GEOPHYSICS, VOL. 52, NO. 9 (SEPTEMBER 1987); P. 1211-1228, 10 FIGS.

Nonlinear two-dimensional elastic inversion of multioffset seismic data

Peter Mora*

ABSTRACT

The treatment of multioffset seismic data as an acoustic wave field is becoming increasingly disturbing to many geophysicists who see a multitude of wave phenomena, such as amplitude-offset variations and shear-wave events, which can only be explained by using the more correct elastic wave equation. Not only are such phenomena ignored by acoustic theory, but they are also treated as undesirable noise when they should be used to provide extra information, such as *S*-wave velocity, about the subsurface.

The problems of using the conventional acoustic wave equation approach can be eliminated via an elastic approach. In this paper, equations have been derived to perform an inversion for *P*-wave velocity, *S*-wave velocity, and density as well as the *P*-wave impedance, *S*-wave impedance, and density. These are better resolved than the Lamé parameters. The inversion is based on nonlinear least squares and proceeds by iteratively updating the earth parameters until a good fit is achieved between the observed data and the modeled data corresponding to these earth parameters. The iterations are based on the preconditioned conjugate gradient algorithm. The fundamental requirement of such a least-squares algorithm is the gradient direction which tells how to update the model parameters. The gradient direction can be derived directly from the wave equation and it may be computed by several wave propagations. Although in principle any scheme could be chosen to perform the wave propagations, the elastic finite-difference method is used because it directly simulates the elastic wave equation and can handle complex, and thus realistic, distributions of elastic parameters. This method of inversion is costly since it is similar to an iterative prestack shot-profile migration. However, it has greater power than any migration since it solves for the *P*-wave velocity, *S*-wave velocity, and density and can handle very general situations including transmission problems.

Three main weaknesses of this technique are that it requires fairly accurate a priori knowledge of the low-wavenumber velocity model, it assumes Gaussian model statistics, and it is very computer-intensive. All these problems seem surmountable. The low-wavenumber information can be obtained either by a prior tomographic step, by the conventional normal-moveout method, by a priori knowledge and empirical relationships, or by adding an additional inversion step for low wavenumbers to each iteration. The Gaussian statistics can be altered by preconditioning the gradient direction, perhaps to make the solution blocky in appearance like well logs, or by using large model variances in the inversion to reduce the effect of the Gaussian model constraints. Moreover, with some improvements to the algorithm and more parallel computers, it is hoped the technique will soon become routinely feasible.

INTRODUCTION

Elastic waves or acoustic waves?

Seismic data contain many features such as shear reflections and amplitude-offset variation that provide useful information about the *P*-wave velocity, *S*-wave velocity, and density. The importance of these features is becoming increasingly evident as the use of longer cables and multicomponent recording becomes more widespread. Conventional seismic processing and inversion based on the acoustic wave equation do not take these features into account, and they therefore treat this useful information as undesirable coherent noise. To use all the information contained in seismic data effectively, one must

Manuscript received by the Editor April 14, 1986; revised manuscript received February 2, 1987.
*Geophysics Department, Stanford University, Stanford, CA 94305.

use the elastic wave equation and build up a new framework (or expand the old one) for the treatment of seismic data. The various seismic events usually considered to be noise, such as ground roll, refractions, multiples, mode conversions, and *S*-wave events, can then be treated as signal and should actually be helpful (I include multiples in this list for completeness and because the elastic wave equation handles their amplitudes more correctly than the acoustic wave equation). In fact, almost all seismic "noise" ceases to be bothersome.

Inversion or conventional processing?

Should inversion or standard processing methods be applied to seismic data? Most practicing geophysicists would say that while the motivation of inversion, to get a quantitative estimate of the physical properties, is appealing, they would prefer to use standard methods because they do not believe inversion can work well in practice. Actually, the devout inverters would reply that standard methods are inversion too because they try to obtain a picture of the subsurface. The main differences are that standard methods are one-step processes, and the result is not a quantitative estimate of the physical properties. For instance, migration gives a relative reflectivity picture and not a quantitative *P*-wave velocity or *P*-wave impedance picture. In general, standard methods do not pay heed to the exact properties of the earth but concentrate on obtaining a good image of the subsurface. In practice, the product of conventional processing applied to reflection seismic data would usually be a quantitative estimate of the low-wavenumber part of the *P*-wave velocities (from velocity analysis) and a relative estimate of the high-wavenumber part of the reflectivities (from migration). I show here that for the case of seismic reflection data, the picture that can be obtained by inversion and that obtained by the conventional approach (migration) are very similar. Transmission data such as those obtained in vertical seismic profiles (VSP) and well-to-well surveys are another matter. The inversion technique can utilize the information contained in the direct waves to improve the results while conventional migration cannot (in particular the direct waves help obtain the low-wavenumber "blocky" components in the *P*- and *S*-wave velocity models). The strength of the inversion philosophy is that it tries to account for the seismic data in terms of the earth properties using the known equations of physics. This strength is also a weakness because it often leads to impractical algorithms.

In this paper, the method of least squares is invoked on a grand scale so that by using the elastic wave equation, it is possible to make use of all the amplitude information in the seismic data and perform an inversion for the *P*-wave velocity, *S*-wave velocity, and density. The result of the inversion is the most likely set of physical parameters that could have given rise to the observed data (provided any energy not accounted for by the elastic wave equation is uncorrelated Gaussian noise). I believe that with the rapid development of more parallel computers this method will soon be routinely feasible.

Inversion philosophies

Inversion is more complicated than the forward problem which attempts to simulate the equations of physics. This is because inversion attempts to solve a problem which is inherently unstable. For instance, in seismic exploration one attempts to predict rock properties based on the appearance of waves which in the past have propagated through some distant rocks. If the effect of the rocks on the waves is not great, then in the presence of any noise one cannot hope to predict the properties of the rocks. In other words, it is sensible to try to predict only those properties which can be resolved given the quality of the data. If a parameter is not well resolved, then the solution should be restricted in a logical way using a priori knowledge.

To make matters more difficult, the seismic inverse problem is nonlinear (i.e., no linear relation exists between the seismogram amplitudes and Earth properties). Even so, most literature on seismic inversion concerns linearized methods (Cohen and Bleistein, 1979; Stolt and Weglein, 1985; Lailly, 1984; Clayton and Stolt, 1981; Ikelle et al., 1986) because through functional analysis, the linearized problem is well understood. The resolution problem has been studied and largely solved by Backus and Gilbert (1967; 1968; 1970) in the linear case.

I first consider the ideal inversion, namely, mapping out the probability functions of the desired physical properties based on all available information (Tarantola and Valette, 1982). However, the ideal inversion is completely impractical for most inverse problems because it would take an almost infinite amount of time on any computer. Sometimes the Monte Carlo methods may be utilized to map the probability functions partially or to solve for the single most probable set of model parameters (Rothman, 1984). Unfortunately, such methods require the forward problem to be solved many times, so for most seismic inverse problems where the forward problem consists of a computationally intensive wave propagation, Monte Carlo methods are too costly. Hence, we must compromise between practicality and elegance of the inversion. This paper constitutes my compromise at the present time. The inversion is based on the relatively fast and tractable least-squares optimization method. It does *not* use a linearized wave equation for the solution of the forward problem, but rather uses the full elastic wave equation. The iterative least-squares algorithm requires a gradient direction which can be derived directly from the wave equation. See also Tarantola (1984) for the elastic case and Lailly (1984) or Kolb (1986) for the acoustic case; Mora (1987a) and McAulay (1985) explicitly obtained Frechét matrices in the elastic and acoustic cases, respectively (this was feasible because they assumed plane layers), enabling them to use the more rapidly converging Newton algorithm.

INVERSE THEORY

In general, the process of inversion can be considered as the location of the single most probable set of model parameters $\mathbf{m}$ given some data observations $\mathbf{d}$ and knowledge of the probability distributions of $\mathbf{m}$ and $\mathbf{d}$ (Tarantola and Valette, 1982; Menke, 1984). For inversion of multioffset seismic data for the elastic properties of rocks $\mathbf{m}$, computation of $\mathbf{d}(\mathbf{m})$ (the common shot gathers) represents elastic wave propagation which consumes large amounts of computer time. Therefore, since inversion schemes based on linear function space theory require relatively few forward problems to be solved, they are preferred over the more general Monte Carlo schemes which would require many forward simulations. The reason is that

linear function space methods assume local linearizability of $\mathbf{d}(\mathbf{m})$ and so can progress steadily uphill on the probability function until the nearest peak is located. Unfortunately, this may not be the biggest peak (most probable solution). This is the price paid for requiring only a few forward simulations. For tractability, the least-squares optimization method, which assumes Gaussian probability distributions for model parameters and data errors, has been chosen. This should often turn out to be a reasonable assumption, at least for the data errors when there are many independent sources of noise. The central limit theorem states that the sum of independent noise tends to have a Gaussian distribution. However, the model parameters (i.e., the physical properties of the earth) in general are not Gaussianly distributed. This can be at least partially handled, if not rigorously, by allowing the model mean to vary with iteration (see Mora, 1987a) or by modifying the solution at each iteration by using some statistical arguments (Harlan, 1986).

Least squares and the preconditioned conjugate gradient method

The conjugate gradient method of nonlinear least squares has been chosen because it is simple, yet it has good convergence properties. I begin with a description of nonlinear least squares and the preconditioned gradient method and later extend this to make use of the so-called conjugate directions to speed convergence. The section on inverse theory defines the notation used throughout and elaborates on the least-squares method applied to the seismic inverse problem.

Consider the Gaussian probability density function

$$P(\mathbf{d}, \mathbf{m}) = \text{constant exp} - \tfrac{1}{2}(\Delta\mathbf{d}^*\underline{\mathbf{C}}_\mathbf{d}^{-1}\Delta\mathbf{d} + \Delta\mathbf{m}^*\underline{\mathbf{C}}_\mathbf{m}^{-1}\Delta\mathbf{m}), \tag{1a}$$

where $\Delta\mathbf{d} = \mathbf{d} - \mathbf{d}_0 = \mathbf{d}(\mathbf{m}) - \mathbf{d}_0$, $\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0$, $\mathbf{d}$ is the data vector, $\mathbf{d}_0$ are the data observations, $\mathbf{m}$ is the model vector, $\mathbf{m}_0$ is the a priori model, and $\underline{\mathbf{C}}_\mathbf{d}$ and $\underline{\mathbf{C}}_\mathbf{m}$ are the covariance matrices for data and model spaces, respectively. [Note that the asterisk indicates conjugate transpose. Normally the data consist of real values, so the asterisk indicates the vector transpose, unless the inversion is carried out in Fourier space (or some other complex space)]. Clearly, the maximum probability solution occurs when the least-squares functional

$$S(\mathbf{d}, \mathbf{m}) = \Delta\mathbf{d}^*\underline{\mathbf{C}}_\mathbf{d}^{-1}\Delta\mathbf{d} + \Delta\mathbf{m}^*\underline{\mathbf{C}}_\mathbf{m}^{-1}\Delta\mathbf{m} \tag{1b}$$

is minimized, so the least-squares solution is equal to the maximum a posteriori solution when the prior distributions are Gaussian.

In the following development, I have chosen to review the least-squares theory in terms of vectors and matrices though this also applies to linear function spaces in the ℓ_2 norm. In the continuous case, vectors would become elements of infinite-dimensional vector spaces, and matrices would become linear operators. The vector-matrix description applies to discretized spaces so that it can be more easily translated into computer code. One further point is that the final algorithm will be iterative, so all vectors and matrices in the following section refer to the nth iteration though subscript n is not given explicitly.

By taking the derivative of the square error functional $S(\mathbf{d}, \mathbf{m})$ with respect to the model vector $\mathbf{m}$ one obtains the gradient vector $\mathbf{g}$ (defined to be $-1/2$ of the steepest descent vector)

$$\mathbf{g} = \frac{1}{2}\frac{\partial S}{\partial \mathbf{m}} = \underline{\mathbf{D}}^*\underline{\mathbf{C}}_\mathbf{d}^{-1}\Delta\mathbf{d} + \underline{\mathbf{C}}_\mathbf{m}^{-1}\Delta\mathbf{m}, \tag{2}$$

where $\underline{\mathbf{D}} = \partial\mathbf{d}/\partial\mathbf{m}$ is the Frechét derivative matrix (later, $\underline{\mathbf{D}}$ for the elastic wave equation will be derived as an operator because the derivation is more straightforward). Note that the factor 1/2 is introduced to simplify later expressions. One way to solve for a minimum of the least-squares functional is by substituting the linearization for $\mathbf{d}(\mathbf{m})$,

$$\delta\mathbf{d} = \mathbf{d}(\mathbf{m}') - \mathbf{d}(\mathbf{m}) = \frac{\partial\mathbf{d}}{\partial\mathbf{m}}(\mathbf{m}' - \mathbf{m}) = \underline{\mathbf{D}}\delta\mathbf{m}, \tag{3}$$

into equation (2) and solving $\mathbf{g} = 0$. This leads to the Newton algorithm

$$\mathbf{m}' = \mathbf{m} - \underline{\mathbf{H}}\mathbf{g}, \tag{4}$$

where

$$\underline{\mathbf{H}} = \left[\underline{\mathbf{D}}^*\underline{\mathbf{C}}_\mathbf{d}^{-1}\underline{\mathbf{D}} + \underline{\mathbf{C}}_\mathbf{m}^{-1}\right]^{-1} \tag{5}$$

is the inverse Hessian matrix. This yields the maximum a posteriori solution in one iteration for linear $\mathbf{d}(\mathbf{m})$ (see also Tarantola and Valette, 1982). The inverse Hessian matrix modifies the gradient direction and chooses a magnitude of the model parameter update such that the best-fit solution is located in one iteration for linear functions $\mathbf{d}(\mathbf{m})$.

The size of the seismic inverse problem is usually too large to handle using matrices (see Mora, 1987a, and McAulay, 1985 for exceptions), but I show that the gradient direction $\mathbf{g}$ can be calculated by only two forward modeling runs in very general situations (i.e., elastic waves, exotic survey specifications, and complex geologic models). Then approximating the inverse Hessian $\underline{\mathbf{H}}$ by $\hat{\underline{\mathbf{H}}} = \eta\underline{\mathbf{C}}_\mathbf{m}$ (where η is called the step length) gives a preconditioned gradient-type iterative least-squares formula

$$\mathbf{m}' = \mathbf{m} - \hat{\underline{\mathbf{H}}}\mathbf{g} = \mathbf{m} - \eta\underline{\mathbf{C}}_\mathbf{m}\mathbf{g}. \tag{6}$$

If this algorithm were iteratively applied, the solution should converge to the same solution as that of equation (4), so the gradient iterations effectively apply the true inverse Hessian $\underline{\mathbf{H}}$. For nonlinear functions $\mathbf{d}(\mathbf{m})$, where many local minima may exist (such as in the seismic problem), the minimum located by the least-squares iterations may not be the global minimum of the error functional. Whether or not the global minimum is located depends upon the proximity of the starting point on the functional surface to the global minimum. This problem of local minima is intrinsic to nonlinear inverse problems and cannot be easily avoided. Actually, there are several techniques that attempt to handle local minima. These include (1) doing several inversions from different starting locations; (2) bounce methods which try to bounce out of local minima; (3) adjustment of the a priori information; (4) additions of new kinds of a priori constraints; and (5) careful choice of model parameters in order to minimize the degree of nonlinearity of the function $\mathbf{d}(\mathbf{m})$ in the region of interest. Of these, the most efficient and meaningful are methods (4) and

(5). I do not include Monte Carlo methods in this list because, for the situation under study where the forward simulation is currently very computer-intensive, Monte Carlo methods are impractical.

Often, by using some a priori knowledge or constraint, it is possible to modify the gradient direction $\mathbf{g}$ in some sensible way, thus greatly speeding convergence. This may also help avoid local minima and resolve the null-space problem. Another way to modify the gradient would be to derive a spatial filter corresponding to the true inverse Hessian $\underline{\mathbf{H}}$. This would (1) perform a spatial deconvolution to make the spatial spectrum comparable to the a priori spectrum; and (2) perform a model parameter deconvolution to resolve between the different physical properties (P-wave velocity, S-wave velocity, and density). However, this deconvolution would not be easy because it would be a model-dependent operator that is spatially variable. In general, a modification to the gradient is termed preconditioning and it need not be a linear operation such as in equation (6) (see Harlan, 1986). Let the new preconditioned gradient be denoted

$$\mathbf{p} = \mathbf{P}(\mathbf{g}). \tag{7}$$

In order to perform the inversion by iteratively updating the model in the direction $\mathbf{p}$, we must determine the step length η. Denote the updated model

$$\mathbf{m}' = \mathbf{m} - \eta\mathbf{p}. \tag{8}$$

Assuming linearity of $\mathbf{d}(\mathbf{m})$ in the vicinity of the current model $\mathbf{m}$, the new data after the model perturbation are

$$\mathbf{d}' = \mathbf{d}(\mathbf{m}') = \mathbf{d} - \eta\underline{\mathbf{D}}\mathbf{p}. \tag{9}$$

Now, solve for the η which minimizes the new error functional S' corresponding to the perturbed model

$$\min_{\eta} S' = \min_{\eta}\Big[(\mathbf{d}' - \mathbf{d}_0)^*\underline{\mathbf{C}}_\mathbf{d}^{-1}(\mathbf{d}' - \mathbf{d}_0) + (\mathbf{m}' - \mathbf{m}_0)^*\underline{\mathbf{C}}_\mathbf{m}^{-1}(\mathbf{m}' - \mathbf{m}_0)\Big]. \tag{10}$$

Solving for the derivative of S' with respect to η yields

$$\frac{\partial S'}{\partial \eta} = -2\Big[\mathbf{p}^*\underline{\mathbf{D}}^*\underline{\mathbf{C}}_\mathbf{d}^{-1}(\Delta\mathbf{d} - \eta\underline{\mathbf{D}}\mathbf{p}) + \mathbf{p}^*\underline{\mathbf{C}}_\mathbf{m}^{-1}(\Delta\mathbf{m} - \eta\mathbf{p})\Big]. \tag{11}$$

Finally, setting this derivative to zero yields the solution for η,

$$\eta = \frac{\mathbf{p}^*\mathbf{g}}{\mathbf{p}^*\underline{\mathbf{D}}^*\underline{\mathbf{C}}_\mathbf{d}^{-1}\underline{\mathbf{D}}\mathbf{p} + \mathbf{p}^*\underline{\mathbf{C}}_\mathbf{m}^{-1}\mathbf{p}}. \tag{12}$$

This value for η may not be very good for highly nonlinear functions $\mathbf{d}(\mathbf{m})$, so it may be necessary to do a line search using the above value as a starting point.

Preconditioned conjugate gradient algorithm

The use of conjugate directions helps to speed convergence by choosing a direction that is a linear combination of the past and current steepest descent directions (Luenberger, 1984). Actually, the use of conjugate directions is not very important if the preconditioning operator is well chosen but is nonetheless worthwhile because it can be included at no extra cost to the algorithm (it can be considered as additional preconditioning). One choice for the conjugate direction $\mathbf{c}_n$ is given by the Polak-Ribiere method (Polak and Ribiere, 1969; Luenberger, 1984) as

$$\mathbf{c}_n = \mathbf{p}_n + \frac{\mathbf{p}_n^*(\mathbf{g}_n - \mathbf{g}_{n-1})}{\mathbf{p}_{n-1}^*\mathbf{g}_{n-1}}\mathbf{c}_{n-1}.$$

This choice is preferred over the Fletcher-Reeves method (Fletcher and Reeves, 1964) because it tends to revert to the simple gradient algorithm in situations where the function is very nonlinear, and so it is more reliable (Powell, 1981).

Thus, the final algorithm for nonlinear least squares by the preconditioned conjugate gradient method is

for $n = 1$ to ∞

$$\left\{\mathbf{d}_n = \mathbf{d}(\mathbf{m}_n)\right. \qquad \text{data calculation}$$

$$\Delta\mathbf{d}_n = \mathbf{d}_n - \mathbf{d}_0,\ \Delta\mathbf{m}_n = \mathbf{m}_n - \mathbf{m}_0, \qquad \text{compute residuals}$$

$$S(\mathbf{d}, \mathbf{m}) = \Delta\mathbf{d}^*\underline{\mathbf{C}}_\mathbf{d}^{-1}\Delta\mathbf{d} + \Delta\mathbf{m}_n^*\underline{\mathbf{C}}_\mathbf{m}^{-1}\Delta\mathbf{m}_n, \qquad \text{square error functional}$$

exit if converged,

$$\mathbf{g}_n = \underline{\mathbf{D}}_n^*\underline{\mathbf{C}}_\mathbf{d}^{-1}\Delta\mathbf{d} + \underline{\mathbf{C}}_\mathbf{m}^{-1}\Delta\mathbf{m}_n, \qquad \text{gradient}$$

$$\mathbf{p}_n = \mathbf{P}(\mathbf{g}_n) \approx \underline{\mathbf{C}}_\mathbf{m}\mathbf{g}_n, \qquad \text{preconditioning}$$

$$\mathbf{c}_n = \mathbf{p}_n + \frac{\mathbf{p}_n^*(\mathbf{g}_n - \mathbf{g}_{n-1})}{\mathbf{p}_n^*\mathbf{g}_n}\mathbf{c}_{n-1},\ \mathbf{c}_1 = \mathbf{p}_1, \qquad \text{conjugate direction}$$

$$\eta_n = \frac{\mathbf{c}_n^*\mathbf{g}_n}{\mathbf{c}_n^*\underline{\mathbf{D}}_n^*\underline{\mathbf{C}}_\mathbf{d}^{-1}\underline{\mathbf{D}}_n\mathbf{c}_n + \mathbf{c}_n^*\underline{\mathbf{C}}_\mathbf{m}^{-1}\mathbf{c}_n}, \qquad \text{calculate step length}$$

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \eta_n\mathbf{c}_n,\Big\}. \qquad \text{update model}$$

(13)

n starts at 1 because I used subscript 0 to indicate the a priori model $\mathbf{m}_0$ and field data $\mathbf{d}_0$.

Computational aspects of the conjugate gradient algorithm

I show in the next section that the operation of $\underline{\mathbf{D}}_n^*$ on a vector can be achieved by a forward simulation (wave propagation), so the conjugate gradient equations may be calculated with three forward simulations, one to compute the synthetic data $\mathbf{d}_n = \mathbf{d}(\mathbf{m}_n)$, one to make the gradient as described in the next section (i.e., to compute the $\underline{\mathbf{D}}_n^*\underline{\mathbf{C}}_\mathbf{d}^{-1}\Delta\mathbf{d}_n$ term of $\mathbf{g}_n$), and one to compute the step length η (i.e., the $\underline{\mathbf{D}}_n\mathbf{c}_n$ term). All other computations are simple dot products. If a line search is required, then several more forward simulations may be necessary in order to optimize the step length η. In the examples given here, I carried out such a line search for the optimal η which tended to be within a factor of two of that predicted by equation (12). Because the predicted and optimal values of η were so close, only two additional forward simulations were required to do the line search. For a particular problem, it

may be found that the optimal η will be always approximately the same so a line search may not always be necessary.

There is a complication in the computation of η. Because the elastic wave equation is nonlinear and the Frechét derivatives $\underset{\sim}{\mathbf{D}}_n$ are never explicitly obtained, $\underset{\sim}{\mathbf{D}}_n \mathbf{c}_n$ cannot be computed directly. Instead, it must be computed using the forward problem $\mathbf{d}(\mathbf{m})$ with the following formula:

$$\begin{aligned}\underset{\sim}{\mathbf{D}}_n \xi \mathbf{c}_n &= \underset{\sim}{\mathbf{D}}_n(\mathbf{m}_n + \xi \mathbf{c}_n) - \underset{\sim}{\mathbf{D}}_n \mathbf{m}_n \\ &= \mathbf{d}(\mathbf{m}_n + \xi \mathbf{c}_n) - \mathbf{d}(\mathbf{m}_n),\end{aligned} \tag{14}$$

where ξ is chosen such that the model perturbation $\xi \mathbf{c}_n$ is sufficiently small relative to the model parameters $\mathbf{m}_n$ (say, 1 percent) that the function $\mathbf{d}(\mathbf{m})$ is almost linear. Therefore, in practice, the formula for η is given by

$$\eta_n = \frac{\mathbf{c}_n^* \mathbf{g}_n}{\frac{1}{\xi^2}\left[(\mathbf{c}_n^* \xi \underset{\sim}{\mathbf{D}}_n^* \underset{\sim}{\mathbf{C}}_{\mathbf{d}}^{-1} \underset{\sim}{\mathbf{D}}_n \xi \mathbf{c})\right] + \mathbf{c}_n^* \underset{\sim}{\mathbf{C}}_{\mathbf{m}}^{-1} \mathbf{c}_n}, \tag{15}$$

where $\underset{\sim}{\mathbf{D}}_n \xi \mathbf{c}_n$ is computed using equation (14).

A further complication is the covariance matrices. If these were assumed to be arbitrary, they would require excessive computer storage and CPU time. Consequently, in practice diagonal matrices are taken. Though diagonal, the covariance matrices need not be constant but may vary along the diagonal which assumes independent but nonconstant noise and independent model parameters. For this assumption to be reasonable, it may be necessary to choose the model parameters (and perhaps data parameters) carefully. A viable alternative to the choice of diagonal covariance matrices is when $\underset{\sim}{\mathbf{C}}_{\mathbf{d}}^{-1}$ and $\underset{\sim}{\mathbf{C}}_{\mathbf{m}}^{-1}$ are constant in the diagonal direction and so represent simple filtering operations.

GRADIENT CALCULATION

This paper uses the implied summation notation for all variables that are not vectors or matrices. The implied summation notation states that repeated subscripts imply there is a sum over that subscript. Also, any implied sums within parentheses must be performed first. Continuous functions are used in this section because they are easy to manipulate mathematically. In practice, seismic data are measured at discrete intervals (for instance, time is sampled every few milliseconds), so calculations are made using discretized versions of the final formulas. Matrices and vectors always refer to discretized operators and functions, respectively.

If the earth is assumed to be perfectly elastic, then the seismic forward problem $\mathbf{d}(\mathbf{m})$ may be computed by solving the elastic wave equation (Aki and Richards, 1980),

$$\rho \ddot{u}_i - \partial_j c_{ijk\ell} \partial_\ell u_k = f_i, \tag{16a}$$

$$c_{ijk\ell} \partial_\ell u_k n_j = T_i, \tag{16b}$$

$$u_i = 0, \qquad t < 0, \tag{16c}$$

$$\dot{u}_i = 0, \qquad t < 0, \tag{16d}$$

where $u_i = u_i(\mathbf{x}_S, \mathbf{x}, t)$ is the ith component of displacement resulting from a shot (i.e., body force f_i and/or traction T_i) located at $\mathbf{x}_S$. If the receivers are located at $\mathbf{x}_R$, then the data are given by

$$d(\mathbf{x}_S, \mathbf{x}_R, t, i, \mathbf{m}) = u_i(\mathbf{x}_S, \mathbf{x}_R, t, \mathbf{m}), \quad i = 1, 2, 3, \tag{17}$$

so $\mathbf{d}(\mathbf{m})$ is $d(\mathbf{m})$ discretized and arranged into a vector.

In order to perform inversion using the conjugate gradient algorithm [equations (13) and (15)], the gradient $\mathbf{g}$ [see equation (2)] corresponding to model parameters $\mathbf{m}$ is required. For the elastic wave equation, the most obvious choices of model parameters $\mathbf{m}$ are the Hooke tensor elements $c_{ijk\ell}$, the density ρ, the body force f_i, and the traction T_i (for an isotropic medium, the Hooke tensor can be described by the two Lamé parameters λ and μ). However, there are other choices of model parameters which are more physically meaningful and (as it turns out) better resolved (Tarantola et al., 1985). In particular, one could choose the P- and S-wave velocities and density or the P- and S-wave impedances and density as model parameters. The gradients in terms of the different choices of model parameters are related, so the development will proceed by deriving the gradient in terms of the simplest choice (λ, μ, and ρ). Subsequently, the expressions for the gradients in terms of the other model parameter choices will be derived and cast in terms of the gradient for λ, μ, and ρ. This is not equivalent to inverting for λ, μ, and ρ followed by transformation to the other parameters because different choices of model parameters affect the shape of the objective function. A poorer choice leads to more elongate minima; consequently, more iterations of the gradient algorithm are required to achieve convergence.

The derivation is based on finding an expression equivalent to the linearized forward problem

$$\delta \mathbf{d} = \underset{\sim}{\mathbf{D}} \delta \mathbf{m}. \tag{18a}$$

In continuous form,

$$\delta \mathbf{d}(D) = \int_M dM \, \frac{\partial \mathbf{d}(D)}{\partial \underset{\sim}{\mathbf{m}}(M)} \, \delta \mathbf{m}(M), \tag{18b}$$

where M indicates the model space. Equation (18b) indicates how to calculate a small perturbation in the wave field $\delta \mathbf{d}$ resulting from a small perturbation in the model parameters $\delta \mathbf{m}$ by integrating over the model space, i.e., it is the linearized Green's function representation of the forward problem or Born approximation. With a representation equivalent to equation (18b), one can identify the Frechét kernel $\underset{\sim}{\mathbf{D}} = \partial \mathbf{d}/\partial \mathbf{m}$ and then compute the adjoint operation

$$\delta \hat{\mathbf{m}} = \underset{\sim}{\mathbf{D}}^* \delta \mathbf{d}, \tag{19a}$$

or, in continuous form

$$\delta \hat{\mathbf{m}}(M) = \int_D dD \left[\frac{\partial \mathbf{d}(D)}{\partial \underset{\sim}{\mathbf{m}}(M)}\right]^* \delta \mathbf{d}(D). \tag{19b}$$

The circumflex is used to make it clear that $\delta \mathbf{m}$ and $\delta \hat{\mathbf{m}}$ are not the same [in fact, $\delta \hat{\mathbf{m}}$ does not even have the same units as $\delta \mathbf{m}$; it has units of [units (D)]^2/units (M) rather than simply units (M)].

Once expression (19b) has been derived, $\delta \mathbf{d}$ can be replaced by $\underset{\sim}{\mathbf{C}}_{\mathbf{d}}^{-1} \Delta \mathbf{d}$ and added to $\underset{\sim}{\mathbf{C}}_{\mathbf{m}}^{-1} \Delta \mathbf{m}$ to obtain the desired gradient

$$\mathbf{g} = \underset{\sim}{\mathbf{D}}^* \underset{\sim}{\mathbf{C}}_{\mathbf{d}}^{-1} \Delta \mathbf{d} + \underset{\sim}{\mathbf{C}}_{\mathbf{m}}^{-1} \Delta \mathbf{m}.$$

Adjoint of the elastic wave equation parameterized by the Lamé parameters and density

This derivation of the adjoint of the elastic wave equation is based upon that of Tarantola (1984). For the seismic problem, the linearized forward problem equivalent to equation (18b) is of the form

$$\delta u_i(\mathbf{x}_S, \mathbf{x}_R, t) = \int_V dV\ (\mathbf{x}) \frac{\partial u_i(\mathbf{x}_S, \mathbf{x}_R, t)}{\partial \mathbf{m}(\mathbf{x})} \delta \mathbf{m}(\mathbf{x}), \tag{20}$$

where $u_i(\mathbf{x}_S, \mathbf{x}_R, t)$ represents the seismogram located at receiver location $\mathbf{x}_R$ which records the ith component of displacement of an elastic wave field due to a shot at $\mathbf{x}_S$ and

$$\mathbf{m}(\mathbf{x}) = \Big[c_{1111}(\mathbf{x}) \cdots c_{ijk\ell}(\mathbf{x}) \cdots c_{3333}(\mathbf{x}), \rho(\mathbf{x}), f_1(\mathbf{x}) \cdots f_3(\mathbf{x}), T_1(\mathbf{x}) \cdots T_3(\mathbf{x}) \Big]^T$$

at location $\mathbf{x}$ in the earth. Therefore, the adjoint problem corresponding to equation (19b) is

$$\delta \hat{\mathbf{m}}(\mathbf{x}) = \sum_S \int dt \sum_R \frac{\partial u_i(\mathbf{x}_S, \mathbf{x}_R, t)}{\partial \mathbf{m}(\mathbf{x})} \delta u_i(\mathbf{x}_S, \mathbf{x}_R, t), \tag{21}$$

i.e., the integral over the data space of the data residuals multiplied by the Frechét kernel. From the two expressions above, it is clear that in order to obtain the adjoint operation (21), all that is required is the integral expression (20) for the forward problem which gives the perturbation in displacement δu_i corresponding to some perturbations in the model parameters $\delta \mathbf{m}$. This expression defines the Frechét kernel so the adjoint operation (21) could then be written by identifying terms in equations (20) and (21). The linearized solution of the elastic wave equation in terms of Green's functions (the Born approximation) supplies the appropriate integral.

To avoid cluttering the equations with the shot location $\mathbf{x}_S$, the following development refers to a single shot. Therefore, the required adjoint operation can be obtained from the result of this development by simply including the sum over shots.

Now consider the elastic wave equation (16). The integral corresponding to equation (20) can be found by making the following substitutions:

$$u_i \to u_i + \delta u_i, \tag{22a}$$

$$\rho \to \rho + \delta\rho, \tag{22b}$$

$$c_{ijk\ell} \to c_{ijk\ell} + \delta c_{ijk\ell}, \tag{22c}$$

$$f_i \to f_i + \delta f_i, \tag{22d}$$

and

$$T_i \to T_i + \delta T_i. \tag{22e}$$

These substitutions yield a new elastic wave equation describing the displacement perturbation δu_i as a function of new force and traction terms Δf_i and ΔT_i:

$$\rho \delta \ddot{u}_i - \partial_j c_{ijk\ell} \partial_\ell \delta u_k = \Delta f_i, \tag{23a}$$

$$c_{ijk\ell} \partial_\ell \delta u_k n_j = \Delta T_i, \tag{23b}$$

$$\delta u_i = 0, \qquad t < 0, \tag{23c}$$

$$\delta \dot{u}_i = 0, \qquad t < 0, \tag{23d}$$

where the new force and traction terms are

$$\Delta f_i = \delta f_i - \delta\rho \ddot{u}_i + \partial_j \delta c_{ijk\ell} \partial_\ell u_k + O_i(\delta\rho, \delta c_{ijk\ell}, \delta f_i, \delta T_i)^2, \tag{24a}$$

and

$$\Delta T_i = \delta T_i - \delta c_{ijk\ell} \partial_\ell u_k n_j + O_i(\delta\rho, \delta c_{ijk\ell}, \delta f_i, \delta T_i)^2. \tag{24b}$$

The new wave equation given in equation (23) has the same form as the elastic wave equation, and hence its solution can be obtained in terms of the Green's functions of the elastic wave equation. The solution in terms of the elastic Green's functions (Aki and Richards, 1980) is

$$\delta u_i(\mathbf{x}_R, t) = \int_V dV\ (\mathbf{x}) G_{ij}(\mathbf{x}_R, t; \mathbf{x}, 0) * \Delta f_j(\mathbf{x}, t) + \int_S dS\ (\mathbf{x}) G_{ij}(\mathbf{x}_R, t; \mathbf{x}, 0) * \Delta T_j(\mathbf{x}, t). \tag{25}$$

This equation has the form of the desired expression for the forward problem, equation (20), and so it defines the Frechét kernel $\partial u_i(\mathbf{x}_S, \mathbf{x}_R, t)/\partial \mathbf{m}(\mathbf{x})$, from which one may obtain the adjoint expression, equation (21). A simplified adjoint expression can be obtained with some algebra.

Substituting the force and traction terms given in equations (24) into equation (25), neglecting the O^2 terms (i.e., assume small perturbations in order to obtain the Frechét derivatives), dropping the arguments of the various functions to avoid cluttering, and using notation $u_{m,\ell} = \partial_\ell u_m$ yield

$$\delta u_i = \int_V dV\ G_{ij} * (\delta f_j - \delta\rho \ddot{u}_j + \partial_k \delta c_{jk\ell m} u_{\ell,m}) + \int_S dS\ G_{ij} * (\delta T_j - \delta c_{jk\ell m} u_{\ell,m} n_k). \tag{26}$$

This equation may be greatly simplified by using the following three results:

(1) $f(t) * \dot{g}(t) = \dot{f}(t) * g(t)$, a property of convolution;

(2) $\int_V dV\ \partial_k F = \int_S dS\ n_k F$, the divergence theorem;

and

(3) $\partial_k \Big[G_{ij} * (\delta c_{jk\ell m} u_{\ell,m}) \Big] = G_{ij} * (\partial_k \delta c_{jk\ell m} u_{\ell,m}) + (\partial_k G_{ij}) * (\delta c_{jk\ell m} u_{\ell,m})$,

by the chain rule. Applying these results to equation (26) yields the formula

$$\delta u_i = \int_V dV\ G_{ij} * \delta f_j + \int_S dS\ G_{ij} * \delta T_j - \int_V dV\ \dot{G}_{ij} * \dot{u}_j \delta\rho - \int_V dV\ (\partial_k G_{ij}) * (\delta c_{jk\ell m} u_{\ell,m}). \tag{27}$$

Finally, assuming isotropy (if anisotropic inversions are de-

sired, this assumption can be dropped from the development),

$$\delta c_{jk\ell m} = \delta\lambda\delta_{jk}\delta_{\ell m} + \delta\mu(\delta_{jm}\delta_{k\ell} + \delta_{j\ell}\delta_{km}), \tag{28}$$

and therefore

$$\begin{aligned}\delta u_i = &\int_V dV\ G_{ij} * \delta f_j + \int_S dS\ G_{ij} * \delta T_j \\ &- \int_V dV\ \dot{G}_{ij} * \dot{u}_j \delta\rho - \int_V dV\ G_{ij,j} * u_{m,m}\delta\lambda \\ &- \int_V dV\ G_{ij,k} * (u_{j,k} + u_{k,j})\delta\mu.\end{aligned} \tag{29}$$

Equation (29) has exactly the same form as equation (20), so clearly it defines the Frechét kernel $\partial u/\partial \mathbf{m}$ for model parameters ρ, λ, μ, f_j, and T_j. Use of equation (29) to solve the forward problem is referred to as the Born approximation. In this paper the Born approximation is not used to solve the forward problem; rather the full elastic wave equation (16) is used. Integrating the Frechét kernel defined by equation (29) over the data space [see equation (21)] produces the adjoint operation

$$\delta\hat{\mathbf{m}} = \left[\delta\hat{\rho},\ \delta\hat{\lambda},\ \delta\hat{\mu},\ (\delta\hat{f}_j, j = 1, 2, 3),\ (\delta\hat{T}_j, j = 1, 2, 3)\right]^T$$

where

$$\delta\hat{\rho} = -\int dt \sum_R \dot{G}_{ij} * \dot{u}_j \delta u_i, \tag{30a}$$

$$\delta\hat{\lambda} = -\int dt \sum_R G_{ij,j} * u_{m,m}\delta u_i, \tag{30b}$$

$$\begin{aligned}\delta\hat{\mu} &= -\int dt \sum_R G_{ij,k} * (u_{j,k} + u_{k,j})\delta u_i \\ &= -\int dt \sum_R \frac{1}{\sqrt{2}}(G_{ij,k} + G_{ik,j}) \\ &\quad * \frac{1}{\sqrt{2}}(u_{j,k} + u_{k,j})\delta u_i,\end{aligned} \tag{30c}$$

$$\delta\hat{f}_j = \int dt \sum_R G_{ij} * \delta u_i, \tag{30d}$$

$$\delta\hat{T}_j = \int dt \sum_R G_{ij} * \delta u_i. \tag{30e}$$

These expressions may be simplified by using the commutativity of convolution (i.e., $f * g = g * f$) and the following property (which shifts the location of the convolution):

$$\int dt\ f(t) * g(t)h(t) = \int dt\ f(-t)g(t) * h(-t). \tag{31}$$

The resulting integrals corresponding to equations (21) for the adjoint in terms of λ, μ, and ρ have the form

$$\begin{aligned}\delta\hat{\gamma}(\mathbf{x}) = &\int dt \sum_R \left[\Omega^\gamma_{ijk} u_j(\mathbf{x}, -t)\right] \\ &\times \left\{\Omega^\gamma_{ijk}\left[G_{ij}(\mathbf{x}_R, t; \mathbf{x}, 0) * \delta u_i(\mathbf{x}_R, -t)\right]\right\} \\ = &\int dt \left[\Omega^\gamma_{ijk} u_j(\mathbf{x}, t)\right] \\ &\times \left\{\Omega^\gamma_{ijk}\left[\sum_R G_{ij}(\mathbf{x}, -t; \mathbf{x}_R, 0) * \delta u_i(\mathbf{x}_R, t)\right]\right\} \\ = &\int dt \left[\Omega^\gamma_{ijk} u_j(\mathbf{x}, t)\right]\left[\Omega^\gamma_{ijk}\psi_j(\mathbf{x}, t)\right],\end{aligned} \tag{32a}$$

where Ω^γ_{ijk} is an operator specified below and $\gamma = \lambda$, μ, or ρ. Remember that implied sums inside brackets or braces must be performed first. I have made use of the reciprocity property of the elastic Green's functions (i.e., interchangeability of $\mathbf{x}$ with $\mathbf{x}_R$), the property

$$\int dt\ f(-t)g(-t) = \int dt\ f(t)g(t),$$

and have introduced the new wave field

$$\psi_j(\mathbf{x}, t) = \sum_R G_{ij}(\mathbf{x}, -t; \mathbf{x}_R, 0) * \delta u_i(\mathbf{x}_R, t), \tag{32b}$$

to be interpreted later. Note that I also introduced an operator Ω^γ_{ijk} which is dependent upon the type of model parameter γ at each location $\mathbf{x}$. Specifically, from equations (30),

$$\begin{aligned}\Omega^\rho_{ijk} &= \sqrt{-1}\ \partial_t, \\ \Omega^\lambda_{ijk} &= \sqrt{-1}\ \partial_j, \\ \Omega^\mu_{ijk} &= \sqrt{\frac{-1}{2}}(\delta_{jk}\partial_i + \delta_{ji}\partial_k).\end{aligned} \tag{33}$$

The meaning of this operator is discussed in detail in the section that compares inversion to migration. The simple form of the adjoint expression given by equation (32a) makes it easy to compare inversion to migration.

Interpreting the adjoint operation

Consider equation (32a) which is the adjoint to the elastic wave equation. There are only two unknowns, the wave fields u_j and ψ_j. The wave field u_j can be computed by forward modeling (i.e., doing a shot simulation). Now consider the definition of ψ_j given by equation (32b). First note that the perturbation wave field δu_i (denoted the "residual wave field") is the difference between the wave field u_j and the observed wave field. Now, observe that the general form of equation (32b) is the same as that of equation (25) except that the Green's function G_{ij} has a minus sign in front of the time variable t. This indicates that here the Green's function corresponds to propagation *backward* in time. Therefore, ψ is generated by applying δu_i as a forcing term in a wave equation that runs backward in time and is therefore called the "back-propagated residual wave field" [for back projection in tomography, Devaney (1984)]. Since the wave equation is symmetric with respect to time, this back propagation can be achieved by doing forward modeling. From this interpretation it is clear that the adjoint can be calculated with the following steps:

(i) Simulate elastic wave propagation to solve for $\mathbf{u}$;

(ii) compute the data residuals $\delta\mathbf{u} = \mathbf{u} - \mathbf{u}_0$;

(iii) back propagate the data residuals (i.e., compute the wave field resulting from a forcing function equal to $\delta\mathbf{u}$ applied in a time reversed wave equation). Simultaneously apply Ω^γ_{ijk} and compute the time integral which represents the computation of the correlation between Ω^γ_{ijk} operated on the forward-propagated field and Ω^γ_{ijk} operated on the back-propagated residual wave field.

The gradient direction

The conjugate gradient inversion algorithm (13) requires the gradient $\mathbf{g} = \underset{\sim}{\mathbf{D}}^* \underset{\sim}{\mathbf{C}}_{\mathbf{d}}^{-1} \Delta\mathbf{d} + \underset{\sim}{\mathbf{C}}_{\mathbf{m}}^{-1} \Delta\mathbf{m}$; however, equation (32) only represents the adjoint operation, i.e., $\underset{\sim}{\mathbf{D}}^* \delta\mathbf{d}$. Assuming, for simplicity of computation, a diagonal data covariance function (uncorrelated noise) and using equation (32) to define the adjoint operator $\underset{\sim}{\mathbf{D}}^*$, the gradient in terms of model parameter γ is

$$\gamma(\mathbf{x}) = \sum_S \int dt \, \frac{1}{C_d^t(t)} \left[\Omega_{ijk}^{\gamma} u_j(\mathbf{x}, t) \right] \left[\Omega_{ijk}^{\gamma} \psi_j(\mathbf{x}, t) \right] + \sum_{\gamma'} C_{\gamma\gamma'}^{-1}(\mathbf{x}) \left[\gamma'(\mathbf{x}) - \gamma_0'(\mathbf{x}) \right], \tag{34a}$$

where I have now reintroduced the shot sum that was dropped in the development and ψ_j is now defined by

$$\psi_j(\mathbf{x}, t) = \sum_R \frac{1}{C_d^{\mathbf{x}_R}(\mathbf{x}_R)} G_{ij}(\mathbf{x}, -t; \mathbf{x}_R, 0) * \Delta u_i(\mathbf{x}_R, t). \tag{34b}$$

ψ_j is the back-propagated residual wave field created by taking the difference Δu_i between the observed data u_{i_0} and the forward-modeled wave field u_i, and using this as a forcing function in a wave equation that runs backward in time. It is assumed that the data covariances can be represented as

$$C_{\mathbf{d}}(\mathbf{x}_R, t) = C_{\mathbf{d}}^{\mathbf{x}_R}(\mathbf{x}_R) C_{\mathbf{d}}^t(t). \tag{34c}$$

The model parameters are assumed to be spatially uncorrelated. If the different kinds of model parameters ($\gamma = \lambda$, μ, etc.) are independent, then the only nonzero term of the sum over γ' is the term where $\gamma' = \gamma$.

Finally, rewriting the gradient (34) using the definition of the operator Ω_{ijk}^{γ} from equation (33) [see also equations (30)], the components of the gradient, now denoted $\delta\hat{\rho}$, $\delta\hat{\lambda}$, $\delta\hat{\mu}$, $\delta\hat{f}_j$, and $\delta\hat{T}_j$, i.e., $\mathbf{g}^T = [\delta\hat{\rho}, \delta\hat{\lambda}, \delta\hat{\mu}, (\delta\hat{f}_j, j = 1, 2, 3), (\delta\hat{T}_j, j = 1, 2, 3)]$

$$\delta\hat{\rho} = -\sum_S \int dt \, \frac{1}{C_{\mathbf{d}}^t} \dot{u}_j \psi_j + C_{\rho\rho}^{-1}(\rho - \rho_0) + \cdots = \delta\rho' + C_{\rho\rho}^{-1}(\rho - \rho_0) + \cdots, \tag{35a}$$

$$\delta\hat{\lambda} = -\sum_S \int dt \, \frac{1}{C_{\mathbf{d}}^t} u_{m,m} \psi_{j,j} + C_{\lambda\lambda}^{-1}(\lambda - \lambda_0) + \cdots = \delta\lambda' + C_{\lambda\lambda}^{-1}(\lambda - \lambda_0) + \cdots, \tag{35b}$$

$$\delta\hat{\mu} = -\sum_S \int dt \, \frac{1}{C_{\mathbf{d}}^t} \frac{1}{\sqrt{2}} (u_{k,j} + u_{j,k}) \frac{1}{\sqrt{2}} (\psi_{k,j} + \psi_{j,k}) + C_{\mu\mu}^{-1}(\mu - \mu_0) + \cdots = \delta\mu' + C_{\mu\mu}^{-1}(\mu - \mu_0) + \cdots, \tag{35c}$$

$$\delta\hat{f}_j = \sum_S \frac{1}{C_{\mathbf{d}}^t} \psi_j + C_{f_j f_j}^{-1}(f_j - f_{j0}) + \cdots = \delta f_j' + C_{f_j f_j}^{-1}(f_j - f_{j0}) + \cdots, \tag{35d}$$

$$\delta\hat{T}_j = \sum_S \frac{1}{C_{\mathbf{d}}^t} \psi_j + C_{T_j T_j}^{-1}(T_j - T_{j0}) + \cdots = \delta T_j' + C_{T_j T_j}^{-1}(T_j - T_{j0}) + \cdots. \tag{35e}$$

These equations define $\delta\lambda'$, $\delta\mu'$, and $\delta\rho'$ which will be used later in the expressions for the gradient in terms of the other choices of model parameters. The dots at the end of each equation represent cross-covariance terms between the different types of model parameters. These cross covariances are usually considered negligible for the sake of simplicity and due to lack of knowledge. Furthermore, constant diagonal covariance functions are normally assumed, i.e., $C_{\rho\rho} = \sigma_{\rho\rho}^2$, etc. (for an exception, see Mora, 1987a). This means that the physical parameters are assumed to be spatially uncorrelated. This is generally false, so biases can be expected in the solutions. Perhaps it would be best to include diagonal cross-covariance functions $C_{\rho\lambda} = \sigma_{\rho\lambda}^2$, etc., if reasonable values can be estimated.

Although the force and traction terms of the gradient have been derived as if they were spatially variable, in seismic exploration it can be assumed that the source locations are known so the force terms need to be solved only at the true source locations [i.e., the force and traction covariance functions C_{f_j} and C_{T_j} are Dirac functions of the form $\delta(\mathbf{x} - \mathbf{x}_S)$]. Actually, it would be best to know the source functions rather than to invert for them because, for reflection data, the source time history cannot be well resolved from the near-surface geology and reverberation effects. However, if the direct wave is known such as is the case for VSP data, the source is much better determined and can be inverted simultaneously with the geology (Harlan, 1984). Normally, with only reflection data it is best to use a preprocessing step to find the source, or more generally, to use relaxation; that is, invert for the source only, followed by inversion for the geology only, and so on.

The time-dependent data covariance function can be approximated as

$$\frac{1}{C_{\mathbf{d}}^t(t)} = \frac{t^{2p}}{\sigma_{\mathbf{d}}^2}, \tag{36}$$

which allows for data errors which decay with time (i.e., errors are assumed to be unexplained seismic waves which diverge with time and hence diminish in amplitude). Typically, for a 2-D problem a value of $p \geq .5$ is reasonable, while in a 3-D problem $p \geq 1$ would be appropriate.

The receiver-dependent inverse covariance function $C_{\mathbf{d}}^{\mathbf{x}_R}(\mathbf{x}_R)$ is usually chosen to be unity, except at the edges of the shot gather where a taper can be applied to decrease artifacts resulting from a sudden edge in the recorded data.

Different model parameters

A theoretical study by Tarantola et al. (1985) indicates that the choice of model parameters is especially important in the elastic case where there are three independent parameters at each earth location. For some choices, three elastic parameters can be easily resolved from one another while for other choices they cannot. The results of the study of Tarantola et al. indicate that for seismic surveys where the source is dominantly *P*-waves, the choice of the Lamé parameters was much worse than the choices of either the velocities or the impedances. Furthermore, the results indicate that the density was not very well resolved (especially from *P*-wave velocity) and that in practice it is probable that only two parameters, such as *P*-wave and *S*-wave velocity or *P*-wave and *S*-wave impedance, will be resolvable. Therefore, I derive expressions for the gradient in terms of the velocities and impedances.

The equations for the gradient in terms of the P-wave velocity α, S-wave velocity β, and density ρ, or the P-wave impedance Z_P, S-wave impedance Z_S, and density ρ, can be easily derived in terms of the gradient for the Lamé parameters and density by changing variables. For example, consider the gradient

$$\mathbf{g}_{\mathbf{m}} = \frac{\partial \mathbf{d}}{\partial \underline{\mathbf{m}}} \underline{\mathbf{C}}_{\mathbf{d}}^{-1} \Delta \mathbf{d} + \underline{\mathbf{C}}_{\mathbf{m}}^{-1} \Delta \mathbf{m},$$

where the subscript $\mathbf{m}$ on $\mathbf{g}$ indicates that the gradient is in terms of model parameters $\mathbf{m}$. The gradient in terms of different model parameters $\mathbf{m}'$ is

$$\mathbf{g}_{\mathbf{m}'} = \frac{\partial \mathbf{d}}{\partial \underline{\mathbf{m}}'} \underline{\mathbf{C}}_{\mathbf{d}}^{-1} \Delta \mathbf{d} + \underline{\mathbf{C}}_{\mathbf{m}'}^{-1} \Delta \mathbf{m}'$$

$$= \frac{\partial \mathbf{d}}{\partial \underline{\mathbf{m}}} \frac{\partial \mathbf{m}}{\partial \underline{\mathbf{m}}'} \underline{\mathbf{C}}_{\mathbf{d}}^{-1} \Delta \mathbf{d} + \underline{\mathbf{C}}_{\mathbf{m}'}^{-1} \Delta \mathbf{m}'. \tag{37}$$

Therefore, all that is needed to evaluate the gradient in terms of different model parameters is the Jacobian $\partial \mathbf{m}/\partial \underline{\mathbf{m}}'$ and the adjoint in terms of the original model parameters $\mathbf{m}$ given by $(\partial \mathbf{d}/\partial \underline{\mathbf{m}})\underline{\mathbf{C}}_{\mathbf{d}}^{-1} \Delta \mathbf{d}$.

For example, the gradient for the P-wave velocity is

$$\delta \hat{\hat{\alpha}} = \frac{\partial \lambda}{\partial \alpha} \delta \lambda' + \frac{\partial \mu}{\partial \alpha} \delta \mu' + \frac{\partial \rho}{\partial \alpha} \delta \rho' + C_{\alpha\alpha}^{-1}(\alpha - \alpha_0) + \cdots, \tag{38}$$

where $\delta\lambda'$, $\delta\mu'$, and $\delta\rho'$ are defined by equations (35). The required elements of the Jacobian $\partial\lambda/\partial\alpha$, etc., can be obtained from the relationships for α and β

$$\alpha = \sqrt{\frac{\lambda + 2\mu}{\rho}}, \qquad \beta = \sqrt{\frac{\mu}{\rho}}, \tag{39}$$

or by solving for λ and μ,

$$\lambda = \rho\alpha^2 - 2\rho\beta^2, \qquad \mu = \rho\beta^2. \tag{40}$$

Using these equations, the gradients in terms of the P- and S-wave velocities and density are

$$\delta \hat{\hat{\alpha}} = 2\rho\alpha\delta\lambda' + C_{\alpha\alpha}^{-1}(\alpha - \alpha_0) + \cdots, \tag{41a}$$

$$\delta \hat{\hat{\beta}} = -4\rho\beta\delta\lambda' + 2\rho\beta\delta\mu' + C_{\beta\beta}^{-1}(\beta - \beta_0) + \cdots, \tag{41b}$$

$$\delta \hat{\hat{\rho}}_{\alpha} = (\alpha^2 - 2\beta^2)\delta\lambda' + \beta^2\delta\mu' + \delta\rho' + C_{\rho\rho}^{-1}(\rho - \rho_0) + \cdots. \tag{41c}$$

The equations relating the impedances Z_P and Z_S to the Lamé parameters are

$$Z_P = \rho\alpha = \sqrt{\rho(\lambda + 2\mu)}, \qquad Z_S = \rho\beta = \sqrt{\rho\mu}, \tag{42}$$

or, solving for λ and μ,

$$\lambda = \frac{1}{\rho}\left(Z_P^2 - 2Z_S^2\right), \qquad \mu = \frac{1}{\rho} Z_S^2. \tag{43}$$

The components of the gradient with respect to the impedances are thus

$$\delta \hat{\hat{Z}}_P = 2\alpha\delta\lambda' + C_{Z_P Z_P}^{-1}(Z_P - Z_{P_0}) + \cdots, \tag{44a}$$

$$\delta \hat{\hat{Z}}_S = -4\beta\delta\lambda' + 2\beta\delta\mu' + C_{Z_S Z_S}^{-1}(Z_S - Z_{S_0}) + \cdots, \tag{44b}$$

$$\delta \hat{\hat{\rho}}_Z = -(\alpha^2 - 2\beta^2)\delta\lambda' - \beta^2\delta\mu' + \delta\rho' + C_{\rho\rho}^{-1}(\rho - \rho_0) + \cdots. \tag{44c}$$

Note that aside from a factor of ρ, the gradient in terms of the velocities [equations (41)] and the gradient in terms of the impedances [equations (44)] are identical except that the density gradient terms have some sign differences. The main effect of the sign differences is to alter the magnitude of the density gradient depending upon which choice of model parameters is used.

Comparison of inversion to migration

The gradient to be used in the conjugate gradient inversion algorithm has the form

$$\delta \hat{\hat{\gamma}} = \sum_s \int dt\; C_{\mathbf{d}}^{-1} \Omega u_j \Omega \psi_j + C_{\gamma}^{-1} \Delta\gamma. \tag{45}$$

If Ω and $C_{\mathbf{d}}^{-1}$ were set to unity and C_{γ}^{-1} were set to zero, then the gradient of equation (45) would do an elastic, prestack, shot-profile migration by correlation of the upgoing reflected waves ψ_j with the downgoing direct wave u_j. Where the correlation is great, there is a reflector, or, in other words, the reflected waves intersect the direct wave at the reflecting interface. This is also similar to Claerbout's U/D concept (Claerbout, 1976) that states that the reflectivity is the ratio of the amplitude of upgoing (reflected) waves U to the downgoing (direct) wave D. Note that sometimes the correlation of U and D is used as a stable approximation to the U/D method. If in addition, the two wave fields u_j and ψ_j represented pressure wave fields, then one would have a conventional acoustic, prestack, shot-profile migration. Hence, conventional migration is equivalent to computation of the least-squares gradient direction of inversion, except that it does not include covariances, the operator Ω, or the damping factor $C_{\gamma}^{-1}\Delta\gamma$ (Lailly, 1984).

It is clear that Ω is a very important operator. It unravels the wave-field information into the desired model parameters using the amplitude information (i.e., the overall amplitude of the reflected waves as well as the amplitude-incidence angle information). Another main difference between the inversion algorithm described here and migration is that the inversion algorithm is iterative. The iterations tend to allow for the fact that the modeling function $\mathbf{d}(\mathbf{m})$ is nonlinear; they also apply the true inverse Hessian of equation (5) which would correct all the relative amplitudes of velocities and densities to their most likely values (i.e., the best-fit values). Note that the iterations also help to regain some of the more poorly resolved low wavenumbers. Without iterations, the inversion algorithm, like migration, would have little hope of obtaining any of the lower wavenumbers except perhaps if a clever preconditioning were constructed. The reason some of the low wavenumbers can be regained is that the inversion tries to match the entire shot profile, which consists of approximately hyperbolic events that can only be matched if the low-wavenumber velocity model is correct. The iterations are necessary because a gradient algorithm is used and such algorithms obtain the best resolved components of the model (high wavenumbers) first and the poorest resolved components (low wavenumbers) last. It will be a challenge of future research to allow for the resolution of low wavenumbers and high wavenumbers simultaneously.

Acceleration by data contraction

Two reasons the inversion method is so costly are that it requires several elastic wave propagations per shot for each

iteration, and the number of shots is typically large for seismic surveys. However, the number of shots can in theory be reduced by about an order of magnitude, thereby bringing the computer time required by this method to a more reasonable figure. Reducing the number of shots is possible by changing the concept of the experiment so that instead of doing 100 shot simulations, only 10 super-shot simulations are done. In this case, a super-shot profile would be obtained by firing 10 shots (every tenth shot) simultaneously using a large fixed-receiver array. Almost the same information is contained in the 10 super-shots, but in a slightly scrambled form so the inversion should give almost the same results but with some additional noise. In particular, because there is presumably more than one seismic wavelength between ten shots, there are still waves illuminating the subsurface at all angles. Hence, spatial resolution should not be seriously affected by using the super-shot method provided the contraction is not so great that the artifacts dominate. In practice, to avoid having to reshoot seismic lines as super shots, every tenth shot profile could be stacked. A problem may be that, due to the different receiver ranges of the different shot profiles, significant artifacts may be introduced into the inversion result. Therefore, some analysis is still required to evaluate whether super-shot inversions would be significantly corrupted by artifacts and noise, especially in the case when shot profiles with different receiver ranges are stacked after recording. The super-shot method will be explored in more detail with some examples in a future paper.

Low wavenumbers in the model solution and the a priori model

The method of inversion presented here is very powerful and can be used for almost any form of seismic data including reflection data (conventional surface seismics) or transmission data (well-to-well or VSP data). When used for transmission problems, it can resolve the low-wavenumber components of the model and does a kind of wave equation tomography (see Gauthier et al., 1986, for an acoustic transmission example and Mora, 1987b, for the elastic case). However, when used to invert reflection seismic data, the low-wavenumber model is not well resolved by the algorithm and would take an excessive number of conjugate gradient iterations to obtain. Inversions of both reflection and transmission data require an a priori model that is adequately close (i.e., has the very low-wavenumber components of the P- and S-wave velocity model such that the kinematics of the waves are approximately correct). Otherwise, the model perturbations located in the inversion will be misplaced (this is comparable to performing depth migration with an inaccurate velocity function). The consequence is that either convergence will be slow or the algorithm will converge to a local minimum. A reasonable a priori (low-wavenumber) P- and S-wave velocity model can be obtained by conventional velocity analysis methods and the a priori density model can be obtained by empirical relationships or some other a priori knowledge. However, rather than relying on an initial traveltime inversion step to help determine the low-wavenumber model, it may be possible to incorporate this step directly into the inversion algorithm. Three possible methods are as follows:

(1) Use of a clever preconditioning operator (i.e., approximate inverse Hessian) to enhance low-wavenumber velocity perturbations. This could be either a boosting of the low-wavenumber components (the small eigenvalues of the problem), or an imposition of non-Gaussian statistics such as blockiness (Harlan, 1986).

(2) Derive a secondary inverse problem that attempts to resolve the low wavenumbers. In this case, one iteration of the "total" inversion algorithm would consist of one iteration for the low-wavenumber part, followed by one iteration for the high-wavenumber part. This method is not unreasonable because the low wavenumbers and high wavenumbers are fairly well decoupled. That is, the high-wavenumber model perturbations cause the amplitudes but not the shapes of the hyperbolas to change, while the low-wavenumber model perturbations cause the shapes but not the amplitudes of hyperbolas to change. Alternatively, an additional term could be included in the objective function that is sensitive to low-wavenumber model perturbations (for example, the square difference between the envelope of the data and the envelope of the modeled data).

(3) Use many inversions steps of low-pass filtered seismic data, gradually relaxing the low-pass filter as steps proceed until finally the full bandwidth is included in the inversion. This method tends to invert for the low-wavenumber model first, followed by the higher wavenumber components. An outer loop gradually increases the length of the time axis to ensure that the shallow part of the model is obtained first. Kolb et al. (1986) illustrated this approach for the case of an acoustic inversion. The drawback with this method is the lack of efficiency.

Now consider the second method. As presented here, the low-wavenumber a priori model $\mathbf{m}_0$ is assumed to be known from a prior "inversion" step (such as NMO), and thereafter it is never changed. However, if NMO is replaced with least squares, i.e., finding a low-wavenumber model that produces hyperbolas that best match the data, and this is included as an extra step at every iteration, the result is a "total" inversion algorithm. One way to do the extra step is by observing that the size of the high-wavenumber model perturbations found in the current algorithm is maximized when the best low-wavenumber model is used (because the kinematics is correct). This is analogous to doing NMO and stacking at different velocity functions. The best quality stack is achieved when the best low-wavenumber velocity model is used. All that would be required for the extra inversion step is the operator $\underline{\mathbf{L}}$ that relates the low-wavenumber model perturbations $\Delta\mathbf{m}_{\text{low}}$ to the high-wavenumber perturbations $\Delta\mathbf{m} = \Delta\mathbf{m}_{\text{high}}$, i.e.,

$$\Delta\mathbf{m}_{\text{high}} = \underline{\mathbf{L}}\Delta\mathbf{m}_{\text{low}}. \tag{46}$$

Then, as for the high-wavenumber derivation, equation (46) can be used to define the adjoint operation and hence the gradient direction for a least-squares inversion. To maximize the perturbations $\Delta\mathbf{m}_{\text{high}}$, one should minimize a functional of the form

$$S(\mathbf{m}_{\text{high}}, \mathbf{m}_{\text{low}}) = -\Delta\mathbf{m}^*_{\text{high}}\,\underline{\mathbf{C}}^{-1}_{\mathbf{m}_{\text{high}}}\Delta\mathbf{m}_{\text{high}} + \Delta\mathbf{m}^*_{\text{low}}\,\underline{\mathbf{C}}^{-1}_{\mathbf{m}_{\text{low}}}\Delta\mathbf{m}_{\text{low}} \tag{47}$$

in the extra inversion step [c.f. equation (1b)]. This method seems preferable to the use of alternative (1) above because it uses the low-wavenumber information in the seismic data (hyperbolic events) to restrict the low-wavenumber model rather than only relying on some additional a priori statistics. Each iteration of the low-wavenumber inversion step is, like NMO, based on the elastic wave equation. Perhaps, the best method of obtaining the low wavenumbers would be a combination of methods (1) and (2).

RESULTS

The following inversions of synthetic reflection seismic data sets demonstrate what information the algorithm can gain about the deeper geology when applied to *reflection* seismic data. Because the use of Rayleigh waves would resolve only the near-surface, they have been excluded from these calculations (i.e., an absorbing boundary condition was used rather than a free-surface condition at the Earth's surface). Note that, together with refracted waves, Rayleigh waves should help resolve the near-surface (and hence the statics problem), and so it is desirable to use a free-surface condition in general. Inversions of Rayleigh and refracted waves and inversions of transmitted-wave data (VSP and/or well-to-well configurations) will be the topic of future papers. One further point is that the sizes of the velocity and density perturbations in these examples were deliberately chosen to be quite small (ranging from 5 to 15 percent) so that the inverse problem would be more linear and hence the convergence would be more rapid. Considering the large CPU requirements of the algorithm and CPU limitations when carrying out this work, rapid convergence was crucial. Therefore, a demonstration of the algorithm when velocity and density perturbations relative to the a priori model are large will have to await a future paper. Even though velocity perturbations are small, the following examples effectively illustrate what can be hoped to be gained by using this inversion procedure.

In particular, these examples illustrate that when *reflection* seismic alone is used in the inversion, the high-wavenumber velocity and density perturbations are resolvable. Furthermore, *P*- and *S*-wave velocities are resolved, while *P*-wave velocity and density are not well resolved. The synthetic data were calculated with elastic finite differences using a similar (faster) version of the algorithm due to Kosloff et al. (1984) and Cerjan et al. (1985) (Mora, 1986). The inversion was carried out with the conjugate gradient algorithm [equations (13) and (15)], using the gradient parameterized in *P*- and *S*-wave velocities and density denoted α, β, and ρ. The gradient was calculated with equations (41a) through (41c), using definitions from equation (34b) and equations (35a) through (35c). To concentrate on the resolvability of α, β, and ρ, the source was assumed to be known and so was *not* varied with equations (35d) and (35e) but was fixed at its true value throughout the inversions. The method of elastic finite differences was used to compute the forward-modeled wave field u_j and the back-propagated residual wave field ψ_j. Equation (36) was used to define the data covariance matrix (this is equivalent to saying that a time varying gain of t^p was applied) where the value of p was .5 or .75 depending upon the example. The data variance was small so the inversion was not heavily damped; hence the solution was not required to stay close to the a priori model $\mathbf{m}_0$ [i.e., little heed was paid in the conjugate gradient algorithm, equation (13), to $\Delta\mathbf{m}_n^* \mathbf{C}_m^{-1} \Delta\mathbf{m}_n$, $\mathbf{C}_m^{-1}\Delta\mathbf{m}$ or $\mathbf{c}_n^* \mathbf{C}_m^{-1}\mathbf{c}_n$]. A preconditioning of

$$\mathbf{C}_m = \begin{bmatrix} C_{\alpha\alpha} & 0 & 0 \\ 0 & C_{\beta\beta} & 0 \\ 0 & 0 & C_{\rho\rho} \end{bmatrix}$$

was applied in algorithm (13) where $C_{\alpha\alpha} = \alpha_{ave}^2$, $C_{\beta\beta} = \beta_{ave}^2$, and $C_{\rho\rho} = \rho_{ave}^2$ and cross-covariances between different model parameters were assumed to be zero. Recall that the model parameters were assumed to be uncorrelated in equation (34a). A spatial deconvolution of the model parameters could be used to remove any spatial correlations that exist in practice. A band limited source wavelet was used (a second derivative of a Gaussian curve, in the far field, with a fundamental frequency of 20 Hz).

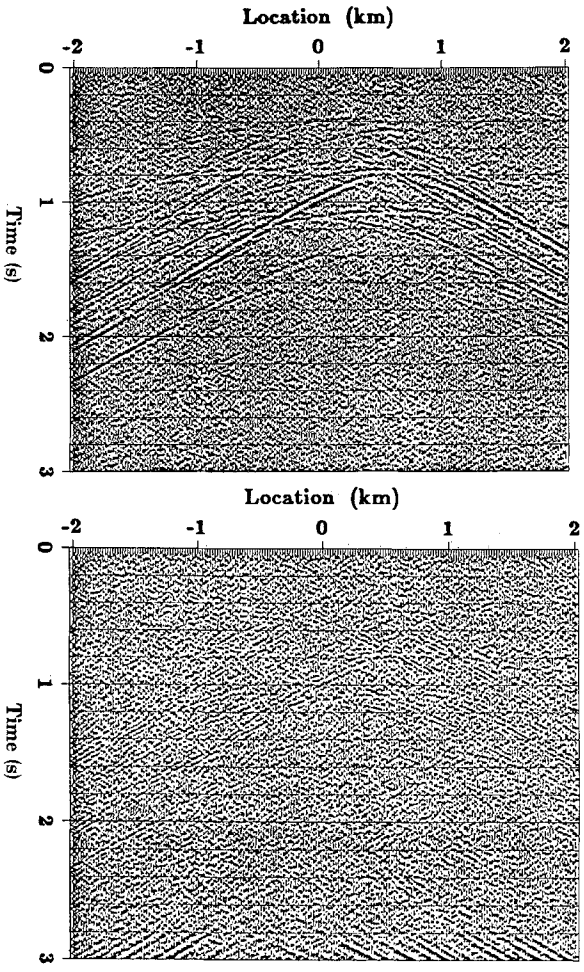

FIG. 1. Noisy diffraction data computed from a twelve-diffractor model and the residual after five iterations. (a) Vertical-component shot gather due to a vertical force with the direct wave removed (i.e., the data generated using the constant-background model have been subtracted so this shot profile is really the residual prior to iterations), and (b) vertical-component residual after five iterations.

Inversion of noisy diffractions

In order to test the algorithm, I carried out an inversion of a synthetic two-component shot profile generated from a model consisting of twelve diffractors in *P*- and *S*-wave velocity and density embedded in a homogeneous space (background properties were $\alpha = 3$ km/s, $\beta = 1.7$ km/s, and $\rho = 2$ g/cm^3). The source was a vertical force located on the top of the model in the middle. The synthetic data were contaminated with about 50–100 percent band-limited Gaussian noise, and an inversion was performed using the homogeneous background model as the initial guess. After five iterations, the residual was mainly Gaussian noise (Figure 1 shows the vertical component data and residual). The true model is plotted alongside the one-iteration and five-iteration inversion results in Figures 2 through 4. Note that the scales of the plots are not the same, but have been chosen for display. The absolute magnitudes of the inversion results are too small, by about a factor of 4, due to problems associated with band limitations which lead to a null space. That is, for band-limited data, more than one physical model have almost identical seismic responses (i.e., the band-limited solution has almost the same seismic response as the true model and so is probabilistically equivalent to the true model). The results show that the diffractors have been well located spatially and that the algorithm has resolved between *P*-wave and *S*-wave velocities but not between *P*-wave velocity and density. This is in agreement with the study of model parameter choice (Tarantola et al., 1985) which indicates that probably only two parameters can be resolved, for instance *P*-wave velocity and *S*-wave velocity or *P*-wave impedance and *S*-wave impedance. Observe that there is some noise in the solution caused by the Gaussian noise in the data, especially in areas around the edges of the model which are more poorly resolved and hence more susceptible to noise contamination (the noise in these regions is mainly seen as smile-like events similar to migration artifacts). However, the level of noise is low considering that only one shot was used and the problem is overdetermined by only a

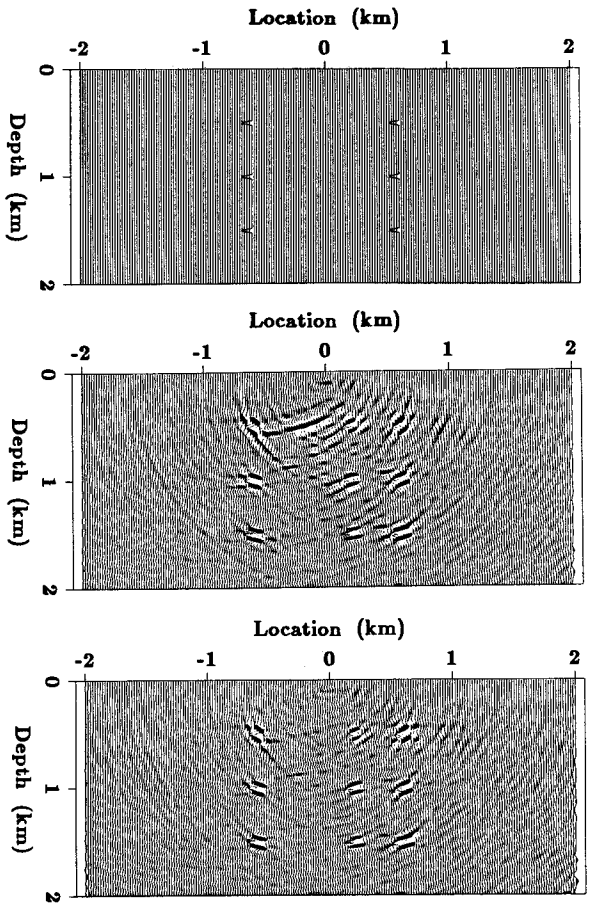

FIG. 2. True model and inversion results for the two-component noisy diffraction data example (i.e., noise-contaminated, vertical-component source, two-component receiver data computed from the twelve-diffractor model). (a) True *P*-wave velocity model, (b) *P*-wave velocity result after one iteration, and (c) *P*-wave velocity result after five iterations.

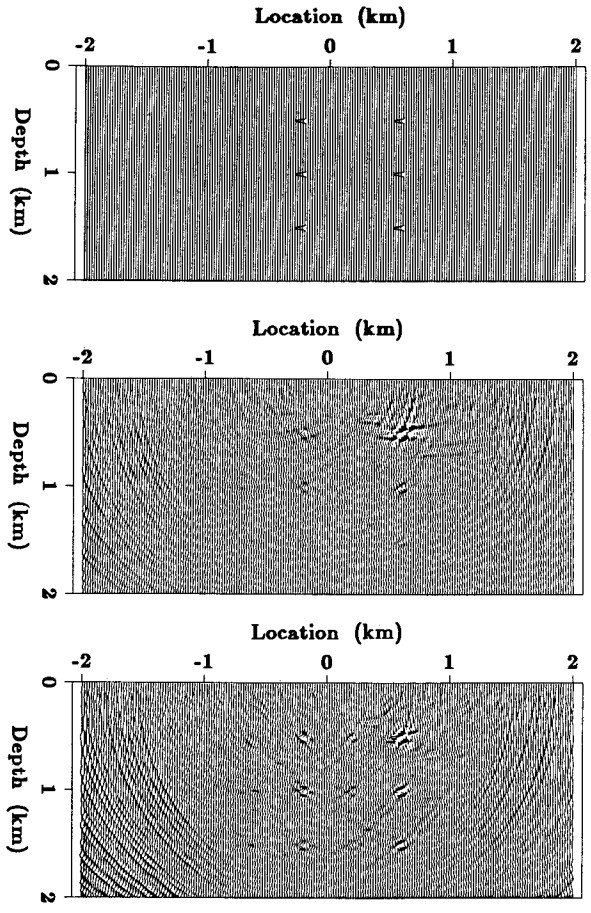

FIG. 3. True model and inversion results for the two-component noisy diffraction data example (i.e., noise-contaminated, vertical-component source, two-component receiver data computed from the twelve-diffractor model). (a) True *S*-wave velocity model, (b) *S*-wave velocity result after one iteration, and (c) *S*-wave velocity result after five iterations.

factor of 2.75. Note that the spatial resolution decreases slightly with depth due to smaller angles of incidence at greater depth. However, the resolution between the different model parameters remains about constant with depth because the incidence angles for the deepest diffractors are still large enough (around 45 degrees) to obtain adequate resolution between the three model parameters. Note that typical seismic surveys have incidence angles up to about 25 degrees for the deepest part of the section, suggesting that spread lengths should be increased in order to achieve similar resolution between the different earth parameters at all depths. The main effect of the iterations is to decrease the size of the artifacts relative to the diffractor perturbations. Actually, the artifacts are mainly generated in the first iteration; later iterations boost the diffractor amplitude without significantly boosting the artifacts. Another main effect of the iterations is to correct the sizes of the *S*-wave velocity solution. This illustrates the advantage of iterative elastic inversion over elastic migration which would give a result similar to the first iteration of the inversion.

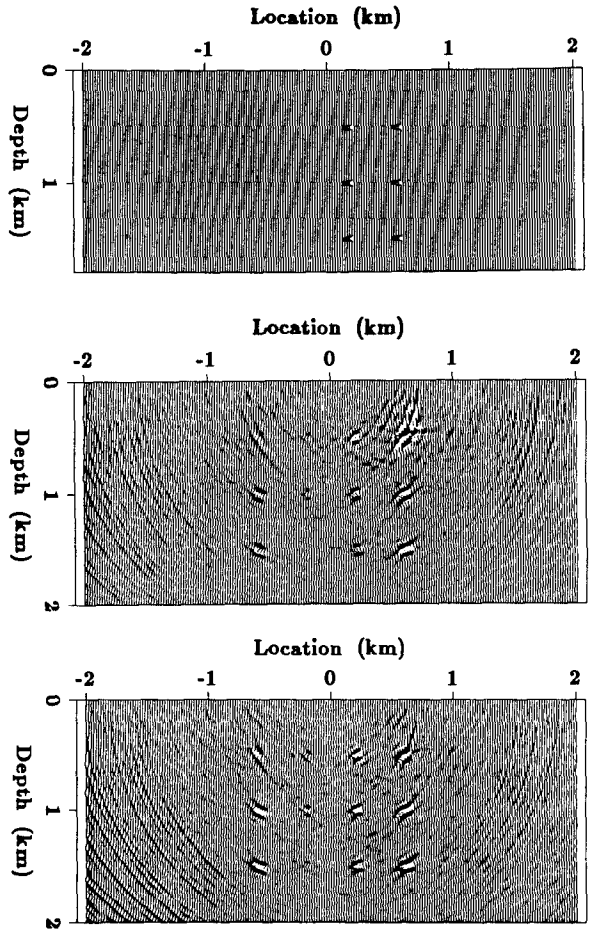

FIG. 4. True model and inversion results for the two-component noisy diffraction data example (i.e., noise-contaminated, vertical-component source, two-component receiver data computed from the twelve-diffractor model). (a) True density model, (b) density result after one iteration, and (c) density result after five iterations.

Inversion of single-component data

To illustrate the importance of two-component data (or equivalently, wide offsets) which contain significant shear-wave energy in the form of *P-S*, *S-P*, and *S-S* diffractions, an inversion was carried out on the synthetic vertical-component data generated from the twelve-diffractor model *without* including the corresponding horizontal-component data. No noise was added to the data (because the effect of noise has already been studied and I wish to focus attention here on the effect on resolution of including different components of data). The inversion results after one and five iterations are plotted in Figures 5 through 7. Compared with the inversion results of the two-component data shown in Figures 2 through 4, the resolution of the *S*-wave velocity is poorer, especially in the early iterations. However, after five iterations there is still fair resolution though it is certainly much worse than when two-component data were used. Probably, in a more realistic example with very slow near-surface velocities and hence even less *P-S*, *S-P*, and *S-S* energy than in this example, the resolution of the *S*-wave velocity would be even worse. The poor resolution could be overcome by recording the data at very wide offsets. Note that the one-iteration *P*-wave velocity result of Figure 5a and the five-iteration result of Figure 2c compare favorably. This indicates that when only a little converted-mode energy exists (such as in most routinely recorded seismic

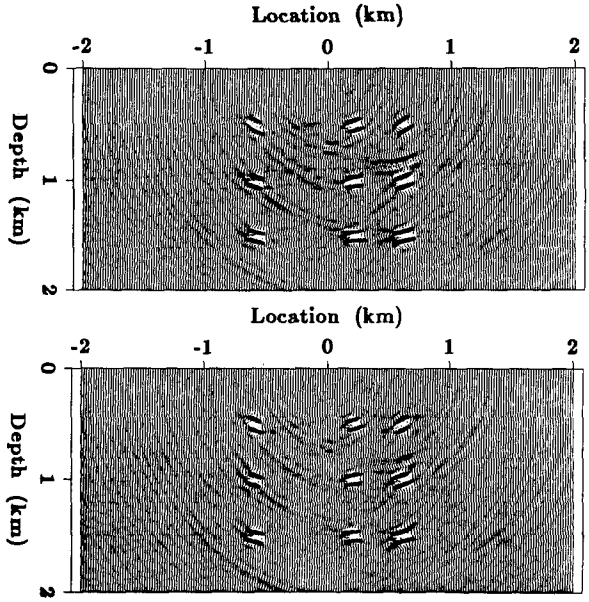

FIG. 5. Inversion results for the single-component diffraction data example (i.e., noise-free, vertical-component source, vertical-component receiver data computed from the twelve-diffractor model). (a) *P*-wave velocity result after one iteration, and (b) *P*-wave velocity result after five iterations.

data), a migration (i.e., one iteration of an inversion) will normally suffice to give a good *P*-wave velocity image and in that case the elastic algorithm is unnecessary. If some converted-mode energy is present, then the iterative elastic inversion algorithm is worthwhile because it obtains an *S*-wave velocity image. Note that if there are converted waves in a data set, then there will be artifacts even in a strictly acoustic inversion (or migration), yielding a result similar to Figure 2b, because the converted waves will be treated as compressional waves. Therefore, the iterations both help resolve the *S*-wave velocity and help reduce artifacts in the *P*-wave velocity result (i.e., take Figure 2b to 2c). If no converted-mode energy is present, then the elastic inversion will give almost the same results for *P*-wave velocity and density as an acoustic inversion (if the *S*-wave velocity is set to zero in the elastic scheme, then the results will be identical). However, if many iterations are done so that the amplitude-offset response becomes well matched in the inversion procedure, then the *S*-wave velocity (and perhaps even density) will begin to be resolved too.

In a marine environment where two-component data are unavailable, long offsets would be required to record significant converted energy. Fortunately, long cables are not very expensive in marine surveys. If long-cable recording were not available, then it is likely that a more efficient algorithm for inversion of *P*- and *S*-wave velocities would be one which converges rapidly using amplitude-offset information alone (e.g., Mora, 1987a). The reasons for the rapid convergence achieved here were the assumptions of plane layers and ray theory, making it relatively cheap to compute the exact inverse Hessian operator of the least-squares algorithm (see McAulay, 1985, for an acoustic example). Unfortunately, the inverse Hessian is too expensive to calculate in the 2-D elastic inversion scheme presented here.

Inversion of four-component data

Is it worth doing *S*-wave surveys when there is plenty of *S*-wave energy contained in normal surveys using large offsets and two-component geophones? With conventional methods of processing, expensive *S*-wave surveys are the only way to get good *S*-wave velocity images because the processing assumes only one wave type (i.e., conventional *S*-wave processing mainly uses the old *P*-wave processing techniques). However, the elastic inversion algorithm uses the elastic wave equation and can correctly handle the presence of both *P*-waves and *S*-waves to produce two separate images. In this case, is the resolution significantly improved by including *S*-wave sources (i.e., horizontal forces) as well as vertical sources? The following example suggests that the answer is no.

Noise-free two-component shot profiles were generated from both a vertical force and a horizontal force using the twelve-diffractor model, and an inversion was carried out. The inversion result is shown in Figure 8. Clearly, the result after five iterations is almost identical to the result when only vertical shots were used. The only significant difference is that when the horizontal source was included, the *S*-wave velocities had almost the correct relative magnitudes from the first iter-

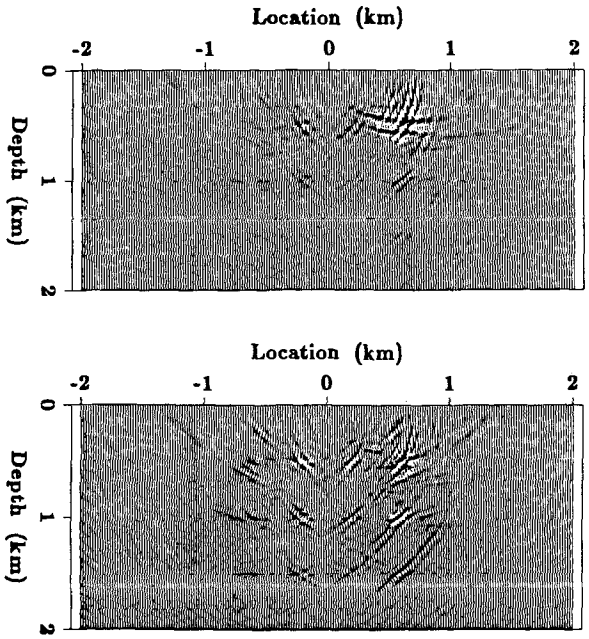

FIG. 6. Inversion results for the single-component diffraction data example (i.e., noise-free, vertical-component source, vertical-component receiver data computed from the twelve-diffractor model). (a) *S*-wave velocity result after one iteration, and (b) *S*-wave velocity result after five iterations.

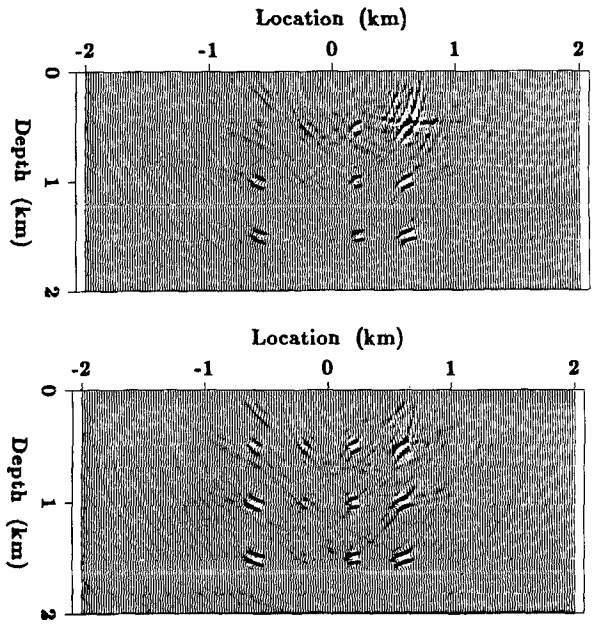

FIG. 7. Inversion results for the single-component diffraction data example (i.e., noise-free, vertical-component source, vertical-component receiver data computed from the twelve-diffractor model). (a) Density result after one iteration, and (b) density result after five iterations.

ation (i.e., convergence was fast). This is because there are now significant *S-S* reflections even at a zero incidence angle (in the horizontal-source shot profile) and this helps resolve the *S*-wave velocity. Previously, the *S*-wave velocity was better resolved where there were more *P-S*, *S-P*, and *S-S* conversions, namely in the shallow part of the model where the incidence angle was higher. Therefore, it seems that by using the iterative inversion approach, one can obtain good S-wave velocity images even when there is no horizontal (shear) source data, provided the vertical source data were recorded to offsets that are long enough that significant mode conversions can take place. Therefore, the data could even be single-component or marine provided the offsets were large enough. Perhaps this means that by basing algorithms on the elastic wave equation, expensive horizontal-source surveys can be eliminated. Of course, the inversion approach is much more expensive than conventional processing and, in any case, the final test will be on field data. One further point is that even with two-component shots and receivers the density is still not perfectly resolved from the *P*- and *S*-wave velocities. Compared to the vertical-source result, *P*-wave velocity and density are better resolved, but *S*-wave velocity and density are more poorly resolved. Therefore, it appears that only two parameters (for instance, *P*- and *S*-wave velocities) can be resolved from seismic data even under almost ideal circumstances. Furthermore, *P*- and *S*-wave velocities are better resolved, whereas there is always fairly poor resolution between density and at least one of the other two parameters. These results suggest it is better to concentrate on elastic inversion for *P*- and *S*-wave velocities rather than on acoustic inversions for *P*-wave velocity and density.

A plane-layered example

The following example demonstrates that the algorithm can resolve some of the low-wavenumber model perturbations but

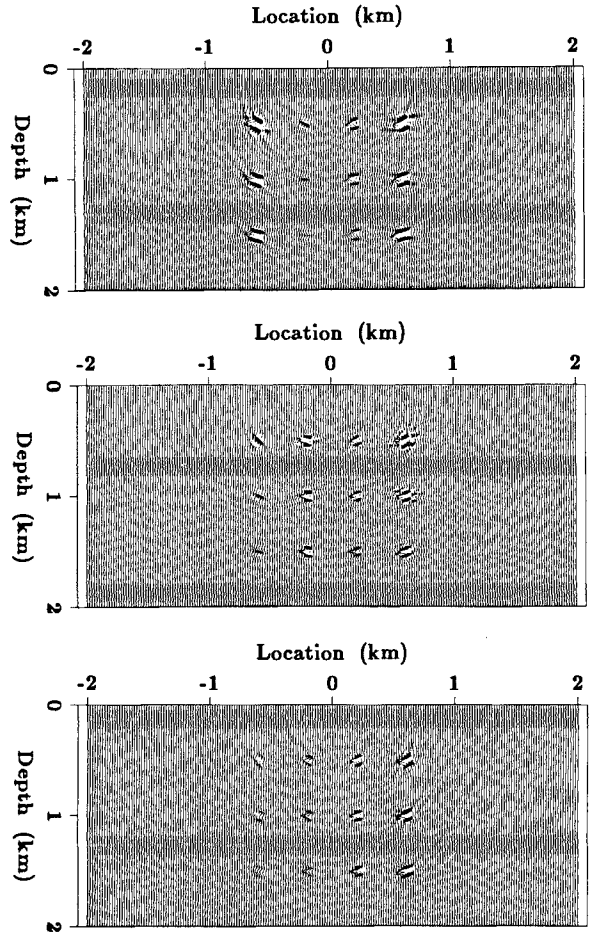

FIG. 8. Inversion results for the four-component diffraction data example (i.e., noise-free, two-component source, two-component receiver data computed from the twelve-diffractor model). (a) *P*-wave velocity result after five iterations, (b) *S*-wave velocity result after five iterations, and (c) density result after five iterations.

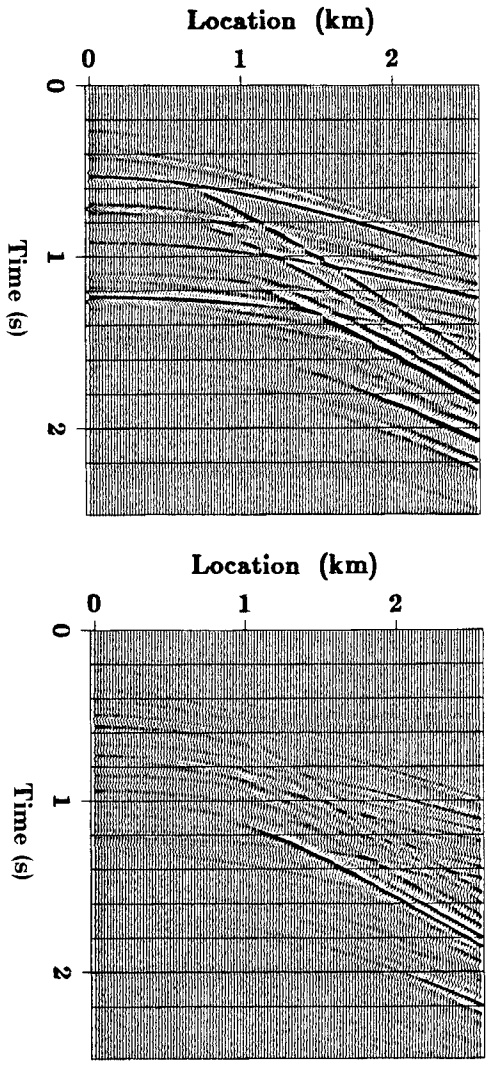

FIG. 9. (a) Vertical-component synthetic shot gather produced from a vertical source for the layered model shown in the solid lines of Figure 10 (the direct wave has been removed by subtracting the data calculated from the starting model), and (b) vertical-component residual after fifteen iterations for the layered example.

that this takes many iterations. Therefore, the algorithm is very inefficient as a method of obtaining the low wavenumbers and some improvements are required (see the section titled "Low wavenumbers in the model solution and the a priori model" for some possible solutions to this problem). The slightly noise-contaminated data shown in Figure 9a were generated from modeling over some plane layers (only the vertical component is shown although the horizontal component was also used in the inversion). The model used to generate these data is shown in Figure 10 in the solid line. A linear fit of this model was used as the starting guess in the inversion. There is still significant energy (especially due to *S*-waves) in the residuals after fifteen iterations (Figure 9b), illustrating slow convergence. The reason for the slow convergence is that the low-wavenumber model perturbations are fairly poorly resolved, so they take many iterations of a gradient algorithm to appear in the solution. The inversion results after one and fifteen iterations are shown in Figure 10, illustrating the gradual creeping of low wavenumbers into the solution. Clearly, the iterations help and the inversion result is much better than the migration result (which would be essentially the same as the first iteration). However, the low wavenumbers only gradually build up so the algorithm requires either a better low-wavenumber starting model or some improvements which help resolve the low wavenumbers. One technique to enhance the low wavenumbers is by finding the boundaries of the main layer from the solution and using these to obtain a blocky solution (i.e., a "blockiness" preconditioning). Note that in this example, unlike the diffractor examples, the magnitudes of the velocity perturbations could be approximately resolved because the blocky model used here has a wavenumber spectrum that has more overlap with the wavelet's spectrum.

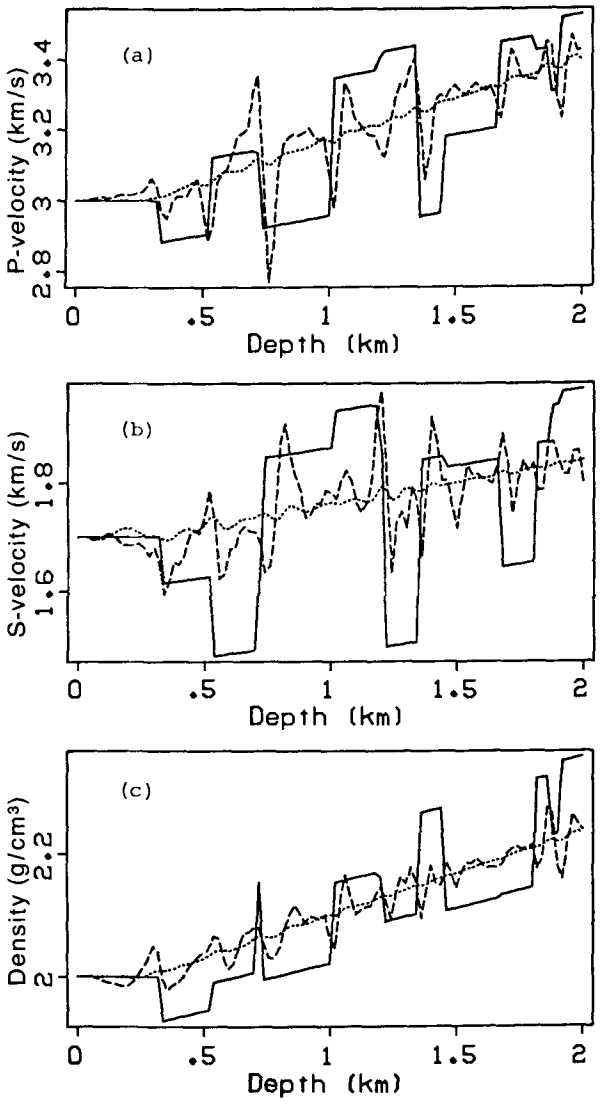

FIG. 10. True model and inversion results for the layered example. (a) True *P*-wave velocity model (solid line) and *P*-wave velocity result after one iteration (dotted line) and fifteen iterations (broken line). (b) True *S*-wave velocity model (solid line) and *S*-wave velocity result after one iteration (dotted line) and fifteen iterations (broken line). (c) True density model (solid line) and density result after one iteration (dotted line) and fifteen iterations (broken line).

DISCUSSION

A method of elastic inversion has been derived using the most obvious choice of model parameters given the elastic wave equation, that is, the Lamé parameters. Equations were also given in terms of two other choices of model parameters, velocities and impedances. The choice of model parameters should be made, where possible, in order to make the model parameters as independent as possible and hence better resolved (see Tarantola et al. (1985), who study resolution of different physical parameter choices or Harlan (1986) who carries out inversion for the depth derivative of impedance in 1-D inversion, thus avoiding problems of knowing a spatially variable mean). Also, the data were parameterized by the displacement amplitude, leading to the Born linearization (i.e., the Frechét derivatives were derived from a Born equation). This has limitations and perhaps better linearizations exist. For example, in transmission problems, the Rytoff linearization which derives Frechét derivatives from phase data rather than amplitude data has a larger linear region. Also, perhaps the model-parameter space can be more heavily restricted in early iterations in order to minimize those components of the model which are clearly nonlinear. For instance, in the Born approximation the kinematics of the wave field must be approximately correct for a good linearization. Therefore, if all wavenumber components of the model were included in an inversion which started relatively far from the solution, false scatterers would be introduced by the inversion in its early iterations, thereby making convergence difficult if not impossible (probably the false scatterers would aggravate the problem of local minima). However, if the inversion proceeded by first inverting only for the lower spatial wavenumber components of the model and then progressively increasing the wavenumber range upward, then no such problem would occur.

Because of the cost of the inversion algorithm, the big area for research and improvements relates to study of the preconditioning operator. Preconditioning should do many things, including source-wavelet deconvolution and restriction of the

estimated model to lie in the part of the null space which intersects with the true model statistics (which are usually non-Gaussian). Preconditioning should greatly speed convergence if it is well chosen. Note that preconditioning is often considered to be an approximation of the inverse Hessian operator, but it can actually do much more. This is because the inverse Hessian is a least-squares concept, while the preconditioning operator can be chosen with no regard for least squares but paying more heed to other factors such as the desired model statistics.

The algorithm outlined is for 2-D inversion. It would be impractical for 3-D inversion at the present time since no one has efficiently solved the 3-D elastic forward problem for relatively arbitrary distributions of physical properties. However, in principle the algorithm could be extended to 2.5-D inversion relatively easily by making some modifications in the computations to allow for out-of-the-plane wave divergence.

CONCLUSIONS

Two-dimensional elastic inversion of multioffset seismic data has been described; with faster convergence it should be feasible, though costly, on today's supercomputers. Synthetic data tests indicate that the elastic inversion works on realistically sized problems with excellent spatial resolution of the high-wavenumber components of the elastic parameters. In the case of reflection seismic data where there is no direct arrival, the low-wavenumber components are at best partially resolved, although they are better resolved than in a migration. The degree of resolution of the low wavenumbers depends mainly on the number of iterations, the degree of noise contamination, and the statistics of the earth model. The resolution between *P*-wave velocity and *S*-wave velocity is quite good, while the resolution between *P*-wave velocity and density is poor. The method has done at least as much as an elastic prestack migration would be expected to do. It provides understanding for a framework of elastic processing and inversion that is badly needed with the increase in multicomponent recordings and wide-angle experiments. The algorithm is very powerful because it can, in principle, be used for almost any situation including inversions of reflected wave fields (e.g., shot-profile inversion as in the examples) and inversions of transmitted wave fields such as in well-to-well tomography (Mora, 1987b). Its limitations are that it is a very computer-intensive process (being an order of magnitude more costly than a prestack migration) and that, like migration, it requires fairly accurate knowledge of the very low-wavenumber velocity model (so that the kinematics of events is approximately correct). Future work is required to test the algorithm more thoroughly under different circumstances, such as for the case of significant multiples, for the case of inversions of Rayleigh waves and refractions for the near-surface, and for the case of transmission data (VSPs and well-to-well surveys). Further tests are required to study the effect of the choice of model parameters on resolution and either to invent a better preconditioning operator to speed convergence and regain the poorly resolved low wavenumbers, or introduce an additional low-wavenumber (model) inversion step.

ACKNOWLEDGMENTS

I acknowledge the support of the sponsors of the Stanford Exploration Project and Jon Claerbout. Also, I acknowledge the CNRS (C_2VR) who supplied some CRAY time. Thanks to Albert Tarantola for many stimulating discussions and making possible an eight-month stay at the Institut de Physique du Globe where this work was initiated. Particular thanks to Alexandre Nercessian for priceless computer aid. Thanks to Odile Gauthier and Antonio Pica who are working on the acoustic equivalent of the algorithm presented and helped me to understand many aspects. Thanks to Jean Remy for computer assistance and to Jean Virieux for help in linking to the CRAY. Finally, thanks to my fellow students at Stanford, particularly John Toldi, Dan Rothman, Kamal Al-Yahya, and Jos Van Trier, for provocative and interesting discussion.

REFERENCES

Aki, K., and Richards, P., 1980, Quantitative seismology: Freeman Press.

Backus, G. E., and Gilbert, J. F., 1967, Numerical application of a formalism for geophysical inverse problems: Geophys. J. Roy. Astr. Soc., **13**, 247–276.

——— 1968, The resolving power of gross earth data: Geophys. J. Roy. Astr. Soc., **16**, 169–205.

——— 1970, Uniqueness in the inversion of gross earth data: Phil. Trans. Roy. Soc. London, Ser. A., **266**, 123–192.

Cerjan, C., Kosloff, D., Kosloff, R., and Reshef, M., 1985, A nonreflecting boundary condition for discrete acoustic and elastic wave equations: Geophysics, **50**, 705–708.

Claerbout, J. F., 1976, Fundamentals of geophysical data processing: McGraw-Hill Book Co.

Clayton, R. W., and Stolt, R. H., 1981, A Born-WKBJ inversion method for acoustic reflection data: Geophysics, **46**, 1559–1567.

Cohen, J. K., and Bleistein, N., 1979, Velocity inversion procedures of acoustic waves: Geophysics, **44**, 1077–1087.

Devaney, A. J., 1984, Geophysical diffraction tomography: Inst. Electr. and Electron. Eng., Trans., Geosci. and Remote Sensing, **GE-22**, 3–13.

Fletcher, R., and Reeves, C. M., 1964, Function minimization by conjugate gradients: The Computer Journal, **7**, 149–154.

Gauthier, O., Virieux, J., and Tarantola, A., 1986, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results: Geophysics, **51**, 1387–1403.

Harlan, W. S., 1986, Signal noise separation and seismic inversion: Ph.D. thesis, Stanford Univ.

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

Kolb, P., and Canadas, G., 1986, Least-squares inversion of prestack data: Simultaneous identification of density and velocity: Presented at the 16th conf. on mathematical geophysics, Oosterbeek, The Netherlands.

Kosloff, D., Reshef, M., and Loewenthal, D., 1984, Elastic wave calculations by the Fourier method: Bull., Seis. Soc. Am., **74**, 875–891.

Lailly, P., 1984, Migration methods: partial but efficient solutions to the seismic inverse problem, *in* Santosa, F., Pao, Y. H., Symes, W., and Holland, Ch., Eds.: Inverse problems of acoustic and elastic waves, Soc. Industr. Appl. Math.

Luenberger, D., 1984, Linear and nonlinear programming, 2nd edition: Addison-Wesley.

McAulay, A. D., 1985, Prestack inversion with plane-layer point source modeling: Geophysics, **50**, 77–89.

Menke, W., 1984, Geophysical data analysis: discrete inverse theory: Academic Press, Inc.

Mora, P., 1986, Elastic finite differences with convolutional operators: Tech. Rep., Stanford Expl. Proj., rep. no. 48, Stanford Univ.

——— 1987a, Elastic inversion of multioffset seismic amplitude data for *P*-wave velocity, *S*-wave velocity and density, *in* Danbom, S., and Domenico, N., Eds.: Shear wave exploration: Soc. Explor. Geophys.

——— 1987b, Elastic wave-field inversion of reflection and transmission data: Geophysics, submitted.

Polak, E., and Ribiére, G., 1969, Notes sur la convergence de méthodes de directions conjugées: Revue Fr. Inf. Rech. Oper, **16-R1**, 35–43.
Powell, M. J. D., 1981, Approximation theory and methods: Cambridge Univ. Press.
Rothman, D. H., 1985, Nonlinear inversion, statistical mechanics, and residual statics estimation: Geophysics, **50**, 2784–2796.
Stolt, R. H., and Weglein, A. B., 1985, Migration and inversion of seismic data: Geophysics, **50**, 2458–2472.
Tarantola, A., and Valette, B., 1982, Inverse problems = quest for information: J. Geophys., **50**, 159–170.
Tarantola, A., 1984, The seismic reflection inverse problem, *in* Santosa, F., Pao, Y. H., Symes, W., and Holland, Ch., Eds., Inverse problems of acoustic and elastic waves: Soc. Industr. Appl. Math.
Tarantola, A., Fosse, I., Mendes, M., and Nercessian, A., 1985, The choice of model parameters for linearized elastic inversion of seismic reflection data: Tech. Rep., Equipe de Tomographie du Geophys. rep. 2, Inst. de Paris du Globe.