A finite-difference algorithm for full waveform teleseismic tomography

S. Roecker, B. Baker and J. McLaughlin
Rensselaer Polytechnic Institute, Troy, NY 12180, USA. E-mail: roecks@rpi.edu

Accepted 2010 February 6. Received 2010 January 14; in original form 2009 August 27

SUMMARY
We adapt a 2-D spectral domain finite difference waveform tomography algorithm previously used in active source seismological imaging to the case of a plane wave propagating through a 2.5-D viscoelastic medium in order to recover P and S wave speed variations from body waves recorded at teleseismic distances. A transferable efficacy that permits recovery of arbitrarily heterogeneous models on moderately sized computers provides the primary motivation for choosing this algorithm. Synthetic waveforms can be generated either by specifying an analytic solution for a background plane wave in a 1-D model and solving for the source distribution that would produce it, or by solving for a scattered field excited by a plane wave source and then adding the background wavefield to it. Because the former approach typically involves a concentration of sources at the free surface, the latter tends to be more stable numerically. We adapt a gradient approach to solve the inverse problem to maintain tractability; calculating the gradient does not require much more computational effort than does the forward problem. The waveform tomography algorithm can be applied in a straightforward way to perform receiver function migration and traveltime inversion.

Key words: Body waves; Seismic tomography; Computational seismology; Wave scattering and diffraction.

1 INTRODUCTION
Analyses of seismograms recorded by regional seismic networks of body waves generated by sources at teleseismic distances have made tremendous contributions to our understanding of the Earth’s interior. Two techniques in particular, traveltime tomography and receiver function analysis, have evolved from their introduction some 30 yr ago (Aki et al. 1977; Vinnik 1977; Langston 1979) to become standard tools for visualizing properties of the medium directly beneath the network. Traveltime tomography techniques attempt to unravel an integrative effect of slowness along a ray path or Fresnel zone, while receiver function analysis focuses more on the location of discrete scatterers, often in the form of continuous interfaces, that generate coherent mode conversions. Each approach has its particular advantages and shortcomings. Traveltime tomography images are generally robust and useful for illuminating large scale features, but the resolution of the images is often low and features such as isolated zones of low wave speed may be missed entirely. Receiver functions provide an excellent means for detecting localized contrasts in medium properties, but the locations of those contrasts in space and their implications for changes in the medium are often poorly defined. For example, the illumination of the Mohorovičić discontinuity is now a routine result of P receiver function analysis, and the lithosphere–asthenosphere boundary, until recently considered an exotic target, has become a fairly robust result of S receiver function analysis (e.g. Oreshin et al. 2002; Kumar et al. 2005). However, an accurate determination of crustal and lithospheric thickness is difficult without ancillary information on background wave speeds, for example from surface wave dispersion analysis (e.g. Herrmann et al. 2001). Even with a reasonable background model, quantitative interpretation of mode conversions can be difficult if the medium is laterally heterogeneous, and as a result this type of analysis tends to fall into one of two end-member categories: the ‘modelling’ end that presumes that the Earth beneath a network is, at least locally, approximately 1-D, and a ‘migration’ end that is interested primarily in locating the sources of the mode conversions with at most a cursory explanation for the polarity and amplitude of the scattered phase. A number of investigators at both ends have helped to reduce this gap. For example, on the modelling end, Vinnik et al. (2004) compensate for lateral heterogeneity by assuming piecewise homogeneity and including complementary information from traveltime tomography. Advances on the migration side include relatively straightforward incorporation of tomographically derived wave speeds (e.g. Wilson & Aster 2003) and a variety of wave equation based algorithms that locate scatterers using a Born (single scatterer) approximation (e.g. Bostock et al. 2001; Poppeliers & Pavlis 2003a,b; Levander et al. 2005) Each of these approaches has the advantage of requiring modest computational resources, but still leaves a substantial gap between modellers and migraters.
In this paper, we attempt to reduce this gap further by introducing a full waveform tomographic technique for quantitative modelling of teleseismic body waves, both primary and multiply scattered, that can be applied to 2.5-D viscoelastic media. The full waveform approach simultaneously models and migrates receiver functions and implicitly includes a form of travelt ime tomography as well.

2 TRACTABLE WAVEFORM TOMOGRAPHY

The term ‘waveform tomography’ refers to a class of algorithms that attempts to recover information about medium properties from the entire seismic waveform. The basic concepts of waveform tomography were introduced through the seminal work of Lailly (1983), Tarantola (1984), Mora (1987) and Woodward (1992), and with a few exceptions (receiver function modelling being one) have for the most part been used in the active source environment. An attractive advantage of waveform tomographic techniques over traditional travelt ime tomography is the potential for significantly improved image resolution: there is simply much more information in a waveform than there is in an arrival time. For example, in a side by side comparison of the two techniques in a simulated cross hole environment, Pratt (2003) demonstrated that while travelt ime resolution was limited by the size of the Fresnel zone, waveform tomography resolution was on the order of the dominant wavelength in the signal.

The algorithm we discuss here is an adaptation of waveform tomography techniques developed over the past 18 yr by R. G. Pratt and coworkers (mostly G. Hicks, L. Siguer, Z. M. Song, I. Stekl and P. R. Williamson; for convenience these techniques are referred to collectively below as the RGP techniques). They are robust enough to have been successfully applied in analysis of real data (e.g. Brittan et al. 1997; Bleibinhaus et al. 2007). Equally as important as their track record, however, is their demonstrated ability to generate reasonably accurate images in heterogeneous media without resorting to computing on a massive scale. For example, they were used to image a hypothetical and highly heterogeneous structure in a competition promoted by a Commission on Controlled-Source Seismology (CCSS) workshop organized in 2003 (Hole et al. 2005; Brenders & Pratt 2007). The image generated by these techniques was easily as good as those generated by higher powered algorithms but had the advantage of being produced using a single Linux box.

While massively parallel computing has been a boon to seismology (e.g. Akc¸elik et al. 2005; Brenders & Pratt 2007), the image generated by these techniques was easily as good as those generated by higher powered algorithms but had the advantage of being produced using a single Linux box.

Our starting point is the acoustic inverse problem described in Pratt (1990, 1999, 2003), Pratt & Worthington (1990), Pratt et al. (1998), Song et al. (1995a,b) and Williamson & Pratt (1995), and an extension of the forward problem to the elastic case by Pratt (1990) and Stekl & Pratt (1998). For clarity we summarize below the key concepts of those studies that have direct bearing on the 2.5-D teleseismic case.

3 WAVEFORM COMPUTATION

We define the forward problem for modelling teleseismic body waves as the response of a locally heterogeneous, viscoelastic medium to an excitation by plane waves generated by sources external to that medium. We start with the Cartesian form of the 3-D elastic wave equation for displacement \( u = (u, v, w) \) in an isotropic, heterogeneous medium in the frequency domain

\[
\omega^2 \rho u + \frac{\partial}{\partial x} \left[ (\lambda + 2\mu) \frac{\partial u}{\partial x} + \lambda \left( \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) \right] + \frac{\partial}{\partial y} \left[ \mu \left( \frac{\partial v}{\partial y} + \frac{\partial u}{\partial x} \right) + \frac{\partial}{\partial z} \left( \frac{\partial w}{\partial z} + \frac{\partial u}{\partial x} \right) \right] = -f_x,
\]

\[
\omega^2 \rho v + \frac{\partial}{\partial y} \left[ (\lambda + 2\mu) \frac{\partial v}{\partial y} + \lambda \left( \frac{\partial w}{\partial z} + \frac{\partial u}{\partial x} \right) \right] + \frac{\partial}{\partial z} \left[ \mu \left( \frac{\partial w}{\partial z} + \frac{\partial v}{\partial y} \right) + \frac{\partial}{\partial x} \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) \right] = -f_y,
\]

\[
\omega^2 \rho w + \frac{\partial}{\partial z} \left[ (\lambda + 2\mu) \frac{\partial w}{\partial z} + \lambda \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) \right] + \frac{\partial}{\partial x} \left[ \mu \left( \frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} \right) + \frac{\partial}{\partial y} \left( \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) \right] = -f_z,
\]

where \( \omega \) is the radial frequency and \( f = (f_x, f_y, f_z) \) is a source term. Conceding that a full 3-D approach is beyond the capabilities of moderately sized computers, we reduce the above equations to 2.5-D by recognizing a plane wave invariance in the \( y \)-direction (Takenaka & Kennett 1996) expressed in the frequency domain as

\[
u(\omega, x, y, z) = A(\omega)e^{i(\omega t - \omega y - k_x x - k_z z)},
\]

where \( p_x \) is the \( x \)-component of the slowness vector. Recognizing that \( \partial u / \partial y = -i\omega p_x u, \partial^2 u / \partial y^2 = -\omega^2 p_x^2 u, \partial \lambda / \partial y = 0 \), and \( \partial \mu / \partial y = 0 \), the wave equations in 2.5-D become

\[
(\rho - p_x^2 \mu) \omega^2 u + \frac{\partial}{\partial x} \left[ (\lambda + 2\mu) \frac{\partial u}{\partial x} + \lambda \left( \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) \right] + \frac{\partial}{\partial z} \left[ \mu \left( \frac{\partial v}{\partial y} + \frac{\partial u}{\partial x} \right) + \mu \left( \frac{\partial w}{\partial z} + \frac{\partial u}{\partial x} \right) \right] - i\omega p_x \left( \frac{\partial}{\partial x} \lambda v + \mu \frac{\partial u}{\partial x} \right) = -f_x,
\]

\[
(\rho - p_x^2 (\lambda + 2\mu)) \omega^2 v + \frac{\partial}{\partial y} \left[ \mu \left( \frac{\partial v}{\partial y} + \frac{\partial u}{\partial x} \right) + \lambda \left( \frac{\partial w}{\partial z} + \frac{\partial u}{\partial x} \right) \right] - i\omega p_x \left( \frac{\partial}{\partial y} \lambda u + \mu \frac{\partial w}{\partial z} \right) = -f_y
\]

\[
(\rho - p_x^2 \mu) \omega^2 w + \frac{\partial}{\partial z} \left[ (\lambda + 2\mu) \frac{\partial w}{\partial z} + \lambda \left( \frac{\partial v}{\partial y} + \frac{\partial u}{\partial x} \right) \right] + \frac{\partial}{\partial x} \left[ \mu \left( \frac{\partial w}{\partial z} + \frac{\partial u}{\partial x} \right) + \mu \left( \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) \right] - i\omega p_x \left( \frac{\partial}{\partial z} \lambda w + \mu \frac{\partial v}{\partial y} \right) = -f_z.
\]

© 2010 The Authors, GJI, 181, 1017–1040
Journal compilation © 2010 RAS
For added stability we include a lumped mass contribution (Marfurt 1984; Šteklo & Pratt 1998) for the central node of a five-point differencing scheme discussed in detail in Appendix B, that minimizes numerical dispersion over all directions of propagation. When $\rho' = 0$ the weights reduce to the forms found in Šteklo & Pratt (1998) for the 2-D elastic case for $P$–$SV$ and to those in Jo et al. (1996) for the acoustic case for $SH$.

The advantage of 2.5-D is that we can model a wide variety of environments of geologic interest that can be expected to be approximately invariant along a particular axis (e.g. mountain belts, rift zones, subduction zones and basins) without significantly increasing the computational demands of the 2-D case, but at the same time relaxing the 2-D restriction that all sources and receivers be in the same plane.

We compute waveforms by solving the 2.5-D system of equations above using finite differences on a 2-D $(x, z)$ grid. The medium is parametrized by specifying a density and either $P$ and $S$ wave speeds or the Lamé parameters $\lambda$ and $\mu$ at each node. The grid is bounded by a vertically dipping rectangle imbedded in background 1-D media (Fig. 1). The media on either vertical side of the rectangle can be vertically stratified but are not necessarily identical. Our method requires only that phases scattered inside the heterogeneous region not re-enter the media above and below the rectangle must be homogeneous half-spaces; typically the top of the rectangle is a free surface. The half-space bounding the bottom of the heterogeneous medium ensures that the time-reduced waveform incident along this boundary is everywhere the same.

A principal concern in solving wave equations using finite differences is the potential for introducing numerically generated dispersion and anisotropy. These artefacts derive from the differences in $P$ and $S$ wave speeds specified in the continuous medium and our finite difference representations of them. The size of these artefacts decreases as the grid density increases, but of course at a cost of increased memory and computational time. Following Šteklo & Pratt (1998) and Jo et al. (1996) we mitigate these artefacts by combining two Cartesian systems related by a $45^\circ$ rotation about the $y$ (invariant) axis (Fig. 1). In the ‘non-rotated’ system, $x$ is horizontal and $z$ is positive down. The rotated $(x', z')$ system is defined by

$$x' = \frac{z + x}{\sqrt{2}} \quad \text{and} \quad z' = \frac{z - x}{\sqrt{2}}.$$  

The $y$ axis is the same for both systems. Positive directions for $x$ and $y$ are arbitrary but conform to $\hat{x} \times \hat{y} = \hat{z}$.

To combine systems, we create an additional set of equations in the $(x', z')$ $45^\circ$ rotated system by substituting

$$\frac{\partial}{\partial x} = \frac{1}{\sqrt{2}} \left( \frac{\partial}{\partial x'} - \frac{\partial}{\partial z'} \right) \quad \text{and} \quad \frac{\partial}{\partial z} = \frac{1}{\sqrt{2}} \left( \frac{\partial}{\partial x'} + \frac{\partial}{\partial z'} \right)$$  

into the 2.5-D version of the wave equation. We then solve this system of equations using finite differences on a weighted sum of the contributions from the two (primed and unprimed) frames

$$\left( \rho - p_1^2 \mu \right) \omega^2 u + aA + (1 - a)A' = -f_i,$$
$$\left[ \rho - p_1^2 (\lambda + 2\mu) \right] \omega^2 v + aB + (1 - a)B' = -f_j,$$
$$\left( \rho - p_1^2 \mu \right) \omega^2 w + aC + (1 - a)C' = -f_z.$$  

For added stability we include a lumped mass contribution (Marfurt 1984; Šteklo & Pratt 1998) for the central node of a five-point differencing star

$$\rho' u' = b p u + \frac{(1 - b)}{4} \left( \rho^+ u^+ + \rho^- u^- + \rho_u u_u + \rho_u u_u \right),$$  

where we have used the differencing notation described in Appendix A. The weights $a$ and $b$ are determined by an optimization procedure, discussed in detail in Appendix B, that minimizes numerical dispersion over all directions of propagation. When $p_1 = 0$ the weights reduce to the forms found in Šteklo & Pratt (1998) for the 2-D elastic case for $P$–$SV$ and to those in Jo et al. (1996) for the acoustic case for $SH$.

---

Figure 1. Schematic of the model. We consider a heterogeneous medium (grey rectangle) bounded by half-spaces on the top and bottom and vertically stratified media on either side. The upper boundary is usually, but not necessarily, a free surface. Plane waves enter the medium from below, but converted waves may also enter the heterogeneous region from the side. The 1-D media on the left and right are used for waves entering from the left and right, respectively, and are not necessarily the same. To prevent diffractions, the wave speeds in the heterogeneous medium match those in the 1-D media at the boundaries. Orientations of the ‘non-rotated’ $(x, y, z)$ and $45^\circ$ rotated $(x', y, z')$ axes are indicated at the right.

---

1 With the exception of an error in their derivation discussed in Appendix B.
The off-axis case is somewhat more complicated as it involves complex terms (specifically the $i\omega$ terms that come from the $\partial^2/\partial y$ partial derivatives). Nevertheless, we find that with appropriate weighting numerical dispersion can be suppressed to less than a few per cent with grid spacing as sparse as four points per wavelength.

After replacing the continuous derivatives with the finite difference stencils of Appendix A, the forward problem can be written as the solution to a system of linear equations:

$$Su = -f$$

If the grid has $n_x \times n_z$ nodes, then $u$ is a $3n_xn_z$ length vector containing the $(u, v, w)$ displacement field at each point in the model, $f$ is a $3n_xn_z$ length vector containing the sources at each point $[f = (f_x, f_y, f_z)]$, and $S$ is a banded, structurally symmetric $3n_xn_z \times 3n_xn_z$ matrix containing the finite difference terms. Following Pratt (1990), we solve for the wavefield

$$u = -S^{-1}f$$

using LU decomposition. The number of non-zero terms (i.e. ‘fill’) in the decomposition is minimized by using the nested decomposition algorithm of George & Liu (1970) as implemented in the Metis package of Karypis & Kumar (1998).

### 3.1 Using a background wavefield to specify forces

In the case of teleseismic body waves the forces in eq. (3) are plane waves generated by distant events external to the modelled medium. Specification of such forces in (3) may seem inappropriate because in the absence of internal sources

$$Su = 0$$

and the solution to the resulting homogeneous eq. (5) is non-unique. Nevertheless, one can ask if there is there a distribution of sources $f$ that will generate a desired $u = u_0$ in a medium described by a finite difference matrix $S_0$. Evidently, an appropriate force distribution can be computed directly from

$$S_0 u_0 = f_0.$$ (6)

Hence if we have a means to specify a field $u_0$ in a medium described by $S_0$, we can use the force distribution above to compute the response $u$ of another medium described by $S$ to the same excitation by solving $u = S^{-1}f$. Eq. (6) shows that in order to generate an appropriate background wavefield we must specify forces at every point in the medium. However, the differencing nature of $S$ means that most of the energy in $f_0$ will be located where the gradient in the displacement field is largest, and this generally will be at the boundaries of the grid. In particular, the forces at the free (top) surface can be several orders of magnitude greater than anywhere else. While not an issue conceptually, such a concentration of source energy all along the free surface can be numerically unstable as the wavefield will necessarily depend on choices we make on how to compute the response of this surface. For this reason we will describe an alternate approach that is theoretically identical but numerically more stable.

We can describe the total displacement field $u_t$ as the sum of a background wavefield $u_b$ and a scattered wavefield $u_s$ generated by perturbations to the background medium

$$u_t = (u_b + u_s).$$ (7)

We can similarly separate our differencing matrix $S$ into a component $S_b$ that produces the background wavefield and a component $S_s$ that produces the scattered field

$$S = S_b + S_s.$$ (8)

Recalling that in the absence of internal sources

$$Su_b = (S_b + S_s)(u_b + u_s) = S_b u_b + S_s u_b + Su_b = 0.$$ (9)

The background plane wave also has no internal sources, hence $S_b u_b = 0$ and

$$Su_s = -S_s u_s = -(S - S_s)u_s.$$ (10)

Using (10), we solve for the scattered wavefield and then add the background to obtain the total field

$$u_t = u_s - S^{-1}[(S - S_s)u_s].$$ (11)

In this case the force distribution is $f = -(S - S_s)u_s$, meaning that forces appear only at perturbations to the background wave speed, which generally will not include any of the boundaries. Hence in this case we avoid having large amplitude sources distributed all along the free surface. To show that the two approaches are conceptually equivalent, recall that in the total field approach (6) we specified a distribution of forces $f_s$ such that

$$Su_s = S_s u_s = f_s.$$ (12)

and so from (9)

$$Su_s = (S_b + S_s)(u_b + u_s) = f_s + S_s u_s + Su_s = f_s.$$ (13)
and finally
\[ Su_h = -S_h u_e \]  
which is eq. (10). Hence we can solve for the total wavefield directly as a response to a distribution of sources that create a wavefield in the background medium, or solve for the scattered wavefield by computing the response a distribution of sources located at the heterogeneities in the medium. We generally choose to use the scattered approach in (10) as it seems to be more stable numerically.

### 3.2 The Plane Wave Source

In implementing (10) we could specify any background wavefield \( u_e \), corresponding to a given \( S_e \). For example, \( u_e \) could be the wavefield computed from a previous iteration of the inverse problem for which we already have stored an appropriate \( S_e \). Generally, however, wavefields for laterally varying media of interest are not known a priori, and so we begin with simpler media for which we have analytical solutions. In the case of teleseismic body waves a suitable \( u_e \) should allow for surface reflections outside the heterogenous (‘modelled’) region. It should also allow for both continuous variations in medium properties with depth and relatively abrupt changes, such as the Moho or sediment/basement interfaces, that would generate large amplitude mode conversions both inside and outside of the heterogenous region.

Fortunately there are several techniques available to compute the response of a 1-D (vertically stratified) medium to plane wave excitation. For our purposes we employ an algorithm based on the Thomson–Haskell approach (Haskell 1953, 1960, 1962), mostly because it is a frequency domain approach that has been used extensively in modelling 1-D \( P-\) and \( S-\)receiver functions (e.g. Kosarev et al. 1993; Oreshin et al. 2002). We adapt it here by computing the response at the free surface to an impulse at the base of the model as described in Haskell (1960, 1962) and then recovering the response at any point in the medium by accumulating the propagator matrices in reverse order and applying an appropriate phase shift \( \theta(\Delta x) = e^{i\omega\Delta x} \) where \( p_h \) is the horizontal slowness) along the base of the model. The computation of the wavefield in this manner is highly efficient.

One can surmise from this approach that it will be desirable to match the external 1-D model to the internal model along the vertical boundaries, otherwise the change in wave speed could generate unwanted diffractions. At the same time, it is not necessary to assign the same 1-D model to both sides of the box. The incident wave is never exactly vertical and so is influenced only by the model on the same side of the heterogeneous medium. Hence we may allow discontinuities to change depth within the medium without having to return to an original depth on the opposite side (Fig. 1).

In applying the analytical form for \( u_e \) one must be aware of potential artefacts caused by differences between an analytical (i.e. continuous) representation of the medium and our finite difference approximation of it. The problem of numerical dispersion was discussed above and a considerable amount of effort was spent minimizing those effects. We note that some investigations in the electromagnetic literature recommend an additional dispersion correction when using analytical plane wave solutions in the time domain (e.g. Schneider 2004; Tan & Potter 2006). We attempted to emulate this recommendation by computing phase velocity corrections in the manner described in Appendix B and applying them to the analytical model prior to computing \( u_e \). However, we found that for the current algorithm such corrections were insignificant at best, and in some cases actually degraded the computed wavefield. Hence we do not apply any analogous correction here.

A more subtle artefact created by the mismatch between continuous and discrete models comes from the appearance of non-physical or ‘ghost’ phases caused when an interface in the background 1-D models is displaced (or otherwise removed) from its original depth within the heterogeneous medium. To understand the cause of this artefact, consider the sources created by application of eq. (6) across an interface separating media with different properties (Fig. 2). Forces of the form \( \mu_{-1/2}(u_+ - u) \) will be created at points just above the interface, and of

![Figure 2. Illustration of the origin of ghost phases generated when an interface aligns with the grid. The two nine-point stars average the medium properties between points connected by the thick black lines. The differences in the forces resulting from eq. (6) are nearly equal but opposite in sign, leading to the creation of a force couple (indicated by horizontal arrows) separated by one grid space at the interface. The waves generated by this force couple can be attenuated either by replacing the interface with a gradual gradient in medium properties or by displacing the interface upwards (to the dashed line) an amount that equates traveltimes in the discrete and continuous representations of the medium between the gridpoints.](https://academic.oup.com/gji/article-abstract/181/2/1017/662526)
the form \( \mu_{1/2}(u - u_-) \) at points along the interface. The net effect is a quasi-couple across the boundary with nearly equal forces in opposite directions. This force couple creates waves that compensate for the mismatch between the analytical and finite difference representations of the medium (as discussed below, waves generated using eq. (6) in a 1-D medium match the analytical form very well), but if the source of that mismatch (i.e. the interface) is removed these compensating waves still exist but no longer play the role of primary phase modifiers. Instead they appear as small amplitude waves that emanate from an invisible interface.

This effect is difficult to eliminate entirely, but because it is caused by the mismatch between continuous and discrete representations of wave speed variations between the gridpoints near the interface, they may be reduced considerably by decreasing this mismatch. One way to do this is to replace the infinitesimal interface by a gradient in wave speed over several gridpoints. Alternatively, one can displace the interface a fraction of a grid spacing upwards from the first gridpoint at or below the interface. An appropriate displacement can be estimated to do this is to replace the infinitesimal interface by a gradient in wave speed over several gridpoints. Alternatively, one can displace the medium (as discussed below, waves generated using eq. (6) in a 1-D medium match the analytical form very well), but if the source of that mismatch (i.e. the interface) is removed these compensating waves still exist but no longer play the role of primary phase modifiers. Instead they appear as small amplitude waves that emanate from an invisible interface.

Without loss of generality, we take

\[
\sigma_{xa} = \mu \left( \frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} \right) = 0
\]

\[
\sigma_{ya} = \mu \left( \frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} \right) = 0
\]

\[
\sigma_{za} = \lambda \left( \frac{\partial v}{\partial x} + \frac{\partial v}{\partial y} \right) + \left( \lambda + 2\mu \right) \frac{\partial w}{\partial z} = 0.
\]

We represent these equations in finite difference form using the templates summarized in Appendix A, and in the 2.5-D case we evaluate the \( \partial \) derivatives above as

\[
\frac{\partial}{\partial z} = \frac{1}{\delta z} \left[ \frac{u(x_0 + \delta z, y, z) - u(x_0 - \delta z, y, z)}{2\delta z} \right]
\]

\[
\frac{\partial}{\partial y} = \frac{1}{\delta y} \left[ \frac{u(x, y_0 + \delta y, z) - u(x, y_0 - \delta y, z)}{2\delta y} \right]
\]

\[
\frac{\partial}{\partial x} = \frac{1}{\delta x} \left[ \frac{u(x + \delta x, y, z) - u(x - \delta x, y, z)}{2\delta x} \right]
\]

3.3 Boundary conditions

The top side of the heterogenous medium generally is taken to be the free surface and we apply the usual traction-free conditions

\[
\sigma_{xa} = \mu \left( \frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} \right) = 0
\]

\[
\sigma_{ya} = \mu \left( \frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} \right) = 0
\]

\[
\sigma_{za} = \lambda \left( \frac{\partial v}{\partial x} + \frac{\partial v}{\partial y} \right) + \left( \lambda + 2\mu \right) \frac{\partial w}{\partial z} = 0.
\]

We now consider the boundary conditions at the heterogenous region.

\[
\delta z = 0.5h \left( \frac{V_{inc}^b}{V_{inc}^a + V_{inc}^b} + 1 \right)
\]

where the superscripts a and b refer to the wave speeds above and below the interface, respectively, and the subscripts inc and con refer to the modes (P or S) of the incident and converted waves, respectively.

\[
\hat{x}_i = x_i^a + \int_{x_i^a}^{x_i^b} e_i(x') dx'
\]

where \( x_i^a \) marks the edge of the PML and \( x_i < x_i^b \) is the normal heterogeneous region. It is straightforward to show that \( e_i \) is related to the d(x) term of Komatitsch & Tromp (2003) and Collino & Tsgoka (2001) by

\[
e_i(x) = 1 - \frac{1}{\alpha} d(x).
\]

Collino & Tsgoka (2001) suggest the following form for d(x)

\[
d(x) = \frac{3\alpha}{2\delta} \left( \frac{x}{R} \right)^2 \log \left( \frac{1}{R} \right) = \frac{Dx^2}{\alpha}
\]

where \( \alpha \) is the P wave speed, \( \delta \) is the width of the PML, and \( R \) is a reflection coefficient generally taken to be 0.001. To apply this concept to the background wavefield we evaluate the integral in (16) to obtain

\[
\hat{x}_i = x_i + \frac{Dx_i^2 + x_i^3}{3}
\]

Without loss of generality, we take \( x_e = 0 \) so that \( x_i \) represents the distance into the PML

\[
\hat{x}_i = x_i \left( 1 - \frac{Dx_i^2}{3} \right)
\]

We then compute \( u(\omega) \) within the PML [\( u_p(\omega) \)] from

\[
u_p(\omega) = A(\omega) \exp[i(\omega t - op_1 y - k_1 \hat{x} - k_\hat{z})]
\]

\[
= u(\omega) \exp \left[ - \frac{Dx_i^2 + x_i^3}{3} \right]
\]

which causes the background wavefield to attenuate exponentially within the PML in a manner consistent with the scattered wavefield.
3.4 The inverse problem

Once a forward problem is implemented there are a number of potential procedures for constructing an inverse problem. Several approaches germane to waveform tomography were investigated for the acoustic case by Pratt et al. (1998). We make use of the gradient approach here partly because of its intuitive character but more because it does not require significantly more computation than the forward problem. We will review the basics of this approach here to illustrate how it can be adapted to the 2.5-D teleseismic case.

We begin by defining a residual vector \( \delta d = u - d \) where \( u \) and \( d \) are the calculated and observed wavefields, respectively. We use \( \delta d \) to create an objective function

\[
E(m) = \frac{1}{2} \delta d^T \delta d,
\]

where the vector \( m \) describes the model. The inverse problem then consists of solving for a perturbation \( \delta m \) to \( m \) that minimizes the objective function. The gradient method estimates \( \delta m \) by taking a step in the direction of the gradient of \( E \)

\[
\delta m = -\gamma \nabla_u E(m),
\]

where \( \gamma \) is a scalar. It is not difficult to show that

\[
\nabla_u E(m) = \frac{1}{2} \nabla_u \delta d^T \delta d = \text{Re} \sum_{i=1}^n \frac{\partial u_i}{\partial m} \delta d^T_i = \text{Re}(J^T \delta d^T),
\]

where \( J \) is the Jacobian. Recalling that the forward problem is \( Su = f \), a derivative with respect to a particular model element \( m_i \) gives

\[
S \frac{\partial u}{\partial m_i} + \frac{\partial S}{\partial m_i} u = 0
\]

and so

\[
\frac{\partial u}{\partial m_i} = -S^{-1} \frac{\partial S}{\partial m_i} u = -S^{-1} f^{(i)},
\]

where we define \( F = f^{(i)} \) as a set of virtual forces. This leads to

\[
\nabla_u E(m) = \text{Re}(J^T \delta d^T) = \text{Re}[(S^{-1} F)^T \delta d^T] = \text{Re}(F^T \nu).
\]

Pratt et al. (1998) argue that as long as reciprocity holds it should be the case that

\[
\nu = (S^{-1})^T \delta d^* = S^{-1} \delta d^*
\]

so that no additional factorization of \( S \) is required to do the inverse problem. When true, this identity significantly reduces the computational effort (LU factorization is the bottleneck in these algorithms), and is why the gradient method was chosen for this application: the inverse problem in this case requires only slightly more computation than the forward problem.

We note that in general if \( S \) is full rank then a singular value decomposition of a complex valued \( S = U \Lambda V^H \) can be manipulated to show that

\[
(S^{-1})^T = (S^T)^{-1}
\]

so that (28) is true as long as \( S \) is symmetric. Inspection of the templates that make up \( S \) (Appendix A) shows that in the 2-D case it is indeed symmetric everywhere except within the PML, the reason being that direction of propagation is consequential within these boundary layers (and so reciprocity no longer applies). It is not difficult to compute \((S^T)^{-1}\), in fact we can do this with the same LU factorization of \( S \) (one simply solves the system of equations using \( S^T = U^T \Lambda^T \)). However, we argue that because the PML is an artificial construct, and because \( S \) would be symmetric in an infinite grid with no artificial boundaries, it is in fact more correct to use \( S^{-1} \) in (28).

In the 2.5-D case \((p_j \neq 0)\), the part of \( S \) outside the PML is Hermitian and thus no longer strictly symmetric (i.e. \( S^T \neq S^H \)). In this case we can write \( S = U \Lambda U^H \) and

\[
(S^{-1})^T = U^H \Lambda^{-1} U^T = U^* \Lambda^{-1} U^T = (S^*)^{-1},
\]

where \( U^* \) refers to the complex conjugate of each element of the eigenvector matrix \( U \) without transposing, while \( U^H \) is the complex conjugate transpose of \( U \). For a Hermitian matrix, \( S^* = S^T \), so (29) still applies, but it is no longer the case that we can simply use the inverse of the \( S \) itself in (28). We should instead use

\[
\nu = S^{-1} \delta d^*
\]

in (28) but recall that \( S^* \) refers only to the part of \( S \) outside the PML. Because terms in the PML are complex, we form \( S^* \) simply by reversing the sign on \( p_j \).

Pratt et al. (1998) identify \( \nu \) in (28) as a wavefield generated by back propagating time reversed scattered waves from the receivers using the same propagator matrix as the forward problem. We retain the same back propagation aspect in (30), and in this case reversing the sign on \( p_j \) is equivalent to sending the wave back in the opposite \( y \) direction from whence it arrived.

The \( F \) term in (27) is the forward propagated wavefield modified by a scattering factor \( \frac{\partial u_i}{\partial m} \). In the acoustic case this factor is a single scalar that is dependent on frequency and medium slowness. As Pratt et al. (1998) point out, because a product of a function with the complex
conjugate of another function in the frequency domain is a correlation of these functions in the time domain, the gradient in (27) is simply a weighted correlation between predicted and observed scattered wavefields at every point in the medium. When the correlation is large, the corresponding perturbation to the model will be large as well. A similar intuitive interpretation applies in the 2.5-D elastic case, although as we have formulated it here, each $m_i$ appears in several elements of $S$. As a result, instead of a single scaler multiplying the forward wavefield, each $\frac{\partial S}{\partial m_i}$ is represented by a $27 \times 27$ submatrix with as many as 369 non-zero elements (Appendix C). This much larger number of elements makes a simple generalization of their effects less straightforward but at the same time does not entail a significant increase in computational effort.

3.5 Source and receiver statics

We consider the plane wave impinging on the heterogenous medium from below to be associated with a particular source time function (STF) which in general includes all of the effects on the incident wave prior to its arrival at the model base. These would include the history of energy release at the source and the interaction of the waves (reflections and mode conversions) that take place near the source region. The STF usually is considered to be an unresolvable ‘nuisance’ parameter in imaging algorithms. In traveltime tomography it takes the form of an unknown origin time that we remove by substituting relative for absolute wave speeds in the inverse problem. In receiver function analysis the STF typically is eliminated by deconvolving one component of motion from another. For example, in $P$-receiver function modelling we represent the radial ($r$) and vertical ($z$) components of observed motion as

$$z(\omega) = s(\omega)G_z(\omega)i_z(\omega)$$

$$r(\omega) = s(\omega)G_r(\omega)i_r(\omega)$$

where $s(\omega)$ is the source time function, $i_r$ and $i_z$ are instrument responses that generally are known and removed prior to analysis, and $G_z$ and $G_r$ are the vertical and radial components of the model Green’s functions. We then employ one of several algorithms in the time or frequency domain to deconvolve the $z$ component from the $r$ component to arrive at a form resembling

$$z(\omega) = \frac{G_z(\omega)}{G_r(\omega)}$$

The STF is a nuisance parameter in teleseismic waveform tomography as well, but if we have an array of stations recording the same event we can adapt a technique suggested by Pratt (1990) for removing the source contribution from explosions in the active source environment. As stated in eq. (31), the wavefield is a convolution of the STF with the medium Green’s functions, and we can write (6) as

$$Su = S[G(\omega)s(\omega)] = f_s.$$  

We estimate the STF by adjusting $s(\omega)$ to minimize the difference between observed and calculated wavefields. That is, if the set $d$ of observations from a given event satisfies

$$Sd = f_s$$

then we seek a STF $\tilde{s}(\omega)$ such that

$$S[G(\omega)\tilde{s}(\omega)] = f_s.$$  

To do so, we modify the STF at any particular frequency by a complex scaler $c$ so that

$$S[G(\omega)c\tilde{s}(\omega)] = S[G(\omega)(c\tilde{s}(\omega))] = S[Su] = cSu = cf_s.$$  

Hence, to approach (34) we minimize the difference between $d$ and $c u$. If we pose this as the minimization of an objective function

$$E(c) = \frac{1}{2} \sum_{i=1}^{n} (d - cu)^T(d - cu)$$

then setting $\frac{\partial E(c)}{\partial c} = 0$ leads to

$$c = \frac{u^T d}{u^T u}$$

which is identical to the source correction term derived by Pratt (1990). Intuitively, (36) shows that $c$ is determined by a correlation of observed ($d$) and predicted ($u = S^{-1} f$) wavefields at each station (normalized by the predicted wavefield). It is similar to stacking, but in this case the stack is weighted by the expected medium response. Hence we solve for $c$ each time we perturb the medium rather than simply determine an average wavefield at the outset (note that if $u$ is a delta function, then $c$ is a simple stack of observed waveforms). It is also analogous to eliminating origin times in traveltime tomography, in that parts of the waveform that appear to be in some sense ‘common’ to all stations are assigned to the source time function. eq. (36) allows for timing adjustments in the form of phase shifts, but these shifts are frequency dependent and an investigator needs to be mindful of the potential for cycle skips. Rather than depend solely on time adjustments calculated from (36) one would be well advised to generate a traveltime tomography image to use as a starting model.

An advantage of this approach over traditional teleseismic receiver functions is that the usual initial step of source deconvolution, with all its attendant assumptions and potential for noise enhancement, is not required. Instead, a source time function can be determined for each
event, and the response of the medium to this function can be calculated explicitly. The computation is economical: we compute the Green's functions in \( S^{-1} \) only once for a given \( p \), and loop through all sources \( f \) to generate the wavefields \( u \) from which we can determine \( c \).

A common concern in teleseismic traveltime tomography is timing offsets at stations caused by local wave speed anomalies, and a standard method for compensating for these offsets is to subtract a station-based average residual (i.e. a 'station correction'). In the waveform tomography approach an algorithm similar to that described above for the STF can be applied to compute receiver statics. In this case we gather the residual wavefields recorded at a given station as a vector \( \delta d = d - u \) and minimize the energy in these residuals by scaling the observables to obtain \( d = r \). Using the same procedure as above, we compute the value of \( r \) that minimizes the energy in \( \delta d = r \) and find that
\[
| r = \frac{ud^*}{dd^*}. \tag{37} \]
Note that a station correction determined in this fashion can correct not only for time delays (again being congnizant of potential cycle skips) but also for frequency dependent modifications to the waveforms such as resonances at stations located in basins.

### 4 RELATING WAVEFORM TOMOGRAPHY TO RECEIVER FUNCTION MIGRATION AND TRAVELTIME TOMOGRAPHY

The waveform tomography approach described here can be expected to improve images of the subsurface by extracting more information from seismograms, but there is a trade off in an elevated sensitivity to noise and a non-linearity that can make reaching a correct minimum through a series of linear steps difficult. As mentioned above, a classic non-linearity in spectral analysis is cycle skips; if we are more than \( \pi \) radians out of phase, most algorithms will shift the phase in the wrong direction. For this reason it is generally advisable to have a reasonable starting model using whatever information can be gleaned from other techniques. As mentioned at the outset, the images that come from traveltime tomography and receiver function migration may be fuzzy or beset by ambiguities, but have an advantage in being relatively simple and robust so that investigators still can manage to recover some useful first order information about the medium.

In the event that background information or noise levels are inadequate to allow a full waveform tomography approach, one can still use parts of the algorithm described above to recover information about the medium in a manner analogous to travel time inversion and receiver function migration, but in ways that are arguably superior to either of these standard techniques. For example, one can generate a finite frequency traveltime tomography image by normalizing amplitudes and solving eq. (27) for phase shifts (i.e. time delays) after windowing the primary (P or S) phase. An analogous approach was employed by Maggi et al. (2009) and Tape et al. (2009). If \( a priori \) information about the subsurface is lacking, cycle skips can be reduced by fitting the low frequency end of the spectrum first before attempting a broad band inversion. We can also emulate receiver function migration by treating the deconvolved waveform in (32) as the residual field \( \delta d^* \) in (27) and (28), and backpropagating it into the medium from the stations. In this way, migration is equivalent to calculating the gradient in (27) without the additional step of solving for \( \delta m \). An advantage of performing migration in this fashion is that we can more accurately account for the effects of lateral heterogeneity and multiple scattering than is possible with most standard techniques.

### 5 EXAMPLES

In this section, we discuss some simple examples that illustrate the basic concepts of teleseismic waveform tomography. The examples were generated using a FORTRAN code that implements the algorithms described above. Most of the computation was done on a Quad core AMD Opteron (2.2 GHz per core) with 16 Gb of RAM, but the examples were also run successfully on a MacBook Pro with a 2.5 GHz Intel Core Duo. The grid in the examples has 400 points in the \( x \) direction and 100 points in the \( z \) direction. The grid spacing is 1 km, which makes the model size about 400 km and also sets the minimum wavelength that we can expect to model without significant numerical dispersion to about 4 km. A 10 point (9 km) wide PML is included along the bottom and sides of the model. For reasons of stability, when we do the inversion we apply a mask to the PML and to the uppermost two gridpoints at the free surface, which keeps the wave-speeds near the boundaries constant.

Computation of the wavefield for a single frequency on the 400 \( \times \) 100 grid required about 20 s on the Opteron and about 50 s on the MacBook Pro. Nominally a complete forward problem would require on the order of 40 frequencies, meaning that an entire suite of Green's functions can be determined for a moderately sized grid on a modest desktop computer in less than an hour.

#### 5.1 Waveform synthetics

In the process of testing the code we generated waveforms for a wide range of trial models, and animations generated from several of these tests can be found online.\(^2\) We review two examples here. The first tests the approach of eq. (11) for computing the plane wave excitation of 2.5-D heterogeneity. The heterogeneity in this case is a slab with wave speeds 5 per cent higher than those in a surrounding constant wave speed background (Fig. 3). While no analytical solution to the wave equation exists for this case, it should happen that the response at stations

\(^2\) [http://gretchen.geo.rpi.edu/rocker/WT/ElasticFD.html](http://gretchen.geo.rpi.edu/rocker/WT/ElasticFD.html)
Figure 3. Model used to verify the computation of the response of heterogeneous media to plane wave excitation using eq. (10). The trial medium (panel B) is a high wave speed (8.4 km s\(^{-1}\)) slab embedded in a constant wave speed (8 km s\(^{-1}\)) background. A scattered field is computed using sources located within the slab, shown in panel C for the real part of the source functions at 0.05 Hz. At points located far from the edges of the slab, the total wavefield should approximate the case of the 1-D layer in panel A, for which analytical solutions are known. The trial plane wave arrives at a 20\(^\circ\) angle of incidence, as shown in the inset in panel B. Closed triangles locate the point where the seismograms shown in Fig. 4 are computed.

far from the edges of the slab should closely approximate the response of a 1-D medium (i.e. an infinite slab), for which the response is known. We use eq. (10) to compute a scattered wavefield (\(u_h\)) from a distribution of sources in the slab (panel C in Fig. 3) that can be added to the (known) half-space response to recover an approximation to the layered medium response. The analytic and numerical waveforms (traces C and F in Fig. 4) are quite similar, the main difference being a slight underestimation of the amplitude of the main phase caused by

Figure 4. Displacement seismograms computed at the free surface of the slab model (Fig. 3) at \(x = 300\) km due to a plane wave entering the medium at a 20\(^\circ\) angle of incidence. Trace ‘A’ is the vertical component of ground motion (\(u_h\)) due to the scattering sources located within the slab (panel C in Fig. 3), while ‘B’ is the displacement due to the constant wave speed half-space (\(u_0\)). Traces ‘C’ show the sum of the upper two seismograms (\(u_t = u_h + u_0\)) plotted as a dashed line on a solid line representing the analytical solution for the 1-D medium shown in panel A of Fig. 3. Traces D, E and F are the analogous results for the horizontal component of motion. Amplitudes of the horizontal components are magnified as indicated by the scales shown on the vertical axes.
approximating a 1-D layer by a bounded slab, which ignores the small but non-zero contributions of scattering sources that would extend out beyond the edges of the slab.

The second example demonstrates how the algorithm works in a model that has large wave speed gradients in both horizontal and vertical directions. The choice of model dimension for this test was intended to represent the Tien Shan region of central Asia, where a significant amount of effort has already gone into generating traveltime and receiver function images (e.g. Roecker et al. 1993; Vinnik et al. 2002a,b; 2004; Li et al. 2009). One of the intriguing results of those investigations was evidence for an undulating Moho with remarkably large excursions from a mean depth of about 45 km (e.g. Oreshin et al. 2002). We reproduce this environment in a simplified manner here by specifying a constant wave speed crust (6 km s$^{-1}$) over a constant wave speed mantle (8 km s$^{-1}$) separated by a thin interface (Fig. 5).

Figure 5. Snapshots of the wavefield generated by an incident plane $P$ wave on a medium composed of a low wave speed medium (6 km s$^{-1}$) over a high wave speed medium (8 km s$^{-1}$) separated by an undulating surface indicated by the black line at about 40 km depth. The incident wave enters from the lower left-hand corner at an angle of 20° from vertical. Left- and right-hand panels show, respectively, vertical and horizontal components of displacement. The top, middle and bottom rows correspond to times of 11, 22 and 33 s, respectively, after the onset at the lower left-hand corner at 100 km depth. Deep reds and blues correspond to large plus and minus amplitude and all displacements are plotted at the same scale.

Figure 6. Record sections of displacement seismograms for the example shown in Fig. 5. The panels on the left-hand side show the plane wave excitation for the case of a horizontal Moho at 40 km depth, which is also the background medium used to compute $u_0$. Panels on the right are the total field for the undulating Moho shown in Fig. 5. The upper and lower rows correspond to vertical and horizontal components of motion, respectively. Note the reduced, but still discernable, presence of the 1-D background mode conversions ('ghost' phases) in the sections on the right-hand side.
Wavefield snapshots (Fig. 5) illustrate the response of this crustal/upper mantle model to an incident P wave approaching from the lower left at a 20° angle of incidence. Clearly visible in the panels are reflections from the curved parts of the Moho and focussing and defocussing caused by wave speed changes and refractions on the boundary. SV mode conversions show up clearly on the horizontal components of motion (panels B and D in Fig. 5). Multiply reflected phases entering from the stratified medium to the left of the modelled region are clearly seen at the left-hand side of panels C and D. Record sections (Fig. 6) computed at points at the surface illustrate the phenomenon of ‘ghost’ phases described above, as the later arriving phases generated in the 1-D background model (left-side panels of Fig. 6) appear as faint, but still discernible, phases in the 2.5-D model. As discussed above, it is difficult to eliminate these artefacts entirely, but we can significantly reduce their amplitude by displacing the 1-D interface upwards by about half a grid spacing.

5.2 Tomography example

To illustrate some of the key concepts of teleseismic waveform tomography we consider a simple model of a half-space composed of a Poisson solid with constant 8 km s\(^{-1}\) compressional wave speed in which a block 40 km on a side with wave speeds 5 per cent higher than the background is embedded (Fig. 7). Synthetic seismograms are generated by calculating the responses to incident plane P waves from six events. Three of these are at a backazimuth of 0° (i.e. waves approaching from the right-hand side in Fig. 7) and three are from 180° (waves approaching from the left-hand side) and with angles of incidence of 20°, 30° and 40° with respect to vertical. The source time function of the incident wave is a Küber wavelet (Küber 1958) with a dominant frequency of 0.5 Hz and a bandwidth from 0.02 to 1 Hz. The wavefield

![Figure 7](https://academic.oup.com/gji/article-abstract/181/2/1017/662526)
Figure 8. Results of a waveform tomography inversion for $S$ wave speed. The target image is the same as that shown in Fig. 7. Scaling of the palette is indicated at the lower left of each image. (A). Reconstruction after the first iteration starting with a constant wave speed background. (B). Revised starting model for second iteration based on the $P$ wave speed model obtained after the first iteration. (C) and (D). Reconstruction after the second and third iterations.

is recorded by 300 receivers placed 1 km apart along the surface between $x = 50$ and 350 km. The starting model in the inversion is the constant background, and three iterations are performed to produce the final results shown in Figs 7 and 8.

Besides being a common type of observation in body wave analysis, modelling an incident $P$ wave and the scattered waves in its coda provides a good example of how waveform tomography incorporates aspects of both traveltime tomography and receiver function migration in creating an image. Because the primary $P$ arrival is much more energetic than the scattered phases, those parts of the Jacobian in (23) related to $P$ wave speeds will be sensitive primarily to the phase shifts caused by the $P$ wave speed anomaly. To see that this is the case, we compare an image reconstructed using the full waveform with one using only phase shifts. The results in Fig. 7 show that while the edges on the full waveform image are somewhat better defined, the phase-only inversion does nearly as well in reconstructing the image.

In contrast to the $P$ wave speeds, there is very little information about $S$ wave speeds in the $P$ coda, and the $Ps$ conversions from the sides of the box are the dominant residual shear waves that the tomography algorithm migrates back into the medium. As shown in panel A in Fig. 8, these residual waves converge in the general vicinity of the edges of the box, but focussing is poor due to the inaccuracy of the
model used to perform the migration. Specifically, the sides of the box appear to dip inwards because the background is 5 per cent slower than it should be. Note also that because all of the residuals are migrated, the misfit to the primary \( P \) phase shows up as large amplitude signals near the top of the model. If we tie the \( S \) wave speeds to the imaged \( P \) wave speeds as shown in Fig. 8(b), the migrated image is much better focussed, and on the third iteration the edges of the box are clearly seen even though the wave speeds within the box are still not well resolved.

We note that while this test reconstructs what appears to be a fairly well resolved image with much less data (six events and 1800 arrivals) than that usually available for tomographic imaging, we compensated for this lack of information by fixing the source time function and not contaminating the data with any but a small amount of numerical noise. The value of tests that quantify the influence of noise and geometry on an image usually depend strongly on the particular data being analysed and so will wait for an investigation involving an actual data set.

6 EXTENSIONS

This paper outlines a basic 2.5-D full waveform teleseismic tomography algorithm, and the discussion has focussed on the case of heterogeneous, elastic media. There are additional techniques that can be adapted from the acoustic waveform tomography approach; for example, an extension from elastic to viscoelastic (i.e. finite \( Q \)) media can be achieved by using complex wave speeds. Simple forms of vertical anisotropy can be included by stretching the coordinate system in the \( z \) direction. More sophisticated methodology like the receiver function approach of Levin & Park (1997) would be possible to employ in computing a background wavefield, but modelling propagation in a general heterogeneous, anisotropic media would most likely require significant computing resources. Topography or submarine environments that involve media with a Solid–liquid interface can be modelled using a combination of elastic and acoustic equations when forming the difference matrix \( S \); methods for computing background fields in these cases are well known. For example, topography may be accounted for using a modified version of the transformed grid technique of Ruud & Hestholm (2001). Finally, teleseismic body waves could be combined with other types of waveforms in a joint inversion. An active source line parallel to the \( x \) axis would be the easiest to incorporate as the original RGP algorithm was designed for this type of application. Surface waves could also be included; as with body waves one need only be able to specify a known background wavefield. Local earthquakes could be included provided both sources and receivers are located within the \((x, z)\) plane because the wavefield in this case is no longer invariant along the \( y \) direction. Nevertheless, one might argue that 2-D sections of the 2.5-D model still allow reasonable approximations of the wavefield.

Our intention in this study was to provide one example of a tractable inverse problem using gradients, but there are several other approaches that one could consider. In particular, Pratt et al. (1998) discuss several alternatives for the acoustic case that could be easily adapted to the viscoelastic teleseismic case. Also, where the location of specific boundaries is an important part of the inverse problem (for example, when focussing on receiver function migration) one might consider inverting for interface topography instead of wave speeds at individual gridpoints, as suggested for example by Dahlen (2005).

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation grants CMG-0327634 and EAR-0838384. We would like to acknowledge Carl Tape and an anonymous reviewer for providing comments that significantly improved the content of this paper. We benefited from the availability of the source code for the acoustic wave algorithms \textsc{Omega} and \textsc{Fullwv} provided by R. G. Pratt, which greatly facilitated
the development of our own code. We acknowledge and appreciate the many years of labour by a number of contributors that went into the development of those codes.

REFERENCES


APPENDIX A: FINITE DIFFERENCE FORMS

Our finite difference forms are defined on a nine-point star shown diagrammatically as

\[ \mu_- \mu_- \mu_-^+ \]
\[ \mu_- \mu \mu_+^+ \]
\[ \mu_+ \mu_+ \mu_+^+ \]

We also define average medium properties between these points as (in this case for \( \mu \))

\[ \mu_- \mu_-^{1/2} \mu_- \mu_-^{1/2} \mu_+ \]
\[ \mu_- \mu_-^{1/2} \mu_- \mu_-^{1/2} \mu_- \]
\[ \mu_- \mu_-^{1/2} \mu_- \mu_-^{1/2} \mu_- \]
\[ \mu_- \mu_-^{1/2} \mu_- \mu_-^{1/2} \mu_- \]
\[ \mu_- \mu_-^{1/2} \mu_- \mu_-^{1/2} \mu_- \]
\[ \mu_- \mu_-^{1/2} \mu_- \mu_-^{1/2} \mu_- \]
\[ \mu_- \mu_-^{1/2} \mu_- \mu_-^{1/2} \mu_- \]
\[ \mu_- \mu_-^{1/2} \mu_- \mu_-^{1/2} \mu_- \]

where for example, \( \mu_-^{1/2} = (\mu + \mu_-) / 2 \). The general forms of the stencils used in the finite difference formulation in the non-rotated frame are

\[ \frac{\partial u}{\partial x} = \left[ \mu_-^{1/2}(u_+ - u) + \mu_-^{1/2}(u - u_-) \right] / 2\Delta \]
\[ \frac{\partial u}{\partial z} = \left[ \mu_-^{1/2}(u_+ - u) + \mu_-^{1/2}(u - u_-) \right] / 2\Delta \]
\[ \frac{\partial}{\partial x} \left[ \mu \frac{\partial u}{\partial x} \right] = \left[ \mu_-^{1/2}(u_+ - u) + \mu_-^{1/2}(u - u_-) \right] / \Delta^2 \]
\[ \frac{\partial}{\partial z} \left[ \mu \frac{\partial u}{\partial z} \right] = \left[ \mu_-^{1/2}(u_+ - u) + \mu_-^{1/2}(u - u_-) \right] / \Delta^2 \]
\[ \frac{\partial}{\partial x} \left[ \mu \frac{\partial u}{\partial x} \right] = \left[ \mu_-^{1/2}(u_+ - u) + \mu_-^{1/2}(u - u_-) \right] / 4\Delta^2 \]
\[ \frac{\partial}{\partial z} \left[ \mu \frac{\partial u}{\partial z} \right] = \left[ \mu_-^{1/2}(u_+ - u) + \mu_-^{1/2}(u - u_-) \right] / 4\Delta^2 \]
and in the 45° rotated frame

$$\frac{\partial u}{\partial x'} = \frac{[\mu_{+1/2}(u_+ + u) + \mu^{-1/2}(u - u_-)]}{2\sqrt{2} \Delta}$$

$$\frac{\partial u}{\partial z'} = \frac{[\mu_{+1/2}(u_+ - u) + \mu^{-1/2}(u_+ - u_-)]}{2\sqrt{2} \Delta}$$

$$\frac{\partial }{\partial x'}[\mu u] = \frac{[\mu_{+1/2}(u_+ + u) - \mu^{-1/2}(u_+ + u_-)]}{2\sqrt{2} \Delta}$$

$$\frac{\partial }{\partial z'}[\mu u] = \frac{[\mu_{+1/2}(u_+ - u) - \mu^{-1/2}(u_+ - u_-)]}{2\sqrt{2} \Delta}$$

$$\frac{\partial }{\partial x'}[\mu u] = \frac{[\mu_{+1/2}(u_+ - u) - \mu^{-1/2}(u_+ - u_-)]}{2\Delta^2}$$

$$\frac{\partial }{\partial z'}[\mu u] = \frac{[\mu_{+1/2}(u_+ + u) - \mu^{-1/2}(u_+ + u_-)]}{2\Delta^2}$$

$$\frac{\partial }{\partial x'}[\mu u] = \frac{[\mu_{+1/2}(u_+ - u) - \mu^{-1/2}(u_+ - u_-)]}{2\Delta^2}$$

$$\frac{\partial }{\partial z'}[\mu u] = \frac{[\mu_{+1/2}(u_+ + u) - \mu^{-1/2}(u_+ + u_-)]}{2\Delta^2}$$

When computing the optimizing weights of the sum of the two frames to minimize numerical dispersion (Appendix B), we use homogeneous versions for the non-rotated frame

$$\frac{\partial u}{\partial x} = \frac{u^+ - u^-}{2\Delta}$$

$$\frac{\partial u}{\partial z} = \frac{u_+ - u_-}{2\Delta}$$

$$\frac{\partial^2 u}{\partial x^2} = \frac{u^+ - 2u + u^-}{\Delta^2}$$

$$\frac{\partial^2 u}{\partial z^2} = \frac{u_+ - 2u + u_-}{\Delta^2}$$

$$\frac{\partial^2 u}{\partial x \partial z} = \frac{u^+ + u^- - u_+ - u_-}{4\Delta^2}$$

and in the 45° rotated frame

$$\frac{\partial u}{\partial x'} = \frac{u^+ - u^-}{2\sqrt{2} \Delta}$$

$$\frac{\partial u}{\partial z'} = \frac{u_+ - u_-}{2\sqrt{2} \Delta}$$

$$\frac{\partial^2 u}{\partial x'^2} = \frac{u^+ - 2u + u^-}{2\Delta^2}$$

$$\frac{\partial^2 u}{\partial z'^2} = \frac{u_+ - 2u + u_-}{2\Delta^2}$$

$$\frac{\partial^2 u}{\partial x' \partial z'} = \frac{u_+ + u^- - u^+ - u^-}{2\Delta^2}$$

**APPENDIX B: OPTIMIZATION OF GRIDS FOR NUMERICAL DISPERSION IN THE 2.5-D ELASTIC CASE**

Following Jo et al. (1996) and Štekl & Pratt (1998) we use a weighted sum of rotated grids to minimize the effects of numerical dispersion. We quantify these effects by determining the difference between true compressional (α) and shear (β) velocities in the medium and the associated phase (ω) and group (κ) velocities predicted by the finite difference forms. We define a set of functionals

$$E = \max \left[ 1 - \frac{v_{grid}}{v_{true}} \right]^2$$

and adjust the weights to minimize an integral of these functionals

$$\frac{1}{2\pi} \int_0^{2\pi/\Omega} \int_0^{\pi/4} \max \left[ E_{\rho \beta}^E, E_{\rho \beta}^{gr}, E_{\rho \beta}^S, E_{\rho \beta}^{gr}, E_{\rho \beta}^S \right] \, d\theta \, dk,$$

© 2010 The Authors, *GJI*, **181**, 1017–1040

Journal compilation © 2010 RAS
where $k^* = \frac{16}{27} = \frac{4}{3}$, $\Delta$ is the grid spacing, $\theta$ is the rotation angle, and $E^2_{\mu\nu}$, etc. refers to the functionals defined above evaluated using the phase (ph) or group (gr) velocities for three phases, one compressional ($C$) and two shear ($S_1$ and $S_2$), that in the 2-D case correspond to $P$, $SV$, and $SH$. We start with the 3-D elastic wave equation in homogeneous media

$$-\omega^2 \rho u_i = (\lambda + \mu) \frac{\partial u_j}{\partial x_j} + \mu \frac{\partial^2 u_i}{\partial x_j^2}$$

(B3)

which in 2.5-D becomes

$$\omega^2 (\beta^2 \rho^2 - 1) u = \alpha^2 \frac{\partial^2 u}{\partial x^2} + \beta^2 \frac{\partial^2 u}{\partial y^2} + (\alpha^2 - \beta^2) \frac{\partial^2 u}{\partial x \partial z} - i \omega \rho (\alpha^2 - \beta^2) \frac{\partial u}{\partial x}$$

(B4)

$$\omega^2 (\alpha^2 \rho^2 - 1) v = \beta^2 \left( \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right) - i \omega \rho (\alpha^2 - \beta^2) \left( \frac{\partial u}{\partial x} + \frac{\partial w}{\partial y} \right)$$

(B5)

$$\omega^2 (\beta^2 \rho^2 - 1) w = \alpha^2 \frac{\partial^2 w}{\partial x^2} + \beta^2 \frac{\partial^2 w}{\partial y^2} + (\alpha^2 - \beta^2) \frac{\partial^2 u}{\partial x \partial z} - i \omega \rho (\alpha^2 - \beta^2) \frac{\partial v}{\partial x}$$

(B6)

We proceed with a von Neumann analysis by introducing a plane wave solution ($u_i, v, w, \phi, \psi, \chi$) into the 3-D elastic wave equation and collecting the finite difference forms we can write the combined wave eq. (B8) as

$$\begin{bmatrix} M_1 & 0 & 0 & u_0 \\ 0 & M_1 & 0 & v_0 \\ 0 & 0 & T_1, M_1 & w_0 \end{bmatrix} \begin{bmatrix} u_0 \\ v_0 \\ w_0 \end{bmatrix} = \frac{1}{\Delta^2} \begin{bmatrix} K_{11} & o \rho, K_{21} \Delta & K_{31} \\ o \rho, K_{21} \Delta & K_{22} & o \rho, K_{23} \Delta \\ K_{31} & o \rho, K_{32} \Delta & K_{33} \end{bmatrix} \begin{bmatrix} u_0 \\ v_0 \\ w_0 \end{bmatrix}$$

(B11)

The Authors, *GJI*, 181, 1017–1040

Journal compilation © 2010 RAS

© 2010 The Authors, *GJI*, 181, 1017–1040

Journal compilation © 2010 RAS
where
\[ M_1 = b + \frac{1 - b}{4} (2 \cos k_\Delta + 2 \cos k_\Delta) \]
\[ = b + \frac{(1 - b)}{2} (\cos (k_\Delta \cos \theta) + \cos (k_\Delta \sin \theta)) \]
\[ K_{11} = a \left[ \frac{\partial^2 u}{\partial x^2} + R \frac{\partial^2 u}{\partial z^2} \right] + (1 - a) \left[ \frac{1 + R^2}{2} \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial z^2} \right) - (1 - R^2) \frac{\partial^2 u}{\partial x \partial z} \right] \]
\[ = 2a [ \cos (k_\Delta \cos \theta) - 1 + R^2 (\cos (k_\Delta \sin \theta) - 1)] \]
\[ + (1 - a) [(1 + R^2) (\cos (k_\Delta \cos \theta) \cos (k_\Delta \sin \theta) - 1) - (1 - R^2) (\cos (k_\Delta \sin \theta) - \cos (k_\Delta \cos \theta))] \]
\[ K_{21} = -i (1 - R^2) \left[ a \frac{dv}{dx} - \frac{(1 - a)}{\sqrt{2}} (\frac{dv}{dx'} - \frac{dv}{dz'}) \right] \]
\[ = -(1 - R^2) [a \sin (k_\Delta \cos \theta) + (1 - a) \sin (k_\Delta \cos \theta) \cos (k_\Delta \sin \theta)] \]
\[ K_{31} = a (1 - R^2) \frac{\partial^2 u}{\partial x \partial z} + (1 - a) \frac{1 - R^2}{2} \left( \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial z^2} \right) \]
\[ = - (1 - R^2) \sin (k_\Delta \cos \theta) \sin (k_\Delta \sin \theta) \]
\[ K_{22} = R^2 \left[ a \left( \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial z^2} \right) + (1 - a) \left( \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial z^2} \right) \right] \]
\[ = 2R^2 [a (\cos (k_\Delta \cos \theta) + \cos (k_\Delta \sin \theta) - 2) + (1 - a) (\cos (k_\Delta \cos \theta) \cos (k_\Delta \sin \theta) - 1)] \]
\[ K_{32} = -i (1 - R^2) \left[ a \frac{dw}{dx} + \frac{(1 - a)}{\sqrt{2}} (\frac{dw}{dx'} + \frac{dw}{dz'}) \right] \]
\[ = -(1 - R^2) [a \sin (k_\Delta \sin \theta) - (1 - a) \sin (k_\Delta \sin \theta) \cos (k_\Delta \cos \theta)] \]
\[ K_{33} = a \left( \frac{\partial^2 w}{\partial z^2} + R \frac{\partial^2 w}{\partial z^2} \right) + (1 - a) \left[ \frac{1 + R^2}{2} \left( \frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial z^2} \right) + (1 - R^2) \frac{\partial^2 w}{\partial x \partial z} \right] \]
\[ = 2a [ \cos (k_\Delta \sin \theta) - 1 + R^2 (\cos (k_\Delta \cos \theta) - 1)] \]
\[ + (1 - a) [(1 + R^2) (\cos (k_\Delta \cos \theta) \sin (k_\Delta \sin \theta) - 1) \]
\[ + (1 - R^2) (\cos (k_\Delta \sin \theta) - \cos (k_\Delta \cos \theta))] \].

Identifying the eigenvalue problem for \( \omega \Delta \) in (B11) gives us the form
\[ \omega^6 \frac{T_{11}^2 T_{21}^2 M_1^3}{\Delta^4} - \omega^4 \frac{T_{11}^2 T_{21}^2 M_1^3 (K_{11} + K_{33})}{\Delta^2} - \omega^4 \frac{T_{11}^2 M_1^3 (K_{21} + K_{11})}{\Delta^4} \]
\[ + \omega^2 \frac{T_{11}^2 M_1 (K_{11} K_{22} + K_{22} K_{33})}{\Delta^2} + \omega^2 \frac{T^2 M_1 (K_{11} K_{33} - K_{21})}{\Delta^2} \]
\[ + \omega^2 \frac{p_1^2 (K_{11}^2 K_{11} + K_{22}^2 K_{33} - 2K_{11} K_{22} K_{33})}{\Delta^4} \]
\[ + \frac{1}{\Delta^4} K_{11}^2 K_{22} - K_{11} K_{22} K_{33} = 0 \]

or rearranging
\[ (\omega \Delta)^6 - (\omega \Delta)^4 \left[ \frac{(K_{11} + K_{33})}{T_{11}^2 M_1^3} + \frac{K_{22}^2}{T_{22}^2 M_1^3} + \frac{p_1^2}{T_{11}^2 T_{22} M_1^3} (K_{11}^2 + K_{22}^2) \right] \]
\[ + (\omega \Delta)^2 \left[ \left( K_{11} K_{22} + K_{22} K_{33} \right) \frac{K_{11} K_{33} - K_{21}}{T_{11}^2 T_{22} M_1^3} + \left( K_{11} K_{33} - K_{21}^2 \right) \frac{K_{11}^2 + K_{22}^2}{T_{11}^2 T_{22} M_1^3} - 2K_{11} K_{22} K_{33} \frac{K_{11} K_{33} - K_{21}}{T_{11}^2 T_{22} M_1^3} \right] \]
\[ + \frac{K_{11}^2 K_{22} - K_{11} K_{22} K_{33}}{T_{11}^2 T_{22} M_1^3} = 0. \]

We now have a problem of the form \( x^3 + ax^2 + bx + c = 0 \) where \( x = (\omega \Delta)^2 \). Following Press et al. (1992), to solve this cubic with real coefficients we first let
\[ Q = \frac{a^2 - 3b}{9}; \quad R = \frac{2a^3 - 9ab + 27c}{54}. \]

B12
Since \( Q \) and \( R \) will in this case always be real and \( R^2 < Q^3 \) the three real roots in this case are given by:

\[
x_1 = -2\sqrt{Q} \cos \left(\frac{\theta}{3}\right) - \frac{a}{3}
\]

(B13a)

\[
x_2 = 2\sqrt{Q} \cos \left(\frac{\theta + \pi}{3}\right) - \frac{a}{3}
\]

(B13b)

\[
x_3 = 2\sqrt{Q} \cos \left(\frac{\theta - \pi}{3}\right) - \frac{a}{3},
\]

(B13c)

where we have defined

\[
\theta \equiv \cos^{-1} \left(\frac{R}{\sqrt{Q}}\right).
\]

(B14)

Evaluation of these roots indicates that \( \theta \) is close to \( \pi \) which makes \( x_1 \) and \( x_2 \) approximately equal and always less than \( x_3 \). Hence we associate the two lower frequency components with shear wave speeds (\( x_1 \Rightarrow S_1, x_2 \Rightarrow S_2 \)) and the higher frequency \( x_3 \) with compressional wave speed (\( x_3 \Rightarrow C \)). In the 2-D case, \( \theta = \pi \) and \( x_1 = x_2 \). The fact that they are different when \( p_i \neq 0 \) implies a directional dependance, or numerical anisotropy, for shear waves. Recalling that \( x_i = (\omega_i \Delta)^2 \), the numerical phase velocities are

\[
v_{pki} = \frac{\omega_i}{k} = \frac{\sqrt{x_i}}{k\Delta}
\]

(B15)

and group velocities are

\[
v_{gri} = \frac{d\omega_i}{dk} = \frac{1}{2\sqrt{x_i\Delta}} \frac{dx_i}{dk} = \frac{1}{2\sqrt{x_i}} \frac{dx_i}{dk}.
\]

(B16)

### B1 Numerical integration

To approximate the integral in B.2 we use a 2-D Simpson’s rule described in Mathews & Fink (2004) as

\[
\int_x f(x, y) \, dy = \int_x^b \int_c^d f(x, y) \, dy \, dx \approx S2D(f, h, k),
\]

where

\[
\frac{9}{h^2} S2D = f(a, c) + f(a, d) + f(b, c) + f(b, d)
\]

\[
+ 4 \sum_{j=1}^{n} f(a, y_{2j-1}) + 2 \sum_{j=1}^{n-1} f(a, y_{2j}) + 4 \sum_{j=1}^{n} f(b, y_{2j-1})
\]

\[
+ 2 \sum_{i=1}^{m} f(b, y_{2i}) + 4 \sum_{i=1}^{m} f(x_{2i-1}, c) + 2 \sum_{i=1}^{m} f(x_{2i}, c)
\]

\[
+ 4 \sum_{i=1}^{m} f(x_{2i-1}, d) + 2 \sum_{i=1}^{m} f(x_{2i}, d)
\]

\[
+ 16 \sum_{j=1}^{n} \left( \sum_{i=1}^{m} f(x_{2i-1}, y_{2j-1}) \right) + 8 \sum_{j=1}^{n-1} \left( \sum_{i=1}^{m} f(x_{2i-1}, y_{2j}) \right)
\]

\[
+ 8 \sum_{j=1}^{n} \left( \sum_{i=1}^{m-1} f(x_{2i}, y_{2j-1}) \right) + 4 \sum_{j=1}^{n-1} \left( \sum_{i=1}^{m-1} f(x_{2i}, y_{2j}) \right)
\]

and the rectangle, \( R = \{ (x, y) : a \leq x \leq b, c \leq y \leq d \} \) where \( x \) is divided into \( 2m \) subintervals by \( h = \frac{b-a}{m}, i = 0, 1, \ldots, 2m \) and \( y \) is divided into \( 2n \) subintervals by \( k = \frac{d-c}{n}, j = 0, 1, \ldots, 2n \). We choose a Simpson’s rule because the error can be shown to be of \( O(h^4) + O(k^4) \). Other schemes, such as Gauss family integration rules, can work just as well provided the integral is ‘smooth’ over the region. In our case, Simpson’s rule has the advantage of ease of implementation.

---

Note: Since formulas differ from those prescribed by Press et al. (1992) by a minus sign and the \( \pi \) factor in the argument of the cosine functions for \( x_2 \) and \( x_3 \). We discovered this error only after a significant amount of aspirin and hair loss.
Figure B1. Optimal weighting coefficients for given $p_y$ from $-0.106954$ to $0.106954$ and Poisson’s ratio from $0.005$ to $0.4$.

Figure B2. Examples of dispersive effects on phase and group velocities using the optimal coefficients in Fig. B1 for $\sigma = 0.27$ and a propagation angle of $22.5^\circ$. The $p_y$ values range from $-0.106954$ to $0.106954$, and the grid values from $0$ to $0.25$.

B2 Results

The optimal $a$ and $b$ coefficients determined from minimizing (B2) depend upon three variables: $p_y$, the local $P$ wave speed ($\alpha$), and Poisson’s ratio ($\sigma$). An exploration through typical ranges of these variables reveals a weak dependence of $a$ and $b$ on $\alpha$. An example of the dependence of $a$ and $b$ on $p_y$ and $\sigma$ is shown in Fig. B1 for $\alpha = 5.0$ km s$^{-1}$. The $p_y$ range of $-0.106954$ to $0.106954$ corresponds approximately to an incident $P$ wave with a variation in angle of incidence of about $0^\circ$ to $40^\circ$ and azimuths between $-30^\circ$ and $30^\circ$, which is a representative range of directions for this application. These plots show that that the optimal values of $a$ and $b$ are much more dependent upon $\sigma$ than on $p_y$. The effect of different choices of $\alpha$ is a slight a compression or stretching of this surface along the $p_y$ axis. To a good approximation then, only $\sigma$ and $p_y$ need to be specified to recover the optimal values at any gridpoint. An exception is ‘fluid-like’ media with small shear moduli, as the minimization procedure becomes unstable when Poisson’s ratio is greater than about $0.40$. As discussed in Stekl & Pratt (1998), in the fluid case one should use $a = 0$ (rotated system only). Functional forms of the error curves (e.g. Fig. B2) for a selection of propagation directions show that for nearly all cases the numerical errors are less that a few percent for $1/G < 0.25$ or 4 gridpoints per wavelength. Somewhat surprisingly, the largest errors were found for horizontally propagating shear waves at large values of $p_y$ implying that numerical anisotropy may be significant in this particular case. Horizontally propagating shear waves are not expected to play a significant...
role in teleseismic tomography, but should be considered if this technique is combined with surface waves. Even in this event, the typically longer wavelengths used in surface wave analysis should keep the total errors below a few per cent.

**B3 Reduction to acoustic and 2-D elastic cases**

The 2.5-D forms shown above can be reduced to 2-D by setting $\rho_y = 0$ in (B11). In this case we arrive at a factored problem of the form

$$
\left( \frac{\omega^2 \Delta^2}{\alpha^2} M_1 + K_{22} \right) \left[ \left( \frac{\omega^2 \Delta^2}{\alpha^2} M_1 + K_{14} \right) \left( \frac{\omega^2 \Delta^2}{\alpha^2} M_1 + K_{33} \right) - K_{23} \right] = 0
$$

and we find that $M_1$ and $K_{22}$ are equivalent to the acoustic forms given by Jo et al. (1996) and correspond to the $SH$ phase in the elastic case. $K_{11}$, $K_{22}$ and $K_{31}$ are equivalent to the forms for $P$–$SV$ in the 2-D elastic case given by Šteklo & Pratt (1998). If we then solve (B17) where $x = \frac{\omega \Delta}{\alpha}$, and match up with the values calculated by the cubic equation, we can identify the roots as

$$
x_1 = \frac{-K_{11} + K_{33}}{M_1} + \sqrt{\frac{(K_{11} + K_{33})^2}{M_1^2} + \frac{K_{11} K_{33} - K_{13}^2}{M_1}}
$$

$$
x_2 = \frac{-K_{11} + K_{33}}{M_1} - \sqrt{\frac{(K_{11} + K_{33})^2}{M_1^2} + \frac{K_{11} K_{33} - K_{13}^2}{M_1}}
$$

$$
x_3 = -\frac{K_{22}}{M_1}
$$

where $x_1$ and $x_2$ are the $P$ and $SV$ coupled 2-D elastic case from the quadratic, and $x_3$ is the $SH$ acoustic case.

**APPENDIX C: CALCULATING THE JACOBIAN FORCES**

In eq. (26), we defined a set of forces $F = f^{(i)}$ that generate the Jacobian

$$
J = \left[ \frac{\partial u}{\partial m_1} \frac{\partial u}{\partial m_2} \ldots \frac{\partial u}{\partial m_n} \right] = S^{-1}[f^{(1)} f^{(2)} \ldots f^{(n)}],
$$

where

$$
f^{(i)} = -\frac{\partial S}{\partial m_i} u.
$$

The displacement field $u$ is determined using eq. (11). Only a very small percentage of the matrix $\partial S / \partial m_i$ is non-zero, and to locate these terms we first examine the structure of $S$.

**C1 Structure of the finite difference matrix**

Each point in the medium is associated with three components of displacement $(u, v, w)$, and in general each component is coupled to the other two. Hence the basic element of $S$ is a $3 \times 3$ matrix $A$ that we order as

$$
A = \begin{bmatrix}
A_{uu} & A_{uv} & A_{uw} \\
A_{uv} & A_{vv} & A_{vw} \\
A_{uw} & A_{vw} & A_{ww}
\end{bmatrix},
$$

where $A_{ij}$ contains the finite difference contributions to displacement in direction $i$ from partial derivatives of displacement in direction $j$. Because we use a nine point star to compute differences, the displacements at any point in the medium are described by nine of these $A$ matrices ordered spatially as

$$
A_1 \quad A_4 \quad A_7 \\
A_2 \quad A_5 \quad A_8 \\
A_3 \quad A_6 \quad A_9
$$

4 The computations of Šteklo & Pratt (1998) differ from 2-D terms we derived independently as their finite difference forms of the cross derivatives in the rotated frame (e.g. $\frac{\partial^2 u}{\partial v \partial w}$) appear to be incorrect.
and arranged to form three rows in the \( S \) matrix as
\[
\begin{bmatrix}
A_1 & A_2 & A_3 & \cdots & A_4 & A_5 & A_6 & \cdots & A_7 & A_8 & A_9 \\
\end{bmatrix}.
\]

For convenience we will refer to this set of three rows as a 'meta-row'.

\section*{C2 Computing derivatives of \( S \)}

As discussed in Appendix A, medium properties are averaged between points in the nine point star. Hence terms involving \( m_i \) will be found in each of the \( A \) matrices in the meta-row corresponding to that particular element. In addition, averaging creates terms involving \( m_i \) that appear in the meta-rows associated with adjacent points in the medium, so that in general there will be eight additional meta-rows with non-zero \( \partial S/\partial m_i \) elements. The locations of these elements can be summarized by first defining a blockette of \( A \) matrices in \( S \) as
\[
\begin{bmatrix}
A_{n-1} & A_n & A_{n+1} & 0 & 0 \\
0 & A_{n-1} & A_n & A_{n+1} & 0 \\
0 & 0 & A_{n-1} & A_n & A_{n+1}
\end{bmatrix},
\]

where the subscript on \( A \) above refers to the position of that matrix within the nine point star associated with a particular meta-row (i.e. \( A_n \) in a given meta-row is not necessarily equal to \( A_n \) in any other meta-row). That part of \( S \) involving \( m_i \) consists of seven of these blockettes arranged as
\[
\begin{bmatrix}
B_5 & \cdots & B_8 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
B_2 & \cdots & B_5 & \cdots & B_8 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & B_2 & \cdots & B_5 \\
\end{bmatrix}.
\]

Typically we take the \( m_i \) to be the wave speeds \( (\alpha_i \text{ and } \beta_i) \) specified at the gridpoints. Let \( m_j \) be a wave speed at point \( j \) of the 9 point star and define \( A_{n,j} \) as the 3 \times 3 matrix of derivatives of \( A_n \) with respect to \( m_j \), that is,
\[
A_{n,j} = \frac{\partial A_n}{\partial m_j}.
\] (C4)

Then for a given \( m_i \), the non-zero elements of the banded \( \partial S/\partial m_i \) matrix in eq. (C3) can be written in a compact form as
\[
\frac{\partial S}{\partial m_i} = \begin{bmatrix}
A_{5,9} & A_{6,9} & 0 & A_{8,9} & A_{9,9} & 0 & 0 & 0 & 0 \\
0 & A_{5,8} & 0 & A_{7,8} & A_{8,8} & A_{9,8} & 0 & 0 & 0 \\
0 & A_{4,7} & A_{5,7} & 0 & A_{7,7} & A_{8,7} & 0 & 0 & 0 \\
0 & A_{3,6} & 0 & A_{5,6} & A_{6,6} & 0 & 0 & A_{9,6} & 0 \\
A_{1,5} & A_{2,5} & A_{3,5} & A_{4,5} & A_{5,5} & A_{6,5} & A_{7,5} & A_{8,5} & A_{9,5} \\
0 & A_{1,4} & 0 & 0 & A_{4,4} & A_{5,4} & 0 & A_{7,4} & 0 \\
0 & 0 & 0 & A_{2,3} & A_{3,3} & 0 & A_{5,3} & A_{6,3} & 0 \\
0 & 0 & 0 & A_{1,2} & A_{2,2} & A_{3,2} & 0 & A_{5,2} & 0 \\
0 & 0 & 0 & 0 & A_{1,1} & A_{2,1} & 0 & A_{4,1} & A_{5,1}
\end{bmatrix},
\]

where we recall that \( A_{5,9} \) references \( m_i \) when it is located at position 9 in the nine point star, \( A_{5,8} \) when \( m_i \) is at position 8, and so on.
Let $D_k$ represent the $3 \times 1$ subvector $(u_k, v_k, w_k)$ of $u$ that is multiplied by the $k$ th meta-row of $\partial S/\partial m_i$. Then from eq. (C2)

$$f^{(i)} = - \sum_{k=1}^{9} A_{k,i} D_i + A_{1,4} D_4 + A_{2,3} D_3 + A_{1,1} D_1 \ .$$

(C5)

The 41 $A_{ni}$ matrices in (C5) require that we specify 369 medium dependent factors to compute the 27 elements of $f^{(i)}$ for each $m_i$. While somewhat more onerous than the acoustic case, where only one such factor is needed, the derivatives in (C4) can be expressed analytically and nearly half of them are zero, so the extra computational effort is not overly burdensome.