Geophys. J. Int. (2007) **170,** 175–181 doi: 10.1111/j.1365-246X.2007.03429.x

FAST TRACK PAPER

Full three-dimensional tomography: a comparison between the scattering-integral and adjoint-wavefield methods

Po Chen,[1] Thomas H. Jordan[2] and Li Zhao[2,3]

[1] *Lamont-Doherty Earth Observatory, Columbia University, USA. E-mail: pochen@ldeo.columbia.edu*
[2] *Department of Earth Sciences, University of Southern California, USA*
[3] *Institute of Earth Sciences, Academia Sinica, Taipei* 115*, Taiwan*

Accepted 2007 February 26. Received 2007 February 26; in original form 2006 July 17

SUMMARY

This paper analyses the computational issues of full 3-D tomography, in which the starting model as well as the model perturbation is 3-D and the sensitivity (Fréchet) kernels are calculated using the full physics of 3-D wave propagation. We compare two formulations of the structural inverse problem: the adjoint-wavefield (AW) method, which back-propagates the data from the receivers to image structure, and the scattering-integral (SI) method, which sets up the inverse problem by calculating and storing the Fréchet kernels for each data functional. The two inverse methods are closely related, but which one is more efficient depends on the overall problem geometry, particularly on the ratio of sources to receivers, as well as trade-offs in computational resources, such as the relative costs of compute cycles to data storage. We find that the SI method is computationally more efficient than the AW method in regional waveform tomography using large sets of natural sources, although it requires more storage.

Key words: adjoint wavefield, 3-D Earth structure, receiver Green tensor, scattering integral, tomography, waveform inversion.

1 INTRODUCTION

In seismic tomography, the data in a vector $\mathbf{d}$ are represented as functionals of the earth model $\mathbf{m}$, which specifies the wave velocities, mass density, and other seismic properties as functions of position $\mathbf{x}$ throughout the model volume V. To set up the inverse problem, the data functionals are calculated for a starting model $\tilde{\mathbf{m}}$, and the forward problem is linearized:

$$\delta\mathbf{d} = \mathbf{A}\delta\mathbf{m} = \int dV(\mathbf{x})\,\mathbf{K}_d(\tilde{\mathbf{m}}, \mathbf{x}) \cdot \delta\mathbf{m}(\mathbf{x}) \approx \mathbf{d}(\mathbf{m}) - \mathbf{d}(\tilde{\mathbf{m}}). \qquad (1)$$

The data sensitivity kernels in $\mathbf{K}_d$ are the functional (Fréchet) derivatives of the data with respect to the model parameters around the starting model. The tomographic inverse problem is to estimate the best representation of the real Earth from observations of $\mathbf{d}$. The approximations applied in tomographic methods thus involve the dimensionalities of the starting model $\tilde{\mathbf{m}}$, the model perturbation $\delta\mathbf{m} = \mathbf{m} - \tilde{\mathbf{m}}$, and the Fréchet kernels $\mathbf{K}_d$.

Early tomographic studies built three-dimensional (3-D) model perturbations on top of 1-D starting models and, in constructing the Fréchet kernels, typically applied geometrical ray theory or 1-D approximations, such as path-averaging and local phase-velocity methods (e.g. Woodhouse & Dziewonski 1984; Li & Tanimoto 1993; Li & Romanowicz 1995; Ritsema & van Heijst 2000). Over the last decade, seismologists have improved the accuracy and spatial resolution of tomographic methods by calculating the Fréchet kernels for finite-bandwidth signals using asymptotic ray theory (Marquering *et al.* 1999; Dahlen *et al.* 2000) and normal-mode summation (Katzman *et al.* 1998; Zhao & Jordan 1998). The latter provides an essentially exact algorithm for computing 3-D Fréchet kernels for a general class of band-limited data functionals (Zhao *et al.* 2000), although it still relies on a 1-D starting model.

Advances in parallel computing technology and numerical methods (Olsen 1994; Graves 1996; Akcelik *et al*. 2003; Olsen *et al.* 2003; Komatitsch *et al.* 2004) have made large-scale 3-D numerical simulations of seismic wavefields much more affordable, and they open up the possibility of 'full 3-D tomography', in which $\tilde{\mathbf{m}}$ as well as $\delta\mathbf{m}$ is 3-D and $\mathbf{K}_d$ is calculated using the full physics of 3-D wave propagation. Full 3-D tomography can account for the non-linearity of the structural inverse problem through iteration, thus providing the most efficient means for assimilating seismic observations into dynamic ground-motion models. This type of tomography was formulated as an adjoint problem in space–time domain in the 1980s (Tarantola 1984, 1988) and in space–frequency domain in the 1990s (Pratt 1990; Pratt *et al*. 1998). It has been applied in the seismic exploration industry for more than a decade (Biondi 2006); however, its adaptation to large-scale structural studies that use earthquake sources remains a work in progress (Tromp *et al.* 2005). We have

recently demonstrated the ability to compute the Fréchet kernels for generalized seismological data functionals at regional distances using a 3-D starting model (Zhao *et al.* 2005), and we have applied this capability to improving existing 3-D models of crustal structure in the Los Angeles region by full 3-D tomography (Chen *et al.* 2007).

In this paper, we compare two formulations of the structural inverse problem: the adjoint-wavefield (AW) method, which back-propagates the data from the receivers to image structure (Ackelik *et al*. 2002; Tromp *et al*. 2005; Liu & Tromp 2006), and the scattering-integral (SI) method, which sets up the inverse problem by calculating and storing the Fréchet kernels for each data functional (Zhao *et al.* 2005; Chen *et al.* 2007). The two inverse methods are closely related (Tromp *et al*. 2005), but which one is more efficient depends on the overall problem geometry, particularly on the ratio of sources to receivers, as well as trade-offs in computational resources, such as the relative costs of compute cycles to data storage.

2 DATA FUNCTIONALS

We presume the seismogram observed on the ith component of the rth receiver from the sth seismic source can be approximated by an instrument-filtered wavefield $u_i^s(\mathbf{x}_r, t)$ calculated from an unknown (target) earth model $\mathbf{m}$. The quality of the approximation will depend on the ambient seismic noise as well as inadequacies in the way $\mathbf{m}$ represents the actual Earth ('signal-generated noise'). For each seismogram, we consider a finite set of data functionals, indexed by n, that measure the misfit between $u_i^s(\mathbf{x}_r, t)$ and the wavefield $\tilde{u}_i^s(\mathbf{x}_r, t)$ synthesized from the starting model $\tilde{\mathbf{m}}$:

$$d_{in}^{sr} = D_n\big[u_i^s(\mathbf{x}_r, t),\ \tilde{u}_i^s(\mathbf{x}_r, t)\big]. \tag{2}$$

The measurement process generally involves non-linear operations on both the observed and synthetic seismograms. We assume it is constructed to satisfy $D_n[\tilde{u}_i^s(\mathbf{x}_r, t),\ \tilde{u}_i^s(\mathbf{x}_r, t)] = 0$, so that the term $\mathbf{d}(\tilde{\mathbf{m}})$ in eq. (1) is zero. Because these differential data functionals are *seismogram-specific*, their Fréchet derivatives with respect to $\tilde{u}_i^s(\mathbf{x}_r, t)$ can be expressed in terms of space-independent kernels,

$$\delta d_{in}^{sr} = \int dt\ J_{in}^{sr}(t)\delta u_i^s(\mathbf{x}_r, t). \tag{3}$$

We consider a starting model $\tilde{\mathbf{m}}$ that comprises estimates of the mass density $\tilde{\rho}(\mathbf{x})$ and elastic moduli $\tilde{c}_{jklm}(\mathbf{x})$ and perturbations that are real and time-independent,

$$\delta\,\mathbf{m}(\mathbf{x}) = \begin{bmatrix} \delta\rho(\mathbf{x}) \\ \delta c_{jklm}(\mathbf{x}) \end{bmatrix}. \tag{4}$$

This model perturbation can be related to seismogram perturbation using the first-order Born approximation (Dahlen & Tromp 1998; Zhao *et al.* 2000):

$$\delta u_i^s(\mathbf{x}_r, t) = -\int dV(\mathbf{x}) \int d\tau \sum_j \Bigg[G_{ij}(\mathbf{x}_r, t-\tau; \mathbf{x})\, \partial_\tau^2 u_j^s(\mathbf{x}, \tau)\, \delta\rho(\mathbf{x}) + \sum_{jklm} \partial_k G_{ij}(\mathbf{x}_r, t-\tau; \mathbf{x})\ \partial_l\, u_m^s(\mathbf{x}, \tau)\, \delta c_{jklm}(\mathbf{x}) \Bigg]. \tag{5}$$

Here, ∂_k is the partial derivative with respect to x_k and ∂_τ is the partial derivative with respect to τ. Substituting (5) into (3) yields the Fréchet derivative of d_{in}^{sr} with respect to $\tilde{\mathbf{m}}$. The resulting kernel is exact (Zhao *et al.* 2000), but the quality with which d_{in}^{sr} can be approximated by a linear perturbation δd_{in}^{sr} from $\tilde{\mathbf{m}}$ depends on the type of data functional. We consider four examples.

1. Perhaps the most obvious data functional for waveform inversion is the difference between the target and starting seismograms at time t_n (Nolet 1987; Tarantola 1988; Ackelik *et al*. 2002):

$$d_{in}^{sr} = u_i^s(\mathbf{x}_r, t_n) - \tilde{u}_i^s(\mathbf{x}_r, t_n). \tag{6}$$

In this case, the seismogram perturbation kernel $J_{in}^{sr}(t)$ is just the Dirac delta function $\delta(t - t_n)$, and the model perturbation kernel $\nabla_{\tilde{\mathbf{m}}} d_{in}^{sr}$ is given directly by the Born eq. (5). However, the Born approximation $u_i^s(\mathbf{x}_r, t_n) \approx \tilde{u}_i^s(\mathbf{x}_r, t_n) + \delta u_i^s(\mathbf{x}_r, t_n)$ is of limited utility in tomography, because it effectively involves the linearization of the Fourier shift operator, $e^{i\omega t}$; that is, for an observed waveform with centroid frequency ω_0, it is valid only when $\omega_0\, \Delta T_{in}^{sr} \ll 1$, where ΔT_{in}^{sr} is the traveltime shift relative to the synthetic.

2. The traveltime shift $d_{in}^{sr} = \Delta T_{in}^{sr}$ of an isolated waveform in a time window t_n, t_n' can be estimated by maximizing the cross-correlation between the observed and synthetic waveforms (e.g. Woodward & Masters 1991). For band-limited signals, the cross-correlation traveltime anomaly is usually a good approximation to the phase delay at the dominant frequency (Gee & Jordan 1992). The perturbation kernel is proportional to the waveform derivative (Luo & Shuster 1991; Zhao & Jordan 1998),

$$J_{in}^{sr}(t) = -\frac{\partial_t u_i^s(\mathbf{x}_r, t)[H(t-t_n) - H(t-t_n')]}{\int_{t_n}^{t_n'} \left|\partial_t u_i^s(\mathbf{x}_r, t)\right|^2 dt}, \tag{7}$$

where $H(t)$ is the Heaviside function. Approximating the Born perturbation in terms of asymptotic ray theory yields the 'banana–doughnut' kernels for ΔT_{in}^{sr} described by Marquering *et al.* (1999) and Dahlen *et al.* (2000). Fréchet kernels for ΔT_{in}^{sr} that fully account for wavefield interactions in a 1-D starting model can be computed by expressing the Born perturbation as a coupled-mode summation (Zhao *et al.* 2000).

3. The amplitude anomaly ΔU_{in}^{sr} of an isolated waveform in a time window $[t_n,\ t_n']$ can also be estimated by waveform cross-correlation (Ritsema *et al.* 2002; Tromp *et al.* 2005). The perturbation kernel for the relative amplitude anomaly, $d_{in}^{sr} = \Delta(\ln U_{in}^{sr}) \approx \Delta U_{in}^{sr}/\tilde{U}_{in}^{sr}$, is proportional to the waveform (Dahlen & Baig 2002),

$$J_{in}^{sr}(t) = \frac{u_i^s(\mathbf{x}_r, t)[H(t-t_n) - H(t-t_n')]}{\int_{t_n}^{t_n'} \left[u_i^s(x_r, t)\right]^2 dt}. \tag{8}$$

4. The cross-correlation traveltime shift and relative amplitude anomaly can be generalized to frequency-dependent quantities that completely describe the discrepancies between the target and synthetic waveforms. In the frequency domain, the synthetic can be mapped into the target by an exponential (Rytov) operator, $e^{i\omega[\Delta\tau_p(\omega) + i\,\Delta\tau_q(\omega)]}$. The *generalized seismological data functionals* (GSDFs) are observational approximations to the phase-delay times $\Delta\tau_p(\omega)$ and amplitude-reduction times $\Delta\tau_q(\omega)$, obtained by time-windowing and narrow-band filtering of the waveform cross-correlagram (Gee & Jordan 1992). The former is the frequency-dependent traveltime shift,

$$d_{in}^{sr} = \Delta\, T_{in}^{sr}(\omega_n), \tag{9}$$

and the latter is proportional to the logarithmic amplitude anomaly,

$$d_{in}^{sr} = -\frac{1}{\omega_n}\Delta\big[\ln U_{in}^{sr}(\omega_n)\big]. \tag{10}$$

The minus sign appears because a positive value of $\Delta\tau_q$ corresponds to an amplitude decrease. The integration kernel $J_{in}^{sr}(t)$ for either type of GSDF measurement can be written as the convolution of the

synthetic waveform with a five-parameter Gaussian wavelet that has an effective centroid frequency ω_n and bandwidth σ_n dependent on the windowing and filtering parameters (Chen 2005).

GSDFs are well suited for the tomographic inverse problem. In particular, their linearization depends on the Rytov approximation, which is valid for large phase-shifts as long as the phase perturbation *per wavelength* is small (Chernov 1960; Snieder & Lomax 1996). This is far less restrictive than the Born approximation, which requires small phase-shifts. A practical aspect of the GSDF method is that the phase-delay measurements need to be properly unwrapped, which can usually be achieved through bootstrapping from low to high frequencies.

3 MISFIT FUNCTIONAL AND ITS FRÉCHET DERIVATIVE

The tomographic methods discussed here are formulated as optimizations of the quadratic misfit functional,

$$\chi^2(\mathbf{m}, \tilde{\mathbf{m}}) = \frac{1}{2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \sum_{i,n} \left| d_{in}^{sr}(\mathbf{m}, \tilde{\mathbf{m}}) \right|^2. \tag{11}$$

We explicitly sum over the N_s sources and N_r receivers that define the geometry of the inverse problem; i and n range across instrumental components and seismogram-specific measurements used in the data vector $\mathbf{d}$. The data-space inner product can be weighted by the data covariance operator $\mathbf{C_d}$, which we omit for notational simplicity.

The perturbation of the misfit functional with respect to $\tilde{\mathbf{m}}$ can be written as

$$\begin{aligned} \delta \chi^2(\mathbf{m}, \tilde{\mathbf{m}}) &= \sum_{srin} d_{in}^{sr} \delta d_{in}^{sr} = \sum_{srin} d_{in}^{sr} \int dt\, J_{in}^{sr}(t)\, \delta u_i^s(\mathbf{x}_r, t) \\ &= \int dV(\mathbf{x}) K_{\chi^2}^{\rho}(\mathbf{x}) \delta \rho(\mathbf{x}) \\ &\quad + \sum_{jklm} \int dV(\mathbf{x}) K_{\chi^2}^{c_{jklm}}(\mathbf{x}) \delta c_{jklm}(\mathbf{x}), \end{aligned} \tag{12}$$

where the Fréchet kernels are

$$\begin{aligned} K_{\chi^2}^{\rho}(\mathbf{x}) &= -\sum_{srin} \int dt \int d\tau\, d_{in}^{sr}\, J_{in}^{sr}(t) \\ &\quad \times \sum_j G_{ji}(\mathbf{x}, t-\tau; \mathbf{x}_r)\, \partial_\tau^2 u_j^s(\mathbf{x}, \tau), \end{aligned} \tag{13}$$

$$K_{\chi^2}^{c_{jklm}}(\mathbf{x}) = -\sum_{srin} \int dt \int d\tau\, d_{in}^{sr} J_{in}^{sr}(t) \partial_k G_{ji}(\mathbf{x}, t-\tau; \mathbf{x}_r)\, \partial_l u_m^s(\mathbf{x}, \tau). \tag{14}$$

In these expressions we have applied the seismic reciprocity principle (Aki & Richards 1980),

$$G_{ij}(\mathbf{x}_r, t-\tau; \mathbf{x}) = G_{ji}(\mathbf{x}, t-\tau; \mathbf{x}_r). \tag{15}$$

The kernels of the misfit functional (13) and (14) are just the data-weighted sum of the Fréchet kernels in eq. (1); that is, $\delta\chi^2 = \mathbf{d}^{\mathrm{T}} \mathbf{A}\, \delta\mathbf{m}$.

4 ADJOINT-WAVEFIELD METHOD I: CONJUGATE-GRADIENT IMPLEMENTATION

One implementation of the AW method minimizes the misfit functional using the kernels $K_{\chi^2}^{\rho}(\mathbf{x})$ and $K_{\chi^2}^{c_{jklm}}(\mathbf{x})$, which give the descent directions in a gradient-based optimization algorithm such as the conjugate-gradient method (Tromp *et al.* 2005). We call this type of implementation 'AW–CG'. If we reverse time by making the substitution $t \to -t$, eqs (13) and (14) can be expressed in terms of the source wavefield and an adjoint wavefield:

$$K_{\chi^2}^{\rho}(\mathbf{x}) = -\sum_s \int d\tau \sum_j \partial_\tau^2 u_j^s(\mathbf{x}, \tau) \left[u_j^s\right]^+(\mathbf{x}, -\tau), \tag{16}$$

$$K_{\chi^2}^{c_{jklm}}(\mathbf{x}) = -\sum_s \int d\tau\, \partial_l u_m^s(\mathbf{x}, \tau) \partial_k \left[u_j^s\right]^+(\mathbf{x}, -\tau). \tag{17}$$

The adjoint wavefield is defined as

$$\left[u_j^s\right]^+(\mathbf{x}, \tau) = \int dt \sum_{rin} G_{ji}(\mathbf{x}, \tau - t; \mathbf{x}_r) J_{in}^{sr}(-t)\, d_{in}^{sr}. \tag{18}$$

In practice, the adjoint wavefield is not constructed using the transposed Green's tensor operator, but rather by solving the adjoint wave equation with the adjoint source field (Tarantola 1988; Ackelik *et al.* 2003; Tromp *et al*. 2005),

$$\left[f_i^s\right]^+(\mathbf{x}, t) = \sum_{rn} J_{in}^{sr}(-t) d_{in}^{sr} \delta(\mathbf{x} - \mathbf{x}_r). \tag{19}$$

The solution is then integrated in time with the derivatives of the forward wavefield to construct $K_{\chi^2}^{\rho}$ and $K_{\chi^2}^{c_{jklm}}$. Because the adjoint source field sums over the receiver index r, only one adjoint calculation is needed for each seismic source s, no matter how many receivers are used. The computational cost for one adjoint calculation is about the same as calculating the forward wavefield.

Because the forward wavefield $\mathbf{u}^s$ and the adjoint wavefield $[\mathbf{u}^s]^\dagger$ are both source-specific, $2N_s$ wavefield simulations are needed for one calculation of $K_{\chi^2}^{\rho}$ and $K_{\chi^2}^{c_{jklm}}$ (N_s forward simulations and N_s adjoint simulations) as well as $N_V N_s$ operations to compute the time-integration of the forward and the adjoint wavefields, where N_V is the number of spatial grid points. If the conjugate-gradient algorithm is used to minimize the misfit functional (Press *et al.* 1992), each iteration requires one calculation of $K_{\chi^2}^{\rho}$ and $K_{\chi^2}^{c_{jklm}}$ to determine the conjugate direction in model space, followed by a line search to determine the optimal step to take in this conjugate direction. A cubic interpolation for the line search (Tromp *et al.* 2005) requires an additional N_s forward calculations and N_s adjoint calculations; hence, the total number of simulations per iteration is $4N_s$, and the total number of time integration operations is $2N_V N_s$ (Table 1). A quadratic interpolation scheme is often adequate, however (Tape *et al*. 2006), which reduces the total number of simulations per iteration to $3N_s$ and the total number of time integration per iteration to $N_V N_s$.

During the construction of $K_{\chi^2}^{\rho}$ and $K_{\chi^2}^{c_{jklm}}$, we need simultaneous access to both the forward wavefield $\mathbf{u}^s$ and the adjoint field $[\mathbf{u}^s]^\dagger$ in order to perform the time integration operation. If we choose to store $\mathbf{u}^s$ and read it back during the construction of $K_{\chi^2}^{\rho}$ and $K_{\chi^2}^{c_{jklm}}$, the storage requirement for the adjoint method is the space–time volume of one wavefield, which is proportional to $N_V N_T$, where N_T is the total number of time steps for the wavefield simulation. If we choose to reconstruct $\mathbf{u}^s$ on the fly, we need to store only the last time frame of $\mathbf{u}^s$ as a function of space, which has volume proportional to N_V, although additional CPU time is then required to recover the forward wavefield by solving the wave equation backward in time, taking the final state of the forward wavefield as its initial condition (Gauthier *et al.* 1986; Tromp *et al.* 2005). The CPU time for doing one backward wave propagation is about the same as doing one forward simulation.

Table 1. Comparison of computational costs for a single optimization step of the scattering-integral (SI) and adjoint-wavefield (AW) methods.

Cost	SI method	AW–CG[a]	AW–GN[b]
Storage requirement	$3N_rN_VN_T$	N_V	N_V
Number of simulations	$3N_r + N_s$	$6N_s$	$4N_{\rm CG}N_{\rm s} + 2N_{\rm s}$
Number of time integrations	$2N_t\,N_V\,N_u$	$2\,N_V\,N_s$	$N_V\,(2N_{\rm CG}N_{\rm s} + N_{\rm s})$
I/O cost	$N_uN_TN_V$	$2N_sN_V$	$(2N_{\rm CG}N_{\rm s} + N_{\rm s})N_V$
Requires solving a linear system?	Yes	No	Yes
Optimization algorithm	Gauss–Newton	Conjugate-Gradient	Gauss–Newton
Number of iterations needed to match one Gauss–Newton step	1	6–7	1

[a]The AW–CG method listed here is based on cubic line search as implemented in Tromp *et al*. (2005). Only the last time frame is stored, and the forward wavefield is reconstructed on-the-fly during time integration. The cost listed here is for a single conjugate-gradient optimization step.
[b]The AW–GN method listed here is based on the implementation by Akcelik *et al.* (2002). The computational cost listed here is for a single Gauss–Newton optimization step, which involves $N_{\rm CG}$ conjugate-gradient iterations for solving the Gauss–Newton normal equation.

The computational cost for reading/writing data from/to storage facilities (I/O cost) depends on the number of time integration operations and whether the forward wavefield $\mathbf{u}^s$ is stored or reconstructed on the fly. If a cubic interpolation scheme is used in the line search and $\mathbf{u}^{\rm s}$ is reconstructed on the fly, the I/O cost is proportional to $2N_sN_V$. The computational cost for a single conjugate-gradient step of this type of AW method is summarized in Table 1 under the column 'AW—CG'.

5 SCATTERING-INTEGRAL METHOD

Substituting the first-order Born approximation (5) into (3) and applying seismic reciprocity yields the Fréchet kernels for the individual data functionals d_{in}^{sr}:

$$K_{d_{in}^{sr}}^{\rho}(\mathbf{x}) = -\int dt \int d\tau\, J_{in}^{sr}(t) \sum_j G_{ji}(\mathbf{x}, t-\tau; \mathbf{x}_r)\partial_\tau^2 u_j^s(\mathbf{x}, \tau), \tag{20}$$

$$K_{d_{in}^{sr}}^{c_{jklm}}(\mathbf{x}) = -\int dt \int d\tau\, J_{in}^{sr}(t)\, \partial_k G_{ji}(\mathbf{x}, t-\tau; \mathbf{x}_r)\, \partial_l u_m^s(\mathbf{x}, \tau), \tag{21}$$

The transposed Green tensor $G_{ji}(\mathbf{x}, t-\tau; \mathbf{x}_r)$ contains the wavefields generated by three orthogonal unit impulsive point forces acting at the receiver location $\mathbf{x}_r$, so we call it the *receiver Green tensor* (RGT). The data-specific Fréchet kernels can be computed by convolving the forward wavefield for source s with the RGT for receiver r and integrating the results against $J_{in}^{sr}(t)$. Eqs (20) and (21) are equivalent to the scattering integrals (16)–(19) in Zhao *et al*. (2005).

We can construct the RGT once for each receiver, store it (at cost $\sim N_VN_T$), and use it for every source in the dataset, which can be efficient when the number of seismic sources is larger than the number of receivers. A complete calculation of the data-specific kernels requires $3N_r$ simulations (considering 3 orthogonal unit forces at each receiver) and one forward simulation for each source—a total of $3N_r + N_s$, plus a storage cost proportional to $3N_r\,N_VN_T$. The number of time integration operations is determined by the convolution between the forward wavefields and the RGTs and is proportional to $2N_t$, where N_t is the number of time steps within the support of $J_{in}^{sr}(t)$, the number of spatial grid points N_V and the total number of seismograms N_u used in the inversion. For three-component receivers, $N_u \le 3N_r\,N_s$. Owing to noise and instrumental reasons for eliminating seismograms, N_u is often much less than this bound. The computational cost for the SI method is summarized in Table 1 under the column 'SI'.

In the SI method, the quadratic misfit is minimized by a Gauss–Newton algorithm. The misfit functional is expanded to second order in the model perturbation,

$$\chi^2(\mathbf{m},\,\tilde{\mathbf{m}}) \approx \chi^2(\tilde{\mathbf{m}},\,\tilde{\mathbf{m}}) + \mathbf{a}(\tilde{\mathbf{m}})\,(\mathbf{m}-\tilde{\mathbf{m}}) + \frac{1}{2}(\mathbf{m}-\tilde{\mathbf{m}})^{\rm T}\mathbf{H}(\tilde{\mathbf{m}})\,(\mathbf{m}-\tilde{\mathbf{m}}), \tag{22}$$

and its gradient with respect to the target model is set to zero,

$$\nabla_{\mathbf{m}}\chi^2(\mathbf{m},\,\tilde{\mathbf{m}}) = \mathbf{a}(\tilde{\mathbf{m}}) + \mathbf{H}(\tilde{\mathbf{m}})\,(\mathbf{m}-\tilde{\mathbf{m}}) = 0. \tag{23}$$

In these expressions, $\mathbf{a} = \nabla_{\tilde{\mathbf{m}}}\chi^2 = -\mathbf{A}(\tilde{\mathbf{m}})^{\rm T}\,\mathbf{d}$ is the data-weighted Fréchet kernels, and $\mathbf{H}$ is the Hessian (Press *et al*. 1992),

$$\mathbf{H} = \nabla_{\tilde{\mathbf{m}}}\nabla_{\tilde{\mathbf{m}}}\chi^2 = \mathbf{A}^{\rm T}\,\mathbf{A} + (\nabla_{\tilde{\mathbf{m}}}\mathbf{A})^{\rm T}\,\mathbf{d}. \tag{24}$$

The second term in the Hessian is routinely neglected (e.g. Nolet, 1987; Sambridge, 1990), which leads to the Gauss–Newton normal equation,

$$\mathbf{A}^{\rm T}\mathbf{A}\,(\mathbf{m}-\tilde{\mathbf{m}}) = \mathbf{A}^{\rm T}\,\mathbf{d}. \tag{25}$$

Because of the positive definiteness of the reduced Hessian $\mathbf{A}^{\rm T}\mathbf{A}$, the solution of eq. (25) will always give a descent direction. Eq. (25) will provide a good approximation to (23) if either the data vector $\mathbf{d}$ or $\nabla_{\tilde{\mathbf{m}}}\mathbf{A}$ is small. The latter will be the case when the data functionals are only weakly non-linear with respect to model parameters (Tarantola 2005). In this respect, the Rytov approximation used in constructing the GSDFs of (9) and (10) is usually much superior to the Born approximation required for the waveform difference functional (6).

If $\mathbf{A}^{\rm T}\mathbf{A}$ approximates the exact Hessian, the Gauss–Newton algorithm converges to a minimum of the misfit functional almost quadratically, a big advantage over gradient-based optimization algorithms such as steepest decent and conjugate gradient, which usually have linear or superlinear convergence rate (Press *et al.* 1992). In practice, the normal equation need not be formed, because its solution can be computed by solving the linear system $\mathbf{A}\delta\mathbf{m} = \mathbf{d}$ via a relaxation method, such as LSQR (Paige & Saunders 1982). This system can be scaled by a data covariance matrix $\mathbf{C_d}$ and augmented to include a model regularization term in the misfit functional (e.g. Snieder & Trampert 1999). A Bayesian solution can be obtained

by referencing the model perturbation to some $\mathbf{m}_{\text{prior}}$ and scaling the regularization using an *a priori* model covariance operator $\mathbf{C_m}$ (Tarantola 2005).

For each iteration of the Gauss–Newton's method, we require $3N_r + N_s$ simulations, N_u convolution operations to construct kernels for individual data functionals, and additional CPU time (though not much) for solving the augmented linear system. The I/O cost for the SI method is proportional to $N_u N_V N_T$ (Table 1).

6 ADJOINT-WAVEFIELD METHOD II: GAUSS–NEWTON IMPLEMENTATION

As pointed out in Tromp *et al*. (2005), the Fréchet kernel for the misfit functional can be expressed as a weighted summation of the Fréchet kernels for individual data functionals,

$$K^{\mathbf{m}}_{\chi^2}(\mathbf{x}) = \sum_{srin} d^{sr}_{in} K^{\mathbf{m}}_{d^{sr}_{in}}(\mathbf{x}) \tag{26}$$

In the AW–CG method, the adjoint sources located at all receivers are propagated back simultaneously in one adjoint calculation, so the integration of the forward and the adjoint wavefields gives only a weighted summation over receivers,

$$\int d\tau \partial_\tau^2 u^s_j(\mathbf{x}, \tau)\left[u^s_j\right]^+(\mathbf{x}, -\tau) = \sum_{rin} d^{sr}_{in} K^{\rho}_{d^{sr}_{in}}(\mathbf{x}), \tag{27}$$

$$\int d\tau \partial_l u^s_m(\mathbf{x}, \tau)\partial_k\left[u^s_j\right]^+(\mathbf{x}, -\tau) = \sum_{rin} d^{sr}_{in} K^{c_{jklm}}_{d^{sr}_{in}}(\mathbf{x}). \tag{28}$$

Therefore, the AW–CG method provides no access to the kernels for individual data functionals, and the Hessian cannot be constructed; only gradient-based algorithms can be used to minimize the misfit functional.

As pointed out in Akcelik *et al.* (2002, 2003), the "Hessian vector product" $\mathbf{p} = \mathbf{Hq}$ can be obtained through one forward and one adjoint simulations and one time integration operation for each source without explicitly forming $\mathbf{H}$ or $\mathbf{A}$. For the forward simulation, we solve the wave equation for seismic source s with the following source field

$$h^s_j(\mathbf{x}, \tau) = \sum_{klm} \partial_k \left[\partial_l u^s_m(\mathbf{x}, \tau) q_{jklm}(\mathbf{x})\right], \tag{29}$$

where $q_{jklm}(\mathbf{x})$ is $\mathbf{q}$ written in component form. We denote the solution at the receiver location $\mathbf{x}_r$ due to this source field as $w^s_i(\mathbf{x}_r, t)$. We then solve the adjoint wave equation for seismic source s with the following adjoint source field

$$[h^s_i]^+(\mathrm{x}, \mathrm{t}) = -\sum_{rn} J^{sr}_{in}(-t)\delta(\mathbf{x} - \mathbf{x}_r) \int d\tau J^{sr}_{in}(\tau) w^s_i(\mathbf{x}_r, \tau), \tag{30}$$

and obtain the adjoint solution $[w_j{}^s]^\dagger(\mathbf{x}, t)$. The Hessian vector product can then be constructed by one integration operation for each seismic source s

$$p_{jklm}(\mathbf{x}) = \sum_s \int dt \partial_l u^s_m(\mathbf{x}, t)\partial_k\left[w^s_j\right]^+(\mathbf{x}, -t), \tag{31}$$

where $p_{jklm}(\mathbf{x})$ is $\mathbf{p}$ written in component form. If we solve the Gauss–Newton normal eq. (25) using the conjugate-gradient method, vector $\mathbf{q}$ is the conjugate direction and each conjugate-gradient step requires two Hessian vector products. To construct the right-hand side of eq. (25) requires one adjoint calculation and N_V time-integration operations for each source. If solving eq. (25) requires N_{CG} conjugate-gradient steps, then each Gauss–Newton iteration will require $4N_{\text{CG}}N_{\text{s}} + 2N_{\text{s}}$ simulations and $N_V(2N_{\text{CG}}N_{\text{s}} + N_{\text{s}})$ time-integration operations. The computational cost for this type of AW method is summarized in Table 1 under the column 'AW–GN'.

7 APPLICATION OF THE SI METHOD TO THE LOS ANGELES REGION

We have applied the SI method to a full 3-D tomography using phase-delay measurements of crustal *P* and *S* phases from 67 small earthquakes recorded by 48 stations of the Southern California Seismic Network. We inverted 7364 frequency-dependent phase delays in the frequency band from 0.2 to 1.0 Hz, which we measured relative to synthetics calculated from version 3.0 of the SCEC 3-D Community Velocity Model (CVM3.0) (Kohler *et al.* 2003). The wavefields and the RGTs were computed on a mesh with 200-m grid spacing (36 million nodes) using a fourth-order, staggered-grid, finite-difference code developed by K. Olsen.

The computational aspects of the inversion cycle are summarized in Table 2. All computations and I/O operations were performed in parallel using the Message Passing Interface (Pacheco 1997). Each wave-propagation simulation took about 90 min on a Linux cluster of 128 Pentium-4 CPUs. Constructing each data kernel required about 4 min to read the RGTs from local disk and 1 min to do the convolutions. To represent $\delta\mathbf{m}$ and compute the kernels, we decimated the simulation grid to a 2-km node spacing. The linearized perturbation equations were scaled by the data-covariance matrix $\mathbf{C_d}$ and augmented to incorporate the data nuisance parameters (source and station anomalies) and the inverse model-covariance operator $\mathbf{C}^{-1}_{\mathbf{m}}$, needed for a Bayesian solution. The resulting linear system, which measured 77,695 columns by 84 896 rows, was inverted by a parallelized version of LSQR in less than 1 hr—a negligible computing expense. The variance reduction, calculated by resynthesizing the seismograms using the perturbed model and remeasuring the GSDFs, was almost 80% for phase-delay times. For the amplitude-reduction times, which were measured but not inverted, the perturbed model achieved a variance reduction of about 40%, consistent with the improvements visible in the waveform fits. A complete description of the inversion results is given in Chen *et al.* (2006).

On the128-processor cluster, about 20 day of CPU time and 24 terabytes (TB) of storage space were required for one complete iteration of the SI method. Based on the scalings in Table 1, we estimate that the AW–CG method would require about 100 day of CPU time for six conjugate-gradient steps, which is probably the minimum number needed to achieve the data variance reduction obtained in a single Gauss–Newton step (Table 1; Tape *et al*. 2006). Numerical experiments conducted in Akcelik (2002) show that for the AW–GN method N_{CG} usually ranges from 10 to 30. Therefore, more CPU time would be needed for the AW–GN method to match one SI step. The AW storage cost would be very small ($\sim$0.04 TB).

Table 2. Computational parameters for the Los Angeles Basin tomography using the SI method.

Number of stations N_r	48
Number of earthquakes N_s	67
Number of seismograms N_u	2000
Number of FD simulations $3N_r + N_s$	211
Simulation grid spacing, time interval	200 m, 0.01 s
Simulation grid points N_V, time steps N_T	36 140 440, 6000
Number of CPUs	128
Total CPU time per iteration	62 000 CPU-hours
Total disk space $3N_r N_V N_T$	24 TB

When disk storage is cheap and the number of earthquakes is larger than the number of receivers, as in our case, the SI method is more economical than the AW methods. As the number of sources used in the inversion increases, the efficiency of the SI method increases. Moreover, the storage requirements and I/O overhead of the SI method can be reduced, perhaps significantly, by deploying more efficient data compression algorithms (e.g. Bassiouni 1985).

Storing the RGTs has other advantages. By applying the reciprocity principle, we can efficiently use them to calculate synthetic seismograms and obtain the source parameters of new earthquakes; we need only extract a small, source-centred volume from each RGT to compute three-component seismograms and the partial derivatives required for source estimation. We have implemented automated procedures that invert the waveform data from small to moderate-sized earthquakes in Southern California for centroid moment tensors (Zhao *et al.* 2006) and finite moment tensors (Chen *et al.* 2005). For example, the source parameters used in the full 3-D tomography described here were refined using the RGTs computed for CVM3.0 prior to the structural inversion.

ACKNOWLEDGMENTS

This research was supported by the Southern California Earthquake Center; SCEC is funded by NSF Cooperative Agreement EAR-0106924 and USGS Cooperative Agreement 02HQAG0008. We thank J. Tromp and J. Bielak for helpful discussions of AW methods. High-throughput computations were supported by the SCEC Community Modeling Environment under NSF grant EAR-0122464 and the USC Center for High Performance Computing and Communications; we thank P. Maechling and other members of CME collaboration and G. Staples from USC/HPCC for help in implementing the simulation and inversion algorithms. The SCEC contribution number for this paper is 1011.

REFERENCES

Akcelik, V., Biros, G. & Omar, G., 2002. Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation, *Proceedings of SC2002,* Baltimore. IEEE/ACM.

Akcelik, V. *et al.*, 2003. High resolution forward and inverse earthquake modeling on terascale computers, *SC2003, Phoenix, AZ*.

Aki, K. & Richards, P.G., 1980. *Quantitative Seismology: Theory and Methods*, W. H. Freeman & Co., San Francisco, California.

Bassiouni, M.A., 1985. Data compression in scientific and statistical databases, *IEEE Trans. Software Eng.,* **11,** 1047–1058.

Biondi, B., 2006. *3-D Seismic Imaging*, Soc. Exploration Geophysicists, in press; http://sepwww.stanford.edu/sep/biondo/3DSI˙frame.html.

Chen, P., 2005. A unified methodology for seismic waveform analysis and inversion, *PhD dissertation,* University of Southern California. http://www-rcf.usc.edu/∼pochen.

Chen, P., Zhao, L. & Jordan, T.H., 2007. Full 3D tomography for the crustal structure in the Los Angeles region, *Bull. Seism. Soc. Am.,* in press.

Chen, P., Jordan, T.H. & Zhao, L., 2005. Finite moment tensor of the 3 September 2002Yorba Linda earthquake, *Bull. Seis. Soc. Am.,* **95,** 1170–1180.

Chernov, L.A., 1960. *Wave Propagation in a Random Medium*. McGraw-Hill, New York, 168 pp.

Dahlen, F.A., Hung, S.-H. & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes—I. Theory, *Geophys. J. Int.,* **141,** 157–174.

Dahlen, F.A. & 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, 1025 pp.

Gauthier, O., Virieux, J. & Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: numerical results, *Geophysics,* **51**(7), 1387–1403.

Gee, L.S. & Jordan, T.H., 1992. Generalized seismological data functionals, *Geophys. J. Int.,* **111,** 363–390.

Graves, R., 1996. Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences, *Bull. Seism. Soc. Am.,* **86,** 1091–1106.

Katzman, R., Zhao, L. & Jordan, T.H., 1998. High-resolution, two-dimensional vertical tomography of the central Pacific mantle using *ScS* reverberations and frequency-dependent travel times, *J. Geophys. Res.,* **103**(B8), 17933–17971.

Kohler, M., Magistrale, H. & Clayton, R., 2003. Mantle heterogeneities and the SCEC three-dimensional seismic velocity model version 3, *Bull. Seismol. Soc. Am.,* **93,** 757–774.

Komatitsch, D., Liu, Q., Tromp, J., Süss, P., Stidham, C. & Shaw, J.H., 2004. Simulations of ground motion in the Los Angeles basin based upon spectral-element method, *Bull. Seismol. Soc. Am.,* **94,** 187–206.

Li, X.-D. & Romanowicz, B., 1995, Comparison of global waveform inversions with and without considering cross-brach modal coupling, *Geophys. J. Int.,* **121,** 695–709.

Li, X.-D. & Tanimoto, T., 1993. Waveforms of long-period body waves in a slightly aspherical Earth model, *Geophys. J. Int.,* **112,** 92–112.

Liu, Q. & Tromp, J., 2006. Finite-frequency kernels based upon adjoint methods, *Bull. Seism. Soc. Am.,* **96**(6), 2383–2397; DOI: 10.1785/0120060041.

Luo, Y. & Shuster, G.T., 1991. Wave-equation traveltime inversion, *Geophysics,* **56,** 645–653.

Marquering, H., Dahlen, F.A. & Nolet, G., 1999. Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox, *Geophys. J. Int.*, **137,** 805–815.

Nolet, G., 1987. Seismic wave propagation and seismic tomography, in *Seismic Tomography*, pp. 1–24, ed. Nolet, G., Reidel, Dordrecht.

Olsen, K.B., 1994. Simulation of three-dimensional wave propagation in the Salt Lake Basin, *PhD thesis*, University of Utah, Salt Lake City, Utah, 157 p.

Olsen, K.B., Day, S.M. & Bradley, C.R., 2003. Estimation of Q for long-period (>2 s) waves in the Los Angeles Basin, *Bull. Seis. Soc. Am.,* **93,** 627–638.

Pacheco, P.S., 1997. *Parallel Programming with MPI*, Morgan Kaufmann Press, San Francisco.

Paige, C.C. & Saunders, M.A., 1982. LSQR: sparse linear equations and least squares problems, *ACM Trans. Math. Software,* **8,** 195–209.

Pratt, R.G., 1990. Inverse theory applied to multi-source cross-hole tomography. Part II: elastic wave-equation method, *Geophys. Prospect.,* **38,** 311–330.

Pratt, R.G., Shin, C. & Hicks, G.J., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophys. J. Int.,* **133,** 341–362.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P., 1992. *Numerical Recipets in C, the Art of Scientific Computing,* 2nd edn, Cambridge University Press, Cambridge.

Ritsema, J. & van Heijst, H.J., 2000. Seismic imaging of structural heterogeneity in Earth's mantle: evidence for large-scale mantle flow, *Sci. Progr.,* **83,** 243–259.

Ritsema, J., Rivera, L., 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

Sambridge, M.S., 1990. Non-linear arrival time inversion: constraining velocity anomalies by seeking smooth models in 3-D, *Geophys. J. Int.,* **102,** 653–677.

Snieder, R. & Lomax, A., 1996. Wavefield smoothing and the effect of rough velocity perturbations on arrival times and amplitudes, *Geophys J. Int.,* **125,** 796–812.

Snieder, R. & Trampert, J., 1999. Inverse problems in geophysics, in *Wavefield Inversion,* pp. 119–190, ed. A. Wirgin, Springer Verlag, New York.

Tape, C., Liu, Q. & Tromp, J., 2006. Finite-frequency tomography using adjoint methods—methodology and examples using membrane surface waves, *Geophys. J. Int.,* **168,** 1105–1129.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics,* **49,** 1259–1266.

Tarantola, A., 1988. Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation, *Pure Appl. Geophys.,* **128**(1/2), 365–399.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, Philadelphia, PA, pp. 342.

Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.,* **160,** 195–216.

Woodhouse, J.H. & Dziewonski, A.M., 1984. Mapping the upper mantle: three dimensional modeling of earth structure by inversion of seismic waveforms, *J. Geophys. Res.,*, **89,** 5953–5986.

Woodward, R.L. & Masters, G., 1991. Global upper mantle structure from long-period differential travel-times, *J. Geophys. Res.,* **96,** 6351–6377.

Zhao, L. & Jordan, T.H., 1998. Sensitivity of frequency-dependent travel-times to laterally heterogeneous, anisotropic Earth structure, *Geophys. J. Int.,* **133,** 683–704.

Zhao, L., Jordan, T.H. & Chapman, C.H., 2000. Three-dimensional Fréchet differential kernels for seismic delay times, *Geophys. J. Int.,* **141,** 58–576.

Zhao, L., Jordan, T.H., Olsen, K.B. & Chen, P., 2005. Fréchet kernels for imaging regional Earth structure based on three-dimensional reference models, *Bull. Seis. Soc. Am.,* **95**(6), 2066–2080.

Zhao, L., Chen, P. & Jordan, T.H., 2006. Strain Green tensors, reciprocity and their applications to seismic source and structure studies, *Bull. Seis. Soc. Am.,* **96**(5), 1753–1763.