GEOPHYSICS, VOL. 51, NO. 7 (JULY 1986); P. 1387-1403, 16 FIGS.

Two-dimensional nonlinear inversion of seismic waveforms: Numerical results

Odile Gauthier*, Jean Virieux*, and Albert Tarantola*

ABSTRACT

The nonlinear problem of inversion of seismic waveforms can be set up using least-squares methods. The inverse problem is then reduced to the problem of minimizing a (nonquadratic) function in a space of many (10^4 to 10^6) variables. Using gradient methods leads to iterative algorithms, each iteration implying a forward propagation generated by the actual sources, a backward propagation generated by the data residuals (acting as if they were sources), and a correlation at each point of the space of the two fields thus obtained, which gives the updated model. The quality of the results of any inverse method depends heavily on the realism of the forward modeling. Finite-difference schemes are a good choice relative to realism because, although they are time-consuming, they give excellent results. Numerical tests performed with multioffset synthetic data from a two-dimensional model prove the feasibility of the approach. If only surface-recorded reflections are used, the high spatial frequency content of the model (but not the low spatial frequencies) is recovered in few ($\simeq 5$) iterations. By using transmitted data also (e.g., between two boreholes), all the spatial frequencies are recovered. Since the problem is nonlinear, if the initial guess is far enough from the true solution, the iterative algorithm may converge into a secondary solution. A nonlinear inversion with 8 shots, each shot recorded at 400 receiver locations, with 700 samples in each seismogram, corresponding to a 2-D model described by 40 000 grid points, takes approximately 1 hour in a CRAY 1S supercomputer.

INTRODUCTION

Linearized inversion of seismic waveforms has received considerable attention in the literature. For instance, Clayton and Stolt (1981) and Tarantola (1984a) propose methods for solving the linearized multidimensional inverse problem with multioffset seismic reflection data. Woodhouse and Dziewonski (1984) and Tanimoto (1984) also use linearized inversion of waveforms to obtain global Earth models.

If a starting model is known to be close enough to the actual medium, linearized inversion will probably perform well. Unfortunately, since there is no practical test to check the accuracy of the linearization, nonlinear inverse techniques are more promising. There are two basic options: a full exploration of the parameter space (systematic or Monte Carlo), or a local descent method. The first approach has the advantage of avoiding local minima, but it is too time-consuming for modern computers. The second approach gives the correct solution when the starting model is inside the valley of the global minimum, irrespective of the choice of starting point. Thus the starting model plays a much less crucial role than in linearized inversion.

Tarantola (1984b) suggested an approach to the general nonlinear inverse problem of interpretation of acoustic seismic waveforms. Our aim here is to prove the feasibility of that method numerically by using synthetic data.

We briefly discuss the choice of the finite-difference technique as the numerical method to solve the forward problem of computing synthetic seismograms. We then review the main steps involved in the inversion method, emphasizing the physical interpretation. A point diffractor is the first numerical example: results clarify the essential features of the inverse procedure. Other geometries help show the circumstances in which the numerical algorithm can recover the parameters of the medium. Mathematical details are in the appendixes. A review of the formulation of the inverse problem can be found in Tarantola (1984c).

THE FORWARD PROBLEM (COMPUTATION OF SYNTHETIC SEISMOGRAMS)

Performance of the inverse procedure depends critically on the technique used for solving the forward problem. Because of their low cost, ray methods and their extensions have been applied for linearized inverse formulations (e.g., Chapman and Orcutt, 1985a) or nonlinear inverse formulations (Mora, 1985). The main limitation of ray methods is that because of the high-frequency approximation, the physical parameters of the medium must vary slowly over several wavelengths. Even so, a low gradient in the velocity structure creates a low-frequency reflection which will be neglected by ray methods, which are only interested in the high-frequency content (Chapman and Orcutt, 1985b). Moreover, zero-order discontinuities in pa-

Manuscript received by the Editor August 15, 1985; revised manuscript received December 13, 1985.
*Laboratoire de Sismologie, Institut de Physique du Globe de Paris, Université de Paris 6 et 7, 4 place Jussieu, F-75252 Paris Cedex 05, France.

rameters (i.e., the interfaces) must be handled explicitly by reflection or transmission coefficients. Whole sets of converted rays cannot be modeled (creeping waves), or approximately modeled (diffracted waves), unless a specific asymptotic theory accounts for them (Keller, 1962).

On the other hand, numerical methods based on discretization of the wave equation synthesize the different waves arriving at the station as a whole. This can make physical interpretation of a forward calculation difficult, but it does not prevent inversion. The main limitation of numerical methods is their computational cost, in both time and memory requirements.

Inside a numerical grid in the space-time domain, the acoustic wave equation (or its equivalent) must be verified at different nodes, approximating the derivatives by finite-differences. Basically, the initial conditions give the value of the pressure field $p(\mathbf{x}, t_0)$ and its velocity $\dot{p}(\mathbf{x}, t_0)$ everywhere in space at time t_0. The wave equation then gives the value at time t_0 of the acceleration of the field $\ddot{p}(\mathbf{x}, t_0)$. Given the density $\rho(\mathbf{x})$ and the bulk modulus $K(\mathbf{x})$, from knowledge of $p(\mathbf{x}, t_0)$, $\dot{p}(\mathbf{x}, t_0)$, and $\ddot{p}(\mathbf{x}, t_0)$, it is possible to estimate $p(\mathbf{x}, t_1)$ and $\dot{p}(\mathbf{x}, t_1)$ with $t_1 = t_0 + \Delta t$. Iterating the procedure gives the values of the pressure field for any time $t_i = t_0 + i\Delta t$.

Since the pioneering work of Alterman and Karal (1968), two different formulations have arisen in the finite-difference literature: the homogeneous formulation and the heterogeneous formulation (Kelly et al., 1976). The homogeneous formulation solves the propagation equation in each homogeneous area, and it verifies explicit boundary conditions between the different areas. The heterogeneous formulation, on the other hand, directly solves the propagation equation in a heterogeneous medium, where physical properties are spatially variable. Therefore, conditions at interfaces are satisfied implicitly. The heterogeneous formulation gives results comparable with the homogeneous formulation when the order of the approximation at interfaces is the same as the order of approximation of the propagation equation. The homogeneous formulation may give more accurate results (Kummer and Behle, 1982), but it considerably increases the complexity of the algorithm for computing derivatives. Discretizing the elastodynamic equation instead of the acoustic equation will lead to the staggered grid used for modeling heterogeneous propagation of *SH* waves (Virieux, 1984) or *P-SV* waves (Virieux, 1986). The discrete values of the medium parameters correspond to a given physical quantity velocity for density and stresses for the bulk modulus K. The bulk modulus K is shifted one-half node with respect to the velocity, in both the horizontal and vertical directions. Because this method uses a first-order hyperbolic system, the computation is well-adapted to vector (and parallel) computers. Appendix A describes the equations used to solve the forward problem and the extensions to the forward problem necessary for the inversion. A complete description of the numerical method may be found in Virieux (1984).

THE INVERSE PROBLEM

An acoustic medium can be described using the density $\rho(\mathbf{x})$ and the bulk modulus $K(\mathbf{x})$. For simplicity, we assume the density known; the problem is to evaluate $K(\mathbf{x})$. We later use the terminology of functional analysis, so the unknown is *the function itself*, which is not assumed to be discretized. Nevertheless, for numerical computations it will be discretized. For instance, in the numerical examples, $K(\mathbf{x})$ is defined on a grid of 200×200 points. In more realistic problems, grids of $1\,000 \times 1\,000$ points should be used.

Let t be a time variable reset to zero at each new shot $\mathbf{x}_s$ ($s = 1, 2, \ldots, NS$) a generic source position, and $\mathbf{x}_r$ ($r = 1, 2, \ldots, NR$) a generic receiver position. The pressure at the receiver location $\mathbf{x}_r$ at time t for a shot at point $\mathbf{x}_s$ is denoted $p(\mathbf{x}_r, t; \mathbf{x}_s)$. Let $p(\mathbf{x}_r, t; \mathbf{x}_s)_{\text{obs}}$ denote the particular measured (observed) values. For a model $K(\mathbf{x})$, $p(\mathbf{x}_r, t; \mathbf{x}_s)_{\text{cal}}$ denotes the predicted seismograms.

The easiest formulation of the inverse problem is as follows: Which Earth's model $K(\mathbf{x})$ predicts seismograms $p(\mathbf{x}_r, t; \mathbf{x}_s)_{\text{cal}}$ which are closest to the observed seismograms $p(\mathbf{x}_r, t; \mathbf{x}_s)_{\text{obs}}$? The simplest results are obtained with a least-squares criterion of closeness. The problem is to obtain the model $K(\mathbf{x})$ for which the misfit function

$$S = \sum_{s=1}^{NS} \int_0^T dt \sum_{r=1}^{NR} \left[p(\mathbf{x}_r, t; \mathbf{x}_s)_{\text{obs}} - p(\mathbf{x}_r, t; \mathbf{x}_s)_{\text{cal}} \right]^2 \qquad (1)$$

is a minimum. This criterion can, of course, be generalized to take into account the estimated observational errors or the a priori information in the model space. Such details are left to Appendix B.

In principle, very different methods can be used to obtain the minimum of the (nonquadratic) functional S, such as (generalized) gradient methods or Monte Carlo (i.e., pseudo-random) methods. Gradient methods have the disadvantage of being local. That is, they converge to the nearest local minimum (if such local minima exist) instead of converging to the global minimum. However, they have the advantage of being tremendously effective. There are many good textbooks about gradient methods of minimization; Walsh (1975), Fletcher (1980), and Scales (1985) are well-documented, classical books, while Céa (1971) emphasizes infinite dimensional (i.e., functional) problems.

The simplest gradient method is the steepest descent method, which gives

$$K(\mathbf{x})_{n+1} = K(\mathbf{x})_n + \alpha_n \gamma(\mathbf{x})_n, \qquad (2)$$

where $\gamma(\mathbf{x})_n$ corresponds to the direction of the steepest descent for the misfit function S in the infinite dimensional model space, and where α_n is simply a constant scaling factor (either analytically estimated or chosen by trial and error). For more details, see Tarantola (1984c). The philosophy of the method is simple. Because $\gamma(\mathbf{x})_n$ is by definition a direction of descent for S, for a sufficiently small α_n the value of S for $K(\mathbf{x})_n + \alpha_n \gamma(\mathbf{x})_n$ must be smaller than its value for $K(\mathbf{x})_n$. Iterating long enough will lead to models with acceptably low values of the misfit function S, i.e., models whose predicted seismograms fit the observed seismograms acceptably well.

The main computational task at each iteration is evaluation of $\gamma(\mathbf{x})_n$. $K(\mathbf{x})_n$ denotes the current model we wish to update, and $p(\mathbf{x}, t; \mathbf{x}_s)_n$ denotes the pressure field predicted from this current model. Then $\delta p(\mathbf{x}_r, t; \mathbf{x}_s)_n = p(\mathbf{x}_r, t; \mathbf{x}_s)_{\text{obs}} - p(\mathbf{x}_r, t; \mathbf{x}_s)_n$ are the data residuals for the current model. A new field $\psi(\mathbf{x}, t; \mathbf{x}_s)_n$ is defined as follows. For a given source point $\mathbf{x}_s$, we consider all the data residuals $\delta p(\mathbf{x}_r, t; \mathbf{x}_s)_n$. At each point $\mathbf{x}_r$ where there was a receiver (for the given source point $\mathbf{x}_s$), we set a source whose time function is the residual $\delta p(\mathbf{x}_r, t; \mathbf{x}_s)_n$. For a given $\mathbf{x}_s$, all the points $\mathbf{x}_r$ radiate in-phase, and the propagation is made backward in time.

Then, as demonstrated by Lailly (1984) and Tarantola (1984b, c) and as outlined in Appendix C,

$$\gamma(\mathbf{x})_n = \frac{1}{K(\mathbf{x})_n^2} \sum_{s=1}^{NS} \int_0^T dt\ \dot{p}(\mathbf{x}, t; \mathbf{x}_s)_n \dot{\psi}(\mathbf{x}, t; \mathbf{x}_s)_n. \quad (3)$$

The physical interpretation of equation (3) is that $p(\mathbf{x}, t; \mathbf{x}_s)_n$ is the predicted field in the current model $K(\mathbf{x})_n$. Because $\psi(\mathbf{x}, t; \mathbf{x}_s)_n$ is obtained by propagating the data residuals backward in time, it represents the missing field, i.e., a field we need to obtain null data residuals. If, for a given shotpoint $\mathbf{x}_s$ at a given point $\mathbf{x}$ of the space, the missing field is correlated in time with the predicted field, this missing field may be created by adding a diffractor at point $\mathbf{x}$. Equations (2) and (3) imply setting the value of the diffractor at point $\mathbf{x}$ as proportional to the time correlation of the time derivatives of the predicted and missing fields. Equation (3) shows that these time correlations have to be computed at each point $\mathbf{x}$ of the space for each of the shots $\mathbf{x}_s$; then the results for each of the shots have simply to be added.

The following is our operational approach. We start with some arbitrary model $K(\mathbf{x})_0$, the closer to the true model, the better. Since $K(\mathbf{x})_0$ is simply the first current model $K(\mathbf{x})_n$, we keep the index n for more generality. We select the first shotpoint $\mathbf{x}_s = \mathbf{x}_1$ and solve the forward problem using the finite-difference scheme introduced above. The result is the field $p(\mathbf{x}, t; \mathbf{x}_1)_n$. In particular, we obtain the predicted seismograms $p(\mathbf{x}_r, t; \mathbf{x}_1)_n$ and the residuals $\delta p(\mathbf{x}_r, t; \mathbf{x}_1)_n = p(\mathbf{x}_r, t; \mathbf{x}_1)_{\text{obs}} - p(\mathbf{x}_r, t; \mathbf{x}_1)_n$ corresponding to the current model for the first shotpoint. These residuals are propagated (simultaneously for all values of r) backward in time, again using the finite-difference scheme, thus giving the field $\psi(\mathbf{x}, t; \mathbf{x}_1)_n$. At each point $\mathbf{x}$ of the space we compute

$$U(\mathbf{x}; \mathbf{x}_1) = \int_0^T dt\ \dot{p}(\mathbf{x}, t; \mathbf{x}_1)_n \dot{\psi}(\mathbf{x}, t; \mathbf{x}_1)_n. \quad (4)$$

The process is repeated for the second shotpoint $\mathbf{x}_s = \mathbf{x}_2$, thus yielding $U(\mathbf{x}, \mathbf{x}_2)$; the process continues until all shotpoints have been considered. $\gamma(\mathbf{x})_n$ is then obtained from the sum

$$\gamma(\mathbf{x})_n = \frac{1}{K(\mathbf{x})_n^2} \sum_{s=1}^{NS} U(\mathbf{x}, \mathbf{x}_s). \quad (5)$$

If we do not have an estimate of the correct value for α_n, we select a few reasonable values (three values, for instance). For each of the selected values of α_n we compute the corresponding value of the misfit function S for $K(\mathbf{x})_{n+1} = K(\mathbf{x})_n + \alpha_n \gamma(\mathbf{x})_n$. A good value for α_n can be chosen by interpolation, and the updated model $K(\mathbf{x})_{n+1}$ is then obtained. This will be the current model for the next iteration (in Appendix C we discuss an alternate, linearized approach for obtaining an adequate value for α_n). Typically, we stop iterations when synthetic seismograms look like the observed ones, or when little advance in the value of the misfit function S is gained between subsequent iterations.

Technical details of the computations have to be modified for an optimal (and practical) utilization of available computer resources. The modifications are left for Appendixes A and B.

NUMERICAL EXAMPLES

Point diffractor

Take as the true medium a single diffracting point ($\delta K \neq 0$) superimposed on a homogeneous medium (K_0) which is also the starting medium. For now, assume that the four edges are absorbing. For a single source, seismograms are recorded at receivers located at each point of the grid just below the surface of the medium. Figure 1 shows the gradient $\gamma(\mathbf{x})$. It is peaked at the true diffracting point, but spreads over several nodes. The "smile" is due to the particular geometry considered (a single source and a line of receivers). Note that this figure does *not* represent the solution of the inverse problem, but only the first iterate.

Considering discrete receivers may lead to some complications: with one receiver at each point of the grid or equally spaced receivers, the waves generated by the residuals are coherent. However, if some receivers are missing, the residuals may destructively interfere and produce some unexpected spikes in the gradient.

In Figure 2, nine sources were used. The peak now clearly predominates over artifacts. The true diffractor is located with a spatial resolution corresponding to the main wavelength of the source.

In the next example, consider a free surface instead of an absorbing surface. Source and receivers are located a few grid points below the surface; Figure 3 shows the corresponding gradient. Essentially the same features are present as in Figure 2. However, the amplitude of the secondary lobes is stronger due to ghosts in the source that are produced by the surface reflection. To a lesser extent, ghosts are also due to multiply reflected (or diffracted) waves between the scatterer and the surface. The amplitude of the artifacts is roughly inversely proportional to the number of sources. Again, the figure represents the first iterate, not the result of the inversion.

Figure 4 illustrates a case with two diffractors. The true model is a homogeneous medium with two point diffractors of equal amplitude ($\delta K_1 = \delta K_2 \neq 0$). The peaks of the gradient have very different amplitudes; the diffractor near the sources and receivers appear predominant. In this example of simulated seismic reflection data, this leads to a decrease of the amplitude of the gradient roughly proportional to the square root of the distance to the surface. The gradient computed here gives the direction of descent for the misfit function S, but the direction of steepest descent is only optimal for infinitesimal moves (which implies an infinite number of iterations). Some preconditioning of the gradient may accelerate convergence in any practical algorithm. The result shown in Figure 4 suggests replacing equation (5) with

$$\gamma(\mathbf{x})_n = \frac{1}{K(\mathbf{x})_n^2} \sum_{s=1}^{NS} \frac{U(\mathbf{x}, \mathbf{x}_s)}{\|\mathbf{x} - \mathbf{x}_s\|^{1/2}}. \quad (6)$$

"Camembert" model

To see what happens when the size of the perturbation is much larger than the mean wavelength of the signal, consider a model composed of a circular perturbation, the diameter of which is about ten wavelengths, superimposed on a homogeneous medium (K_0). As shown in Figure 5, we use a grid of 200×200 points to describe our model, which is thus defined by the values of $K(\mathbf{x})$ at 40 000 points. The relative perturbation of K in the disk (with respect to the surrounding medium) is $\varepsilon = \delta K/K$.

Before performing any inversion, we look at the range of ε in which the problem is linearizable around the medium K_0 to determine where to start our iterative algorithm. We first placed sources and receivers all around the Camembert, and the computation of $S(K)$ for different perturbations ε leads to

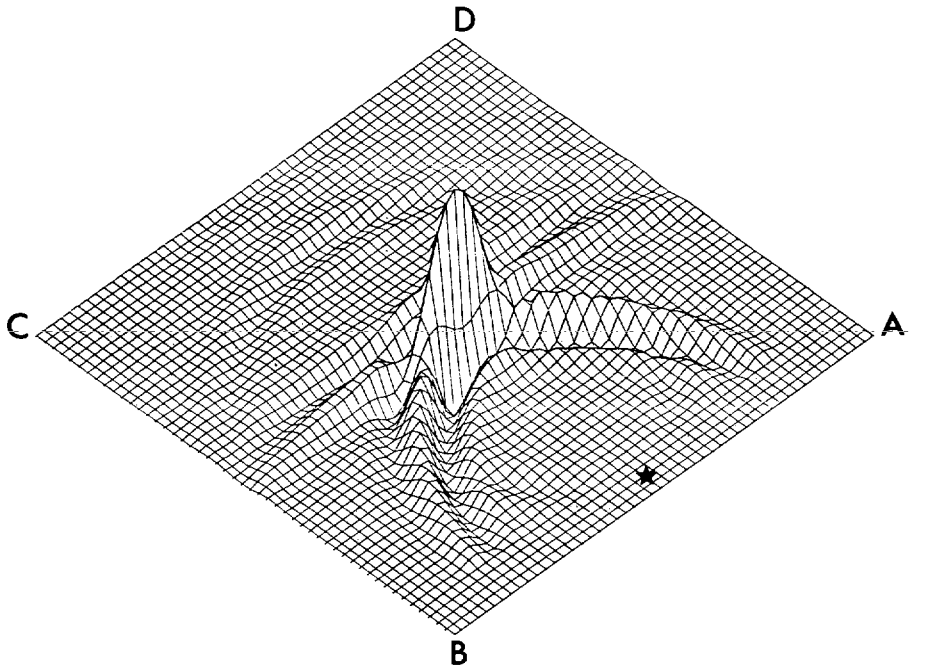

FIG. 1. A "true" model has been defined by an homogeneous medium with a single point diffractor ($\delta K \neq 0$). The medium is defined by a grid of 50 × 50 points, and the four edges satisfy absorbing boundary conditions. A single source has been shot (a star indicates its location), and the seismograms recorded at each grid point on line A-B. This data set is used for a synthetic inversion. The starting model is an homogeneous medium with the right background values. The figure shows the gradient obtained at the first iteration. It is peaked at the right position, but shows spatial spreading and some artifacts.

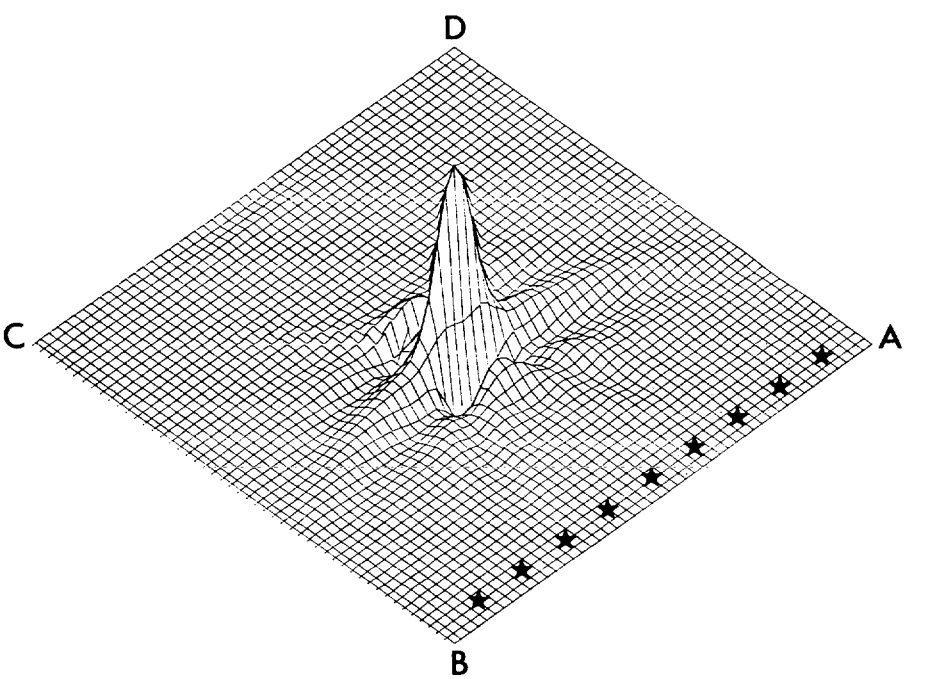

FIG. 2. Same as Figure 1 but using nine sources (stars). The smile of Figure 1 disappears.

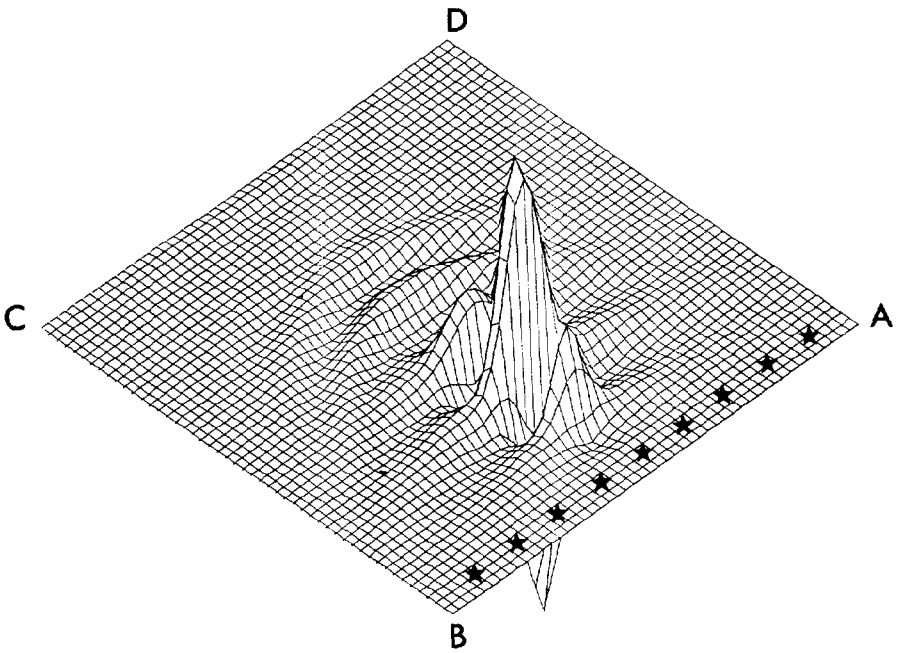

FIG. 3. Same as Figure 3 but using a free surface boundary condition above sources and receivers. The ghosts (surface reflections) of the source introduce some artifacts.

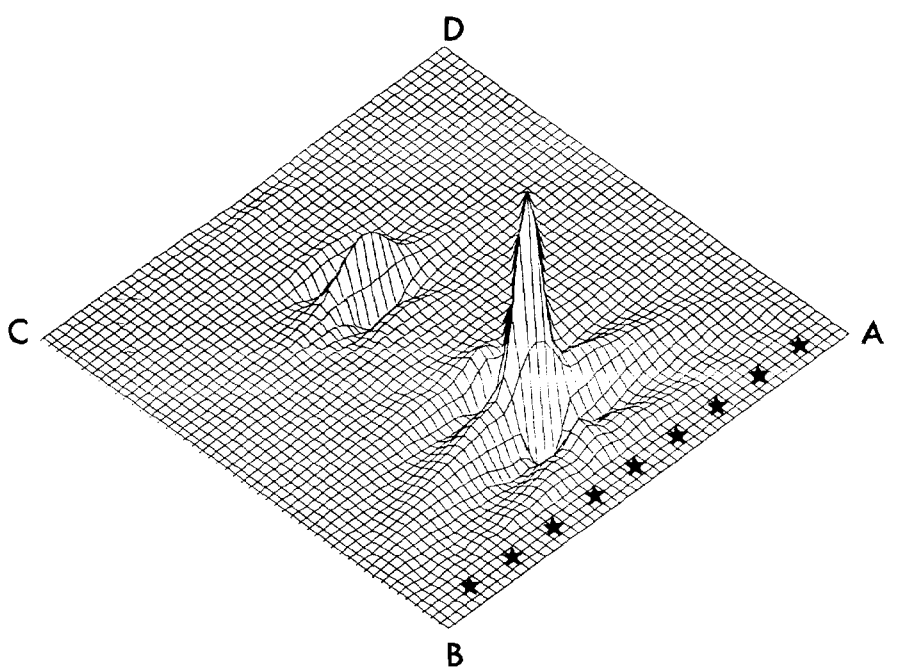

FIG. 4. Same as in Figure 2, but the true model now contains two identical diffractors. The amplitude of the peaks is roughly proportional to the square root of the distance to the surface. This suggests a "preconditioned" gradient, as discussed in the text.

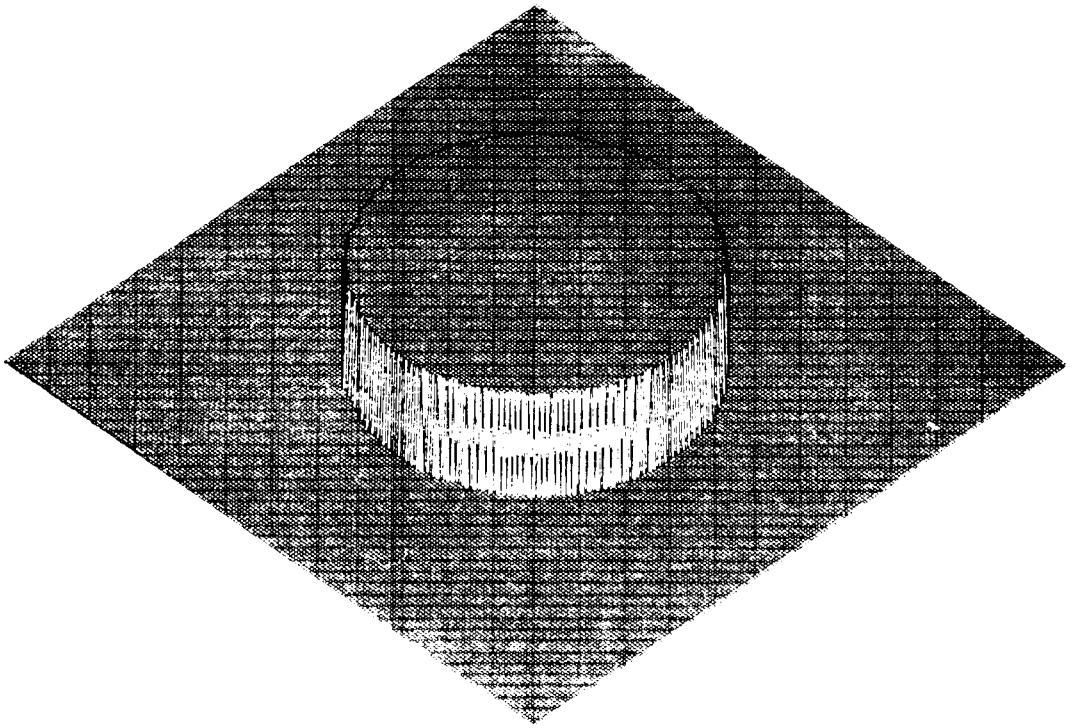

FIG. 5. The model is now a circular inclusion (the "Camembert") in a homogeneous medium. The size of the Camembert is about 10 wavelengths. The model is numerically defined in a grid with 200 × 200 points, so the model contains 10^4 parameters (unknowns for the inversion).

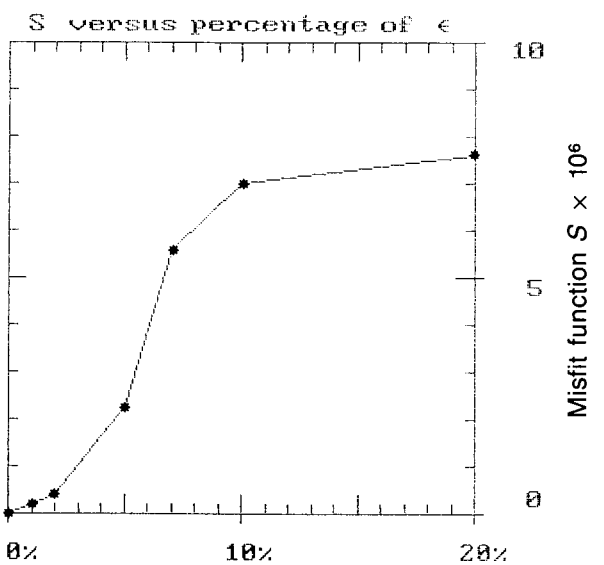

FIG. 6. Using sources and receivers all around the Camembert (see Figure 11), the seismograms have been computed for different values of the perturbation $\varepsilon = \delta K/K$ inside the disk. The figure represents the value obtained for the misfit function S when comparing the seismograms obtained with a given value of ε with the seismograms obtained for $\varepsilon = 0$. For values of ε great than 10 percent, the function S clearly departs from a parabola. The saturation of S corresponds to a shift of the seismograms of the order of a wavelength. Thus, for $\varepsilon > 10$ percent, the problem is strongly nonlinear.

Figure 6. As soon as ε reaches about 10 percent, the values of S saturate and the curve clearly diverges from a quadratic function. This behavior corresponds to the phase shift of seismograms reaching the mean period of the signal.

In the finite-difference propagations, we use the following numerical values: $\Delta x = \Delta z = 5$ m, $\Delta t = 1.15$ ms, $f_{\text{source}} \simeq 50$ Hz, $K_0 = 0.25 \times 10^{11}$ N/m^2, $L_0 = 0.25 \times 10^{-3}$ m^3/kg. Lines of receivers are defined with one receiver every two points of the grid. Absorbing boundary conditions are set at the four edges of the medium. We generate the data using eight sources, and the relative position of the sources and lines of receivers will be chosen to deal with different geometries. The total propagation time T will be chosen to cut the reflections from the corners, due to imperfectly absorbing edges. The medium is assumed perfectly known in a band of width λ around the sources and receivers (so that the gradient is not computed there).

The following examples present the results of the inversion using different source-receiver configurations, and different values of ε.

Seismic reflection data

Eight sources and 100 receivers are located a few grid points below the surface. Figure 7 shows the synthetic seismograms for one of the shots (at corner A of Figure 8). The direct wave has been suppressed in the representation, so that these seismograms correspond to the starting *residuals* with respect to the homogeneous model (K_0). Primary reflections at the top of the Camembert are followed by primary reflections at the bottom. Between them, small wrinkles are the surface manifes-

tation of the curved shape of the Camembert.

A complete physical description is difficult, so we give an intuitive point of view. Each point of the edge of the Camembert acts as a diffractor and gives a hyperbolic signature at the surface. The sum of the horizontal top and bottom of the Camembert gives the strong primary reflections, while isolated signatures from other points of the edge are the previously noted wrinkles. We expect the signature to be strong for the vertical part of the edge in transmission. Reflected and creeping waves propagating at curved interfaces deserve a complete study where different methods such as finite-difference methods or ray theory are used for a correct interpretation (George et al., 1986).

Figure 8 shows the first iteration, while Figure 9 shows the fifth iteration. The envelope of the perturbation is accurately found after the first iteration, although the bottom necessarily is not accurately described because the velocity has not changed in the interior of the disk. Vertical edges are also located. Iteration number 5 does not greatly improve the result of iteration number 1. In particular, the low spatial frequency content is not recovered after five iterations. The artifacts around the main diffracting zones and at the bottom of the model are due to the small number of sources used and to parasitic reflections coming from the imperfectly absorbing edges. The final residuals (not shown) present essentially the same features as the initial ones, with smaller amplitude. The misfit function S, given by equation (1), takes the values 14.2×10^3, 7.5×10^3, 5.7×10^3, 4.9×10^3, 4.3×10^3, and 3.9×10^3, respectively.

Tomographic data

Sources and receivers are placed all around the disk. Figure 10 shows the data corresponding to one of the eight sources, for which receivers are placed at the same and opposite edges.

The results of the first, third, and fifth iterations are shown in Figure 11, 12, and 13 respectively. All the essential features of the model are recovered well by the inversion. In particular, unlike the previous example, the low spatial frequency component is obtained. There remain some artifacts in certain parts of the medium, particularly at the corners, which are not well-resolved because of the number and positioning of the eight sources. The average amplitude of the inverted Camembert is about 90 percent of the true value. The misfit function S takes the values 38.7×10^6, 10.0×10^6, 3.5×10^6, 1.4×10^6, 0.7×10^6, and 0.5×10^6, respectively. After five iterations, the final value of S is about 1 percent of its initial value.

Highly nonlinear inversion

To explore the limitations of inversion, we study the tomographic configuration, which has a perturbation of $\varepsilon = 20$ percent. In Figure 14, the initial residuals show phase shifts of the order of λ. We performed five iterations, as in the previous case. Figures 15 and 16 show the first and fifth iterations, respectively. The misfit function S takes the values 12.4×10^7, 11.9×10^7, 11.2×10^7, 10.5×10^7, 9.9×10^7, and 9.4×10^7, respectively.

As for all nonlinear problems, the present problem may have secondary minima. The results of Figure 16 suggest that this is indeed the case here. Due to the strong nonlinearity of this problem, correlation between the incident field and residual field takes the wrong sign in the vicinity of the disk, and the gradient iterative algorithm is trapped. The only way to deal with multiple solutions is to start the iterations at different points and check for a "deep" global minimum. Limited computer time prevented us from searching for a global minimum. Imposing a priori smoothness on our solution, or using better preconditioning, might help avoid the problem of local minima.

CONCLUSION

The two-dimensional non-linear inversion of seismic waveforms is feasible even if it takes a lot of computer time (3 500 s on a CRAY 1S for the example of Figure 13).

For a problem using only surface sources and only surface-recorded reflected waves, the proposed gradient algorithm cannot recover the low spatial frequencies of the model, at least with the few iterations (five) performed here. Only the high spatial frequencies of the model are mapped, and the inaccuracies are due to the errors in the low spatial frequencies (and thus of the velocity field). The situation is simi-

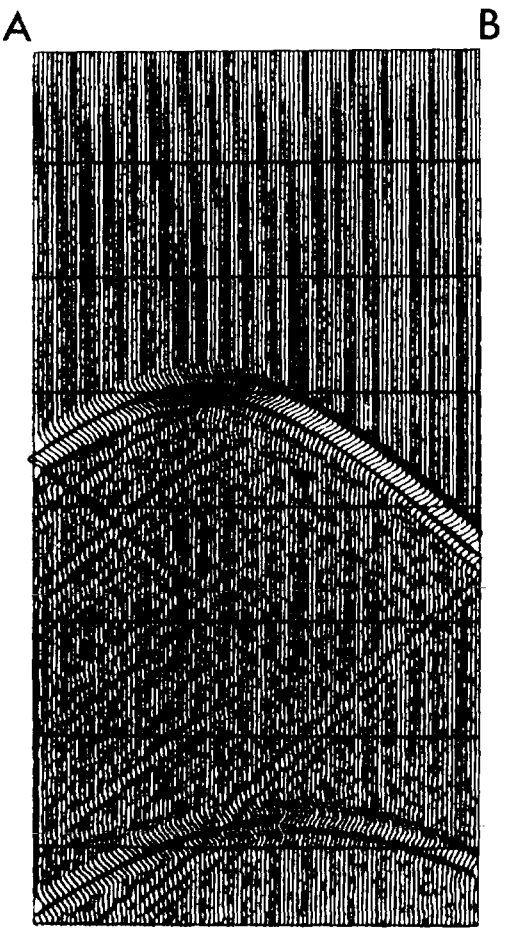

FIG. 7. For the first inversion, eight sources were used (the stars in Figure 8), and each source has been "observed" by 100 receivers in line A-B (one receiver at every other grid point). This figure shows the data corresponding to one of the sources (the star near corner A). The direct wave has been substracted. A value of $\varepsilon = 5$ percent was used.

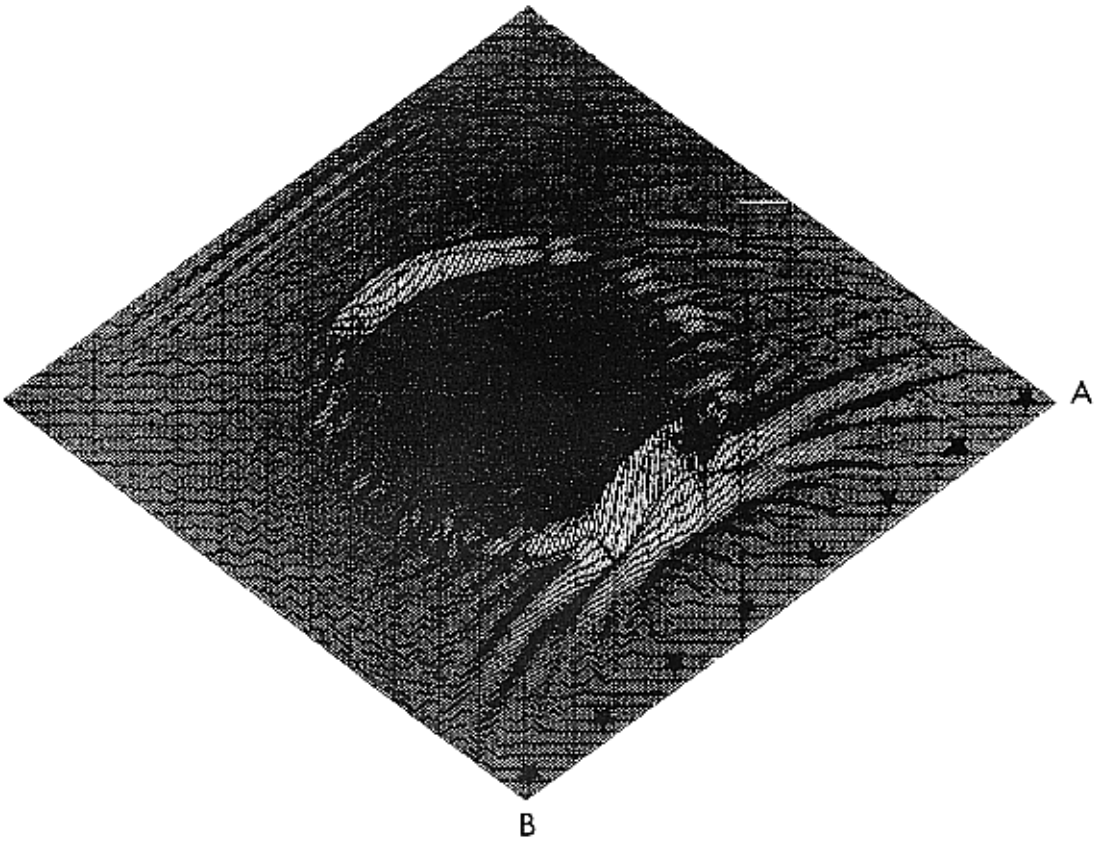

FIG. 8. Model obtained at the first iteration. The scale is 0.33 compared to Figure 5.

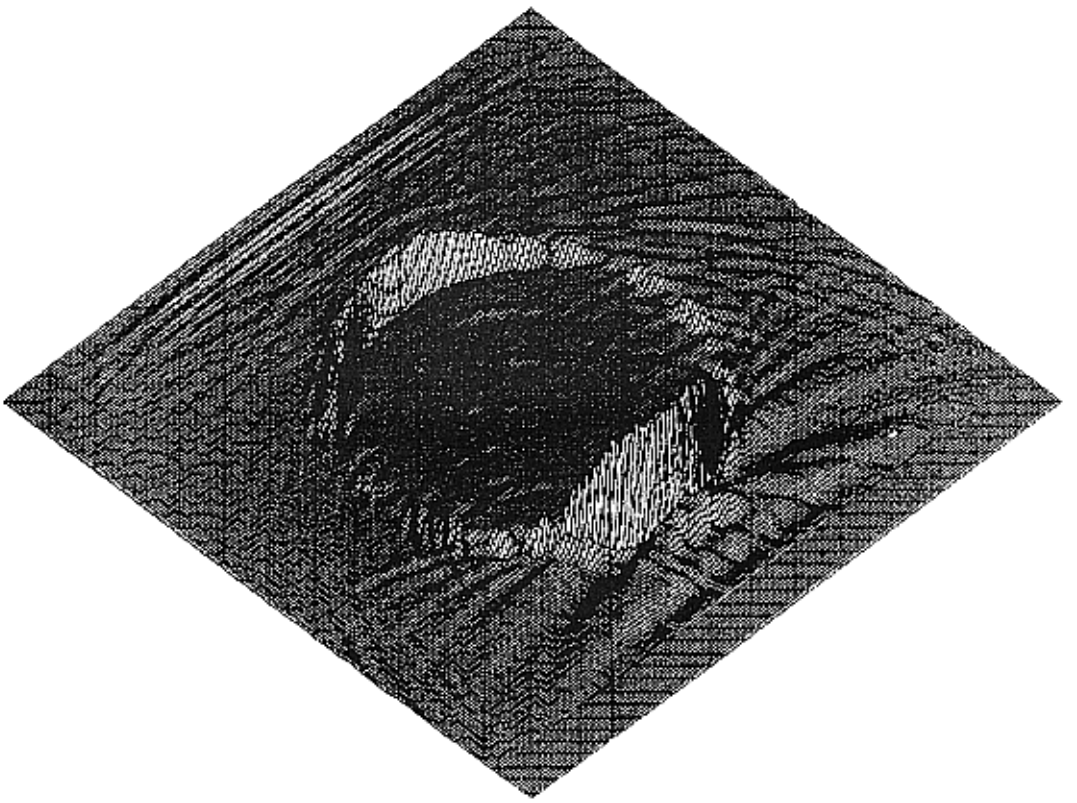

FIG. 9. Model obtained after five iterations. The geometry of the Camembert is well-recovered, but the low spatial frequency content is not. The gradient algorithm converges very slowly with surface seismic reflection data configuration.

lar to migration with an incorrect velocity. It does not seem that the algorithm has been trapped in a secondary minimum; rather, the experiment was too time-consuming to iterate further. Our preliminary conclusion is that, although in principle the formulation of the problem allows for all the spatial frequencies of the model to be obtained, in practice only the high frequencies converge rapidly enough. The numerical experiments by Kolb et al. (1986) for 1-D models (with multioffset data) confirm that by iterating long enough the low frequencies can be recovered also. In other words, the components of the direction of steepest descent corresponding to the high spatial frequency content of the model are reliable, but the components corresponding to the low spatial frequency content are not. Because each iteration of a gradient method needs the resolution of twice as many forward problems as there are sources, it is out of the question to iterate more than twice with modern computers. The solution of the problem will probably be found through better use of the traveltime information in the data set.

Using both reflection and transmission data, the results give us confidence in the recovery of medium parameters for offset seismic profiles.

If the starting model is far enough from the true model, secondary solutions exist, but they are easily identified.

Future interpretation of seismic records will certainly be performed through very accurate forward modeling. The method described here shows that accurate forward modeling codes can be used for waveform matching.

ACKNOWLEDGMENTS

We thank V. Farra, T. George, P. Lailly, P. Mora, and A. Pica for helpful discussions and suggestions. This work was supported by the sponsors of the Equipe de Tomographie Géophysique of the Institut de Physique du Globe de Paris (CGG, IFP, Schlumberger, SNEA, TOTAL CFP), by the CNRS ATP Géophysique Appliquée, and by the INSU ASP Tomographie Géophysique. We particularly thank TOTAL CFP (J. J. Raoult and M. Riguidel) for providing computing time on the CRAY 1S.

REFERENCES

Alterman, Z., and Karal, F. C., 1968, Propagation of elastic waves in layered media by finite-difference methods: Bull., Seis. Soc. Am., **58**, 367–398.

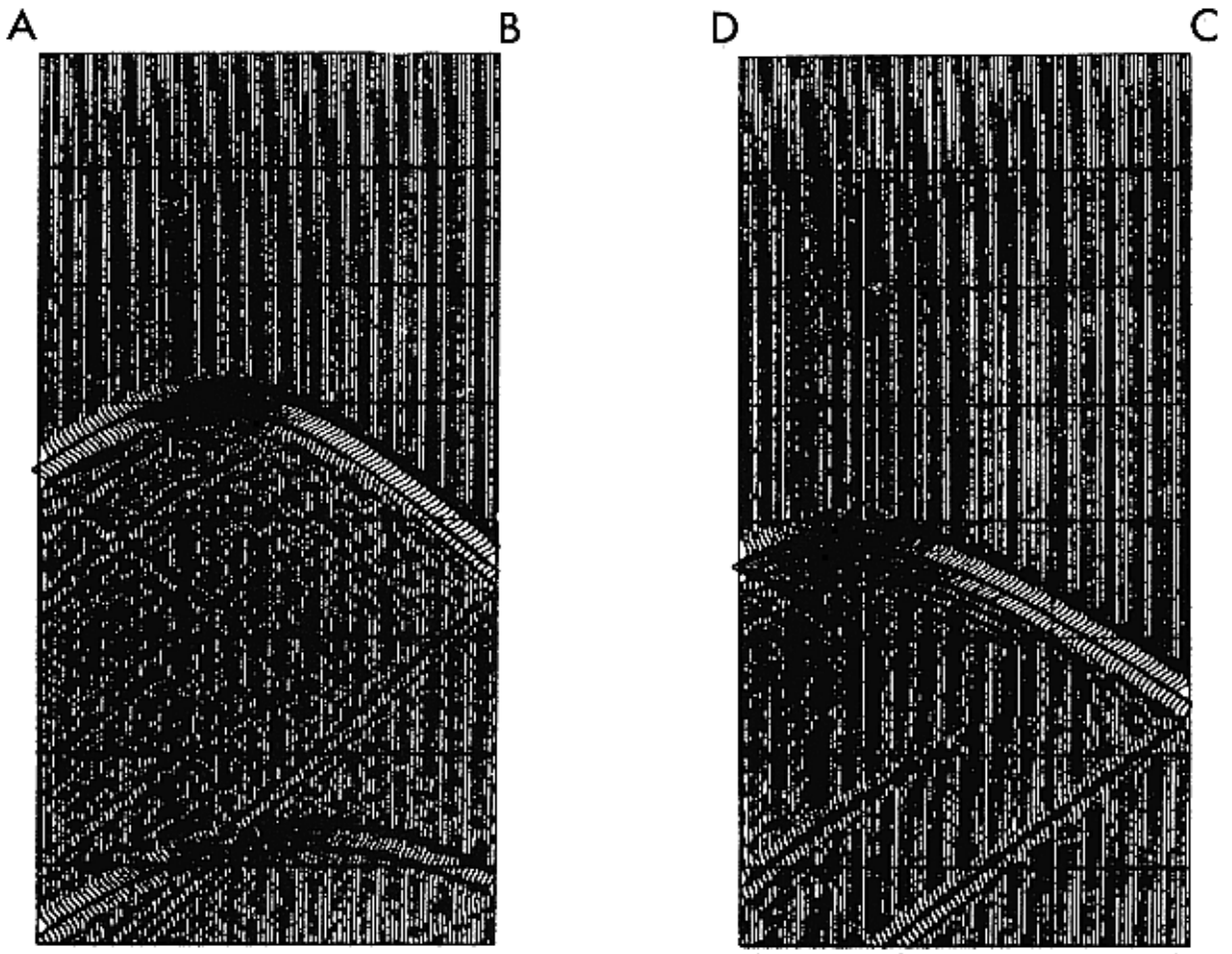

FIG. 10. For the second inversion, eight sources were also used, but they were placed all around the Camembert (see Figure 11). Each source is also observed all around (400 receivers). This figure shows some of the seismograms corresponding to one of the sources (the star near corner A). A value of $\varepsilon = 5$ percent was used.

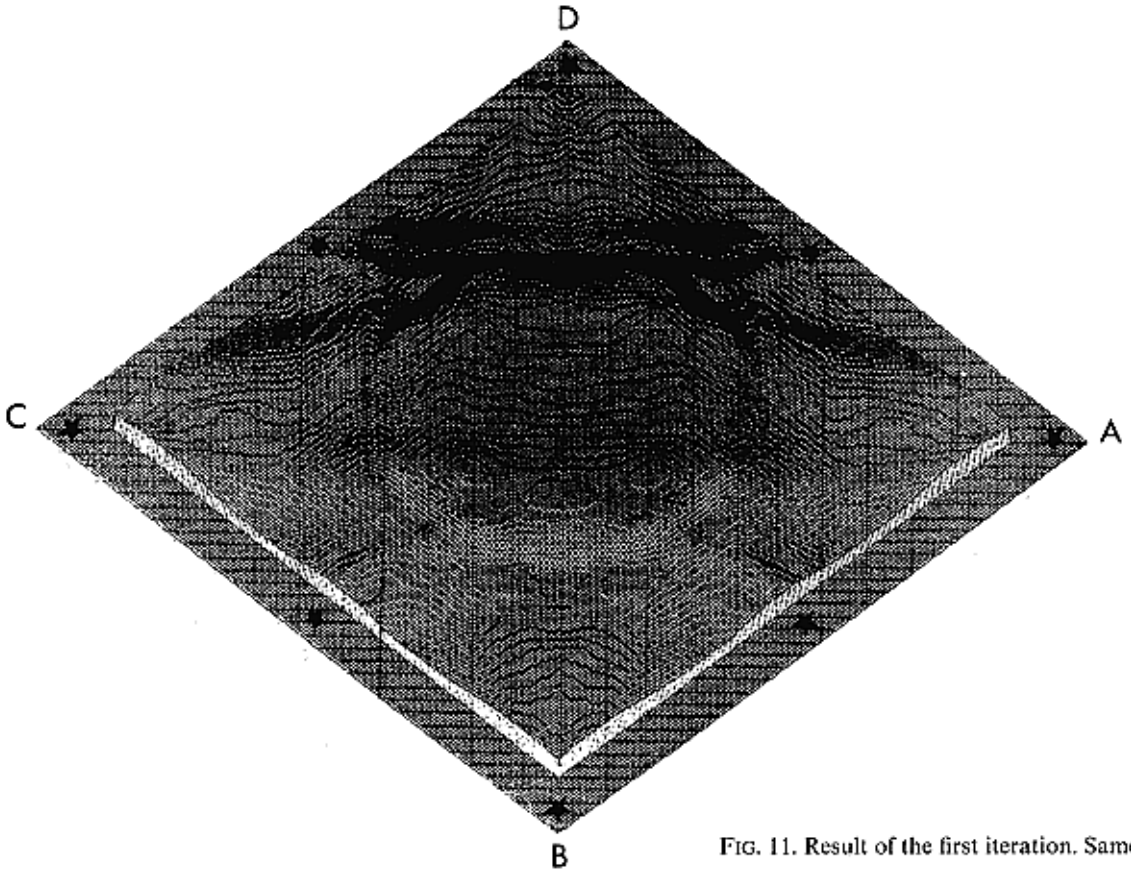

FIG. 11. Result of the first iteration. Same scale as Figure 5.

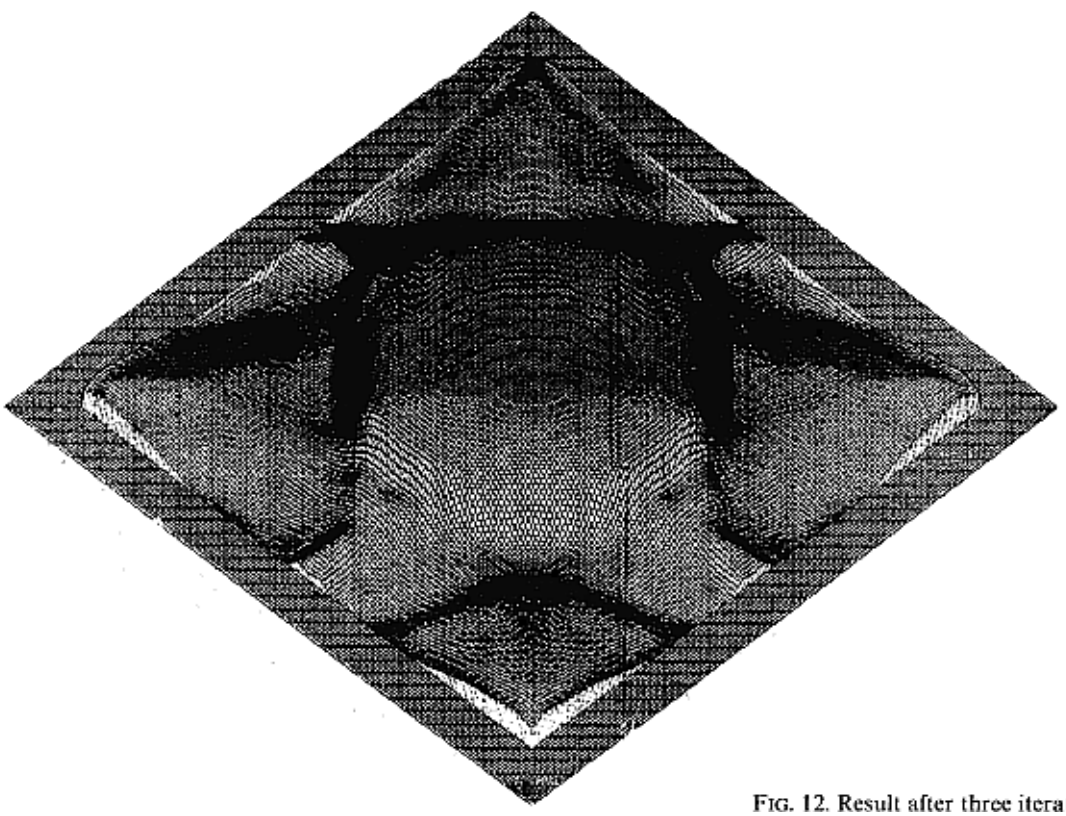

FIG. 12. Result after three iterations.

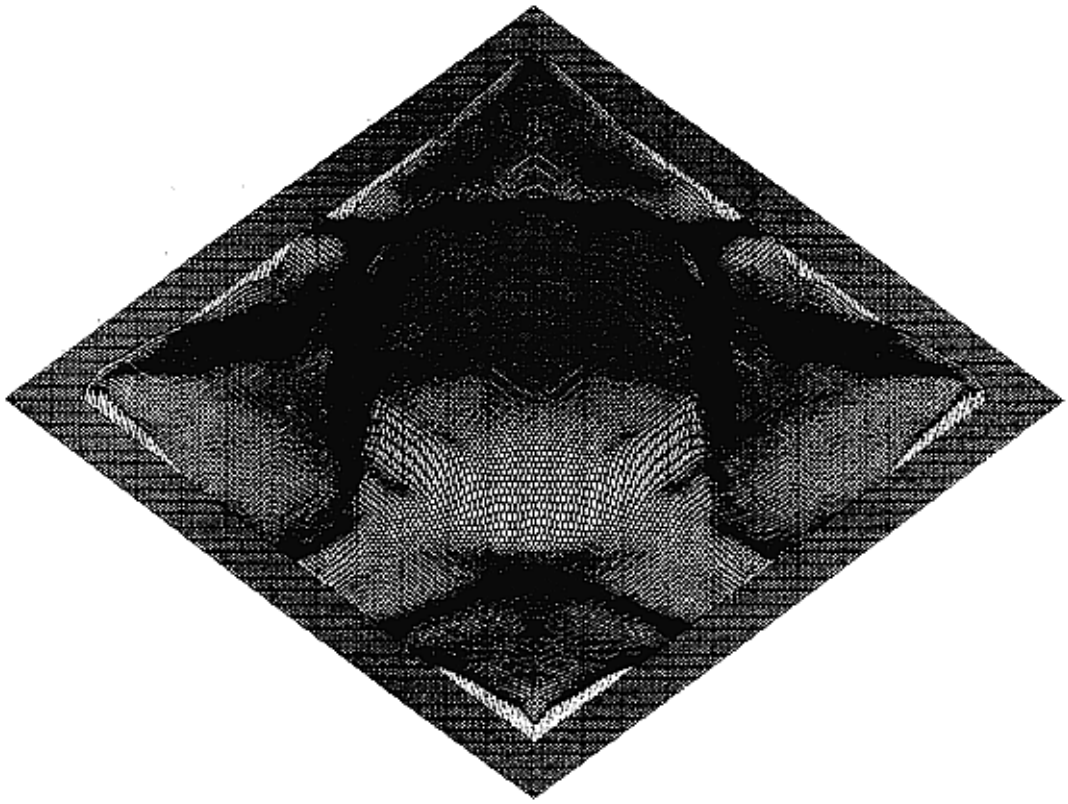

FIG. 13. Result after five iterations. This tomographic example has a highly redundant data set, containing in particular transmitted waves. Both the geometry of the Camembert and the low spatial frequencies are well-recovered by the inversion. Some numerical artifacts remain.

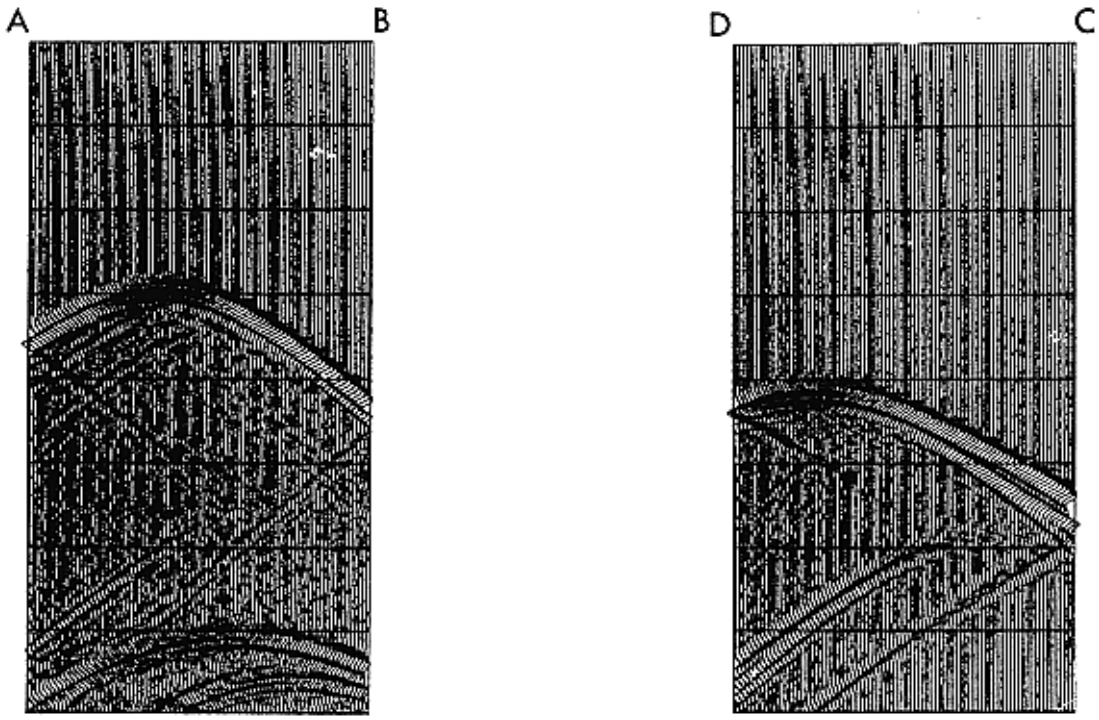

FIG. 14. Same as Figure 10, but a value of $\varepsilon = \delta K/K = 20$ percent was used. The problem is highly nonlinear.

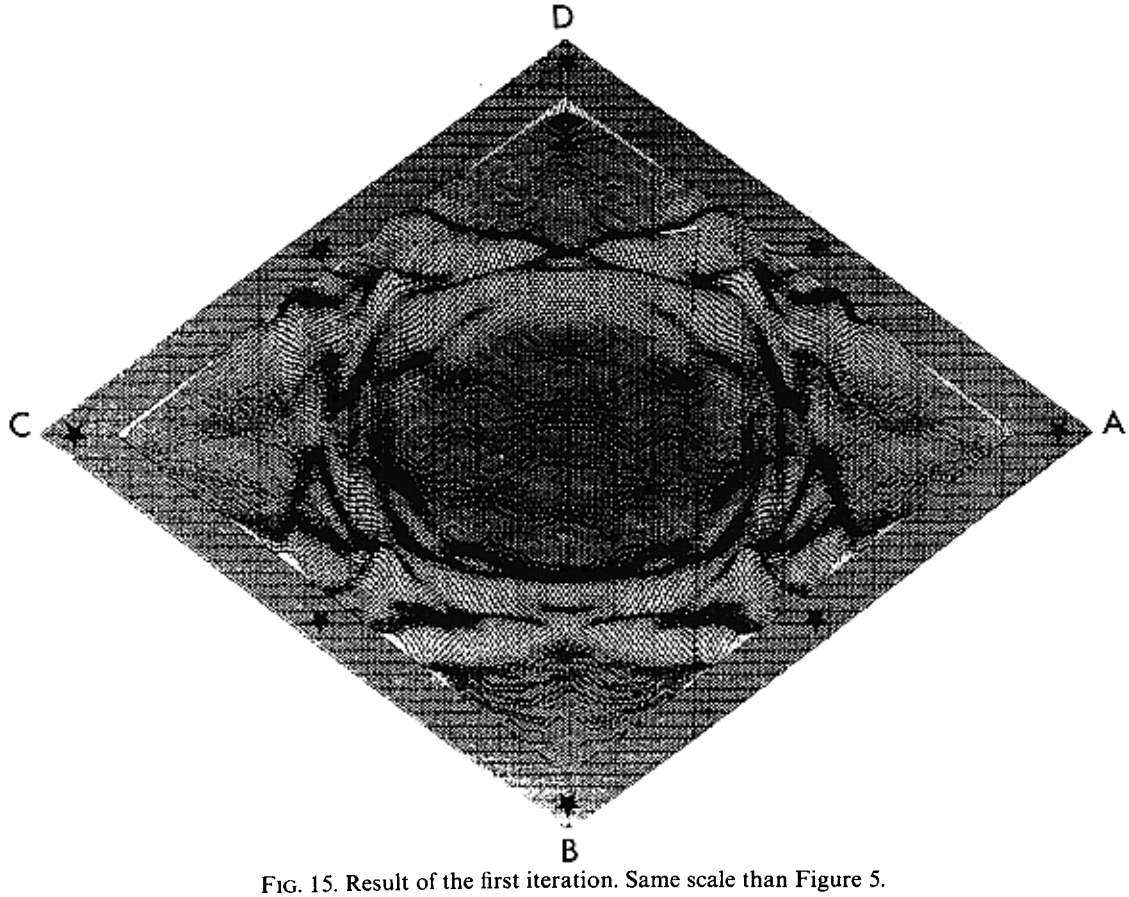

FIG. 15. Result of the first iteration. Same scale than Figure 5.

FIG. 16. Result after five iterations. This figure suggests that the algorithm converges into a secondary minimum of the misfit function. This kind of secondary minimum may probably be avoided by introducing some a priori smoothing condition on the model.

Céa, J., 1971, Optimisation théorie et algorithmes: Dunod.
Chapman, C. H., and Orcutt, J. A., 1985a, Least-squares fitting of marine seismic refraction data: Geophys. J. Roy. Astr. Soc., **82**, 339–374.
– – 1985b, The computation of body wave synthetic seismograms in laterally homogeneous media: Rev. Geophys., **23**, 105–163.
Clayton, R. W., and Stolt, R. H., 1981, A Born-WKBJ inversion method for acoustic reflection data: Geophysics, **46**, 1559–1567.
Fawcett, J. A., and Clayton, R. W., 1984, Tomographic reconstruction of velocity anomalies: Bull., Seis. Soc. Am., **74**, 2201–2219.
Fletcher, R., 1980, Practical methods of optimization, **1**: Wiley and Sons.
George, T., Virieux, J., and Madariaga, R., 1986, Gaussian beam of complex structures: comparison with other techniques, in press.
Keller, J. B., 1962, Geometrical theory of diffraction: J. Opt. Soc. Am., **52**, 116–130.
Kelly, K. R., Ward, R. W., Treitel, S., and Alford, R. M., 1976, Synthetic seismograms: A finite-difference approach: Geophysics, **41**, 2 27.
Kolb, P., Collino, F., and Lailly, P., 1986, Prestack inversion of a 1-D medium: Proceedings Inst. Electr. Electron. Eng. special issue on seismic inversion.
Kummer, B., and Behle, A., 1982, Second-order finite-difference modeling of SH-wave propagation in laterally inhomogeneous media: Bull., Seis. Soc. Am., **72**, 793–808.
Lailly, P., 1984, The seismic inverse problem as a sequence of before stack migrations, *in* Santosa, F., Pao, Y.-H., Symes, W. W., and Holland, C., Eds., Inverse problems of acoustic and elastic waves: Soc. Industr. Appl. Mech., 104–181.
Morse, P. M., and Feshbach, H., 1953, Methods of theoretical physics: McGraw-Hill Book Co.
Mora, P., 1984, Elastic inversion using ray theory: Stanford Exploration Project, rep. 41, Stanford, Univ.
Scales, L. E., 1985, Introduction to nonlinear optimization: Macmillan.
Tanimoto, T., 1984, Waveform inversion of mantle Love waves: the Born seismogram approach: Geophys. J. Roy. Astr. Soc., **78**, 641–660.
Tarantola, A., 1984a, Linearized inversion of seismic 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. Mech., 104–181.
Taylor, A. E., and Lay, D. C., 1980, Introduction to functional analysis: Wiley and Sons.
Virieux, J., 1984, *SH*-wave propagation in heterogeneous media, velocity-stress finite-difference method: Geophysics, **49**, 1933–1957.
1986, *PSV*-wave propagation in heterogeneous media, velocity-stress finite-difference method: Geophysics, **51**, 000–000.
Walsh, G. R., 1975, Methods of optimization: Wiley and Sons.
Woodhouse, J. H., and Dziewonski, A. M., 1984, Mapping the upper mantle: three dimensional modeling of earth's structure by inversion of seismic waveforms: J. Geophys. Res., **89**, 5953–5986.

APPENDIX A

THE FORWARD PROBLEM

Finite-difference scheme

To use a staggered velocity-stress finite-difference method (Virieux, 1984), instead of the wave equation, discretize the system

$$\frac{\partial p(x,\, t)}{\partial t} = K(\mathbf{x})\ \mathrm{div}\ (\mathbf{w}(\mathbf{x},\, t)) + \tilde{S}(\mathbf{x},\, t;\, \mathbf{x}_s), \tag{A-1}$$

and

$$\frac{\partial \mathbf{w}(\mathbf{x},\, t)}{\partial t} = L(\mathbf{x})\ \mathbf{grad}[p(\mathbf{x},\, t)], \tag{A-2}$$

where $p(\mathbf{x},\, t)$ is the pressure in the medium and $\mathbf{w}(\mathbf{x},\, t)$ the particle velocity. K is the bulk modulus, while L is the inverse of density, namely, the lightness. Equation (A-1) represents Hooke's law differentiated with respect to time, and the equation (A-2) is the fundamental equation of dynamics. The source term $\tilde{S}(\mathbf{x},\, t)$, included in the scheme as a second member of Hooke's law, corresponds to an explosive source located at $\mathbf{x} = \mathbf{x}_s$ when $\tilde{S}(\mathbf{x},\, t) = \delta(\mathbf{x} - \mathbf{x}_s)\tilde{S}(t)$. Differentiating equation (A-1) with respect to time and introducing equation (A-2) leads to the following wave equation, obeyed by the pressure $p(\mathbf{x},\, t)$

$$\frac{1}{K(\mathbf{x})}\frac{\partial^2 p}{\partial t^2}(\mathbf{x},\, t;\, \mathbf{x}_s) - \mathrm{div}\left[\frac{1}{\rho(\mathbf{x})}\mathbf{grad}\ p(\mathbf{x},\, t;\, \mathbf{x}_s)\right] = S(\mathbf{x},\, t;\, \mathbf{x}_s), \tag{A-3}$$

where

$$S(\mathbf{x},\, t;\, \mathbf{x}_s) = \frac{1}{K(\mathbf{x})}\frac{\partial \tilde{S}(\mathbf{x},\, t;\, \mathbf{x}_s)}{\partial t}. \tag{A-4}$$

If $S(\mathbf{x},\, t;\, \mathbf{x}_s) = 0$ for $t \leq 0$ with identical initial conditions and boundary conditions, the pressure solution of our system is identical to the pressure solution of the wave equation provided the relation (A-4) between source terms if fulfilled. We simply reformulated the acoustic problem inside the frame of the *SH* propagation equations used by Virieux (1984). No benefits for the inverse problem are obtained by this manipulation.

The numerical scheme belongs to the explicit heterogeneous formulation of finite-difference techniques. The precision of the scheme is of second-order, implying the rule of thumb of ten nodes per wavelength. This numerical scheme allows generation of pressure waves [and, correspondingly, a set of data $p_0(\mathbf{x}_G,\, t;\, \mathbf{x}_S)$] for any given medium ($K$, L) and any geometry of shots and receivers.

Extensions of the forward problem

As shown in the text, computation of the direction of the steepest descent γ_n requires (1) knowledge of the field $p_n(\mathbf{x},\, t;\, \mathbf{x}_S)$, which is the solution of the forward problem; (2) knowledge of the field $\psi_n(\mathbf{x};\, t;\, \mathbf{x}_S)$ obtained by propagating the residuals $\delta p_n(\mathbf{x}_G,\, t;\, \mathbf{x}_S)$ backward in time (Appendix B), with the initial conditions $\psi_n(\mathbf{x},\, T;\, \mathbf{x}_s) = \dot{\psi}_n(\mathbf{x},\, T;\, \mathbf{x}_s) = 0$; and (3) the correlation of the two fields p_n and ψ_n at each point of the medium.

Computing p_u, then ψ_n, and correlating them after propagation would require storing the whole time history of the p_n fields. A better way is to perform the correlation for each source, starting from the final time T. ψ_n is computed by a single propagation (with $\Delta t < 0$), starting from the final time of the computation T. Since the wave equation is reversible in time, p_n can be recovered by moving backward in time, taking as the "initial" state the field $p_n(\mathbf{x}, T; \mathbf{x}_S)$ and using the same source $\delta(\mathbf{x} - \mathbf{x}_S)S(t)$. For this field p_n, explicit boundary conditions such as Neumann or Dirichlet conditions are left unchanged; in the case of absorbing boundary conditions the values of $p_n(\mathbf{x}, t; \mathbf{x}_s)$, which can be known for all $t \in [0, T]$ from the forward propagation, must be imposed on the absorbing boundary. The required storage is far less than for the whole medium. We checked the field p_n so obtained against the one calculated in the forward propagation. They are the same to a relative precision of 10^{-6}. Therefore, the correlation can be performed for each point of the medium and for each decreasing time.

APPENDIX B

THE INVERSE PROBLEM

Formulation of the problem

The data set is composed of the seismograms $p(\mathbf{x}_r, t; \mathbf{x}_s)$, where $\mathbf{x}_r$ and $\mathbf{x}_s$ denote receiver and source (discrete) position, and where the time variable t, (reset to zero for each new shot) runs from 0 to T. Actually, because seismic data are digital, t is also discrete, but to simplify the notation, we do not introduce the sampling explicitly.

We assume the source is exactly known. An acoustic Earth model may then be described using density $\rho(\mathbf{x})$ and bulk modulus $K(x)$. We assume that $\rho(\mathbf{x})$ is known, so that the only remaining unknown is $K(\mathbf{x})$.

For compactness, an Earth model is denoted $\mathbf{K}$ and a data set is denoted $\mathbf{p}$. The computation of the seismograms corresponding to the model $\mathbf{K}$ is written

$$\mathbf{p} = \mathbf{g}(\mathbf{K}), \tag{B-1}$$

where the operator $\mathbf{g}$ is nonlinear. Let $\mathbf{p}_{\text{obs}}$ represent the observed data set and $\mathbf{C}_p$ the covariance operator describing experimental uncertainties. The kernel of the covariance operator is diagonal:

$$C_p(\mathbf{x}_r, t; \mathbf{x}_s; \mathbf{x}'_r, t'; \mathbf{x}'_s) = \sigma^2(\mathbf{x}_r, t; \mathbf{x}_s)\delta^{rr'}\delta(t - t')\delta^{ss'}, \tag{B-2}$$

so that the expression

$$\delta\mathbf{p} = \mathbf{C}_p\,\delta\hat{\mathbf{p}} \tag{B-3}$$

is written, explicitly,

$$\begin{aligned}\delta p(\mathbf{x}_r, t; \mathbf{x}_s) &= \sum_{s'} \int_0^T dt' \sum_{r'} C_p(\mathbf{x}_r, t; \mathbf{x}_s; \mathbf{x}'_r, t'; \mathbf{x}'_s) \\ &\quad \times \delta\hat{p}(\mathbf{x}'_r, t'; \mathbf{x}'_s) \\ &= \sigma^2(\mathbf{x}_r, t; \mathbf{x}_s)\delta\hat{p}(\mathbf{x}_r, t; \mathbf{x}_s).\end{aligned} \tag{B-4}$$

The reciprocal relation

$$\delta\hat{\mathbf{p}} = \mathbf{C}_p^{-1}\delta\mathbf{p} \tag{B-5}$$

simply gives

$$\delta\hat{p}(\mathbf{x}_r, t; \mathbf{x}_s) = \frac{\delta p(\mathbf{x}_r, t; \mathbf{x}_s)}{\sigma^2(\mathbf{x}_r, t; \mathbf{x}_s)}. \tag{B-6}$$

The best model (in the least-squares sense) is defined by the minimization of the squared norm

$$S(\mathbf{K}) = \tfrac{1}{2}\| \mathbf{g}(\mathbf{K}) - \mathbf{d}_{\text{obs}} \|^2 \tag{B-7}$$

where, with the notation

$$\begin{aligned}\langle \delta\hat{\mathbf{p}}_1, \delta\mathbf{p}_2 \rangle &= \delta\hat{\mathbf{p}}'_1\delta\mathbf{p}_2 = \delta\mathbf{p}'_2\,\delta\hat{\mathbf{p}}_1 \\ &= \sum_s \int_0^T dt \sum_r \delta\hat{p}(\mathbf{x}_r, t; \mathbf{x}_s)_1 \delta p(\mathbf{x}_r, t; \mathbf{x}_s)_2,\end{aligned} \tag{B-8}$$

the norm $\| \delta\mathbf{p} \|$ is defined by

$$\| \delta\mathbf{p} \|^2 = \langle \mathbf{C}_p^{-1}\delta\mathbf{p}, \delta\mathbf{p} \rangle = \delta\mathbf{p}'\mathbf{C}_p^{-1}\delta\mathbf{p}. \tag{B-9}$$

Then

$$S(\mathbf{K}) = \tfrac{1}{2}[\mathbf{g}(\mathbf{K}) - \mathbf{d}_{\text{obs}}]'\mathbf{C}_p^{-1}[\mathbf{g}(\mathbf{K}) - \mathbf{d}_{\text{obs}}]. \tag{B-10}$$

The Fréchet derivative of the nonlinear operator $\mathbf{g}$ at a point $\mathbf{K}_n$ of the model space is the linear operator $\mathbf{G}_n$ that associates the data perturbation $\mathbf{G}_n\delta\mathbf{K}$ defined by the first-order development to any model perturbation $\delta\mathbf{K}$. $\mathbf{G}_n\delta\mathbf{K}$ is written

$$\mathbf{g}(\mathbf{K}_n + \delta\mathbf{K}) = \mathbf{g}(\mathbf{K}_n) + \mathbf{G}_n\delta\mathbf{K} + \text{higher-order terms}. \tag{B-11}$$

Introducing in the model space notation corresponding to that in equation (B-8),

$$\begin{aligned}\langle \delta\hat{\mathbf{K}}_1, \delta\mathbf{K}_2 \rangle &= \delta\hat{\mathbf{K}}'_1\delta\mathbf{K}_2 \\ &= \delta\mathbf{K}'_2\,\delta\hat{\mathbf{K}}_1 \\ &= \int_V dV(\mathbf{x})\,\delta\hat{K}(\mathbf{x})_1\delta K(\mathbf{x})_2,\end{aligned} \tag{B-12}$$

and given a linear operator $\mathbf{G}_n$, the transpose operator $\mathbf{G}'_n$ can be defined by the identity (Taylor and Lay, 1980)

$$\langle \delta\hat{\mathbf{p}}, \mathbf{G}_n\delta\mathbf{K} \rangle = \langle \mathbf{G}'_n\delta\hat{\mathbf{p}}, \delta\mathbf{K} \rangle \tag{B-13}$$

for any $\delta\hat{\mathbf{p}}$ and $\delta\mathbf{K}$.

The least-squares minimization problem can then be solved using a gradient algorithm. This easily gives (Tarantola, 1984b)

$$\mathbf{K}_{n+1} = \mathbf{K}_n - \mu_n \hat{\underset{\sim}{\mathbf{S}}}_0 \underset{\sim}{\mathbf{G}}_n^t \underset{\sim}{\mathbf{C}}_d^{-1}(\mathbf{g}(\mathbf{K}_n) - \mathbf{d}_{obs}), \quad \text{(B-14)}$$

where $\hat{\underset{\sim}{\mathbf{S}}}_0$ is an arbitrary, positive definite operator, the "preconditioning operator" which is chosen for accelerating the convergence. In our case, we choose an operator $\hat{\underset{\sim}{\mathbf{S}}}_0$ whose application corresponds to an amplification increasing with depth. Equation (B-14) corresponds to a steepest descent algorithm. Use of a conjugate gradient did not improve our results; rather, the preconditioning was of primary importance.

Fréchet derivative

As explained in the text, we use different boundary conditions in our simulations of wave propagation. The formulas derived here are only valid for free boundaries; the demonstrations for rigid boundaries are similar to those in the text, and the absorbing boundaries can be accounted for by conceptually replacing them with distant boundaries of any sort.

For free boundaries, we solve the forward problem by discretizing the system (Appendix A)

$$\frac{\partial p}{\partial t}(\mathbf{x}, t; \mathbf{x}_s) = K(\mathbf{x}) \operatorname{div} \boldsymbol{\omega}(\mathbf{x}, t; \mathbf{x}_s) + \tilde{S}(\mathbf{x}, t; \mathbf{x}_s), \quad \text{(B-15a)}$$

$$\frac{\partial \boldsymbol{\omega}}{\partial t}(\mathbf{x}, t; \mathbf{x}_s) = \frac{1}{\rho(\mathbf{x})} \mathbf{grad}\, p(\mathbf{x}, t; \mathbf{x}_s), \quad \text{(B-15b)}$$

$$p(\mathbf{x}, t; \mathbf{x}_s) = 0 \quad \text{(for } \mathbf{x} \text{ belonging to the boundary } \mathbf{S}), \quad \text{(B-15c)}$$

$$p(\mathbf{x}, 0; \mathbf{x}_s) = 0, \quad \text{(B-15d)}$$

and

$$\dot{p}(\mathbf{x}, 0; \mathbf{x}_s) = 0, \quad \text{(B-15e)}$$

where $\tilde{S}(\mathbf{x}, t; \mathbf{x}_s)$ is the source function describing the sth shot. Defining

$$\tilde{S}(\mathbf{x}, t; \mathbf{x}_s) = K(\mathbf{x}) \int_0^T dt'\, S(\mathbf{x}, t', \mathbf{x}_s) \quad \text{(B-16)}$$

gives

$$\frac{1}{K(\mathbf{x})} \frac{\partial^2 p}{\partial t^2}(\mathbf{x}, t; \mathbf{x}_s) - \operatorname{div}\left[\frac{1}{\rho(\mathbf{x})} \mathbf{grad}\, p(\mathbf{x}, t; \mathbf{x}_s)\right] = S(\mathbf{x}, t; \mathbf{x}_s), \quad \text{(B-17a)}$$

$$p(\mathbf{x}, t; \mathbf{x}_s) = 0 \quad (\text{for } \mathbf{x} \in \mathbf{S}), \quad \text{(B-17b)}$$

$$p(\mathbf{x}, 0; \mathbf{x}_s) = 0, \quad \text{(B-17c)}$$

and

$$\dot{p}(\mathbf{x}, 0; \mathbf{x}_s) = 0. \quad \text{(B-17d)}$$

The Green's function $\Gamma(\mathbf{x}, t; \mathbf{x}_s, t')$ is defined by

$$\frac{1}{K(\mathbf{x})} \frac{\partial^2 \Gamma}{\partial t^2}(\mathbf{x}, t; \mathbf{x}', t') - \operatorname{div}\left[\frac{1}{\rho(\mathbf{x})} \mathbf{grad}\, \Gamma(\mathbf{x}, t; \mathbf{x}', t')\right] = \delta(\mathbf{x} - \mathbf{x}_s)\delta(t - t'), \quad \text{(B-18a)}$$

where

$$\Gamma(\mathbf{x}, t; \mathbf{x}', t') = 0 \quad (\text{for } \mathbf{x} \in \mathbf{S}), \quad \text{(B-18b)}$$

$$\Gamma(\mathbf{x}, t; \mathbf{x}', t') = 0 \quad (\text{for } t < t'), \quad \text{(B-18c)}$$

and

$$\dot{\Gamma}(\mathbf{x}, t; \mathbf{x}', t') = 0 \quad (\text{for } t < t'), \quad \text{(B-18d)}$$

and we have the integral representation (Morse and Feshbach, 1953)

$$p(\mathbf{x}, t; \mathbf{x}_s) = \int_V dV(\mathbf{x}')\, \Gamma(\mathbf{x}, t; \mathbf{x}', 0) * S(\mathbf{x}', t; \mathbf{x}_s). \quad \text{(B-19)}$$

To obtain the Fréchet derivative of the displacements with respect to the bulk modulus [as defined by equation (B-11)], introduce the wave field $p(\mathbf{x}, t; \mathbf{x}_s)_n$ propagating in the medium $K(\mathbf{x})_n$. Then

$$\frac{1}{K(\mathbf{x})_n} \frac{\partial^2 p(\mathbf{x}, t; \mathbf{x}_s)_n}{\partial t^2} - \operatorname{div}\left[\frac{1}{\rho(\mathbf{x})} \mathbf{grad}\, p(\mathbf{x}, t; \mathbf{x}_s)_n\right] = S(\mathbf{x}, t; \mathbf{x}_s), \quad \text{(B-20a)}$$

$$p(\mathbf{x}, t; \mathbf{x}_s)_n = 0 \quad (\text{for } \mathbf{x} \in \mathbf{S}), \quad \text{(B-20b)}$$

$$p(\mathbf{x}, 0; \mathbf{x}_s)_n = 0, \quad \text{(B-20c)}$$

and

$$\dot{p}(\mathbf{x}, 0; \mathbf{x}_s)_n = 0. \quad \text{(B-20d)}$$

The corresponding Green's function obeys

$$\frac{1}{K(\mathbf{x})_n} \frac{\partial^2 \Gamma(\mathbf{x}, t; \mathbf{x}', t')_n}{\partial t^2} - \operatorname{div}\left[\frac{1}{\rho(\mathbf{x})} \mathbf{grad}\, \Gamma(\mathbf{x}, t; \mathbf{x}', t')_n\right] = \delta(\mathbf{x} - \mathbf{x}_s)\delta(t - t'), \quad \text{(B-21a)}$$

where

$$\Gamma(\mathbf{x}, t; \mathbf{x}', t')_n = 0 \quad (\text{for } \mathbf{x} \in \mathbf{S}), \quad \text{(B-21b)}$$

$$\Gamma(\mathbf{x}, t; \mathbf{x}', t')_n = 0 \quad (\text{for } t < t'), \quad \text{(B-21c)}$$

and

$$\dot{\Gamma}(\mathbf{x}, t; \mathbf{x}', t')_n = 0 \quad (\text{for } t < t'). \quad \text{(B-21d)}$$

A perturbation of bulk modulus $K(\mathbf{x})_n \to K(\mathbf{x})_n + \delta K(\mathbf{x})$ will produce a field $p(\mathbf{x}, t; \mathbf{x}_s)_n + \delta p(\mathbf{x}, t; \mathbf{x}_s)$ defined by

$$\frac{1}{K(\mathbf{x})_n + \delta K(\mathbf{x})} \frac{\partial^2 [p(\mathbf{x}, t; \mathbf{x}_s)_n + \delta p(\mathbf{x}, t; \mathbf{x}_s)]}{\partial t^2} - \operatorname{div}\left\{\frac{1}{\rho(\mathbf{x})} \mathbf{grad}\, [p(\mathbf{x}, t; \mathbf{x}_s)_n + \delta p(\mathbf{x}, t; \mathbf{x}_s)]\right\} = S(\mathbf{x}, t; \mathbf{x}_s), \quad \text{(B-22a)}$$

where

$$p(\mathbf{x}, t; \mathbf{x}_s)_n + \delta p(\mathbf{x}, t; \mathbf{x}_s) = 0 \quad (\text{for } \mathbf{x} \in \mathbf{S}), \quad \text{(B-22b)}$$

$$p(\mathbf{x}, 0; \mathbf{x}_s)_n + \delta p(\mathbf{x}, 0; \mathbf{x}_s) = 0, \quad \text{(B-22c)}$$

and

$$\dot{p}(\mathbf{x}, 0; \mathbf{x}_s)_n + \delta \dot{p}(\mathbf{x}, 0; \mathbf{x}_s) = 0. \quad \text{(B-22d)}$$

This gives

$$\frac{1}{K(x)_n} \frac{\partial^2 \delta p(\mathbf{x}, t; \mathbf{x}_s)}{\partial t^2} - \operatorname{div}\left[\frac{1}{\rho(\mathbf{x})} \mathbf{grad}\, \delta p(\mathbf{x}, t; \mathbf{x}_s)\right] = \frac{\partial^2 p(\mathbf{x}, t; \mathbf{x}_s)_n}{\partial t^2} \frac{\delta K(\mathbf{x})}{K(\mathbf{x})_n^2} + O(\delta \mathbf{K}^2), \quad \text{(B-23a)}$$

where

$$\delta p(\mathbf{x}, t; \mathbf{x}_s) = 0 \quad (\text{for } \mathbf{x} \in \mathbf{S}), \quad \text{(B-23b)}$$

$$\delta p(\mathbf{x}, 0; \mathbf{x}_s) = 0, \quad \text{(B-23c)}$$

and

$$\delta\dot{p}(\mathbf{x}, 0; \mathbf{x}_s) = 0. \tag{B-23d}$$

Using theorem (B-19),

$$\delta p(\mathbf{x}_r, t; \mathbf{x}_s) = \int_V dV(\mathbf{x})\, \Gamma(\mathbf{x}_r, t; \mathbf{x}, 0)_n * \ddot{p}(\mathbf{x}, t; \mathbf{x}_s)_n \frac{\delta K(\mathbf{x})}{K(\mathbf{x})_n^2} + O(\delta\mathbf{K}^2). \tag{B-24}$$

The Fréchet derivative operator $\mathbf{G}_n$ introduced in equation (B-11) is then defined by

$$(\mathbf{G}_n \delta\mathbf{K})(\mathbf{x}_r, t; \mathbf{x}_s) = \int_V dV(\mathbf{x})\, \Gamma(\mathbf{x}_r, t; \mathbf{x}, 0)_n * \ddot{p}(\mathbf{x}, t; \mathbf{x}_s)_n \frac{\delta K(\mathbf{x})}{K(\mathbf{x})_n^2}, \tag{B-25}$$

where $\Gamma(\mathbf{x}, t; \mathbf{x}', t')_n$ is the Green's function defined by equation (21), corresponding to the medium $K(\mathbf{x})_n$ and $p(\mathbf{x}, t; \mathbf{x}_s)_n$ is the wave field defined by equation (B-20) corresponding to $K(\mathbf{x})_n$.

Transpose operator

We have defined the Fréchet derivative operator $\mathbf{G}_n$. Its transpose $\mathbf{G}_n^t$ was defined by equation (B-13). To solve the inverse problem, we need to be able to compute $\mathbf{G}_n^t \delta\hat{\mathbf{p}}$ for arbitrary $\delta\hat{\mathbf{p}}$. Using the notation introduced in equations (B-8) and (B-12), equation (B-13) is written as

$$\sum_s \int_0^T dt \sum_r \delta\hat{p}(\mathbf{x}_r, t; \mathbf{x}_s)(\mathbf{G}_n \delta\mathbf{K})(\mathbf{x}_r, t; \mathbf{x}_s) = \int_V dV(\mathbf{x})\, (\mathbf{G}_n^t \delta\hat{\mathbf{p}})(\mathbf{x}) \delta K(\mathbf{x}). \tag{B-26}$$

Using equation (B-25), this becomes

$$\sum_s \int_0^T dt \sum_r \delta\hat{p}(\mathbf{x}_r, t; \mathbf{x}_s) \int_V dv(\mathbf{x})\, \Gamma(\mathbf{x}_r, t; \mathbf{x}, 0)_n * \ddot{p}(\mathbf{x}, t; \mathbf{x}_s)_n \frac{\delta K(\mathbf{x})}{K(\mathbf{x})_n^2} = \int_V dV(\mathbf{x})\, (\mathbf{G}_n^t \delta\hat{\mathbf{p}})(\mathbf{x}) \delta K(\mathbf{x}), \tag{B-27}$$

i.e.,

$$\int_V dV(\mathbf{x})\, \delta K(\mathbf{x}) \left[(\mathbf{G}_n^t \delta\hat{\mathbf{p}})(\mathbf{x}) - \frac{1}{K(\mathbf{x})_n^2} \sum_s \int_0^T dt \sum_r \Gamma(\mathbf{x}_r, t; \mathbf{x}, 0)_n * \ddot{p}(\mathbf{x}, t; \mathbf{x}_s)_n \delta\hat{p}(\mathbf{x}_r, t; \mathbf{x}_s) \right] = 0. \tag{B-28}$$

Because this is valid for any $\delta K(\mathbf{x})$, we obtain

$$(\mathbf{G}_n^t \delta\hat{\mathbf{p}})(\mathbf{x}) = \frac{1}{K(\mathbf{x})_n^2} \sum_s \int_0^T dt \sum_r \Gamma(\mathbf{x}_r, t; \mathbf{x}, 0)_n * \ddot{p}(\mathbf{x}, t; \mathbf{x}_s)_n \delta\hat{p}(\mathbf{x}_r, t; \mathbf{x}_s). \tag{B-29}$$

We introduce a field $\Psi(\mathbf{x}, t; \mathbf{x}_s)_n$ defined by the differential system

$$\frac{1}{K(\mathbf{x})_n} \frac{\partial^2 \Psi(\mathbf{x}, t; \mathbf{x}_s)_n}{\partial t^2} - \operatorname{div}\left[\frac{1}{\rho(\mathbf{x})} \mathbf{grad}\, \Psi(\mathbf{x}, t; \mathbf{x}_s)_n \right] = \delta\hat{p}(\mathbf{x}, t; \mathbf{x}_s), \tag{B-30a}$$

where

$$\Psi(\mathbf{x}, t; \mathbf{x}_s)_n = 0 \qquad (\text{for } \mathbf{x} \in \mathbf{S}), \tag{B-30b}$$

$$\Psi(\mathbf{x}, T; \mathbf{x}_s)_n = 0, \tag{B-30c}$$

and

$$\dot{\Psi}(\mathbf{x}, T; \mathbf{x}_s)_n = 0. \tag{B-30d}$$

$\Psi(\mathbf{x}, t; \mathbf{x}_s)_n$ satisfies *final* (instead of initial) conditions. Using the property

$$\Gamma(\mathbf{x}, t; \mathbf{x}_r, t')_n = \Gamma(\mathbf{x}, t + \tau; \mathbf{x}_r, t' + \tau)_n, \tag{B-31}$$

and reversing time in theorem (B-19),

$$\Psi(\mathbf{x}, t; \mathbf{x}_s)_n = \sum_r \int_0^T dt'\, \Gamma(\mathbf{x}, 0; \mathbf{x}_r, t - t')_n \delta\hat{p}(\mathbf{x}_r, t'; \mathbf{x}_s). \tag{B-32}$$

We then have

$$\begin{aligned} \Psi(\mathbf{x}, t; \mathbf{x}_s) &= \sum_r \int_0^T dt' \frac{\partial}{\partial t} \Gamma(\mathbf{x}, 0; \mathbf{x}_r, t - t') \delta\hat{p}(\mathbf{x}_r, t', \mathbf{x}_s), \\ &= \sum_r \int_0^T dt' \frac{\partial}{\partial t} \Gamma(\mathbf{x}, t' - t; \mathbf{x}_r, 0) \delta\hat{p}(\mathbf{x}_r, t', \mathbf{x}_s), \\ &= -\sum_r \int_0^T dt'\, \dot{\Gamma}(\mathbf{x}, t' - t; \mathbf{x}_r, 0) \delta\hat{p}(\mathbf{x}_r, t'; \mathbf{x}_s), \end{aligned} \tag{B-33}$$

where

$$\dot{\Gamma}(\mathbf{x}, t; \mathbf{x}', t') = \frac{\partial}{\partial t} \Gamma(\mathbf{x}, t; \mathbf{x}', t'). \tag{B-34}$$

Using integration by parts, we have

$$\begin{aligned} &\Gamma(\mathbf{x}_r, t; \mathbf{x}, 0) * \ddot{p}(\mathbf{x}, t; \mathbf{x}_s) \\ &\quad = \int_0^T dt'\, \Gamma(\mathbf{x}_r, t - t'; \mathbf{x}, 0) \dot{p}(\mathbf{x}, t'; \mathbf{x}_s), \\ &\quad = \Gamma(\mathbf{x}_r, t - T; \mathbf{x}, 0) \dot{p}(\mathbf{x}, T; \mathbf{x}_s) - \Gamma(\mathbf{x}_r, t; \mathbf{x}, 0) \dot{p}(\mathbf{x}, 0; \mathbf{x}_s) \\ &\qquad - \int_0^T dt'\, \dot{\Gamma}(\mathbf{x}_r, t - t'; \mathbf{x}, 0) \dot{p}(\mathbf{x}, t'; \mathbf{x}_s), \end{aligned}$$

and using the initial conditions (B-20d) and (B-21c),

$$\Gamma(\mathbf{x}_r, t; \mathbf{x}, 0) * \ddot{p}(\mathbf{x}, t; \mathbf{x}_s) = -\dot{\Gamma}(\mathbf{x}_r, t; \mathbf{x}, 0) * \dot{p}(\mathbf{x}, t; \mathbf{x}_s). \tag{B-35}$$

Using equation (B-35) gives

$$\sum_s \int_0^T dt \sum_r \Gamma(\mathbf{x}_r, t; \mathbf{x}, 0)_n * \ddot{p}(\mathbf{x}, t; \mathbf{x}_s) \delta\hat{p}(\mathbf{x}_r, t; \mathbf{x}_s) = -\sum_s \int_0^T dt \sum_r \dot{\Gamma}(\mathbf{x}_r, t; \mathbf{x}, 0) * \dot{p}(\mathbf{x}, t; \mathbf{x}_s) \delta\hat{p}(\mathbf{x}_r, t; \mathbf{x}_s),$$

$$= -\sum_s \int_0^T dt \int_0^T dt' \sum_r \hat{\Gamma}(\mathbf{x}_r, t - t'; \mathbf{x}, 0) \times \dot{p}(\mathbf{x}, t'; \mathbf{x}_s)\delta\hat{p}(\mathbf{x}_r, t; \mathbf{x}_s),$$

from which, using equation (B-33), we obtain

$$\sum_s \int_0^T dt \sum_r \Gamma(\mathbf{x}_r, t; \mathbf{x}, 0)_n * \dot{p}(\mathbf{x}, t; \mathbf{x}_s)\delta\hat{p}(\mathbf{x}_r, t; \mathbf{x}_s) = \sum_s \int_0^T dt\ \dot{p}(\mathbf{x}, t; \mathbf{x}_s)\Psi(\mathbf{x}, t; \mathbf{x}_s). \quad \text{(B-36)}$$

From equation (B-29) we then obtain the result

$$(\mathbf{G}_n^t \delta\hat{\mathbf{p}})(\mathbf{x}) = \frac{1}{K(\mathbf{x})_n^2} \sum_s \int_0^T dt\ \dot{p}(\mathbf{x}, t; \mathbf{x}_s)_n \Psi(\mathbf{x}, t; \mathbf{x}_s)_n. \quad \text{(B-37)}$$

APPENDIX C
COMPUTATION OF THE STEP LENGTH

The gradient algorithm gives [equation (B-14)]

$$\mathbf{K}_{n+1} = \mathbf{K}_n - \mu_n \hat{\mathbf{S}}_0 \mathbf{G}_n^t \mathbf{C}_d^{-1}[\mathbf{g}(\mathbf{K}_n) - \mathbf{p}_{\text{obs}}]. \quad \text{(C-1)}$$

Defining

$$\hat{\gamma}_n = \mathbf{G}_n^t \mathbf{C}_d^{-1}[\mathbf{g}(\mathbf{K}_n) - \mathbf{p}_{\text{obs}}) \quad \text{(C-2)}$$

and

$$\mathbf{d}_n = \hat{\mathbf{S}}_n \hat{\gamma}_n \quad \text{(C-3)}$$

gives

$$\mathbf{K}_{n+1} = \mathbf{K}_n - \mu_n \mathbf{d}_n. \quad \text{(C-4)}$$

This brings the problem of obtaining an adequate estimation of the scalar μ_n. For given $\mathbf{K}_n$,

$$S(\mathbf{K}_n - \mu_n \mathbf{d}_n) = \tfrac{1}{2}\{[\mathbf{g}(\mathbf{K}_n - \mu_n \mathbf{d}_n) - \mathbf{p}_{\text{obs}}]^t \times \mathbf{C}_p^{-1}[\mathbf{g}(\mathbf{K}_n - \mu_n \mathbf{d}_n) - \mathbf{p}_{\text{obs}}]\}. \quad \text{(C-5)}$$

If μ_n is small enough, using the definition of Fréchet derivative [equation (B-11)] we have

$$\mathbf{g}(\mathbf{K}_n - \mu_n \mathbf{d}_n) \simeq \mathbf{g}(\mathbf{K}_n) - \mu_n \mathbf{G}_n \mathbf{d}_n, \quad \text{(C-6)}$$

which gives

$$S(\mathbf{K}_n - \mu_n \mathbf{d}_n) \simeq S(\mathbf{K}_n) - \mu_n(\mathbf{G}_n \mathbf{d}_n)^t \mathbf{C}_p^{-1}[\mathbf{g}(\mathbf{K}_n) - \mathbf{p}_{\text{obs}}] + \tfrac{1}{2}\mu_n^2(\mathbf{G}_n \mathbf{d}_n)^t \mathbf{C}_p^{-1}(\mathbf{G}_n \mathbf{d}_n),$$

and the condition $\partial S/\partial \mu_n = 0$ gives

$$\mu_n \simeq \frac{(\mathbf{G}_n \mathbf{d}_n)^t \mathbf{C}_p^{-1}[\mathbf{g}(\mathbf{K}_n) - \mathbf{p}_{\text{obs}}]}{(\mathbf{G}_n \mathbf{d}_n)^t \mathbf{C}_p^{-1}(\mathbf{G}_n \mathbf{d}_n)}. \quad \text{(C-8)}$$

Using the definition of the transpose operator [equation (B-13)] we finally obtain

$$\mu_n \simeq \frac{\mathbf{d}_n^t \mathbf{G}_n^t \mathbf{C}_p^{-1}[\mathbf{g}(\mathbf{K}_n) - \mathbf{p}_{\text{obs}}]}{(\mathbf{G}_n \mathbf{d}_n)^t \mathbf{C}_p^{-1}(\mathbf{G}_n \mathbf{d}_n)} = \frac{\mathbf{d}_n^t \hat{\gamma}_n}{(\mathbf{G}_n \mathbf{d}_n)^t \mathbf{C}_p^{-1}(\mathbf{G}_n \mathbf{d}_n)}. \quad \text{(C-9)}$$

To compute $\mathbf{G}_n \mathbf{d}_n$, we could use the result equation (B-25), but it is more practical to invoke the definition of derivative operator and use a finite-difference approximation. We thus use

$$\mathbf{G}_n \mathbf{d}_n \simeq \frac{1}{\varepsilon}[\mathbf{g}(\mathbf{K}_n + \varepsilon \mathbf{d}_n) - \mathbf{g}(\mathbf{K}_n)], \quad \text{(C-10)}$$

with a sufficiently small value of ε.