Seismic tomography, adjoint methods, time reversal, and banana-doughnut kernels


Jeroen Tromp, Carl Tape, and Qinya Liu

Seismological Laboratory, California Institute of Technology, Pasadena, California 91125

**Contents**

Summary 3

1 Introduction 4

2 Waveform Tomography 5
   2.1 Topography on internal discontinuities 9

3 Adjoint equations 11

4 Traveltime Tomography 13
   4.1 Banana-doughnut kernels 13
   4.2 Misfit kernels 15
   4.3 Differential traveltime tomography 16

5 Amplitude Tomography 16
   5.1 Attenuation 18

6 Generalizations 19
   6.1 Banana-doughnut kernels 19
   6.2 Misfit kernels 21

7 Kernel Gallery 21
   7.1 Model setup 22
   7.2 Banana-doughnut kernels 22
      7.2.1 SH waves 22
      7.2.2 P-SV waves 23
   7.3 Amplitude kernels 24
   7.4 Q kernels 24

8 Source Inversions 24
   8.1 Example 25

9 Joint Inversions 26

10 Conclusions 27

Acknowledgments 28

References 28
List of Figures

1. Sketch of 2D model set-up and ray paths ........................................... 31
2. SH seismograms .................................................................................. 31
3. SHs sequence of s-s† interaction. ......................................................... 32
4. Six SH kernels for reversing S ............................................................... 33
5. Reversing different time windows of the SH wavefield ......................... 34
6. Frequency dependence of sensitivity kernels ......................................... 35
7. P-SV seismograms .................................................................................. 36
8. P-SVPS+SP sequence of s-s† interaction. .............................................. 37
9. Six P-SV kernels for reversing PS+SP .................................................... 38
10. Reversing different time windows of P-SV .......................................... 39
11. Traveltime, amplitude, and attenuation kernels ..................................... 40
12. Source inversion with conjugate gradient method ............................... 41

Table 1: Adjoint fields discussed in the text, with corresponding equations. MF and BD denote “misfit kernels” and “banana-doughnut kernels”, respectively. Banana-doughnut kernels do not incorporate data in their computation and are limited to a single receiver. Misfit kernels incorporate data and may include multiple receivers.

<table>
<thead>
<tr>
<th>adjoint field</th>
<th>source variation of kernels</th>
<th>kernels</th>
</tr>
</thead>
<tbody>
<tr>
<td>s† or s†</td>
<td>f† or f†</td>
<td>ΔF K† or K†</td>
</tr>
<tr>
<td>waveform</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>traveltime</td>
<td>44</td>
<td>45</td>
</tr>
<tr>
<td>combined travelt ime</td>
<td>56</td>
<td>57</td>
</tr>
<tr>
<td>combined differential travelt ime</td>
<td>60</td>
<td>61</td>
</tr>
<tr>
<td>amplitude</td>
<td>66</td>
<td>67</td>
</tr>
<tr>
<td>combined amplitude</td>
<td>69</td>
<td>71</td>
</tr>
<tr>
<td>GSDF</td>
<td>83</td>
<td>84</td>
</tr>
<tr>
<td>combined GSDF</td>
<td>89</td>
<td>90</td>
</tr>
</tbody>
</table>
Summary

We draw connections between seismic tomography, adjoint methods popular in climate and ocean dynamics, time-reversal imaging, and finite-frequency ‘banana-doughnut’ kernels. We demonstrate that Fréchet derivatives for tomographic and (finite) source inversions may be obtained based upon just two numerical simulations for each earthquake: one calculation for the current model and a second, ‘adjoint’, calculation that uses time-reversed signals at the receivers as simultaneous, fictitious sources. For a given model \( m \), we consider objective functions \( F(m) \) that minimize differences between waveforms, traveltimes, or amplitudes. For tomographic inversions we show that the Fréchet derivatives of such objective functions may be written in the generic form \( \delta F = \int_V K_m(x) \delta \ln m(x) \, d^3x \), where \( \delta \ln m = \delta m / m \) denotes the relative model perturbation. The volumetric kernel \( K_m \) is defined throughout the model volume \( V \) and is determined by time-integrated products between spatial and temporal derivatives of the regular displacement field \( s \) and the adjoint displacement field \( s^\dagger \); the latter is obtained by using time-reversed signals at the receivers as simultaneous sources. In waveform tomography the time-reversed signal consists of differences between the data and the synthetics, in traveltime tomography it is determined by synthetic velocities, and in amplitude tomography it is controlled by synthetic displacements. For each event, the construction of the kernel \( K_m \) requires one forward calculation for the regular field \( s \) and one adjoint calculation involving the fields \( s \) and \( s^\dagger \). For multiple events the kernels are simply summed. The final summed kernel is controlled by the distribution of events and stations. Fréchet derivatives of the objective function with respect to topographic variations \( \delta h \) on internal discontinuities may be expressed in terms of two-dimensional kernels \( K_h \) and \( K_h \) in the form \( \delta F = \int_\Sigma K_h(x) \delta h(x) \, d^2x + \int_{\Sigma_{FS}} K_h(x) \cdot \nabla^\Sigma \delta h(x) \, d^2x \), where \( \Sigma \) denotes a solid-solid or fluid-solid boundary and \( \Sigma_{FS} \) a fluid-solid boundary, and \( \nabla^\Sigma \) denotes the surface gradient. In the case of traveltime tomography, the kernels \( K_m \) are weighted combinations of banana-doughnut kernels. We illustrate how amplitude anomalies may be inverted for lateral variations in elastic and anelastic structure. In the context of a finite-source inversion, the model vector consists of the time-dependent elements of the moment-density tensor \( \mathbf{m}(x, t) \). We demonstrate that the Fréchet derivatives of the objective function \( F \) may in this case be written in the form \( \delta F = \int_0^T \int_\Sigma \epsilon^\dagger(x, T - t) : \delta \mathbf{m}(x, t) \, d^2x \, dt \), where \( \epsilon^\dagger \) denotes the adjoint strain tensor on the finite-fault plane \( \Sigma \). In the case of a point source this result reduces further to the calculation of the time-dependent adjoint strain tensor \( \epsilon^\dagger \) at the location of the point source \( x_s \), an approach reminiscent of an acoustic time-reversal mirror. The theory is illustrated for both tomographic and source inversions using a two-dimensional spectral-element method.
1 Introduction

Recent advances in high-performance computing and numerical techniques have facilitated fully three-dimensional (3D) simulations of global and regional seismic wave propagation at unprecedented resolution and accuracy (Komatitsch et al., 2002; Tsuboi et al., 2003; Capdeville et al., 2003; Komatitsch et al., 2004). The challenge now lies in harnessing these new found capabilities to enhance the quality of tomographic images of the Earth’s interior, in conjunction with improving models of the rupture process during an earthquake. On the face of it, this seems like a herculean task because hundreds or even thousands of model parameters are involved in such inversions. In principle, the Fréchet derivatives that represent the sensitivity of a seismogram with respect to the model parameters may be calculated numerically, but this would require a number of forward calculations equal to the number of model parameters.

Tarantola (1984, 1987, 1988) demonstrates that the seismic inverse problem may be solved iteratively by numerically calculating the Fréchet derivatives of a waveform misfit function. The construction of these derivatives involves the interaction between the wavefield for the current model and a wavefield obtained by using the time-reversed differences between the data and the current synthetics as simultaneous sources. Only two numerical simulations are required to calculate the necessary Fréchet derivatives: one for the current model and a second for the time-reversed differences between the data and the synthetics for this model. The acoustic theory developed by Tarantola (1984) was implemented and illustrated numerically by Gauthier et al. (1986). The theory was extended to the (an)elastic wave equation by Tarantola (1987, 1988), and implemented by Crase et al. (1990). Other applications of the theory may be found in Mora (1987, 1988), Pratt (1999), and Akçelik et al. (2002, 2003).

The theory developed by Tarantola (1984) predates an influential paper by Talagrand and Courtier (1987), in which the concept of an ‘adjoint’ calculation is introduced as a means of explicitly determining the gradient of a misfit function. In fact, propagating waveform residuals backward in time is an example of an adjoint calculation. Tarantola shows how the gradient of a misfit function defined in terms of a squared residual integrated over time may be rewritten as a spatial integral involving a 3D kernel and a model perturbation, precisely as articulated by Talagrand and Courtier (1987) for the more general adjoint problem. Adjoint calculations can be quite complicated, because they involve interactions between the regular and adjoint fields, see e.g., Bunge et al. (2003) for an example in geodynamics, but for the seismic inverse problem they are relatively straightforward (Akçelik et al., 2002, 2003). In this paper we will demonstrate that the construction of the 3D kernels that characterize the gradient of a misfit function involves interactions between the regular and adjoint wavefields.

The concept of ‘time-reversal mirrors’, in which an acoustic signal is recorded by an array of transducers, time-reversed, and retransmitted, has been made popular by Fink (Fink et al., 1989; Fink, 1992, 1997). Time-reversal acoustics is actively used in medical therapy and nondestructive testing. As was shown by Tarantola (1984, 1987, 1988), and we will see again in the present paper, time-reversal is an integral part of the seismic inverse problem. One effectively uses the Green’s function of the current model to locate the origin of the discrepancies between the data and the
current synthetics. Using these time-reversed residuals as sources results in the illumination of the parts of the model that are inadequate, much like a time-reversal mirror refocuses on the source of an acoustic signal.

The use of finite-frequency kernels for traveltime or amplitude inversions, rather than ray theory, is advocated by Marquering et al. (1999), Zhao et al. (2000), Dahlen et al. (2000), Hung et al. (2000), and Dahlen and Baig (2002). These kernels recognize the fact that traveltimes and amplitude anomalies are frequency dependent, and incorporate effects associated with wavefront healing. Simple 3D traveltime kernels for phases like P and S are shaped like bananas with a doughnut-like cross-section, and thus the kernels are whimsically referred to as ‘banana-doughnut’ kernels. Such kernels were recently implemented for compressional-wave tomography by Montelli et al. (2004). In this paper we show that the Fréchet derivatives of traveltime or amplitude misfit functions consist of weighted sums of 3D banana-doughnut kernels. The weights in these sums are determined by the traveltime or amplitude misfit at a particular station for a particular event.

The purpose of this paper is to bring waveform tomography, adjoint methods, time-reversal imaging, and banana-doughnut theory under the same (theoretical) umbrella. The characteristics of the associated kernels are illustrated for a variety of waveforms based upon a two-dimensional (2D) spectral-element method.

2 Waveform Tomography

Motivated by the work of Tarantola (1984, 1987, 1988), we begin by considering waveform tomography. We seek to minimize the differences between three-component waveform data \( \mathbf{d}(x_r, t) \) recorded at \( N \) stations \( x_r, r = 1, \ldots, N \), and the corresponding synthetics \( \mathbf{s}(x_r, t, m) \) for a given model vector \( m \). As a measure of the goodness of fit, we introduce the least-squares waveform misfit function (Nolet, 1987)

\[
F(m) = \frac{1}{2} \sum_{r=1}^{N} \int_0^T \| \mathbf{s}(x_r, t, m) - \mathbf{d}(x_r, t) \|^2 dt.
\]

(1)

In practice, both the data \( \mathbf{d} \) and the synthetics \( \mathbf{s} \) will be windowed and filtered on the time interval \([0, T]\). This may involve windowing and bandpassing multiple body- and surface-wave arrivals. In what follows we will implicitly assume that such filtering operations have been performed, i.e., the symbols \( \mathbf{d} \) and \( \mathbf{s} \) will denote processed data and synthetics, respectively. The differences between the data and synthetics may also be weighted based upon a data covariance matrix, but for the sake of brevity and simplicity we will not explicitly incorporate this weighting. The \( M \)-dimensional vector \( m \) characterizes the current model. An iterative inversion requires the calculation of the Fréchet derivatives

\[
\delta F = \sum_{r=1}^{N} \int_0^T [\mathbf{s}(x_r, t, m) - \mathbf{d}(x_r, t)] \cdot \delta \mathbf{s}(x_r, t, m) dt,
\]

(2)
where $\delta s$ denotes the perturbation in the displacement field $s$ due to a model perturbation $\delta m$. In seismic tomography the model vector $m$ is long, and therefore a brute-force calculation of the derivatives (2) is impractical and generally not feasible with limited computational resources. Note that for multiple earthquakes the Fréchet derivatives are simply summed. To avoid clutter we omit the explicit dependence of the displacement field $s$ on the model $m$ from here on.

The misfit function (1) is of course just one possible measure of goodness of fit. In subsequent sections we will consider a number of other choices that may in practice lead to better convergence. Similarly, norms other than the $L^2$ norm may be readily introduced (Crase et al., 1990).

In seismic tomography, Fréchet derivatives may be determined based upon the Born approximation (Hudson, 1977; Wu and Aki, 1985). Using index notation and the Einstein summation convention (unless stated otherwise), suppose we have a generic anisotropic background model \{\rho, c_{jklm}\}, where $\rho$ denotes the density and $c_{jklm}$ an element of the fourth-order elastic tensor, upon which we superimpose perturbations \{\delta \rho, \delta c_{jklm}\}. Then the associated perturbed displacement $\delta s$ field may be expressed as

$$
\delta s_i(x, t) = - \int_0^t \int_V \left[ \delta \rho(x') G_{ij}(x, x'; t - t') \partial_r^2 s_j(x', t') + \delta c_{jklm}(x') \partial_k^t G_{ij}(x, x'; t - t') \partial_l^t s_m(x', t') \right] d^3 x' \ dt',
$$

where $V$ denotes the model volume. Upon substitution of (3) into (2) we obtain

$$
\delta F = - \sum_{r=1}^N \int_0^T \left[ s_i(x_r, t, m) - d_i(x_r, t) \right] \int_0^t \int_V \left[ \delta \rho(x') G_{ij}(x_r, x'; t - t') \partial_r^2 s_j(x', t') + \delta c_{jklm}(x') \partial_k^t G_{ij}(x_r, x'; t - t') \partial_l^t s_m(x', t') \right] d^3 x' \ dt' \ dt.
$$

Let us define the field

$$
\Phi_k(x', t') = \sum_{r=1}^N \int_{t'}^T G_{ik}(x_r, x'; t - t') \left[ s_i(x_r, t) - d_i(x_r, t) \right] dt.
$$

Taking advantage of the reciprocity of the Green’s tensor (Aki and Richards, 1980; Dahlen and Tromp, 1998),

$$
G_{ik}(x_r, x'; t - t') = G_{ki}(x_r', x_r; t - t'),
$$

we may rewrite (5) as

$$
\Phi_k(x', t') = \sum_{r=1}^N \int_{t'}^T G_{ki}(x_r', x_r; t - t') \left[ s_i(x_r, t) - d_i(x_r, t) \right] dt.
$$

Upon reversing time by making the substitution $t \rightarrow T - t$ in (7) we obtain

$$
\Phi_k(x', t') = \sum_{r=1}^N \int_0^{T-t'} G_{ki}(x_r', x_r; T - t' - t) \left[ s_i(x_r, T - t) - d_i(x_r, T - t) \right] dt.
$$
Next, we define what we shall refer to as the waveform adjoint source

$$f_i^\dagger(x, t) = \sum_{r=1}^{N} [s_i(x_r, T - t) - d_i(x_r, T - t)]\delta(x - x_r). \quad (9)$$

Note that this distributed source involves using the time-reversed differences between the data and the current synthetics as $N$ simultaneous point sources located at the stations. With this definition we may rewrite (8) as

$$\Phi_k(x', t') = \int_0^{T-t'} \int_V G_{ki}(x', x; T - t - t')f_i^\dagger(x, t) d^3x dt. \quad (10)$$

Using the standard Green’s function approach, we introduce the waveform adjoint field $s_i^\dagger$ generated by the waveform adjoint source (9):

$$s_k^\dagger(x', t') = \int_{t'}^t \int_V G_{ki}(x', x; t' - t)f_i^\dagger(x, t) d^3x dt. \quad (11)$$

The relationship between (10) and (11) is then

$$\Phi_k(x', t') = s_k^\dagger(x', T - t'). \quad (12)$$

In seismology, the concept of an adjoint field was first introduced by Tarantola (1984) for the acoustic wave equation, and by Tarantola (1987, 1988) for the (an)elastic wave equation. It is analogous to the approach taken in time-reversal imaging (Fink et al., 1989; Fink, 1992, 1997), where one retransmits a time-reversed acoustic signal to locate its origin. In the seismic case, the waveform adjoint field $s^\dagger$ will illuminate the source of the discrepancy between the data, $d(x_r, t)$, and the synthetics for the current model, $s(x_r, t)$. Alternatively, from the perspective of adjoint methods (Talagrand and Courtier, 1987), the corresponding adjoint equations of motion are discussed in the next section.

With the introduction of the adjoint field (11), we can rewrite (4) in the form

$$\delta F = \int_V \left[ K_\rho(x) \delta \ln \rho(x) + K_{c_{jklm}}(x) \delta \ln c_{jklm}(x) \right] d^3x, \quad (13)$$

where $\delta \ln \rho = \delta \rho / \rho$ and $\delta \ln c_{jklm} = \delta c_{jklm} / c_{jklm}$ (no summation) denote relative model perturbations. Notice how the gradient of the misfit function is transformed from an integral over time in (2) to an integral over the model volume $V$ in (13), precisely as demonstrated more generally in the context of adjoint methods by Talagrand and Courtier (1987). The 3D waveform misfit kernels $K_\rho$ and $K_{c_{jklm}}$ represent Fréchet derivatives with respect to density and the elastic parameters, respectively, and are defined by

$$K_\rho(x) = -\int_0^T \rho(x)s^\dagger(x, T - t) \cdot \partial_t^2 s(x, t) dt, \quad (14)$$
\begin{equation}
K_{jklm}(x) = -\int_0^T \varepsilon^\dagger_{jk}(x, T-t)c_{jklm}(x)\epsilon_{lm}(x, t) dt, \quad (\text{no summation})
\end{equation}

where $\epsilon_{lm}$ and $\varepsilon^\dagger_{jk}$ denote elements of the strain and the waveform adjoint strain tensors, respectively.

For an isotropic material we have

\begin{equation}
c_{jklm} = (\kappa - \frac{2}{3}\mu)\delta_{jk}\delta_{lm} + \mu(\delta_{jl}\delta_{km} + \delta_{jm}\delta_{kl}),
\end{equation}

and thus (13) becomes\(^1\)

\begin{equation}
\delta F = \int_V \left[ K_{\rho(\kappa\mu)}(x) \delta \ln \rho(x) + K_{\mu(\kappa\rho)}(x) \delta \ln \mu(x) + K_{\kappa(\mu\rho)}(x) \delta \ln \kappa(x) \right] d^3x.
\end{equation}

The isotropic misfit kernels $K_{\mu(\kappa\rho)}$ and $K_{\kappa(\mu\rho)}$ represent Fréchet derivatives with respect to the bulk and shear moduli $\kappa$ and $\mu$, respectively, and are given by

\begin{equation}
K_{\mu(\kappa\rho)}(x) = -\int_0^T 2\mu(x)D^\dagger(x, T-t) : D(x, t) dt,
\end{equation}

\begin{equation}
K_{\kappa(\mu\rho)}(x) = -\int_0^T \kappa(x)[\nabla \cdot s^\dagger(x, T-t)][\nabla \cdot s(x, t)] dt,
\end{equation}

where $D$ and $D^\dagger$ denote the traceless strain deviator and its waveform adjoint, respectively.

Alternatively, we may express the Fréchet derivatives (16) in terms of variations in density $\rho$, shear-wave speed $\beta$, and compressional-wave speed $\alpha$ as

\begin{equation}
\delta F = \int_V \left[ K_{\rho(\alpha\beta)}(x) \delta \ln \rho(x) + K_{\beta(\alpha\rho)}(x) \delta \ln \beta(x) + K_{\alpha(\beta\rho)}(x) \delta \ln \alpha(x) \right] d^3x,
\end{equation}

where the misfit kernels $K_{\rho(\alpha\beta)}$, $K_{\beta(\alpha\rho)}$, and $K_{\alpha(\beta\rho)}$ represent Fréchet derivatives with respect to density, shear-, and compressional-wave speed, respectively; they are given in terms of the kernels (14), (17) and (18) by

\begin{align*}
K_{\rho(\alpha\beta)} &= K_{\rho(\kappa\mu)} + K_{\kappa(\mu\rho)} + K_{\mu(\kappa\rho)}, \\
K_{\beta(\alpha\rho)} &= 2 \left( K_{\mu(\kappa\rho)} - \frac{4}{3\kappa}K_{\kappa(\mu\rho)} \right), \\
K_{\alpha(\beta\rho)} &= 2 \left( 1 + \frac{4\mu}{3\kappa} \right) K_{\kappa(\mu\rho)}.
\end{align*}

All the kernels presented in this section are symmetric with regards to the interchange $s(x, t) \leftrightarrow s^\dagger(x, T-t)$ (assuming $\partial_0 s(x, 0) = 0$ and $s^\dagger(x, 0) = 0$).

Note in (13), (16), and (19) that the kernels are constructed based upon two calculations: one for the displacement field $s$ and a second for the waveform adjoint field $s^\dagger$. The only interactions between $s$ and its adjoint $s^\dagger$ occur through the time-reversed source (9). The Fréchet derivatives of the penalty function $F$ with respect to any model coefficient $m$ may be obtained by performing

\(^1\)The two variables in parentheses in the subscript are the quantities that remain fixed. For example, the kernel $K_{\alpha(b \epsilon)}$ is the sensitivity kernel for variable $\alpha$ with variables $b$ and $c$ fixed.
the volume integrations in (13), (16), or (19) involving the kernels. The choice of model basis functions has not been specified. In fact, this choice should be guided by the spatial characteristics of the kernels: where the kernels are large one obtains the largest Fréchet derivatives. The results (13), (16), and (19), in combination with a standard conjugate-gradient algorithm (Fletcher and Reeves, 1964; Mora, 1987, 1988), may be used to determine the (local) minimum of the objective function (1).

The computational challenge in the construction of the kernels lies in the fact that we need to have simultaneous access to the regular field \( s \) and the waveform adjoint field \( s^\dagger \). Therefore, either the regular field \( s \) needs to be stored as a function of space and time so that it can be read back from the hard disk during the calculation of the waveform adjoint field \( s^\dagger \) and the construction of the kernels, or it needs to be reconstructed on the fly, using the final displacement field \( s(x, T) \) as a starting point for and integration backward in time (Gauthier et al., 1986). This backward integration of the regular field \( s \) involves undoing the effects of attenuation, a process that is numerically stable (Tarantola, 1988). For a model with absorbing boundaries one also needs to store the regular field on the boundaries as a function of time in order to reconstruct the entire field \( s(x, t) \) backward in time (Gauthier et al., 1986).

### 2.1 Topography on internal discontinuities

In the context of global tomography one may wish to consider the effects of topography on the Moho, upper mantle discontinuities, the core-mantle boundary, or the inner-core boundary on the misfit function. Similarly, in regional tomography one may be interested in the effects of topographic variations of the basement or the Moho. Let \( \delta h \) denote topographic perturbations in the direction of the unit outward normal \( \hat{n} \) on solid-solid discontinuities \( \Sigma_{SS} \) or fluid-solid discontinuities \( \Sigma_{FS} \). It is shown by Dahlen (2005) that the perturbed displacement field \( \delta s \) due to topographic perturbations \( \delta h \) may be written in the form

\[
\delta s_i(x, t) = \int_0^t \int_\Sigma \left[ \rho(x') G_{ij}(x, x'; t - t') \partial_n^2 s_j(x', t') + \partial'_k G_{ij}(x, x'; t - t') c_{jklm}(x') \partial'_l s_m(x', t') \right.
- \hat{n}_k(x') \partial'_k G_{ij}(x, x'; t - t') c_{jklm}(x') \partial'_l s_m(x', t') \\
- \hat{n}_k(x') c_{jklm}(x') \partial'_l G_{im}(x, x'; t - t') \partial_n s_j(x', t') \left] \delta h(x') \right]_{-}^{+} d^2 x' dt'
+ \int_0^t \int_{\Sigma_{FS}} [ G_{ik}(x, x'; t - t') \hat{n}_j(x') \hat{n}_p(x') c_{jklm}(x') \partial'_l s_m(x', t') \\
+ s_k(x', t') \hat{n}_j(x') \hat{n}_p(x') c_{jklm}(x') \partial'_l G_{im}(x, x'; t - t') ] \hat{n}_l \delta h(x') \right]_{-}^{+} \nabla^2 \delta h(x') \right]_{-}^{+} d^2 x' dt',
\]

where \( \Sigma = \Sigma_{SS} + \Sigma_{FS} \) denotes all solid-solid and fluid-solid discontinuities. The surface gradient is defined by \( \nabla^\Sigma = (I - \hat{n} \hat{n}) \cdot \nabla \), and the normal derivative \( \partial_n \) is given by \( \partial_n = \hat{n} \cdot \nabla \). The notation \( \left[ \right]_{-}^{+} \) denotes the jump in the enclosed quantity when going from the outward (+) side of the discontinuity to the inward (−) side of the discontinuity. Upon substitution of (21) into (2), the gradient of the misfit function due to topographic perturbations \( \delta h \) may be written in
the form

\[ \delta F = \int \sum K_h(x) \delta h(x) \, d^2x + \int \sum_{\Sigma_{\text{fs}}} K_h(x) \cdot \nabla^F \delta h(x) \, d^2x, \quad (22) \]

where

\[ K_h(x) = \int_0^T \left[ \rho(x)s^{\dagger}(x, T - t) \cdot \partial_t^2 s(x, t) + \epsilon^{\dagger}(x, T - t) : c(x) : \epsilon(x, t) \\
- \hat{n}(x) \partial_n s^{\dagger}(x, T - t) : c(x) : \epsilon(x, t) - \hat{n}(x) \partial_n s(x, t) : c(x) : \epsilon^{\dagger}(x, T - t) \right]^+ dt, \quad (23) \]

\[ K_h(x) = \int_0^T \left[ s^{\dagger}(x, T - t) \hat{n}(x) \hat{n}(x) : c(x) : \epsilon(x, t) + s(x, t) \hat{n}(x) \hat{n}(x) : c(x) : \epsilon^{\dagger}(x, T - t) \right]^+ dt. \quad (24) \]

In an isotropic Earth model (23) and (24) reduce to

\[ K_h(x) = \int_0^T \left[ \rho(x)s^{\dagger}(x, T - t) \cdot \partial_t^2 s(x, t) + \kappa(x) \nabla \cdot s^{\dagger}(x, T - t) \nabla \cdot s(x, t) + 2\mu(x)D^{\dagger}(x, T - t) : D(x, t) \\
- \kappa(x) \hat{n}(x) \cdot \partial_n s^{\dagger}(x, T - t) \nabla \cdot s(x, t) - 2\mu(x) \hat{n}(x) \partial_n s(x, t) : D^{\dagger}(x, T - t) \right]^+ dt, \quad (25) \]

\[ K_h(x) = \int_0^T \left[ s^{\dagger}(x, T - t) [\kappa(x) \nabla \cdot s(x, t) + 2\mu(x) \hat{n}(x) \cdot D(x, t) \cdot \hat{n}(x)] \\
+ s(x, t) [\kappa(x) \nabla \cdot s^{\dagger}(x, T - t) + 2\mu(x) \hat{n}(x) \cdot D^{\dagger}(x, T - t) \cdot \hat{n}(x)] \right]^+ dt. \quad (26) \]

The result (22) may be used to invert for lateral variations in topography on an internal discontinuity.

The effects of small topographic perturbations on the Earth’s surface were considered in the Born approximation by Hudson (1977). On the Earth’s free surface the traction \( \hat{n}c_{jk}l_m \epsilon_{lm} \) vanishes, and in this case the kernel (23) reduces to:

\[ K_h(x) = -\int_0^T \left[ \rho(x)s^{\dagger}(x, T - t) \cdot \partial_t^2 s(x, t) + \epsilon_{jk}(x, T - t)c_{jk}l_m(x)\epsilon_{lm}(x, t) \right] dt. \quad (27) \]

In the isotropic case this becomes

\[ K_h(x) = -\int_0^T \left[ \rho(x)s^{\dagger}(x, T - t) \cdot \partial_t^2 s(x, t) + \kappa(x) \nabla \cdot s(x, t) \nabla \cdot s^{\dagger}(x, T - t) + 2\mu(x)D(x, t) : D^{\dagger}(x, T - t) \right] dt. \quad (28) \]

As in the previous section, all the kernels in this section are symmetric with regards to the interchange \( s(x, t) \leftrightarrow s^{\dagger}(x, T - t) \).
3 Adjoint equations

In the context of meteorological inverse problems, Talagrand and Courtier (1987) introduced the concept of an adjoint calculation as a means of explicitly determining the gradient of a misfit function. Such adjoint equations can be quite complicated, see e.g., Bunge et al. (2003) for an example in geodynamics, but we will see that for the seismological inverse problem they are rather straightforward.

For a detailed introduction to the numerical simulation of seismic wave propagation in anelastic materials the reader is referred to Komatitsch and Tromp (1999, 2002a); complications due to self-gravitation and rotation are addressed in Komatitsch and Tromp (2002b). Here we will simply summarize their results to contrast them with the adjoint equations.

The equation of motion that needs to be solved in an anelastic Earth model is

$$\rho \partial_t^2 \mathbf{s} = \nabla \cdot \mathbf{T} + \mathbf{f}. \tag{29}$$

Here $\rho$ denotes the distribution of density and the earthquake is represented by the force $\mathbf{f}$. The symmetric stress tensor $\mathbf{T}$ in an anelastic material may be expressed in terms of the unrelaxed elastic tensor $\mathbf{c}^U$, the displacement gradient $\nabla \mathbf{s}$, and a sum of $L$ memory variable tensors $\mathbf{R}^\ell$, $\ell = 1, \ldots, L$, as

$$\mathbf{T} = \mathbf{c}^U : \nabla \mathbf{s} - \sum_{\ell=1}^L \mathbf{R}^\ell. \tag{30}$$

The symmetric memory variable tensors $\mathbf{R}^\ell$ represent standard linear solids which are chosen to mimic an absorption-band solid. For each standard linear solid one needs to solve a first-order ordinary differential in time:

$$\partial_t \mathbf{R}^\ell = -\mathbf{R}^\ell / \tau_{\sigma \ell} + \delta \mathbf{c}^\ell : \nabla \mathbf{s} / \tau_{\sigma \ell}. \tag{31}$$

The components of the unrelaxed modulus $c_{ijkl}^U$ are given in terms of the relaxed modulus $c_{ijkl}^R$ by

$$c_{ijkl}^U = c_{ijkl}^R \left[ 1 - \sum_{\ell=1}^L (1 - \tau_{ijkl}^\ell / \tau_{\sigma \ell}) \right]. \tag{32}$$

The stress relaxation times $\tau_{\sigma \ell}$ are chosen to be the same for all components of the anelastic tensor, whereas the strain relaxation times $\tau_{ijkl}^\ell$ are unique to each individual component. In (31), the modulus defect $\delta \mathbf{c}^\ell$ associated with each individual standard linear solid is determined by

$$\delta c_{ijkl}^\ell = -c_{ijkl}^R (1 - \tau_{ijkl}^\ell / \tau_{\sigma \ell}). \tag{33}$$

Ignoring possible complications associated with absorbing boundary conditions, equations (29)–
(31) need to be solved subject to the stress-free surface boundary condition \( \mathbf{n} \cdot \mathbf{T} = 0 \), where \( \mathbf{n} \) denotes the unit outward normal to the surface. At solid-solid boundaries both the displacement \( \mathbf{s} \) and the traction \( \mathbf{n} \cdot \mathbf{T} \) need to be continuous, whereas at fluid-solid boundaries the traction \( \mathbf{n} \cdot \mathbf{T} \) and the normal component of displacement \( \mathbf{n} \cdot \mathbf{s} \) are continuous.

To obtain the waveform adjoint field (11), we need to solve a momentum equation in which the source \( \mathbf{f} \) is replaced with the waveform adjoint source (9), i.e., the sum over the time-reversed differences \( \mathbf{f}^\dagger(x,t) = \sum_{r=1}^{N} [s(x_r, T-t) - d(x_r, T-t)] \delta(x - x_r) \). Thus we obtain the following system of adjoint equations:

\[
\rho \partial_t^2 \mathbf{s}^\dagger = \nabla \cdot \mathbf{T}^\dagger + \mathbf{f}^\dagger
\]  
\[
\mathbf{T}^\dagger = c^U : \nabla \mathbf{s}^\dagger - \sum_{\ell=1}^{L} \mathbf{R}^{\ell \dagger},
\]  
\[
\partial_t \mathbf{R}^{\ell \dagger} = -\mathbf{R}^{\ell \dagger} / \tau^\ell + \delta c^\ell : \nabla \mathbf{s}^\dagger / \tau^\ell.
\]  

Equations (34)–(36) need to be solved subject to the stress-free surface boundary condition \( \mathbf{n} \cdot \mathbf{T}^\dagger = 0 \). At solid-solid boundaries both the adjoint displacement \( \mathbf{s}^\dagger \) and the adjoint traction \( \mathbf{n} \cdot \mathbf{T}^\dagger \) need to be continuous, whereas at fluid-solid boundaries the adjoint traction \( \mathbf{n} \cdot \mathbf{T}^\dagger \) and the normal component of adjoint displacement \( \mathbf{n} \cdot \mathbf{s}^\dagger \) are continuous. Mora (1987, 1988) solved a system of adjoint equations to improve seismic reflection images. In the context of regional simulations of acoustic and elastic wave propagation, Akçelik et al. (2002, 2003) solved a system of equations similar to (34)–(36).

For the sake of completeness we note that the adjoint momentum equation for a rotating, self-gravitating Earth model is given by

\[
\rho(\partial_t^2 \mathbf{s}^\dagger - 2\Omega \times \partial_t \mathbf{s}^\dagger) = \nabla \cdot \mathbf{T}^\dagger + \nabla(\rho \mathbf{s}^\dagger \cdot \mathbf{g}) - \rho \nabla \phi^\dagger - \nabla \cdot (\rho \mathbf{s}^\dagger) \mathbf{g} + \mathbf{f}^\dagger
\]  

where \( \Omega \) denotes the angular velocity of the Earth model and \( \mathbf{g} \) the equilibrium gravitational acceleration. The adjoint perturbation to the gravitational potential \( \phi^\dagger \) is determined by

\[
\nabla^2 \phi^\dagger = -\nabla \cdot (\rho \mathbf{s}^\dagger).
\]  

On a rotating, anelastic Earth model we need to invoke a generalized principle of reciprocity that involves the Green’s tensor of an Earth model with the opposite sense of rotation (Dahlen and Tromp, 1998). This implies that the adjoint equation (37) involves an Earth model that rotates in the opposite direction.
4 Traveltime Tomography

In this section we make the connection between the results obtained thus far and the finite-frequency traveltime tomography advocated by Zhao et al. (2000), Dahlen et al. (2000), and Hung et al. (2000). To accomplish this we introduce the traveltime misfit function

\[ F(m) = \frac{1}{2} \sum_{r=1}^{N} [T_r(m) - T_r^{\text{obs}}]^2, \quad (39) \]

where \( T_r(m) \) denotes the predicted traveltime at station \( r \) for model \( m \), and \( T_r^{\text{obs}} \) the observed traveltime. The gradient of this misfit function is

\[ \delta F = \sum_{r=1}^{N} [T_r(m) - T_r^{\text{obs}}] \delta T_r. \quad (40) \]

We begin with an analysis of the relationship between the traveltime perturbations \( \delta T_r \) and the model perturbations \( \delta m \).

4.1 Banana-doughnut kernels

The Fréchet derivative of the traveltime, \( \delta T_r \), defined in terms of the cross-correlation of an observed and synthetic waveform, may be written in the form (Luo and Schuster, 1991; Marquering et al., 1999; Dahlen et al., 2000)

\[ \delta T_r = \frac{1}{M_{T_r}} \int_0^T w_r(t) \partial_t s_i(x_r, t) \delta s_i(x_r, t) \, dt, \quad \text{(no summation)} \quad (41) \]

where \( w_r \) denotes the cross-correlation window, \( \delta s_i \) the change in displacement due to a model perturbation \( \delta m \), and \( M_{T_r} \) the normalization factor

\[ M_{T_r} = \int_0^T w_r(t) s_i(x_r, t) \partial_t^2 s_i(x_r, t) \, dt. \quad \text{(no summation)} \quad (42) \]

Upon substitution of the Born approximation (3) in (41) we obtain

\[ \begin{align*}
\delta T_r &= -\frac{1}{M_{T_r}} \int_0^T w_r(t) \partial_t s_i(x_r, t) \int_0^t \left[ \delta \rho(x') G_{ij}(x_r, x'; t - t') \partial_{\tau} s_j(x', t') \\
&\quad + \delta c_{ijklm}(x') \partial_{\tau} G_{ij}(x_r, x'; t - t') \partial_t s_m(x', t') \right] d^3x' \, dt'. \quad \text{(no summation over \( i \))} \quad (43)
\end{align*} \]

Next, again using the reciprocity relation (6), we define the traveltime adjoint field

\[ \bar{s}_j^i(x', x_r, T - t') = \frac{1}{M_{T_r}} \int_0^{T-t'} G_{ji}(x', x_r; T - t - t') w_r(T - t) \partial_t s_i(x_r, T - t) \, dt. \quad \text{(no summation)} \quad (44) \]
The associated traveltime adjoint source is given by

\[ f_i^r(x, t) = \frac{1}{M_r} w_r(T - t) \partial_t s_i(x_r, T - t) \delta(x - x_r). \]  

(45)

Note that the traveltime adjoint field \( s_i^J \) is generated by time-reversing the predicted \( i \)th-component of ground velocity at receiver \( r \) only, i.e., unlike the waveform adjoint field defined in (11), the traveltime adjoint field (44) does not depend on the data. With this definition the isotropic traveltime Fréchet derivatives may be rewritten in the form

\[ \delta T_r = \int_V [\tilde{K}_{\rho(\kappa\mu)}(x, x_r) \delta \ln \rho(x) + \tilde{K}_{\mu(\kappa\rho)}(x, x_r) \delta \ln \mu(x) + \tilde{K}_{\kappa(\mu\rho)}(x, x_r) \delta \ln \kappa(x)] \, d^3x. \]  

(46)

The banana-doughnut kernels \( \tilde{K}_{\rho(\kappa\mu)} \), \( \tilde{K}_{\mu(\kappa\rho)} \), and \( \tilde{K}_{\kappa(\mu\rho)} \) are given by

\[ \tilde{K}_{\rho(\kappa\mu)}(x, x_r) = -\int_0^T \rho(x)[\tilde{s}^\dagger(x, x_r, T - t) \cdot \partial_t \hat{s}(x, t)] \, dt, \]  

(47)

\[ \tilde{K}_{\mu(\kappa\rho)}(x, x_r) = -\int_0^T 2\mu(x)\hat{D}^\dagger(x, x_r, T - t) : D(x, t) \, dt, \]  

(48)

\[ \tilde{K}_{\kappa(\mu\rho)}(x, x_r) = -\int_0^T \kappa(x)[\nabla \cdot \tilde{s}^\dagger(x, x_r, T - t)][\nabla \cdot \hat{s}(x, t)] \, dt, \]  

(49)

where \( \hat{D}^\dagger \) denotes the traveltime adjoint strain deviator associated with \( \tilde{s}^\dagger \). The Fréchet derivative of the traveltime \( \delta T_r \) has been recast from an integration over a particular time window \( w_r \) in (41) to a volumetric integral involving 3D kernels in (46). Alternatively, and more sensibly, we may use a representation in terms of wave speeds and density, in which case the traveltime Fréchet derivatives become

\[ \delta T_r = \int_V [\tilde{K}_{\rho(\alpha\beta)}(x, x_r) \delta \ln \rho(x) + \tilde{K}_{\beta(\alpha\rho)}(x, x_r) \delta \ln \beta(x) + \tilde{K}_{\alpha(\beta\rho)}(x, x_r) \delta \ln \alpha(x)] \, d^3x, \]  

(50)

where

\[ \tilde{K}_{\rho(\alpha\beta)} = \tilde{K}_{\rho(\kappa\mu)} + \tilde{K}_{\kappa(\mu\rho)} + \tilde{K}_{\mu(\kappa\rho)}, \]

\[ \tilde{K}_{\beta(\alpha\rho)} = 2\left( \tilde{K}_{\mu(\kappa\rho)} - \frac{4\mu}{3\kappa} \tilde{K}_{\kappa(\mu\rho)} \right), \]

\[ \tilde{K}_{\alpha(\beta\rho)} = 2\left( \frac{3}{1 + \frac{4\mu}{3\kappa}} \right) \tilde{K}_{\kappa(\mu\rho)}. \]  

(51)

The 3D kernels \( \tilde{K}_m \) may be calculated on the fly, using the final displacement field \( \hat{s}(x, T) \) as a starting point for a calculation backward in time, while at the same time generating the traveltime adjoint field \( \tilde{s}^\dagger \). This doubles the memory and CPU requirements compared to a single forward calculation. Note that the kernels \( \tilde{K}_m \) are specific to a particular combination between an earthquake and a station. Thus, for each event one needs to perform an adjoint calculation for every receiver. Equation (50) is identical to the finite-frequency traveltime expression derived by Zhao et al. (2000) and Dahlen et al. (2000). The corresponding banana-doughnut kernels were
visualized by Marquering et al. (1999), Zhao et al. (2000), and Hung et al. (2000).

4.2 Misfit kernels

The Fréchet derivative of the traveltime misfit function (40) may be written in the form

$$\delta F = \int_V \left[ K_{\rho(\alpha\beta)}(x) \delta \ln \rho(x) + K_{\beta(\alpha\rho)}(x) \delta \ln \beta(x) + K_{\alpha(\beta\rho)}(x) \delta \ln \alpha(x) \right] d^3x,$$

where in this case the traveltime misfit kernels $K_{\rho(\alpha\beta)}$, $K_{\beta(\alpha\rho)}$, and $K_{\alpha(\beta\rho)}$ are given by

$$K_{\rho(\alpha\beta)}(x) = \sum_{r=1}^{N} (T_r - T_r^{\text{obs}}) \bar{K}_{\rho(\alpha\beta)}(x, x_r),$$

$$K_{\beta(\alpha\rho)}(x) = \sum_{r=1}^{N} (T_r - T_r^{\text{obs}}) \bar{K}_{\beta(\alpha\rho)}(x, x_r),$$

$$K_{\alpha(\beta\rho)}(x) = \sum_{r=1}^{N} (T_r - T_r^{\text{obs}}) \bar{K}_{\alpha(\beta\rho)}(x, x_r).$$

The kernels (53)–(55) are weighted sums of the banana-doughnut kernels (51) for all the receivers $r = 1, \ldots, N$, with weights determined by the traveltime anomaly $T_r - T_r^{\text{obs}}$ at a particular receiver $r$. Note that by defining the combined traveltime adjoint field

$$s^\dagger(x, t) = \sum_{r=1}^{N} (T_r - T_r^{\text{obs}}) \bar{s}^\dagger(x, x_r, t),$$

the kernels (53)-(55) may be calculated based upon a single adjoint calculation for each earthquake by using the combined traveltime adjoint source

$$f^\dagger_i(x, t) = \sum_{r=1}^{N} (T_r - T_r^{\text{obs}}) \frac{1}{M_{Tr}} w_i(T - t) \partial_t s_i(x_r, T - t) \delta(x - x_r).$$

Thus, if there are $n_{\text{receivers}}$ receivers and $n_{\text{events}}$ earthquakes, the calculation of the 3D kernels that determine the gradient of the traveltime misfit function requires just two 3D simulations for each earthquake, i.e., a total of $2n_{\text{events}}$ simulations. In contrast, the calculation of receiver-specific banana-doughnut kernels requires $(n_{\text{receivers}} + 1)n_{\text{events}}$ calculations. However, an advantage of having access to individual kernels is that one can calculate not only the gradient, but also the Hessian for the inverse problem.

We note that any tomographic inversion based upon the finite-frequency expression (46), e.g., Montelli et al. (2004), will involve the construction of these weighted averages of banana-doughnut kernels during the inversion process: if we write the linear inverse problem as $A\mathbf{m} = \mathbf{d}$, where $\mathbf{m}$ denotes the model vector, $\mathbf{d}$ the traveltime anomaly vector, and $A$ the matrix of traveltime Fréchet derivatives, then the generalized inverse may be expressed as $\mathbf{m} = (A^T A)^{-1} A^T \mathbf{d}$. The quantity $A^T \mathbf{d}$ is precisely the weighted sum of banana-doughnut kernels.
4.3 Differential traveltime tomography

Suppose we have a data set of observed differential traveltimes \( \Delta T_{r}^{\text{obs}}, r = 1, \ldots, N \), between two phases A and B, e.g., PP and P. For a given model \( m \), the corresponding predicted differential traveltimes are \( \Delta T_r(m) = T_r^A(m) - T_r^B(m), r = 1, \ldots, N \). We seek to minimize the differential traveltime misfit function

\[
F(m) = \frac{1}{2} \sum_{r=1}^{N} [\Delta T_r(m) - \Delta T_r^{\text{obs}}]^2. \tag{58}
\]

The gradient of this misfit function is

\[
\delta F = \sum_{r=1}^{N} [\Delta T_r(m) - \Delta T_r^{\text{obs}}] \delta \Delta T_r, \tag{59}
\]

where \( \delta \Delta T_r = \delta T_r^A - \delta T_r^B \). Thus the differential traveltime equivalent of (46) or (50) may be obtained by forming the difference \( \delta T_r^A - \delta T_r^B \), which implies that banana-doughnut kernels for differential traveltime measurements may be obtained by simply subtracting the banana-doughnut kernel for phase B, involving the travelt ime adjoint field \( \bar{s}_B^\dagger \), from the banana-doughnut kernel for phase A, involving the travelt ime adjoint field \( \bar{s}_A^\dagger \) (Dahlen et al., 2000). It requires \( 2(n_{\text{receivers}} + 1)n_{\text{events}} \) calculations to construct the differential banana-doughnut kernels. Alternatively, by defining the combined differential traveltime adjoint field

\[
\Delta s^\dagger(x, t) = \sum_{r=1}^{N} (\Delta T_r - \Delta T_r^{\text{obs}}) \left[ \bar{s}_A^\dagger(x, x_r, t) - \bar{s}_B^\dagger(x, x_r, t) \right], \tag{60}
\]

the gradient of the differential traveltime misfit function (59) may be rewritten in the form (52), but using the combined differential traveltime adjoint field (60) generated by the combined differential traveltime adjoint source

\[
f_i^\dagger(x, t) = \sum_{r=1}^{N} (\Delta T_r - \Delta T_r^{\text{obs}}) \left[ \frac{1}{M_A^T} w_r^A(T - t) \partial_t s_i^A(x_r, T - t) - \frac{1}{M_B^T} w_r^B(T - t) \partial_t s_i^B(x_r, T - t) \right] \delta(x - x_r) \tag{61}
\]

in the construction of the associated kernels. This approach involves only one adjoint calculation per event, for a total of \( 2n_{\text{events}} \) simulations. Basically, (60) leads to the construction of weighted differential banana-doughnut kernels, with weights determined by the current differential traveltime anomalies.

5 Amplitude Tomography

We can also choose to measure the misfit between amplitudes. Let \( A_r^{\text{obs}}/A_r(m) \) denote the ratio between the observed and the predicted amplitude of a particular body-wave arrival at station \( r \).
determined based upon cross-correlation \((\text{Ritsema et al., 2002})\). We introduce the amplitude misfit function

\[
F(m) = \frac{1}{2} \sum_{r=1}^{N} [A_r^{\text{obs}}/A_r(m) - 1]^2. \tag{62}
\]

The gradient of this misfit function is

\[
\delta F = \sum_{r=1}^{N} [A_r^{\text{obs}}/A_r(m) - 1] \delta \ln A_r. \tag{63}
\]

Using the definitions in \(\text{Ritsema et al. (2002)}\), assuming per usual that \(d_i = s_i + \delta s_i\), it is readily shown that \((\text{Dahlen and Baig, 2002})\)

\[
\delta \ln A_r = \frac{1}{M_{Ar}} \int_{0}^{T} w_r(t) s_i(x_r, t) \delta s_i(x_r, t) \, dt, \quad \text{(no summation)} \tag{64}
\]

where \(w_r\) denotes the cross-correlation window, \(\delta s_i\) the change in displacement due to a model perturbation \(\delta m\), and \(M_{Ar}\) the normalization factor

\[
M_{Ar} = \int_{0}^{T} w_r(t) s_i^2(x_r, t) \, dt. \tag{65}
\]

Next, we define the amplitude adjoint field

\[
\bar{s}_j^\dagger(x', x_r, T-t') = \frac{1}{M_{Ar}} \int_{0}^{T-t'} G_{ji}(x', x_r; T-t-t') w_r(T-t) s_i(x_r, T-t) \, dt. \quad \text{(no summation)} \tag{66}
\]

The associated amplitude adjoint source is given by

\[
\bar{f}_i^\dagger(x, t) = \frac{1}{M_{Ar}} w_r(T-t) s_i(x_r, T-t) \delta(x - x_r). \tag{67}
\]

Equation (64) may now be written in the form

\[
\delta \ln A_r = \int_V \left[ K_{\rho(\alpha \beta)}(x, x_r) \delta \ln \rho(x) + K_{\beta(\alpha \rho)}(x, x_r) \delta \ln \beta(x) + K_{\alpha(\beta \rho)}(x, x_r) \delta \ln \alpha(x) \right] \, d^3x, \tag{68}
\]

where the amplitude kernels \(K_{\rho(\alpha \beta)}\), \(K_{\beta(\alpha \rho)}\), and \(K_{\alpha(\beta \rho)}\) are calculated based upon the amplitude adjoint field (66). These amplitude kernels are the finite-frequency kernels discussed extensively by \(\text{Dahlen and Baig (2002)}\).

By defining the combined amplitude adjoint field

\[
s_j^\dagger(x, t) = \sum_{r=1}^{N} (A_r^{\text{obs}}/A_r - 1) \bar{s}_j^\dagger(x, x_r, t), \tag{69}
\]

17
we may rewrite the gradient of the amplitude misfit function (63) as

\[
\delta F = \int_V [K_\rho(\alpha\beta)(x) \delta \ln \rho(x) + K_\beta(\alpha\rho)(x) \delta \ln \beta(x) + K_\alpha(\beta\rho)(x) \delta \ln \alpha(x)] \, d^3x. \tag{70}
\]

The amplitude misfit kernels \(K_\rho(\alpha\beta), K_\beta(\alpha\rho)\) and \(K_\alpha(\beta\rho)\) may be calculated based upon a single adjoint calculation for each earthquake by using the combined amplitude adjoint source

\[
f_i^\dagger(x, t) = \sum_{r=1}^N \left( \frac{A_r^{\text{obs}}}{A_r} - 1 \right) \frac{1}{M_{Ar}} w_r(T - t) s_i(x_r, T - t) \delta(x - x_r). \tag{71}\]

### 5.1 Attenuation

A significant portion of the amplitude anomaly may be due to lateral variations in intrinsic attenuation. Dissipation in the Earth is dominated by shear, and therefore we will only consider the effects of shear attenuation represented by the shear quality factor \(Q_\mu\). For an absorption-band solid, i.e., an Earth model in which \(Q\) is observed to be constant over a wide range of angular frequencies \(\omega\), the shear modulus \(\mu\) may be written in the form \((\text{Liu et al., 1976})\)

\[
\mu(\omega) = \mu(\omega_0)[1 + (2/\pi) Q_\mu^{-1} \ln(|\omega|/\omega_0) - i \text{sgn}(\omega) Q_\mu^{-1}], \tag{72}\]

where \(\omega_0\) denotes the reference angular frequency, \(\text{sgn}(\omega)\) denotes the sign of \(\omega\), and we have used the Fourier convention of \(\text{Dahlen and Tromp} (1998)\). Note that \(\mu(-\omega) = \mu^*(\omega)\). The change in the shear modulus \(\delta\mu\) due to perturbations in shear attenuation \(\delta Q_\mu^{-1}\) may be written in the form

\[
\delta\mu(\omega) = \mu(\omega_0) [(2/\pi) \ln(|\omega|/\omega_0) - i \text{sgn}(\omega)] \delta Q_\mu^{-1}. \tag{73}\]

Using (73) in the frequency domain version of the Born approximation (3), defining the wavefield

\[
\psi_i(x, t) = \frac{1}{2\pi} \int_{-\infty}^\infty [(2/\pi) \ln(|\omega|/\omega_0) - i \text{sgn}(\omega)] s_i(x, \omega) \exp(i\omega t) \, d\omega, \tag{74}\]

and introducing the \(Q\) adjoint field \(\bar{s}^\dagger\) generated by the \(Q\) adjoint source

\[
f_i^\dagger(x, t) = \frac{1}{M_{Ar}} w_r(T - t) \psi_i(x_r, T - t) \delta(x - x_r), \tag{75}\]

the amplitude anomaly (64) may be recast in the form

\[
\delta \ln A_r = \int_V \bar{K}_{\mu(s\rho)}(x, x_r) \delta Q_\mu^{-1}(x) \, d^3x, \tag{76}\]

where the kernel \(\bar{K}_{\mu(s\rho)}\) is calculated based upon (48) using the \(Q\) adjoint field \(\bar{s}^\dagger\) generated by (75). Note that when physical dispersion, represented by the term involving \(\ln(|\omega|/\omega_0)\), is ignored, the \(Q\) adjoint field \(\bar{s}^\dagger\) involves the time-reversed Hilbert transform of the displacement field at the receiver.
Introducing the combined Q adjoint field \( s^\dagger \) generated by the combined Q adjoint source

\[
f_i^1(x,t) = \sum_{r=1}^{N} \left( A_r^{\text{obs}}/A_r - 1 \right) \frac{1}{M_{A_r}} w_r(T-t) \psi_i(x_r,T-t) \delta(x - x_r), \tag{77}
\]

the gradient of the misfit function may be written in the form

\[
\delta F = \int_V K_{\mu(\kappa\rho)}(x) \delta Q_{\mu}^{-1}(x) \, d^3x, \tag{78}
\]

where the kernel \( K_{\mu(\kappa\rho)} \) is of the form (17) and may be calculated based upon one adjoint calculation for each earthquake. The result (78) may be used to invert for lateral variations in attenuation. In practice, since amplitude variations are caused by lateral variations in both elastic and anelastic heterogeneity, one should combine the elastic and anelastic Fréchet derivatives (70) and (78), which requires the calculation of two adjoint fields.

6 Generalizations

The results obtained thus far are readily generalized to other representations of the misfit between seismic data and synthetics. Consider for example the Generalized Seismological Data Functionals (GSDF) introduced by Gee and Jordan (1992) and extended by Chen et al. (2004). The objective of the GSDF methodology is to represent the misfit between an observed and a predicted waveform in terms of frequency dependent traveltime, \( \tau_p \), and amplitude, \( \tau_q \), anomalies. Let \( \tau_r(\omega_\lambda) \) denote either the frequency dependent traveltime anomaly \( \tau_p \) or the frequency dependent amplitude anomaly \( \tau_q \) at receiver \( r, r = 1, \ldots, N \), determined at \( L \) discrete angular frequencies \( \omega_\lambda \), \( \lambda = 1, \ldots, L \), for the current model \( m \). We may then define the GSDF misfit function

\[
F(m) = \frac{1}{2} \sum_{r=1}^{N} \sum_{\lambda=1}^{L} [\tau_r(\omega_\lambda)]^2. \tag{79}
\]

The gradient of this misfit function is

\[
\delta F = \sum_{r=1}^{N} \sum_{\lambda=1}^{L} \tau_r(\omega_\lambda) \delta \tau_r(\omega_\lambda). \tag{80}
\]

6.1 Banana-doughnut kernels

For a given receiver location \( x_r \) and target frequency \( \omega_\lambda \), one needs to determine the time-dependent function \( \Psi_i(x_r,t,\omega_\lambda) \) that relates perturbations in the GSDF parameters \( \delta \tau_r(\omega_\lambda) \) to perturbations in the \( i \)th component of the seismogram \( \delta s_i \):

\[
\delta \tau_r(\omega_\lambda) = \int_0^T \Psi_i(x_r,t,\omega_\lambda) \delta s_i(x_r,t) \, dt. \quad \text{(no summation)} \tag{81}
\]
Specific expressions for $\Psi_i(x_r, t, \omega)$ are given in Chen et al. (2004). Note from (41) that in traveltime tomography this relationship is $\Psi_i(x_r, t, \omega) = \frac{w_r(t) \partial_t s_i(x_r, t)}{M_{rT}}$, and from (64) that in amplitude tomography it is $\Psi_i(x_r, t, \omega) = \frac{w_r(t) s_i(x_r, t)}{M_{rA}}$. Substitution of the Born approximation (3) into (81) yields

$$\delta \tau_r(\omega) = - \int_0^T \Psi_i(x_r, t, \omega) \int_0^T \left[ \delta \rho(x') G_{ij}(x_r, x'; t - t') \partial^2_t s_j(x', t') + \delta c_{ijklm} G_{ij}(x_r, x'; t - t') \partial^2_t s_m(x', t') \right] d^3x' dt' dt. \quad \text{(no summation over } i)$$  

(82)

Next, again using the reciprocity relation (6), we define the GSDF adjoint field

$$\bar{s}^\dagger_{ij}(x', x_r, T - t', \omega) = \int_0^{T - t'} G_{ji}(x', x_r; T - t - t') \Psi_i(x_r, T - t, \omega) dt, \quad \text{(no summation)}$$  

(83)

generated by the GSDF adjoint source

$$\bar{j}^\dagger_i(x, t) = \Psi_i(x_r, T-t, \omega) \delta(x - x_r).$$  

(84)

Analogous to (44), this GSDF adjoint field is independent of the data. Now equation (82) may be rewritten in the familiar form

$$\delta \tau_r(\omega) = \int_V \left[ \bar{K}_{\rho(\kappa)}(x, x_r, \omega) \delta \ln \rho(x) + \bar{K}_{\mu(\kappa)}(x, x_r, \omega) \delta \ln \mu(x) + \bar{K}_{\kappa(\mu)}(x, x_r, \omega) \delta \ln \kappa(x) \right] d^3x,$$  

(85)

where the GSDF kernels $\bar{K}_{\rho(\kappa)}$, $\bar{K}_{\mu(\kappa)}$, and $\bar{K}_{\kappa(\mu)}$ are given by

$$\bar{K}_{\rho(\kappa)}(x, x_r, \omega) = - \int_0^T \rho(x) \bar{s}^\dagger(x, x_r, T - t, \omega) \cdot \partial^2_t \bar{s}(x, t) dt,$$  

(86)

$$\bar{K}_{\mu(\kappa)}(x, x_r, \omega) = - \int_0^T 2\mu(x) \bar{D}^\dagger(x, x_r, T - t, \omega) : \bar{D}(x, t) dt,$$  

(87)

$$\bar{K}_{\kappa(\mu)}(x, x_r, \omega) = - \int_0^T \kappa(x) [\nabla \cdot \bar{s}^\dagger(x, x_r, T - t, \omega)] [\nabla \cdot \bar{s}(x, t)] dt,$$  

(88)

and where $\bar{D}^\dagger$ denotes the GSDF adjoint strain deviator associated with the GSDF adjoint field (83). Using the relationships (51), the result (85) may be rewritten in terms of kernels $\tilde{K}_{\rho(\alpha\beta)}$, $\tilde{K}_{\beta(\alpha\mu)}$, and $\tilde{K}_{\alpha(\beta\mu)}$. Kernels for differential GSDF measurements may be obtained by simply subtracting the GSDF kernels for the two phases.
6.2 Misfit kernels

Let us introduce the combined GSDF adjoint field

\[ s^\dagger(x, t) = \sum_{r=1}^{N} \sum_{\lambda=1}^{L} \tau_r(\omega_\lambda) \bar{s}^\dagger(x, x_r, t, \omega_\lambda), \]  

(89)

generated by the combined GSDF adjoint source

\[ f^\dagger_i(x, t) = \sum_{r=1}^{N} \sum_{\lambda=1}^{L} \tau_r(\omega_\lambda) \Psi_i(x_r, T - t, \omega_\lambda) \delta(x - x_r). \]  

(90)

Note that the combined GSDF adjoint field \( s^\dagger \) involves a summation over all stations \( x_r \) and all angular frequencies \( \omega_\lambda \), resulting in a single calculation for each earthquake, a total of \( 2n_{\text{events}} \) simulations per iteration. This should be contrasted with the determination of the receiver-specific GSDF adjoint fields \( \bar{s}^\dagger \), which involves \( L(n_{\text{receivers}} + 1)n_{\text{events}} \) calculations.

The gradient (80) may be written in the form

\[ \delta F = \int_V \left[ K_{\rho(\kappa\mu)}(x) \delta \ln \rho(x) + K_{\mu(\kappa\rho)}(x) \delta \ln \mu(x) + K_{\kappa(\mu\rho)}(x) \delta \ln \kappa(x) \right] d^3x, \]  

(91)

where the combined GSDF kernels \( K_{\rho(\kappa\mu)} \), \( K_{\mu(\kappa\rho)} \), and \( K_{\kappa(\mu\rho)} \) are given by

\[ K_{\rho(\kappa\mu)}(x) = \sum_{r=1}^{N} \sum_{\lambda=1}^{L} \tau_r(\omega_\lambda) \bar{K}_{\rho(\kappa\mu)}(x, x_r, \omega_\lambda), \]  

(92)

\[ K_{\mu(\kappa\rho)}(x) = \sum_{r=1}^{N} \sum_{\lambda=1}^{L} \tau_r(\omega_\lambda) \bar{K}_{\mu(\kappa\rho)}(x, x_r, \omega_\lambda), \]  

(93)

\[ K_{\kappa(\mu\rho)}(x) = \sum_{r=1}^{N} \sum_{\lambda=1}^{L} \tau_r(\omega_\lambda) \bar{K}_{\kappa(\mu\rho)}(x, x_r, \omega_\lambda). \]  

(94)

Per usual, invoking the relationships (20), the gradient (91) may be recast in the more desirable form (19).

7 Kernel Gallery

We use a two-dimensional (2D) elastic wave propagation code to illustrate the construction of sensitivity kernels using the adjoint methodology discussed in this paper. Each kernel is based upon the interaction between a regular wavefield \( s \) and an adjoint wavefield \( s^\dagger \). Changing the adjoint source \( f^\dagger \) results in a different adjoint field \( s^\dagger \) and, hence, different kernels. For example, we can use the residuals between the data and the synthetics as the waveform adjoint source (9) to construct misfit kernels (20), or we can use the synthetic velocity field as the traveltime adjoint source (45) to construct banana-doughnut kernels (51). In this section we present examples of finite-frequency traveltime and amplitude kernels.
7.1 Model setup

We simulate 2D elastic wave propagation using a spectral-element method, which combines the flexible spatial parameterization of finite-element methods with the accuracy of pseudospectral methods (e.g., Komatitsch and Tromp, 1999). The source-receiver geometry and the various SH and P-SV body-wave arrivals are illustrated in Figure 1. The top boundary is a free surface, whereas the remaining three boundaries are absorbing to mimic a half space. The model extends 200 km in width and 80 km in depth, and is homogeneous with density $\rho = 2600 \text{ kg m}^{-3}$, bulk modulus $\kappa = 5.20 \times 10^{10} \text{ Pa}$, and shear modulus $\mu = 2.66 \times 10^{10} \text{ Pa}$; these values correspond to a compressional wave speed of $\alpha = 5800 \text{ m s}^{-1}$ and a shear wave speed of $\beta = 3199 \text{ m s}^{-1}$. We use a simple one-way treatment for the implementation of the absorbing boundary conditions (Komatitsch and Tromp, 1999). For pedagogical reasons, both the source and the receiver are located at a depth of 40 km to generate direct and surface reflected waves, leading to a variety of interesting phases and associated kernels.

The source-time function used in the simulations is a Gaussian of the form

$$h(t) = \left(-2\alpha^3/\sqrt{\pi}\right) (t - t_0) \exp[-\alpha^2(t - t_0)^2],$$

(95)

where $t_0 = 8.0 \text{ s}$, $\alpha = 2\tau_0/\tau$, $\tau_0 = 2.628 \text{ s}$, and $\tau$ is the duration of $h(t)$ (e.g., Figure 2a). The source duration is $\tau = 4.0 \text{ s}$ in each example, with the exception of Figure 6, where we also used $\tau = 8.0 \text{ s}$. In each simulation the source is applied in the $x$ and $y$ directions to generate both P-SV and SH motions (which are of course completely decoupled). Changing the orientation of the source results in different sensitivity kernels.

7.2 Banana-doughnut kernels

Banana-doughnut traveltime kernels are constructed by using the time-reversed velocity field at one particular receiver as the adjoint source, as in (45). Kernels calculated in this manner may be compared with the finite-frequency kernels presented in recent studies using ray-based methods (e.g., Hung et al., 2000). As discussed earlier, the construction of each kernel is based on the interaction between the time-reversed regular field and the adjoint field; hence the ‘interaction field’ can be thought of as propagating from the receiver to the source in reverse time from $t = T$ to $t = 0$.

7.2.1 SH waves

We begin with the simplest case, the SH wavefield. The experimental setup is depicted in Figure 1. Because both the source and the receiver are located at depth, there are two possible arrivals, which we label S and SS. The source-time function used to generate the regular wavefield is shown in Figure 2a, and the associated seismogram with distinct S and SS arrivals is displayed in Figure 2b. Figure 3 illustrates the construction of the $\tilde{K}_{\beta(\alpha\rho)}$ kernel from the interaction between the regular field $s$ and the S adjoint field $s^\dagger$, whose source is shown in Figure 2d. Keep in mind that for increasing time $t$ the regular field propagates from the source to the receiver,
whereas the adjoint field propagates from the receiver to the source. Marching backward in
time from \( t = T \), the traveltime adjoint source (located at the receiver) ‘turns on’ at the precise
moment that the regular S wavefield passes over it (between Figure 3a and 3b). At each moment
in time the two wavefields are combined via (47), (48), (49), or (51) to form the ‘interaction’
field, which is integrated to construct the kernel. In other words, the interaction field represents
the time-dependent integrand in the kernel definition. Once the regular source is ‘extinguished’,
no further contributions are made toward \( \tilde{K}_{\beta(\alpha \rho)} \). Note that \( \tilde{K}_{\beta(\alpha \rho)} \) is cigar-shaped rather than
banana-shaped because the model is homogeneous, and there is no doughnut hole because we are
dealing with 2D rather than 3D kernels. We refer to this example as SH\(_S\), where SH designates
participation of only the \( y \)-component of the wavefields, and the subscript \( S \) denotes the phase
that is being reversed. The pulse is tapered within the time window \( w_r \) using a Welch window
(\textit{Press et al.}, 1994).

Figure 4 shows all six kernels for the SH\(_S\) scenario. These kernels are constructed simulta-
necessarily via the process illustrated in Figure 3 for \( \tilde{K}_{\beta(\alpha \rho)} \). Notice that the relative amplitudes
of the kernels are consistent with what is expected from the relationships in (51). For example,
since \( \tilde{K}_{\kappa(\mu \rho)} = 0 \) and \( \tilde{K}_{\mu(\kappa \rho)} \approx -\tilde{K}_{\rho(\kappa \mu)} \), we see that \( \tilde{K}_{\rho(\alpha \beta)} = \tilde{K}_{\kappa(\mu \rho)} + \tilde{K}_{\kappa(\mu \rho)} + \tilde{K}_{\rho(\kappa \mu)} \) is very
weak. Note that for SH waves we have \( \tilde{K}_{\beta(\alpha \rho)} = 2 \tilde{K}_{\mu(\kappa \rho)} \). Figure 5 shows the effect of reversing
two different time windows of the synthetic velocity field. Reversing the SS pulse gives a kernel
in the shape of a ‘folded-over cigar’ (Figure 5f). The ellipse surrounding the source and receiver
represents SS scatterers with comparable traveltimes to the SS wave reflected at the surface.

Figure 6 illustrates the effect of changing the source duration, \( \tau \) in (95), on the kernels. We
see that the width of the kernel shrinks at higher frequencies. We expect this since in the limit
of infinite frequency the kernel should collapse onto the ray path. Note that the amplitude of the
kernel increases with increasing frequency. This frequency dependence was illustrated by \textit{Hung
et al.} (2000) using a different technique to construct the kernels. Cross-sections of the kernels
(Figure 6d) help to highlight the Fresnel zones. In the case of the SH\(_S\) \( \beta \)-kernel, the broad, low-
sensitivity red zone represents the first Fresnel zone, whereas the sidelobes defined by the narrow,
high-sensitivity green zone correspond to the second Fresnel zone (e.g., \textit{Hung et al.}, 2000).

\subsection{P-SV waves}

The P-SV wavefield is more complicated than the SH wavefield (Figure 1), and even in the ho-

mogeneous case Rayleigh waves arise through interactions at the free surface. Figure 7 illustrates
the construction of the P-SV adjoint source for the PS+SP arrival, and Figure 8 shows the
 corresponding formation of the \( \tilde{K}_{\alpha(\beta \rho)} \) kernel. Notice how the interaction between SP and P\textsuperscript{\dagger}
forms the right portion of the sensitivity kernel, e.g., at \( t = 32.0 \) s (Figure 8b), whereas the left
portion results from the interaction between P and SP\textsuperscript{\dagger}, e.g., at \( t = 16.0 \) s (Figure 8d). This can
be deduced simply by matching up the portions of the regular and adjoint wavefields that are
contributing to the interaction field. These interactions ‘paint’ the resultant sensitivity kernel.
Figure 9 shows all six kernels for the P-SV\(_{PS+SP}\) scenario.

Figure 10 shows the effect of reversing four distinct time windows of the P-SV synthetic
velocity field: the P, PP, PS+SP, and SS arrivals. In Figure 10b we see that the $\alpha$-kernel for P-SV$_P$ is wider than the $\beta$-kernel for SH$_S$ (Figure 5c) for the same source period. This is due to the relatively longer wavelengths of the P waves: $\lambda_\alpha = \alpha T > \lambda_\beta = \beta T$. Figures 10c-f illustrate examples of $\alpha$ and $\beta$ kernels for the P-SV$_{PP}$, P-SV$_{PS+SP}$, and P-SV$_{SS}$ scenarios. Note that, as expected, the $\alpha$-kernel for the SS wave (Figure 10e) is insignificant relative to the $\beta$-kernel (Figure 10f).

### 7.3 Amplitude kernels

Amplitude kernels are constructed by using the time-reversed displacement field at one particular receiver as the adjoint source, as in (67). Thus, the regular displacement field interacts with a time-reversed version of itself, normalized by $M_{Ar}$ in (65), to form the amplitude kernels. Figure 11 shows a comparison of traveltime, amplitude, and attenuation kernels for two examples: SH$_S$ and SH$_{SS}$. The amplitude kernels are qualitatively similar to the examples in Dahlen and Baig (2002). (Note that Dahlen and Baig (2002) use slowness perturbations, not wave speed, and that their color scale is reversed; these two effects ‘cancel’ each other so that their figures are qualitatively similar to ours.) The amplitude sensitivity is maximal along the ray paths, implying that a slow S-wave speed anomaly ($\delta\mu < 0$) situated on the ray path will lead to an increased amplitude measurement at the receiver.

### 7.4 Q kernels

Attenuation kernels are constructed based upon the Q adjoint field generated by the source (75). For the dispersion term in (73) we select $\omega_0 = 1.86$ radians/s, which corresponds to a central period of $T_0 = 3.37$ s. Figure 11e–f show two attenuation kernels. The attenuation kernels are similar to the traveltime kernels, which is expected, since the time-reversed velocity seismogram is a similar adjoint source to the Hilbert-transformed displacement seismogram.

### 8 Source Inversions

The response $s(x, t)$ due to a finite source represented by a moment-density distribution $m(x, t)$ on a fault plane $\Sigma$ may be written in terms of the Green’s tensor $G(x, x'; t - t')$ as (Aki and Richards, 1980; Dahlen and Tromp, 1998)

$$s_i(x, t) = \int_0^t \int_\Sigma \partial_j G_{ik}(x, x', t - t') m_{jk}(x', t') d^2 x' dt'. \quad (96)$$

Therefore, the change in displacement $\delta s$ due to a change in the moment-density tensor $\delta m$ may be written as

$$\delta s_i(x, t) = \int_0^t \int_\Sigma \partial_j G_{ik}(x, x'; t - t') \delta m_{jk}(x', t') d^2 x' dt'. \quad (97)$$
Upon substitution of the derivatives (97) into (2) we find that the change in the waveform misfit function may be expressed as

$$\delta F = \int_0^T \epsilon^t(x, T - t) : \delta \mathbf{m}(x, t) d^2x dt,$$  \hspace{1cm} (98)

where $\epsilon^t = \frac{1}{2}[\nabla s^t + (\nabla s^t)^T]$ denotes the waveform adjoint strain tensor on the finite-fault plane $\Sigma$, and a superscript $T$ denotes the transpose. Alternatively, one may choose to perform source inversions based upon the traveltime misfit function (39), the amplitude misfit function (62), or the GSDF misfit function (79); this simply changes the nature of the adjoint field.

For a point source located at $\mathbf{x}_s$ we seek to determine the Fréchet derivatives with respect to the time-dependent centroid-moment tensor $\mathbf{m}(t)$. In that case the result (98) reduces to

$$\delta F = \int_0^T \epsilon^t(\mathbf{x}_s, T - t) : \delta \mathbf{m}(\mathbf{x}_s, t) dt.$$  \hspace{1cm} (99)

Depending on the choice of misfit function, this implies that one needs to keep track of the time-dependence of the waveform, traveltime, amplitude, or GSDF adjoint strain tensor at the location of the point source.

**8.1 Example**

We consider a simple example of a source inversion in which the response $\mathbf{s}(\mathbf{x}, t)$ is due to a finite point force $\mathbf{f}$ at location $\mathbf{x}_s$. The source may be expressed as

$$\mathbf{f}(\mathbf{x}, t) = h(t)\delta(\mathbf{x} - \mathbf{x}_s)\mathbf{v},$$  \hspace{1cm} (100)

where $\mathbf{x}_s$ denotes its location, $h(t)$ the normalized source-time function, and $\mathbf{v}$ its magnitude and direction. Let us assume that we know the source location $\mathbf{x}_s$ and the source-time function $h(t)$, and that we seek to determine its magnitude and direction $\mathbf{v}$. Then, following the procedure outlined in the previous section, the Fréchet derivative of the misfit function may be expressed as

$$\delta F = \int_0^T h(t)s^t(\mathbf{x}_s, T - t) \cdot \delta \mathbf{v} dt.$$  \hspace{1cm} (101)

Using this result, together with a conjugate-gradient algorithm (Fletcher and Reeves, 1964), we can iteratively solve for the point force vector $\mathbf{v}$.
The algorithm we use may be summarized as follows: given an initial vector \( v^0 \), calculate \( F(v^0) \), \( g^0 = \partial F / \partial v(v^0) \), and set the initial conjugate-gradient search direction equal to minus the initial gradient of the misfit function, \( p^0 = -g^0 \). If \( ||p^0|| < \epsilon \), where \( \epsilon \) is a suitably small number, then \( v^0 \) is the vector we seek to determine, otherwise:

1. Perform a line search to obtain the scalar \( \lambda^k \) that minimizes the function \( \tilde{F}(\lambda) = F(v^k + \lambda p^k) \):
   - Choose a test parameter \( \lambda^k_t = -2\tilde{F}(0)/\tilde{g}(0) \), where \( \tilde{g}(\lambda) = g(v^k + \lambda p^k) \cdot p^k \).
   - Calculate \( F(v^k + \lambda^k_p^k) \) and \( g(v^k + \lambda^k_p^k) \).
   - Interpolate the function \( \tilde{F}(\lambda) \) by a cubic polynomial and obtain the \( \lambda^k \) that gives the minimum value of this polynomial.
2. Update the vector \( v^{k+1} = v^k + \lambda^k p^k \), and calculate \( g^{k+1} = \partial F / \partial v(v^{k+1}) \).
3. Update the conjugate-gradient search direction \( p^{k+1} = -g^{k+1} + \beta_{k+1} p^k \), where \( \beta_{k+1} = g^{k+1} \cdot (g^{k+1} - g^k) / g^k \cdot g^k \).
4. If \( ||p^{k+1}|| < \epsilon \), then \( v^{k+1} \) is the desired vector, otherwise replace \( k \) with \( k + 1 \) and restart from 1.

We apply this source inversion procedure to a synthetic problem based upon the waveform misfit function (1). The 2D model properties are the same as in section 7.1, and the source-receiver geometry is shown in Figure 12a. We use the Ricker wavelet source-time function (95), apply the force in the \( x \) and \( z \) directions to generate P-SV motions, and create ‘data’ for a particular choice of the source vector \( v \) denoted by the triangle in Figure 12b. Figure 12b shows the inverted vector \( v^k \) for the first and second iterations, starting from a zero vector. Starting with a zero initial vector implies that we use the time-reversed data as the initial adjoint source in the inverse problem, precisely as in time-reversal imaging. After two iterations the force vector coincides nicely with the direction and magnitude of the true source. Figure 12c shows that the value of the normalized misfit function \( F(v^k) \) decreases rapidly during the first and second iterations.

Notice from the above algorithm that every conjugate-gradient iteration involves two regular calculations and two adjoint calculations, totaling \( 4n_{\text{iterations}} \) calculations for a complete inversion, where \( n_{\text{iterations}} \) denotes the number of iterations.

9 Joint Inversions

The results obtained thus far may be used as a basis for inversions that involve structural, topographic, and source parameters. For example, the waveform misfit function (1) may be jointly minimized with respect to structural, topographic, and source parameters. In that case, its gradient (2) involves perturbations \( \delta s \) due to structural (3), topographic (21), and source
parameters (97). The gradient of the waveform misfit function may thus be rewritten in the form

$$\delta F = \int_V \left[ K_{\rho(\alpha\beta)}(x) \delta \ln \rho(x) + K_{\beta(\alpha\rho)}(x) \delta \ln \beta(x) + K_{\alpha(\beta\rho)}(x) \delta \ln \alpha(x) \right] d^3x$$

$$+ \int_{\Sigma} K_h(x) \delta h(x) d^2x + \int_{\Sigma_{PS}} K_h(x) \cdot \nabla \delta h(x) d^2x,$$

$$+ \int_0^T \int_{\Sigma} \epsilon \delta m(x, t) d^2x dt,$$

which simply combines (19), (22), and (98). Similar results may be obtained for traveltime, amplitude, and GSDF misfit functions. For a point source one would use (99) instead.

10 Conclusions

We have demonstrated that seismic tomography, adjoint methods, time-reversal imaging, and banana-doughnut kernels are closely related. Tomographic inversions involve 3D kernels that may be calculated based upon interactions between the wavefield for the current model and a wavefield obtained by using time-reversed signals as simultaneous sources at the receivers. In waveform tomography the time-reversed signal consists of the differences between the data and the current synthetics at all the receivers, in traveltime tomography it consists of the predicted velocities at the stations, and in amplitude tomography one uses the predicted displacements. These kernels should be used as a guide for the choice of model parameterization by selecting basis functions that result in large gradients of the misfit function.

The results in this paper may be readily generalized to inversions that involve norms other than the $L^2$-norm (Crase et al., 1990), as well as to objective functions that measure misfit in the frequency or wavelet domain (Pratt, 1999). Different norms or misfit criteria simply change the nature of the time-reversed field that is used in the adjoint calculation, something we illustrated for the Generalized Seismological Data Functionals (GSDF) advocated by Gee and Jordan (1992) and Chen et al. (2004).

Let $n_{events}$ denote the number of earthquakes used in the inverse problem. We have shown that Fréchet derivatives for source and tomographic problems may be obtained based upon just two forward calculations for each earthquake, a total of $2n_{events}$ simulations, independent of the number of receivers. As illustrated for the source problem, a typical conjugate-gradient algorithm involves two Fréchet derivative calculations for each iteration. Let $n_{iterations}$ denote the number of iterations used in the inverse problem. Then the total number of forward calculations involved in a complete inversion is $4n_{events}n_{iterations}$. Considering the fact that on modern parallel computers regional simulations may be performed in less than one hour and global simulations may be performed in a matter of hours on tens of processors, finite-source inversion based upon the adjoint methodology are practical on modest PC clusters, and 3D tomographic inversions should be feasible on the largest parallel machines, e.g., the Earth Simulator (www.es.jamstec.go.jp).
Acknowledgments

This paper was inspired by the delightful 2003 summer school on ‘Imaging of Complex Media with Acoustic and Seismic Waves’ in Cargese, Corsica. We gratefully acknowledge extensive discussions with Tony Dahlen, Tapio Schneider and the members of a Caltech-USC reading class on waveform tomography. This is contribution No. 9070 of the Division of Geological & Planetary Sciences of the California Institute of Technology. We acknowledge support by the National Science Foundation under grant EAR-0309576.

References


Figure 1: Sketch of the 2D model dimensions and the source-receiver geometry for Figures 3–11. The solid line denotes a free surface, whereas dashed lines are absorbing boundaries. The source is indicated by the ⋆ and the receiver by the □. Left: The two possible ray paths for the SH wavefield are labeled S and SS. The ▽ denotes the SS bounce point. Right: The possible body-wave ray paths for the P-SV wavefield. The ray paths are based on a homogeneous model.

Figure 2: Construction of the adjoint source-time function used in calculating SH banana-doughnut kernels. All traces represent the y-component. (a) Source for the regular wavefield. (b) Regular seismogram recorded at the receiver. (c) Velocity seismogram at the receiver. (d) Source for the adjoint wavefield constructed by time-reversing (c) and Welch tapering the S arrival. Note that this includes the normalization factor $M_{T^r} \leq 0$ defined in (42).
Figure 3: Sequence of interactions between the regular and adjoint SH wavefields during the construction of the banana-doughnut kernel $\bar{K}_\beta(\alpha \rho)$. This particular $\bar{K}_\beta(\alpha \rho)$ kernel is for SH$_S$, i.e., the SH $\beta$–kernel obtained by time-reversing the S arrival. The regular and adjoint sources are shown in Figure 2; the model is a homogeneous half space. Each row represents an instantaneous interaction between the regular and adjoint fields. From the left column to the right column are shown the regular field, the adjoint field, the interaction field, and the instantaneous sensitivity to shear velocity perturbations, $\bar{K}_\beta(\alpha \rho)$. The $K_{\beta(\alpha \rho)}$ kernel is constructed by integrating the interaction field, shown in the third column, over time. (a) At this point in time there is no interaction between the regular field and the adjoint field, since the S wave has yet to reach the receiver. (b) Adjoint wavefield ‘lights up’ as the regular wavefield S phase passes over the receiver (traveling toward the source in reverse time). The label S$\sim$S$^\dagger$ indicates interaction between the regular and adjoint S waves, respectively. (c)–(d) The sensitivity kernel forms via the interaction between the regular and adjoint wavefields. (e) Time of regular source initiation, before which no interaction occurs. The source is labeled by the $\star$ and the receiver by the $\square$. 
Figure 4: The six SH$_S$ banana-doughnut kernels. Each kernel is constructed simultaneously as shown in Figure 3 for $\bar{K}_\beta(\alpha\rho)$. Note that $\bar{K}_\alpha(\beta\rho) = \bar{K}_\kappa(\mu\rho) = 0$, $\bar{K}_\beta(\alpha\rho) = 2\bar{K}_\mu(\kappa\rho)$, $\bar{K}_\mu(\kappa\rho) \approx -\bar{K}_\rho(\kappa\mu)$, and $\bar{K}_\rho(\alpha\beta) \approx 0$ for SH propagation.
Figure 5: The effect of time window selection on sensitivity kernels, using $\hat{K}_{\beta(\alpha \rho)} (SH)$ as an example. (a) Source for the regular wavefield. (b) Velocity recorded at the receiver showing the arrivals S and SS. (c) Adjoint source for SH$_S$, constructed by time-reversing S in (b) and normalizing by $M_{Tr}$ defined in (42). (d) $\hat{K}_{\beta(\alpha \rho)}$ for reversing S only. (e) Adjoint source for SH$_{SS}$, constructed by time-reversing SS in (b) and normalizing by $M_{Tr}$. (f) $\hat{K}_{\beta(\alpha \rho)}$ for reversing SS only. Each point on the ellipse represents a scattering point for a path with a comparable traveltime to the SS path. The SS bounce point is labeled by the $\bigtriangledown$, the source by the $\star$, and the receiver by the $\square$. 

Traveltime kernel $\bar{K}_{\beta}$ ($10^{-8}$ s m$^{-2}$)
Figure 6: Frequency dependence of the $\tilde{K}_{\beta(\alpha\rho)}$ sensitivity kernel for SH$_S$. (a) Two source-time functions for the regular wavefield with durations of $\tau = 8.0$ s (dashed) and $\tau = 4.0$ s (solid). (b) $\tilde{K}_{\beta(\alpha\rho)}$ for $\tau = 8.0$ s. (c) $\tilde{K}_{\beta(\alpha\rho)}$ for $\tau = 4.0$ s (see also Figure 3). (d) Depth cross-sections of (b) and (c) at a horizontal distance of $x = 100$ km. As expected, the higher-frequency kernel is narrower in width and greater in amplitude.
Figure 7: Construction of the adjoint source-time function used in calculating P-SV banana-doughnut kernels. (a) Source-time function responsible for the regular wavefield ($x$-component; the $z$-component is zero). (b) Regular seismogram ($x$-component). (c) Velocity seismogram ($x$-component). (d) Source-time function for the adjoint source constructed by time-reversing (c) and Welch tapering the PS+SP arrival ($x$-component). Note that this includes the normalization factor $M_{Tr}$ defined by (42). This is the source-time function used in Figure 8.
Figure 8: Sequence of interactions between the regular and adjoint P-SV wavefields to produce the banana-doughnut kernel $\bar{K}_{\alpha(\beta\rho)}$. This particular $\bar{K}_{\alpha(\beta\rho)}$ is for P-SV$_{PS+SP}$, i.e., the P-SV $\alpha$-kernel obtained by time-reversing the PS+SP arrival. Given the geometry in Figure 1, the SP and PS phases arrive simultaneously, at nearly the same time as the S arrival (Figure 7b). The $x$-$z$ grid in each snapshot is 200 km in width and 80 km in depth. Wavefield snapshots capture the $x$-component of displacement. See Section 7.2 for details, and compare with Figure 3. (a) No interaction between the regular and adjoint fields, since the PS+SP phase has yet to reach the receiver. (b) Adjoint wavefield ‘lights up’ as the regular wavefield PS+SP phase, depicted by the X-shaped crossing of the two green wavefields, passes over the receiver (traveling toward the source). (c)–(d) Sensitivity kernel forms via the interaction between the regular and adjoint fields. (e) Time of regular source initiation, before which no interaction occurs. We use labels $\star$ for the source, $\square$ for the receiver, and $\bigtriangledown$ for the PS (right) and SP (left) bounce points.
Figure 9: The six P-SV PS+SP banana-doughnut kernels. Each kernel is constructed simultaneously, as shown in Figure 8. Notice that the predominant shape of the $K_{\alpha(\beta\rho)}$ kernel is that of two adjacent, folded-over ‘cigars’, the right one for PS and the left for SP. The labels $\nabla$ denotes the PS (right) and SP (left) bounce points.
Figure 10: The effect of time-window selection on sensitivity kernels, using $\tilde{K}_{\alpha(\beta\rho)}$ as an example. See Figure 1 for labeling and Section 7.2 for details. (a) Velocity recorded at the receiver (x-component) showing the consecutive arrivals of $P$, $PP$, $PS+SP$, and $SS$. The $S$ phase is expected to arrive nearly simultaneous with $PS+SP$, but is insignificant on this component. In (b)–(f) we Welch taper one of the pulses and time-reverse it as the adjoint source via the method explained in Figure 7. The color scale varies for each plot according to the value ‘max’. (b) $\tilde{K}_{\alpha(\beta\rho)}$ for reversing the $P$ arrival ($P-SV_P$, max = 1.0). (c) $\tilde{K}_{\alpha(\beta\rho)}$ for reversing the $PP$ arrival ($P-SV_{PP}$, max = 5.0). (d) $\tilde{K}_{\alpha(\beta\rho)}$ for reversing the $PS+SP$ arrival ($P-SV_{PS+SP}$, max = 1.5). (e) $\tilde{K}_{\alpha(\beta\rho)}$ for reversing the $SS$ arrival ($P-SV_{SS}$, max = 2.75). (f) $\tilde{K}_{\beta(\alpha\rho)}$ for reversing the $SS$ arrival ($P-SV_{SS}$, max = 2.75).
Figure 11: Comparison of SH $\bar{K}_{\mu(\kappa\rho)}$ kernels for traveltime, amplitude, and attenuation for the S and SS arrivals. See Figure 1 for labeling and Section 7.3 for details. The color scale varies for each pair of plots according to the value ‘max’. (a–b) Traveltime kernels $\bar{K}_{\mu(\kappa\rho)}$ for reversing the S arrival (a) and the SS arrival (b) (max = $1.75 \times 10^{-8}$ s m$^{-2}$). (c–d) Amplitude kernels $\bar{K}_{\mu(\kappa\rho)}$ for reversing the S arrival (c) and the SS arrival (d) (max = $2.25 \times 10^{-8}$ m$^{-2}$). (e–f) Q kernels $\bar{K}_{\mu(\kappa\rho)}$ for reversing the S arrival (e) and the SS arrival (f) (max = $1.00 \times 10^{-4}$).
Figure 12: (a) Source–receiver geometry for the point force inversion. The dimensions of the model are the same as in Section 7.1. (b) Results of the source inversion based upon the procedures outlined in Section 8.1. Circles indicate consecutive locations of the force vector in the $x$-$z$ plane, with the labels indicating the iteration number (the initial force is zero). The triangle denotes the true force location we used to generate the ‘data’ for this synthetic problem. (c) Decrease of the waveform misfit function $F$ with the number of iterations ($F$ has been normalized by $\frac{1}{2} \sum_{r=1}^{N} \int_{0}^{T} \|d(x_r, t)\|^2 dt$).