Chapter 3

Elastic inversion theory

3.1 Overview

The treatment of multi-offset seismic data as an acoustic wavefield is becoming increasingly disturbing to many geophysicists who see a multitude of wave phenomena, such as amplitude-offset variations and shear wave events, which can only be explained by using the more correct elastic wave equation. Not only are these phenomena ignored by acoustic theory, but they are also treated as undesirable noise when they should be used to provide extra information about the sub-surface such as S-velocity. The problems of using the conventional acoustic wave equation approach can be eliminated by starting afresh with an elastic approach.

In this chapter, equations have been derived to perform an inversion for P-velocity, S-velocity and density as well as the P-impedance, S-impedance and density. These are better resolved than the Lamé parameters. The inversion is based on nonlinear least squares and proceeds by iteratively updating the Earth parameters until a good fit is achieved between the observed data and the modeled data corresponding to these Earth parameters. The iterations are based on the preconditioned conjugate-gradient algorithm. The fundamental requirement of such a least squares algorithm is the gradient direction which tells how to update the model parameters. This can be derived directly from the wave equation and it may be computed by several wave propagations. Although any scheme could in principle be chosen to perform the wave propagations, the elastic finite difference method is used because it directly simulates the elastic wave equation and can handle complex, and thus realistic, distributions of elastic parameters. This method of inversion is costly since it

is similar to an iterative prestack shot-profile elastic migration. However, it has greater power than any migration since it solves for the P-velocity, S-velocity and density and can handle very general situations including transmission problems. Also, both high and low wavenumbers of the P- and S-wave velocities are obtained because both amplitudes and shapes of hyperbolas are matched.

3.2 Philosophy of inversion

3.2.1 Elastic waves or acoustic waves?

Seismic data contain many features such as shear reflections and amplitude-offset variation that provide useful information about the P-velocity, S-velocity and density. These are becoming increasingly evident as the use of longer cables and multi-component recording becomes more widespread. Conventional seismic processing and inversion based on the acoustic wave equation do not take these features into account and they therefore treat this useful information as undesirable coherent noise. To effectively use all the information contained in seismic data one must use the elastic wave equation and build up a new framework (or expand the old one) for the treatment of seismic data. Simply taking the old acoustic algorithms and applying them to P-S and S-S reflections does not use the constraint that the different modes are related through the elastic wave equation. Separate images may be obtained by processing the different modes but they pay no heed to the elastic wave equation. How do the different images relate to physical properties? Once an elastic approach is adopted, the various seismic events usually considered to be noise such as ground roll, refractions, multiples, mode conversions and S-wave events can then be treated as signal and should theoretically be helpful (I include multiples in this list for completeness and because the elastic wave equation handles their amplitudes more correctly than the acoustic wave equation). In principle, almost all seismic "noise" ceases to be bothersome!

3.2.2 Inversion or conventional processing?

Should inversion or standard processing methods be applied to seismic data? Most practicing geophysicists would say that while the philosophy of inversion, to get a quantitative estimate of the physical properties, is appealing, they would prefer to use standard methods such as NMO stack and migration because they do not believe inversion can work well

in practice. Devout inverters would perhaps reply that standard methods are inversion too because they try to obtain a picture of the subsurface. The main difference is that standard methods are one step processes and that they yield a qualitative estimate of the physical properties. For instance, migration gives a relative reflectivity picture and not a quantitative P-velocity or P-impedance picture. In general, standard methods do not pay heed to the exact properties of the Earth but concentrate on obtaining a good image of the subsurface. In practice, the product of conventional processing applied to reflection seismic data would usually be a quantitative estimate of the low-wavenumber part the P-velocities (from velocity analysis) and a relative estimate of the high-wavenumber part of the reflectivities (from migration). Ultimately, it will be shown, that for the case of seismic reflection data, the picture that can be obtained by inversion is similar to the expected result of an iterative simultaneous reflection tomography and migration. The inversion accounts for all elastic waves and is wave equation based. The inversion can also handle transmission data such as VSP and well-to-well data. Unlike migration, the inversion utilizes the information contained in direct waves and reflection traveltime curves to help obtain the interval velocities. The strength of the inversion philosophy is that it tries to account for the seismic data in terms of the Earth properties using the known equations of physics.

The method of least squares is invoked on a grand scale so that by using the elastic wave equation it is possible to make use of all the amplitude information in the seismic data and perform an inversion for the P-velocity, S-velocity and density. The result of the inversion is the most likely set of physical parameters that could have given rise to the observed data (provided any energy not accountable by the elastic wave equation is uncorrelated Gaussian noise).

3.2.3 Linear or nonlinear?

What resolves Earth properties

Inversion is more complicated than the forward problem which attempts to simulate the equations of physics. This is because inversion attempts to solve a problem which is inherently unstable. For instance, in seismic exploration one attempts to predict rock properties based on the appearance of waves which in the past have propagated through some distant rocks. If the effect of the rocks on the waves is not great, then in the presence

of any noise, one cannot hope to predict the properties of the rocks. In other words, it is sensible to only try to predict those properties which can be resolved given the quality of the data. If a parameter is not well resolved then the solution should be restricted in a logical way using a priori knowledge.

What was done and why

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

What it would be nice to do

To advance, I first consider the ideal inversion, namely, mapping out the probability functions of the desired physical properties based on all available information (Tarantola and Valette, 1982). However, even if we were patient enough to wait for these to be calculated, how would such high dimensional probability functions be displayed and understood (typically, the dimension of seismic inverse problems is 10^6 or more). Sometimes the Monte Carlo methods may be utilized to partially map the probability functions or solve for the single most probable set of model parameters (e.g. Rothman, 1985). Unfortunately, such methods require the forward problem to be solved many times. Thus, Monte Carlo methods are too costly for most seismic inverse problems where the forward problem consists of a computationally intensive wave propagation. Hence, in choosing the method, we must compromise between practicality and completeness of the inversion.

What I have chosen to do and why

I have chosen to base the inversion on the relatively fast and tractable least squares optimization method. It does *not* use a linearized wave equation for the solution of the forward problem but rather uses the full elastic wave equation. The iterative least squares algorithm requires a gradient direction which can be derived directly from the wave equation.

The gradient is then the correlation between two wavefields calculated by doing a forward simulation using the wave equation. This principle was used by Tarantola (1984) for elastic waves and Lailly (1984) for acoustic waves, Kolb and Canadas (1986) for acoustic waves in a stratified medium and Gauthier et al. (1986) for acoustic waves in a two dimensional medium.

I will not explicitly calculate the Frechet derivatives because the number of parameters in two-dimensional elastic inversion is too large and hence the derivatives are too costly to compute. Mora (1987a) and McAulay (1985) explicitly obtained Frechet matrices respectively in the elastic and acoustic cases because they were restricted to a one dimensional medium. The Frechet derivative matrix enabled them to use the more rapidly converging Newton algorithm. However, in this thesis I found the Newton method was unnecessary. By understanding the physics of the inverse calculations and taking this into account, I obtain rapid convergence to the inverse solution using the conjugate gradient algorithm.

3.3 Probability theory and least squares

3.3.1 Probabilistic inversion

In general, the process of inversion can be considered as the location of the single most probable set of model parameters $\mathbf{m}$ given some data observations $\mathbf{d}$ and knowledge of the probability distributions of $\mathbf{m}$ and $\mathbf{d}$ (Tarantola and Valette, 1982; Menke, 1984). In the case of multi-offset seismic data inversion for the elastic properties of rocks $\mathbf{m}$, computation of $\mathbf{d}(\mathbf{m})$ (shot gathers etc.) represents elastic wave propagation which consumes large amounts of computer time. Therefore, the inversion schemes based on linear function space theory require relatively few forward problems (wave propagations) to be solved are preferred over the more general Monte Carlo schemes which would require many forward simulations. The reason is that linear function space methods assume local linearizability of $\mathbf{d}(\mathbf{m})$ and so can progress steadily uphill on the probability function until the nearest peak is located. Unfortunately, this may not be the biggest peak (most probable solution). This is the price paid for requiring only a few forward simulations. For tractability, the least squares optimization method, which assumes Gaussian probability distributions for model parameters and data errors (noise), has been chosen. This should often turn out to be a reasonable assumption, at least for the data errors when there are many sources of independent identically distributed noise. The central limit theorem states that the sum

of independent noise tends to be Gaussian distributed. However, the model parameters (i.e. the physical properties of the Earth) are in general non-Gaussianly distributed. This can be at least partially handled, if not rigorously, by allowing the model mean to vary with iteration (see Mora (1987a)) or by modifying the solution at each iteration by using some statistical arguments (Harlan, 1986).

3.3.2 Nonlinear least squares

I begin with a description of nonlinear least squares applied to the seismic inverse problem. The preconditioned conjugate-gradient method of nonlinear least squares has been chosen because of its simplicity yet good convergence properties.

Consider the Gaussian probability density function

$$P(\mathbf{d},\mathbf{m}) \quad = \quad constant \ \exp\left[-\frac{1}{2}\left(\Delta\mathbf{d}^*\mathbf{C}_\mathbf{d}^{-1}\Delta\mathbf{d} \ + \ \Delta\mathbf{m}^*\mathbf{C}_\mathbf{m}^{-1}\Delta\mathbf{m}\right)\right] \ , \qquad (3.1)$$

where $\Delta\mathbf{d} = \mathbf{d} - \mathbf{d}_0 = \mathbf{d}(\mathbf{m}) - \mathbf{d}_0$ is the difference between the synthetic data $\mathbf{d}(\mathbf{m})$ and the the observed seismograms $\mathbf{d}_0$, $\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0$ is the difference between the Earth model $\mathbf{m}$ and the a priori model $\mathbf{m}_0$, and $\mathbf{C_d}$ and $\mathbf{C_m}$ are the covariance matrices for data and model spaces respectively. (Note that * indicates conjugate transpose. Normally the data consists of real values so $^* = {}^T$, unless the inversion is carried out in Fourier space (or some other complex space)). Clearly, the maximum probability solution occurs when the least squares functional

$$S(\mathbf{d},\mathbf{m}) \quad = \quad \Delta\mathbf{d}^*\mathbf{C}_\mathbf{d}^{-1}\Delta\mathbf{d} \ + \ \Delta\mathbf{m}^*\mathbf{C}_\mathbf{m}^{-1}\Delta\mathbf{m} \ , \qquad (3.2)$$

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

In the following development, I have chosen to review the least squares theory in terms of vectors and matrices though this also applies to linear function spaces in the L2 norm. In the continuous case, vectors would become elements of infinite-dimensional vector spaces, and matrices would become linear operators. The vector/matrix description applies to discretized spaces which is more easily translated into computer code. One further point is that the final algorithm will be iterative and so all vectors and matrices in the following section refer to the n-th iteration though subscript n is not given explicitly to avoid cluttering.

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

$$\mathbf{g} = \frac{1}{2}\frac{\partial S}{\partial \mathbf{m}^*} = \mathbf{D}^*\mathbf{C}_\mathbf{d}^{-1}\Delta\mathbf{d} + \mathbf{C}_\mathbf{m}^{-1}\Delta\mathbf{m} \ , \tag{3.3}$$

where $\mathbf{D} = \partial\mathbf{d}/\partial\mathbf{m}$ is the Frechet derivative matrix. Later, $\mathbf{D}$ for the elastic wave equation will be derived as an operator because the derivation is more straightforward. Actually, in this the elastic inversion examples of this thesis with about 10^6 data points and 10^5 model points, it is infeasible to explicitly compute matrix $\mathbf{D}$ but its operation on a vector $\delta\mathbf{m}$ containing perturbations of elastic parameters relative to an Earth model can be achieved by solving the elastic wave equation. I will also show how to achieve the operation of $\mathbf{D}^*$ on a vector $\delta\mathbf{d}$ by similar calculations.

Note that the factor 1/2 in equation 3.3 is introduced to simplify later expressions. One way to solve for a minimum of the least-squares functional is by substituting the linearization for $\mathbf{d}(\mathbf{m})$,

$$\delta\mathbf{d} = \mathbf{d}(\mathbf{m}') - \mathbf{d}(\mathbf{m}) = \frac{\partial\mathbf{d}}{\partial\mathbf{m}}(\mathbf{m}' - \mathbf{m}) = \mathbf{D}\delta\mathbf{m} \ , \tag{3.4}$$

into equation (3.3) and solving $\mathbf{g} = 0$. This leads to the Newton algorithm,

$$\mathbf{m}' = \mathbf{m} - \mathbf{H}\mathbf{g} \ , \tag{3.5}$$

where

$$\mathbf{H} = \left(\mathbf{D}^*\mathbf{C}_\mathbf{d}^{-1}\mathbf{D} + \mathbf{C}_\mathbf{m}^{-1}\right)^{-1} \ , \tag{3.6}$$

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

I will show that the gradient direction $\mathbf{g}$ can be calculated by only two forward modeling runs in very general situations (i.e. elastic waves, exotic survey specifications and complex geologic models). Then, approximating the inverse Hessian $\mathbf{H}$ by $\hat{\mathbf{H}} = \eta \ diag\{\mathbf{S}\}\,\mathbf{C}_\mathbf{m}$ (where η is called the steplength and this choice is motivated below) which is to say that the posteriori covariance matrix is approximately a space variable scale factor multiplied by the the prior covariance matrix, we arrive at a preconditioned gradient-type iterative least squares formula,

$$\mathbf{m}' = \mathbf{m} - \hat{\mathbf{H}}\mathbf{g} = \mathbf{m} - \eta \ diag\{\mathbf{S}\}\,\mathbf{C}_\mathbf{m} \ . \tag{3.7}$$

Common choices for the approximate Hessian are:

1. $\eta\mathbf{I}$: This choice of $\hat{\mathbf{H}}$ is the simplest but leads to a slow rate of convergence of equation 3.7 if some of the model parameters have different physical units (such as in elastic inversion for different elastic parameters).

2. $\eta\mathbf{C_m}$: rescales the gradient to be a perturbation having the appropriate physical units avoiding the slow convergence of choice 1. It is adequate provided most of the model parameters are equally well resolved.

3. η *diag* $\{\mathbf{S}\}\,\mathbf{C_m}$: better rescaling than 2 because it essentially allows for how well the properties are resolved at different spatial locations due varying seismic illumination (e.g. the Earth properties are better resolved in areas illuminated by high amplitude seismic waves than in areas illuminated by low amplitude seismic waves). The *diag*$\{\mathbf{S}\}$ is a slowly varying scale factor.

4. $diag\left\{\left(\mathbf{C_m}\mathbf{D}^*\mathbf{C_d}^{-1}\mathbf{D}+\mathbf{I}\right)^{-1}\right\}\mathbf{C_m}$: better rescaling than 3 because it more accurately allows for how well the properties are resolved at different spatial locations.

5. $band\left\{\left(\mathbf{C_m}\mathbf{D}^*\mathbf{C_d}^{-1}\mathbf{D}+\mathbf{I}\right)^{-1}\right\}\mathbf{C_m}$: better than 4 because it essentially does a local spatial deconvolution unraveling apparent spatial correlations (e.g. due to band-limitations of the seismic source wavelet).

6. $\left(\mathbf{D}^*\mathbf{C_d}^{-1}\mathbf{D}+\mathbf{C_m}^{-1}\right)^{-1}$: exact inverse Hessian.

Note that I never explicitly evaluate matrix $\mathbf{D}$ in this thesis but I can compute $\mathbf{D}$ operated on an Earth model by solving the elastic wave equation using finite differences. In principle, $\mathbf{D}^*\mathbf{C_d}^{-1}\mathbf{D}$ could be computed by n_m such operations but in the examples, the number of Earth model parameters n_m is about 10^5 and finite differences is slow so choices 4 through 6 above are computationally infeasible. Therefore, I use choice 3 which at least converts the gradient units into physical units such as velocity or density and does a space variable scaling thereby speeding convergence relative to choice 1.

If this algorithm (equation 3.7) is iteratively applied, the solution converges to the same as that of equation (3.5) so the true Hessian is not necessary although the suggested approximate Hessian is useful to speed convergence. For nonlinear functions $\mathbf{d}(\mathbf{m})$, where many local minima may exist (such as in the seismic problem), the minimum located by the least-squares iterations may not be the global minimum of the error functional. Whether

or not the global minima is located depends on the proximity of the starting point on the functional surface to the global minimum. Local minima are largely the result of band and aperture limitations in the seismic data so "cycle skips" are possible. Cycle skips occur when the kinematics implied by the starting model are so bad that the traveltime of wavefronts are incorrect by about a cycle. However, it will be later shown that the inversion algorithm can converge on gross variations in velocity. Thus, it may be possible to escape from the cycle skip kind of local minima starting from the top of the model and progressing down as iterations proceed if good quality events are found at many levels on the seismograms.

3.3.3 Preconditioning

Often, by using some a priori knowledge or constraint, it is possible to modify the gradient direction $\mathbf{g}$ in some sensible way thus greatly speeding convergence. This may also help avoid local minima and resolve the null space problem. Another way to modify the gradient would be to derive a spatial filter corresponding to the true inverse Hessian $\mathbf{H}$. This would (1) perform a spatial deconvolution to make the spatial spectrum comparable to the a priori spectrum; (2) perform a model parameter deconvolution to resolve between the different physical properties (P-wave velocity, S-wave velocity and density); and (3) do a gradually varying spatial scaling to allow for illumination of the different parts of the Earth model by different amounts of seismic energy. However, this would not be easy because it would be a model-dependent operator that is spatially variable. In general, a modification to the gradient is termed preconditioning and it need not be a linear operation such as in equation (3.7) (see for example Harlan (1986)). Let the new preconditioned gradient be denoted

$$\mathbf{p} = \mathbf{P}(\mathbf{g}) \ . \tag{3.8}$$

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

$$\mathbf{m}' = \mathbf{m} - \eta\mathbf{p} \ . \tag{3.9}$$

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

$$\mathbf{d}' = \mathbf{d}(\mathbf{m}') = \mathbf{d} - \eta\mathbf{D}\mathbf{p} \ . \tag{3.10}$$

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

$$\min_{\eta} S' = \min_{\eta} \left[(\mathbf{d}' - \mathbf{d}_0)^* \mathbf{C}_\mathbf{d}^{-1} (\mathbf{d}' - \mathbf{d}_0) + (\mathbf{m}' - \mathbf{m}_0)^* \mathbf{C}_\mathbf{m}^{-1} (\mathbf{m}' - \mathbf{m}_0) \right] . \quad (3.11)$$

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

$$\frac{\partial S'}{\partial \eta} = -2 \left[\mathbf{p}^* \mathbf{D}^* \mathbf{C}_\mathbf{d}^{-1} (\Delta \mathbf{d} - \eta \mathbf{D}\mathbf{p}) + \mathbf{p}^* \mathbf{C}_\mathbf{m}^{-1} (\Delta \mathbf{m} - \eta \mathbf{p}) \right] . \quad (3.12)$$

Finally, by setting this derivative to zero we obtain the solution for η

$$\eta = \frac{\mathbf{p}^* \mathbf{g}}{\mathbf{p}^* \mathbf{D}^* \mathbf{C}_\mathbf{d}^{-1} \mathbf{D}\mathbf{p} + \mathbf{p}^* \mathbf{C}_\mathbf{m}^{-1} \mathbf{p}} . \quad (3.13)$$

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

3.3.4 Conjugate directions

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

$$\mathbf{c}_n = \mathbf{p}_n + \frac{\mathbf{p}_n^* (\mathbf{g}_n - \mathbf{g}_{n-1})}{\mathbf{p}_{n-1}^* \mathbf{g}_{n-1}} \mathbf{c}_{n-1} . \quad (3.14)$$

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

3.3.5 Inversion algorithm

The algorithm for nonlinear least squares of seismic amplitude data by the preconditioned conjugate-gradient method is

$$
\begin{aligned}
&\textit{for } n = 1 \textit{ to } \infty \quad \{ \\
&\quad \mathbf{d}_n = \mathbf{d}(\mathbf{m}_n) \quad , && \textit{data calculation} \\
&\quad \Delta\mathbf{d}_n = \mathbf{d}_n - \mathbf{d}_0 \ , \ \Delta\mathbf{m}_n = \mathbf{m}_n - \mathbf{m}_0 \quad , && \textit{compute residuals} \\
&\quad S(\mathbf{d},\mathbf{m}) = \Delta\mathbf{d}^* \mathbf{C}_\mathbf{d}^{-1} \Delta\mathbf{d} + \Delta\mathbf{m}_n^* \mathbf{C}_\mathbf{m}^{-1} \Delta\mathbf{m}_n \quad , && \textit{square error functional} \\
&\quad \textit{exit if converged} \quad , \\
&\quad \mathbf{g}_n = \mathbf{D}_n^* \mathbf{C}_\mathbf{d}^{-1} \Delta\mathbf{d} + \mathbf{C}_\mathbf{m}^{-1} \Delta\mathbf{m}_n \quad , && \textit{gradient} \\
&\quad \mathbf{p}_n = \mathbf{P}(\mathbf{g}_n) \approx \mathbf{C}_\mathbf{m} \mathbf{g}_n \quad , && \textit{preconditioning} \\
&\quad \mathbf{c}_n = \mathbf{p}_n + \frac{\mathbf{p}_n^*(\mathbf{g}_n - \mathbf{g}_{n-1})}{\mathbf{p}_n^* \mathbf{g}_n} \mathbf{c}_{n-1} \ , \ \mathbf{c}_1 = \mathbf{p}_1 \quad , && \textit{conjugate direction} \\
&\quad \eta_n = \frac{\mathbf{c}_n^* \mathbf{g}_n}{\mathbf{c}_n^* \mathbf{D}_n^* \mathbf{C}_\mathbf{d}^{-1} \mathbf{D}_n \mathbf{c}_n + \mathbf{c}_n^* \mathbf{C}_\mathbf{m}^{-1} \mathbf{c}_n} \quad , && \textit{calculate steplength} \\
&\quad \mathbf{m}_{n+1} = \mathbf{m}_n - \eta_n \mathbf{c}_n \quad , && \textit{update model} \\
&\} \quad .
\end{aligned}
\tag{3.15}
$$

In equation (3.15), n starts at 1 because I used subscript 0 to indicate the a priori model $\mathbf{m}_0$ and field data $\mathbf{d}_0$.

3.3.6 Computational aspects for nonlinear functions

It is shown in the next section that the operation of $\mathbf{D}_n^*$ on a vector can be achieved by a forward simulation (wave propagation) so the conjugate gradient equations may be calculated with three forward simulations, one to compute the synthetic data $\mathbf{d}_n = \mathbf{d}(\mathbf{m}_n)$, one to make the gradient as described in the next section (i.e. to compute the $\mathbf{D}_n^* \mathbf{C}_\mathbf{d}^{-1} \Delta\mathbf{d}_n$ term of $\mathbf{g}_n$), and one to compute the steplength η (i.e. the $\mathbf{D}_n \mathbf{c}_n$ term). All other computations are simple dot products. If a line search is required, then several more forward simulations may be necessary in order to optimize the steplength η. In the examples in this thesis, I carried out such a line search for the optimal η which tended to be within a factor of two of that predicted by equation (3.13). Because the predicted and optimal values of η were so close, only two additional forward simulations were required to

do the line search. For a particular problem it may be found that the optimal η is always approximately the same so a line search may not always be necessary.

There is a complication in the computation of η. Because the elastic wave equation is nonlinear and we never explicitly obtain the Frechet derivatives $\mathbf{D}_n$, we cannot compute $\mathbf{D}_n c_n$ directly. Instead, it must be computed using the forward problem $\mathbf{d}(\mathbf{m})$ with the following formula

$$\mathbf{D}_n \xi \mathbf{c}_n \;=\; \mathbf{D}_n(\mathbf{m}_n + \xi \mathbf{c}_n) - \mathbf{D}_n \mathbf{m}_n \;=\; \mathbf{d}(\mathbf{m}_n + \xi \mathbf{c}_n) - \mathbf{d}(\mathbf{m}_n) \;\;, \tag{3.16}$$

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

$$\eta_n \;=\; \frac{\mathbf{c}_n^* \mathbf{g}_n}{\frac{1}{\xi^2}\left[\mathbf{c}_n^* \xi \mathbf{D}_n^* \mathbf{C}_\mathbf{d}^{-1} \mathbf{D}_n \xi \mathbf{c}\right] + \mathbf{c}_n^* \mathbf{C}_\mathbf{m}^{-1} \mathbf{c}_n} \;\;, \tag{3.17}$$

where $\mathbf{D}_n \xi \mathbf{c}_n$ is computed using equation (3.16).

A further complication is the covariance matrices. If these are assumed to be arbitrary they would require excessive computer storage and CPU time so in practice diagonal matrices are taken. Though diagonal, the covariance matrices need not be constant but may vary along the diagonal which assumes independent but non-constant noise and independent model parameters. In order that this assumption be reasonable, it may be necessary to carefully choose the model parameters (and perhaps data parameters). A viable alternative to the choice of diagonal covariance matrices is when $\mathbf{C}_\mathbf{d}^{-1}$ and $\mathbf{C}_\mathbf{m}^{-1}$ banded so they would represent spatially variable filtering operations or, if the bands were constant, space invariant filtering. Later, I will specify the exact diagonal forms of the covariance matrices I used in the calculations in this thesis.

3.4 Inversion achieved by modeling

3.4.1 Overview

Seismic modeling is well understood. In this section I show how to achieve an inversion by doing seismic modeling a few times. Complicated and unstable inversion formulas are not required, only a method to perform modeling (i.e. solve the equations of physics describing elastic wave propagation).

3.4.2 What is the adjoint operation?

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

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

$$\begin{aligned} \rho\partial_{tt}u_i - \partial_j c_{ijkl}\partial_l u_k &= f_i \quad , \\ c_{ijkl}\partial_l u_k n_j &= T_i \quad , \\ u_i = 0 \quad , \quad t < 0 \quad , \\ \dot{u}_i = 0 \quad , \quad t < 0 \quad , \end{aligned} \tag{3.18}$$

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

$$d(\mathbf{x}_S, \mathbf{x}_R, t, i, \mathbf{m}) = u_i(\mathbf{x}_S, \mathbf{x}_R, t, \mathbf{m}) \quad , \quad \textit{where } i = 1, 2, 3 \quad , \tag{3.19}$$

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

In order to perform inversion using the conjugate-gradient algorithm (equations (3.15) and (3.17)), the gradient $\mathbf{g}$ (see equation (3.3)) corresponding to model parameters $\mathbf{m}$ is required. For the elastic wave equation, the most obvious choices of model parameters $\mathbf{m}$ are the Hooke tensor elements c_{ijkl}, the density ρ, the body force f_i and the traction T_i (for an isotropic medium the Hooke tensor can be described by the two Lamé parameters, λ and μ). However, there are other choices of model parameters which are more experimentally observable and (as it turns out) better resolved (Tarantola et al., 1985). For example, traveltimes depend on the seismic velocities and reflection amplitudes depend on impedance contrasts. This suggests that good model parameter choices are the P- and S-wave velocities and density or the P- and S-wave impedances and density as

model parameters. The gradients in terms of the different model parameter choices are related, so the development will proceed by deriving the gradient in terms of the simplest choice (λ, μ and ρ). Subsequently, the expressions for the gradients in terms of the other model parameter choices will be derived and cast in terms of the gradient for λ, μ and ρ. This is not equivalent to inverting for λ, μ and ρ followed by transformation to the other parameters because different choices of model parameters affect the shape of the objective function. A poorer choice leads to more elongate minima; consequently, more iterations of the gradient algorithm are required to achieve convergence.

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

$$\delta \mathbf{d} \quad = \quad \mathbf{D}\delta\mathbf{m} \ . \tag{3.20}$$

In continuous form,

$$\delta \mathbf{d}(D) \quad = \quad \int_M dM \frac{\partial \mathbf{d}}{\partial \mathbf{m}} \frac{(D)}{(M)} \delta \mathbf{m}(M) \ , \tag{3.21}$$

where M indicates the model space. This expression indicates how to calculate a small perturbation in the wavefield $\delta\mathbf{d}$ resulting from a small perturbation in the model parameters $\delta\mathbf{m}$ by integrating over the model space, i.e. it is the linearized Green function representation of the forward problem or first order Born approximation. With a representation equivalent to equation (3.21), one can identify the Frechet kernel $\mathbf{D} = \partial\mathbf{d}/\partial\mathbf{m}$ and then compute the adjoint operation (see also Tarantola, 1987)

$$\delta \hat{\mathbf{m}} \quad = \quad \mathbf{D}^* \delta \mathbf{d} \ , \tag{3.22}$$

or, in continuous form

$$\delta \hat{\mathbf{m}}(M) \quad = \quad \int_D dD \left(\frac{\partial \mathbf{d}}{\partial \mathbf{m}} \frac{(D)}{(M)} \right)^* \delta \mathbf{d}(D) \ . \tag{3.23}$$

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

Once expression (3.23) has been derived, $\delta\mathbf{d}$ can be replaced by $\mathbf{C}_\mathbf{d}^{-1}\Delta\mathbf{d}$ and added to $\mathbf{C}_\mathbf{m}^{-1}\Delta\mathbf{m}$ to obtain the desired gradient

$$\mathbf{g} \quad = \quad \mathbf{D}^* \mathbf{C}_\mathbf{d}^{-1} \Delta \mathbf{d} \ + \ \mathbf{C}_\mathbf{m}^{-1} \Delta \mathbf{m} \ . \tag{3.24}$$

3.4.3 What is the adjoint operation for the seismic problem?

For the seismic problem, the linearized forward problem equivalent to equation (3.21) is of form

$$\delta u_i(\mathbf{x}_S, \mathbf{x}_R, t) = \int_V dV(\mathbf{x}) \frac{\partial u_i(\mathbf{x}_S, \mathbf{x}_R, t)}{\partial \mathbf{m}(\mathbf{x})} \delta \mathbf{m}(\mathbf{x}) \ , \tag{3.25}$$

where $u_i(\mathbf{x}_S, \mathbf{x}_R, t)$ represents the seismogram located at receiver location $\mathbf{x}_R$ which records the i-th component of displacement of an elastic wavefield due to a shot at $\mathbf{x}_S$ and $\mathbf{m}(\mathbf{x}) = [c_{1111}(\mathbf{x})...c_{ijkl}(\mathbf{x})...c_{3333}(\mathbf{x}), \rho(\mathbf{x}), f_1(\mathbf{x})...f_3(\mathbf{x}), T_1(\mathbf{x})...T_3(\mathbf{x})]^T$ at location $\mathbf{x}$ in the Earth. Therefore, the adjoint problem corresponding to equation (3.23) is

$$\delta \hat{\mathbf{m}}(\mathbf{x}) = \sum_S \int dt \sum_R \left(\frac{\partial u_i(\mathbf{x}_S, \mathbf{x}_R, t)}{\partial \mathbf{m}(\mathbf{x})} \right)^* \delta u_i(\mathbf{x}_S, \mathbf{x}_R, t) \ , \tag{3.26}$$

i.e. the integral over the data space of the data residuals multiplied by the Frechet kernel. From the above two expressions, it is clear that the only requirement to define the adjoint operation (equation 3.26) is the integral expression (equation 3.25) for the forward problem which gives the perturbation in displacement δu_i corresponding to some perturbations in the model parameters $\delta \mathbf{m}$. This expression would define the Frechet kernel so the adjoint operation (3.26) could then be written by identifying terms in equations (3.25) and (3.26). The linearized solution of the elastic wave equation in terms of Green's functions (the first order Born approximation) supplies the appropriate integral.

3.4.4 What is the linearized elastic forward problem?

To avoid cluttering the equations with the shot location $\mathbf{x}_S$, the following development refers to a single shot. Therefore, the required adjoint operation can be obtained from the result of this development by simply including the sum over shots. Now consider the elastic wave equation (3.18). The integral corresponding to equation (3.25) can be found by making the following substitutions:

$$\begin{aligned}
u_i &\rightarrow u_i + \delta u_i \ , \\
\rho &\rightarrow \rho + \delta \rho \ , \\
c_{ijkl} &\rightarrow c_{ijkl} + \delta c_{ijkl} \ , \\
f_i &\rightarrow f_i + \delta f_i \ , \\
T_i &\rightarrow T_i + \delta T_i \ .
\end{aligned} \tag{3.27}$$

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

$$\begin{aligned} \rho\partial_{tt}\delta u_i - \partial_j c_{ijkl}\partial_l \delta u_k &= \Delta f_i \quad , \\ c_{ijkl}\partial_l \delta u_k n_j &= \Delta T_i \quad , \\ \delta u_i = 0 \quad , \quad t < 0 \quad , \\ \dot{\delta} u_i = 0 \quad , \quad t < 0 \quad , \end{aligned} \tag{3.28}$$

where the new force and traction terms are

$$\Delta f_i = \delta f_i - \delta\rho\partial_{tt}u_i + \partial_j \delta c_{ijkl}\partial_l u_k + O_i(\delta\rho, \delta c_{ijkl}, \delta f_i, \delta T_i)^2 \ , \tag{3.29}$$

and

$$\Delta T_i = \delta T_i - \delta c_{ijkl}\partial_l u_k n_j + O_i(\delta\rho, \delta c_{ijkl}, \delta f_i, \delta T_i)^2 \ . \tag{3.30}$$

The new wave equation given in equation (3.28) has the same form as the elastic wave equation and hence its solution can be obtained in terms of the Green's functions of the elastic wave equation. The solution in terms of the elastic Green's functions (Aki and Richards, 1980) is

$$\begin{aligned} \delta u_i(\mathbf{x}_R, t) &= \int_V dV(\mathbf{x})\ G_{ij}(\mathbf{x}_R, t; \mathbf{x}, 0) * \Delta f_j(\mathbf{x}, t) \\ &+ \int_S dS(\mathbf{x})\ G_{ij}(\mathbf{x}_R, t; \mathbf{x}, 0) * \Delta T_j(\mathbf{x}, t) \ . \end{aligned} \tag{3.31}$$

This equation has the form of the desired expression for the forward problem, equation (3.25), and so it defines the Frechet kernel $\partial u_i(\mathbf{x}_R, t)/\partial \mathbf{m}(\mathbf{x})$ from which one may obtain the adjoint expression, equation (3.26). A simplified adjoint expression can be obtained with some algebra.

Substituting the force and traction terms given in equations (3.29) and (3.30) into equation (3.31), neglecting the O^2 terms (i.e. assume small perturbations in order to obtain the Frechet derivatives), dropping the arguments of the various functions to avoid cluttering, and using notation $u_{m,l} = \partial_l u_m$ yields

$$\delta u_i = \int_V dV\ G_{ij} * [\delta f_j - \delta\rho\ \partial_{tt}u_j + \partial_k \delta c_{jklm} u_{l,m}] + \int_S dS\ G_{ij} * [\ \delta T_j - \delta c_{jklm} u_{l,m} n_k] \ . \tag{3.32}$$

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

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

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

(iii) $\partial_k [G_{ij} * (\delta c_{jklm} u_{l,m})] = G_{ij} * (\partial_k \delta c_{jklm} u_{l,m}) + (\partial_k G_{ij}) * (\delta c_{jklm} u_{l,m})$, by the chain rule.

Applying these results to equation (3.32) yields the formula

$$\delta u_i = \int_V dV\ G_{ij} * \delta f_j + \int_S dS\ G_{ij} * \delta T_j$$

$$- \int_V dV \dot{G}_{ij} * \dot{u}_j \delta\rho - \int_V dV\ (\partial_k G_{ij}) * (\delta c_{jklm} u_{l,m}) \quad . \tag{3.33}$$

Finally, assuming isotropy (if anisotropic inversions are desired, this assumption can be dropped from the development),

$$\delta c_{jklm} = \delta\lambda \delta_{jk} \delta_{lm} + \delta\mu(\delta_{jm}\delta_{kl} + \delta_{jl}\delta_{km}) \quad , \tag{3.34}$$

and therefore

$$\delta u_i = \int_V dV\ G_{ij} * \delta f_j + \int_S dS\ G_{ij} * \delta T_j - \int_V dV \dot{G}_{ij} * \dot{u}_j \delta\rho -$$

$$\int_V dV\ G_{ij,j} * u_{m,m} \delta\lambda - \int_V dV\ G_{ij,k} * (u_{j,k} + u_{k,j}) \delta\mu \quad . \tag{3.35}$$

This equation has exactly the same form as equation (3.25), so clearly it defines the Frechet kernel $\partial u / \partial \mathbf{m}$ for model parameters ρ, λ, μ, f_j and T_j. Use of this equation to solve the forward problem is referred to as the Born approximation. Note that the full elastic wave equation (3.18) is used to solve the forward problem in this thesis rather than the Born approximation (the linearized forward problem). The linearized forward problem is only specified in order to derive the adjoint operation.

3.4.5 What is the elastic adjoint operation?

Integrating the Frechet kernel defined by equation (3.35) over the data space (see equation (3.26)) produces the adjoint operation in terms of the Lamé moduli

$\delta\hat{\mathbf{m}} = \left[\delta\hat{\rho}\ ,\ \delta\hat{\lambda}\ ,\ \delta\hat{\mu}\ ,\ (\delta\hat{f}_j, j = 1,2,3)\ ,\ (\delta\hat{T}_j, j = 1,2,3)\right]^T$ (see also Tarantola (1984)), where

$$\delta\hat{\rho} = -\int dt \sum_R \dot{G}_{ij} * \dot{u}_j \delta u_i \ , \tag{3.36}$$

$$\delta\hat{\lambda} = -\int dt \sum_R G_{ij,j} * u_{m,m} \delta u_i \ , \tag{3.37}$$

$$\begin{aligned} \delta\hat{\mu} &= -\int dt \sum_R G_{ij,k} * (u_{j,k} + u_{k,j}) \delta u_i \\ &= -\int dt \sum_R \frac{1}{\sqrt{2}} (G_{ij,k} + G_{ik,j}) * \frac{1}{\sqrt{2}} (u_{j,k} + u_{k,j}) \, \delta u_i \ , \end{aligned} \tag{3.38}$$

$$\delta\hat{f}_j = \int dt \sum_R G_{ij} * \delta u_i \ , \tag{3.39}$$

$$\delta\hat{T}_j = \int dt \sum_R G_{ij} * \delta u_i \ . \tag{3.40}$$

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

$$\int dt\ f(t) * g(t) h(t) = \int dt\ f(-t) g(t) * h(-t) \ . \tag{3.41}$$

The resulting integrals corresponding to equation (3.26) for the adjoint in terms of λ, μ and ρ have the form

$$\begin{aligned} \delta\hat{\gamma}(\mathbf{x}) &= \sum_S \int dt \sum_R \left(\Omega^{\gamma}_{ijk} u_j(\mathbf{x}_S, \mathbf{x}, -t)\right) \left(\Omega^{\gamma}_{ijk} \left[G_{ij}(\mathbf{x}_R, t; \mathbf{x}, 0) * \delta u_i(\mathbf{x}_S, \mathbf{x}_R, -t)\right]\right) \\ &= \sum_S \int dt \left(\Omega^{\gamma}_{ijk} u_j(\mathbf{x}_S, \mathbf{x}, t)\right) \left(\Omega^{\gamma}_{ijk} \left[\sum_R G_{ij}(\mathbf{x}, -t; \mathbf{x}_R, 0) * \delta u_i(\mathbf{x}_S, \mathbf{x}_R, t)\right]\right) \\ &= \sum_S \int dt \left(\Omega^{\gamma}_{ijk} u_j(\mathbf{x}_S, \mathbf{x}, t)\right) \left(\Omega^{\gamma}_{ijk} \psi_j(\mathbf{x}_S, \mathbf{x}, t)\right) \ . \end{aligned} \tag{3.42}$$

where the shot sum $\sum_S$ dropped earlier for convenience is reintroduced, Ω^{γ}_{ijk} is an operator that is dependent on the physical property type denoted $\gamma = \lambda$, μ, or ρ, u_j is the background wavefield computed using Earth model $\mathbf{m}$, and ψ_j is called the "back propagated residual wavefield" and is defined by equation 3.43. This is the general form of the adjoint operation for any choice of Earth properties γ. The development so far has been for an

isotropic solid with Earth properties specified as the Lamé parameters so $\gamma = \lambda$, μ or ρ. Later I will extend these formulas to some more experimentally observable choices of model parameters such as the compressional and shear wavespeeds and density. Note that the implied sums inside ()'s or []'s must be performed first in equations 3.42 and 3.43 and the derivation made use of the reciprocity property of the elastic Green's functions, (i.e. interchangeability of $\mathbf{x}$ with $\mathbf{x}_R$), and the property $\int dt\ f(-t)g(-t) = \int dt\ f(t)g(t)$. The "back propagated residual wavefield" is computed by applying the data residuals as a forcing function in the elastic wave equation but backwards in time, i.e.

$$\psi_j(\mathbf{x}_S, \mathbf{x}, t) = \sum_R G_{ij}(\mathbf{x}, -t; \mathbf{x}_R, 0) * \delta u_i(\mathbf{x}_S, \mathbf{x}_R, t) \ . \tag{3.43}$$

The operator Ω^{γ}_{ijk} is dependent on the type of model parameter γ at each location $\mathbf{x}$. From equations (3.36) through (3.40) we have

$$\Omega^{\rho}_{ijk} = \sqrt{-1}\,\partial_t \ , \ \Omega^{\lambda}_{ijk} = \sqrt{-1}\,\partial_j \ , \ \Omega^{\mu}_{ijk} = \sqrt{-1/2}\,(\delta_{jk}\partial_i + \delta_{ji}\partial_k) \ . \tag{3.44}$$

The meaning of this operator is discussed in detail in the section that compares inversion to migration. The simple form of the adjoint expression given by equations (3.42) and (3.43) makes it easy to compare inversion to migration.

3.4.6 Interpreting the adjoint

Consider equation (3.42) which is the adjoint to the elastic wave equation. There are only two unknowns, the wavefields u_j and ψ_j. The wavefield u_j can be computed by forward modeling (i.e. doing a shot simulation). Now consider the definition of ψ_j given by equation (3.43). Firstly, note that the data perturbation δu_i (denoted the *data residuals*) is the difference between the data u_i and the observed seismograms u_{i0}. Now, observe that the general form of equation (3.42) is the same as that of equation (3.31) except that the Green's function G_{ij} in equation (3.43) has a minus sign in front of the time variable t. This indicates that here the Green's function corresponds to propagation *backwards* in time. Therefore, ψ_j is generated by applying δu_i as a forcing term in a wave equation that runs backwards in time and is therefore called the *back-propagated residual wavefield* (c.f. back projection in tomography (Devaney, 1984)). Since the wave equation denoted $E(\mathbf{x}, t)u_i(\mathbf{x}, t) = f_i(\mathbf{x}, t)$ has a time symmetric operator (i.e. $E(\mathbf{x}, t) = E(\mathbf{x}, -t)$), this back propagation can be achieved by forward modeling using a time reversed forcing function (i.e. solving $E(\mathbf{x}, t)u_i(\mathbf{x}, t) = f_i(\mathbf{x}, -t)$).

From this interpretation, it is clear that the adjoint operation corresponds to:

(i) Propagate elastic waves through Earth model $\mathbf{m}$ using equation (3.18) to solve for the background wavefield $u_j(\mathbf{x}_S, \mathbf{x}, t)$ and synthetic seismograms $u_i(\mathbf{x}_S, \mathbf{x}_R, t)$.

(ii) Compute residual seismograms $\delta u_i(\mathbf{x}_S, \mathbf{x}_R, t) = u_i(\mathbf{x}_S, \mathbf{x}_R, t) - u_{i0}(\mathbf{x}_S, \mathbf{x}_R, t)$. These measure the difference between the synthetic and observed data.

(iii) Back propagate the residual seismograms by applying $\delta u_i(\mathbf{x}_S, \mathbf{x}_R, t)$ as a forcing function in a time reversed wave equation to evaluate the back propagated residual wavefield $\psi_j(\mathbf{x}_S, \mathbf{x}, t)$. Note that because the elastic wave equation is time symmetric, equation (3.18) is also used to do the back propagation.

(iv) Apply operator Ω^{γ}_{ijk} to u_j and ψ_j and calculate the time integral and shot sum in equation (3.42). In practice, this step is done during the back propagation to avoid storage of the wavefield ψ_j. This step represents stacking (over shots) the correlations between the background wavefield $\Omega^{\gamma}_{ijk} u_j$ and the back propagated residual wavefield $\Omega^{\gamma}_{ijk} \psi_j$.

Numerical calculations and accuracy

Note that in practice, the wave propagations are achieved by numerically solving the elastic wave equation via finite differences (see Mora, 1986). Appendix A compares the finite difference solution of the wave equation with the solution via the Haskell-Thompson method. The differences between the results of the two methods are consistent with the differences in their assumptions and are in any case minor for the typical offset-depth ratios. (The Haskell-Thompson method assumes the medium consists of homogeneous plane layers separated by sharp interfaces while the finite difference method assumes the medium has an upper spatial frequency cutoff). I conclude that finite differences can simulate elastic waves to an adequate accuracy to be useful for seismic inversions which is lucky because it can simulate waves in complex media.

The derivatives of operator Ω^{γ}_{ijk} are done finite differences and the time integral becomes a sum because the domains are discrete. Spatial derivatives of operator Ω^{γ}_{ijk} are done by the same eight point convolutional derivative operator used in the finite difference program while time derivatives are done by two point centered finite differences. The spatial derivative operator is accurate in the frequency band of the seismic wavelets used

in the examples while the time derivatives are less accurate at high frequencies. The examples in Chapter 4 will demonstrate that the finite difference approximations used to perform the operation of Ω^{γ}_{ijk} seem to be adequate in that the inversion results are good and converge rapidly.

Appendix B numerically verifies steps (i) through (iv) above corresponds to the adjoint operation to a high precision as a test of the interpretation of the adjoint equations and the numerical implementation.

3.4.7 The gradient direction

The conjugate-gradient inversion algorithm (3.15) requires the gradient $\mathbf{g} = \mathbf{D}^*\mathbf{C}_\mathbf{d}^{-1}\Delta\mathbf{d} + \mathbf{C}_\mathbf{m}^{-1}\Delta\mathbf{m}$ but equation (3.42) only represents the adjoint operation, i.e. $\mathbf{D}^*\delta\mathbf{d}$. Assuming, for simplicity of computation, a diagonal data covariance function (uncorrelated noise) and using equation (3.42) to define the adjoint operator $\mathbf{D}^*$, we obtain for the gradient in terms of model parameter γ

$$\delta\hat{\gamma}(\mathbf{x}) = \sum_S \int dt \frac{1}{C^t_\mathbf{d}(t)} \left(\Omega^{\gamma}_{ijk} u_j(\mathbf{x},t)\right)\left(\Omega^{\gamma}_{ijk}\psi_j(\mathbf{x},t)\right) + \sum_{\gamma'} C^{-1}_{\gamma\gamma'}(\mathbf{x})\left(\gamma'(\mathbf{x}) - \gamma'_0(\mathbf{x})\right) \quad , \tag{3.45}$$

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

$$\psi_j(\mathbf{x},t) = \sum_R \frac{1}{C^{\mathbf{x}_R}_\mathbf{d}(\mathbf{x}_R)} G_{ij}(\mathbf{x},-t;\mathbf{x}_R,0) * \Delta u_i(\mathbf{x}_R,t) \quad . \tag{3.46}$$

ψ_j is the back-propagated residual wavefield created by taking the difference Δu_i between the observed data u_{i0} and the forward modeled wavefield u_i, and using this as a forcing function in a wave equation that runs backwards in time. This assumes that the data covariances can be represented as

$$C_\mathbf{d}(\mathbf{x}_R,t) = C^{\mathbf{x}_R}_\mathbf{d}(\mathbf{x}_R) C^t_\mathbf{d}(t) \quad . \tag{3.47}$$

The model parameters are assumed to be spatially uncorrelated. If the different kinds of model parameters ($\gamma = \lambda$, μ etc.) are independent of one another then the only non-zero term of the $\sum_{\gamma'}$ is the term where $\gamma' = \gamma$.

Finally, rewriting the gradient (3.45) using the definition of the operator Ω^{γ}_{ijk} from equation (3.44) (see also equations (3.36) through (3.40)), the components of the gradient,

now denoted $\delta\hat{\hat{\rho}}$, $\delta\hat{\hat{\lambda}}$, $\delta\hat{\hat{\mu}}$, $\delta\hat{\hat{f}}_j$ and $\delta\hat{\hat{T}}_j$,
(i.e. $\mathbf{g}^T = \left[\delta\hat{\hat{\rho}}\ ,\ \delta\hat{\hat{\lambda}}\ ,\ \delta\hat{\hat{\mu}}\ ,\ (\delta\hat{\hat{f}}_j\ , j = 1,2,3)\ ,\ (\delta\hat{\hat{T}}_j\ , j = 1,2,3)\right]$)

$$\begin{aligned}\delta\hat{\hat{\rho}} &= -\sum_S \int dt \frac{1}{C_{\mathbf{d}}^t} \dot{u}_j \dot{\psi}_j + C_{\rho\rho}^{-1}(\rho - \rho_0) + \ldots \\ &= \delta\rho' + C_{\rho\rho}^{-1}(\rho - \rho_0) + \ldots \ , \end{aligned} \tag{3.48}$$

$$\begin{aligned}\delta\hat{\hat{\lambda}} &= -\sum_S \int dt \frac{1}{C_{\mathbf{d}}^t} u_{m,m}\psi_{j,j} + C_{\lambda\lambda}^{-1}(\lambda - \lambda_0) + \ldots \\ &= \delta\lambda' + C_{\lambda\lambda}^{-1}(\lambda - \lambda_0) + \ldots \ , \end{aligned} \tag{3.49}$$

$$\begin{aligned}\delta\hat{\hat{\mu}} &= -\sum_S \int dt \frac{1}{C_{\mathbf{d}}^t} \frac{1}{\sqrt{2}}(u_{k,j} + u_{j,k}) \frac{1}{\sqrt{2}}(\psi_{k,j} + \psi_{j,k}) + C_{\mu\mu}^{-1}(\mu - \mu_0) + \ldots \\ &= \delta\mu' + C_{\mu\mu}^{-1}(\mu - \mu_0) + \ldots \ , \end{aligned} \tag{3.50}$$

$$\begin{aligned}\delta\hat{\hat{f}}_j &= \sum_S \frac{1}{C_{\mathbf{d}}^t}\psi_j + C_{f_j f_j}^{-1}(f_j - f_{j0}) + \ldots \\ &= \delta f_j' + C_{f_j f_j}^{-1}(f_j - f_{j0}) + \ldots \ , \end{aligned} \tag{3.51}$$

$$\begin{aligned}\delta\hat{\hat{T}}_j &= \sum_S \frac{1}{C_{\mathbf{d}}^t}\psi_j + C_{T_j T_j}^{-1}(T_j - T_{j0}) + \ldots \\ &= \delta T_j' + C_{T_j T_j}^{-1}(T_j - T_{j0}) + \ldots \ . \end{aligned} \tag{3.52}$$

These equations define $\delta\lambda'$, $\delta\mu'$ and $\delta\rho'$ which will be used later in the expressions for the gradient in terms of the other choices of model parameters.

3.4.8 Model covariances

The dots at the ends of equations 3.48 through 3.52 represent cross-covariance terms between the different types of model parameters. These cross covariances are usually considered to be negligible for the sake of simplicity and due to lack of accurate statistical knowledge. It would be better to at least include diagonal cross covariance functions $C_{\rho\lambda} = \sigma_{\rho\lambda}^2$ etc., if reasonable values could be estimated from rock physical, empirical or statistical knowledge. Furthermore, constant diagonal covariance functions are normally assumed, i.e. $C_{\rho\rho} = \sigma_{\rho\rho}^2$ etc. (for an exception see Mora (1987a)). This means that

the physical parameters are assumed to be spatially uncorrelated. This is generally an oversimplification so biases can be expected in the solutions. If the spatial correlations of the properties were known and could be represented as banded covariance matrices, the operation of the covariance matrices could be easily achieved by space variable filtering.

Although the force and traction terms of the gradient have been derived as if they are spatially variable, in seismic exploration we can assume that the source locations are known so the force terms need to be solved only at the true source locations (i.e. the force and traction covariance functions C_{f_j} and C_{T_j} are Dirac functions of form $\delta(\mathbf{x}-\mathbf{x}_S)$). Actually, it would be best to know the source functions rather than invert for them because for reflection data, the source time history cannot be well resolved from the near surface geology and reverberation effects. However, if the direct wave is known such as is the case for VSP data, the source is much better determined and can be inverted simultaneously with the geology (Harlan, 1986). Normally, with only reflection data it is best to use a preprocessing step to find the source, or more generally, to use relaxation; that is, invert for the source only followed by inversion for the geology only and so on.

3.4.9 Data covariances

The time-dependent data covariance function can be approximated as

$$\frac{1}{C_{\mathbf{d}}^{t}(t)} = \frac{t^{2p}}{\sigma_{\mathbf{d}}^{2}} \quad , \tag{3.53}$$

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

The receiver dependent inverse covariance function $1/C_{\mathbf{d}}^{\mathbf{x}_R}(\mathbf{x}_R)$ is usually chosen to be unity except at the edges of the shot gather where a taper can be applied to decrease artifacts resulting from a sudden edge in the recorded data. In real data examples geophones have different couplings to the Earth which effectively scales the amplitudes in the seismic traces (seismic signal recorded as a function of time at geophone location $\mathbf{x}_R$). In this case, the noise amplitude and hence $1/C_{\mathbf{d}}^{\mathbf{x}_R}(\mathbf{x}_R)$ depends on $\mathbf{x}_R$. One way to look at this is to consider that $1/C_{\mathbf{d}}^{\mathbf{x}_R}(\mathbf{x}_R)$ is doing a "trace balancing" to normalize the energy in the seismic traces at different $\mathbf{x}_R$ locations. Failure to allow for the geophone couplings results in artifacts similar to those caused by sudden data truncations.

3.4.10 Different model parameters

The choice of model parameters is especially important in the elastic case where there are three independent parameters at each Earth location. The reason is that for some choices, three elastic parameters can be easily resolved from one another while for other choices they cannot. One may expect that the resolvable parameters will turn out to be easily measurable and observable properties such as the velocities and densities which determine the amplitude and traveltime of waves.

The results of the study of Tarantola et al. (1985) indicated that for seismic surveys where the source is dominantly P-waves, the choice of the Lamé parameters was much worse than the choices of either the velocities or the impedances. Furthermore, the results indicated that the density was not very well resolved (especially from P-wave velocity) and that in practice it is probable that only two parameters such as P- and S-velocity or P- and S-impedance will be resolvable.

Therefore, I will now derive expressions for the gradient in terms of the velocities and impedances. Note that the reason I did the original derivation in terms of the moduli was because these are the obvious parameters in the wave equation. This turned out to be a good choice because the derivation is messy unless it is done in terms of the Lamé parameters and it is easy to convert the final result to be in terms of other parameters.

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

$$\mathbf{g_m} = \frac{\partial \mathbf{d}}{\partial \mathbf{m}} \mathbf{C_d^{-1}} \Delta \mathbf{d} + \mathbf{C_m^{-1}} \Delta \mathbf{m} \ , \tag{3.54}$$

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

$$\mathbf{g_{m'}} = \frac{\partial \mathbf{d}}{\partial \mathbf{m}'} \mathbf{C_d^{-1}} \Delta \mathbf{d} + \mathbf{C_{m'}^{-1}} \Delta \mathbf{m}' = \frac{\partial \mathbf{d}}{\partial \mathbf{m}} \frac{\partial \mathbf{m}}{\partial \mathbf{m}'} \mathbf{C_d^{-1}} \Delta \mathbf{d} + \mathbf{C_{m'}^{-1}} \Delta \mathbf{m}' \ . \tag{3.55}$$

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

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

$$\delta \hat{\tilde{\alpha}} = \frac{\partial \lambda}{\partial \alpha} \delta \lambda' + \frac{\partial \mu}{\partial \alpha} \delta \mu' + \frac{\partial \rho}{\partial \alpha} \delta \rho' + C_{\alpha\alpha}^{-1} (\alpha - \alpha_0) + \ldots \ , \tag{3.56}$$

where $\delta\lambda'$, $\delta\mu'$ and $\delta\rho'$ are defined by equations (3.48) through (3.52). The required elements of the Jacobian $\partial\lambda/\partial\alpha$ etc. can be obtained from the relationships for α and β

$$\alpha = \sqrt{\frac{\lambda+2\mu}{\rho}} \ , \ \beta = \sqrt{\frac{\mu}{\rho}} \ , \tag{3.57}$$

or solving for λ and μ,

$$\lambda = \rho\alpha^2 - 2\rho\beta^2 \ , \ \mu = \rho\beta^2 \ . \tag{3.58}$$

Using these equations we obtain the gradient in terms of the P- and S-wave velocities and density as

$$\delta\hat{\hat{\alpha}} = 2\rho\alpha\delta\lambda' + C_{\alpha\alpha}^{-1}(\alpha-\alpha_0) + \ldots \ , \tag{3.59}$$

$$\delta\hat{\hat{\beta}} = -4\rho\beta\delta\lambda' + 2\rho\beta\delta\mu' + C_{\beta\beta}^{-1}(\beta-\beta_0) + \ldots \ , \tag{3.60}$$

$$\delta\hat{\hat{\rho}}_\alpha = (\alpha^2-2\beta^2)\delta\lambda' + \beta^2\delta\mu' + \delta\rho' + C_{\rho\rho}^{-1}(\rho-\rho_0) + \ldots \ . \tag{3.61}$$

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

$$Z_P = \rho\alpha = \sqrt{\rho(\lambda+2\mu)} \ , \ Z_S = \rho\beta = \sqrt{\rho\mu} \ , \tag{3.62}$$

or solving for λ and μ

$$\lambda = \frac{1}{\rho}\left(Z_P^2 - 2Z_S^2\right) \ , \ \mu = \frac{1}{\rho}Z_S^2 \ , \tag{3.63}$$

so the components of the gradient with respect to the impedances are

$$\delta\hat{\hat{Z}}_P = 2\alpha\delta\lambda' + C_{Z_PZ_P}^{-1}(Z_P - Z_{P0}) + \ldots \ , \tag{3.64}$$

$$\delta\hat{\hat{Z}}_S = -4\beta\delta\lambda' + 2\beta\delta\mu' + C_{Z_SZ_S}^{-1}(Z_S - Z_{S0}) + \ldots \ , \tag{3.65}$$

$$\delta\hat{\hat{\rho}}_Z = -(\alpha^2-2\beta^2)\delta\lambda' - \beta^2\delta\mu' + \delta\rho' + C_{\rho\rho}^{-1}(\rho-\rho_0) + \ldots \ . \tag{3.66}$$

Note that aside from a factor of ρ, the gradient in terms of the velocities (equations (3.59) through (3.61)) and the gradient in terms of the impedances (equations (3.64) through (3.66)) are identical except that the density gradient terms have some sign differences. The main effect of these sign differences is to alter the magnitude of the density gradient depending on which choice of model parameters is used.

3.5 Inversion described by pictures

3.5.1 Introduction

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

$$\delta\hat{\gamma} \quad = \quad \sum_{s} \int dt \; C_{\mathbf{d}}^{-1} (\Omega_{ijk}^{\gamma} u_j)(\Omega_{ijk}^{\gamma} \psi_j) \; + \; C_{\gamma\gamma}^{-1} \Delta\gamma \quad . \tag{3.67}$$

It is evaluated by first initializing the gradient $\delta\hat{\gamma}$ to zero. Then, for each shot, some calculations are done to update the gradient. These proceed by first computing and storing the background wavefield u_j. Subsequently, the back-propagated residual wavefield, ψ_j is evaluated in reverse time using elastic finite differences by feeding in the residuals into the Earth as a forcing function starting from the latest time. At each time step in the finite difference algorithm, a contribution to the time integral is added in and the gradient slowly builds up as time evolves backwards towards the zero time when the shot was detonated. Finally, after all shots have been summed, the damping term $C_{\gamma\gamma}^{-1}\Delta\gamma$ is added.

This section describes this process for a single shot using pictures for two different kinds of data, namely transmission data and reflection data.

3.5.2 Transmission data

Tutorial

Transmission seismic data includes VSP data or well-well surveys. The distribution of sources and receivers is similar to that of medical imaging and one may therefore expect to recover a fairly complete image of the Earth. Here, I will walk through the calculations with the help of pictures.

Figure 3.1 shows a circular body embedded in a homogeneous space. Relative to the homogeneous background medium, the body had a small (few percent) positive P-wave velocity anomaly and a small negative S-wave velocity anomaly.

Elastic finite differences were used to do a shot simulation to generate some data. Two-component data was recorded on all sides of the model due to a vertical shot located at the base of the model. The source excitation function had time history shaped like a second derivative of a Gaussian curve. Direct waves traveling from the shot to receivers that pass through the anomalous region will be advanced (P-waves) or delayed (S-waves). This results in a significant residual (difference between the synthetic data and data generated

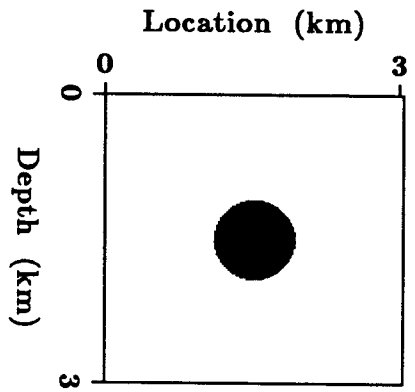

Figure 3.1: The Tarantola/Gauthier "camembert" velocity model relative to a homogeneous background used to generate transmission data.

using the homogeneous background model) as shown in Figure 3.2. The residual is large for transmitted waves that have passed through the anomalous region.

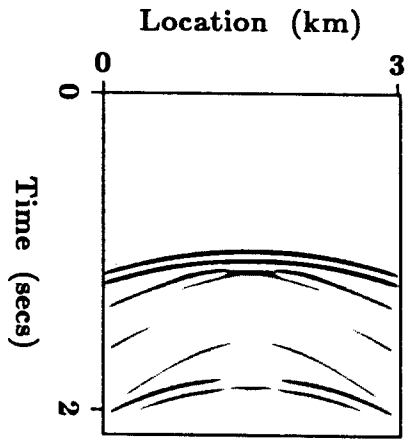

Figure 3.2: Vertical component residual transmission data.

Figure 3.3 illustrates graphically how the gradient is computed.

The residuals are fed back into the model backwards in time to generate the back propagated residual wavefield ψ_j. At any instant in (backward) time, the stored background wavefield u_j and the back propagated residual wavefield ψ_j are operated on by Ω^{γ}_{ijk} and multiplied together. The time integral is evaluated by adding this contribution into the adjoint fields $\delta\lambda'$, $\delta\mu'$ and $\delta\rho'$ (see equations 3.48 through 3.50). Finally, after the Lamé adjoint fields are complete, the gradient is computed in terms of P-wave velocity, S-wave velocity and density through equations 3.59 through 3.60 (see Figure 3.4). Because the

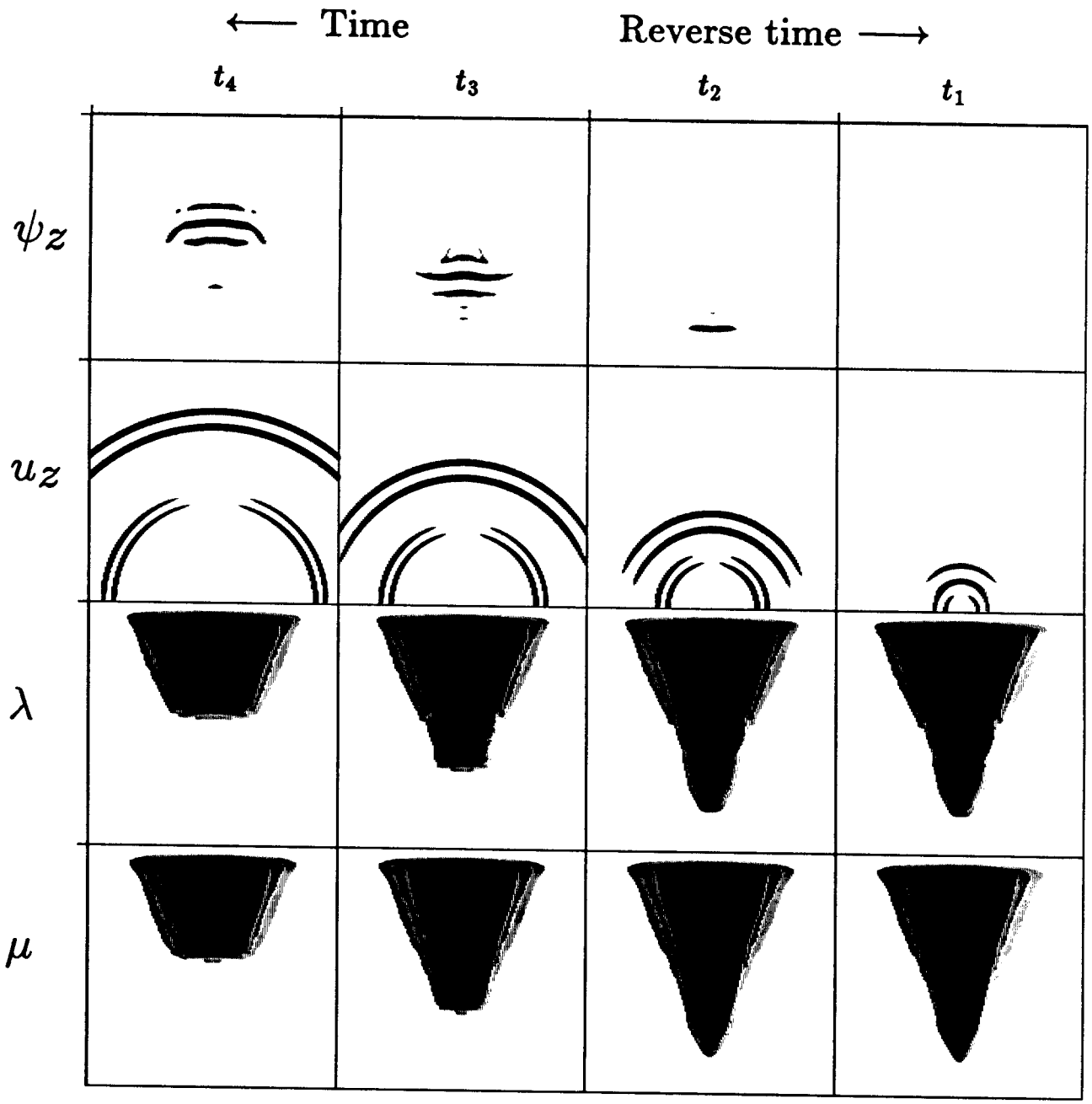

Figure 3.3: Pictorial representation of the gradient calculations for the transmission data.

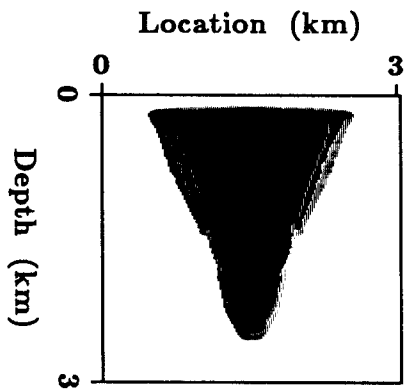

Figure 3.4: (a). P-wave velocity perturbation for the transmission data inversion.

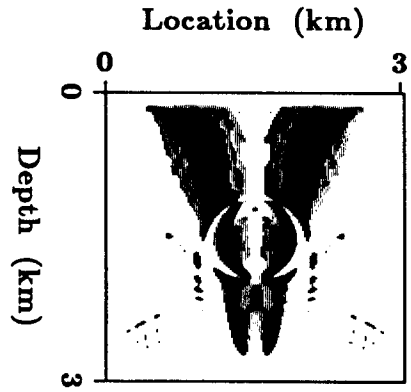

Figure 3.4: (b). S-wave velocity perturbation for the transmission data inversion.

S-wave velocity gradient was mainly negative, it is displayed in reverse polarity (otherwise most of the display would be white).

Comparison to tomography

Observe that the back propagated wavefield ψ_j was traveling in the same direction as the background wavefield u_j (because the main data residual was in the transmitted waves) and so the gradient is a smear running through the anomalous region. This can be compared with traveltime tomography which would also try to account for traveltime delays by distributing a velocity perturbation through the anomalous region (see Toldi (1985)).

From the above discussion it is clear that the inversion algorithm is similar to an elastic wave equation tomography when transmitted waves are used in the inversion. Consequently, low-wavenumber components of the model can be resolved if transmitted waves are present in the data set. For a complete reconstruction of the circular anomaly, many shots would be required at different locations. To see this, imagine summing the P-wave velocity gradient field in Figure 3.4 (a) with copies of itself rotated through various angles. The greater the number of shots and hence angles of waves going through the anomaly, the greater the magnitude of the central region relative to the surrounding smear. Similarly for the S-wave velocity gradient in Figure 3.4 (b).

As expected, S-wave velocity is poorly resolved where S-waves are weak as seen by the lack of energy in the S-wave velocity result directly above the vertical component source. However, the S-wave velocity throughout the entire region will be resolved if many shots at different locations are used.

3.5.3 Reflection data

Tutorial

The most abundant kind of data used in seismic prospecting for oil is reflection seismic data usually in the form of shot gathers. This section illustrates how the inversion calculations proceed using this kind of data.

Figures 3.5 through 3.8 illustrate the adjoint calculations when using reflection data (c.f. Figures 3.1 through 3.4 which illustrated the calculations using transmission data). The data was calculated by doing a shot simulation using elastic finite differences with the two layer model of Figure 3.5.

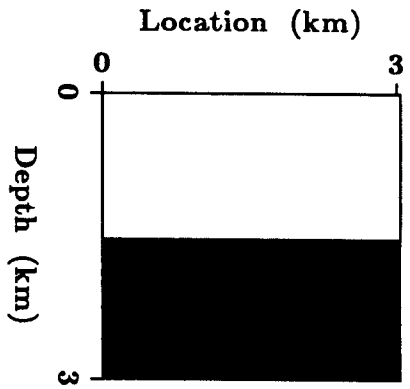

Figure 3.5: Layer/halfspace velocity model relative to a homogeneous background used to generate reflection data.

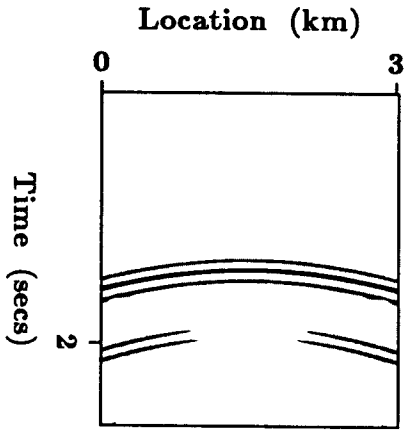

Figure 3.6: Vertical component residual shot gather for layer/halfspace model.

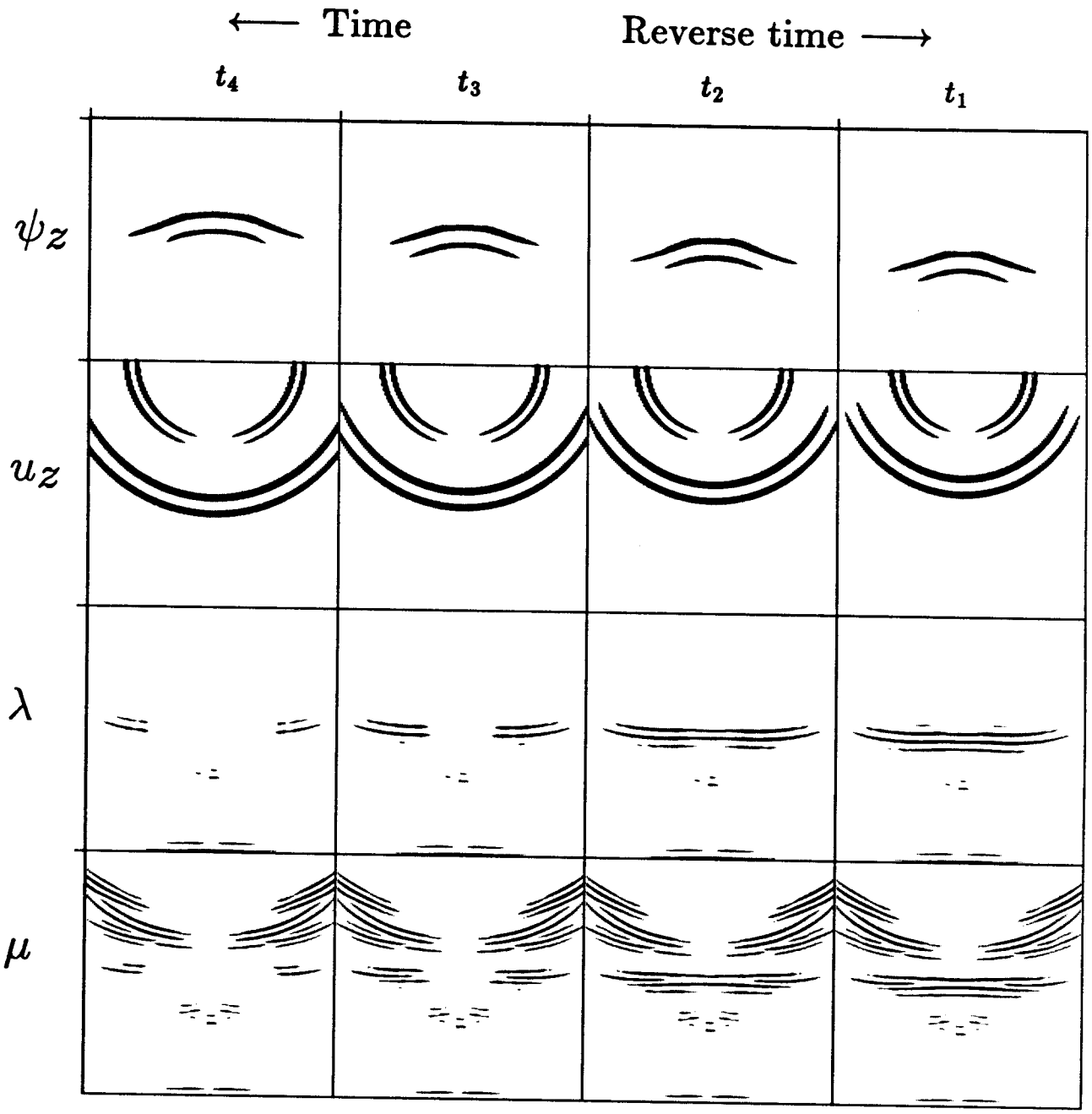

Figure 3.7: Pictorial representation of the gradient calculations for the reflection data.

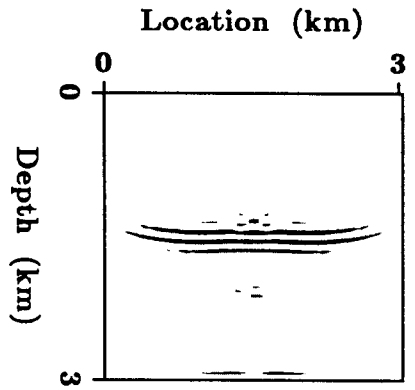

Figure 3.8: (a). P-wave velocity perturbation for the reflection data inversion.

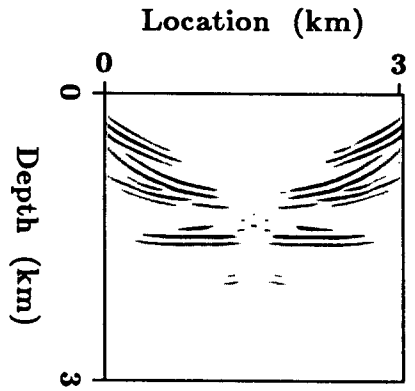

Figure 3.8: (b). S-wave velocity perturbation for the reflection data inversion.

The main difference between the calculations here using reflection data versus the previous example using transmission data is that now the back propagated wavefield ψ_j is going in the opposite direction to the background wavefield u_j (see Figure 3.7). This is similar to a migration using the U/D principle (see Claerbout (1976)) to obtain reflectivities by dividing the back scattered waves with the direct wave.

The P- and S-wave velocity gradients are shown in Figure 3.8. The P-wave velocity result looks similar to a derivative operator located at the interface between the layers while the S-wave velocity looks like a negative derivative operator. This is a band limited version of the positive and negative P- and S-wave velocity perturbations.

As in the transmission example, the S-wave velocity was unresolved in a region directly below the vertical component shot where the S-waves were weak.

Another feature is the occurrence of smile-like artifacts on the S-wave velocity plot. These result from non-zero correlations between different wave modes at positions other than at the real reflection locations (c.f. elliptical prestack migration impulse responses).

Comparison to migration and linearized inversion

The inversion algorithm is based on a least-squares gradient direction of the form given by equation (3.67).

If Ω and $C_{\mathbf{d}}^{-1}$ were set to unity and $C_{\gamma\gamma}^{-1}$ was set to zero then the gradient corresponds to an elastic prestack shot-profile migration by correlation of the upgoing reflected waves $U = \psi_j$ with the downgoing direct wave $D = u_j$ (migration is sometimes done by using the correlation of U and D as a stable approximation to U/D which estimates reflectivities (see Claerbout, 1976)). Where the correlation is great, there is a reflector, or, in other words, the reflected waves intersect the direct wave at the reflecting interface.

If in addition, the two wavefields u_j and ψ_j represented pressure wavefields, then the inversion is identical to a conventional acoustic prestack shot-profile migration.

Hence, prestack elastic migration (e.g. Kuo and Dai (1984), Dai and Kuo 1986, Wapenaar (1986) and Wapenaar et al. (1987)) is essentially equivalent to the computation of the least squares gradient direction of inversion but excluding covariances, the operator Ω and the damping factor $C_{\gamma\gamma}^{-1}\Delta\gamma$ (see also Lailly (1984) for a comparison of migration and inversion).

Strictly speaking, the discussion so far is only true if the wavefields u_j and ψ_j represent one-way waves. However, they are calculated by the full elastic wave equation and so have

waves traveling both upwards and downwards. Chapter 5 shows how the up and down traveling waves are necessary in order to resolve the low vertical-wavenumber components of the velocities.

The operator Ω is very important. It unravels the wavefield information into the desired model parameters using the amplitude information (i.e. the overall amplitude of the reflected waves as well as the amplitude-incidence angle information).

Another main difference between the inversion algorithm described here and migration is that it is iterative. The iterations tend to allow for the fact that the modeling function $\mathbf{d}(\mathbf{m})$ is nonlinear as well as performing the same function as the inverse Hessian of equation (3.6) which would correct all the relative amplitudes of velocities and densities to their most likely values (i.e. the best fit values) and remove the source signature (i.e. do a spatial deconvolution). The iterations also help resolve the interval velocities. If the starting model is smoothly varying and does not generate reflections, the inversion algorithm, like migration, would have little hope of obtaining any of the lower wavenumbers except perhaps if a very clever preconditioning were constructed. The reason the low wavenumbers can be regained is because the inversion is iterative and tries to match shapes of reflection traveltime curves to modeled events (by matching the amplitudes of the seismic data observations).

I will show in chapter 5 that contributions in the inversion formulas to the low- and high-wavenumbers components of the velocities are separable and that each separated component has the same signal to noise ratio. Thus, the low wavenumber which are smaller in magnitude can be boosted without boosting noise.