Elastic inversion using ray theory

Peter Mora

INTRODUCTION

This paper extends the SEP-38 paper titled "Inversion of CMP gathers for P and S velocity", in which I introduced a method of elastic inversion using ray theory. The inversion method was based on nonlinear least squares and proceeded iteratively, forward modeling and updating the physical parameters, until a minimum error solution was located. The forward modeling part of the algorithm was a ray theoretical method, which computed P-wave primary reflections with the restriction that the physical model be homogeneous plane layers. The main reason for using the ray theoretical method was for computational speed of forward modeling. Another reason was that it enabled the Frechet derivatives used in the inversion algorithm to be computed analytically. (The Frechet derivatives represent the linearization of the nonlinear function.)

The elastic inversion algorithm solves for P velocity, S velocity, density and source wavelet using the amplitude information contained in a common midpoint gather. It requires that a sample rate for the plane layers be specified, which enables the modeled wavefield to accurately simulate the common midpoint gather data. Therefore, layer boundaries need not be specified but will be located as part of the inversion process.

This paper extends the ray equations of the earlier paper to include all primary ray modes and the ray spreading term. (See Mora, SEP-41 for a simple derivation of the ray spreading term.) The primary rays are the PP and SS reflections and the converted modes, PS and SP. The equations have been reformulated from the frequency domain to the time domain. This improves the efficiency because the convolution of the source wavelet with the impulse response is faster in the time domain than in the frequency domain if the source wavelet is short relative to the total length of the time series. The ray algorithm was also modified so the layers have constant vertical P-wave traveltime.

This helps to stabilize the inversion by avoiding movements (at zero offset) of the dominant PP reflection hyperbolae from one iteration to the next. Another improvement to the algorithm is that the reflection and transmission coefficient derivatives are computed analytically whereas previously they were evaluated using a finite difference formula.

Some automatic methods for choosing data variances are described. The model variances are chosen manually and are crucial to stable inversions. Usually several trials must be carried out in order to obtain model variances that result in a satisfactory inversion. A discussion of the meanings of variances in nonlinear inverse problems is included and finally an example of an inversion of synthetic data is given. In order to study the inversion process, and particularly the sensitivities of the different model parameters in a controlled situation, the synthetic data was generated using the same ray modeling algorithm as was used in the inversion and no noise was added.

A 100 layer model was generated with random but non-Gaussian distributed physical properties. The layers were 8 milliseconds thick measured in vertical two-way P-wave traveltime and the inversion was carried out for P velocity, S velocity and density. The data used in the inversion was a 16 fold CMP gather sampled at 8 milliseconds (ie: with the same sample rate as the layers). The solution is not perfect in this example. There are at least four possible reasons. (1) The iterations may have been terminated too early, (2) The choice of the model covariances may not be ideal, (3) The example may contain a null space that corresponds to those model perturbations that do not significantly alter the modeled wavefield and hence cannot be resolved. The null space problem is tied in with both the choice of data and model covariances and the implicit resolvability of the physical parameters, (4) Degradation of accuracy due to use of the data space contraction method (see Mora, SEP-41).

FORWARD MODELING AND FRECHET DERIVATIVES

Ray modeling

The ray theoretical CMP wavefield recorded at the earths surface is the sum of wavefields due to all possible raypaths between the source and receiver pairs. Therefore, the CMP wavefield denoted $u(\mathbf{m},x,\omega)$ is given by the sum

$$u(\mathbf{m},x,\omega) = \sum_j u_j(\mathbf{m},x,\omega) = \sum_j u_j(\mathbf{m},p_j(\mathbf{m},x),\omega) \tag{1}$$

where $\mathbf{m}$ is the model vector, p is the ray parameter, x is offset and ω is angular frequency. From ray theory, the wavefield due to the j-th raypath through a one

dimensional earth is

$$u_j(\mathbf{m},x,\omega) = s(\omega)a_j(\mathbf{m},x)exp(-i\omega\tau_j(\mathbf{m},x)) \tag{2}$$
$$= s(\omega)a_j(\mathbf{m},p_j(\mathbf{m},x))exp(-i\omega\tau_j(\mathbf{m},p_j(\mathbf{m},x)))$$

where $s(\omega)$ is the source, a_j is the complex amplitude and τ_j is the traveltime.

Frechet derivatives

The Frechet kernel is the partial derivative matrix of the CMP wavefield with respect to model parameters. Using the chain rule, the derivative of $u_j(\mathbf{m},x,\omega)$ with respect to the k-th model parameter is

$$\frac{\partial u}{\partial m_k}(\mathbf{m},x,\omega) = \sum_j \frac{\partial u_j}{\partial m_k}(\mathbf{m},x,\omega)$$

$$= \sum_j \left(\frac{\partial u_j}{\partial p_j}(\mathbf{m},p_j,\omega)\frac{\partial p_j}{\partial m_k}(\mathbf{m},x) + \frac{\partial u_j}{\partial m_k}(\mathbf{m},p_j,\omega) \right) \tag{3}$$

The second term represents the derivative of the wavefield in the p-domain, which is the domain where the ray tracing is performed. The first term is the adjustment that converts the derivative from the p-domain to the x-domain. (This is required because x_j and p_j are dependent, i.e. $p_j = p_j(\mathbf{m},x)$.) From equation (2)

$$\frac{\partial u_j}{\partial p_j}(\mathbf{m},p_j,\omega) = u_j(\mathbf{m},p_j,\omega)\left(\frac{1}{a_j}\frac{\partial a_j}{\partial p_j}(\mathbf{m},p_j) - i\omega\frac{\partial \tau_j}{\partial p_j}(\mathbf{m},p_j) \right) \tag{4a}$$

$$\frac{\partial u_j}{\partial m_k}(\mathbf{m},p_j,\omega) = u_j(\mathbf{m},p_j,\omega)\left(\frac{1}{a_j}\frac{\partial a_j}{\partial m_k}(\mathbf{m},p_j) - i\omega\frac{\partial \tau_j}{\partial m_k}(\mathbf{m},p_j) \right) \tag{4b}$$

The $\frac{\partial p_j}{\partial m_k}(\mathbf{m},x)$ term of equation (3) may be computed from the equation for $p_j(\mathbf{m},x)$. However, usually the offset $x_j(\mathbf{m},p_j)$ as a function of ray parameter is known. In this case the derivative of the ray parameter with respect to the k-th model parameter $\frac{\partial p_j}{\partial m_k}(\mathbf{m},x)$, is evaluated by setting the total differential dx_j to zero.

$$dx_j = \frac{\partial x_j}{\partial m_k}(\mathbf{m},p_j)dm_k + \frac{\partial x_j}{\partial p_j}(\mathbf{m},p_j)dp_j = 0 \tag{5}$$

This equation defines the relation between a model parameter and the ray parameter for the j-th raypath in order that the offset x_j be fixed. Rearranging equation (5) yields

$$\frac{\partial p_j}{\partial m_k}(\mathbf{m},x) = \left.\frac{dp_j}{dm_k}\right|_{dx_j=0} = -\left(\frac{\partial x_j}{\partial m_k}(\mathbf{m},p_j) \right)\left(\frac{\partial x_j}{\partial p_j}(\mathbf{m},p_j) \right)^{-1} \tag{6}$$

Substituting equations (4a) and (4b) into equation (3) yields the Frechet kernel in the frequency domain:

$$\frac{\partial u}{\partial m_k}(\mathbf{m},x,\omega) = \sum_j u_j(\mathbf{m},p_j,\omega)\left[\begin{array}{c}\left(\frac{1}{a_j}\frac{\partial a_j}{\partial p_j}(\mathbf{m},p_j) - i\omega\frac{\partial \tau_j}{\partial p_j}(\mathbf{m},p_j)\right)\frac{\partial p_j}{\partial m_k}(\mathbf{m},x) \\ + \frac{1}{s}\frac{\partial s}{\partial m_k}(\omega) + \frac{1}{a_j}\frac{\partial a_j}{\partial m_k}(\mathbf{m},p_j) - i\omega\frac{\partial \tau_j}{\partial m_k}(\mathbf{m},p_j)\end{array}\right] \quad (7)$$

Equations (6) and (7) define the analytic ray theoretical Frechet derivatives. However, equation (7) is not a computationally efficient form of the Frechet derivatives if the source wavelet is short in relation to the time axis. This is because convolution of a short wavelet with a long time series may involve less computations than multiplication in the frequency domain. The $-i\omega$ terms of equation (7), which correspond to time derivatives, are now grouped together and after some substitutions the expression is inverse Fourier transformed to yield the time domain Frechet derivatives:

$$\frac{\partial u}{\partial m_k}(\mathbf{m},x,\omega) =$$

$$\sum_j u_j(\mathbf{m},p_j,\omega)\left[\begin{array}{c} - i\omega\left(\frac{\partial \tau_j}{\partial p_j}(\mathbf{m},p_j)\frac{\partial p_j}{\partial m_k}(\mathbf{m},x) + \frac{\partial \tau_j}{\partial m_k}(\mathbf{m},p_j)\right) \\ + \frac{1}{s}\frac{\partial s}{\partial m_k}(\omega) + \frac{1}{a_j}\frac{\partial a_j}{\partial p_j}(\mathbf{m},p_j)\frac{\partial p_j}{\partial m_k}(\mathbf{m},x) + \frac{1}{a_j}\frac{\partial a_j}{\partial m_k}(\mathbf{m},p_j)\end{array}\right]$$

$$= \sum_j u_j(\mathbf{m},p_j,\omega)\left[-i\omega A_j(\mathbf{m},p_j,m_k) + B_j(\mathbf{m},p_j,m_k) + \frac{1}{s}\frac{\partial s}{\partial m_k}(\omega)\right]$$

$$= \sum_j s(\omega)a_j(\mathbf{m},p_j)exp\left(-i\omega\tau_j(\mathbf{m},p_j)\right)\left[-i\omega A_j(\mathbf{m},p_j,m_k) + B_j(\mathbf{m},p_j,m_k) + \frac{1}{s}\frac{\partial s}{\partial m_k}(\omega)\right]$$

$$= \sum_j s(\omega)exp\left(-i\omega\tau_j(\mathbf{m},p_j)\right)\left[-i\omega C_j(\mathbf{m},p_j,m_k) + D_j(\mathbf{m},p_j,m_k) + a_j(\mathbf{m},p_j)\frac{1}{s}\frac{\partial s}{\partial m_k}(\omega)\right]$$

$$= F\left\{\begin{array}{l} s(t)*\sum_j \left[\begin{array}{l}\mathrm{Re}\left\{C_j(\mathbf{m},p_j,m_k)\right\}\delta'(t-\tau_j(\mathbf{m},p_j)) + \mathrm{Im}\left\{C_j(\mathbf{m},p_j,m_k)\right\}H\left\{\delta'(t-\tau_j(\mathbf{m},p_j))\right\} \\ + \mathrm{Re}\left\{D_j(\mathbf{m},p_j,m_k)\right\}\delta(t-\tau_j(\mathbf{m},p_j)) + \mathrm{Im}\left\{D_j(\mathbf{m},p_j,m_k)\right\}H\left\{\delta(t-\tau_j(\mathbf{m},p_j))\right\}\end{array}\right] \\ + \frac{\partial s}{\partial m_k}(t)*\sum_j \left[\mathrm{Re}\left\{a_j(\mathbf{m},p_j)\right\}\delta(t-\tau_j(\mathbf{m},p_j)) + \mathrm{Im}\left\{a_j(\mathbf{m},p_j)\right\}H\left\{\delta(t-\tau_j(\mathbf{m},p_j))\right\}\right]\end{array}\right\} \quad (8)$$

where H represents the Hilbert transform, δ is the Dirac delta function and the $C_j(\mathbf{m},p_j,m_k)$ and $D_j(\mathbf{m},p_j,m_k)$ terms are defined in equations (10a) and (10b). Therefore, the time domain Frechet kernel is:

$$\frac{\partial u}{\partial m_k}(\mathbf{m},x,t) = \begin{cases} \sum_j \left[\begin{array}{l} \mathrm{Re}\left\{ C_j(\mathbf{m},p_j,m_k) \right\} s'\,(t-\tau_j(\mathbf{m},p_j)) \\ + \mathrm{Im}\left\{ C_j(\mathbf{m},p_j,m_k) \right\} H\left\{ s'\,(t-\tau_j(\mathbf{m},p_j)) \right\} \\ + \mathrm{Re}\left\{ D_j(\mathbf{m},p_j,m_k) \right\} s\,(t-\tau_j(\mathbf{m},p_j)) \\ + \mathrm{Im}\left\{ D_j(\mathbf{m},p_j,m_k) \right\} H\left\{ s\,(t-\tau_j(\mathbf{m},p_j)) \right\} \end{array} \right] & , \quad m_k = v_{Pl}, v_{Sl}, \rho_l \\ \sum_j \left[\begin{array}{l} \mathrm{Re}\left\{ a_j(\mathbf{m},p_j) \right\} \delta(t-T-\tau_j(\mathbf{m},p_j)) \\ + \mathrm{Im}\left\{ a_j(\mathbf{m},p_j) \right\} H\left\{ \delta(t-T-\tau_j(\mathbf{m},p_j)) \right\} \end{array} \right] & , \quad m_k = s(T) \end{cases} \tag{9}$$

where the derivatives with respect to physical model parameters v_{Pl}, v_{Sl} and ρ_l are shown separately from the derivatives with respect to the source wavelet $s(T)$ and $C_j(\mathbf{m},p_j,m_k)$ and $D_j(\mathbf{m},p_j,m_k)$ are defined below.

$$C_j(\mathbf{m},p_j,m_k) = a_j(\mathbf{m},p_j)\left(\frac{\partial \tau_j}{\partial p_j}(\mathbf{m},p_j)\frac{\partial p_j}{\partial m_k}(\mathbf{m},x) + \frac{\partial \tau_j}{\partial m_k}(\mathbf{m},p_j) \right) \tag{10a}$$

$$D_j(\mathbf{m},p_j,m_k) = \frac{\partial a_j}{\partial p_j}(\mathbf{m},p_j)\frac{\partial p_j}{\partial m_k}(\mathbf{m},x) + \frac{\partial a_j}{\partial m_k}(\mathbf{m},p_j) \tag{10b}$$

So far the formulas are generally applicable to rays travelling a 1-D earth. The following derivations specialize the formulation to primary reflections only and so includes only the PP, PS, SP and SS primary rays. Henceforth, subscripts u and d are used to represent the upgoing and downgoing wave types so both u and d may be either P or S. Other subscripts used are l for a ray reflected from the l-th interface and j for the j-th layer. Also, for brevity, let $p = p_{udl}(\mathbf{m},x)$. The equations for ray amplitude $a_{udl}(\mathbf{m},p)$, traveltime $\tau_{udl}(\mathbf{m},p)$ and offset $x_{udl}(\mathbf{m},p)$ are

$$a_{udl}(\mathbf{m},p) = d_{udl}(\mathbf{m},p)R_{udl}(\mathbf{m},p)\prod_{j=1}^{l-1}\left(T_{uj}(\mathbf{m},p)T_{dj}(\mathbf{m},p) \right) \tag{11}$$

$$\tau_{udl}(\mathbf{m},p) = \sum_{j=1}^{l} \Delta z_j \left(\frac{1}{v_{uj}\cos\theta_{uj}} + \frac{1}{v_{dj}\cos\theta_{dj}} \right) \tag{12a}$$

$$x_{udl}(\mathbf{m},p) = \sum_{j=1}^{l} \Delta z_j \left(\tan\theta_{dj} + \tan\theta_{uj} \right) \tag{12b}$$

where d is the ray divergence factor and R and T are the complex elastic reflection and transmission coefficients respectively. The ray divergence factor (from Mora, SEP-41, "A simple geometric derivation of the ray spreading factor") is

$$d_{udl}(\mathbf{m},p) = \left[\left(\frac{\cos\theta_{dl}\cos\theta_{u1}}{\cos\theta_{d1}\cos\theta_{ul}} \right) \frac{\cos^2\theta_{d1}}{v_{d1}} \sum_{j=1}^{l} \Delta z_j \left(\frac{v_{uj}}{\cos^3\theta_{uj}} + \frac{v_{dj}}{\cos^3\theta_{dj}} \right) \right]^{-1} \tag{13}$$

If P-wave traveltime through layers is held constant and the trigonometric factors are expressed in terms of ray parameter p and velocity, equations (12) and (13) become

$$\tau_{udl}(\mathbf{m},p) = \sum_{j=1}^{l} v_{Pj}\Delta t_j \left(\frac{1}{v_{uj}\sqrt{1-p^2v_{uj}^2}} + \frac{1}{v_{dj}\sqrt{1-p^2v_{dj}^2}} \right) \tag{14a}$$

$$x_{udl}(\mathbf{m},p) = \sum_{j=1}^{l} v_{Pj}\Delta t_j \left(\frac{pv_{dj}}{\sqrt{1-p^2v_{dj}^2}} + \frac{pv_{uj}}{\sqrt{1-p^2v_{uj}^2}} \right) \tag{14b}$$

$$d_{udl}(\mathbf{m},p) =$$

$$\left[\frac{\sqrt{1-p^2v_{dl}^2}\sqrt{1-p^2v_{u1}^2}}{\sqrt{1-p^2v_{d1}^2}\sqrt{1-p^2v_{ul}^2}} \frac{(1-p^2v_{d1}^2)}{v_{d1}} \sum_{j=1}^{l} \Delta z_j \left(\frac{v_{uj}}{(1-p^2v_{uj}^2)^{3/2}} + \frac{v_{dj}}{(1-p^2v_{dj}^2)^{3/2}} \right) \right]^{-1}$$

$$= f_{udl}^{-1}(\mathbf{m},p) = \left(F_1F_2F_3 \right)^{-1} \tag{15}$$

The terms required for the Frechet derivatives (see equations (6), (9) and (10)) are

$$\frac{\partial\tau_{udl}}{\partial m_k}(\mathbf{m},p) = \begin{cases} \dfrac{Np^2v_{Pj}\Delta t_j}{(1-p^2v_{Pj}^2)^{3/2}} & , \quad m_k = v_{Pj} \ , \ j \le l \\ \dfrac{N(2p^2v_{Sj}^2-1)v_{Pj}\Delta t_j}{v_{Sj}^2(1-p^2v_{Sj}^2)^{3/2}} & , \quad m_k = v_{Sj} \ , \ j \le l \\ 0 & , \quad otherwise \end{cases} \tag{16}$$

$$\frac{\partial x_{udl}}{\partial m_k}(\mathbf{m},p) = \begin{cases} \dfrac{Npv_{Pj}\Delta t_j(2-p^2v_{Pj}^2)}{(1-p^2v_P^2)^{3/2}} & , \quad m_k = v_{Pj} \ , \ j \le l \\ \dfrac{Npv_{Pj}\Delta t_j}{(1-p^2v_S^2)^{3/2}} & , \quad m_k = v_{Sj} \ , \ j \le l \\ 0 & , \quad otherwise \end{cases} \tag{17}$$

where $N = 2$ if $u = d$ and $N = 1$ if $u \neq d$; and

$$\frac{\partial x_{udl}}{\partial p}(\mathbf{m},p) = \sum_{j=1}^{l} \left(\frac{v_{Pj}\, v_{uj}\, \Delta t_j}{(1 - p^2 v_{uj}^2)^{3/2}} + \frac{v_{Pj}\, v_{dj}\, \Delta t_j}{(1 - p^2 v_{dj}^2)^{3/2}} \right) \tag{18}$$

$$\frac{\partial a_l}{\partial m_k}(\mathbf{m},p) =$$

$$\begin{cases} a_l(\mathbf{m},p) \left(\dfrac{1}{d_{udl}} \dfrac{\partial d_{udl}}{\partial m_k}(\mathbf{m},p) + \sum\limits_{i=j}^{\min(j+1,l)} \left(\dfrac{1}{T_{dj}} \dfrac{\partial T_{dj}}{\partial m_k}(\mathbf{m},p) + \dfrac{1}{T_{uj}} \dfrac{\partial T_{uj}}{\partial m_k}(\mathbf{m},p) \right) \right) & m_k \rightarrow \text{layer } i \text{, } i < l \\ a_l(\mathbf{m},p) \left(\dfrac{1}{d_{udl}} \dfrac{\partial d_{udl}}{\partial m_k}(\mathbf{m},p) + \dfrac{1}{R_{udl}} \dfrac{\partial R_{udl}}{\partial m_k}(\mathbf{m},p) + \sum\limits_{i=j}^{\min(j+1,l)} \left(\dfrac{1}{T_{dj}} \dfrac{\partial T_{dj}}{\partial m_k}(\mathbf{m},p) + \dfrac{1}{T_{uj}} \dfrac{\partial T_{uj}}{\partial m_k}(\mathbf{m},p) \right) \right) & m_k \rightarrow \text{layer } i \text{, } i = l \\ a_l(\mathbf{m},p) \dfrac{1}{R_{udl}} \dfrac{\partial R_{udl}}{\partial m_k}(\mathbf{m},p) & m_k \rightarrow \text{layer } i \text{, } i = l+1 \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

The remaining terms required in order to evaluate the Frechet derivatives are the amplitude derivative terms of equation (19). From equation (15), the ray spreading derivative is

$$\frac{1}{d_{udl}} \frac{\partial d_{udl}}{\partial m_k}(\mathbf{m},p) = \frac{1}{d_{udl}} \frac{\partial d_{udl}}{\partial f_{udl}}(f_{udl}) \frac{\partial f_{udl}}{\partial m_k}(\mathbf{m},p) = -d_{udl} \frac{\partial f_{udl}}{\partial m_k}(\mathbf{m},p) \tag{20}$$

where from equation (15) and using the product rule we have:

$$\frac{\partial f_{udl}}{\partial m_k}(\mathbf{m},p) = \frac{\partial F_1}{\partial m_k} F_2 F_3 + \frac{\partial F_2}{\partial m_k} F_1 F_3 + \frac{\partial F_3}{\partial m_k} F_1 F_2 \tag{21}$$

From equation (15), which implicitly defines F_1, F_2 and F_3, we obtain:

$$\frac{\partial F_3}{\partial m_k} = \begin{cases} \dfrac{N v_{Pj}\, \Delta t_j (2 + p^2 v_{Pj}^2)}{(1 - p^2 v_{Pj}^2)^{3/2}} & , \quad m_k = v_{Pj} \\ \dfrac{N v_{Pj}\, \Delta t_j (1 + 2p^2 v_{Sj}^2)}{(1 - p^2 v_{Sj}^2)^{3/2}} & , \quad m_k = v_{Sj} \\ 0 & , \quad \text{otherwise} \end{cases} \tag{22}$$

$$\frac{\partial F_2}{\partial m_k} = \begin{cases} \dfrac{p^2 v_{d1}^2 - 1}{v_{d1}^2} & , \quad m_k = v_{d1} \\ 0 & , \quad otherwise \end{cases} \tag{23}$$

$$\frac{\partial F_1}{\partial m_k} = \begin{cases} \dfrac{-p^2 v_j^2}{(1 - p^2 v_j^2)} F_1 & m_k = v_j = v_{u1} \quad or \quad v_{dl} \\ \dfrac{p^2 v_j^2}{(1 - p^2 v_j^2)} F_1 & m_k = v_j = v_{d1} \quad or \quad v_{ul} \\ 0 & otherwise \end{cases} \tag{24}$$

As may be expected, the ray spreading derivatives given in equations (20) to (24) were found to have a second order effect on the Frechet kernel and so can be ignored in most practical computations. This was done in the inversion example in this paper.

All the derivatives with respect to p in equation (10) such as $\frac{\partial \tau_j}{\partial p_j}(\mathbf{m}, p_j)$ are computed numerically using cubic splines.

Finally, the reflection and transmission coefficient derivatives in equation (19) are evaluated by solving the equations below. Let the elastic boundary conditions be given by the linear equation

$$AX = B \tag{25}$$

where X is the 4×4 complex scattering matrix. Taking the derivative of both sides

$$A' X + AX' = B' \tag{26}$$

Rearranging equation (26)

$$X' = A^{-1}(B' - A' X) \tag{27}$$

Therefore, the elastic boundary conditions A and B and their derivatives A' and B' are all that is required in order to compute the derivatives of reflection and transmission coefficients. (Note: A is made non-singular by addition of an infinitesimal amount of attenuation, thus avoiding problems with Rayleigh poles etc.)

INVERSION

Discussion

The following is a review of the inversion theory applicable to this paper. Inversion may be considered to be the process of locating the model $\mathbf{m}$ that maximizes the posterior probability given the prior probabilities and knowledge of the theoretical

relationship between data **d** and model **m** (Tarantola, 1982). Many simplifications result from choosing particular prior distributions; in particular, the choice of Gaussian "priors" leads to the well known least squares solution. In this paper, I have chosen Gaussian priors in order utilize the tractable least squares formulae even though neither data nor model errors may actually be Gaussian distributed. Although it would seem a major flaw to assume probability functions that are most likely not correct, a reasonable solution may still be obtained by sacrificing some of the a priori model constraints and probabilistic rigor. This is because the inversion can be performed iteratively (even for linear theories) while allowing the Gaussian a priori information to vary during the iterations. (This is contrary to the philosophy presented by Tarantola, 1982.) For example, the background model may be allowed to change during the iterations. Therefore, the standard deviation and mean of the Gaussian probability functions is not held constant during the iterations. The true prior information about the model is essentially ignored and the Gaussian model statistics are only used to control the model perturbations at each iteration. In some cases this iterative approach would yield the correct probabilistic solution; for example, suppose the model were expected to have Gaussian statistics but about an unknown mean. (This unknown mean can often be given constraints such as smoothness.) A problem with allowing the background model to vary is that nonlinearity is introduced so the iterations may become unstable. This problem can often be solved by constraining the background model using some criterion such as smoothness.

The problem of non-Gaussian data errors can be partially solved by enabling the data covariances to vary spatially and also as a function of iteration in a manner similar to the way the background model is allowed to vary with each iteration. This method was used by Thorson in his thesis (SEP-39).

To summarize, inversion by least squares is based on Gaussian prior probability density functions. In practice, the method may be utilized when the errors are non-Gaussian or the equations are nonlinear by performing an iterative inversion in which the prior information is allowed to vary with each iteration. Specifically, the data and model covariance matrices and the background model may be changed at each iteration. This introduces an additional nonlinearity to the problem. The way that the background model and covariances are updated during the iterations is crucial to the stability of the least squares inversion as well as the accuracy of the final solution. Henceforth, the Gaussian prior information will be considered to be a function of iteration and is termed pseudo prior information to distinguish it from the true prior information, which may or may not be Gaussian.

Review of least squares inversion

Consider the case of Gaussian prior probabilities where the covariance matrices are $\mathbf{C_d}$ and $\mathbf{C_m}$ for the data and model errors respectively. The posterior probability is given by the product of the prior probability functions (see Tarantola, 1982) so we have

$$P = constant\ \exp -\frac{1}{2}\left[\Delta\mathbf{d}^*\mathbf{C_d}^{-1}\Delta\mathbf{d} + \Delta\mathbf{m}^*\mathbf{C_m}^{-1}\Delta\mathbf{m}\right] \tag{28}$$

where $\Delta\mathbf{d}$ is the data error and $\Delta\mathbf{m}$ is the model error. The posterior probability is maximized when the weighted least squares functional F given below is minimized

$$F = \Delta\mathbf{d}^*\mathbf{C_d}^{-1}\Delta\mathbf{d} + \Delta\mathbf{m}^*\mathbf{C_m}^{-1}\Delta\mathbf{m} \tag{29}$$

Let the a priori data or observations be denoted $\mathbf{d}_0$ and the a priori model be denoted $\mathbf{m}_0$. Therefore, the data and model errors are respectively

$$\Delta\mathbf{d} = \mathbf{d}(\mathbf{m}) - \mathbf{d}_0 \quad \text{and} \quad \Delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0 \tag{30}$$

where $\mathbf{d}(\mathbf{m})$ denotes the assumed exact theoretical relationship between data and model. At a minimum in the error functional, its derivative with respect to the model is zero so

$$\frac{\partial F}{\partial \mathbf{m}} = \mathbf{D}^*\mathbf{C_d}^{-1}\Delta\mathbf{d} + \mathbf{C_m}^{-1}\Delta\mathbf{m} = 0 \tag{31}$$

where $\mathbf{D} = \dfrac{\partial \mathbf{d}(\mathbf{m})}{\partial \mathbf{m}}$. Assuming a linear theoretical relationship in the vicinity of $\mathbf{m}_0$, we have $\mathbf{D} = \left.\dfrac{\partial \mathbf{d}(\mathbf{m})}{\partial \mathbf{m}}\right|_{(\mathbf{m}=\mathbf{m}_0)}$ so

$$\mathbf{d}(\mathbf{m}) = \mathbf{d}(\mathbf{m}_0) + \mathbf{D}\Delta\mathbf{m} \tag{32}$$

The model perturbation $\Delta\mathbf{m}$ required for a minimum in the error functional is obtained by substituting equation (32) into equation (31) and rearranging terms.

$$\Delta\mathbf{m} = \left(\mathbf{D}^*\mathbf{C_d}^{-1}\mathbf{D} + \mathbf{C_m}^{-1}\right)^{-1}\mathbf{D}^*\left(\mathbf{d}_0 - \mathbf{d}(\mathbf{m}_0)\right) \tag{33}$$

For nonlinear functions $\mathbf{d}(\mathbf{m})$ this equation can be applied iteratively in order to obtain a solution. The nonlinear inversion equation is derived by performing appropriate substitutions for model error $\Delta\mathbf{m}$ and data error $\Delta\mathbf{d}$ and using some linearization $\mathbf{D}$ of the theoretical relationship $\mathbf{d}(\mathbf{m})$ valid in the vicinity of the a priori model. The model error at the $(n+1)$-th iteration is

$$\Delta\mathbf{m} = \mathbf{m}_{n+1} - \mathbf{m}_0 \tag{34}$$

Similarly, the data error at the $(n+1)$-th iteration is

$$\Delta\mathbf{d} = \mathbf{d}(\mathbf{m}_{n+1}) - \mathbf{d}_0 \tag{35}$$

In the vicinity of the a priori model we have the linear relationship

$$\mathbf{d}(\mathbf{m}_{n+1}) \;=\; \mathbf{d}(\mathbf{m}_n) + \mathbf{D}(\mathbf{m}_0)\left(\mathbf{m}_{n+1} - \mathbf{m}_n\right) \tag{36}$$

Equation (36) is abbreviated to

$$\mathbf{d}_{n+1} \;=\; \mathbf{d}_n + \mathbf{D}_0\left(\mathbf{m}_{n+1} - \mathbf{m}_n\right) \tag{37}$$

Substituting equations (34), (35) and (37) into equation (31) and rearranging (see also Tarantola, 1982) yields

$$\mathbf{m}_{n+1} \;=\; \mathbf{m}_n + \left(\mathbf{D}_0^{*}\mathbf{C}_\mathbf{d}^{-1}\mathbf{D}_0 + \mathbf{C}_\mathbf{m}^{-1}\right)^{-1}\left[\mathbf{D}_0^{*}\mathbf{C}_\mathbf{d}^{-1}(\mathbf{d}_0 - \mathbf{d}_n) - \mathbf{C}_\mathbf{m}^{-1}(\mathbf{m}_n - \mathbf{m}_0)\right] \tag{38}$$

This equation is generalized to allow the a priori information $\mathbf{C}_\mathbf{d}^{-1}$, $\mathbf{C}_\mathbf{m}^{-1}$, $\mathbf{d}_0$ and $\mathbf{m}_0$ to vary with iteration n, so

$$\mathbf{m}_{n+1} = \mathbf{m}_n + \left(\mathbf{D}_{0n}^{*}\mathbf{C}_{\mathbf{d}n}^{-1}\mathbf{D}_{0n} + \mathbf{C}_{\mathbf{m}n}^{-1}\right)^{-1}\left[\mathbf{D}_{0n}^{*}\mathbf{C}_{\mathbf{d}n}^{-1}(\mathbf{d}_{0n} - \mathbf{d}_n) - \mathbf{C}_{\mathbf{m}n}^{-1}(\mathbf{m}_n - \mathbf{m}_{0n})\right] \tag{39a}$$

Usually it makes little sense to alter the observations $\mathbf{d}_0$ as iterations proceed so we would normally have $\mathbf{d}_{0n} = \mathbf{d}_0$. The Frechet derivatives $\mathbf{D}_{0n}$ represent the linearization of the nonlinear equation $\mathbf{d}(\mathbf{m})$ computed at the pseudo a priori model $\mathbf{m}_{0n}$. Sometimes it is advantageous to use the approximation

$$\mathbf{D}_{0n} \;\approx\; \mathbf{D}_n \;=\; \mathbf{D}(\mathbf{m}_n) \tag{39b}$$

For example, if $\mathbf{D}_n$ was computed as part of the forward modeling, then it would be more efficient to use approximation (39b) than to solve the forward problem again using the model $\mathbf{m}_{0n}$ in order to compute the exact value of $\mathbf{D}_{0n}$. (This approximation was used in the inversion example in this paper). For highly nonlinear functions $\mathbf{d}(\mathbf{m})$, other approximations for $\mathbf{D}_{0n}$ may be useful such as the expectation of $\mathbf{D}(\mathbf{m})$ given the prior probability density function for $\mathbf{m}$, ie:

$$\mathbf{D}_{0n} \;\approx\; E\left(\mathbf{D}(\mathbf{m})\right) \tag{39c}$$

This approximation would effectively smooth nonlinearities in $\mathbf{d}(\mathbf{m})$ in a probabilistic sense and hence lead to a more stable inversion. Unfortunately, it is computationally expensive to evaluate $E\left(\mathbf{D}(\mathbf{m})\right)$. Smoothing of $\mathbf{D}(\mathbf{m})$ should be useful in general because it makes the function $\mathbf{d}(\mathbf{m})$ appear more linear in a local sense.

The biggest practical difficulty in an inversion using equation (39a) is the choice of the pseudo a priori information, $\mathbf{C}_{\mathbf{d}n}^{-1}$ and $\mathbf{C}_{\mathbf{m}n}^{-1}$, and $\mathbf{m}_{0n}$, as iterations proceed. (The qualifier "pseudo" is used to emphasize that this information is not truly prior information because it is changed during the iterations.) This choice is especially important for

nonlinear functions $\mathbf{d}(\mathbf{m})$ where the pseudo a priori information is used to stabilize the least squares iterations. Constraints on the pseudo a priori information sometimes achieve stability. These constraints, coupled with the pseudo a priori information, may be considered to be the effective a priori state of knowledge.

One final note on the method is that it cannot handle the problems associated with multiple minima in the error functional. Therefore, the solution obtained will be a local minimum and not a global minimum. However, if the starting guess is within a closed contour on the error functional around the global minima, then equation (39a) will locate the least error solution provided the equation is carefully implemented.

Automatic choice of data covariance

The data covariance matrix required by the least squares inversion formula is not known in general but with some assumptions, it may be estimated from the data. The simplest assumption is that noise is independent and Gaussian with stationary statistics. The inverse covariance matrix in this situation has the form

$$\mathbf{C}_{\mathbf{d}n}^{-1} = \frac{1}{\sigma_{\mathbf{d}n}^2}\mathbf{I} \tag{40}$$

where $\mathbf{I}$ is the identity matrix and $\sigma_{\mathbf{d}n}^2$ is the estimated data variance at the n-th iteration. The data variance at each iteration $\sigma_{\mathbf{d}n}^2$ can be estimated from the data error $\mathbf{d}_0$–$\mathbf{d}_n$ using

$$\sigma_{\mathbf{d}n}^2 = \frac{1}{n_{\mathbf{d}}}(\mathbf{d}_0 - \mathbf{d}_n)^*(\mathbf{d}_0 - \mathbf{d}_n) \tag{41}$$

where $n_{\mathbf{d}}$ is the number of data points. This formula has the advantage that the data variance will decrease as iterations proceed and the error decreases. The data variance essentially damps the model perturbations evaluated by the least squares inversion formula. Therefore, more damping occurs when the error is large and the model is far from the solution and less damping occurs when the model becomes close to the solution. As the least squares solution is approached, the estimate of $\sigma_{\mathbf{d}n}^2$ should approach the true noise variance. This method decreases the likelihood that the least squares iterations will become unstable due to nonlinearity of the geophysical theory. This is because the damping is largest when the model is far from the solution and the effect of nonlinearity is strongest. Unfortunately, stability is not assured by using this method of choosing data variance but at least it chooses the variance in a qualitatively sensible manner. (This method is due to Dale Morgan of Stanford, pers. comm., 1984)

A generalization of Dale Morgan's method outlined above is due to Jeff Thorson who allowed for spatially variable data variances. In this case the covariance matrix is diagonal but with a non-constant value along the diagonal. Thorson used the smoothed envelope of the error in his estimation of spatially variable data variance. This method would be useful when doing field data inversions in order to suppress the influence from the noisiest regions in the data. For example, regions containing non-Gaussian coherent noise such as ground roll will be treated as areas of Gaussian noise with a large variance.

If the geophysical theory is almost linear, it may be better to simply estimate the noise level N, and use a constant value for $\sigma_{\mathbf{d}n}^2$ as a function of iteration. Then the value of $\sigma_{\mathbf{d}n}^2$ would be

$$\sigma_{\mathbf{d}n}^2 \quad = \quad N \frac{1}{n_{\mathbf{d}}} \mathbf{d}_0{}^* \mathbf{d}_0 \tag{42}$$

Choice of model covariance

It is more difficult to choose model covariances for iterative inversions than data covariances. This is because useful "pseudo" model covariances, which are used for damping, may bear no relation to the a priori values especially for highly nonlinear functions or when statistics are very non-Gaussian. (This is discussed in the section that reviewed least squares inversion.) A good choice of the model covariances should lead to a stable inversion. (This may be considered a definition of a "good choice".) The covariances should be chosen to damp each physical parameter in a sensible way from one iteration to the next. If the model parameters are assumed to be independent then the covariance matrix would be diagonal. Sometimes it makes sense to use non-diagonal covariance matrices. For example, if you desired the perturbations in the model parameters to be smooth, some covariance between adjacent layers could be introduced. Another possibility is the introduction of covariance between different physical parameters. For instance, suppose P velocity was known to correlate well with S velocity. Then introduction of a covariance between P velocity and S velocity may help stabilize or even speed up the inversion.

A useful rule of thumb is that model perturbations from one iteration to the next will tend to be smaller than the standard deviation implied by the choice of covariance matrix provided the data covariances are reasonably accurate. For example, if a diagonal covariance matrix of form $\sigma_{\mathbf{m}}^2 \mathbf{I}$ was chosen, the model perturbations would tend to be less than about $\sigma_{\mathbf{m}}$. (ie: Perturbations greater than $\sigma_{\mathbf{m}}$ would have a low probability of occurrence.)

Summarizing, an inversion would normally begin with some fairly small model variance in order to stabilize the iterations. As the algorithm converges, the model variance should be gradually increased until either the true a priori variance is reached, or any model parameters begin to diverge.

Choice of a background model

The background model is the a priori state of knowledge of the physical model. This may be varied from one least squares iteration to the next in order to achieve a non-Gaussian model solution. (See the discussion in the section which reviews least squares inversion.) A sensible choice of this background model is crucial for stable inversions. The simplest choice for the background model is the previous iteration. A better choice has been found to be a smoothed version of the previous iteration. (This is done in the example given in this paper.) The smoothing is carried out independently for each physical parameter over depth. This avoids buildup of the poorly resolved Nyquist jitter in the model from one iteration to the next. Nyquist jitter in the model is seen as model perturbations that alternate between positive and negative from one layer to the next. The Nyquist frequency for layers is poorly resolved when there is little or no energy in the seismic source wavelet at this frequency. Note that traveltime and amplitude information contained in reflection hyperbolas helps to constrain the low to middle frequency components in the model so these components tend to be fairly well resolved. (In practice, one could simply choose the layer sampling such that the layer properties are fairly well resolved, for instance the 1/2 Nyquist sample interval or the fundamental period.)

EXAMPLE

Synthetic data inversion example

Random properties for a stack of 100 layers were generated using non-Gaussian statistics. The resultant logs for P velocity, S velocity and density are shown as the solid line in figure 1. A synthetic CMP gather was generated using the ray equations summarized by equation (2). An explosive source type and vertical component geophones were used in the forward modeling. All primary ray modes were computed but because the source was explosive, the only non-zero modes were the PP and PS. The synthetic gather is shown in figure 2 with a time varying gain of $t^{1.5}$.

A least squares inversion was performed using equation (39a) and the approximation (39b) with data variances computed assuming 10% noise (ie: $N = .1$ in equation (42)). The inversion was carried out in a contracted data space in order to improve the

efficiency. (See Mora, SEP-41 for details of the method of data-space contraction). The properties of the first layer were fixed at their true values during the inversion. This effectively fixes the ray amplitudes incident upon the second layer. Equivalently, the complex ray directivity function at the top of the second layer was assumed to be known. (This directivity can be evaluated in the inversion if it is unknown thereby avoiding a detailed knowledge of the layers above the zone of interest.) The starting guess used in the iterative inversion was a linear fit for the P velocity, S velocity and density shown as the broken lines in figure 1. After a little experimentation, suitable model variances were found which resulted in a stable inversion. The initial standard deviation used to damp the iterations was .1 km/sec for P velocity, .15 km/sec for S velocity and .05 gm/cc for density. As the iterations proceeded, these values were increased by a factor of ten. (This decreases the damping as the solution is approached.) In order to achieve stability, the density variance was required to be less than the variances used for the P or S velocities which possibly indicates that density is not as well determined as P or S velocity. (These variances may be compared because the units chosen for velocity and density are such that the numerical values of the different properties are similar.) Also, the method of using a smooth pseudo a priori model was employed. The square error, normalized using the data energy, is graphed as a function of iteration in figure 3. The error decreases to about zero by the 10-th iteration. The inversion results after 10 iterations (broken line) are plotted on the same graph as the true model (solid line) in figure 4. From figure 4, is is clear that the inversion recovered the main features of all three logs. As expected, the inverted P velocity log is the most accurate while the inverted density log is the least accurate. The forward modeled wavefield and the error are plotted in figure 5 for several iterations to illustrate the way the error decreases with iteration.

Notes on field data inversion

When inverting field data, it may be necessary to simultaneously invert for complex directivity as a function of ray parameter at some specified depth. This would have the advantage that source directivity as well as all the amplitude effects caused by the overlying layers would be removed. Therefore, problems of a complicated near surface and lack of accurate knowledge of the true source and receiver directivity would be avoided. It may be necessary to include a preprocessing step to remove any surface-consistent, shot-consistent, or offset-consistent amplitude effects from the data to be inverted.

SUMMARY AND CONCLUSIONS

In this paper, extensions have been to the ray scheme of Mora (SEP-38) that performed an iterative inversion of a PP wavefield for the physical properties of plane layers and the source wavelet. The scheme performs an iterative least squares inversion with the Newton algorithm using a Frechet derivative matrix that was computed analytically during the forward ray modeling. The process is an amplitude inversion of the complete CMP wavefield based on the ray theoretical elastic amplitudes of primary reflections. Hence, the amplitude versus offset information is utilized in the inversion. An expected practical drawback in using such an inversion technique is that many factors affecting amplitudes in field data cannot be accurately simulated. If this were a severe problem, the results of field data inversion would be inaccurate. The extensions to the inversion scheme, which are presented in this paper, are:

(1) Inclusion of other primary reflection mode types, PS, SP and SS.

(2) Inclusion of the ray spreading term.

(3) Transformation of the equations from the frequency domain to the time domain in order to speed the convolution of a short source wavelet with a long impulse response.

(4) The ray equations were reformulated for constant vertical P-wave traveltime layers. This decreases nonlinearity by diminishing movement of the dominant PP reflection hyperbolae as P velocity is varied.

(5) Analytic calculation of reflection and transmission coefficient derivatives.

(6) Use of the method of data space contraction (see Mora, SEP-41, Data space contraction in overdetermined inverse problems)

A 16 fold synthetic CMP gather sampled at .008 seconds was computed by performing ray modeling through a random 100 layer model, which had layers sampled at .008 secs of two way P traveltime. An inversion of this synthetic data was carried out for P-velocity, S-velocity and density of the 100 layer model. The source wavelet, a first derivative Gaussian curve, was assumed to be known exactly and so was fixed throughout the inversion. The inversion results for this noise free synthetic example matched fairly well to the true physical model used to generate the data. This verifies a correct implementation and also gives some clues about the resolution of the different parameters. The P-velocity is the best resolved, which is to be expected considering the dominant energy in the data is PP reflections. However, the main features in both the S-velocity and density logs have also been resolved.

ACKNOWLEDGMENTS

Thanks to my SEP colleagues for many interesting discussions. Particularly, thanks to Dan Rothman and John Toldi.

REFERENCES

Mora, P., Data space contraction in overdetermined inverse problems: SEP-41.

Mora, P., A simple geometric derivation of the ray spreading factor: SEP-41.

Mora, P., Inversion of CMP gathers for P and S velocity: SEP-38.

Tarantola, A., and Valette, B., 1982, Generalized nonlinear inverse problems solved using the least squares criterion, Reviews of geophysics and space physics, v. 20, no. 2, p 219-232.

Tarantola, A., 1983, Principles of seismic imagery, preprint.

Tarantola, A., and Valette, B., 1982, Inverse problems = quest for information, Journal of geophysics, v. 50, p. 159-170.

Thorson, J., Velocity stack and slant stack inversion methods: SEP-39.

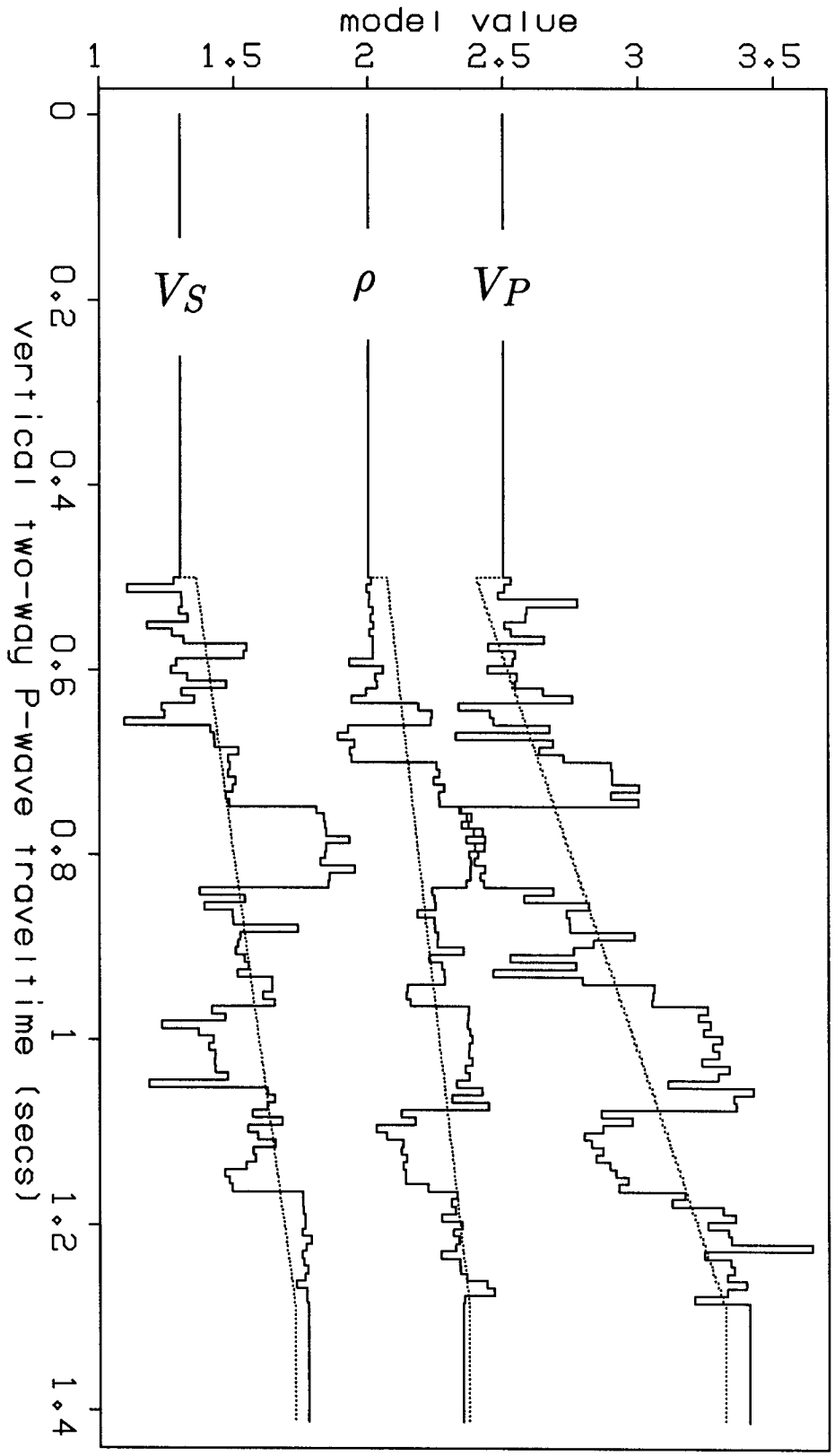

FIG. 1. Random, non-Gaussian logs used to generate the synthetic data (solid line) and the linear fit used as the first guess for the least squares iterative inversion (broken line).

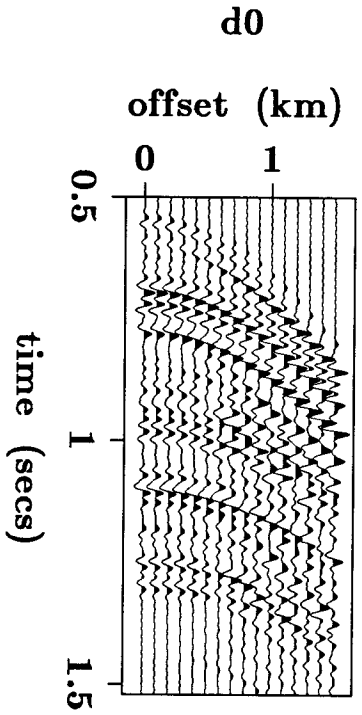

FIG. 2. The synthetic common midpoint gather generated by ray modeling using the random logs of figure 1. The ray modeling includes both PP and PS primary reflections and was computed for an explosive source and vertical component geophones.

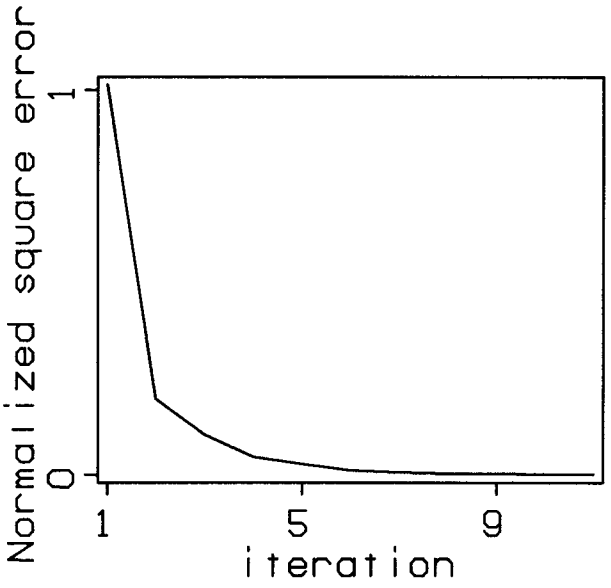

FIG. 3. The square error as a function of iteration normalized to the energy in the synthetic data. The normalized error at the first iteration is approximately one so the initial error has about the same magnitude as the synthetic data. The error decreases to about zero by the 10-th iteration.

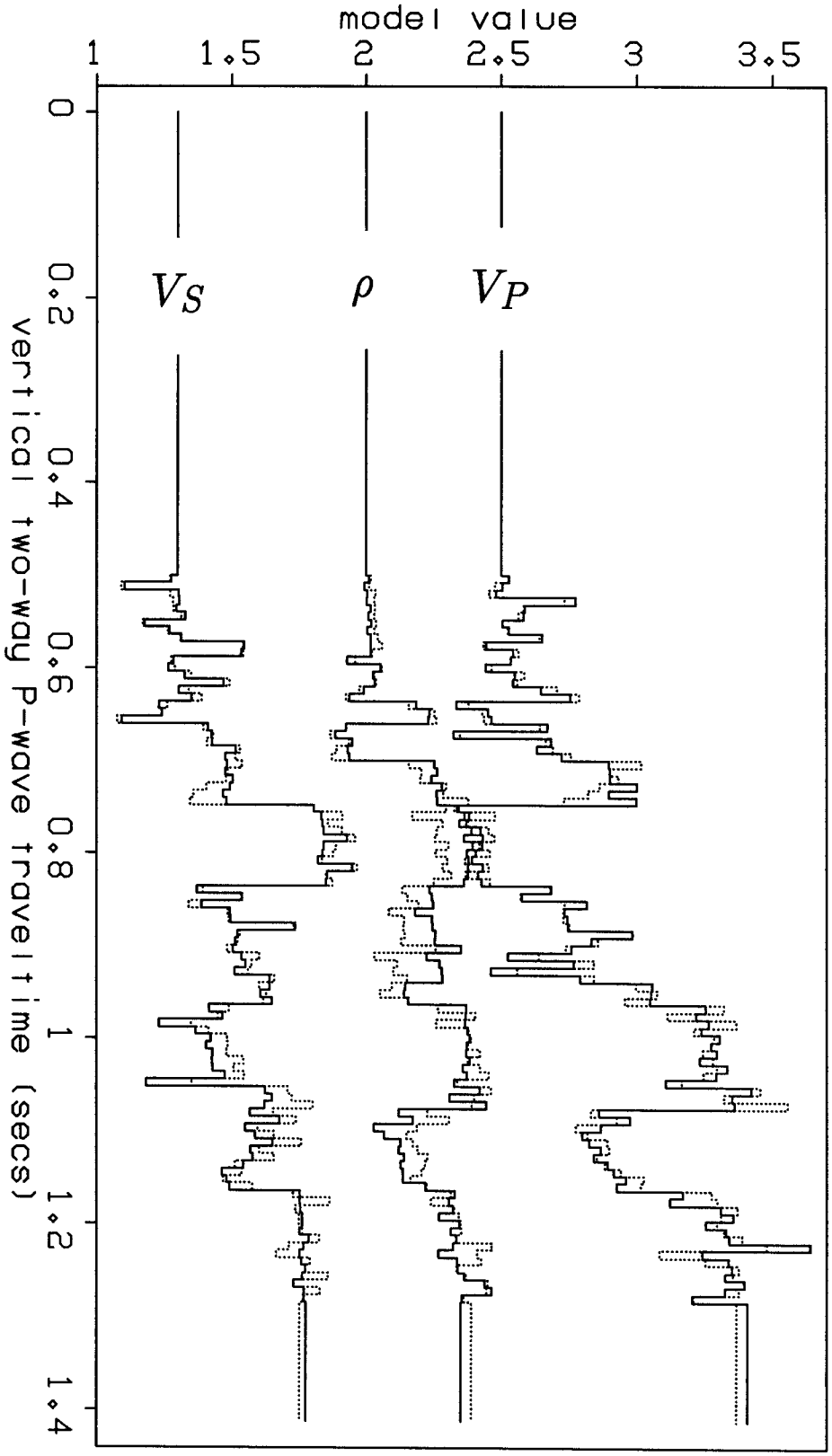

FIG. 4. Random, non-Gaussian logs used to generate the synthetic data (solid line) and the result of the inversion after 10 iterations (broken line).

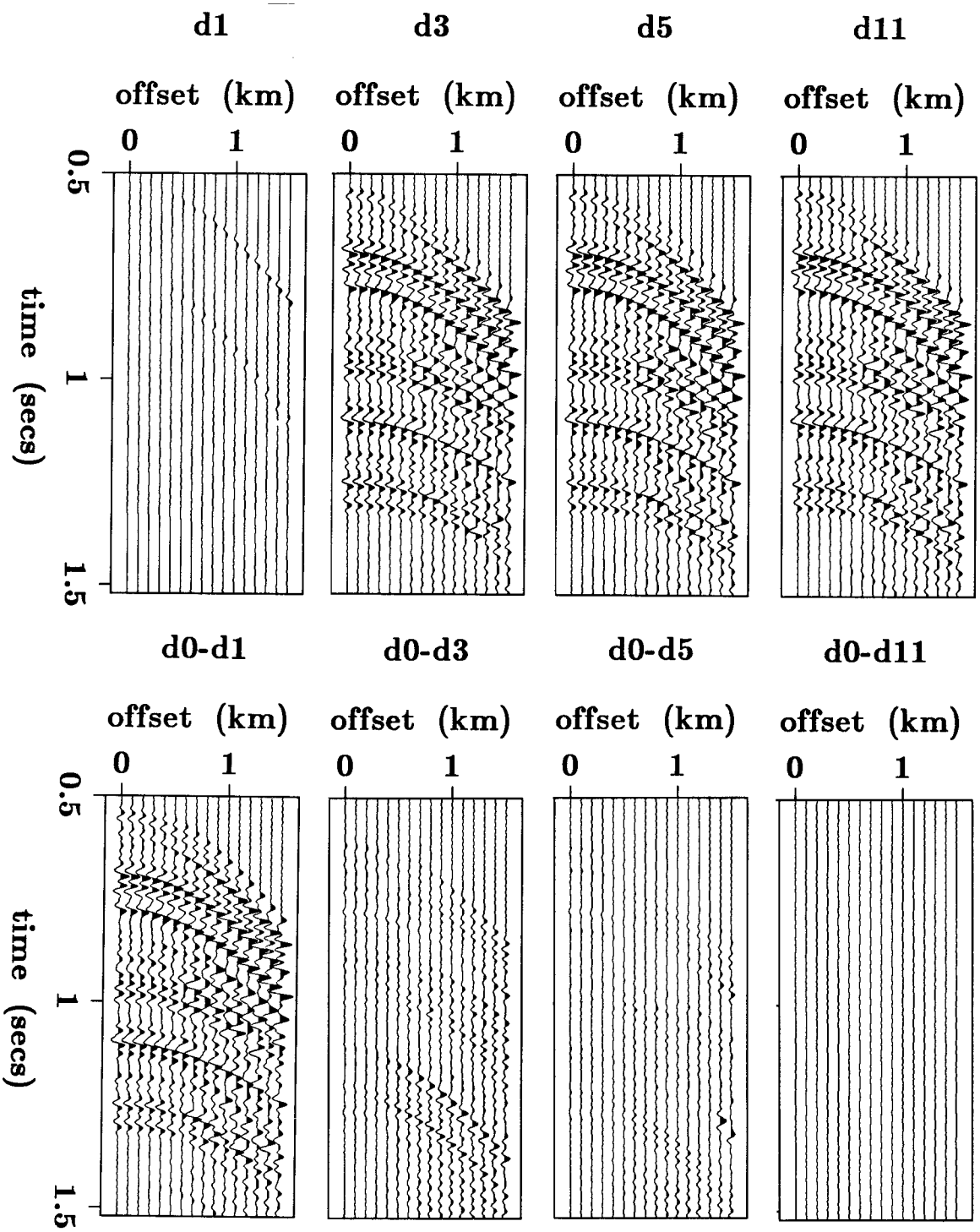

FIG. 5. The modeled gathers $\mathbf{d}_n$ and the error $\mathbf{d}_0$–$\mathbf{d}_n$ at various iterations. After about 10 iterations, the gather generated using the inversion result as the physical model matches the synthetic gather of figure 2. Note that the modeled gather immediately prior to the n-th iteration is denoted $\mathbf{d}_n$ so the gather generated using the initial guess is denoted $\mathbf{d}_1$.


San
Andreas
San
Andreas