Geophys. J. R. astr. Soc. (1986) **87**, 701–730

Iterative inversion for velocity using waveform data

Robert L. Nowack *Purdue University, Department of Geosciences, West Lafayette, IN 47907, USA*

Keiiti Aki *Department of Geological Sciences, University of Southern California, University Park, CA 90089-0741, USA*

Accepted 1986 March 25. Received 1986 March 25; in original form 1986 January 24

Summary. In this paper, velocity inversion using waveform data is investigated. A linearized approach is used in which a linear sensitivity operator must be derived. This operator can be computed economically using reciprocity of the Green's function. In order to avoid a large matrix inversion, several descent algorithms are described. Data errors and *a priori* model information are incorporated using covariance operators. A fast and reasonably accurate forward modelling scheme is required and here the Gaussian beam method for a slowly varying heterogeneous medium is used. Several types of linearizations can be done including the Born approximation, a linearization in terms of the field, and the Rytov approximation, a linearization in terms of the log field. Field linearizations are expected to be useful for small-scale heterogeneities which result in scattering effects that are additive in the field. For small perturbations from a homogeneous background, a linearized inversion in terms of the field is equivalent to a sequence of Kirchhoff migrations. Log field linearizations may be more robust for large-scale heterogeneities where forward scattering predominates, but phase unwrapping may be difficult numerically. Several numerical examples using a field linearization are performed in which transmitted body waves through a model with small velocity variations are used. The results using the waveform data identify the trial structures and are comparable or slightly better than the travel-time inversion results.

Key words: seismology, inversion

Introduction

In this paper, an iterative inversion method for each structure is investigated. Previous inversions for velocity using array data have mostly used travel times in constructing the tomography problem (see Aki 1977; Aki, Christoffersson & Husebye 1977). Recently

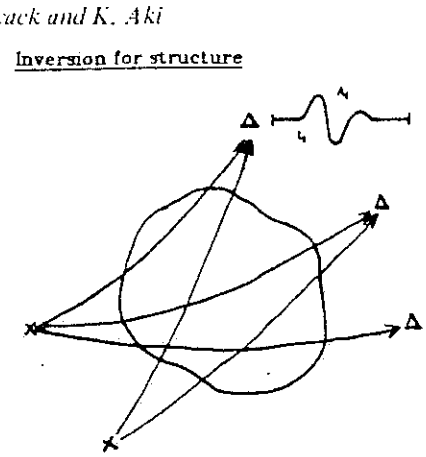

Figure 1. Sketch of heterogeneous body with surrounding sources and receivers. A simple transmitted waveform is shown.

iterative methods, such as the algebraic reconstruction method, have been applied to travel time inversions (see McMechan 1983; Clayton & Comer 1983; Humphreys, Clayton & Hager 1984). Several of these iterative algorithms are described by Censor (1981). The iterative methods have advantages over generalized inversion methods when dealing with large sparse matrices, but can have convergence problems when small eigenvalues are present (Ivansson 1983). Still, there is much more information in seismic data than just first arrival travel times, including amplitudes and waveforms. Fig. 1 shows a sketch of a heterogeneous region with surrounding sources and receivers. A simple transmitted waveform is shown.

A number of studies have incorporated amplitude and waveform data in seismic inversion and interpretation. For example, spatial variations of phase time and amplitude from teleseismic body waves recorded at the Montana LASA were interpreted by Larner (1970) in terms of a dented Moho. Haddon & Husebye (1978) and Thomson & Gubbins (1982) derived joint interpretations of travel time and amplitude to infer the structure beneath NORSAR. Waveform inversion using refraction data that assume a vertically varying structure has been performed by Shaw (1983). Brown (1984) made a comparison of travel-time inversions and waveform inversions using synthetic data also in a vertically varying structure. Migration of reflection seismic data is a form of inversion (see Schneider 1978), and it has been recently shown to be kinematically similar to a linearized Born inversion (see Tarantola 1984b, Miller, Oristaglio & Beylkin 1984). Direct inversions based on the linearized model have been investigated by Cohen & Bleistein (1979) and Raz (1981). There has also been some progress in generalizing exact 1-D inversion methods to higher dimensions (see Newton 1983).

A recent formulation by Tarantola (1984a) uses an iterative linearized approach and is described in the next section. A linear sensitivity operator must be derived, and this can be done economically by using reciprocity of the Green's function. Several descent algorithms are described which avoid a large matrix inversion. A fast and reasonably accurate forward modelling scheme is required, and here we make use of the Gaussian beam method in a slowly varying medium. Data errors and *a priori* information are incorporated using covariance operators. Different types of linearization are possible including linearization in terms of the field, the Born approximation, and a linearization in terms of the log of the field, the Rytov approximation. Their relative merits are discussed. Finally several numerical examples are performed using a linearization in terms of the field and transmitted waveforms in order to test the method.

Linearized inversion

In this section a review of the linearized inverse approach is given following Madden (unpublished lecture notes), Tarantola (1984a, b, c) and Lailly (1983). In the linearized approach, a linear sensitivity or Frechet derivative operator must be derived first and then used in an iterative procedure for the model parameters. In addition, several descent algorithms are described which do not require the inversion of a large matrix.

The forward problem can be written in the form

$$L[p(\mathbf{r}, t)] = S(\mathbf{r}, t), \tag{1}$$

where L is a model dependent differential operator, $p(\mathbf{r}, t)$ is the field variable, and $S(\mathbf{r}, t)$ is the source term. For example, for the scalar wave equation

$$L = \left[\frac{1}{v(\mathbf{r})^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right] . \tag{2}$$

The solution of (1) can be written

$$p(\mathbf{r}, t) = \int dV(\mathbf{r}')g(\mathbf{r}', t; \mathbf{r}) * S(\mathbf{r}', t) \quad + \text{(boundary terms)},$$

where $g(\mathbf{r}', t, \mathbf{r})$ is the Green's function from $\mathbf{r}$ to $\mathbf{r}'$, and * is a time convolution. This can be derived from the bilinear identity (see Lanczos 1961). For the scalar wave equation, the Kirchhoff integral is the relevant boundary term, in the following, homogeneous boundary conditions will be assumed giving zero for the boundary term. Finally, for a wave equation operator L with homogeneous boundary conditions, the Green's function is reciprocal with respect to source and receiver location. Thus,

$$g(\mathbf{r}, t; \mathbf{r}_g) = g(\mathbf{r}_g, t; \mathbf{r}).$$

In order to obtain a linear sensitivity operator, a perturbed problem is constructed

$$(L + \delta L)(p + \delta p) = S + \delta S \tag{3}$$

or

$$L(\delta p) = -\delta L(p + \delta p) + \delta S.$$

Assuming that δL (δp) is small is equivalent to the Born approximation which requires that the perturbed field, δp, be much smaller than the unperturbed primary field, p. To first order, (3) can be written

$$L(\delta p) \approx -\delta L(p) + \delta S, \tag{4}$$

where $-\delta L(p)$ is an equivalent source term for the medium perturbations and δS is the term for the source perturbations. Here we will investigate the 'inverse medium problem' for a given source, thus $\delta S = 0$. The solution for the perturbed field can then be written as a space integration over secondary equivalent sources

$$\delta p(\mathbf{r}_g, t; \mathbf{r}_s) = -\int dV(\mathbf{r})g(\mathbf{r}, t; \mathbf{r}_g) * \delta L\,[p(\mathbf{r}, t; \mathbf{r}_s)], \tag{5}$$

where δL includes the model perturbations, and p is the incident primary field. Equation (5) can be written as

$$\delta p(\mathbf{r}_g, t; \mathbf{r}_s) = \frac{\partial p}{\partial L}\delta L = F\,\delta L, \tag{6}$$

where F is a linear sensitivity or Frechet derivative operator, and the operator δL has imbedded within it the model variations. Assuming that the primary field can be written as $p(\mathbf{r}, t; \mathbf{r}_s) = g(\mathbf{r}, t; \mathbf{r}_s)$, then

$$\delta p(\mathbf{r}_g, t; \mathbf{r}_s) = -\int dV(\mathbf{r}) g(\mathbf{r}, t; \mathbf{r}_g) * \delta L[g(\mathbf{r}, t; \mathbf{r}_s)].$$

Thus, the linear sensitivity operator can be constructed by computing $g(\mathbf{r}, t; \mathbf{r}_s)$ which propagates the field from the source $\mathbf{r}_s$ to each model point, $\mathbf{r}$, and $g(\mathbf{r}, t; \mathbf{r}_g)$ which back-propagates the field back from the geophones into the model at $\mathbf{r}$. This requires the computation of $N_g + N_s$ forward problems evaluated at the interior points of the model where N_g is the number of geophones and N_s is the number of sources.

For the scalar wave equation from equations (2) and (5), we have

$$\delta p(\mathbf{r}_g, t; \mathbf{r}_s) = \int dV(\mathbf{r}) \left\{ \frac{2}{v^3(\mathbf{r})} \ddot{p}(\mathbf{r}, t; \mathbf{r}_s) * g(\mathbf{r}, t; \mathbf{r}_g) \right\} \delta v(\mathbf{r}), \tag{7}$$

where the term in the brackets is the linear sensitivity operator, $F = \partial p / \partial v$, $\delta v(\mathbf{r})$ is the velocity model perturbation at $\mathbf{r}$, and $\ddot{p}(\mathbf{r}, t; \mathbf{r}_s)$ is the second time derivative of the primary field computed from the source point to the point $\mathbf{r}$ in the medium. The linear sensitivity operators for the elastic wave equation are given by Tarantola (1984c). The linear sensitivity operators for the acoustic wave equation are given by Tarantola (1984a) and used in a simple 1-D example in Appendix C.

In the following, the linearized problem (5) will be imbedded within an iterative procedure. The effect of ignoring the δL (δp) term in the linearization will result in no multiple interactions with the perturbations within δL, or no 'cross talk' between perturbations during each iteration. Multiple interactions from previous iterations are included via the Green's functions. This approach requires the calculation of Green's functions, which are recalculated at each iteration. In principle, an iterative strategy should be able to handle large velocity contrasts. This approach requires the calculation of Green's function in inhomogeneous media inexpensively and reasonably accurately. Various forward modelling approaches could be considered including ray theoretical methods, the discrete wavenumber method, or the finite difference and finite element methods. Any inversion method is only as good as the forward modelling scheme on which it is based.

The forward problem as a function of model parameters is in general a nonlinear function and can be written

$$\mathbf{p} = f(\mathbf{m}), \tag{8}$$

where $\mathbf{p}$ is the field and $\mathbf{m}$ is the model. In general, $\mathbf{p}_{obs} \neq f(\mathbf{m}_{prior})$, where $\mathbf{p}_{obs}$ is the observed field and $\mathbf{m}_{prior}$ is the *a priori* model based on previous information. The objective is to find the combined vector $[\mathbf{p}, \mathbf{m}]$ that satisfies $\mathbf{p} = f(\mathbf{m})$ and is closest to $[\mathbf{p}_{obs}, \mathbf{m}_{prior}]$ in some sense. This distance could be defined in various ways depending on the noise structure of the problem, including the L_2 norm, the L_1 norm, or the L_∞ norm. Here the weighted L_2 norm will be considered with its associated induced inner product. This results in a least-squares formulation. Thus we want to minimize the functional (see Tarantola & Valette 1982; Tarantola 1984a)

$$S(\mathbf{m}) = \tfrac{1}{2} \{ |\mathbf{p}_{obs} - f(\mathbf{m})|^2 + |\mathbf{m} - \mathbf{m}_{prior}|^2 \} , \tag{9}$$

where the 1/2 has been introduced for later convenience. The linear sensitivity operator,

F_k at $\mathbf{m}_k$ then satisfies

$$f(\mathbf{m}_k + \delta\mathbf{m}) = f(\mathbf{m}_k) + F_k\delta\mathbf{m}_k + O(|\delta\mathbf{m}_k|^2).$$

The negative of the gradient of $S(\mathbf{m}_k)$ gives the local direction of maximum descent at $\mathbf{m}_k$. To first order this can be written

$$\gamma_k = -\nabla S(\mathbf{m}_k) = \{F_k^*\,\delta\mathbf{p}_k - (\mathbf{m}_k - \mathbf{m}_{\text{prior}})\}, \tag{10}$$

where $\delta\mathbf{p}_k = \mathbf{p}_{\text{obs}} - f(\mathbf{m}_k)$, and F_k^* is the adjoint of the linear sensitivity operator, F_k. The adjoint is defined from the bilinear identity

$$<\delta\mathbf{p}, F\delta\mathbf{m}>_{\delta p} - <F^*\delta\mathbf{p}, \mathbf{m}>_{\delta m} = \text{boundary term}, \tag{11}$$

where $<,>$ is a defined inner product. The boundary term is assumed to be zero using homogeneous boundary conditions. Thus, F is a linear mapping from model to data space, and the adjoint, F^*, is a linear mapping from data space back to model space. Inverse operators are in general difficult to construct, but adjoint operators through the bilinear identity are straightforward to obtain. In addition, adjoint operators have very useful properties which aid in the construction of inverse and generalized inverse operators.

Consider a linear problem, $F\mathbf{m} \approx \mathbf{p}_{\text{obs}}$, with noisy observations but where the null space of F is zero, $N(F) = \mathbf{0}$. The null space, $N(F)$, represents the subspace of the model space that satisfies $F\mathbf{m} = \mathbf{0}$. The standard least-squares solution results by using the adjoint theorem, $R^{\perp}(F) = N(F^*)$, where R is the range of $F\mathbf{m}$ in data space. Thus mapping $F\mathbf{m}$ and $\mathbf{p}_{\text{obs}}$ through F^* annihilates the component of $\mathbf{p}_{\text{obs}}$ in $R^{\perp}(F)$ resulting in a consistent set of equations. Then, $F^*F\mathbf{m} = F^*\mathbf{p}_{\text{obs}}$ or $\mathbf{m} = (F^*F)^{-1}F^*\mathbf{p}_{\text{obs}}$. For $N(F) = \mathbf{0}$, (F^*F) is full rank and an inverse exists. The standard least-squares solutions thus results from two operations on the data, $\mathbf{p}_{\text{obs}}$. First, the data is operated on by the adjoint operator, F^*, which projects $\mathbf{p}_{\text{obs}}$ from data space to model space. This blurred image in model space is then filtered with the operator $(F^*F)^{-1}$. When $N(F) \neq \mathbf{0}$, a generalized inverse can be computed by using additional adjoint theorems. This is equivalent to the Lanczos formulation which results in a minimum–norm least-squares solution (Aki & Richards 1980). An alternative is to stabilize the (F^*F) operator using a maximum likelihood procedure which reduces to minimizing a functional like (9) with appropriate data and model covariance operators defined. Any *a priori* knowledge about the component of the model in the null space can be incorporated into the final solution. This might include smoothness of the resulting solution.

The maximum likelihood algorithm of Tarantola & Valette (1982) is a linear iterative procedure for solving the non-linear inverse problem, (8), by minimizing the functional (9). It can be written

$$\mathbf{m}_{k+1} = \mathbf{m}_k + [H_k]^{-1}\,\{F_k^*\delta\mathbf{p} - (\mathbf{m}_k - \mathbf{m}_{\text{prior}})\}\,, \tag{12}$$

where the term in the brackets is the negative gradient of the functional $S(\mathbf{m})$, and

$$H_k = [I + F_k^*F_k]\,. \tag{13}$$

Covariance operators can be introduced by a suitable definition of the data and model space scalar products (see Tarantola 1984c). Thus

$$<\mathbf{m}_1, \mathbf{m}_2>_m = \mathbf{m}_1^T C_m^{-1}\,\mathbf{m}_2$$

$$<\mathbf{p}_1, \mathbf{p}_2>_p = \mathbf{p}_1^T C_p^{-1}\,\mathbf{p}_2\,,$$

where $\mathbf{m}^T$ is the transpose of $\mathbf{m}$, C_m is the model covariance operator, and C_p is the data

covariance operator. From this definition of the scalar products, the adjoint operator, F_k^*, of F_k is related to the transpose by

$$F_k^* = C_m F_k^T C_p^{-1}. \tag{14}$$

The adjoint is equal to the transpose when $C_m = I$ and $C_p = I$. For a complex operator, F, the transpose is replaced by the conjugate transpose operator.

A major problem with the iterative scheme (12) is that in discretized form it requires the inversion of a possibly large matrix. Since the goal is to iteratively solve the original problem, (8), only an approximate solution of the imbedded linearized problem may be required. Thus, a simple iterative scheme could be written

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k W_k \gamma_k, \tag{15}$$

where γ_k is the negative gradient of the functional $S(\mathbf{m})$, W_k is a preconditioner, and α_k is a scalar that approximately solves a 1-D line search of $S(\mathbf{m})$ in the direction $W_k\gamma_k$. α_k is thus chosen to minimize $S(\mathbf{m}_k + \alpha_k W_k \gamma_k)$ with respect to α_k giving

$$\alpha_k = \frac{<W_k\gamma_k, \gamma_k>}{<W_k\gamma_k, H_k W_k \gamma_k>}, \tag{16}$$

where H_k is given in (13). For the maximum likelihood algorithm, (11), $W_k = H_k^{-1}$ giving $\alpha_k = 1$. This would then be similar to a modified Newton's method for solving the non-linear problem, (8), dependent on the choice of the covariance operators used.

Various simplifying choices for W_k could be considered. Choosing $W_k = I$ results in the steepest descent method. Thus,

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \gamma_k, \tag{17}$$

where

$$\alpha_k = \frac{<\gamma_k, \gamma_k>}{<\gamma_k, H_k\gamma_k>},$$

with H_k given in (13). For a quadratic $S(m)$, γ_{k+1} and γ_k will be perpendicular. One thus moves in the direction γ_k until tangent to a level curve of $S(\mathbf{m})$. At this point, γ_{k+1} is chosen perpendicular to the level curve in a direction of maximum descent.

Another possible choice for W_k is (see Tarantola 1984a)

$$W_k = [\mathrm{DIAG}\,(H_k)]^{-1}. \tag{18}$$

This is similar to a modified Jacobi method with suitably defined covariance operators.

In any of the preconditioning strategies, the single step convergence properties for quadratic functionals, $S(\mathbf{m})$, are governed by the difference in the smallest and largest eigenvalues of $W_k H_k$ with H_k given in (13). From Luenberger (1984),

$$S(\mathbf{m}_{k+1}) \leq \left[\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}}\right]^2 S(\mathbf{m}_k).$$

The closer W_k is to H_k^{-1} the better the single step convergence. A general overall strategy is to construct a preconditioning scheme that is easy to compute and possesses a favourable eigenvalue structure at each step.

Instead of using the gradient directions in the steepest descent approach, improved convergence for very little extra effort can be obtained by using the so called conjugate gradient

directions defined with respect to a new inner product, $\langle x, x\rangle = x^T H_k x$ where H_k is given in (13). The Fletcher-Reeves implementation of the conjugate gradient method to non-quadratic problems is briefly outlined. First, given $\mathbf{m}_{prior}$ and $\gamma_0 = -\nabla S(\mathbf{m}_{prior})$, set $\mathbf{d}_0 = \gamma_0$. Now let $\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \mathbf{d}_k$ where α_k minimizes $S(\mathbf{m}_k + \alpha_k \mathbf{d}_k)$. The next direction of descent, $\mathbf{d}_{k+1}$ is chosen to be H_k – orthogonal to the previous direction $\mathbf{d}_k$. This is done by setting $\mathbf{d}_{k+1} = \gamma_{k+1} + \beta_k \mathbf{d}_k$, where $\beta_k = \langle \gamma_{k+1}, \gamma_{k+1}\rangle / \langle \gamma_k, \gamma_k\rangle$. $\mathbf{d}_{k+1}$ is thus a direction modified from the steepest descent direction. In a similar fashion to steepest descent, $\mathbf{d}_{k+1}$ and $\mathbf{d}_k$ are orthogonal but with respect to H_k. The conjugate gradient method has the important property that the estimate $\mathbf{m}_k$ minimizes a quadratic functional $S(m)$ over a subspace spanned by all the previous directions $(\mathbf{d}_0, \ldots, \mathbf{d}_{k-1})$ (see Luenberger 1984).

All of the above methods require the linear sensitivity operator F_k and the adjoint of this operator F_k^*. For example, for the scalar wave equation (2)

$$F_k = \left\{ \frac{2}{v^3(r')} \ddot{p}(r', t, r_s) * g(r', t; r_g) \right\} = \frac{\partial p}{\partial v}, \tag{19}$$

where velocity v is the model parameter. Thus to first order $\delta p = (\partial p/\partial v)\delta v$. Using the bilinear identity, (11), with $C_m = I$ and $C_p = I$ and homogeneous boundary conditions, then $F^* = F^T$ and

$$\langle \delta p, F\delta v\rangle_p = \sum_{r_s}\sum_{r_g} \int dt\, \delta p F(\delta v)$$

$$= \sum_{r_s}\sum_{r_g} \int dt\, \delta p(r_g, t; r_s) \int dV(r') \left[\frac{2}{v^3(r')} \ddot{p}(r', t; r_s) * g(r', t; r_g) \right] \delta v(r').$$

Then,

$$\langle F^* \delta p, \delta v\rangle_v = \int dV(r')\, F^*(\delta p)\, \delta v$$

$$= \int dV(r') \left[\sum_{r_s}\sum_{r_g} \int dt \left[\frac{2}{v^3(r')} \ddot{p}(r', t; r_s) * g(r', t; r_g) \right] \delta p(r_g, t; r_s) \right] \delta v(r').$$

Thus F_k^* is just F_k with summations now over geophone and source locations and an integration in time.

The final adjoint operator includes the data and model covariance operators. Assuming uncorrelated errors over source and receiver locations and time, then,

$$C_p(\mathbf{r}_g, t; \mathbf{r}_s | \mathbf{r}_g', t; \mathbf{r}_s') = \sigma_{gs}^2 \delta_{gg'} \delta_{ss'} \delta_{tt'}, \tag{20}$$

where σ_{gs} represents the estimated error in the trace corresponding to the gth receiver and the sth source. For the model covariance operator, a spatial Gaussian correlation is a commonly used choice (see Aki & Richards 1980; Tarantola 1984a). Thus,

$$C_v(\mathbf{r}, \mathbf{r}') = (2\pi)^{-3/2} \frac{\sigma_v^2}{L^3} \exp\left[\frac{-1}{2} \frac{(\mathbf{r} - \mathbf{r}')^2}{\Delta^2} \right] \tag{21}$$

where v is the model parameter $\mathbf{m}$, L is the medium correlation length for the Gaussian correlation function, and σ_v represents the estimated departures of $\mathbf{m}(\mathbf{r})$ from $\mathbf{m}_{prior}(\mathbf{r})$. As L goes to zero, then $C_v(\mathbf{r}, \mathbf{r}') = \sigma_v^2 \delta(\mathbf{r} - \mathbf{r}')$. The model covariance operator will constrain the iterated model parameters as well as act like a spatial smoothing filter. The complete adjoint can be written, $F_k^* = C_v F_k^T C_p^{-1}$, as in (14).

For the scalar wave equation, the preconditioned descent method can be written

$$v_{k+1} = v_k + \alpha_k W_k \left[C_v F^T \delta \mathbf{p}' - (v_k - v_{\text{prior}})\right]. \tag{22}$$

where

$$\delta \mathbf{p}' = C_p^{-1} \delta \mathbf{p}$$

are the weighted data residuals. Now let

$$\delta \hat{v} = C_v F_k^T \delta \mathbf{p}'; \tag{23}$$

Then,

$$\delta \hat{v} = C_v \sum_{r_g} \sum_{r_s} \int dt \left[\frac{2}{v^3(\mathbf{r})} \ddot{p}(\mathbf{r}, t; \mathbf{r}_s) * g(\mathbf{r}, t; r_g)\right] \delta p'(\mathbf{r}_g, t; \mathbf{r}_s). \tag{24}$$

This can be rewritten as

$$\delta \hat{v}(\mathbf{r}) = C_v \sum_{r_s} \sum_{r_g} \int dt \frac{2}{v^3(\mathbf{r})} \ddot{p}(\mathbf{r}, t; \mathbf{r}_s) \left[g(\mathbf{r}, -t; \mathbf{r}_g) * \delta p'(\mathbf{r}_g, t, \mathbf{r}_s)\right], \tag{25}$$

where the term in the brackets corresponds to the propagation of the weighted residuals from the geophones into the model backwards in time. This is then compared with the computed primary field from the sources to the model point. A summation over all source and receiver locations is then performed and an integration over time. Finally, the *a priori* model covariance operator is applied. This is reminiscent of an imaging principle in which the reflector (velocity perturbation) exists where the downgoing and upgoing waves coincide in time (Claerbout 1976; Tarantola 1984b).

In order to implement this algorithm, $p(\mathbf{r}, t; \mathbf{r}_s)$ and $g(\mathbf{r}, t; \mathbf{r}_g)$ must be computed where $\mathbf{r}$ is a point in the model (see Fig. 2). Since for an inhomogeneous medium an inexpensive forward modelling scheme is required, one possibility is that paraxial ray theory or the Gaussian beam method be used to calculate p and g (see Červený, Popov & Pšenčík 1982; Nowack & Aki 1984). An advantage of using these methods is that no two-point ray tracing is required. In addition, the Gaussian beam method produces a smoothed field with no unphysical singularities in amplitude resulting from caustics which may have adverse effects on an inversion.

Using the ray approximation in 3-D,

$$g(\mathbf{r}, t; r_g) \approx A_{rr_g} \delta(t - T_{rr_g})$$

$$\ddot{p}(\mathbf{r}, t; \mathbf{r}_s) \approx A_{rr_s} \ddot{S}(t - T_{rr_s}),$$

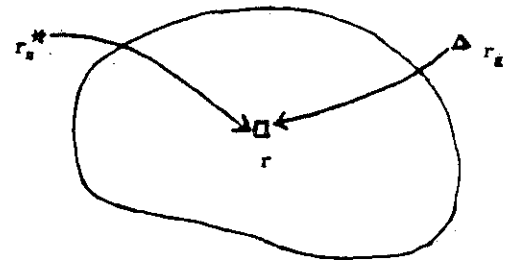

Figure 2. The computation of the linear sensitivity operator requires a forward problem for each source and receiver evaluated at the interior points of the model.

where A_{rr_i} and T_{rr_i} are the amplitude and travel time computed by ray theoretical methods and $\ddot{S}(t)$ is the second derivative of the source time pulse. For the 2-D case, an additional $\pi/4$ phase shift in the far field must be included. Equation (24) is then

$$\delta\hat{v}_k = C_v F_k^T \delta p' = C_v \sum_{r_s} \sum_{r_g} \frac{2}{v^3(\mathrm{r})} A_{rr_s} A_{rr_g} \int dt\, \ddot{S}(t - T_{rr_s}) \delta p'(r_g, t + T_{rr_g}; r_s) \tag{26}$$

This can be rewitten as

$$\delta\hat{v}_k = C_v \sum_{r_s} \sum_{r_g} \frac{2}{v^3(\mathrm{r})} A_{rr_s} A_{rr_g} \delta\tilde{p}_k(r_g, t = T_{rr_g} + T_{rr_s}; r_s), \tag{27}$$

where,

$$\delta\tilde{p}_k = \ddot{S}(-t) * \delta p'_k(r_g, t; r_s): \tag{28}$$

$\delta\tilde{p}$ is the cross correlation of the weighted residuals with the second derivative of the source time pulse. $\delta\hat{v}_k$ is the result of the weighted sum over the phase surface through the data defined by $t = T_{rr_g} + T_{rr_s}$ for each model point, r.

When the *a priori* starting model is close enough to the true velocity model so that the Born approximation is strictly valid, then $F_k^* = F_0^*$ for all iterations. In addition, if the *a priori* model is homogeneous, then

$$\delta\hat{v}_k = C_v F_0^T \delta p' = C_v \sum_{r_s} \sum_{r_g} \frac{1}{v^3 8\pi^2 R_{rr_s} R_{rr_g}} \delta\tilde{p}\left(r_g, t = \frac{R_{rr_s} + R_{rr_g}}{v}; r_s\right), \tag{29}$$

where $\delta\tilde{p}$ is defined above, R_{rr_s} is the distance between source and model point, and R_{rr_g} is the distance between the geophone and the model point. This is similar to a Kirchhoff migration and a cross correlation with the source time pulse (Schneider 1978; Tarantola 1984b). Thus, assuming the Born approximation is strictly valid, then a velocity inversion using a descent approach is equivalent to an iterative sequence of Kirchhoff migrations.

A complete iterative step involves a preconditioner, W_k, and a scalar α_k, which solves a 1-D line search. W_k is assumed to be some approximation to $(I + F_k^* F_k)^{-1}$. Approximating F_k by a matrix, then

$$F = \{B_{lj}\}$$

where

$$B_{lj} = \left\{\frac{2}{v^3(r_j)} \ddot{p}(r_j, t; r_s) * g(r_j, t; r_g)\right\},$$

with $l = 1, N_p$ where N_p is the number of data values $= N_t^* N_s^* N_g$, and $j = 1, N_v$ where N_v is the number of velocity model values. Assuming

$$F_k^* = \frac{\sigma_v^2}{\sigma_p^2} F_k^T$$

with $C_p = \sigma_p^2 I$ and $C_v = \sigma_v^2 I$, then

$$(H_k)_{ij} = (I + F_k^* F_k)_{ij} = \left\{\delta_{ij} + \frac{\sigma_v^2}{\sigma_p^2} \sum_{l=1}^{N_p} B_{li} B_{lj}\right\}.$$

Thus,

$$(I + F_k^* F_k)_{ij} = \delta_{ij} + \sum_{r_s} \sum_{r_g} \int dt \; \frac{4\sigma_v^2}{\sigma_p^2 v^3(r_i)\, v^3(r_j)} \left\{ \ddot{p}(r_i, t; r_s) * g(r_i, t; r_g) \right\} \left\{ \ddot{p}(r_j, t; r_s) * g(r_j, t; r_g) \right\}. \tag{30}$$

In the ray approximation this can be written

$$(I + F_k^* F_k)_{ij} = \delta_{ij} + \frac{4\sigma_v^2}{\sigma_p^2 v^3(r_i)\, v^3(r_j)} \sum_{r_s} \sum_{r_g} A_{r_i r_s} A_{r_i r_g} A_{r_j r_s} A_{r_j r_g} \, \tilde{S}(t = T_{r_i r_s} + T_{r_i r_g} - T_{r_j r_s} - T_{r_j r_g}), \tag{31}$$

where,

$$\tilde{S}(t) = \ddot{S}(-t) * \ddot{S}(t).$$

With a high-frequency source, this matrix will be sparse. For a diagonally dominant operator, H_k, then an improved single-step convergence can be obtained using $W_k = [\mathrm{DIAG}(H_k)]^{-1}$, where

$$\mathrm{DIAG}\,(H_k)_{ii} = \left\{ 1 + \frac{4\sigma_v^2 \tilde{S}(0)}{\sigma_p^2 v^6(r_i)} \sum_{r_s} \sum_{r_g} A_{r_i r_s}^2 A_{r_i r_g}^2 \right\}. \tag{32}$$

Thus, a simple preconditioned iterative scheme can be written

$$v_{k+1}(r_i) = v_k(r_i) + \alpha_k \left\{ 1 + \frac{4\sigma_v^2 \tilde{S}(0)}{\sigma_p^2 v^6(r_i)} \sum_{r_s} \sum_{r_g} A_{r_i r_s}^2 A_{r_i r_g}^2 \right\}^{-1} \times C_v \sum_{r_s} \sum_{r_g} \frac{2}{v^3(\mathbf{r})} A_{r_i r_s} A_{r_i r_g} \int dt \, \ddot{S}(t - T_{r_i r_s}) \, \delta p'(r_g, t + T_{r_i r_s}; r_s). \tag{33}$$

When the *a priori* model is homogeneous, the starting model is sufficiently close to the true model so that the Born approximation is strictly valid, and with $\sigma_v \to \infty$, then

$$v_{k+1}(r_i) = v_k(r_i) + \alpha_k 8\pi^2 v^3(r_i) \left\{ \frac{C_v}{\sigma_v^2} \right\} \left[\sum_{r_s} \sum_{r_g} \left\{ \frac{1}{R_{r_i r_s}^2} \frac{1}{R_{r_i r_g}^2} \right\} \right]^{-1} \times \sum_{r_s} \sum_{r_g} \left\{ \frac{1}{R_{r_i r_s}} \frac{1}{R_{r_r r_g}} \right\} \frac{\sigma_p^2}{\tilde{S}(0)} \, \delta\tilde{p}\left(r_g, t = \frac{R_{r_i r_s} + R_{r_i r_g}}{v}; r_s \right), \tag{34}$$

where $\{C_v/\sigma_v^2\}$ includes the spatial filtering part of the Covariance operator, $\delta\tilde{p}$ is given in equation (28) and is normalized by $\tilde{S}(0)/\sigma_p^2$, and α_k is obtained from equation (16). The factor

$$\left[\sum_{r_s} \sum_{r_g} \left\{ \frac{1}{R_{r_i r_s}^2} \frac{1}{R_{r_i r_g}^2} \right\} \right]^{-1}$$

approximately corrects for geometric spreading. A simple variant of this formula will be used in the examples to follow.

Alternative linearizations

The formulation in the previous section was applied to the linearization of the field variables. General limitations to this approach are summarized in Appendix A. A simple example of the limitations in using a linearization in terms of the field can be seen as follows. Consider a field in an unperturbed homogeneous medium (see Brown 1984),

$$u_0(r, t) = \frac{\delta(t - R/v_0)}{4\pi R}$$

where $R = |r - r_0|$. In a slightly perturbed homogeneous model with velocity $v_1 = v_0 + \delta v$, the field is

$$u_1(r, t) = \frac{\delta(t - R/v_1)}{4\pi R}.$$

The perturbed-minus-unperturbed field is then approximated by the linearisation

$$u_1 - u_0 = \delta u \approx \frac{\partial u}{\partial v}\delta v \tag{34}$$

or

$$\frac{\delta(t - R/v_1)}{4\pi R} - \frac{\delta(t - R/v_0)}{4\pi R} \approx \frac{\partial}{\partial v}\left[\frac{\delta(t - R/v_0)}{4\pi R}\right]\delta v$$

which reduces to

$$\delta(t - R/v_1) - \delta(t - R/v_0) \approx \delta(t - R/v_0)\left(\frac{R}{v_0^2}\right)\delta v.$$

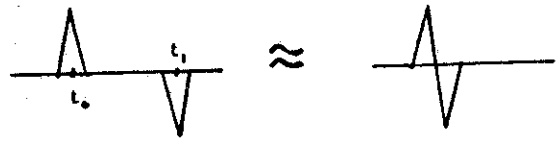

Figure 3. A linearization in terms of the field approximates the differential seismogram on the left with the doublet on the right. This will be a good approximation only for low enough frequencies.

Thus the actual differential seismogram is approximated by a constant times the derivative of a delta function. As seen from Fig. 3, depending on the frequency content, this will be a poor approximation when

$$\delta t = \left(\frac{R}{v_1} - \frac{R}{v_0}\right)$$

is large. For a single harmonic frequency in a homogeneous medium, then

$$u = \frac{\exp(i\omega R/v)}{4\pi R}.$$

The linearization in terms of the field can then be written

$$\frac{\exp(i\omega R/v_1)}{4\pi R} - \frac{\exp(i\omega R/v_0)}{4\pi R} \approx \frac{\partial u}{\partial v}\delta v = \left(\frac{-i\omega R}{v_0^2}\right)\frac{\exp(i\omega R/v_0)}{4\pi R}\delta v$$

or

$$\exp\left[i\omega R\left(\frac{1}{v_1} - \frac{1}{v_0}\right)\right] - 1 \approx \left\{\frac{-i\omega R}{v_0^2}\right\}\delta v,$$

for small

$$\omega R\left\{\frac{1}{v_1} - \frac{1}{v_0}\right\} \ll 1,$$

then the exponential can be approximated as

$$\exp(i\omega\delta t) \approx 1 + i\omega R\left\{\frac{1}{v_1} - \frac{1}{v_0}\right\}.$$

Thus for sufficiently low frequencies, such that $\omega\delta t \ll 1$, the linearization will be a good approximation to the actual differential field.

An alternative linearization could be performed in terms of the log of the field variables. This has been successfully applied to electromagnetic problems by Madden (private communication). By perturbing the velocity in a homogeneous medium, we again consider the extreme case of $ka \gg 1$. With

$$\ln(u_0) = \ln\left(\frac{\exp(i\omega R/v_0)}{4\pi R}\right),$$

then

$$\ln(u_1) - \ln(u_0) \approx \frac{\partial[\ln(u)]}{\partial(\ln v)}\delta(\ln v),$$

or,

$$i\omega R\left\{\frac{1}{v_1} - \frac{1}{v_0}\right\} \approx i\omega R\left(\frac{-1}{v_0^2}\right)(v_0)\,\delta(\ln v)$$

$$\left\{\frac{1}{v_1} - \frac{1}{v_0}\right\} \approx \frac{-1}{v_0^2}(v_1 - v_0).$$

Thus for this simple example of a homogeneous medium perturbation there is no dependence on frequency for the complex phase linearization. The region of validity for the log field parameterization is more involved in the general case where there is an amplitude perturbation as well. In any event, this type of linearization has been used successfully applied in travel time inversions (see Aki *et al.* 1977; Clayton & Comer 1983), where only the imaginary part of the complex phase is used.

The linear sensitivity operator in terms of the log field is now derived and compared with the Born linearization. The log field linearization is additive in the phase, and is equivalent to

Rytov's method of smooth perturbations (see Tatarskii 1971, section 45; Ishimaru 1978). Rytov's approximation is compared in Appendix 2 with other asymptotic, multiplicative approximations including geometric optics and the parabolic equation method. We start with the Helmholz equation

$$\nabla^2 u + \frac{\omega^2}{v^2(\mathbf{x})} u = 0$$

and let $\psi = \ln [u(\mathbf{x}, \omega)]$, then the following non-linear Riccati equation is obtained

$$\nabla^2 \psi + (\nabla\psi)^2 + \frac{\omega^2}{v(\mathbf{x})^2} = 0.$$

Now a perturbed problem is derived with $\psi = \psi_0 + \epsilon\psi_1$ and $v = v_0(\mathbf{x}) + \epsilon v_1(\mathbf{x})$, assuming $\epsilon \ll 1$, $\psi_0 \sim \psi_1$, and $v_0 \sim v_1$. Then

$$\left[\nabla^2\psi_0 + (\nabla\psi_0)^2 + \frac{\omega^2}{v_0^2(\mathbf{x})}\right] + \epsilon\left[\nabla^2\psi_1 + 2\nabla\psi_0\nabla\psi_1 - \frac{2\omega^2 v_1(\mathbf{x})}{v_0^3(\mathbf{x})}\right]$$

$$+ \epsilon^2\left[(\nabla\psi_1)^2 + \frac{3\omega^2 v_1^2(\mathbf{x})}{v_0^4(\mathbf{x})}\right] = 0.$$

The first bracket is set to zero by solving the forward problem in the unperturbed medium, $v_0(\mathbf{x})$. Dropping the third bracket in ϵ^2 constitutes the Rytov approximation. This equation is now linear in ψ_1 and v_1. Letting $\delta v(\mathbf{x}) = \epsilon v_1$ and $\delta\psi = \epsilon\psi_1$ gives

$$\nabla^2(\delta\psi) + 2\nabla\psi_0\nabla(\delta\psi) = \frac{2\omega^2\delta v(\mathbf{x})}{v_0^3(\mathbf{x})}.$$

This is a linear equation for $\delta\psi$ which can be simplified by the substitution, $\delta\psi = \bar{\psi}\exp(-\psi_0)$,

$$\nabla^2\bar{\psi} + \frac{\omega^2}{v_0^2(\mathbf{x})}\bar{\psi} = \frac{2\omega^2\delta v}{v_0^3}\exp(\psi_0)$$

which has a solution

$$\bar{\psi}(r_g, r_s) = -\int dV(r')\, g(r_g, r')\exp[\psi_0(r', r_s)]\frac{2\omega^2\delta v}{v_0^3}$$

or

$$\delta\psi(r_g, r_s) = -\int dV(r')\, g(r_g, r')\exp[\psi_0(r', r_s) - \psi_0(r_g, r_s)]\frac{2\omega^2\delta v}{v_0^3}. \tag{35}$$

This gives the linear sensitivity operator in terms of the log field variables in the form

$$\delta\psi = \frac{\partial\psi}{\partial v}\delta v.$$

To see how this compares to the Born approximation or a linearization in terms of the field, equation (35) is exponentiated

$$\exp(\delta\psi) = \exp(\psi - \psi_0) = \exp\left[-\int dV(r')(\ldots)\right].$$

For small $-\int dV(r')(\ldots)$,

$$\exp(\psi) \approx \exp(\psi_0) - \int dV(r') p(r', r_s) g(r_g, r') \frac{2\omega^2 \delta v(r')}{v_0^3(r')},$$

which is equivalent to the frequency transformed equation (7). Thus equation (7) can be derived from (35) assuming that the exponentiated argument is small. It appears that the Rytov approximation includes some multiple scattering effects not included in the Born approximation. However, the Rytov approximation was originally derived assuming $\delta\psi \ll 1$. Thus, there has been ongoing controversy as to whether the Rytov approximation is in fact any better than the Born approximation (see Tatarskii 1971; Barabanenko *et al.* 1971; Aki 1973; Brown 1966, 1967; Taylor 1967; Yura *et al.* 1983; Devaney 1984). In fact, Aki & Richards (1980) derive their expressions for the log amplitude and phase from the exponentiated Born series (EBS).

However, asymptotic equivalence may not correspond to a practical equivalence between the two approaches as shown in the previous simple example. Keller (1969) showed that the Born and the Rytov approximations are equally accurate in their dependence on the small velocity parameter, ϵ, but show quite different asymptotic properties as a function of range. Mueller, Kaveh & Wade (1979) noted that both the Born and the Rytov approximations required that some norm of the velocity perturbations over the volume be small compared to unity. In addition, the Born approximation required small ka, where a is the scale of the heterogeneity, in agreement with the range of validity for the single scattering formulation given in Appendix A. Thus low spatial wavenumber, band limited velocity perturbations appear to favour the Rytov approximation. In the context of an inverse problem, it is natural to estimate the low spatial wavenumbers of the velocity first via a multiple scale type of procedure. For small-scale residual heterogeneities, the Born approximation to the field may have an advantage.

Experiments in diffraction tomography comparing the Born and the Rytov formulations have been conducted by Kaveh *et al.* (1982, 1984) who showed that when a large forward scattering component exists, the linearized reconstructed images based on the Born approximation were more in error than the Rytov, but the Rytov required the determination of the phase of the scattered waves. There are numerical difficulties in obtaining the log of the field since the phase of the complex log must be unwrapped. Still, numerical algorithms for this exist (see Tribolet 1978). Also, algorithms based on the smoothness of the phase at adjacent spatial points have had some initial success (see Kaveh *et al.* 1982, 1984).

Another consideration as to whether to use a linearization based on the Born (field) or the Rytov (log of the field) approximations is the noise structure of the problem (see Tatarskii 1971). If the data are expressed as a sum of signals plus independent Gaussian noise, N, then the linearized inversion should be applied to the waveform or to the real and imaginary Fourier components. If the noise structure is multiplicative and proportional to the signal, then $\log N$ will behave in a Gaussian manner around $\log u$. In this case, the log should be taken (see Aki & Richards 1980, p. 639). Detailed work on the effects of additive and multiplicative noise for moment tensor inversion of surface wave data was performed by Patton & Aki (1979).

If the noise is Gaussian and additive in the field, then the fluctuations in the amplitude should follow a Rayleigh distribution. If the noise is Gaussian and additive in the complex phase, then the fluctuations in the amplitude should follow a lognormal distribution. Experiments of light propagation in the atmosphere show closer agreement with the log-normal distribution (see Tatarskii 1971).

In order to investigate the linearization in terms of the log field for a large-scale heterogeneity, a homogeneous velocity perturbation, δv, is again considered. Equation (35) can then be written

$$\delta\psi(r_g, r_s) = \frac{-2\omega^2\delta v}{v_0^3}\int dV(r')\ \frac{\exp(ikR_{r_g r'})}{4\pi R_{r_g r'}}\ \frac{\exp(ikR_{r' r_s})}{4\pi R_{r' r_s}}\, 4\pi R_{r' r_s}\exp(-ikR_{r_g r_s}).$$

Using the Fresnel approximation with the z-axis oriented from the source to receiver

$$\delta\psi(r_g, r_s) = \frac{-\omega^2\delta v z}{2\pi v_0^3}\int_0^z \frac{dz'}{(z-z')z'}\int_{-\infty}^{\infty} dx' \exp\left[ikzx'^2/2z'(z-z')\right]\int_{-\infty}^{\infty} dy'$$

$$\times \exp\left[ikzy'^2/2z'(z-z')\right].$$

Using the relation

$$\int_{-\infty}^{\infty}\exp(-ax^2) = (\pi/a)^{1/2}, \text{ then}$$

$$\delta\psi(r_g, r_s) = \frac{-\omega^2\delta v z}{2\pi v_0^3}\int_0^z \frac{dz'}{z'(z-z')}\,\frac{2\pi z'(z-z')}{-ikz}$$

or

$$\delta\psi(r_g, r_s) = \frac{-i\omega z\delta v}{v_0^2}.$$

Now let

$$\delta\psi = i\omega\delta t = i\omega z\left\{\frac{1}{v_1} - \frac{1}{v_0}\right\} \approx \frac{-i\omega z\delta v}{v_0^2}.$$

Thus, the linearization in terms of log variables will retrieve a global velocity change in one iteration without the apparent low frequency requirement of the Born approximation of equation (7).

There are several basic linearizations that can be used in inversion studies, including those based on the additive nature of the field, Born, or the log field, Rytov. Travel-time inversions are more similar to the Rytov approximation except that only the imaginary part of the complex phase is utilized and usually only first arrival times. Keller (1969) noted that any advantages that the Rytov approximation may have over the Born approximation may be lost when the waveform contains more than one wave. This would occur when there is high wavenumber velocity fluctuations causing significant backscatter. This may be the reason for the success of simple migration algorithms which map small-scale reflectors and which are kinematically equivalent to Born inversion (see Miller *et al.* 1984). Another consideration as to whether to use the log field data, or the amplitude and phase, is the numerical difficulties in unwrapping the phase when taking the complex log, although work on this has been done (Tribolet 1978; Kaveh *et al.* 1982, 1984).

One strategy for inversion of material parameters is to use a multiple scale procedure in which the large-scale fluctuations are inverted first, and compensated for. This might include several Rytov iterations followed by a series of Born iterations. A compromise approach would be to use an iterative travel-time inversion first to adjust for the major phase

differences and then use the iterative Born inversion on the waveforms to incorporate more of the data and fine tune the model. In the following numerical examples, the Born linearization is used on residual waveform data assuming small velocity perturbations.

Examples

In this section, several examples are given using a linearization in terms of the field as described previously. A simple descent method is used

$$v_{k+1} = v_k + \alpha W_k (F_k^* \delta p_k) \tag{36}$$

where $F_k^* = C_v F_k^T C_p^{-1}$, W_k is a preconditioner, and α is a scaling factor. For an initial homogeneous model close to the true model, then $F_k^* \approx F_0^*$. Operating on the residuals by the adjoint is then given in (29) as

$$\delta \hat{v}_k = F_k^* \delta p_k \approx \frac{C_v}{v^3 8\pi^2} \sum_{r_s} \sum_{r_g} \left[\frac{1}{R_{rr_s} R_{rr_g}} \right] \delta \tilde{p} \left(r_g, t = \frac{R_{rr_s} + R_{rr_g}}{v}, r_s \right),$$

where $\delta \tilde{p}$ is the cross-correlation of the second derivative of the source time function with the weighted data residuals, C_v is the model covariance operator and includes a spatial smoothing function. Here, a Gaussian smoothing function as in (21) is used. Operating on the residuals by the adjoint is kinematically similar to a Kirchhoff migration. If the preconditioner, W_k, is chosen as I, then the velocity inversion in (36) with $F_k^* = F_0^*$ is equivalent to a sequence of Kirchhoff migrations. An alternative choice for the preconditioner, W_k, is the inverse of the diagonal of $(I + F^* F)$ as in (32). The part of this preconditioner that approximately corrects for geometric spreading is

$$W_k = \left\{ \sum_{r_s} \sum_{r_g} \left[\frac{1}{R_{rr_s}^2 R_{rr_g}^2} \right] \right\}^{-1}. \tag{37}$$

This simple preconditioner will be used in addition to the identity in the next example.

The geometry of the first example is shown in Fig. 4. There are nine sources marked by x's and 19 receivers marked by triangles. The axes are in km to give specific units and the

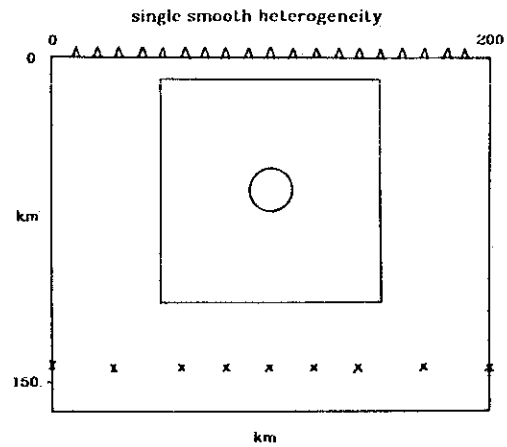

Figure 4. Model geometry for the single smooth heterogeneity numerical example.

initial velocity is 8 km s^{-1}. For this example, a single smooth heterogeneity with a radius of 10 km is reduced in velocity by -0.01 km s^{-1}. This is shown by the circle in Fig. 4. The interior box in Fig. 4 shows the region where the velocity is to be inferred.

Fig. 5 shows the seismograms computed using the Gaussian beam method for the single smooth heterogeneity shown in Fig. 4 for three of the nine source locations. The source wavelet is the Gabor wavelet

$$f(t) = \exp\left[-(2\pi f_0 t/\gamma)^2\right] \cos\left(2\pi f_0 t + \phi_0\right),$$

where $f_0 = 4$ Hz, $\gamma = 3, \phi_0 = 0$. This results in a 2 km wavelength. The correction given by Červený *et al.* (1982) is used to approximately compute the 3-D response. In Fig. 5, a slightly larger amplitude can be seen on some seismograms for different source locations. For example, for $S = (30., 140.)$, the amplitude increases due to focusing at a range of 180 km and a time of 24 s.

The residual, perturbed minus unperturbed, seismograms are shown in Fig. 6 for three source locations. The amplitudes of the residuals are about 10 times smaller than the

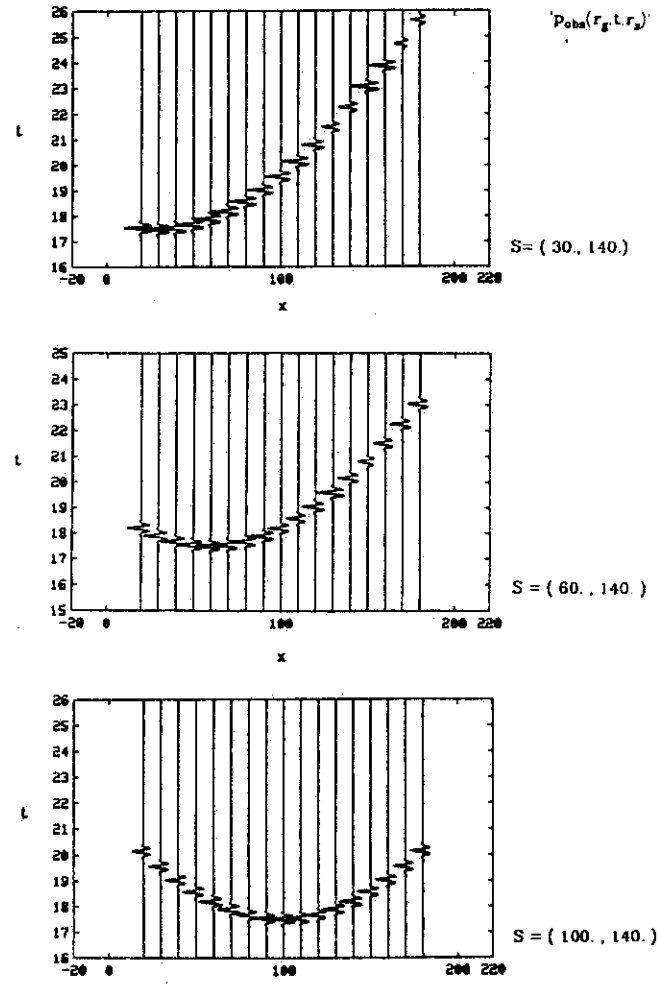

Figure 5. Gaussian beam seismograms for three source locations for a single smooth heterogeneity.

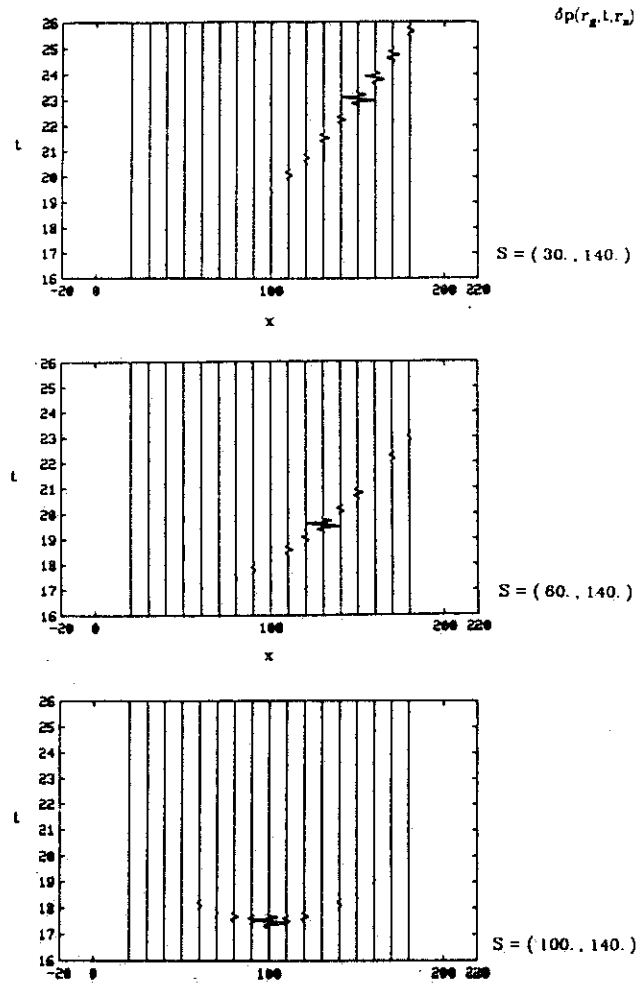

Figure 6. Residual seismograms for three source locations for a single smooth heterogeneity.

amplitudes of the unperturbed seismograms. This satisfies the Born requirement that the scattered field be small compared to the primary field. The wavelet cross correlation with the data results in a symmetric wavelet centred on the arrival time, and the second derivative approximately results in a factor, $-\omega_0^2$, assuming a Gabor wavelet.

Fig. 7 shows the result of back propagating the residuals into model space using the adjoint operator. This would be the first iteration in a descent procedure with $W_k = I$ and $C_v = \sigma_v \delta(\mathbf{r} - \mathbf{r}')$. A grid of 100 by 100 points has been evaluated in model space. The result in Fig. 7 is scaled from -1 to $+1$. The contours are at $(-0.75, -0.25, 0.25, 0.75)$ with the -0.75 contour in the centre. The parallelogram and the small plus represent the location of the true heterogeneity. The data residuals appear to be streaked along lines from the receivers to the stations in model space. The response to the simple transpose operator for a single heterogeneity thus has a streaked appearance in model space.

There are additional operations that can be included in an individual velocity inversion step. This includes a more involved preconditioner as in equation (37) which approximately corrects for geometric spreading, and is an approximation to filtering the backpropagation.

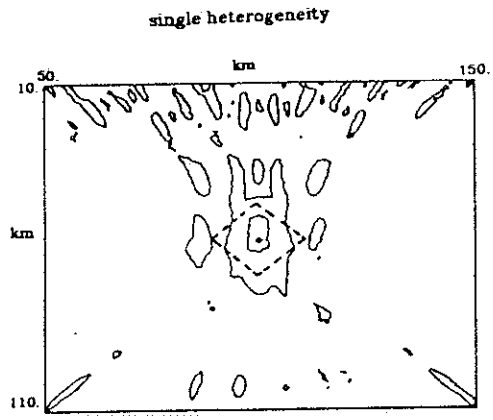

Figure 7. The backpropagation of the residuals into model space for a single smooth heterogeneity.

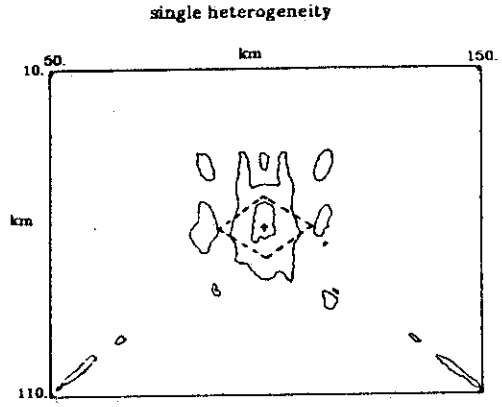

Figure 8. The first iteration of a velocity inversion for a single smooth heterogeneity with W_k given by equation (37) and $C_v = \sigma_v \delta(\mathbf{r} - \mathbf{r}')$.

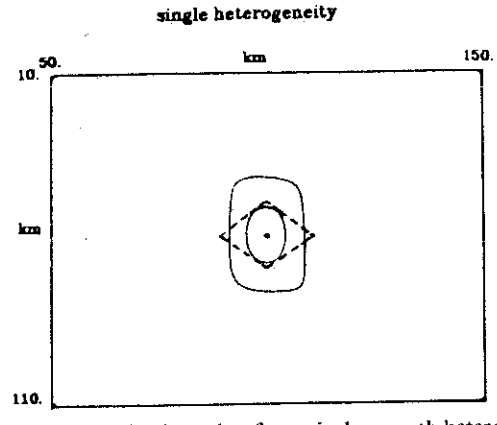

Figure 9. The first iteration of a velocity inversion for a single smooth heterogeneity with $W_k = I$ and C_v given by a Gaussian function with a 5 km radius.

Fig. 8 shows the result of using this preconditioner. It has the effect of reducing some of the streaks in the backpropagation. Again, the true heterogeneity is represented by the parallelogram.

Another operation which could be applied to the backpropagation is spatial smoothing by the model covariance operator based on *a priori* information. Fig. 9 shows the result of smoothing the backpropagation by a Gaussian smoothing filter with a 5 km radius. This filtering operation eliminates the streaks and localizes the heterogeneity. The derived heterogeneity is elongated in the z direction as compared to the true heterogeneity represented by the parallelogram. This results from the source and receiver geometry with sources below and receivers above the heterogeneous region.

To check the sign of the derived heterogeneity, the example was done with a positive instead of negative velocity heterogeneity. This resulted in the inferred velocity also being positive. A final step in an iteration is the choice of the scalar, α, which specifies how far in a particular descent direction one moves. A particular value for the scaling factor can be computed from equation (16), although other values might be used to over or under relax the problem. The scale factor for each iteration must be chosen small enough to ensure convergence of the process.

In the next example, three heterogeneities are used, each with a velocity heterogeneity of -0.01 km s^{-1} lower than the background. The geometry is shown in Fig. 10 with the sources below the heterogeneity and the stations above. An iterative procedure is then used to reconstruct the velocity structure. The scale factor here was chosen as a certain fraction of the value used to match the single heterogeneity case.

For each iteration, the residuals were backpropagated into model space and then smoothed using a 4 km smoothing filter in order to reduce the streaked character of the backpropagation. To avoid any broadening of the main anomalies at each iteration due to smoothing, only values greater than 50 per cent of the maximum for that iteration were multipled by the scale factor and added to the cumulative anomaly from previous iterations.

An example using a scale factor of 0.50 times the factor used to match the single heterogeneity case is shown in Fig. 11. The contour interval is scaled to go from -1.0 to $+1.0$. In the first iteration shown in Fig. 11(a), the lower heterogeneity is evident and the upper anomalies are emerging. Fig. 11(b) shows the fifth iteration in which the upper heterogeneities are more well defined. The eighth iteration is shown in Fig. 11(c) where the lower

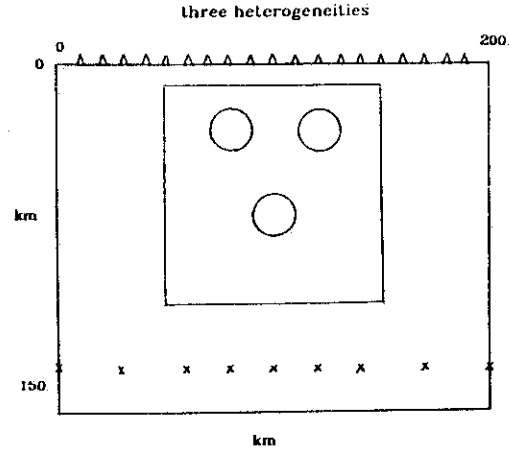

Figure 10. Model geometry for a numerical example with three smooth heterogeneities.

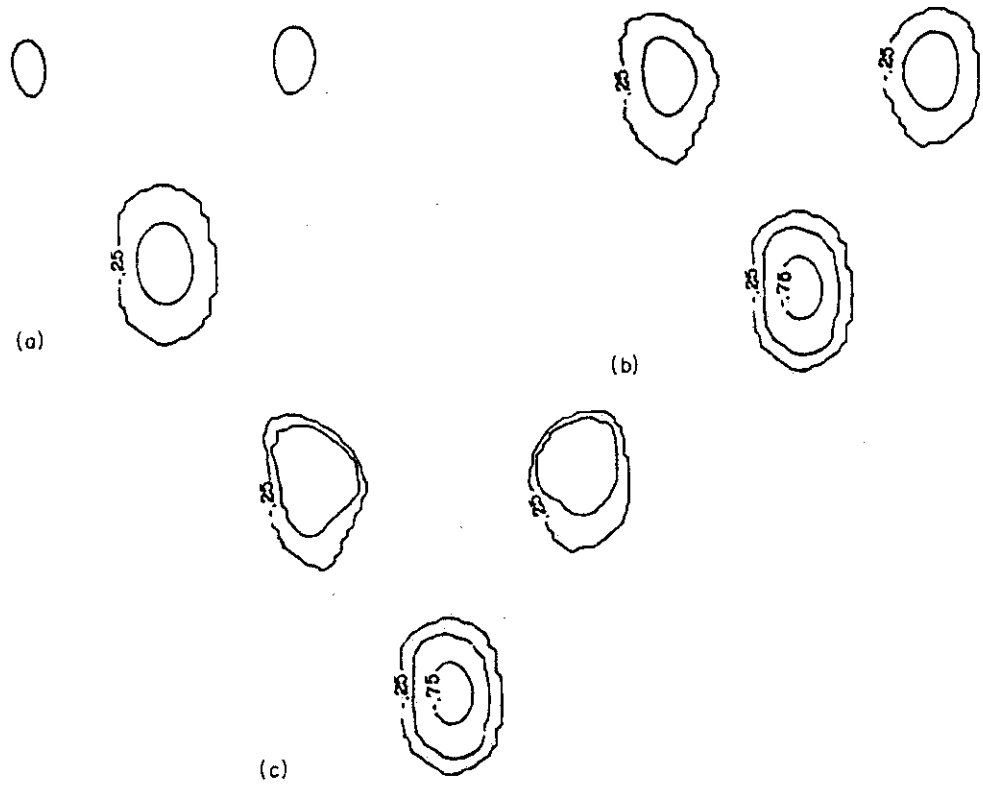

Figure 11. Iterative inversion results for three smooth heterogeneities with $W_k = I$ and C_v given by a Gaussian function with a 4 km radius. (a) First iteration; (b) fifth iteration; (c) eighth iteration.

heterogeneity is still slightly more compensated for than the upper heterogeneities. All three heterogeneities are also elongated in the vertical direction. This results from the geometry of the sources and receivers. In Fig. 12, the observed-minus-predicted waveforms are shown for the first, fifth and eight iterations for the source coordinate at (100., 140.). The root mean square error was decreased by 49 per cent at the fifth iteration and by 60 per cent at the eighth iteration.

Other scale factors were attempted in the iteration process. Smaller values for the scale factor resulted in more iterations. For values much larger than 0.5, the scale factor for the single heterogeneity case, divergent results occurred. Although this steepest descent procedure is straightforward, faster convergence might be obtained from the more elaborate conjugate gradient methods.

Finally, a simple comparison with travel-time inversions was made using the same geometries. The model was divided into 10 by 10 blocks or 100 model parameters. One hundred and seventy-one travel-time values were used, not all of which provided new information about the model. The results for the single heterogeneity are shown in Fig. 13(a). This result compares well with the waveform inversion shown in Fig. 9. The travel-time inversion for three heterogeneities is shown in Fig. 13(b) and this compares with the waveform results shown in Fig. 11(c). The travel-time results appear to be slightly less localized than the waveform results. Thus, in this case, the waveform results are comparable or better. As discussed in the previous section, log field linearizations may be slightly more robust over field linearizations in certain cases. Travel time inversions are in the log field

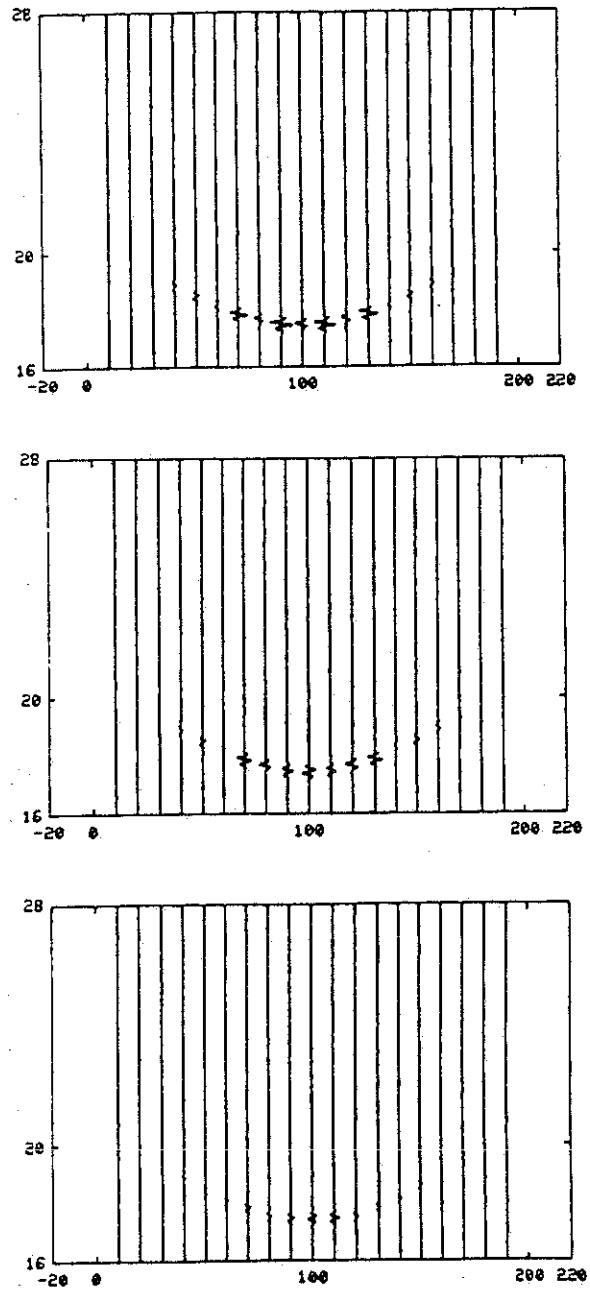

Figure 12. Observed minus predicted residuals with a source at (100.,140.) for (a) first iteration; (b) fifth iteration; (c) eighth iteration.

class in which only part of the data, the imaginary part of the complex phase, derived from the first arrivals is used. However, for small percent velocity perturbations, the field and log field linearization results are expected to be comparable. In actual cases, waveform inversions will allow for a more complete use of the data when the phase perturbations are small, or have been previously compensated for.

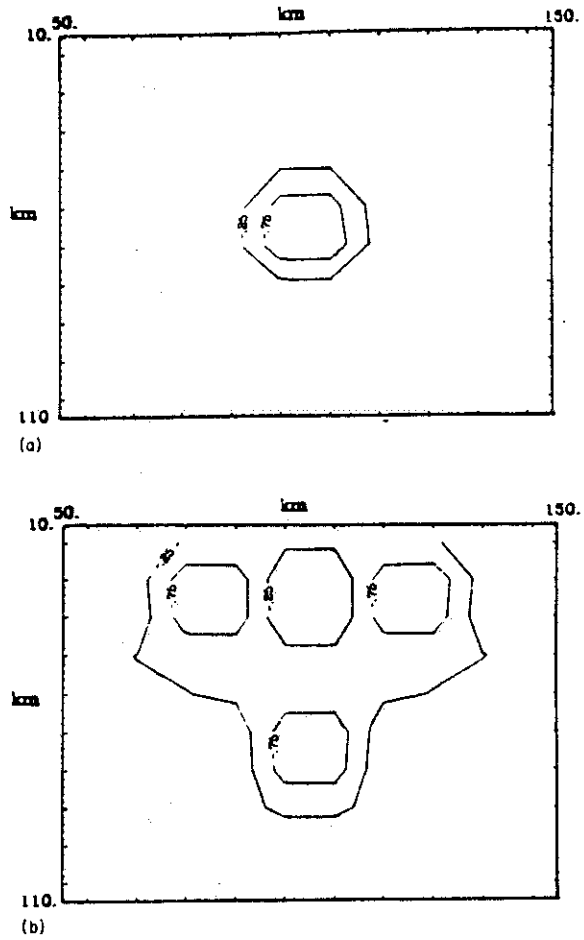

Figure 13. Travel-time inversions. (a) Single heterogeneity; (b) three heterogeneities.

Conclusions

In this paper, a linearized inversion procedure for material parameters has been investigated. In this procedure a linear sensitivity operator must be derived and can be computed economically by using reciprocity of the Green's function. Data errors and *a priori* model information are included via covariance operators. Large matrix inversions are avoided by using descent algorithms. A fast forward modelling scheme is required and here the Gaussian beam method for a laterally varying medium is used. This allows for inhomogeneous initial models.

Different types of linearizations can be performed including the Born approximation, a linearization in terms of the field, and the Rytov approximation, a linearization in terms of the log field. Travel-time inversions are in the class of linearizations for the log field, where only the imaginary part of the complex phase is used with first arrivals. Log field linearizations may be more robust than field linearizations for large-scale heterogeneities where forward scattering predominates, but phase unwrapping may be difficult numerically. The linearization in terms of the field for small perturbations from a homogeneous back-

ground is kinematically equivalent to a sequence of Kirchhoff migrations. Field linearizations are expected to be useful for small-scale heterogeneities which result in scattering effects that are additive in the field. Several numerical examples using a field linearization are performed in which transmitted body waves through a model with small velocity variations are used. The results using the waveform data identify the trial structures, and are comparable or slightly better than the travel-time inversion results.

Acknowledgments

The authors wish to thank A. Tarantola for sending us preprints. Thanks are also given to T. Madden and R. Wu for informative discussions. This work was supported by the Defense Advanced Research Projects Agency under grant F49620-82-K-0004 monitored by the Air Force Office of Scientific Research, and by Schlumberger-Doll research funds.

References

Aki, K., 1977. Three dimensional seismic velocity anomalies in the lithosphere, *J. Geophys.*, 43, 235–242.

Aki, K., 1973. Scattering of *P* waves under the Montana Lasa, *J. geophys. Res.*, 78, 1334–1346.

Aki, K., Christoffersson, A. & Husebye, E. S., 1977. Determination of the three-dimensional seismic structure of the lithosphere, *J. geosphys. Res.*, 82, 277–296.

Aki, K. & Richards, P. G., 1980. *Quantitative Seismology: Theory and Methods*, W. H. Freeman, San Francisco.

Barabanenkov, Yu.N., Kravtsov, Yu.A., Rytov, S. M. & Tamarskii, V. I., 1971. Status of the theory of propagation of waves in a randomly inhomogeneous medium, *Soviet Phys. Usp.*, 13, 551–575.

Brown, G. B., 1984. Linearized travel time, intensity, and waveform inversion in the ocean sound channel – a comparison, *J. acoust. Soc. Am.*, 75, 1451–1461.

Brown, W. P., 1966. Validity of the Rytov approximation in optical propagation calculations, *J. opt. Soc. Am.*, 56, 1045–1052.

Brown, W. P., 1967. Validity of the Rytov approximation, *J. opt. Soc. Am.*, 57, 1539–1543.

Censor, Y., 1981. Row-action methods for huge sparse systems and their applications, *SIAM Rev.*, 23, 444–466.

Červený, V., Popov, M. M. & Pšenčík, I., 1982. Computation of wave fields in inhomogeneous media – Gaussian beam approach, *Geophys. J. R. astr. Soc.*, 70, 109–128.

Clayton, R. & Comer, R., 1983. A tomographic analysis of mantle heterogeneities from body wave travel times, *EOS Trans. Am. geophys. Un.*, 64, 776.

Claerbout, J. F., 1976. *Fundamentals of Geophysical Data Processing*, McGraw-Hill, New York.

Cohen, J. K. & Bleistein, N., 1979. Velocity inversion procedure of acoustic waves, *Geophysics*, 36, 467–481.

Devaney, H. J., 1984. Geophysical diffraction tomography, *IEEE Trans. Geosci. Remote Sensing, Ge-22*, No. 1.

Haddon, R. A. W. & Husebye, E. S., 1978. Joint interpretation of *P*-wave time and amplitude anomalies in terms of lithospheric heterogeneities, *Geophys. J. R. astr. Soc.*, 55, 19–43.

Humphreys, E., Clayton, R. & Hager, B. H., 1984. A tomographic image of the mantle structure beneath southern California, *Geophys. Res. Lett.*, 11, 625–627.

Ishimaru, A., 1978. *Wave Propagation and Scattering in Random Media*, Academic Press, New York.

Ivansonn, S., 1983. Remark on an earlier proposed iterative tomographic algorithm, *Geophys. J. R. astr. Soc.*, 75, 855–860.

Kaveh, M., Soumekh, M. & Greenleaf, J. F., 1984. Signal processing for diffraction tomography, *IEEE Trans. Sonics Ultrasonics, SU-31*, 230–239.

Kaveh, M., Soumekh, M., Lu, Z. Q., Mueller, R. K. & Greenleaf, J. F., 1982. Further results on diffraction tomography using Rytov's approximation, in *Acoustical Imaging*, 12, Plenum Press, New York.

Keller, J. B., 1969. Accuracy and validity of the Born and Rytov approximations, *J. opt. Soc. Am.*, 59, 1003–1004.

Lailly, P., 1983. The seismic inverse problem as a sequence of before stack migrations, in *Conference on Inverse Scattering: Theory and Applications*, eds Bednar *et al.*, SIAM, Philadelphia, 206–220.

Lanczos, C., 1961. *Linear Differential Operators*, Van Nostrand-Reinhold, Princeton, New Jersey.

Larner, K. L., 1970. Near-receiver scattering of teleseismic body waves in layered crust–mantle models having irregular interfaces, *PhD thesis*, MIT.

Luenberger, D. G., 1984. *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, London.

McMechan, G. A., 1983. Seismic tomography in boreholes, *Geophys. J. R. astr. Soc.*, 74, 601–612.

Miller, D., Oristaglio, M. & Beylkin, G., 1984. A new formalism and an old heuristic for seismic migration, *SEG* (extended abstract).

Mueller, R. K., Kaveh, M. & Wade, G., 1979. Reconstructive tomography and application to ultrasonics, *Proc. IEEE*, 67, 567–587.

Newton, R. G., 1983. The Marchenko and Gelfand–Levitan methods in the inverse scattering problem in one and three dimensions, in *Con. Inverse Scattering: Theory and Applications*, pp. 1–74, eds Bednar *et al.*, SIAM, Philadelphia.

Nowack, R. & Aki, K., 1984. The 2-D Gaussian beam synthetic method: testing and application, *J. geophys. Res.*, 89, 7797–7819.

Patton, H. & Aki, K., 1979. Bias in the estimate of seismic moment tensor by the linear inversion method, *Geophys. J. R. astr. Soc.*, 59, 479–495.

Raz, S., 1981. Three-dimensional velocity profile inversion from finite offset scattered data, *Geophysics*, 46, 837–842.

Schneider, W. S., 1978. Integral formulation for migration in two and three dimensions, *Geophysics*, 43, 49–76.

Shaw, P., 1983. Waveform inversion of explosion data, *PhD thesis*, University of California, San Diego.

Tarantola, A., 1984a. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, 49, 1259–1266.

Tarantola, A., 1984b. Linearized problem for an homogeneous reference model, *Geophys Prospect*, 32, 998–1015.

Tarantola, A., 1984c. The seismic reflection inverse problem, in *Inverse Problems for Acoustic and Elastic Waves*, ed. Santosa, F., Yih-Hsing, P., Symes, W. W. & Holland, Ch., SIAM, Philadelphia.

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

Tatarskii, V. I., 1971. *The Effects of the Turbulent Atmosphere on Wave Propagation*, (trans. from Russian by Israel Program for Scientific Translations Ltd).

Taylor, L. S., 1967. On Rytov's method, *Radio Sci.*, 2, 437–441.

Thomson, C. J. & Gubbins, D., 1982. Three-dimensional lithosphere modelling at NORSAR: linearity of the method and amplitude variations of the anomalies, *Geophys. J. R. astr. Soc.*, 71, 1–36.

Tribolet, J. M., 1977. A new phase unwrapping algorithm, *IEEE Trans. acoust. Speech Signal Processing*, *ASSP-25*, 170–177.

Yura, H. T., Sung, C. C., Clifford, S. F. & Hill, R. J., 1983. Second-order Rytov approximation, *J. opt. Soc. Am.*, 73, 500–502.

Appendix A

In this appendix a brief review is given for the conditions of validity of the Born approximation in a continuous fluctuating medium characterized by a spatial Gaussian correlation function where

$$N(r) = \frac{E[\mu(r)\mu(r\ r\ .}{E[\mu^2]} = \exp(-|r|^2/a^2)$$

where $\mu(r) = -\delta v(r)/v(r)$, and $E[\ldots]$ represents the expectation over an ensemble of random media, and a is the correlation length. If the process is ergodic, then this expectation can be replaced by a spatial average over one realization. The fractional loss of energy to the

primary wave for a medium with velocities with a Gaussian correlation function is (see Aki & Richards 1980)

$$\frac{\Delta I}{I} = \sqrt{\pi} E[\mu^2]\, kakL\, [1 - \exp(-k^2a^2)].$$

For the Born approximation to be valid then $\frac{\Delta I}{I} \ll 1$, or

$$\sqrt{\pi} E[\mu^2] kakL \ll 1 \quad ka \gg 1$$

$$\sqrt{\pi} E[\mu^2] k^3a^3kL \ll 1 \quad ka \ll 1$$

where L is the propagation distance and a is the correlation distance. Thus, $ka \gg 1$ refers to large scale scatterers and $ka \ll 1$ refers to small-scale scatterers with respect to the wavelength. $ka \ll 1$ corresponds to Rayleigh scattering. For large-scale scatterers, $ka \gg 1$, then the mean square velocity variation must be small for the strict validity of the Born approximation. For small-scale scatterers, relatively large mean square velocity fluctuations can be accommodated by the Born approximation. Thus

$ka = 10.$ $E[\mu^2]kL \ll 0.056$

$ka = 2.0$ $E[\mu^2]kL \ll 0.280$

$ka = 0.5$ $E[\mu^2]kL \ll 4.50$

$ka = 0.1$ $E[\mu^2]kL \ll 564.$

For a given propagation distance, L, linearizations based on the Born approximation will favour small scale fluctuations.

Appendix B

In this appendix a comparison is made between perturbation solutions which to first order are additive in the complex phase or multiplicative in the field. These include the first order geometric optics approximation, the parabolic equation method, and the Rytov approximation. These types of approximations work best in a smoothly varying media with predominantly forward scattering. First, the geometric optics approximation is outlined and compared with the standard parabolic approximation. These are then compared to the Rytov approximation.

Starting with the Helmholtz equation, the zeroth order geometric optics approximation has the following general trial solution

$$u = \exp(k\varphi_0 + \varphi_1) \tag{B1}$$

where $\bar{k}$ is a large parameter and $\varphi_0 \sim \varphi_1$. In this approximation $\bar{k}$ is usually identified with an average wavenumber and large $\bar{k}$ implies a small wavelength. By substituting this into the Helmholtz equation then

$$\bar{k}^2 \{\nabla\varphi_0 \nabla\varphi_0 + k^2/\bar{k}^2\} + \bar{k}\{\nabla^2\varphi_0 + 2\nabla\varphi_0 \nabla\varphi_1\} + \{\nabla^2\varphi_1 + \nabla\varphi_1 \nabla\varphi_1\} = 0 \tag{B2}$$

where $k = \omega/v(\mathrm{x})$. The particular trial solution usually used for geometric optis is $u = A(\mathrm{x})$ exp $[i\bar{k}S(\mathrm{x})]$ where $\bar{k} = \omega/\langle v \rangle$. Thus $\varphi_0 = iS(\mathrm{x})$ and $\varphi_1 = \ln A(\mathrm{x})$ and (B2) can then be written as

$$\bar{k}^2 [-(\nabla S)^2 + k^2/\bar{k}^2] + 2ik \left(\frac{1}{2}\nabla^2 S + \nabla S \nabla \ln A\right) + (\nabla^2 A/A) = 0. \tag{B3}$$

For large $\bar{k}$, the first geometric optics solution involves dropping the $\{\nabla^2 A/A\}$ term. The Eikonal equation results from equating the first bracket to zero and the transport equation results from equating the second bracket to zero. The transport equation relates the imaginary part of the phase to the real part of the phase.

The standard parabolic equation method is derived by using for (B1) the particular trial solution $u = Ue^{i\bar{k}x}$ where U is assumed to be a slowly varying function, then $\varphi_0 = ix$ and $\varphi_1 = \ln U(x)$. (B2) can then be written

$$\bar{k}^2 \{-1 + k^2/\bar{k}^2\} + 2i\bar{k}\left\{0 + \frac{U_x}{U}\right\} + \left\{\frac{U_{,xx} + U_{,yy} + U_{,zz}}{U}\right\} = 0. \tag{B4}$$

For large $\bar{k}$, the first bracket gives $k = \bar{k}$, and the second bracket gives $U_{,x} = 0$. The parabolic approximation results from neglecting the $U_{,xx}/U$ in the final bracket but retaining the transverse part of the Laplacian, $\nabla^2_{\text{perp}} = U_{,yy} + U_{,zz}$. This is not entirely consistent within the perturbation scheme used but it gives a one way operator in x,

$$2i\bar{k}U_{,x} + \nabla^2_{\text{perp}}U + \{k^2 - \bar{k}^2\}\, U = 0. \tag{B5}$$

The transverse Laplacian is called the diffraction term and the term $\{k^2 - \bar{k}^2\}$ is called the 'thin lens term' and locally adjusts for the refractive index.

The Rytov approximation results by using the trial solution

$$u = \exp(\psi_0 + \epsilon\psi_1), \tag{B6}$$

where ϵ is a small parameter, and $\psi_0 \sim \psi_1$. The velocity is also expanded as $v(x) = v_0(x) + \epsilon v_1(x)$, where $v_0 \sim v_1$. Substituting this into the Helmholtz equation gives

$$[\nabla^2\psi_0 + (\nabla\psi_0)^2 + \omega^2/v_0^2(x)] + \epsilon\,[\nabla^2\psi_1 + 2\nabla\psi_0\nabla\psi_1 - 2\omega^2 v_1(x)/v_0^3(x)] + \epsilon^2\,[(\nabla\psi_1)^2 + 3\omega^2 v_1^2(x)/v_0^4(x)] = 0, \tag{B7}$$

where the first term in the brackets is a solution of the Helmholtz equation with $v = v_0(x)$ and $u = \exp(\psi_0)$. In the first Rytov approximation, the ϵ^2 term is dropped. Equating the second term to zero gives

$$\nabla^2\psi_1 + 2\nabla\psi_0\nabla\psi_1 = 2v_1\omega^2/v_0^3. \tag{B8}$$

In order to compare with the previous approximations let $\psi_0 = \bar{k}\varphi_0$ and $\psi_1 = \varphi_1/\epsilon = \bar{k}\varphi_1$, then (B7) can be written

$$\left\{\bar{k}\nabla^2\varphi_0 + \bar{k}^2(\nabla\varphi_0)^2 + \bar{k}^2\left(\frac{k_0^2}{\bar{k}^2}\right)\right\} + \left\{\nabla^2\varphi_1 + \bar{k}2\nabla\varphi_0\nabla\varphi_1 + \bar{k}\left(\frac{k_0^2}{\bar{k}^2}\frac{2v_1}{v_0}\right)\right\} + \left\{(\nabla\varphi_1)^2 + \left(\frac{k_0^2}{\bar{k}^2}\frac{3v_1^2}{v_0^2}\right)\right\} = 0, \tag{B9}$$

where $\bar{k} = \omega/\langle v_0(x)\rangle$, $k_0 = \omega/v_0(x)$, and $\bar{k} \approx k_0$. The Rytov approximation again retains the first two brackets, where the first bracket is assumed known. The $\bar{k}^2$ terms relates to the first model in equation (B2), only now related to the velocity model $v_0(x)$. The $\bar{k}$ terms correspond to the second model in (B2) plus a correction term to the velocity linear in $v_1(x)$. Finally, the term $\nabla^2\varphi_1$ is retained, but not the term $(\nabla\varphi_1)^2$ and the higher order terms containing the velocity correction $v_1(x)$. This is similar to the parabolic approximation where again only a part of the third bracket in (B2) is retained. In that case the

transverse Laplacian perpendicular to the predominant direction of the incident wave is retained. For example if we let v_0 be a homogeneous velocity, and $\psi_0 = i\bar{k}x$, then by dropping terms in ϵ^2, (B7) can be written in terms of $U = \exp(\epsilon\psi_1)$.

$$2i\bar{k}U_{,x} + \left\{\nabla^2 U - \frac{(\nabla U)^2}{U}\right\} - \frac{2\epsilon v_1 \omega^2}{v_0^3} U = 0.$$

This is similar to the parabolic equation given in (B5) with the thin lens term approximated to first order in $\epsilon v_1(\mathbf{x})$ and the term $\nabla^2_{\text{perp}} U$ replaced by $\nabla^2 U - (\nabla U)^2/U$. Using the transverse Laplacian results in a simpler equation but the Rytov equation above is consistent to order ϵ.

Appendix C

In this appendix, the acoustic wave equation is investigated in a simple 1-D example, where

$$L = \left\{\frac{1}{K(r)}\frac{\partial^2}{\partial t^2} - \nabla\left(\frac{1}{\rho(r)}\nabla\right)\right\}.$$

The transposed linear sensitivity operators for a perturbation in density and bulk modulus from the field are given in Tarantola (1984a) where

$$\delta K'_k = U_k^T \delta p = \frac{1}{K^2(r)} \sum_{r_s}\sum_{r_g} \int dt \{\ddot{p}(r, t; r_s) * g(r, t; r_g)\} \delta p(r_g, t; r_s)$$

$$\delta \rho'_k = V_k^T \delta p = \frac{1}{\rho^2(r)} \sum_{r_s}\sum_{r_g} \int dt \nabla\{p(r, t; r_g) * \nabla g(r, t; r_g)\} \delta p(r_g, t; r_s)$$

or

$$\delta \bar{K}'_k = \frac{1}{K^2(r)} \int dt \sum_{r_s} \ddot{p}(r, t; r_s) \left\{\sum_{r_g} g(r, -t; r_g) * \delta p(r_g, t; r_s)\right\} \tag{C1}$$

$$\delta \rho'_k = \frac{1}{\rho^2(r)} \int dt \sum_{r_s} \nabla p(r, t; r_s) \left\{\sum_{r_g} \nabla g(r, -t; r_g) * \delta p(r_g, t; r_s)\right\}. \tag{C2}$$

In the 1-D case, a coincident source and receiver are considered where $r_g = r_s = z = 0$. Small perturbations of a homogeneous starting model to the true model will be assumed. The observed reflection seismogram is

$P_{\text{obs}}(r_g, t; r_s) = P_{\text{obs}}(t)$

and the predicted seismogram is

$P_k(r_g, t; r_s) = P_k(t)$.

The differential seismogram is then

$\delta p_k(r_g, t; r_s) = P_{\text{obs}}(t) - P_k(t)$.

The operation in (C1) can then be written

$$\delta K_k' = \frac{1}{K_k^2(z)} \int dt\, \ddot{p}_k(z, t; 0) \langle g(z, t; 0)\, \delta p_k(t) \rangle$$

where $\langle \ldots \rangle$ denotes a cross correlation. The analogous operation in $\delta\rho_k'$ in (C2) can be written

$$\delta\rho_k' = \frac{1}{\rho_k^2(z)} \int dt\, \frac{\partial}{\partial z} p_k(z, t; 0) \frac{\partial}{\partial z} \langle g(z, t; 0)\, \delta p_k(t) \rangle .$$

With an initial homogeneous model with velocity v then

$$g(z, t; 0) = \delta(t - z/v)$$

and

$$\langle g(z, t; 0)\, \delta p_k(t) \rangle = \delta p_k(t + z/v).$$

Then

$$\delta K_k'(z) = \frac{1}{K_k^2(z)} \int dt\, \frac{\partial^2}{\partial t^2} p_k(z, t; 0)\, \delta p_k(t + z/v)$$

$$\delta\rho_k'(z) = \frac{1}{\rho_k^2(z)} \int dt\, \frac{\partial}{\partial z} p_k(z, t; 0) \frac{\partial}{\partial z} \delta p_k(t + z/v).$$

Now with $p_k(z, t; 0) = \delta(t - z/v)$ then

$$\delta K_k' = \frac{1}{K_k^2(z)}\, \delta\ddot{p}_k(t = 2z/v)$$

$$\delta\rho_k'(z) = \frac{1}{v^2 \rho_k^2(z)}\, \delta\ddot{p}_k(t = 2z/v)$$

where $t = 2z/v$ places the model perturbation at $z = vt/2$. Since $v = \sqrt{K/\rho}$ and $\eta = \rho v =$ impedance, then

$$\frac{\delta v}{v} = \left\{ \frac{\delta K}{K} - \frac{\delta\rho}{\rho} \right\}$$

$$\frac{\delta\eta}{\eta} = \left\{ \frac{\delta K}{K} + \frac{\delta\rho}{\rho} \right\} .$$

Including the model covariances this can be written assuming, $C_K \approx \sigma_k^2 \approx K_k^2(z)$

$$\delta K_k(z) = C_K \delta K_k'(z) \approx \frac{\sigma_k^2}{K_k^2(z)}\, \delta\ddot{p}(2z/v) \approx \delta\ddot{p}(2z/v)$$

and with $\sigma_\rho^2 \approx \rho_k^2(z)$,

$$\delta\rho_k = C_\rho \delta p_k' \approx \frac{\sigma_\rho^2}{v^2 \rho_k^2(z)}\, \delta\ddot{p}(2z/v) \approx \frac{1}{v^2}\, \delta\ddot{p}(2z/v).$$

Perturbations in velocity and impedance can then be written

$$\frac{\delta v}{v} = \frac{1}{2}\left[\frac{1}{K} - \frac{1}{v^2 \rho}\right] \delta \ddot{p}(2z/v) = 0$$

and

$$\frac{\delta \eta}{\eta} = \frac{1}{2}\left[\frac{1}{K} + \frac{1}{K}\right] \delta \ddot{p}(2z/v) = \frac{1}{K}\, \delta \ddot{p}(2z/v).$$

Thus for the 1-D reflection seismogram modelled using the acoustic wave equation, there is no velocity perturbation, but there is an impedance perturbation. This is as expected from the 1-D exact inversion results.