume any particular numerical solution of the forward problem, although it is clear that the finite-differencing technique is well adapted to the problem. I must emphasize that if the forward scheme accounts for surface, refracted, or multiply reflected waves, the inverse solution presented here will use these waves.

Numerical examples are now being implemented and will be the subject of another paper. The generalization of the present results to the viscoelastic case is presently being studied.

ACKNOWLEDGMENTS

I thank my colleagues P. Bernard, A. Cisternas, G. Jobert, P. Lailly, F. Landre, and A. Nercessian for very helpful comments and suggestions. This work has been partially supported by the R.C.P. 264 (Etude Interdisciplinaire des Problèmes Inverses) and by the A.T.P. Calcul Vectoriel 1982. Contribution I.P.G. no. 688.

REFERENCES

Aki, K., and Richards, P., 1980, Quantitative seismology: W. H. Freeman and Co.
Ben-Menahem, A., and Singh, S. J., 1981, Seismic waves and sources: Springer-Verlag.
Claerbout, J. F., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481.
Claerbout, J. F., and Muir, F., 1973, Robust modelling with erratic data: Geophysics, **38**, 826–844.
Clayton, R. W., Seismic Tomography. Preprint 1982.
French, W. S., 1974, Two-dimensional and three-dimensional migration of model-experiment reflection profiles: Geophysics, **39**, 265–277.
Lailly, P., 1984, The seismic inverse problem as a sequence of before stack migrations: Conference on Inverse Scattering, Tulsa, Oklahoma: Philadelphia, SIAM.
Morse, P. M., and Feshbach, H., 1953, Methods of theoretical physics: McGraw-Hill Book Co., Inc.
Schneider, W. S., 1978, Integral formulation for migration in two and three dimensions: Geophysics, **43**, 49–76.
Tarantola, A., 1984, Linearized inversion of seismic reflection data: Geophys. Prosp. (in press).
Tarantola, A., and Valette, B., 1982, Generalized non linear inverse problems solved using the least-squares criterion: Rev. of Geophys. and Space Physics, **20**, p. 219–232.
Walsh, G. R., 1975, Methods of optimization: John Wiley and Sons.

APPENDIX

In this Appendix I examine the Fréchet derivatives of the pressure field and show which is the result of the action of their transposes on an arbitrary vector.

First [equation (6)]

$$\mathbf{p} = \mathbf{f}(\mathbf{K}, \boldsymbol{\rho}, \mathbf{S}).$$

The operators U, V, and T were defined by equation (26):

$$\delta\mathbf{p} = U\delta\mathbf{K} + V\delta\boldsymbol{\rho} + T\delta\mathbf{S} + \mathbf{o}(\delta\mathbf{K}, \delta\boldsymbol{\rho}, \delta\mathbf{S})^2, \qquad \text{(A-1)}$$

where

$$\delta\mathbf{p} = \mathbf{f}(\mathbf{K} + \delta\mathbf{K}, \boldsymbol{\rho} + \delta\boldsymbol{\rho}, \mathbf{S} + \delta\mathbf{S}) - \mathbf{f}(\mathbf{K}, \boldsymbol{\rho}, \mathbf{S}).$$

Introducing the kernels of U, V, and T, equation (A-1) becomes

$$\begin{aligned}\delta p(\mathbf{r}_g, t; \mathbf{r}_s) = &\int d\mathbf{r}\; U(\mathbf{r}_g, t; \mathbf{r}_s \mid \mathbf{r})\delta K(\mathbf{r}) \\ &+ \int d\mathbf{r}\; V(\mathbf{r}_g, t; \mathbf{r}_s \mid \mathbf{r})\delta\rho(\mathbf{r}) \\ &+ \int dt'\; T(\mathbf{r}_g, t; \mathbf{r}_s \mid t')\delta S(t') \\ &+ \mathbf{o}(\delta\mathbf{K}, \delta\boldsymbol{\rho}, \delta\mathbf{S})^2. \qquad \text{(A-2)}\end{aligned}$$

To give a physical meaning to these kernels, one can formally write

$$U(\mathbf{r}_g, t; \mathbf{r}_s \mid \mathbf{r}) = \frac{\partial p(\mathbf{r}_g, t; \mathbf{r}_s)}{\partial K(\mathbf{r})},$$

$$V(\mathbf{r}_g, t; \mathbf{r}_s \mid \mathbf{r}) = \frac{\partial p(\mathbf{r}_g, t; \mathbf{r}_s)}{\partial \rho(\mathbf{r})},$$

and (A-3)

$$T(\mathbf{r}_g, t; \mathbf{r}_s \mid t') = \frac{\partial p(\mathbf{r}_g, t; \mathbf{r}_s)}{\partial S(t')}.$$

By definition of $\mathbf{p}$ and $\delta\mathbf{p}$, I have

$$\left\{\frac{1}{K(\mathbf{r})}\frac{\partial^2}{\partial t^2} - \operatorname{div}\left[\frac{1}{\rho(\mathbf{r})}\operatorname{grad}\right]\right\} p(\mathbf{r}, t; \mathbf{r}_s) = \delta(\mathbf{r} - \mathbf{r}_s)S(t)$$

and

$$\left\{\frac{1}{K(\mathbf{r}) + \delta K(\mathbf{r})}\frac{\partial^2}{\partial t^2} - \operatorname{div}\left[\frac{1}{\rho(\mathbf{r}) + \delta\rho(\mathbf{r})}\operatorname{grad}\right]\right\} \cdot [p(\mathbf{r}, t; \mathbf{r}_s) + \delta p(\mathbf{r}, t; \mathbf{r}_s)] = \delta(\mathbf{r} - \mathbf{r}_s)[S(t) + \delta S(t)].$$

After some easy manipulations, and using

$$\frac{1}{h(\mathbf{r}) + \delta h(\mathbf{r})} = \frac{1}{h(\mathbf{r})} - \frac{\delta h(\mathbf{r})}{h^2(\mathbf{r})} + \mathbf{o}(\delta h^2),$$

then

$$\left\{\frac{1}{K(\mathbf{r})}\frac{\partial^2}{\partial t^2} - \operatorname{div}\left[\frac{1}{\rho(\mathbf{r})}\operatorname{grad}\right]\right\} \delta p(\mathbf{r}, t; r_s) = \Delta s(\mathbf{r}, t; \mathbf{r}_s),$$

where $\Delta\mathbf{s}$ are the "secondary sources"

$$\begin{aligned}\Delta s(\mathbf{r}, t; \mathbf{r}_s) = &\;\delta(\mathbf{r} - \mathbf{r}_s)\delta S(t) \\ &+ \frac{\delta K(\mathbf{r})}{K^2(\mathbf{r})}\frac{\partial^2 p}{\partial t^2}(\mathbf{r}, t; \mathbf{r}_s) \\ &- \operatorname{div}\left[\frac{\delta\rho(\mathbf{r})}{\rho^2(\mathbf{r})}\operatorname{grad} p(\mathbf{r}, t; \mathbf{r}_s)\right] \\ &+ \mathbf{o}(\delta\mathbf{K}, \delta\boldsymbol{\rho}, \delta\mathbf{S})^2. \qquad \text{(A-4)}\end{aligned}$$

Since the solution of equation (1) was given by equation (5), the solution for $\delta\mathbf{p}$ will be given by

$$\delta p(\mathbf{r}, t; \mathbf{r}_s) = \int d\mathbf{r}'\; g(\mathbf{r}, t; \mathbf{r}', 0) * \Delta s(\mathbf{r}', t; \mathbf{r}_s),$$

or using the reciprocity theorem for the Green's function (see Morse and Feshbach, 1953; Aki and Richards, 1980; Ben-Menahem and Singh, 1981)

$$\delta p(\mathbf{r}, t; \mathbf{r}_s) = \int d\mathbf{r}'\; g(\mathbf{r}', t; \mathbf{r}, 0) * \Delta s(\mathbf{r}', t; \mathbf{r}_s).$$

Using equation (A-4) this gives

$$[U\delta K](\mathbf{r}_g, t; \mathbf{r}_s) = \int d\mathbf{r}\; g(\mathbf{r}, t; \mathbf{r}_g, 0) * \left[\frac{\delta K(\mathbf{r})}{K^2(\mathbf{r})}\frac{\partial^2 p}{\partial t^2}(r, t; \mathbf{r}_s)\right], \qquad \text{(A-5a)}$$

$$[V\delta\rho](\mathbf{r}_g, t; \mathbf{r}_s) = \int d\mathbf{r}\; g(\mathbf{r}, t; \mathbf{r}_g, 0) * \left\{-\operatorname{div}\left[\frac{\delta\rho(\mathbf{r})}{\rho^2(\mathbf{r})}\operatorname{grad} p(\mathbf{r}, t; \mathbf{r}_s)\right]\right\}, \qquad \text{(A-5b)}$$

and

$$[T\delta S](\mathbf{r}_g, t; \mathbf{r}_s) = g(\mathbf{r}_g, t; \mathbf{r}_s, 0) * \delta S(t). \qquad \text{(A-5c)}$$

Assuming that g and p vanish for $t \to \infty$, equation (A-5a) can be rewritten

$$[U\delta\mathbf{K}](\mathbf{r}_g, t; \mathbf{r}_s) = \int d\mathbf{r}\; \frac{1}{K^2(\mathbf{r})}\dot{g}(\mathbf{r}, t; \mathbf{r}_g, 0) * \dot{p}(\mathbf{r}, t; \mathbf{r}_s)\,\delta k(\mathbf{r}). \qquad \text{(A-5a')}$$

Using

$$\alpha \operatorname{div}(\mathbf{v}) = -(\operatorname{grad}\alpha)\mathbf{v} + \operatorname{div}(\alpha\mathbf{v}),$$

equation (A-5b) can also be written

$$\begin{aligned}[V\delta\rho](\mathbf{r}_g, t; \mathbf{r}_s) = &\int d\mathbf{r}\left[\frac{\delta\rho(\mathbf{r})}{\rho^2(\mathbf{r})}\right]\operatorname{grad} g(\mathbf{r}, t; \mathbf{r}_g, 0) \;\dot{*}\; \operatorname{grad} p(\mathbf{r}, t; \mathbf{r}_s) \\ &- \int d\mathbf{r}\operatorname{div}\left[\frac{\delta\rho(\mathbf{r})}{\rho^2(\mathbf{r})}g(\mathbf{r}, t; \mathbf{r}_g, 0) * \operatorname{grad} p(\mathbf{r}, t; \mathbf{r}_s)\right], \qquad \text{(A-5b')}\end{aligned}$$

where the gradient of the Green's function is always taken with respect to the first spatial variable, and where the symbol $\dot{*}$ stands for a time convolution of the scalar product. Using

$$\int d\mathbf{r}\operatorname{div}(\mathbf{v}) = \int_S \mathbf{v}\cdot d\mathbf{S},$$

the last sum in equation (A-5b′) can be written as an integral over the surface of the Earth, which vanishes because homogeneous boundary conditions are assumed. This lets

$$[V\delta\rho](\mathbf{r}_g, t; \mathbf{r}_s) = \int d\mathbf{r} \left[\frac{\delta\rho(\mathbf{r})}{\rho^2(\mathbf{r})}\right] [\text{grad } g(\mathbf{r}, t; \mathbf{r}_g, 0)] \underset{\cdot}{\ast} [\text{grad } p(\mathbf{r}, t; \mathbf{r}_s)]. \tag{A-5b″}$$

Using equations (A-2), (A-3), (A-5a′), (A-5b″), and (A-5c), one can now directly obtain the kernels of U, V, and T:

$$\frac{\partial p(\mathbf{r}_g, t; \mathbf{r}_s)}{\partial K(\mathbf{r})} = U(\mathbf{r}_g, t; \mathbf{r}_s \mid \mathbf{r}) = \frac{1}{K^2(\mathbf{r})} \dot{g}(\mathbf{r}, t; \mathbf{r}_g, 0) \ast \dot{p}(\mathbf{r}, t; \mathbf{r}_s),$$

$$\frac{\partial p(\mathbf{r}_g, t; \mathbf{r}_s)}{\partial \rho(\mathbf{r})} = V(\mathbf{r}_g, t; \mathbf{r}_s \mid \mathbf{r}) = \frac{1}{\rho^2(\mathbf{r})} \text{grad } g(\mathbf{r}, t; \mathbf{r}_g, 0) \underset{\cdot}{\ast} \text{grad } p(\mathbf{r}, t; \mathbf{r}_s), \tag{A-6}$$

and

$$\frac{\partial p(\mathbf{r}_g, t; \mathbf{r}_s)}{\partial S(t')} = T(\mathbf{r}_g, t; \mathbf{r}_s \mid t') = g(\mathbf{r}_g, t; \mathbf{r}_s, t').$$

We now turn to look for the action of the operators U^*, V^*, and T^*. For an arbitrary $\delta\tilde{\mathbf{p}}$ define

$$\delta\tilde{\mathbf{K}} = U^*\delta\tilde{\mathbf{p}},$$

$$\delta\tilde{\boldsymbol{\rho}} = V^*\delta\tilde{\mathbf{p}},$$

and

$$\delta\tilde{\mathbf{S}} = T^*\delta\tilde{\mathbf{p}}.$$

Let $U^*(\mathbf{r} \mid \mathbf{r}_g, t; \mathbf{r}_s)$, $V^*(\mathbf{r} \mid \mathbf{r}_g, t; \mathbf{r}_s)$, and $T^*(\mathbf{r} \mid \mathbf{r}_g, t; \mathbf{r}_s)$ be, respectively, the kernels of U^*, V^*, and T^*. From the definition of the transpose of a linear operator,

$$U^*(\mathbf{r} \mid \mathbf{r}_g, t; \mathbf{r}_s) = U(\mathbf{r}_g, t; \mathbf{r}_s \mid \mathbf{r}),$$

$$V^*(\mathbf{r} \mid \mathbf{r}_g, t; \mathbf{r}_s) = V(\mathbf{r}_g, t; \mathbf{r}_s \mid \mathbf{r}),$$

and

$$T^*(t' \mid \mathbf{r}_g, t; \mathbf{r}_s) = T(\mathbf{r}_g, t; \mathbf{r}_s \mid t').$$

Then write

$$\delta\tilde{K}(\mathbf{r}) = \sum_g \int dt \sum_s U^*(\mathbf{r} \mid \mathbf{r}_g, t; \mathbf{r}_s)\delta\tilde{p}(\mathbf{r}_g, t; \mathbf{r}_s) = \sum_g \int dt \sum_s U(\mathbf{r}_g, t; \mathbf{r}_s \mid \mathbf{r})\delta\tilde{p}(\mathbf{r}_g, t; \mathbf{r}_s),$$

$$\delta\tilde{\rho}(\mathbf{r}) = \sum_g \int dt \sum_s V^*(\mathbf{r} \mid \mathbf{r}_g, t; \mathbf{r}_s)\delta\tilde{p}(\mathbf{r}_g, t; \mathbf{r}_s) = \sum_g \int dt \sum_s V(\mathbf{r}_g, t; \mathbf{r}_s \mid \mathbf{r})\delta\tilde{p}(\mathbf{r}_g, t; \mathbf{r}_s),$$

and

$$\delta\tilde{S}(t') = \sum_g \int dt \sum_s T^*(t' \mid \mathbf{r}_g, t; \mathbf{r}_s)\delta\tilde{p}(\mathbf{r}_g, t; \mathbf{r}_s) = \sum_g \int dt \sum_s T(\mathbf{r}_g, t; \mathbf{r}_s \mid t')\delta\tilde{p}(\mathbf{r}_g, t; \mathbf{r}_s).$$

Using equation (A-6),

$$\delta\tilde{K}(\mathbf{r}) = \frac{1}{K^2(\mathbf{r})} \sum_g \int dt \sum_s \dot{g}(\mathbf{r}, t; \mathbf{r}_g, 0) \ast \dot{p}(\mathbf{r}, t; \mathbf{r}_s)\delta\tilde{p}(\mathbf{r}_g, t; \mathbf{r}_s),$$

$$\delta\tilde{\rho}(\mathbf{r}) = \frac{1}{\rho^2(\mathbf{r})} \sum_g \int dt \sum_s \text{grad } g(\mathbf{r}, t; \mathbf{r}_g, 0) \underset{\cdot}{\ast} \text{grad } p(\mathbf{r}, t; \mathbf{r}_s)\delta\tilde{p}(\mathbf{r}_g, t; \mathbf{r}_s),$$

and

$$\delta\tilde{S}(t') = \sum_g \int dt \sum_s g(\mathbf{r}_g, t; \mathbf{r}_s, t')\delta\tilde{p}(\mathbf{r}_g, t; \mathbf{r}_s).$$

Using

$$\int dt\, [f(t) \ast g(t)]h(t) = \int dt\, g(t)[f(-t) \ast h(t)],$$

and

$$g(\mathbf{r}, -t; \mathbf{r}', 0) = g(\mathbf{r}, 0; \mathbf{r}', t),$$

$$\delta\tilde{K}(\mathbf{r}) = \frac{1}{K^2(\mathbf{r})} \int dt \sum_s p(\mathbf{r}, t; \mathbf{r}_s) \cdot \cdot \left[\sum_g \dot{g}(\mathbf{r}, 0; \mathbf{r}_g, t) \ast \delta\tilde{p}(\mathbf{r}_g, t; \mathbf{r}_s)\right],$$

$$\delta\tilde{\rho}(\mathbf{r}) = \frac{1}{\rho^2(\mathbf{r})} \int dt \sum_s \text{grad } p(\mathbf{r}, t; \mathbf{r}_s) \cdot \cdot \left[\sum_g \text{grad } g(\mathbf{r}, 0; \mathbf{r}_g, t) \ast \delta\tilde{p}(\mathbf{r}_g, t; \mathbf{r}_s)\right],$$

and

$$\delta\tilde{S}(t) = \sum_s \sum_g g(\mathbf{r}_g, 0; \mathbf{r}_s, t) \ast \delta\tilde{p}(\mathbf{r}_g, t; \mathbf{r}_s).$$

One can write

$$\delta\tilde{K}(\mathbf{r}) = \frac{1}{K^2(\mathbf{r})} \sum_s \int dt\, \dot{p}(\mathbf{r}, t; \mathbf{r}_s)\dot{p}'(\mathbf{r}, t; \mathbf{r}_s),$$

$$\delta\tilde{\rho}(\mathbf{r}) = \frac{1}{\rho^2(\mathbf{r})} \sum_s \int dt\, \text{grad } p(\mathbf{r}, t; \mathbf{r}_s) \cdot \text{grad } p'(\mathbf{r}, t; \mathbf{r}_s), \tag{A-7}$$

and

$$\delta\tilde{S}(t) = \sum_s p'(\mathbf{r}_s, t; \mathbf{r}_s),$$

where

$$p'(\mathbf{r}, t; \mathbf{r}_s) = \sum_g g(\mathbf{r}, 0; \mathbf{r}_g, t) \ast \delta\tilde{p}(\mathbf{r}_g, t; \mathbf{r}_s).$$

Defining

$$\delta s(\mathbf{r}, t; \mathbf{r}_s) = \sum_g \delta(\mathbf{r} - \mathbf{r}_g)\delta\tilde{p}(\mathbf{r}_g, t; \mathbf{r}_s), \tag{A-8}$$

and

$$p'(\mathbf{r}, t; \mathbf{r}_s) = \int d\mathbf{r}'\, g(\mathbf{r}, 0; \mathbf{r}', t) \ast \delta s(\mathbf{r}', t; \mathbf{r}_s), \tag{A-9}$$

equations (35) to (37) are then easily obtained.