

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

Jeroen Tromp, Carl Tape and Qinya Liu

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

Accepted 2004 August 5. Received 2004 March 25

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 $\chi(m)$ 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 $\delta\chi = \int_V K_m(\mathbf{x})\delta \ln m(\mathbf{x}) d^3\mathbf{x}$, 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 \mathbf{s} and the adjoint displacement field \mathbf{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 travelttime 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 \mathbf{s} and one adjoint calculation involving the fields \mathbf{s} and \mathbf{s}^\dagger . In the case of travelttime tomography, the kernels K_m are weighted combinations of banana-doughnut kernels. 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 δh on internal discontinuities may be expressed in terms of 2-D kernels K_h and \mathbf{K}_h in the form $\delta\chi = \int_\Sigma K_h(\mathbf{x})\delta h(\mathbf{x}) d^2\mathbf{x} + \int_{\Sigma_{FS}} \mathbf{K}_h(\mathbf{x}) \cdot \nabla^\Sigma \delta h(\mathbf{x}) d^2\mathbf{x}$, where Σ denotes a solid-solid or fluid-solid boundary and Σ_{FS} a fluid–solid boundary, and ∇^Σ denotes the surface gradient. 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 $\mathbf{m}(\mathbf{x}, t)$. We demonstrate that the Fréchet derivatives of the objective function χ may in this case be written in the form $\delta\chi = \int_0^t \int_\Sigma \epsilon^\dagger(\mathbf{x}, T-t) : \delta\mathbf{m}(\mathbf{x}, t) d^2\mathbf{x} dt$, where ϵ^\dagger denotes the adjoint strain tensor on the finite-fault plane Σ . In the case of a point source this result reduces further to the calculation of the time-dependent adjoint strain tensor ϵ^\dagger at the location of the point source, an approach reminiscent of an acoustic time-reversal mirror. The theory is illustrated for both tomographic and source inversions using a 2-D spectral-element method.

1 INTRODUCTION

Recent advances in high-performance computing and numerical techniques have facilitated fully three-dimensional (3-D) 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) pre-dates an influential paper by Talagrand & 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 backwards 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 3-D kernel and a model perturbation, precisely as articulated by Talagrand & 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 3-D 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 non-destructive 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 those 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 traveltimes 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 & Baig (2002). These kernels recognize the fact that traveltimes and amplitude anomalies are frequency dependent, and incorporate effects associated with wave front healing. Simple 3-D travelttime 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 travelttime or amplitude misfit functions consist of weighted sums of 3-D banana-doughnut kernels. The weights in these sums are determined by the travelttime 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 (2-D) 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 $\mathbf{d}(\mathbf{x}_r, t)$ recorded at N stations $\mathbf{x}_r, r = 1, \dots, N$, and the corresponding synthetics $\mathbf{s}(\mathbf{x}_r, t, \mathbf{m})$ for a given model vector \mathbf{m} . As a measure of the goodness of fit, we introduce the least-squares *waveform misfit function* (Nolet 1987)

$$\chi(m) = \frac{1}{2} \sum_{r=1}^N \int_0^T \|\mathbf{s}(\mathbf{x}_r, t, \mathbf{m}) - \mathbf{d}(\mathbf{x}_r, t)\|^2 dt. \quad (1)$$

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 \mathbf{m} characterizes the current model. An iterative inversion requires the calculation of the Fréchet derivatives

$$\delta\chi = \sum_{r=1}^N \int_0^T [\mathbf{s}(\mathbf{x}_r, t, \mathbf{m}) - \mathbf{d}(\mathbf{x}_r, t)] \cdot \delta\mathbf{s}(\mathbf{x}_r, t, \mathbf{m}) dt, \quad (2)$$

where $\delta\mathbf{s}$ denotes the perturbation in the displacement field \mathbf{s} due to a model perturbation $\delta\mathbf{m}$. In seismic tomography the model vector \mathbf{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 \mathbf{s} on the model \mathbf{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 & Aki 1985). Using index notation and the Einstein summation convention (unless stated otherwise), suppose we have a generic anisotropic background model $\{\rho, c_{ijklm}\}$, where ρ denotes the density and c_{ijklm} an element of the fourth-order elastic tensor, upon which we superimpose perturbations $\{\delta\rho, \delta c_{ijklm}\}$. Then the associated perturbed displacement $\delta\mathbf{s}$ field may be expressed as

$$\delta s_i(\mathbf{x}, t) = - \int_0^t \int_V [\delta\rho(\mathbf{x}') G_{ij}(\mathbf{x}, \mathbf{x}'; t - t') \partial_t^2 s_j(\mathbf{x}', t') + \delta c_{ijklm}(\mathbf{x}') \partial'_k G_{ij}(\mathbf{x}, \mathbf{x}'; t - t') \partial'_l s_m(\mathbf{x}', t')] d^3 \mathbf{x}' dt', \quad (3)$$

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

$$\begin{aligned} \delta\chi = - \sum_{r=1}^N \int_0^T [s_i(\mathbf{x}_r, t) - d_i(\mathbf{x}_r, t)] \int_0^t \int_V & [\delta\rho(\mathbf{x}') G_{ij}(\mathbf{x}_r, \mathbf{x}'; t - t') \partial_t^2 s_j(\mathbf{x}', t') \\ & + \delta c_{ijklm}(\mathbf{x}') \partial'_k G_{ij}(\mathbf{x}_r, \mathbf{x}'; t - t') \partial'_l s_m(\mathbf{x}', t')] d^3 \mathbf{x}' dt' dt. \end{aligned} \quad (4)$$

Let us define the field

$$\Phi_k(\mathbf{x}', t') = \sum_{r=1}^N \int_{t'}^T G_{ik}(\mathbf{x}_r, \mathbf{x}'; t - t')[s_i(\mathbf{x}_r, t) - d_i(\mathbf{x}_r, t)] dt. \quad (5)$$

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

$$G_{ik}(\mathbf{x}_r, \mathbf{x}'; t - t') = G_{ki}(\mathbf{x}', \mathbf{x}_r; t - t'), \quad (6)$$

we may rewrite (5) as

$$\Phi_k(\mathbf{x}', t') = \sum_{r=1}^N \int_{t'}^T G_{ki}(\mathbf{x}', \mathbf{x}_r; t - t')[s_i(\mathbf{x}_r, t) - d_i(\mathbf{x}_r, t)] dt. \quad (7)$$

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

$$\Phi_k(\mathbf{x}', t') = \sum_{r=1}^N \int_0^{T-t'} G_{ki}(\mathbf{x}', \mathbf{x}_r; T - t - t')[s_i(\mathbf{x}_r, T - t) - d_i(\mathbf{x}_r, T - t)] dt. \quad (8)$$

Next, we define what we shall refer to as the *waveform adjoint source*

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

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

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

Using the standard Green's function approach, we introduce the *waveform adjoint field* \mathbf{s}^\dagger generated by the waveform adjoint source (9):

$$s_k^\dagger(\mathbf{x}', t') = \int_0^{t'} \int_V G_{ki}(\mathbf{x}', \mathbf{x}; t' - t) f_i^\dagger(\mathbf{x}, t) d^3 \mathbf{x} dt. \quad (11)$$

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

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

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 \mathbf{s}^\dagger will illuminate the source of the discrepancy between the data, $\mathbf{d}(\mathbf{x}_r, t)$, and the synthetics for the current model, $\mathbf{s}(\mathbf{x}_r, t)$. Alternatively, from the perspective of adjoint methods (Talagrand & 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

$$\delta\chi = \int_V [K_\rho(\mathbf{x}) \delta \ln \rho(\mathbf{x}) + K_{c_{ijklm}}(\mathbf{x}) \delta \ln c_{ijklm}(\mathbf{x})] d^3 \mathbf{x}, \quad (13)$$

where $\delta \ln \rho = \delta \rho / \rho$ and $\delta \ln c_{ijklm} = \delta c_{ijklm} / c_{ijklm}$ (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 & Courtier (1987). The 3-D *waveform misfit kernels* K_ρ and $K_{c_{ijklm}}$ represent Fréchet derivatives with respect to density and the elastic parameters, respectively, and are defined by

$$K_\rho(\mathbf{x}) = - \int_0^T \rho(\mathbf{x}) \mathbf{s}^\dagger(\mathbf{x}, T - t) \cdot \partial_t^2 \mathbf{s}(\mathbf{x}, t) dt, \quad (14)$$

$$K_{ijklm}(\mathbf{x}) = - \int_0^T \epsilon_{jk}^\dagger(\mathbf{x}, T-t) c_{jklm}(\mathbf{x}) \epsilon_{lm}(\mathbf{x}, t) dt, \quad (\text{no summation}) \quad (15)$$

where ϵ_{lm} and ϵ_{jk}^\dagger denote elements of the strain and the waveform adjoint strain tensors, respectively.

For an isotropic material we have $c_{jklm} = (\kappa - 2\mu/3)\delta_{jk}\delta_{lm} + \mu(\delta_{jl}\delta_{km} + \delta_{jm}\delta_{kl})$, and thus (13) becomes

$$\delta\chi = \int_V [K_\rho(\mathbf{x})\delta \ln \rho(\mathbf{x}) + K_\mu(\mathbf{x})\delta \ln \mu(\mathbf{x}) + K_\kappa(\mathbf{x})\delta \ln \kappa(\mathbf{x})] d^3\mathbf{x}. \quad (16)$$

The isotropic misfit kernels K_μ and K_κ represent Fréchet derivatives with respect to the bulk and shear moduli κ and μ , respectively, and are given by

$$K_\mu(\mathbf{x}) = - \int_0^T 2\mu(\mathbf{x}) \mathbf{D}^\dagger(\mathbf{x}, T-t) : \mathbf{D}(\mathbf{x}, t) dt, \quad (17)$$

$$K_\kappa(\mathbf{x}) = - \int_0^T \kappa(\mathbf{x}) [\nabla \cdot \mathbf{s}^\dagger(\mathbf{x}, T-t)] [\nabla \cdot \mathbf{s}(\mathbf{x}, t)] dt, \quad (18)$$

where \mathbf{D} and \mathbf{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 ρ , shear wave speed β and compressional wave speed α as

$$\delta\chi = \int_V [K'_\rho(\mathbf{x})\delta \ln \rho(\mathbf{x}) + K'_\beta(\mathbf{x})\delta \ln \beta(\mathbf{x}) + K'_\alpha(\mathbf{x})\delta \ln \alpha(\mathbf{x})] d^3\mathbf{x}, \quad (19)$$

where the misfit kernels K'_ρ , K'_β and K'_α represent Fréchet derivatives with respect to density, shear wave and compressional wave speed, respectively; they are given in terms of the kernels (14), (17) and (18) by

$$K'_\rho = K_\rho + K_\kappa + K_\mu, \quad K'_\beta = 2 \left(K_\mu - \frac{4}{3} \frac{\mu}{\kappa} K_\kappa \right), \quad K'_\alpha = 2 \left(\frac{\kappa + \frac{4}{3}\mu}{\kappa} \right) K_\kappa. \quad (20)$$

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

Note in (13), (16), and (19) that the kernels are constructed based upon two calculations: one for the displacement field \mathbf{s} and a second for the waveform adjoint field \mathbf{s}^\dagger . The only interactions between \mathbf{s} and its adjoint \mathbf{s}^\dagger occur through the time-reversed source (9). The Fréchet derivatives of the penalty function χ with respect to any model coefficient m may be obtained by performing 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 & 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 \mathbf{s} and the waveform adjoint field \mathbf{s}^\dagger . Therefore, either the regular field \mathbf{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 \mathbf{s}^\dagger and the construction of the kernels, or it needs to be reconstructed on the fly, using the final displacement field $\mathbf{s}(\mathbf{x}, T)$ as a starting point for and integration backward in time (Gauthier *et al.* 1986). This backward integration of the regular field \mathbf{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 $\mathbf{s}(\mathbf{x}, t)$ 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 δh denote topographic perturbations in the direction of the unit outward normal $\hat{\mathbf{n}}$ on solid–solid discontinuities Σ_{SS} or fluid–solid discontinuities Σ_{FS} . It is shown by Dahlen (2004) that the perturbed displacement field $\delta\mathbf{s}$ due to topographic perturbations δh may be written in the form

$$\begin{aligned} \delta s_i(\mathbf{x}, t) = & \int_0^t \int_{\Sigma} [\rho(\mathbf{x}') G_{ij}(\mathbf{x}, \mathbf{x}'; t-t') \partial_t^2 s_j(\mathbf{x}', t') + \partial_k' G_{ij}(\mathbf{x}, \mathbf{x}'; t-t') c_{jklm}(\mathbf{x}') \partial_l' s_m(\mathbf{x}', t')] \\ & - \hat{n}_k(\mathbf{x}') \partial_n' G_{ij}(\mathbf{x}, \mathbf{x}'; t-t') c_{jklm}(\mathbf{x}') \partial_l' s_m(\mathbf{x}', t') \\ & - \hat{n}_k(\mathbf{x}') c_{jklm}(\mathbf{x}') \partial_l' G_{im}(\mathbf{x}, \mathbf{x}'; t-t') \partial_n' s_j(\mathbf{x}', t')]_-^+ \delta h(\mathbf{x}') d^2\mathbf{x}' dt' \\ & + \int_0^t \int_{\Sigma_{FS}} [G_{ik}(\mathbf{x}, \mathbf{x}'; t-t') \hat{n}_j(\mathbf{x}') \hat{n}_p(\mathbf{x}') c_{jplm}(\mathbf{x}') \partial_l' s_m(\mathbf{x}', t')] \\ & + s_k(\mathbf{x}', t') \hat{n}_j(\mathbf{x}') \hat{n}_p(\mathbf{x}') c_{jplm}(\mathbf{x}') \partial_l' G_{im}(\mathbf{x}, \mathbf{x}'; t-t')]_-^+ \nabla_k^\Sigma \delta h(\mathbf{x}') d^2\mathbf{x}' dt', \end{aligned} \quad (21)$$

where $\Sigma = \Sigma_{SS} + \Sigma_{FS}$ denotes all solid–solid and fluid–solid discontinuities. The surface gradient is defined by $\nabla^\Sigma = (\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}) \cdot \nabla$, and the normal derivative ∂_n is given by $\partial_n = \hat{\mathbf{n}} \cdot \nabla$. The notation $[\cdot]^\pm$ 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, δh , may be written in the form

$$\delta\chi = \int_\Sigma K_h(\mathbf{x})\delta h(\mathbf{x})d^2\mathbf{x} + \int_{\Sigma_{FS}} \mathbf{K}_h(\mathbf{x}) \cdot \nabla^\Sigma \delta h(\mathbf{x})d^2\mathbf{x}, \quad (22)$$

where

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

$$\mathbf{K}_h(\mathbf{x}) = \int_0^T [\mathbf{s}^\dagger(\mathbf{x}, T-t)\hat{\mathbf{n}}(\mathbf{x})\hat{\mathbf{n}}(\mathbf{x}) : \mathbf{c}(\mathbf{x}) : \epsilon(\mathbf{x}, t) + \mathbf{s}(\mathbf{x}, t)\hat{\mathbf{n}}(\mathbf{x})\hat{\mathbf{n}}(\mathbf{x}) : \mathbf{c}(\mathbf{x}) : \epsilon^\dagger(\mathbf{x}, T-t)]_-^\pm dt. \quad (24)$$

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

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

$$\mathbf{K}_h(\mathbf{x}) = \int_0^T [\mathbf{s}^\dagger(\mathbf{x}, T-t)[\kappa(\mathbf{x})\nabla \cdot \mathbf{s}(\mathbf{x}, t) + 2\mu(\mathbf{x})\hat{\mathbf{n}}(\mathbf{x}) \cdot \mathbf{D}(\mathbf{x}, t) \cdot \hat{\mathbf{n}}(\mathbf{x})] \\ + \mathbf{s}(\mathbf{x}, t)[\kappa(\mathbf{x})\nabla \cdot \mathbf{s}^\dagger(\mathbf{x}, T-t) + 2\mu(\mathbf{x})\hat{\mathbf{n}}(\mathbf{x}) \cdot \mathbf{D}^\dagger(\mathbf{x}, T-t) \cdot \hat{\mathbf{n}}(\mathbf{x})]]_-^\pm dt. \quad (26)$$

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 $\hat{n}_k c_{jklm} \epsilon_{lm}$ vanishes, and in this case the kernel (23) reduces to:

$$K_h(\mathbf{x}) = - \int_0^T [\rho(\mathbf{x})\mathbf{s}^\dagger(\mathbf{x}, T-t) \cdot \partial_t^2 \mathbf{s}(\mathbf{x}, t) + \epsilon^\dagger(\mathbf{x}, T-t) : \mathbf{c}(\mathbf{x}) : \epsilon(\mathbf{x}, t)] dt. \quad (27)$$

In the isotropic case this becomes

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

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

3 ADJOINT EQUATIONS

In the context of meteorological inverse problems, Talagrand & 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 & Tromp (1999, 2002a); complications due to self-gravitation and rotation are addressed in Komatitsch & 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

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

Here ρ 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}^ℓ , $\ell = 1, \dots, L$, as

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

The symmetric memory variable tensors \mathbf{R}^ℓ 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:

$$\partial_t \mathbf{R}^\ell = -\mathbf{R}^\ell / \tau^{\sigma\ell} + \delta \mathbf{c}^\ell : \nabla \mathbf{s} / \tau^{\sigma\ell}. \quad (31)$$

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

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

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 $\tau_{ijkl}^{\epsilon\ell}$ are unique to each individual component. In (31), the modulus defect δc^ℓ associated with each individual standard linear solid is determined by

$$\delta c_{ijkl}^\ell = -c_{ijkl}^R \left(1 - \tau_{ijkl}^{\epsilon\ell} / \tau^{\sigma\ell}\right). \quad (33)$$

Ignoring possible complications associated with absorbing boundary conditions, equations (29)–(31) need to be solved subject to the stress-free surface boundary condition $\hat{\mathbf{n}} \cdot \mathbf{T} = \mathbf{0}$, where $\hat{\mathbf{n}}$ denotes the unit outward normal to the surface. At solid–solid boundaries both the displacement \mathbf{s} and the traction $\hat{\mathbf{n}} \cdot \mathbf{T}$ need to be continuous, whereas at fluid–solid boundaries the traction $\hat{\mathbf{n}} \cdot \mathbf{T}$ and the normal component of displacement $\hat{\mathbf{n}} \cdot \mathbf{s}$ 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 $\mathbf{f}^\dagger(\mathbf{x}, t) = \sum_{r=1}^N [\mathbf{s}(\mathbf{x}_r, T-t) - \mathbf{d}(\mathbf{x}_r, T-t)]\delta(\mathbf{x} - \mathbf{x}_r)$. Thus we obtain the following system of *adjoint equations*:

$$\rho \partial_t^2 \mathbf{s}^\dagger = \nabla \cdot \mathbf{T}^\dagger + \mathbf{f}^\dagger \quad (34)$$

$$\mathbf{T}^\dagger = \mathbf{c}^U : \nabla \mathbf{s}^\dagger - \sum_{\ell=1}^L \mathbf{R}^{\ell\dagger}, \quad (35)$$

$$\partial_t \mathbf{R}^{\ell\dagger} = -\mathbf{R}^{\ell\dagger} / \tau^{\sigma\ell} + \delta \mathbf{c}^\ell : \nabla \mathbf{s}^\dagger / \tau^{\sigma\ell}. \quad (36)$$

Eqs (34)–(36) need to be solved subject to the stress-free surface boundary condition $\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger = \mathbf{0}$. At solid–solid boundaries both the adjoint displacement \mathbf{s}^\dagger and the adjoint traction $\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger$ need to be continuous, whereas at fluid–solid boundaries the adjoint traction $\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger$ and the normal component of adjoint displacement $\hat{\mathbf{n}} \cdot \mathbf{s}^\dagger$ 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

$$\rho (\partial_t^2 \mathbf{s}^\dagger - 2\Omega \times \partial_t \mathbf{s}^\dagger) = \nabla \cdot \mathbf{T}^\dagger + \nabla(\rho \mathbf{s}^\dagger \cdot \mathbf{g}) - \rho \nabla \phi^\dagger - \nabla \cdot (\rho \mathbf{s}^\dagger) \mathbf{g} + \mathbf{f}^\dagger \quad (37)$$

where Ω denotes the angular velocity of the earth model and \mathbf{g} the equilibrium gravitational acceleration. The adjoint perturbation to the gravitational potential ϕ^\dagger is determined by

$$\nabla^2 \phi^\dagger = -\nabla \cdot (\rho \mathbf{s}^\dagger). \quad (38)$$

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 & Tromp 1998). This implies that the adjoint eq. (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*

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

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

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

We begin with an analysis of the relationship between the traveltime perturbations δT_r and the model perturbations δm .

4.1 Banana-doughnut kernels

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

$$\delta T_r = \frac{1}{N_r} \int_0^T w_r(t) \partial_t s_i(\mathbf{x}_r, t) \delta s_i(\mathbf{x}_r, t) dt, \quad (\text{no summation}) \quad (41)$$

where w_r denotes the cross-correlation window, δs_i the change in displacement due to a model perturbation δm , and N_r the normalization factor

$$N_r = \int_0^T w_r(t) s_i(\mathbf{x}_r, t) \partial_t^2 s_i(\mathbf{x}_r, t) dt, \quad (\text{no summation}). \quad (42)$$

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

$$\delta T_r = -\frac{1}{N_r} \int_0^T w_r(t) \partial_t s_i(\mathbf{x}_r, t) \iint_V [\delta \rho(\mathbf{x}') G_{ij}(\mathbf{x}_r, \mathbf{x}'; t - t') \partial_t^2 s_j(\mathbf{x}', t') \\ + \delta c_{jklm}(\mathbf{x}') \partial_k' G_{ij}(\mathbf{x}_r, \mathbf{x}'; t - t') \partial_l' s_m(\mathbf{x}', t')] d^3 \mathbf{x}' dt' dt, \quad (\text{no summation over } i). \quad (43)$$

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

$$\bar{s}_j^\dagger(\mathbf{x}', \mathbf{x}_r, T - t') = \frac{1}{N_r} \int_0^{T-t'} G_{ji}(\mathbf{x}', \mathbf{x}_r; T - t - t') w_r(T - t) \partial_t s_i(\mathbf{x}_r, T - t) dt, \quad (\text{no summation}). \quad (44)$$

The associated *travelttime adjoint source* is given by

$$\bar{f}_i^\dagger(\mathbf{x}, t) = \frac{1}{N_r} w_r(T - t) \partial_t s_i(\mathbf{x}_r, T - t) \delta(\mathbf{x} - \mathbf{x}_r). \quad (45)$$

Note that the travelttime adjoint field \bar{s}_j^\dagger 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 travelttime adjoint field (44) does not depend on the data. With this definition the isotropic travelttime Fréchet derivatives may be rewritten in the form

$$\delta T_r = \int_V [\bar{K}_\rho(\mathbf{x}, \mathbf{x}_r) \delta \ln \rho(\mathbf{x}) + \bar{K}_\mu(\mathbf{x}, \mathbf{x}_r) \delta \ln \mu(\mathbf{x}) + \bar{K}_\kappa(\mathbf{x}, \mathbf{x}_r) \delta \ln \kappa(\mathbf{x})] d^3 \mathbf{x}. \quad (46)$$

The *banana-doughnut kernels* \bar{K}_ρ , \bar{K}_μ and \bar{K}_κ are given by

$$\bar{K}_\rho(\mathbf{x}, \mathbf{x}_r) = - \int_0^T \rho(\mathbf{x}) [\bar{s}^\dagger(\mathbf{x}, \mathbf{x}_r, T - t) \cdot \partial_t^2 \mathbf{s}(\mathbf{x}, t)] dt, \quad (47)$$

$$\bar{K}_\mu(\mathbf{x}, \mathbf{x}_r) = - \int_0^T 2\mu(\mathbf{x}) \bar{\mathbf{D}}^\dagger(\mathbf{x}, \mathbf{x}_r, T - t) : \mathbf{D}(\mathbf{x}, t) dt, \quad (48)$$

$$\bar{K}_\kappa(\mathbf{x}, \mathbf{x}_r) = - \int_0^T \kappa(\mathbf{x}) [\nabla \cdot \bar{s}^\dagger(\mathbf{x}, \mathbf{x}_r, T - t)] [\nabla \cdot \mathbf{s}(\mathbf{x}, t)] dt, \quad (49)$$

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

$$\delta T_r = \int_V [\bar{K}'_\rho(\mathbf{x}, \mathbf{x}_r) \delta \ln \rho(\mathbf{x}) + \bar{K}_\beta(\mathbf{x}, \mathbf{x}_r) \delta \ln \beta(\mathbf{x}) + \bar{K}_\alpha(\mathbf{x}, \mathbf{x}_r) \delta \ln \alpha(\mathbf{x})] d^3 \mathbf{x}, \quad (50)$$

where

$$\bar{K}'_\rho = \bar{K}_\rho + \bar{K}_\kappa + \bar{K}_\mu, \quad \bar{K}_\alpha = 2 \left(\frac{\kappa + \frac{4}{3}\mu}{\kappa} \right) \bar{K}_\kappa, \quad \bar{K}_\beta = 2 \left(\bar{K}_\mu - \frac{4}{3} \frac{\mu}{\kappa} \bar{K}_\kappa \right). \quad (51)$$

The 3-D kernels \bar{K}_m may be calculated on the fly, using the final displacement field $\mathbf{s}(\mathbf{x}, T)$ as a starting point for a calculation backward in time, while at the same time generating the travelttime adjoint field \bar{s}^\dagger . This doubles the memory and CPU requirements compared with a single forward calculation. Note that the kernels \bar{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. Eq. (50) is identical to the finite-frequency travelttime 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 travelttime misfit function (40) may be written in the form

$$\delta \chi = \int_V [K'_\rho(\mathbf{x}) \delta \ln \rho(\mathbf{x}) + K_\beta(\mathbf{x}) \delta \ln \beta(\mathbf{x}) + K_\alpha(\mathbf{x}) \delta \ln \alpha(\mathbf{x})] d^3 \mathbf{x}, \quad (52)$$

where in this case the *travelttime misfit kernels* K'_ρ , K_β and K_α are given by

$$K'_\rho(\mathbf{x}) = \sum_{r=1}^N (T_r - T_r^{\text{obs}}) \bar{K}'_\rho(\mathbf{x}, \mathbf{x}_r), \quad (53)$$

$$K_\beta(\mathbf{x}) = \sum_{r=1}^N (T_r - T_r^{\text{obs}}) \bar{K}_\beta(\mathbf{x}, \mathbf{x}_r), \quad (54)$$

$$K_\alpha(\mathbf{x}) = \sum_{r=1}^N (T_r - T_r^{\text{obs}}) \bar{K}_\alpha(\mathbf{x}, \mathbf{x}_r). \quad (55)$$

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

$$\mathbf{s}^\dagger(\mathbf{x}, t) = \sum_{r=1}^N (T_r - T_r^{\text{obs}}) \bar{\mathbf{s}}^\dagger(\mathbf{x}, \mathbf{x}_r, t), \quad (56)$$

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

$$f_i^\dagger(\mathbf{x}, t) = \sum_{r=1}^N (T_r - T_r^{\text{obs}}) \frac{1}{N_r} w_r(T-t) \partial_t s_i(\mathbf{x}_r, T-t) \delta(\mathbf{x} - \mathbf{x}_r). \quad (57)$$

Thus, if there are $n_{\text{receivers}}$ receivers and n_{events} earthquakes, the calculation of the 3-D kernels that determine the gradient of the travelttime misfit function requires just two 3-D 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 $\mathbf{Am} = \mathbf{d}$, where \mathbf{m} denotes the model vector, \mathbf{d} the travelttime anomaly vector and \mathbf{A} the matrix of travelttime Fréchet derivatives, then the generalized inverse may be expressed as $\mathbf{m} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{d}$. The quantity $\mathbf{A}^T \mathbf{d}$ is precisely the weighted sum of banana-doughnut kernels.

4.3 Differential travelttime tomography

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

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

The gradient of this misfit function is

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

where $\delta \Delta T_r = \delta T_r^A - \delta T_r^B$. Thus the differential travelttime 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 travelttime measurements may be obtained by simply subtracting the banana-doughnut kernel for phase B, involving the travelttime adjoint field $\bar{\mathbf{s}}^{B\dagger}$, from the banana-doughnut kernel for phase A, involving the travelttime adjoint field $\bar{\mathbf{s}}^{A\dagger}$ (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 travelttime adjoint field*

$$\Delta \mathbf{s}^\dagger(\mathbf{x}, t) = \sum_{r=1}^N (\Delta T_r - \Delta T_r^{\text{obs}}) [\bar{\mathbf{s}}^{A\dagger}(\mathbf{x}, \mathbf{x}_r, t) - \bar{\mathbf{s}}^{B\dagger}(\mathbf{x}, \mathbf{x}_r, t)], \quad (60)$$

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

$$f_i^\dagger(\mathbf{x}, t) = \sum_{r=1}^N (\Delta T_r - \Delta T_r^{\text{obs}}) \left[\frac{1}{N_r^A} w_r^A(T-t) \partial_t s_i^A(\mathbf{x}_r, T-t) - \frac{1}{N_r^B} w_r^B(T-t) \partial_t s_i^B(\mathbf{x}_r, T-t) \right] \delta(\mathbf{x} - \mathbf{x}_r) \quad (61)$$

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 travelttime 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 (Ritsema *et al.* 2002). We introduce the *amplitude misfit function*

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

The gradient of this misfit function is

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

Using the definitions in Ritsema *et al.* (2002), assuming as per usual that $d_i = s_i + \delta s_i$, it is readily shown that (Dahlen & Baig 2002)

$$\delta \ln A_r = \frac{1}{M_r} \int_0^T w_r(t) s_i(\mathbf{x}_r, t) \delta s_i(\mathbf{x}_r, t) dt, \quad (\text{no summation}) \quad (64)$$

where w_r denotes the cross-correlation window, δs_i the change in displacement due to a model perturbation δm , and M_r the normalization factor

$$M_r = \int_0^T w_r(t) s_i^2(\mathbf{x}_r, t) dt. \quad (65)$$

Next, we define the *amplitude adjoint field*

$$\bar{s}_j^\dagger(\mathbf{x}', \mathbf{x}_r, T - t') = \frac{1}{M_r} \int_0^{T-t'} G_{ji}(\mathbf{x}', \mathbf{x}_r; T - t - t') w_r(T - t) s_i(\mathbf{x}_r, T - t) dt, \quad (\text{no summation}). \quad (66)$$

The associated *amplitude adjoint source* is given by

$$\bar{f}_i^\dagger(\mathbf{x}, t) = \frac{1}{M_r} w_r(T - t) s_i(\mathbf{x}_r, T - t) \delta(\mathbf{x} - \mathbf{x}_r). \quad (67)$$

Eq. (64) may now be written in the form

$$\delta \ln A_r = \int_V [\bar{K}'_\rho(\mathbf{x}, \mathbf{x}_r) \delta \ln \rho'(\mathbf{x}) + \bar{K}_\beta(\mathbf{x}, \mathbf{x}_r) \delta \ln \beta(\mathbf{x}) + \bar{K}_\alpha(\mathbf{x}, \mathbf{x}_r) \delta \ln \alpha(\mathbf{x})] d^3 \mathbf{x}, \quad (68)$$

where the *amplitude kernels* \bar{K}'_ρ , \bar{K}_β and \bar{K}_α are calculated based upon the amplitude adjoint field (66). These amplitude kernels are the finite-frequency kernels discussed extensively by Dahlen & Baig (2002).

By defining the *combined amplitude adjoint field*

$$\mathbf{s}^\dagger(\mathbf{x}, t) = \sum_{r=1}^N (A_r^{\text{obs}} / A_r - 1) \bar{s}_r^\dagger(\mathbf{x}, \mathbf{x}_r, t), \quad (69)$$

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

$$\delta \chi = \int_V [K'_\rho(\mathbf{x}) \delta \ln \rho(\mathbf{x}) + K_\beta(\mathbf{x}) \delta \ln \beta(\mathbf{x}) + K_\alpha(\mathbf{x}) \delta \ln \alpha(\mathbf{x})] d^3 \mathbf{x}. \quad (70)$$

The amplitude misfit kernels K'_ρ , K_β and K_α may be calculated based upon a single adjoint calculation for each earthquake by using the *combined amplitude adjoint source*

$$f_i^\dagger(\mathbf{x}, t) = \sum_{r=1}^N (A_r^{\text{obs}} / A_r - 1) \frac{1}{M_r} w_r(T - t) s_i(\mathbf{x}_r, T - t) \delta(\mathbf{x} - \mathbf{x}_r). \quad (71)$$

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_μ . 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 ω , the shear modulus μ may be written in the form (Liu *et al.* 1976)

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

where ω_0 denotes the reference angular frequency, $\text{sgn}(\omega)$ denotes the sign of ω , and we have used the Fourier convention of Dahlen & Tromp (1998). Note that $\mu(-\omega) = \mu^*(\omega)$. The change in the shear modulus $\delta \mu$ due to perturbations in shear attenuation δQ_μ^{-1} may be written in the form

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

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

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

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

$$\bar{f}_i^\dagger(\mathbf{x}, t) = \frac{1}{M_r} w_r(T - t) \psi_i(\mathbf{x}_r, T - t) \delta(\mathbf{x} - \mathbf{x}_r), \quad (75)$$

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

$$\delta \ln A_r = \int_V \bar{K}_\mu(\mathbf{x}, \mathbf{x}_r) \delta Q_\mu^{-1}(\mathbf{x}) d^3 \mathbf{x}, \quad (76)$$

where the kernel \bar{K}_μ 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* \mathbf{s}^\dagger generated by the *combined Q adjoint source*

$$f_i^\dagger(\mathbf{x}, t) = \sum_{r=1}^N \left(A_r^{\text{obs}} / A_r - 1 \right) \frac{1}{M_r} w_r(T-t) \psi_i(\mathbf{x}_r, T-t) \delta(\mathbf{x} - \mathbf{x}_r), \quad (77)$$

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

$$\delta\chi = \int_V K_\mu(\mathbf{x}) \delta Q_\mu^{-1}(\mathbf{x}) d^3\mathbf{x}, \quad (78)$$

where the kernel K_μ 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 & 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 traveltimes, τ_p , and amplitudes, τ_q , anomalies. Let $\tau_r(\omega_\lambda)$ denote either the frequency-dependent traveltimes τ_p or the frequency-dependent amplitude anomalies τ_q at receiver r , $r = 1, \dots, N$, determined at L discrete angular frequencies ω_λ , $\lambda = 1, \dots, L$, for the current model m . We may then define the *GSDF misfit function*

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

The gradient of this misfit function is

$$\delta\chi = \sum_{r=1}^N \sum_{\lambda=1}^L \tau_r(\omega_\lambda) \delta\tau_r(\omega_\lambda). \quad (80)$$

6.1 Banana-doughnut kernels

For a given receiver location \mathbf{x}_r and target frequency ω_λ , one needs to determine the time-dependent function $\Psi_i(\mathbf{x}_r, t, \omega_\lambda)$ that relates perturbations in the GSDF parameters $\delta\tau_r(\omega_\lambda)$ to perturbations in the i th component of the seismogram δs_i :

$$\delta\tau_r(\omega_\lambda) = \int_0^T \Psi_i(\mathbf{x}_r, t, \omega_\lambda) \delta s_i(\mathbf{x}_r, t) dt, \quad (\text{no summation}). \quad (81)$$

Specific expressions for $\Psi_i(\mathbf{x}_r, t, \omega_\lambda)$ are given in Chen *et al.* (2004). Note from (41) that in traveltime tomography this relationship is $\Psi_i(\mathbf{x}_r, t, \omega_\lambda) = w_r(t) \partial_t s_i(\mathbf{x}_r, t) / N_r$, and from (64) that in amplitude tomography it is $\Psi_i(\mathbf{x}_r, t, \omega_\lambda) = w_r(t) s_i(\mathbf{x}_r, t) / M_r$. Substitution of the Born approximation (3) into (81) yields

$$\begin{aligned} \delta\tau_r(\omega_\lambda) = & - \int_0^T \Psi_i(\mathbf{x}_r, t, \omega_\lambda) \int_0^t \int_V [\delta\rho(\mathbf{x}') G_{ij}(\mathbf{x}_r, \mathbf{x}'; t-t') \partial_t^2 s_j(\mathbf{x}', t')] d^3\mathbf{x}' dt' dt, \\ & + \delta c_{jklm}(\mathbf{x}') \partial_k' G_{ij}(\mathbf{x}_r, \mathbf{x}'; t-t') \partial_l' s_m(\mathbf{x}', t')] d^3\mathbf{x}' dt' dt, \quad (\text{no summation over } i). \end{aligned} \quad (82)$$

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

$$\bar{s}_j^\dagger(\mathbf{x}', \mathbf{x}_r, T-t', \omega_\lambda) = \int_0^{T-t'} G_{ji}(\mathbf{x}', \mathbf{x}_r; T-t-t') \Psi_i(\mathbf{x}_r, T-t, \omega_\lambda) dt, \quad (\text{no summation}) \quad (83)$$

generated by the *GSDF adjoint source*

$$\bar{f}_i^\dagger(\mathbf{x}, t) = \Psi_i(\mathbf{x}_r, T-t, \omega_\lambda) \delta(\mathbf{x} - \mathbf{x}_r). \quad (84)$$

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

$$\delta\tau_r(\omega_\lambda) = \int_V [\bar{K}_\rho(\mathbf{x}, \mathbf{x}_r, \omega_\lambda) \delta \ln \rho(\mathbf{x}) + \bar{K}_\mu(\mathbf{x}, \mathbf{x}_r, \omega_\lambda) \delta \ln \mu(\mathbf{x}) + \bar{K}_\kappa(\mathbf{x}, \mathbf{x}_r, \omega_\lambda) \delta \ln \kappa(\mathbf{x})] d^3\mathbf{x}, \quad (85)$$

where the *GSDF kernels* \bar{K}_ρ , \bar{K}_μ and \bar{K}_κ are given by

$$\bar{K}_\rho(\mathbf{x}, \mathbf{x}_r, \omega_\lambda) = - \int_0^T \rho(\mathbf{x}) [\bar{s}^\dagger(\mathbf{x}, \mathbf{x}_r, T-t, \omega_\lambda) \cdot \partial_t^2 \mathbf{s}(\mathbf{x}, t)] dt, \quad (86)$$

$$\bar{K}_\mu(\mathbf{x}, \mathbf{x}_r, \omega_\lambda) = - \int_0^T 2\mu(\mathbf{x}) \bar{\mathbf{D}}^\dagger(\mathbf{x}, \mathbf{x}_r, T-t, \omega_\lambda) : \mathbf{D}(\mathbf{x}, t) dt, \quad (87)$$

$$\bar{K}_\kappa(\mathbf{x}, \mathbf{x}_r, \omega_\lambda) = - \int_0^T \kappa(\mathbf{x}) [\nabla \cdot \bar{s}^\dagger(\mathbf{x}, \mathbf{x}_r, T-t, \omega_\lambda)] [\nabla \cdot \mathbf{s}(\mathbf{x}, t)] dt, \quad (88)$$

and where $\bar{\mathbf{D}}^\dagger$ 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 $\bar{K}_{\rho'}$, \bar{K}_β and \bar{K}_α . 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*

$$\mathbf{s}^\dagger(\mathbf{x}, t) = \sum_{r=1}^N \sum_{\lambda=1}^L \tau_r(\omega_\lambda) \bar{\mathbf{s}}^\dagger(\mathbf{x}, \mathbf{x}_r, t, \omega_\lambda), \quad (89)$$

generated by the *combined GSDF adjoint source*

$$f_i^\dagger(\mathbf{x}, t) = \sum_{r=1}^N \sum_{\lambda=1}^L \tau_r(\omega_\lambda) \Psi_i(\mathbf{x}_r, T - t, \omega_\lambda) \delta(\mathbf{x} - \mathbf{x}_r). \quad (90)$$

Note that the combined GSDF adjoint field \mathbf{s}^\dagger involves a summation over all stations \mathbf{x}_r and all angular frequencies ω_λ , 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 $\bar{\mathbf{s}}^\dagger$, which involves $L(n_{\text{receivers}} + 1)n_{\text{events}}$ calculations.

The gradient (80) may be written in the form

$$\delta\chi = \int_V [K_\rho(\mathbf{x}) \delta \ln \rho(\mathbf{x}) + K_\mu(\mathbf{x}) \delta \ln \mu(\mathbf{x}) + K_\kappa(\mathbf{x}) \delta \ln \kappa(\mathbf{x})] d^3 \mathbf{x}, \quad (91)$$

where the *combined GSDF kernels* K_ρ , K_μ and K_κ are given by

$$K_\rho(\mathbf{x}) = \sum_{r=1}^N \sum_{\lambda=1}^L \tau_r(\omega_\lambda) \bar{K}_\rho(\mathbf{x}, \mathbf{x}_r, \omega_\lambda), \quad (92)$$

$$K_\mu(\mathbf{x}) = \sum_{r=1}^N \sum_{\lambda=1}^L \tau_r(\omega_\lambda) \bar{K}_\mu(\mathbf{x}, \mathbf{x}_r, \omega_\lambda), \quad (93)$$

$$K_\kappa(\mathbf{x}) = \sum_{r=1}^N \sum_{\lambda=1}^L \tau_r(\omega_\lambda) \bar{K}_\kappa(\mathbf{x}, \mathbf{x}_r, \omega_\lambda). \quad (94)$$

As 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 (2-D) 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 \mathbf{s} and an adjoint wavefield \mathbf{s}^\dagger . Changing the adjoint source \mathbf{f}^\dagger results in a different adjoint field \mathbf{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 set-up

We simulate 2-D elastic wave propagation using a spectral-element method, which combines the flexible spatial parametrization of finite-element methods with the accuracy of pseudospectral methods (e.g. Komatsch & Tromp 1999). The source-receiver geometry and the various *SH* and *P-SV* body-wave arrivals are illustrated in Fig. 1. The top boundary is a free surface, whereas the remaining three boundaries

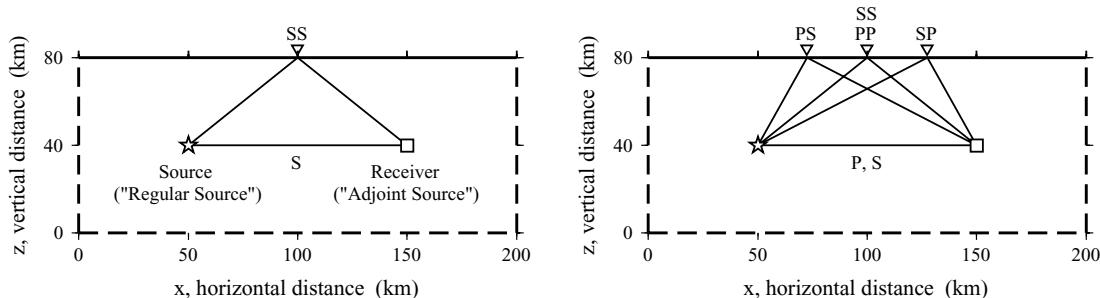


Figure 1. Sketch of the 2-D model dimensions and the source-receiver geometry for Figs 3–11. The solid line denotes a free surface, whereas dashed lines are absorbing boundaries. The source is indicated by the \star and the receiver by the \square . Left: The two possible ray paths for the *SH* wavefield are labelled *S* and *SS*. The ∇ denotes the *SS* bounce point. Right: The possible body wave ray paths for the *P-SV* wavefield.

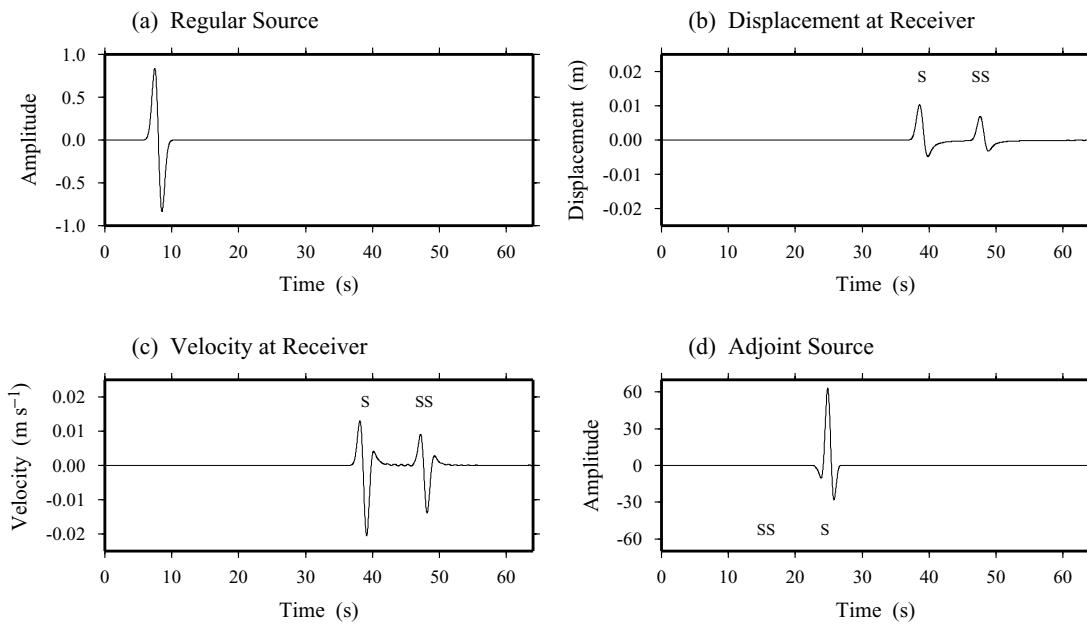


Figure 2. Construction of the adjoint source–time function used in calculating SH banana-doughnut kernels. All traces represent the y component. (a) Source for the regular wavefield. (b) Regular seismogram recorded at the receiver. (c) Velocity seismogram at the receiver. (d) Source for the adjoint wavefield constructed by time-reversing (c) and Welch tapering the S arrival. Note that this includes the normalization factor N , defined in (42).

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 $\mu = 2.66 \times 10^{10} \text{ Pa}$. We use a simple one-way treatment for the implementation of the absorbing boundary conditions (Komatitsch & 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 Ricker wavelet of the form

$$h(t) = (-2\alpha^3/\pi)(t - t_0) \exp[-\alpha^2(t - t_0)^2], \quad (95)$$

where $t_0 = 8.0 \text{ s}$, $\alpha = 2\tau_0/\tau$, $\tau_0 = 2.628 \text{ s}$ and τ is the duration of $h(t)$ (e.g. Fig. 2a). The source duration is $\tau = 4.0 \text{ s}$ in each example, with the exception of Fig. 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 traveltimes 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 set-up is depicted in Fig. 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 Fig. 2(a), and the associated seismogram with distinct S and SS arrivals is displayed in Fig. 2(b). Fig. 3 illustrates the construction of the \bar{K}_β kernel from the interaction between the regular field s and the S adjoint field s^\dagger , whose source is shown in Fig. 2(d). 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 backwards 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 Figs 3a and b). 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 \bar{K}_β . Note that \bar{K}_β is cigar-shaped rather than banana-shaped because the model is homogeneous, and there is no doughnut hole because we are dealing with 2-D rather than 3-D kernels. We refer to this example as SH_S , where SH designates participation of only the y -component of

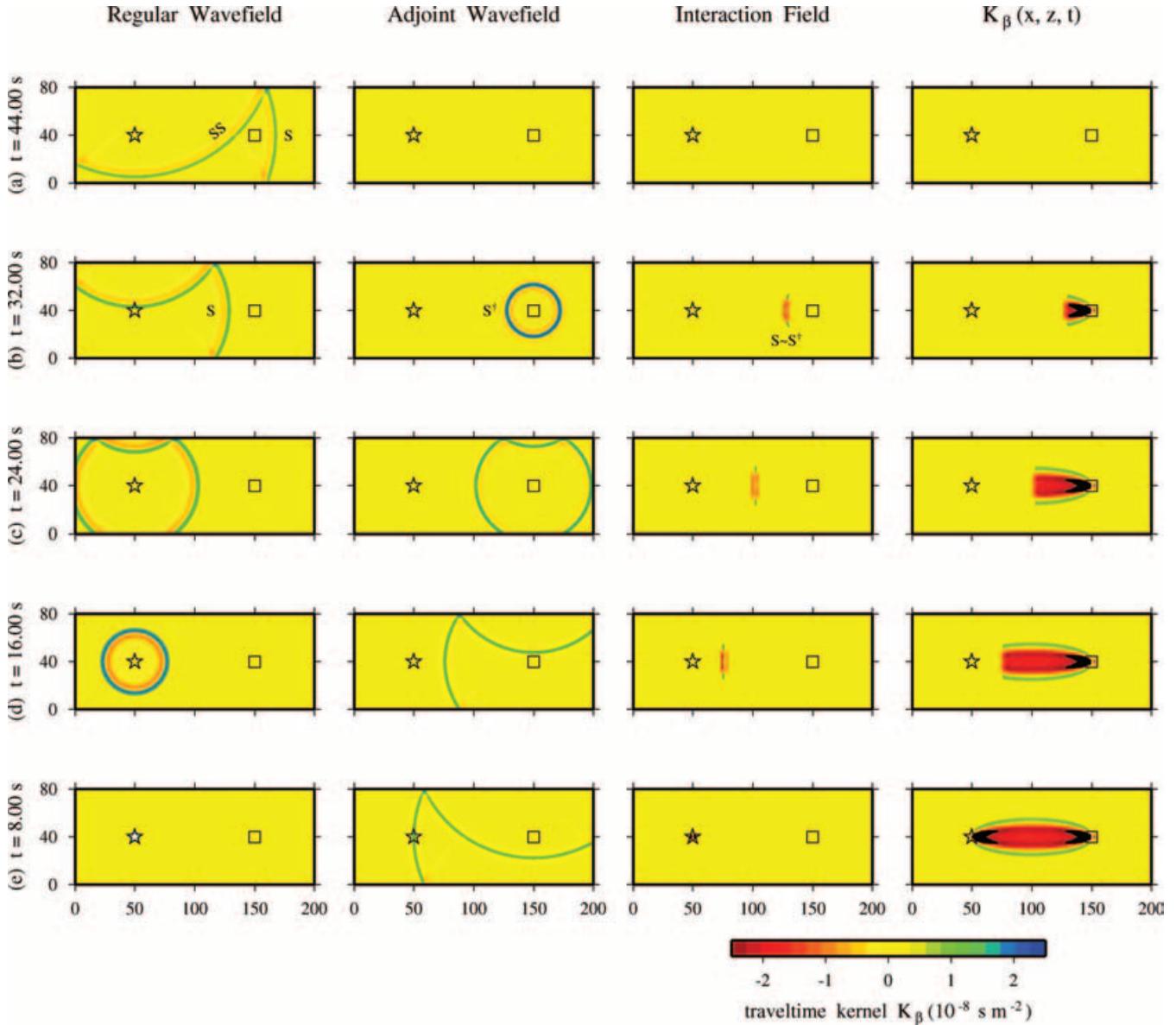


Figure 3. Sequence of interactions between the regular and adjoint SH wavefields during the construction of the banana-doughnut kernel \bar{K}_β . This particular \bar{K}_β kernel is for SH_S , i.e. the $SH \beta$ kernel obtained by time-reversing the S arrival. The regular and adjoint sources are shown in Fig. 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}_β . The \bar{K}_β 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 (travelling toward the source in reverse time). The label $S \sim S^\dagger$ indicates interaction between the regular and adjoint S waves. (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 labelled by the \star and the receiver by the \square .

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 (Press *et al.* 1994).

Fig. 4 shows all six kernels for the SH_S scenario. These kernels are constructed simultaneously via the process illustrated in Fig. 3 for \bar{K}_β . Notice that the relative amplitudes of the kernels are consistent with what is expected from the relationships in (51). For example, since $\bar{K}_\kappa = 0$ and $\bar{K}_\mu \approx -\bar{K}_\rho$, we see that $\bar{K}_{\rho'} = \bar{K}_\mu + \bar{K}_\kappa + \bar{K}_\rho$ is very weak. Note that for SH waves we have $\bar{K}_\beta = 2\bar{K}_\mu$. Fig. 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’ (Fig. 5f). The ellipse surrounding the source and receiver represents SS scatterers with comparable traveltimes to the SS wave reflected at the surface.

Fig. 6 illustrates the effect of changing the source duration, τ 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

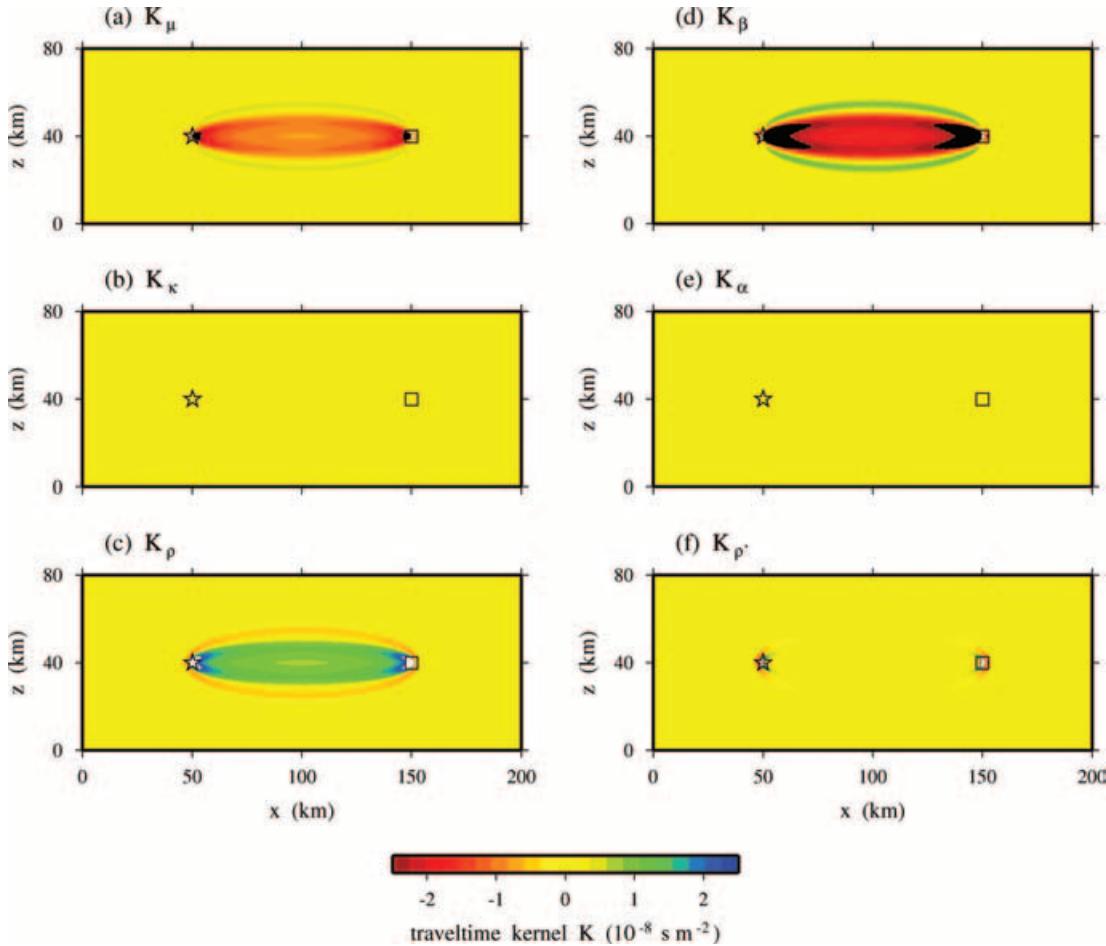


Figure 4. The six SH_S banana-doughnut kernels. Each kernel is constructed simultaneously as shown in Fig. 3 for \bar{K}_β . Note that $\bar{K}_\alpha = \bar{K}_\kappa = 0$, $\bar{K}_\beta = 2\bar{K}_\mu$, $\bar{K}_\mu \approx -\bar{K}_\rho$ and $\bar{K}_{\rho'} \approx 0$ for SH propagation.

kernel increases with increasing frequency. This frequency dependence was illustrated by Hung *et al.* (2000) using a different technique to construct the kernels. Cross-sections of the kernels (Fig. 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. Hung *et al.* 2000).

7.2.2 P-SV waves

The P - SV wavefield is more complicated than the SH wavefield (Fig. 1), and even in the homogeneous case Rayleigh waves arise through interactions at the free surface. Fig. 7 illustrates the construction of the P - SV adjoint source for the $PS+SP$ arrival, and Fig. 8 shows the corresponding formation of the \bar{K}_α kernel. Notice how the interaction between SP and P^\dagger forms the right portion of the sensitivity kernel, e.g. at $t = 32.0$ s (Fig. 8b), whereas the left portion results from the interaction between P and SP^\dagger , e.g. at $t = 16.0$ s (Fig. 8d). This can be deduced simply by matching up the P -wave portions of the regular and adjoint wavefields that are contributing to the interaction field. These interactions ‘paint’ the resultant sensitivity kernel. Fig. 9 shows all six kernels for the P - SV_{PS+SP} scenario.

Fig. 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 Fig. 10(b) we see that the α kernel for P - SV_P is wider than the β kernel for SH_S (Fig. 5c) for the same source period. This is due to the relatively longer wavelengths of the P waves: $\lambda_\alpha = \alpha\tau > \lambda_\beta = \beta\tau$. Figs 10(c)–(f) illustrate examples of α and β kernels for the P - SV_{PP} , P - SV_{PS+SP} and P - SV_{SS} scenarios. Note that, as expected, the α kernel for the SS wave (Fig. 10e) is insignificant relative to the β kernel (Fig. 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_r in (65), to form the amplitude kernels. Fig. 11 shows a comparison of traveltimes, amplitude and attenuation kernels for two examples: SH_S and SH_{SS} . The amplitude kernels are

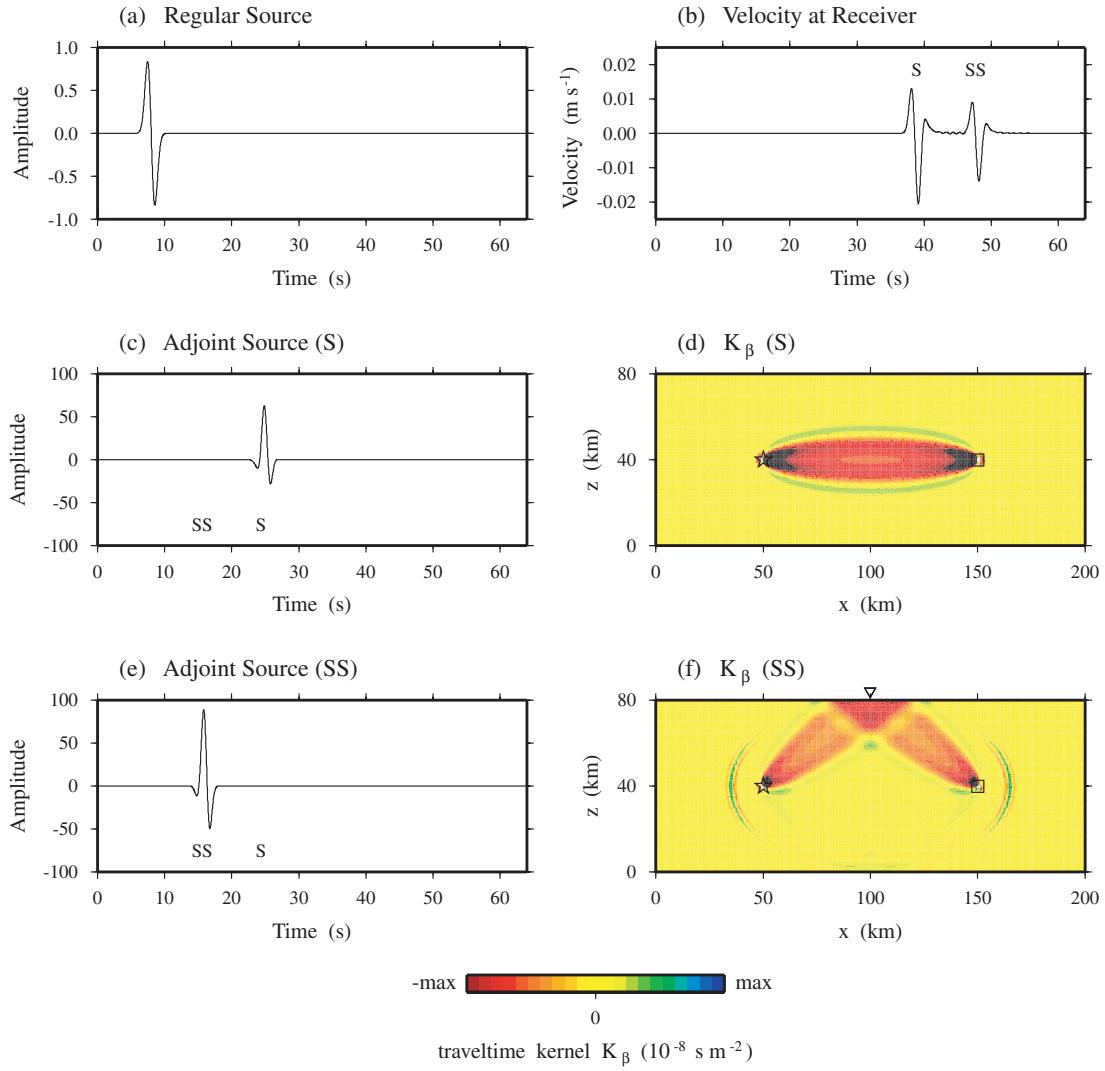


Figure 5. The effect of time window selection on sensitivity kernels, using $\bar{K}_\beta(SH)$ as an example. (a) Source for the regular wavefield. (b) Velocity recorded at the receiver showing the arrivals S and SS . (c) Adjoint source for SH_S , constructed by time-reversing S in (b) and normalizing by N_r defined in (42). (d) \bar{K}_β for reversing S only. (e) Adjoint source for SH_{SS} , constructed by time-reversing SS in (b) and normalizing by N_r . (f) \bar{K}_β for reversing SS only. Each point on the ellipse represents a scattering point for a path with a comparable traveltimeto the SS path. The SS bounce point is labelled by the ∇ , the source by the $*$, and the receiver by the \square .

qualitatively similar to the examples in Dahlen & Baig (2002). (Note that Dahlen & Baig (2002) use slowness perturbations, not wave speed, and that their colour scale is reversed; these two effects ‘cancel’ each other so that their figures are qualitatively similar to ours.) The amplitude sensitivity is negative 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 $\omega_0 = 1.86 \text{ rad s}^{-1}$, which corresponds to a central period of $h(t)$ of $T_0 = 3.37 \text{ s}$. Figs 11(e)–(f) show two attenuation kernels. The attenuation kernels are similar to the travelttime 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 $\mathbf{s}(\mathbf{x}, t)$ due to a finite source represented by a moment-density distribution $\mathbf{m}(\mathbf{x}, t)$ on a fault plane Σ may be written in terms of the Green’s tensor $\mathbf{G}(\mathbf{x}, \mathbf{x}'; t - t')$ as (Aki & Richards 1980; Dahlen & Tromp 1998)

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

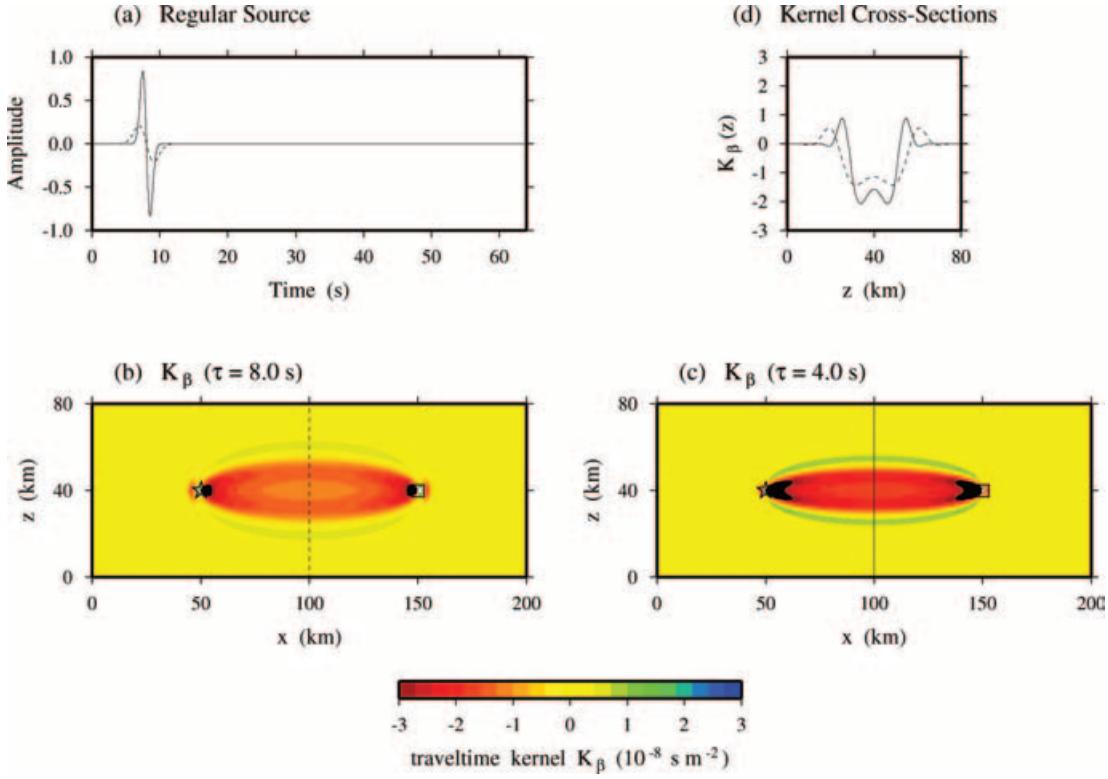


Figure 6. Frequency dependence of the \bar{K}_β sensitivity kernel for SH_S . (a) Two source–time functions for the regular wavefield with durations of $\tau = 8.0$ s (dashed) and $\tau = 4.0$ s (solid). (b) \bar{K}_β for $\tau = 8.0$ s. (c) \bar{K}_β for $\tau = 4.0$ s (see also Fig. 3). (d) Depth cross-sections of (b) and (c) at a horizontal distance of $x = 100$ km. As expected, a higher-frequency kernel is narrower in width and larger in amplitude.

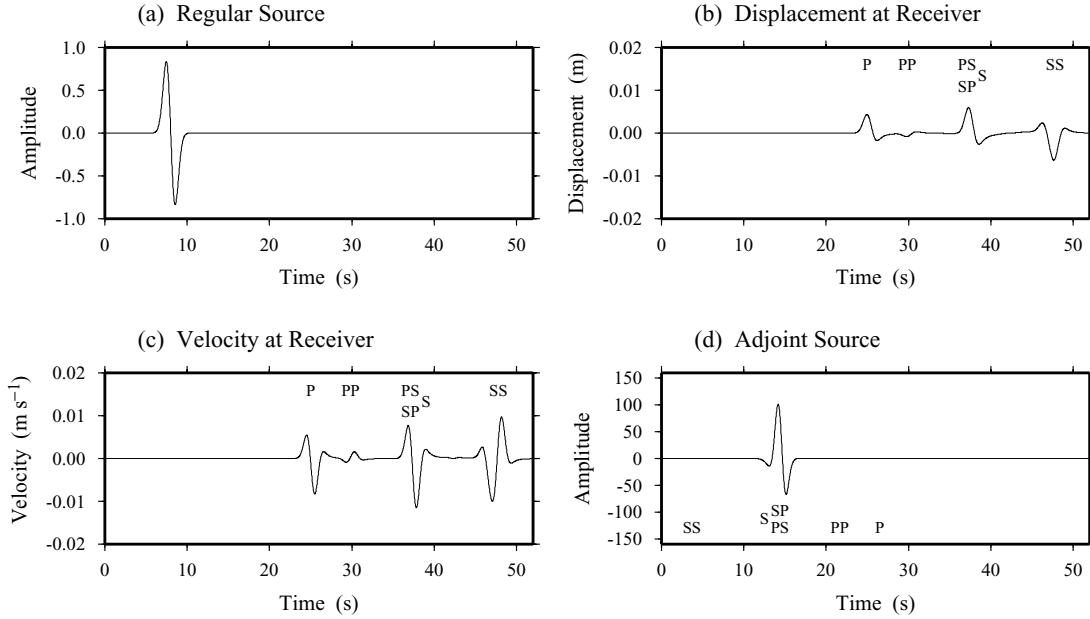


Figure 7. Construction of the adjoint source–time function used in calculating P - SV banana-doughnut kernels. (a) Source–time function responsible for the regular wavefield (x component; the z component is zero). (b) Regular seismogram (x component). (c) Velocity seismogram (x component). (d) Source–time 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 N_r defined by (42). This is the source–time function used in Fig. 8.

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

$$\delta s_i(\mathbf{x}, t) = \int_0^t \int_{\Sigma} \partial'_j G_{ik}(\mathbf{x}, \mathbf{x}'; t - t') \delta m_{jk}(\mathbf{x}', t') d^2 \mathbf{x}' dt'. \quad (97)$$

Upon substitution of the derivatives (97) into (2) we find that the change in the waveform misfit function may be expressed as

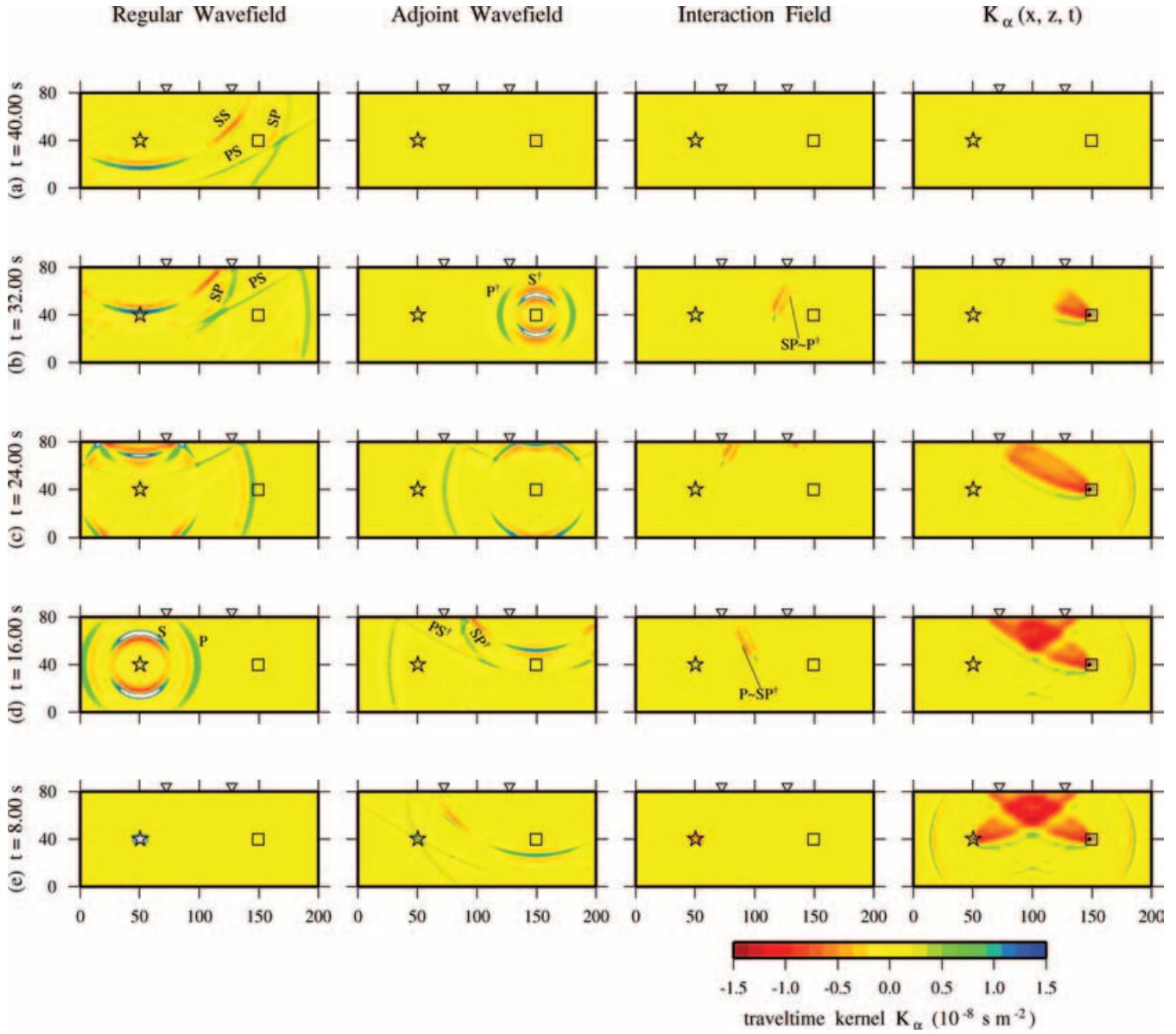


Figure 8. Sequence of interactions between the regular and adjoint P - SV wavefields to produce the banana-doughnut kernel \bar{K}_α . This particular \bar{K}_α is for P - SV_{PS+SP} , i.e. the P - SV α kernel obtained by time-reversing the $PS + SP$ arrival. Given the geometry in Fig. 1, the SP and PS phases arrive simultaneously, at nearly the same time as the S arrival (Fig. 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 Fig. 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 (travelling 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 ∇ for the PS (right) and SP (left) bounce points.

$$\delta\chi = \int_0^T \int_{\Sigma} \epsilon^\dagger(\mathbf{x}, T-t) : \delta\mathbf{m}(\mathbf{x}, t) d^2\mathbf{x} dt, \quad (98)$$

where $\epsilon^\dagger = \frac{1}{2}[\nabla\mathbf{s}^\dagger + (\nabla\mathbf{s}^\dagger)^T]$ denotes the waveform adjoint strain tensor on the finite-fault plane Σ , 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

$$\delta\chi = \int_0^T \epsilon^\dagger(\mathbf{x}_s, T-t) : \delta\mathbf{M}(\mathbf{x}_s, t) dt. \quad (99)$$

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.

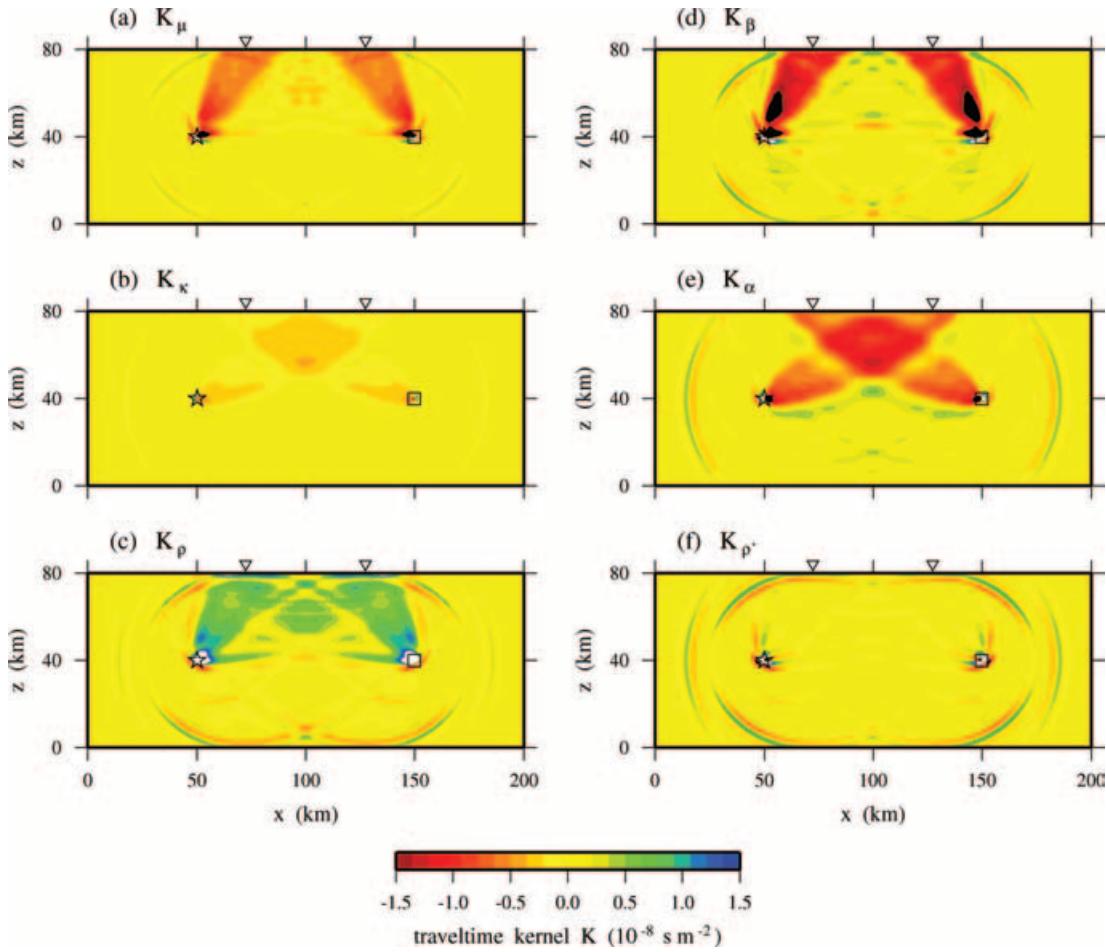


Figure 9. The six $P\text{-}SV_{PS+SP}$ banana-doughnut kernels. Each kernel is constructed simultaneously, as shown in Fig. 8. Notice that the predominant shape of the \bar{K}_α kernel is that of two truncated ‘cigars’, the right one for PS and the left for SP . The labels ∇ denotes the PS (right) and SP (left) bounce points.

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

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

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

$$\delta\chi = \int_0^T h(t)\mathbf{s}^\dagger(\mathbf{x}_s, T-t) \cdot \delta\mathbf{v} dt. \quad (101)$$

Using this result, together with a conjugate-gradient algorithm (Fletcher & 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 \mathbf{v}^0 , calculate $\chi(\mathbf{v}^0)$, $\mathbf{g}^0 = \partial\chi/\partial\mathbf{v}(\mathbf{v}^0)$, and set the initial conjugate-gradient search direction equal to minus the initial gradient of the misfit function, $\mathbf{p}^0 = -\mathbf{g}^0$. If $\|\mathbf{p}^0\| < \epsilon$, where ϵ is a suitably small number, then \mathbf{v}^0 is the vector we seek to determine, otherwise:

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

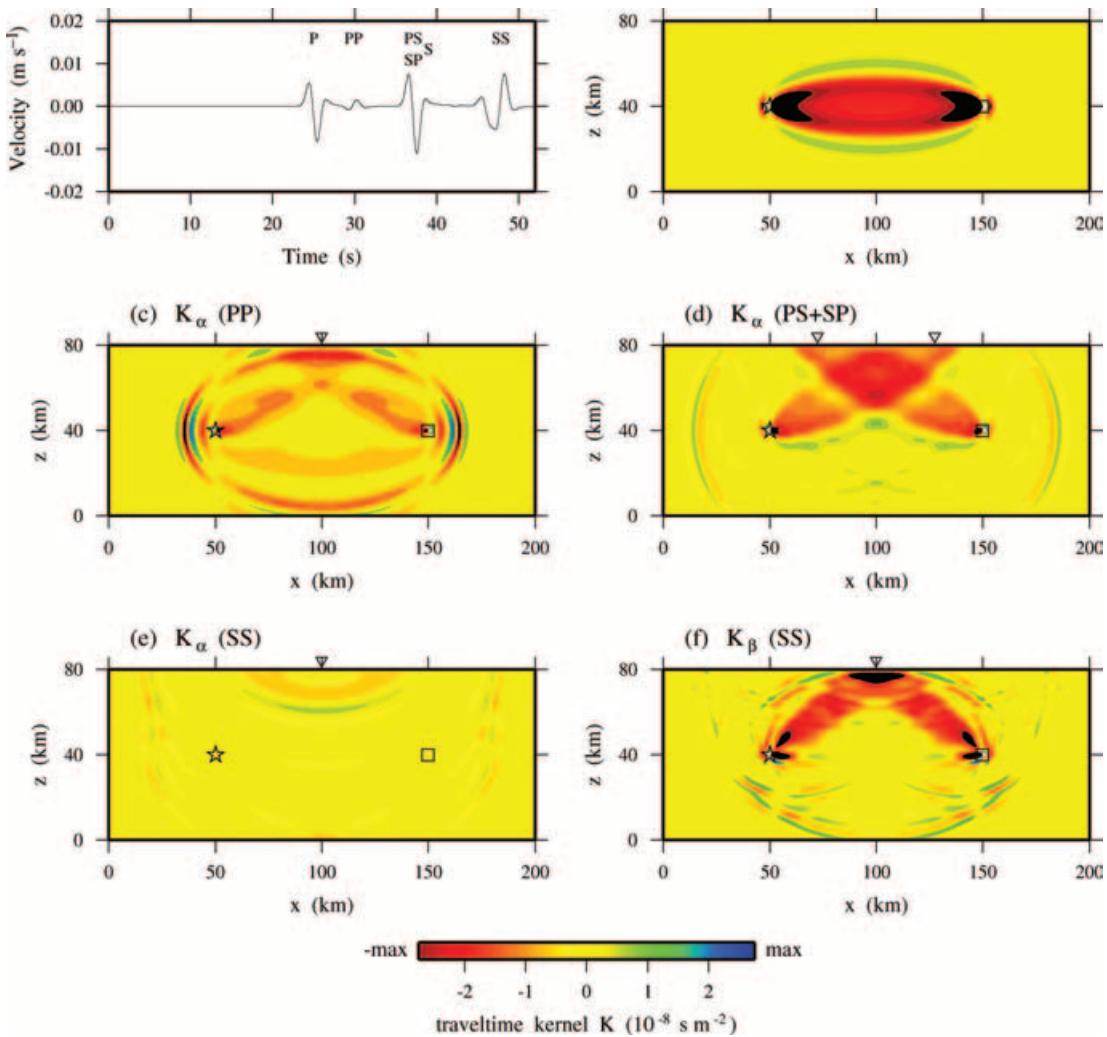


Figure 10. The effect of time-window selection on sensitivity kernels, using \tilde{K}_α as an example—see Fig. 1 for labelling and Section 7.2 for details. (a) Velocity recorded at the receiver (x component) showing the consecutive arrivals of P , PP , PS and SS . The S phase is expected to arrive nearly 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 Fig. 7. The colour scale varies for each plot according to the value ‘max’. (b) \tilde{K}_α for reversing the P arrival ($P-SV_P$, max = 1.0). (c) \tilde{K}_α for reversing the PP arrival ($P-SV_{PP}$, max = 5.0). (d) \tilde{K}_α for reversing the $PS + SP$ arrival ($P-SV_{PS+SP}$, max = 1.5). (e) \tilde{K}_α for reversing the SS arrival ($P-SV_{SS}$, max = 2.75). (f) \tilde{K}_β for reversing the SS arrival ($P-SV_{SS}$, max = 2.75).

We apply this source inversion procedure to a synthetic problem based upon the waveform misfit function (1). The 2-D model properties are the same as in Section 7.1, and the source–receiver geometry is shown in Fig. 12(a). 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 \mathbf{v} denoted by the triangle in Fig. 12(b). Fig. 12(b) shows the inverted vector \mathbf{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. Fig. 12(c) shows that the value of the normalized misfit function $\chi(\mathbf{v}^k)$ 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, totalling $4n_{\text{iterations}}$ 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 $\delta \mathbf{s}$ due to structural (3), topographic (21) and source parameters (97). The gradient of the waveform misfit

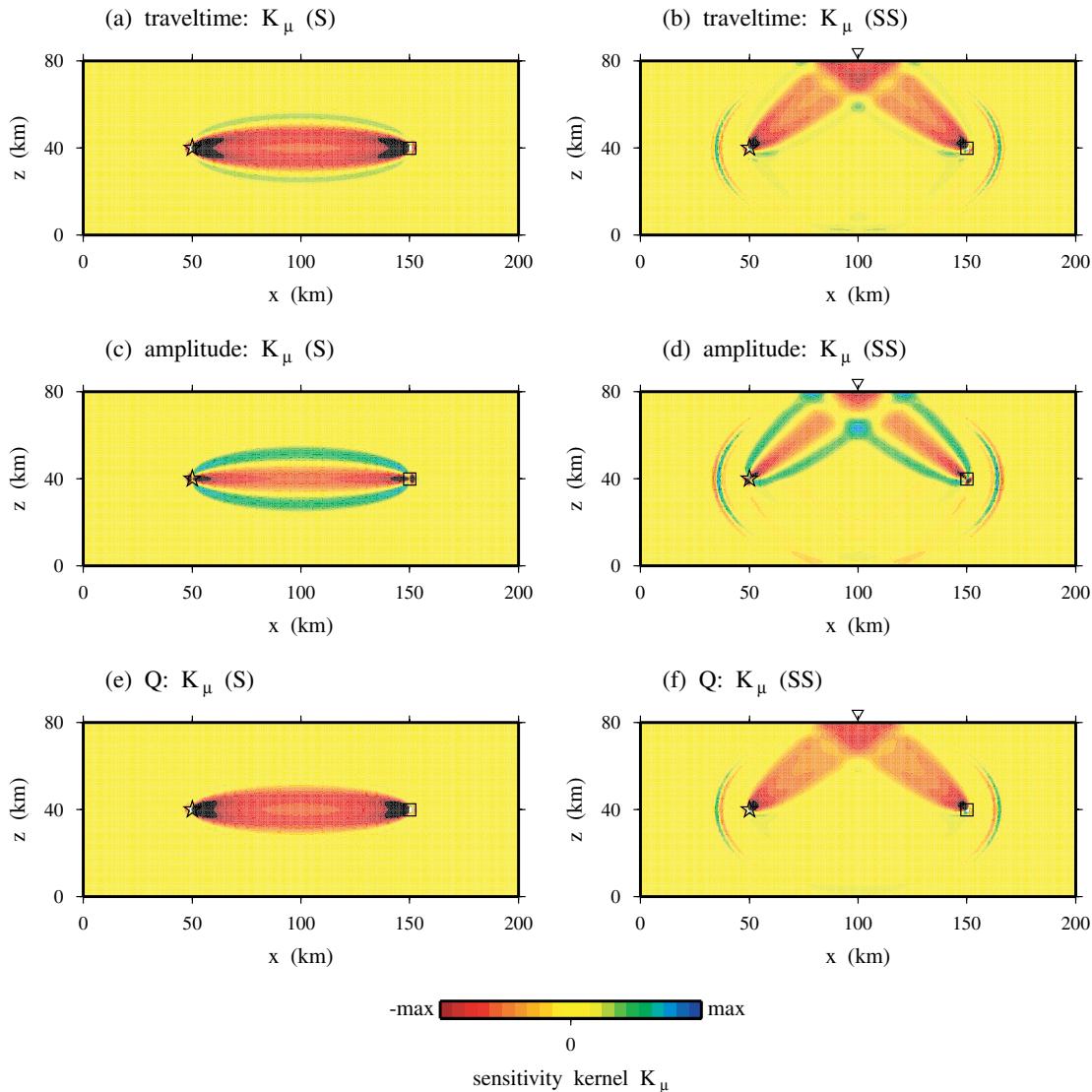


Figure 11. Comparison of $SH \bar{K}_\mu$ kernels for traveltimes, amplitude and attenuation for the S and SS arrivals—see Fig. 1 for labelling and Section 7.3 for details. The colour scale varies for each pair of plots according to the value ‘max’. (a, b) Traveltime kernels \bar{K}_μ for reversing the S arrival (a) ($\text{max} = 1.4 \times 10^{-8} \text{ s m}^{-2}$) and the SS arrival (b) ($\text{max} = 2.6 \times 10^{-8} \text{ s m}^{-2}$). (c, d) Amplitude kernels \bar{K}_μ for reversing the S arrival (c) ($\text{max} = 2.0 \times 10^{-8} \text{ m}^{-2}$) and the SS arrival (d) ($\text{max} = 3.2 \times 10^{-8} \text{ m}^{-2}$). (e, f) Q kernels \bar{K}_μ for reversing the S arrival (e) ($\text{max} = 0.75 \times 10^{-4} \text{ m}^{-2}$) and the SS arrival (f) ($\text{max} = 1.5 \times 10^{-4} \text{ m}^{-2}$).

function may thus be rewritten in the form

$$\begin{aligned} \delta\chi = & \int_V [K'_\rho(\mathbf{x})\delta \ln \rho(\mathbf{x}) + K_\beta(\mathbf{x})\delta \ln \beta(\mathbf{x}) + K_\alpha(\mathbf{x})\delta \ln \alpha(\mathbf{x})] d^3\mathbf{x} \\ & + \int_{\Sigma} K_h(\mathbf{x})\delta h(\mathbf{x}) d^2\mathbf{x} + \int_{\Sigma_{FS}} \mathbf{K}_h(\mathbf{x}) \cdot \nabla^\Sigma \delta h(\mathbf{x}) d^2\mathbf{x} + \int_0^T \int_{\Sigma} \epsilon^\dagger(\mathbf{x}, T-t) : \delta \mathbf{m}(\mathbf{x}, t) d^2\mathbf{x} dt, \end{aligned} \quad (102)$$

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 3-D 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 parametrization by selecting basis functions that result in large gradients of the misfit function.

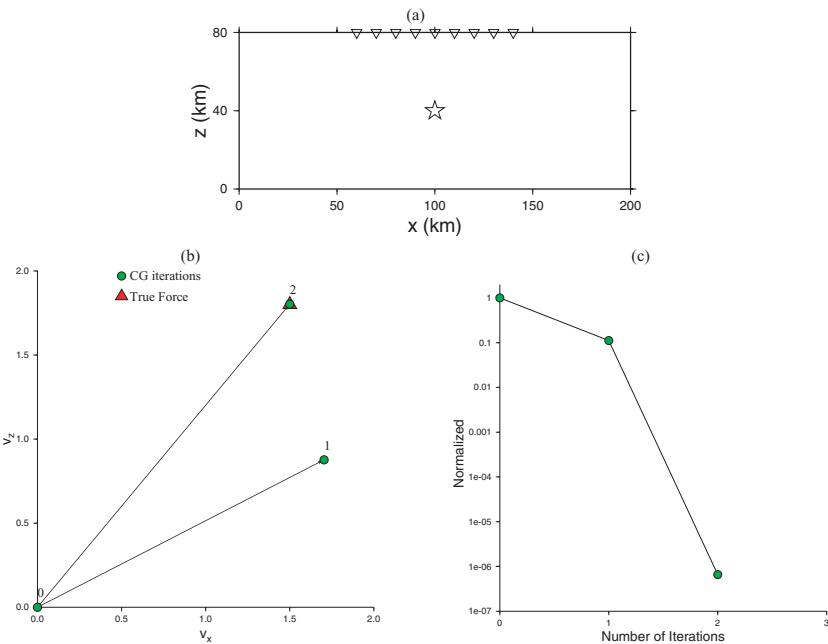


Figure 12. (a) Source–receiver geometry for the point force inversion. The dimensions of the model are the same as in Section 7.1. (b) Results of the source inversion based upon the procedures outlined in Section 8.1. Circles indicate consecutive locations of the force vector in the x – z plane, with 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 χ with the number of iterations (χ has been normalized by $\frac{1}{2} \sum_{r=1}^N \int_0^T \| \mathbf{d}(\mathbf{x}_r, t) \|^2 dt$).

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 & 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 $2n_{\text{events}}$ 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 $n_{\text{iterations}}$ denote the number of iterations used in the inverse problem. Then the total number of forward calculations involved in a complete inversion is $4n_{\text{events}}n_{\text{iterations}}$. Considering the fact that on modern parallel computers regional simulations may be performed in less than 1 hr 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 3-D tomographic inversions should be feasible on the largest parallel machines, e.g. the Earth Simulator (www.es.jamstec.go.jp).

ACKNOWLEDGEMENTS

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

- Akçelik, V., Biros, G. & Ghattas, O., 2002. Parallel multiscale Gauss–Newton–Krylov methods for inverse wave propagation, *Proceedings of the ACM/IEEE Supercomputing SC’2002 Conference*, published on CD-ROM and at <http://www.sc-conference.org/sc2002/>
- Akçelik, V. *et al.*, 2003. High resolution forward and inverse earthquake modeling on terascale computers, *Proceedings of the ACM/IEEE Supercomputing SC’2003 Conference*, published on CD-ROM and at <http://www.sc-conference.org/sc2003/paperpdfs/pap298.pdf>
- Aki, K. & Richards, P.G., 1980. *Quantitative Seismology, Theory and Methods*, W. H. Freeman, San Francisco, CA.
- Bunge, H.-P., Hagelberg, C.R. & Travis, B.J., 2003. Mantle circulation models with variational data assimilation: inferring past mantle flow and structure from plate motion histories and seismic tomography, *Geophys. J. Int.*, **152**, 280–301.
- Capdeville, Y., Chaljub, E., Villette, J.P. & Montagner, J.P., 2003. Coupling the spectral element method with a modal solution for elastic wave propagation in global Earth models, *Geophys. J. Int.*, **152**, 34–67.
- Chen, P., Jordan, T. & Zhao, L., 2004. Generalized Seismological Data Functionals (GSDF) for broadband seismic waveform analysis and inversion, *Geophys. J. Int.*, in preparation.
- Crase, E., Pica, A., Noble, M., McDonald, J. & Tarantola, A., 1990. Robust elastic non-linear waveform inversion: application to real data, *Geophysics*, **55**, 527–538.
- Dahlen, F.A., 2004. Finite-frequency sensitivity kernels for boundary topography perturbations, *Geophys. J. Int.*, in preparation.

- Dahlen, F. & Baig, A.M., 2002. Fréchet kernels for body wave amplitudes, *Geophys. J. Int.*, **150**, 440–466.
- Dahlen, F.A. & Tromp, J., 1998. *Theoretical Global Seismology*, Princeton University Press, Princeton, NJ.
- Dahlen, F.A., Nolet, G. & Hung, S.-H., 2000. Fréchet kernels for finite-frequency traveltimes—I. Theory, *Geophys. J. Int.*, **141**, 157–174.
- Fink, M., 1992. Time reversal of ultrasonic field—Part I: Basic principles, *IEEE Trans. Ultras. Ferroelectr. Freq. Contr.*, **39**, 555–566.
- Fink, M., 1997. Time reversed acoustics, *Phys. Today*, **50**, 34–40.
- Fink, M., Prada, C., Wu, F. & Cassereau, D., 1989. Self focusing in inhomogeneous media with time reversal acoustic mirrors, *IEEE Ultrason. Symp. Proc.*, **1**, 681–686.
- Fletcher, R. & Reeves, C., 1964. Function minimization by conjugate gradients, *Comput. J.*, **7**, 149–154.
- Gauthier, O., Virieux, J. & Tarantola, A., 1986. Two-dimensional non-linear inversion of seismic waveforms: numerical results, *Geophysics*, **51**, 1387–1403.
- Gee, L. & Jordan, T., 1992. Generalized seismological data functionals, *Geophys. J. Int.*, **111**, 363–390.
- Hudson, J.A., 1977. Scattered waves in the coda of P, *J. Geophys.*, **43**, 359–374.
- Hung, S.-H., Dahlen, F.A. & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes—II. Examples, *Geophys. J. Int.*, **141**, 175–203.
- Komatitsch, D. & Tromp, J., 1999. Introduction to the spectral-element method for 3-D seismic wave propagation, *Geophys. J. Int.*, **139**, 806–822.
- Komatitsch, D. & Tromp, J., 2002a. Spectral-element simulations of global seismic wave propagation—I. Validation, *Geophys. J. Int.*, **149**, 390–412.
- Komatitsch, D. & Tromp, J., 2002b. Spectral-element simulations of global seismic wave propagation—II. 3-D models, oceans, rotation, and self-gravitation, *Geophys. J. Int.*, **150**, 303–318.
- Komatitsch, D., Ritsema, J. & Tromp, J., 2002. The spectral-element method, Beowulf computing, and global seismology, *Science*, **298**, 1737–1742.
- Komatitsch, D., Liu, Q., Tromp, J., Süss, P., Stidham, C. & Shaw, J.H., 2004. Simulations of strong ground motion in the Los Angeles Basin based upon the spectral-element method, *Bull. seism. Soc. Am.*, **94**, 187–206.
- Liu, H.P., Anderson, D.L. & Kanamori, H., 1976. Velocity dispersion due to anelasticity: implications for seismology and mantle composition, *Geophys. J. R. astr. Soc.*, **47**, 41–58.
- Luo, Y. & Schuster, G.T., 1991. Wave-equation traveltime tomography, *Geophysics*, **56**, 645–653.
- Marquering, H., Dahlen, F. & Nolet, G., 1999. Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox, *Geophys. J. Int.*, **137**, 805–815.
- Montelli, R., Nolet, G., Dahlen, F., Masters, G., Engdahl, E.R. & Hung, S.-H., 2004. Finite-frequency tomography reveals a variety of plumes in the mantle, *Science*, **303**, 338–343.
- Mora, P., 1987. Nonlinear two-dimensional elastic inversion of multioffset seismic data, *Geophysics*, **52**, 1211–1228.
- Mora, P., 1988. Elastic wave-field inversion of reflection and transmission data, *Geophysics*, **53**, 750–759.
- Nolet, G., 1987. Waveform tomography, in *Seismic Tomography, with Applications in Global Seismology and Exploration Geophysics*, pp. 301–322, ed. Nolet, G., Reidel, Dordrecht.
- Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model, *Geophysics*, **64**, 888–901.
- Press, W., Teukolsky, S., Vetterling, W. & Flannery, B., 1994. *Numerical Recipes: the Art of Scientific Computing*, Cambridge University Press, Cambridge.
- Ritsema, J., Rivera, L.A., Komatitsch, D., Tromp, J. & van Heijst, H.J., 2002. The effects of crust and mantle heterogeneity on PP/P and SS/S amplitude ratios, *Geophys. Res. Lett.*, **29**, doi:10.1029/2001GL013831.
- Talagrand, O. & Courtier, P., 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Q. J. R. Meteorol. Soc.*, **113**, 1311–1328.
- Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**, 1259–1266.
- Tarantola, A., 1987. Inversion of travel times and seismic waveforms, in *Seismic Tomography*, pp. 135–157, ed. Nolet, G., Reidel, Dordrecht.
- Tarantola, A., 1988. Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation, *Pure appl. Geophys.*, **128**, 365–399.
- Tsuboi, S., Komatitsch, D., Chen, J. & Tromp, J., 2003. Spectral-element simulations of the November 3, 2002, Denali, Alaska earthquake on the Earth Simulator, *Phys. Earth planet. Inter.*, **139**, 305–312.
- Wu, R. & Aki, K., 1985. Scattering characteristics of elastic waves by an elastic heterogeneity, *Geophysics*, **50**, 582–595.
- Zhao, L., Jordan, T.H. & Chapman, C.H., 2000. Three-dimensional Fréchet differential kernels for seismic delay times, *Geophys. J. Int.*, **141**, 558–576.