Frozen Gaussian approximation for three-dimensional seismic tomography

Lihui Chai* Ping Tong† Xu Yang‡

Abstract

Three-dimensional (3D) wave-equation-based seismic tomography is computationally challenging in large scales and high-frequency regime. In this paper, we apply the frozen Gaussian approximation (FGA) method to compute 3D sensitivity kernels and seismic tomography of high-frequency. Rather than standard ray theory used in seismic inversion (e.g. Kirchhoff migration and Gaussian beam migration), FGA is used to compute the 3D high-frequency sensitivity kernels for travel-time or full waveform inversions. Specifically, we reformulate the equations of the forward and adjoint wavefields for the purpose of convenience to apply FGA, and with this reformulation, one can efficiently compute the Green's functions whose convolutions with source time function produce wavefields needed for the construction of 3D kernels. Moreover, a fast summation method is proposed based on local fast Fourier transform which greatly improves the speed of reconstruction as the last step of FGA algorithm. We apply FGA to both the travel-time adjoint tomography and full waveform inversion (FWI) on synthetic crosswell seismic data with dominant frequencies as high as those of real crosswell data, and confirm again that FWI requires a more sophisticated initial velocity model for the convergence than travel-time adjoint tomography. We also numerically test the accuracy of applying FGA to local earthquake tomography. This study paves the way to directly apply wave-equation-based seismic tomography methods into real data around their dominant frequencies.

1 Introduction

Seismic tomography is a widely used method for imaging the subsurface structures of Earth at a variety of scales. Crucial information has been extracted from tomographic images to understand tectonics, volcanism, and geodynamics (e.g. Romanowicz, 1991; Rawlinson *et al.*, 2010; Zhao, 2012) since the pioneering work of Aki and Lee (1976). In the meantime, there have been significant developments of methodology and data usage in seismic tomography

*Department of Mathematics, University of California, Santa Barbara, CA 93106, USA. (chailh@math.ucsb.edu)

†School of Physical and Mathematical Sciences & Asian School of the Environment, Nanyang Technological University, Singapore 639798. (tongping@ntu.edu.sg)

‡Department of Mathematics, University of California, Santa Barbara, CA 93106, USA. (xuyang@math.ucsb.edu)

(Tong *et al.*, 2017). At the beginning, ray theory is the main tool used in seismic tomography based on the assumption that seismic travel-time is only determined by the structure that infinitely thin ray passes. However, this assumption is only valid for extremely high-frequency waves (Rawlinson *et al.*, 2010), and otherwise, 3D structures off the ray path will affect wavefront healing and measurements of travel-time and amplitude made on broadband recordings due to wave scattering (e.g. Dahlen *et al.*, 2000; Liu and Gu, 2012).

Ray-based finite-frequency tomography (e.g. Dahlen *et al.*, 2000; Tong *et al.*, 2011) and full wave-equation-based inversion methods (Tromp *et al.*, 2005; Chen *et al.*, 2007b; Tong *et al.*, 2014a,b) have been proposed to include the influence of off-ray structures by constructing 2D and 3D travel-time and amplitude sensitivity kernels. Improved resolutions of tomographic results have been reported in ray-based finite-frequency tomographic studies (e.g. Montelli *et al.*, 2004; Hung *et al.*, 2004). The resolutions can be further increased if full wave-equation-based methods are used to invert full or even partial waveform contents (Virieux and Operto, 2009; Huang *et al.*, 2016). However, solving 3D full wave equations is always computationally expensive. Only the recent rapid developments in high-performance computing and forward modeling techniques make the computation of sensitivity kernels based on simulation of full seismic wavefield become feasible (e.g. Komatitsch and Tromp, 2002; Fichtner *et al.*, 2009; Tong *et al.*, 2014a; Chai *et al.*, 2017), which also make full wave-equation-based seismic inversions applicable in realistic situations (e.g. Tape *et al.*, 2009; Zhu *et al.*, 2012; Huang *et al.*, 2016).

Among popular wave-equation-based tomographic inversion methods, time-domain adjoint tomography (Tromp *et al.*, 2005; Liu and Gu, 2012) and frequency-domain full waveform inversion (e.g. Pratt and Shipp, 1999; Virieux and Operto, 2009) have been intensively studied based on our current knowledge. This study focuses on the time-domain adjoint tomography method. Adjoint tomography iteratively optimizes velocity models by fully solving 3D wave-equations for the accurate computation of seismograms and sensitivity kernels in complex models (Liu and Tromp, 2008; Tong *et al.*, 2014a). Successful applications of adjoint tomography include but are not limited to a detailed velocity model of the southern California crust (Tape *et al.*, 2009, 2010), a tomographic model of the European upper mantle (Zhu *et al.*, 2012), and a 3D S-wave velocity model of the North Atlantic region (Rickers *et al.*, 2013). Although fully solving wave-equations provides accurate synthetic seismograms and sensitivity kernels, it also leads to demanding and even unaffordable computational cost which prevents wide application of adjoint tomography in 3D seismic inversion problems. We should be noted that most of the previous real applications used significantly low-frequency data compared to the dominant frequencies of the originally observed data in order to be computationally feasible (e.g. Tape *et al.*, 2009; Zhu *et al.*, 2012). Usually, a low-frequency earthquake has its dominant frequency around 1 Hz, however, a typical earthquake has a dominant frequency around 5 Hz (e.g. Nakamichi *et al.*, 2003). The simulation of high-frequency seismic waves requires much more computational resources on both memories and CPU time than the low-frequency simulation. So improving the efficiency and accuracy of simulation methods is constantly necessary to use real seismic data around their dominant frequencies.

To balance the computational efficiency and accuracy of full wave-equation based tomographic methods, we propose to use the frozen Gaussian approximation (FGA) to compute the 3D forward modeling and sensitivity kernel. FGA approximates seismic wavefield by

a summation of frozen (fixed-width) Gaussian wave-packets propagating along ray paths (Chai *et al.*, 2017). Rigorous asymptotic analysis on phase plane shows that the asymptotic error is proportional to the ratio of short wavelength to large propagation distance (Lu and Yang, 2011, 2012a,b; Yang *et al.*, 2013; Li *et al.*, 2015; Chai *et al.*, 2017). One can use a relatively small number of Gaussians to get accurate approximations of high-frequency wavefield. Meanwhile, FGA algorithm can be perfectly parallelized, which speeds up the computation drastically with a high-performance computing station. The accuracy and efficiency are illustrated in Chai *et al.* (2017) in high-frequency regime by comparisons between FGA and the spectral element method for 3D seismic wave propagations.

In this paper, we first reformulate the equations of the forward and adjoint wavefields in adjoint tomography methods, in which one can easily use FGA to compute the Green's function and get the forward and adjoint wavefields by convoluting the Green's function with source seismogram. This reformulation provides an analytical form of initial weight function for FGA, and thus avoids the computational cost of decomposing wavefield into Gaussians in standard FGA formulation (Yang *et al.*, 2013; Chai *et al.*, 2017); details of discussions are provided in Section 3.2. We also propose a fast summation method based on local fast Fourier transform which greatly improves the speed of reconstruction as the last step of FGA algorithm. Then we reconstruct the 3D finite-frequency kernels for travel-time or full waveform inversions based on the forward and adjoint wavefields and apply FGA to both the travel-time adjoint tomography and full waveform inversion (FWI) on synthetic crosswell seismic data. Our numerical results show that FWI requires a more sophisticated initial velocity model for the convergence than travel-time adjoint tomography. We also numerically test the accuracy of applying FGA to local earthquake tomography. One of the key features of this study is that the FGA method enables us to conduct 3D wave-equation-based inversions with high-frequency data around the dominant frequencies of the observed data in real situations.

2 Tomographic equation

In this section, we summarize the tomographic equations used to determine the linear relationship between the perturbation of observed data (e.g. travel-time or full wavefield) and velocity perturbation in a reference model.

Denote $\delta \boldsymbol{d} = \boldsymbol{d}_{\text{obs}} - \boldsymbol{d}_{\text{sys}}$, where $\boldsymbol{d}_{\text{obs}}$ is the recorded data at the receivers and $\boldsymbol{d}_{\text{sys}}$ is the predicted data by simulation on a reference velocity model $\boldsymbol{c}$. Remark that in real-life data applications, $\boldsymbol{d}_{\text{obs}}$ implicitly contains an error term e that accounts for the measurement error as well as noises, but since this will not affect the introduction of the algorithm, we shall take the error matrix to be zero for a sake of simplicity. One defines the misfit functional by

$$\chi(\boldsymbol{c}) = \frac{1}{2}\langle \delta \boldsymbol{d}, \delta \boldsymbol{d} \rangle, \tag{1}$$

where $\langle \cdot, \cdot \rangle$ is an (symmetric) inner product defined on the data space. To minimize the misfit functional χ over $\boldsymbol{c}$, one can use the Born approximation around the reference model

$\boldsymbol{c}_0$,

$$\chi(\boldsymbol{c}_0 + \delta\boldsymbol{c}) \approx \chi(\boldsymbol{c}_0) + \frac{\delta\chi}{\delta\boldsymbol{c}}(\boldsymbol{c}_0) \cdot \delta\boldsymbol{c} + \frac{1}{2}\delta\boldsymbol{c} \cdot \frac{\delta^2\chi}{\delta^2\boldsymbol{c}}\delta\boldsymbol{c}, \tag{2}$$

and therefore

$$\frac{\delta\chi}{\delta\boldsymbol{c}}(\boldsymbol{c}_0 + \delta\boldsymbol{c}) \approx \frac{\delta\chi}{\delta\boldsymbol{c}}(\boldsymbol{c}_0) + \frac{\delta^2\chi}{\delta^2\boldsymbol{c}}\delta\boldsymbol{c}. \tag{3}$$

Then $\dfrac{\delta\chi}{\delta\boldsymbol{c}} = 0$ implies

$$\delta\boldsymbol{c} = -\left(\frac{\delta^2\chi}{\delta^2\boldsymbol{c}}(\boldsymbol{c}_0)\right)^{-1} \frac{\delta\chi}{\delta\boldsymbol{c}}(\boldsymbol{c}_0). \tag{4}$$

One computes the gradient by

$$\frac{\delta\chi}{\delta\boldsymbol{c}} = -\left\langle \frac{\delta\boldsymbol{d}_{\text{sys}}}{\delta\boldsymbol{c}}, \delta\boldsymbol{d} \right\rangle, \tag{5}$$

and the Hessian by

$$\frac{\delta^2\chi}{\delta^2\boldsymbol{c}} = \left\langle \frac{\delta\boldsymbol{d}_{\text{sys}}}{\delta\boldsymbol{c}}, \frac{\delta\boldsymbol{d}_{\text{sys}}}{\delta\boldsymbol{c}} \right\rangle - \left\langle \frac{\delta^2\boldsymbol{d}_{\text{sys}}}{\delta^2\boldsymbol{c}}, \delta\boldsymbol{d} \right\rangle. \tag{6}$$

In the linear perturbation model, the second term on the right-hand-side of (6) is zero, and thus

$$\frac{\delta^2\chi}{\delta^2\boldsymbol{c}} = \left\langle \frac{\delta\boldsymbol{d}_{\text{sys}}}{\delta\boldsymbol{c}}, \frac{\delta\boldsymbol{d}_{\text{sys}}}{\delta\boldsymbol{c}} \right\rangle, \tag{7}$$

which implies

$$\delta\boldsymbol{c} = \left(\left\langle \frac{\delta\boldsymbol{d}_{\text{sys}}}{\delta\boldsymbol{c}}, \frac{\delta\boldsymbol{d}_{\text{sys}}}{\delta\boldsymbol{c}} \right\rangle\right)^{-1} \left\langle \frac{\delta\boldsymbol{d}_{\text{sys}}}{\delta\boldsymbol{c}}, \delta\boldsymbol{d} \right\rangle. \tag{8}$$

2.1 Traveltime adjoint tomography

In travel-time adjoint tomography,

$$\boldsymbol{d}_{\text{obs/sys}} = T_{\text{obs/sys}},\ \delta\boldsymbol{d} = \delta T = T_{\text{obs}} - T_{\text{sys}},\ \text{and}\ \langle \delta T, \delta T\rangle = |\delta T|^2.$$

Define the forward field $u(t, \boldsymbol{x})$ that satisfies the forward propagating wave equation,

$$\begin{cases} \partial_t^2 u - c^2(\boldsymbol{x})\Delta_{\boldsymbol{x}} u = s(t)\delta_d(\boldsymbol{x} - \boldsymbol{x}_s), \quad \boldsymbol{x} = (x, y, z) \in \mathbb{R}^3, \\ u(0, \boldsymbol{x}) = \partial_t u(0, \boldsymbol{x}) = 0, \end{cases} \tag{9}$$

where $c(\boldsymbol{x})$ is the reference velocity, $\Delta_{\boldsymbol{x}}$ is the Laplace operator in $\boldsymbol{x}$, $s(t)$ is the source time function at $\boldsymbol{x}_s$ with compact support on $[0, \infty)$, and δ_d is the Dirac delta function.

Define the adjoint wavefield $u^\dagger(t,\boldsymbol{x})$ that satisfies the backward propagating wave equation,

$$\begin{cases} \partial_t^2 u^\dagger - c^2(\boldsymbol{x})\Delta_{\boldsymbol{x}} u^\dagger = s^\dagger(t)\delta_d(\boldsymbol{x}-\boldsymbol{x}_r), \quad \boldsymbol{x}=(x,y,z)\in\mathbb{R}^3, \\ u^\dagger(T,\boldsymbol{x}) = \partial_t u^\dagger(T,\boldsymbol{x}) = 0, \end{cases} \tag{10}$$

where $s^\dagger(t) = w(t)\partial_t u(t,\boldsymbol{x})/N_r$ with $w(t)$ as a weight function over $[0,T]$ and

$$N_r = \int_0^T w(t)u(t,\boldsymbol{x}_r)\partial_t^2 u(t,\boldsymbol{x}_r)\mathrm{d}t.$$

Then the Fréchet derivative of the travel-time δT_{obs} is (Dahlen *et al.*, 2000; Tromp *et al.*, 2005),

$$\begin{aligned} \delta T_{\text{sys}} &= \frac{1}{N_r}\int_0^T w(t)\partial_t u(t,\boldsymbol{x}_r)\delta u(t,\boldsymbol{x}_r)\mathrm{d}t \\ &= \int_0^T \frac{w(t)\partial_t u(t,\boldsymbol{x}_r)}{N_r}\int_\Omega\int_0^t G(t,\boldsymbol{x}_r;\tau,\boldsymbol{y})\,\nabla\cdot\Big(2c(\boldsymbol{y})\delta c(\boldsymbol{y})\nabla u(\tau,\boldsymbol{y})\Big)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}t \\ &= \int_0^T \mathrm{d}\tau\int_\Omega \mathrm{d}\boldsymbol{y}\int_\tau^T \mathrm{d}t\,\frac{w(t)\partial_t u(t,\boldsymbol{x}_r)}{N_r}G(t,\boldsymbol{x}_r;\tau,\boldsymbol{y})\,\nabla\cdot\Big(2c(\boldsymbol{y})\delta c(\boldsymbol{y})\nabla u(\tau,\boldsymbol{y})\Big) \\ &= -\int_0^T\int_\Omega 2\,\delta c(\boldsymbol{x})\,c(\boldsymbol{x})\,\nabla u^\dagger(\tau,\boldsymbol{x})\cdot\nabla u(\tau,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t \\ &=: \int_\Omega \mathcal{K}(\boldsymbol{x})\,\delta c(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \end{aligned} \tag{11}$$

which implies $\dfrac{\delta T_{\text{sys}}}{\delta c} = \mathcal{K}$ and then solving (8) is equivalent to find the least squares solution to the linear system

$$\delta T = \int_\Omega \mathcal{K}(\boldsymbol{x})\,\delta c(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}. \tag{12}$$

Remark 2.1 *1. The propagation of adjoint wavefield is defined as time-reversal in* (10)*, i.e. from time T to* 0*.*
2. The least squares problem (12) *is in general ill-posed, and instead one has to minimize the regularized objective function $\|\mathcal{K}\delta c - \delta T\|^2 + \epsilon\|\delta\boldsymbol{c}\|^2 + \eta\|D\delta\boldsymbol{c}\|^2$ with D as a first derivative smoothing operator and ϵ and η as two relaxation parameters.*

2.2 Full waveform inversion

In the full waveform inversion (FWI),

$$\boldsymbol{d}_{\text{obs}} = u_{\text{obs}}(t,\boldsymbol{x}_r),\ \boldsymbol{d}_{\text{sys}} = u(t,\boldsymbol{x}_r),\ \text{and}\ \langle\delta\boldsymbol{d},\delta\boldsymbol{d}\rangle = \int_0^T |\delta\boldsymbol{d}(t)|^2\,\mathrm{d}t.$$

Then the Fréchet derivative of the misfit functional $\delta\chi$ is

$$\begin{aligned}\delta\chi &= -\int_0^T (u_{\text{obs}}(t,\boldsymbol{x}_r) - u(t,\boldsymbol{x}_r))\delta u(t,\boldsymbol{x}_r)\mathrm{d}t \\ &= -\int_0^T [u_{\text{obs}} - u](t,\boldsymbol{x}_r) \int_\Omega \int_0^t G(t,\boldsymbol{x}_r;\tau,\boldsymbol{y})\, \nabla\cdot\Big(2c(\boldsymbol{y})\delta c(\boldsymbol{y})\nabla u(\tau,\boldsymbol{y})\Big)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}t \\ &= -\int_0^T [u_{\text{obs}} - u](t,\boldsymbol{x}_r) \int_\Omega \mathcal{J}(t,\boldsymbol{y})\,\delta c(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}t.\end{aligned}$$

then solving (8) is equivalent to find the least square solution to the linear system

$$u_{\text{obs}}(t,\boldsymbol{x}_r) - u(t,\boldsymbol{x}_r) = \int_\Omega \mathcal{J}(t,\boldsymbol{y})\delta c(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}. \tag{13}$$

Multiplying (13) by $u_{\text{obs}}(t,\boldsymbol{x}_r) - u(t,\boldsymbol{x}_r)$ and integrating over t produce

$$\begin{aligned}&\|u_{\text{obs}}(t,\boldsymbol{x}_r) - u(t,\boldsymbol{x}_r)\|_2^2 \\ &= \int_0^T [u_{\text{obs}} - u](t,\boldsymbol{x}_r) \int_\Omega \int_0^t G(t,\boldsymbol{x}_r;\tau,\boldsymbol{y})\, \nabla\cdot\Big(2c(\boldsymbol{y})\delta c(\boldsymbol{y})\nabla u(\tau,\boldsymbol{y})\Big)\,\mathrm{d}\tau\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}t \\ &= \int_0^T \mathrm{d}\tau \int_\Omega \mathrm{d}\boldsymbol{y} \int_\tau^T \mathrm{d}t\, [u_{\text{obs}} - u](t,\boldsymbol{x}_r) G(t,\boldsymbol{x}_r;\tau,\boldsymbol{y})\, \nabla\cdot\Big(2c(\boldsymbol{y})\delta c(\boldsymbol{y})\nabla u(\tau,\boldsymbol{y})\Big) \\ &= \int_0^T \mathrm{d}\tau \int_\Omega \mathrm{d}\boldsymbol{y}\, u^\dagger(\tau,\boldsymbol{y})\, \nabla\cdot\Big(2c(y)\delta c(y)\nabla u(\tau,\boldsymbol{y})\Big) \\ &= -\int_0^T \mathrm{d}\tau \int_\Omega \mathrm{d}\boldsymbol{y}\, 2\,\delta c(\boldsymbol{y})\, c(\boldsymbol{y})\, \nabla u^\dagger(\tau,\boldsymbol{y})\cdot\nabla u(\tau,\boldsymbol{y}),\end{aligned} \tag{14}$$

where $u^\dagger(t,\boldsymbol{x})$ satisfies (10) with the adjoint source function given by

$$s^\dagger(t,\boldsymbol{x}) = [u_{\text{obs}} - u](t,\boldsymbol{x}_r).$$

Define

$$\mathcal{K} = -\int_0^T 2c(\boldsymbol{y})\,\nabla u^\dagger(\tau,\boldsymbol{y})\cdot\nabla u(\tau,\boldsymbol{y})\mathrm{d}\tau,$$

then one can find the least squares solution to the following linear system for δc instead of using (13),

$$\|u_{\text{obs}}(t,\boldsymbol{x}_r) - u(t,\boldsymbol{x}_r)\|_2^2 = \int_\Omega \mathcal{K}(\boldsymbol{y})\delta c(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}. \tag{15}$$

3 Frozen Gaussian approximation

In this section, we introduce the formulation of frozen Gaussian approximation (FGA) to solve (9)-(10) following the notations used in Chai *et al.* (2017).

3.1 Reformulation of the forward and adjoint wavefields

We first reformulate (9)-(10) using Green's function, which will improve the efficiency of FGA algorithm described later. For simplicity, we restrict ourselves to the discussions on (9) only, and (10) can be treated in the essentially same way.

The solution to (9) can be written as

$$u(t,\boldsymbol{x}) = \int_0^t s(\tau) G(t,\boldsymbol{x};\tau,\boldsymbol{x}_s)\,\mathrm{d}\tau, \tag{16}$$

where $G(t,\boldsymbol{x};\tau,\boldsymbol{x}_s)$ is the Green's function given by

$$\begin{cases} \partial_t^2 G - c^2(\boldsymbol{x})\Delta_{\boldsymbol{x}} G = \delta_d(t-\tau)\delta_d(\boldsymbol{x}-\boldsymbol{x}_s), \\ G(\tau,\boldsymbol{x};\tau,\boldsymbol{x}_s) = 0, \\ \partial_t G(\tau,\boldsymbol{x};\tau,\boldsymbol{x}_s) = 0. \end{cases} \tag{17}$$

By Duhamel's principle, (17) is equivalent to

$$\begin{cases} \partial_t^2 G - c^2(\boldsymbol{x})\Delta_{\boldsymbol{x}} G = 0, \\ G(\tau,\boldsymbol{x};\tau,\boldsymbol{x}_s) = 0, \\ \partial_t G(\tau,\boldsymbol{x};\tau,\boldsymbol{x}_s) = \delta_d(\boldsymbol{x}-\boldsymbol{x}_s). \end{cases} \tag{18}$$

3.2 FGA algorithm for Green's function

In this subsection, we describe specific FGA formulation to approximate the Green's function in (18) by the following integral (Chai *et al.*, 2017),

$$u_{\mathrm{F}}^k(t,\boldsymbol{x}) = \sum_{\pm} \iint a_{\pm} \mathrm{e}^{\mathrm{i}k\boldsymbol{P}_{\pm}\cdot(\boldsymbol{x}-\boldsymbol{Q}_{\pm})-\frac{k}{2}|\boldsymbol{x}-\boldsymbol{Q}_{\pm}|^2}\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p}, \tag{19}$$

where $\mathrm{i} = \sqrt{-1}$ is the imaginary unit, the superscripts "+" and "−" indicate two wave branches respectively, and we use superscripts k to indicate quantities depending on wave number k. Remark that $u_{\mathrm{F}}^k(t,\boldsymbol{x})$ in (19) can be treated as an approximation to seismic wavefield by a summation of dynamic localized wave-packets with fixed width centered at $\boldsymbol{Q}$ with propagation vector $\boldsymbol{P}$. Associated with each frozen Gaussian wave-packet, the *time-dependent* quantities are the position center $\boldsymbol{Q}_{\pm}$, propagation vector $\boldsymbol{P}_{\pm}$, and amplitude $a_{\pm}$, which satisfy the following dynamic ODEs,

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{Q}_{\pm}}{\mathrm{d}t} = \partial_{\boldsymbol{P}_{\pm}} H_{\pm}, & \boldsymbol{Q}_{\pm}(0,\boldsymbol{q},\boldsymbol{p}) = \boldsymbol{q}, \\ \dfrac{\mathrm{d}\boldsymbol{P}_{\pm}}{\mathrm{d}t} = -\partial_{\boldsymbol{Q}_{\pm}} H_{\pm}, & \boldsymbol{P}_{\pm}(0,\boldsymbol{q},\boldsymbol{p}) = \boldsymbol{p}, \\ \dfrac{\mathrm{d}a_{\pm}}{\mathrm{d}t} = a_{\pm}\dfrac{\partial_{\boldsymbol{P}_{\pm}} H_{\pm}\cdot\partial_{\boldsymbol{Q}} H_{\pm}}{H_{\pm}} + \dfrac{a}{2}\mathrm{Tr}\left(Z_{\pm}^{-1}\dfrac{\mathrm{d}Z_{\pm}}{\mathrm{d}t}\right), & a_{\pm}(0,\boldsymbol{q},\boldsymbol{p}) = \omega_{\pm}^k(\boldsymbol{q},\boldsymbol{p}), \end{cases} \tag{20}$$

with $H_{\pm}(\boldsymbol{Q},\boldsymbol{P}) = \pm c(\boldsymbol{Q})|\boldsymbol{P}|$, and the shorthand notations

$$\partial_{\boldsymbol{z}} = \partial_{\boldsymbol{q}} - \mathrm{i}\partial_{\boldsymbol{p}}, \qquad Z_{\pm} = \partial_{\boldsymbol{z}}(\boldsymbol{Q}_{\pm} + \mathrm{i}\boldsymbol{P}_{\pm}), \tag{21}$$

and the initial condition $\omega_\pm^k$ for $a_\pm$ is given by the projection weight of initial wavefield onto each frozen Gaussian wave-packet $\mathrm{e}^{\mathrm{i}k\boldsymbol{p}\cdot(\boldsymbol{y}-\boldsymbol{q})-\frac{k}{2}|\boldsymbol{y}-\boldsymbol{q}|^2}$,

$$\omega_\pm^k(\boldsymbol{q},\boldsymbol{p}) = \frac{1}{2}\Big(\mathcal{T}[u_0](\boldsymbol{q},\boldsymbol{p}) \pm \frac{\mathrm{i}}{kc(\boldsymbol{q})|\boldsymbol{p}|}\mathcal{T}[u_1](\boldsymbol{q},\boldsymbol{p})\Big), \tag{22}$$

with $u(0,\boldsymbol{x}) = u_0(\boldsymbol{x})$ and $u_t(0,\boldsymbol{x}) = u_1(\boldsymbol{x})$. Here $\mathcal{T}[\cdot]$ is the local FBI transform (Martinez, 2002) defined as, for any function g,

$$\mathcal{T}[g](\boldsymbol{q},\boldsymbol{p}) = \frac{2^{3/2}}{(2\pi/k)^{9/2}}\int g(\boldsymbol{y})\mathrm{e}^{-\mathrm{i}k\boldsymbol{p}\cdot(\boldsymbol{y}-\boldsymbol{q})-\frac{k}{2}|\boldsymbol{y}-\boldsymbol{q}|^2}\,\mathrm{d}\boldsymbol{y},$$

and it satisfies

$$g(\boldsymbol{x}) = \iint \mathcal{T}[g](\boldsymbol{q},\boldsymbol{p})\mathrm{e}^{\mathrm{i}k\boldsymbol{p}\cdot(\boldsymbol{x}-\boldsymbol{q})-\frac{k}{2}|\boldsymbol{x}-\boldsymbol{q}|^2}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}.$$

When $u_0(\boldsymbol{x}) \equiv 0$ and $u_1(\boldsymbol{x}) = \delta_d(\boldsymbol{x}-\boldsymbol{x}_s)$ as in (18), $\omega_\pm^k$ can be computed explicitly by

$$\omega_\pm^k(\boldsymbol{q},\boldsymbol{p}) = \pm\mathrm{i}\frac{2^{3/2}}{(2\pi/k)^{9/2}}\frac{\mathrm{e}^{-\mathrm{i}k\boldsymbol{p}\cdot(\boldsymbol{x}_s-\boldsymbol{q})-\frac{k}{2}|\boldsymbol{x}_s-\boldsymbol{q}|^2}}{2kc(\boldsymbol{q})|\boldsymbol{p}|}, \tag{23}$$

which saves the computational cost required in standard FGA algorithm for general initial conditions (Chai *et al.*, 2017; Yang *et al.*, 2013).
In summary, the FGA algorithm consists of three steps to solve (9)-(10):

Step 1. *FGA propagation.* Solve (20) with the initial weight given in (23);

Step 2. *Computation of Green's function.* Approximate the high frequency Green's function by a numerical discretization of the integral in (19);

Step 3. *Computation of the forward and adjoint wavefields.* Convolve the Green's function with the source time functions $s(t)$ and $s^\dagger(t)$ to get the wavefields.

Remark 3.1 *To compute* (18) *on a bounded domain, one needs reflection and transmission conditions at interfaces of background media, free surface condition on top boundary, and transparent boundary condition on side and bottom boundaries, which are discussed in Section 4 of Chai* et al. *(2017) with a systematic derivation.*

3.3 Fast Gaussian summation algorithm

In Step 2 as described in Section 3.2, FGA computes the Green's function at time t by the following approximation,

$$u_{\mathrm{F}}^k(t,\boldsymbol{x}) \approx \sum_{\alpha\in\Xi} a_\alpha \mathrm{e}^{\mathrm{i}k\boldsymbol{P}_\alpha\cdot(\boldsymbol{x}-\boldsymbol{Q}_\alpha)-\frac{k}{2}|\boldsymbol{x}-\boldsymbol{Q}_\alpha|^2}\delta\boldsymbol{q}\delta\boldsymbol{p}, \tag{24}$$

where Ξ is a set for the triple index $\alpha = (\pm, m, n)$, $a_\alpha = a_\pm(t,\boldsymbol{q}_m,\boldsymbol{p}_n)$, $\boldsymbol{Q}_\alpha = \boldsymbol{Q}_\pm(t,\boldsymbol{q}_m,\boldsymbol{p}_n)$, $\boldsymbol{P}_\alpha = \boldsymbol{P}_\pm(t,\boldsymbol{q}_m,\boldsymbol{p}_n)$, and $\{\boldsymbol{q}_m\}$ and $\{\boldsymbol{p}_n\}$ are discretizations for $\boldsymbol{q}$ and $\boldsymbol{p}$, respectively. When

the number of α in Ξ is small, direct summation process in (24) can be fast and efficient. However, when one needs a well-resolved wavefield in space, FGA requires a large number (usually at the order of $\mathcal{O}(k^{3/2})$) of Gaussian for the summation, which makes direct summation extremely expensive.

Here we propose a fast Gaussian summation algorithm based on fast Fourier transform. If one focuses on one single frozen Gaussian indexed by α, by the FBI transform there exists a b_α such that

$$\mathrm{e}^{\mathrm{i}k\boldsymbol{P}_\alpha\cdot(\boldsymbol{x}-\boldsymbol{Q}_\alpha)-\frac{k}{2}|\boldsymbol{x}-\boldsymbol{Q}_\alpha|^2} = \iint b_\alpha(\boldsymbol{q},\boldsymbol{p})\mathrm{e}^{\mathrm{i}k\boldsymbol{p}\cdot(\boldsymbol{x}-\boldsymbol{q})-\frac{k}{2}|\boldsymbol{x}-\boldsymbol{q}|^2}\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p}, \tag{25}$$

where

$$\begin{aligned} b_\alpha(\boldsymbol{q},\boldsymbol{p}) &= \frac{2^{3/2}}{(2\pi/k)^{9/2}}\int \mathrm{e}^{\mathrm{i}k\boldsymbol{P}_\alpha\cdot(\boldsymbol{y}-\boldsymbol{Q}_\alpha)-\frac{k}{2}|\boldsymbol{y}-\boldsymbol{Q}_\alpha|^2}\mathrm{e}^{-\mathrm{i}k\boldsymbol{p}\cdot(\boldsymbol{y}-\boldsymbol{q})-\frac{k}{2}|\boldsymbol{y}-\boldsymbol{q}|^2}\,\mathrm{d}\boldsymbol{y} \\ &= \frac{2^{3/2}}{(2\pi/k)^{9/2}}\exp\left(-\frac{\mathrm{i}k}{2}(\boldsymbol{P}_\alpha+\boldsymbol{p})\cdot(\boldsymbol{Q}_\alpha-\boldsymbol{q})-\frac{k}{4}\left(|\boldsymbol{Q}_\alpha-\boldsymbol{q}|^2+|\boldsymbol{P}_\alpha-\boldsymbol{p}|^2\right)\right). \end{aligned} \tag{26}$$

Substituting (25) into (24) gives

$$\begin{aligned} u_{\mathrm{F}}^k(t,\boldsymbol{x}) &\approx \sum_{\alpha\in\Xi}\delta\boldsymbol{q}\delta\boldsymbol{p}\,a_\alpha\iint b_\alpha(\boldsymbol{q},\boldsymbol{p})\mathrm{e}^{\mathrm{i}k\boldsymbol{p}\cdot(\boldsymbol{x}-\boldsymbol{q})-\frac{k}{2}|\boldsymbol{x}-\boldsymbol{q}|^2}\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} \\ &= \delta\boldsymbol{q}\delta\boldsymbol{p}\iint\sum_{\alpha\in\Xi}a_\alpha b_\alpha(\boldsymbol{q},\boldsymbol{p})\mathrm{e}^{\mathrm{i}k\boldsymbol{p}\cdot(\boldsymbol{x}-\boldsymbol{q})-\frac{k}{2}|\boldsymbol{x}-\boldsymbol{q}|^2}\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} \\ &= \iint A(\boldsymbol{q},\boldsymbol{p})\mathrm{e}^{\mathrm{i}k\boldsymbol{p}\cdot(\boldsymbol{x}-\boldsymbol{q})-\frac{k}{2}|\boldsymbol{x}-\boldsymbol{q}|^2}\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p}, \\ &\approx \sum_{j,l}A(\boldsymbol{q}_j,\boldsymbol{p}_l)\mathrm{e}^{\mathrm{i}k\boldsymbol{p}_l\cdot(\boldsymbol{x}-\boldsymbol{q}_j)-\frac{k}{2}|\boldsymbol{x}-\boldsymbol{q}_j|^2}d\boldsymbol{q}d\boldsymbol{p}, \end{aligned} \tag{27}$$

where we have defined

$$A(\boldsymbol{q},\boldsymbol{p}) = \delta\boldsymbol{q}\delta\boldsymbol{p}\sum_{\alpha\in\Xi}a_\alpha b_\alpha(\boldsymbol{q},\boldsymbol{p}). \tag{28}$$

Notice that, (27) is in the inverse FWI form and thus can be numerically implemented by a local fast Fourier transform, provided $A(\boldsymbol{q}_j,\boldsymbol{p}_l)$. Moreover, for each $(\boldsymbol{q}_j,\boldsymbol{p}_l)$, the equation (26) implies that $b_\alpha(\boldsymbol{q}_j,\boldsymbol{p}_l)$ would be small if $(\boldsymbol{Q}_\alpha,\boldsymbol{P}_\alpha)$ is far away enough from $(\boldsymbol{q}_j,\boldsymbol{p}_l)$, and thus only a finite small number of b_α contribute to the summation in (28). These observations allow us to design a fast summation algorithm in calculating the wavefield in (24) by first doing a finite-band summation in (28) and then calculating (27) using local fast Fourier transform in the summation of $\boldsymbol{p}$.

4 Numerical examples

In this section, we apply frozen Gaussian approximation (FGA) to compute 3D finite-frequency banana-doughnut kernels and subsequently conduct travel-time adjoint tomography and full waveform inversion (FWI) on synthetic crosswell seismic data. We also numerically test the accuracy by applying FGA to local earthquake tomography for the synthetic data of the 1992 Landers earthquake sequence.

4.1 Crosswell seismic tomography

Crosswell seismic tomography is often used for high-resolution reservoir characterization in exploration geophysics. In the first example, we test the performance of FGA in seismic tomography with a crosswell setup. The three-layered velocity model is set up as follows, with a low-velocity region located at the second layer and homogeneity in the y-direction,

$$c_{\rm sys}(x,y,z) = c_{\rm sys}(x,z) = \begin{cases} C_1, & \text{if } z_0 < z < z_1, \\ C_2\left(1 - \alpha \mathrm{e}^{-\beta\left((x-x_c)^2+(z-z_c)^2\right)}\right), & \text{if } z_1 < z < z_2, \\ C_3, & \text{if } z > z_2, \end{cases} \tag{29}$$

where the background velocities $C_1 = 1800$ m/s, $C_2 = 2000$ m/s, $C_3 = 2200$ m/s, and the interfaces locate at $z_0 = 0$ m, $z_1 = 100$ m, $z_2 = 200$ m. The low-velocity region is centered at $x_c = 75$ m, $z_c = 150$ m, and $\beta = 1/450$ m^{-2}, and here α is a positive number to indicate the largest magnitude of the low-velocity perturbation from the background velocity; see Figure 1 for an illustration. In Figure 1, we also show the positions of 16 seismic resources (as stars) with an equal spacing of 16 m in one well and 32 seismic receivers (as dots) with an equal spacing of 8 m in the other well.

4.1.1 Least squares optimization with constraint of y-homogeneity

In this subsection, we solve the least squares optimization (12) and (15) with homogeneous constraint in y-direction, *i.e.* we only consider $c(x,y,z) = c(x,z)$. The synthetic seismograms are recorded at the receivers for each source-receiver pair and are used for $T_{\rm obs}$ and $u_{\rm obs}$ in the computation of (12) and (15) respectively. The forward, adjoint wavefields and kernel functions are computed by frozen Gaussian approximation (FGA) described in Section: 3.2. We choose the initial model as

$$c_0(x,y,z) = c_0(x,z) = \begin{cases} C_1, & \text{if } z_0 < z < z_1, \\ C_2, & \text{if } z_1 < z < z_2, \\ C_3, & \text{if } z > z_2. \end{cases} \tag{30}$$

We show the seismograms of dominant frequency around 543 Hz generated by an event at depth 161.3 m and recorded at depth 37.5 m and 277.3 m in Figure 2. We also plot the banana-doughnut kernels computed by FGA in Figure 3 for the initial velocity model (30).

The numerical experiment consists of three parts: (i) one-step iteration for $\alpha = 2\%$; (ii) one-step iteration for $\alpha = 5\%$; (iii) multiple iterations for $\alpha = 10\%$. The first two parts are numerical tests on small-perturbed velocity models. The purpose is to check the performance of FGA, travel-time tomography and full waveform inversion (FWI), with an emphasized comparison of travel-time and FWI on the dependence of convergence on the quality of initial velocity model. Then we perform a full iteration on a 10%-perturbation model, i.e. $\alpha = 10\%$ in (29), with a combination of travel-time tomography and FWI. The least squares problems (12) and (15) are solved by minimizing the following misfit function:

$$\min_{\boldsymbol{c}(x,z)} \|\mathcal{K}\delta\boldsymbol{c} - \delta\boldsymbol{d}\|^2 + \epsilon\|\delta\boldsymbol{c}\|^2 + \eta\|D\delta\boldsymbol{c}\|^2, \tag{31}$$

where $\boldsymbol{c}$ stands for the discretized value of c on x, z-grids, $\delta\boldsymbol{d}$ equals to δT and $\|u_{\text{obs}}(t, \boldsymbol{x}_r) - u(t, \boldsymbol{x}_r)\|_2$ for travel-time tomography and FWI respectively, D is a first derivative smoothing operator, and ϵ and η are two relaxation parameters.

$\alpha = 2\%$ *model*. We perform the travel-time tomography and FWI in the velocity model as described in (29) with $\alpha = 2\%$. We take $\epsilon = 0.076$ and $\eta = 0.005$ for travel-time tomography, and $\epsilon = 0.001$ and $\eta = 0.00003$ for FWI. Note that these parameters are chosen empirically in order to get the best image as we can. The one-step inversion results are given in Figure 4, where one observes that both travel-time tomography and FWI converges but FWI produces less artifacts than travel-time tomography.

$\alpha = 5\%$ *model*. We perform the travel-time tomography and FWI in the velocity model as described in (29) with $\alpha = 5\%$ using the initial model (30). The one-step inversion results are shown in Figure 5, from which one can see that when the initial model is not close enough to the true velocity model (29), FWI does not converge in the correct direction while travel-time tomography still shows convergence.

By comparing the performances of travel-time tomography and FWI in the 2%- and 5%-models, we notice that FWI provides better accuracy with fewer artifacts but requires more sophisticated initial model for convergence. Therefore, for the 10%-*model*, we use the travel-time tomography for the first few iterations until the results get closer to the true velocity, and then perform the FWI for later iterations to refine the results. We start the iterations with the initial model c_0 given in (30), and perform three travel-time tomography iterations to get models c_1, c_2, and c_3, as shown in Figure 6 (a), (b), and (c), respectively. One observes that the travel-time tomography captures the low-velocity region and improves the result in each iteration. Then from model c_3, we perform both the travel-time tomography and the full waveform inversion, to get model c_4^{TI} and c_4^{FWI}, which are shown in Figure 6 (d) and (e) respectively. Notice that there is not much significant improvement from c_3 to c_4^{TI}, which indicates that the travel-time tomography has converged. On the other hand, performing the FWI from model c_3 to c_4^{FWI}, removes some artifacts and further refines the result. The damping and smoothing parameters ε and η are chosen empirically by forcing the variation in each iteration less than 3%. Note that artifacts are visible in the final images, which are unavoidable and mainly caused by the uneven data coverages.

4.1.2 3D least squares optimization

In this subsection, we study the full 3D least squares optimization by considering c as a function of (x, y, z). We make 16 copies of source-receiver pairs used in the previous crosswell seismic tomography in the y-direction, and consider (12) using travel-time tomography, i.e., to solve

$$\min_{\boldsymbol{c}(x,y,z)} \|\mathcal{K}\delta\boldsymbol{c} - \delta T\|^2 + \epsilon\|\delta\boldsymbol{c}\|^2 + \eta\|D\delta\boldsymbol{c}\|^2, \tag{32}$$

where $\boldsymbol{c}$ stands for the 3D discretized value of c on x, y, z-grids. The 3D inversion result is shown in Figure 7. The wiggling structures in y-direction indicate the 3D effect when one uses duplicated source-receiver pairs instead of assuming the velocity is homogeneous in y.

4.2 Local earthquake tomography

Local earthquake tomography has been widely used to investigate the heterogeneous structures of the crust and upper mantle in seismically active regions (e.g. Aki and Lee, 1976; Chen *et al.*, 2007a; Tong *et al.*, 2017). Since the finite-frequency effects and waveform information are considered, using travel-time adjoint tomography or FWI rather than the conventional ray-based inversion in local earthquake tomography studies can potentially generate higher-resolution images of the subsurface structures (e.g. Dahlen *et al.*, 2000; Virieux and Operto, 2009). As a performance test for the future use, we conduct a synthetic local earthquake tomography investigation using the travel-time adjoint tomography and full waveform inversion together with the FGA modeling.

In this example, we choose 64 earthquakes and 16 receivers in the investigated region, for which the distributions are shown in Figure 8. The relative locations of these earthquakes and receivers resemble the distributions of 64 earthquakes and 16 seismic stations in the source area of the 1992 Landers earthquake (Tong *et al.*, 2014c). For simplicity, the synthetic velocity model is chosen as follows

$$c_{\text{sys}}(x,y,z) = c_{\text{sys}}(x,z) = C_1 \tanh(\gamma(z-z_c)) + C_2 \left(1 - \alpha \mathrm{e}^{-\left(\beta_1 (x-x_c)^2 + \beta_2 (z-z_c)^2\right)}\right), \qquad (33)$$

where the $C_1 = 0.25$ km/s, $C_2 = 6.0$ km/s, $x_c = 75$ km, $z_c = 10$ km, $\gamma = 0.4\ \text{km}^{-1}$, $\beta_1 = 1/32\ \text{km}^{-2}$ and $\beta_2 = 1/8\ \text{km}^{-2}$, and α is selected as 0.1 so that the peak relative velocity perturbation is 10%; see Figure 8 for an illustration. The dominant frequency of the using data is 6.1 Hz. Travel-time tomography and FWI are applied by using all the 1024 first arrivals, with dominant frequency 6.1 Hz. The initial model is chosen as (33) with $\alpha = 0$. We then perform 3 travel-time tomography iterations to get models c_1, c_2, and c_3, as shown in Figure 9 (a), (b), and (c), respectively. One can see the travel-time tomography captures the low-velocity region but is not good enough in getting the accurate velocity model. On the other hand, as shown in Figure 9 (d), an FWI iteration from model c_2 to c_3^{FWI} improves the result and gives a better quality in the inversion than that of the travel-time tomography.

5 Discussion and conclusions

Traveltime adjoint tomography and full waveform inversion are able to generate higher-resolution images of the Earth's interior than the conventional ray-based tomography method (Virieux and Operto, 2009). However, mainly restricted by the drastic computational cost, these wave-equation-based inversion methods have not been widely used so far, and in some real applications only low-frequency data are modeled and inverted (e.g. Tape *et al.*, 2007; Zhu *et al.*, 2012). To address the computation challenge, Chai *et al.* (2017) developed the FGA method to model seismic wave propagation and compute sensitivity kernels. The FGA method is computationally efficient and ideal for high-frequency wave simulation (Chai *et al.*, 2017).

In this study, the FGA method is reformulated and applied to travel-time adjoint tomography and full waveform inversion. Synthetic examples are given to show the feasibility of using the FGA method in these wave-equation-based inversions. A three-dimensional investigation is conducted. The dominant frequencies of the synthetic data mimic the real

dominant frequencies of typical crosswell data and local earthquakes. These are impossible if one solves full wave equations using a finite-difference or spectral-element method. In both crosswell seismic tomography and local earthquake tomography examples, artifacts are visible in the final images. These artifacts are unavoidable and mainly caused by the uneven data coverages. Our synthetic examples also confirm that travel-time inversion has a wider convergence domain than full waveform inversion does, but within the convergence domain, full waveform inversion has a better performance in suppressing artifacts. This suggests a hierarchical approach which first uses travel-time adjoint tomography to create a macro-scale model and then adopts full waveform inversion to generate a high-resolution micro-scale model. Our next plan will be the application of travel-time adjoint tomography and full waveform inversion with the FGA method into real data around their dominant frequencies.

Acknowledgments

We acknowledge support from the Center for Scientific Computing from the CNSI, MRL: an NSF MRSEC (DMR-1121053) and NSF CNS-0960316. LC and XY were partially supported by the NSF grants DMS-1418936 and DMS-1107291.

References

Aki K and Lee W 1976 Determination of the three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes 1. A homogeneous intial model *J. Geophys. Res.* **81** 4381–4399

Chai L, Tong P and Yang X 2017 Frozen Gaussian approximation for 3-D seismic wave propagation *Geophysical Journal International* **208** 59–74

Chen P, Jordan T H and Zhao L 2007a Full 3d waveform tomography: A comparison between the scattering-integral and adjoint-wavefield methods *Geophys. J. Int.* **170** 175–181

Chen P, Zhao L and Jordan T H 2007b Full 3D tomography for the crustal structure of the Los Angeles region *Bull. Seismol. Soc. Am.* **97** 1094–1120

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

Fichtner A, Igel H, Bunge H P and Kennett B L N 2009 Simulation and inversion of seismic wave propagation on continental scales based on a spectral-element method *Journal of Numerical Analysis, Industrial and Applied Mathematics* **4** 11–22

Huang X, Yang D, Tong P, Badal J and Liu Q 2016 Wave equation-based reflection tomography of the 1992 landers earthquake area *Geophysical Research Letters* **43** 1884–1892

Hung S, Shen Y and Chiao L 2004 Imaging seismic velocity structure beneath the iceland hot spot: a finite frequency approach *J. Geophys. Res.* **109** B08305

Komatitsch D and Tromp J 2002 Spectral-element simulations of global seismic wave propagation-I.Validation *Geophys. J. Int.* **149** 390–412

Li J, Lin G and Yang X 2015 A frozen gaussian approximation-based multi-level particle swarm optimization for seismic inversion *J. Comput. Phys.* **296** 58–71

Liu Q and Gu Y J 2012 Seismic imaging: From classical to adjoint tomography *Tectonophysics* **566-567** 31–66

Liu Q and Tromp J 2008 Finite-frequency Sensitivity Kernels for Global Seismic Wave Propagation based upon Adjoint Methods *Geophys. J. Int.* **174** 265–286

Lu J and Yang X 2011 Frozen Gaussian approximation for high frequency wave propagation *Commun. Math. Sci.* **9** 663–683

Lu J and Yang X 2012a Convergence of frozen Gaussian approximation for high frequency wave propagation *Comm. Pure Appl. Math.* **65** 759–789

Lu J and Yang X 2012b Frozen Gaussian approximation for general linear strictly hyperbolic systems: Formulation and Eulerian methods *Multiscale Model. Simul.* **10** 451–472

Martinez A 2002 *An introduction to semiclassical and microlocal analysis* (New York: Springer-Verlag)

Montelli R, Nolet G, Dahlen A, Masters G, Robert Engdahl E and Hung S 2004 Finite-frequency tomography reveals a variety of plumes in the mantle *Science* **303** 338–343

Nakamichi H, Hamaguchi H, Tanaka S, Ueki S, Nishimura T and Hasegawa A 2003 Source mechanisms of deep and intermediate-depth low-frequency earthquakes beneath iwate volcano, northeastern japan *Geophysical Journal International* **154** 811–828

Pratt R G and Shipp R M 1999 Seismic waveform inversion in the frequency domain, part 2: fault delineation in sediments using crosshole data *Geophysics* **64** 902–914

Rawlinson N, Pozgay S and Fishwick S 2010 Seismic tomography: A window into deep Earth *Phys. Earth Planet. Inter.* **178** 101–135

Rickers F, Fichtner A and Trampert J 2013 The Iceland-Jan Mayen plume system and its impact on mantle dynamics in the North Atlantic region: Evidence from full-waveform inversion *Earth Planet. Sci. Lett.* **367** 39–51

Romanowicz B 1991 Seismic tomography of the Earth's mantle *Annu. Rev. Earth Planet. Sci.* **19** 77–99

Tape C, Liu Q, Maggi A and Tromp J 2009 Adjoint tomography of the southern California crust *Science* **325** 988–992

Tape C, Liu Q, Maggi A and Tromp J 2010 Seismic tomography of the southern California crust based on spectral-element and adjoint methods *Geophys. J. Int.* **180** 433–462

Tape C, Liu Q and Tromp J 2007 Finite-frequency tomography using adjoint methods-Methodology and examples using membrane surface waves *Geophys. J. Int.* **168** 1105–1129

Tong P, Chen C W, Komatitsch D, Basini P and Liu Q 2014a High-resolution seismic array imaging based on an SEM-FK hybrid method *Geophys. J. Int.* **197** 369–395

Tong P, Yang D, Li D and Liu Q 2017 Time-evolving seismic tomography: The method and its application to the 1989 loma prieta and 2014 south napa earthquake area, california *Geophysical Research Letters* **44** 3165–3175

Tong P, Zhao D and Yang D 2011 Tomography of the 1995 kobe earthquake area: comparison of finite-frequency and ray approaches *Geophys. J. Int.* **187** 278–302

Tong P, Zhao D, Yang D, Yang X, Chen J and Liu Q 2014b Wave-equation-based travel-time seismic tomography—Part 1: Method *Solid Earth* **5** 1151–1168

Tong P, Zhao D, Yang D, Yang X, Chen J and Liu Q 2014c Wave-equation-based travel-time seismic tomography—Part 2: Application to the 1992 Landers earthquake area *Solid Earth* **5** 1169–1188

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

Virieux J and Operto S 2009 An overview of full-waveform inversion in exploration geophysics *Geophysics* **74** WCC1–WCC26

Yang X, Lu J and Fomel S 2013 Seismic modeling using the frozen Gaussian approximation *SEG Technical Program Expanded Abstracts 2013* 4677–4682

Zhao D 2012 Tomography and dynamics of Western-Pacific subduction zones *Monogr. Environ. Earth Planets* **1** 1–70

Zhu H, Bozdag E, Peter D and Tromp J 2012 Structure of the European upper mantle revealed by adjoint tomography *Nature Geoscience* **5** 493–498

List of Figures

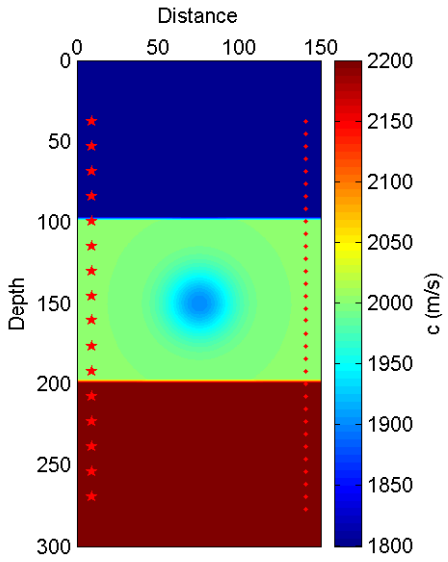

Figure 1: The three-layered crosswell velocity model with homogeneity in horizontal y-direction. The background velocities from top to bottom are $C_1 = 1800$ m/s, $C_2 = 2000$ m/s, $C_3 = 2200$ m/s, and the interfaces locate at $z_0 = 0$ m, $z_1 = 100$ m, $z_2 = 200$ m. The low-velocity region has a Gaussian shape centered at $x_c = 75$ m, $z_c = 150$ m, with standard deviation equal to 1/30 m. 16 stars indicate the locations of seismic sources, and 32 dots indicate the locations of seismic receivers.

(a) (b)

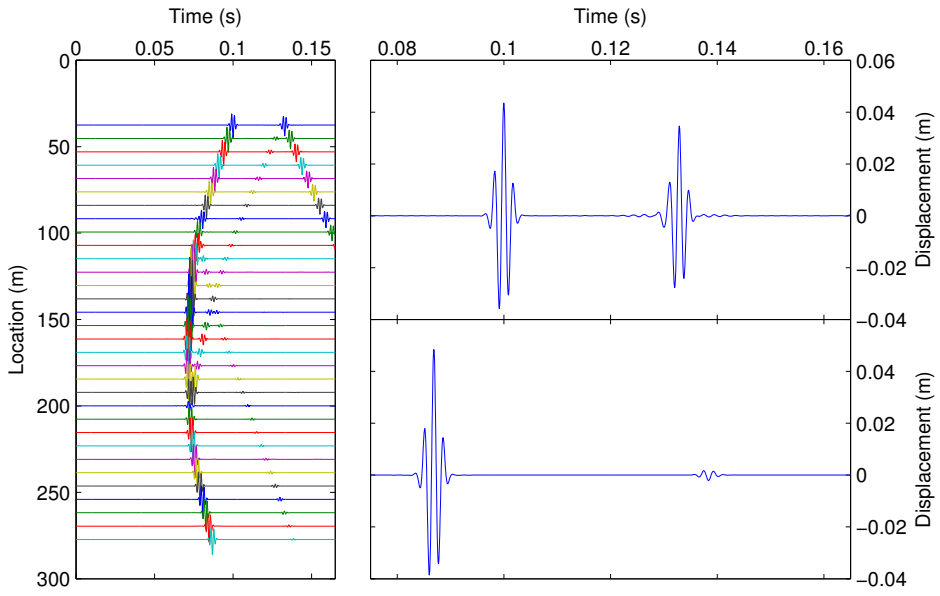

Figure 2: (a) Seismograms generated by an event at depth 161.3 m (middle layer) with the dominant frequency around 543 Hz. (b) Seismograms recorded at depth 37.5 m (top) and 277.3 m (bottom).

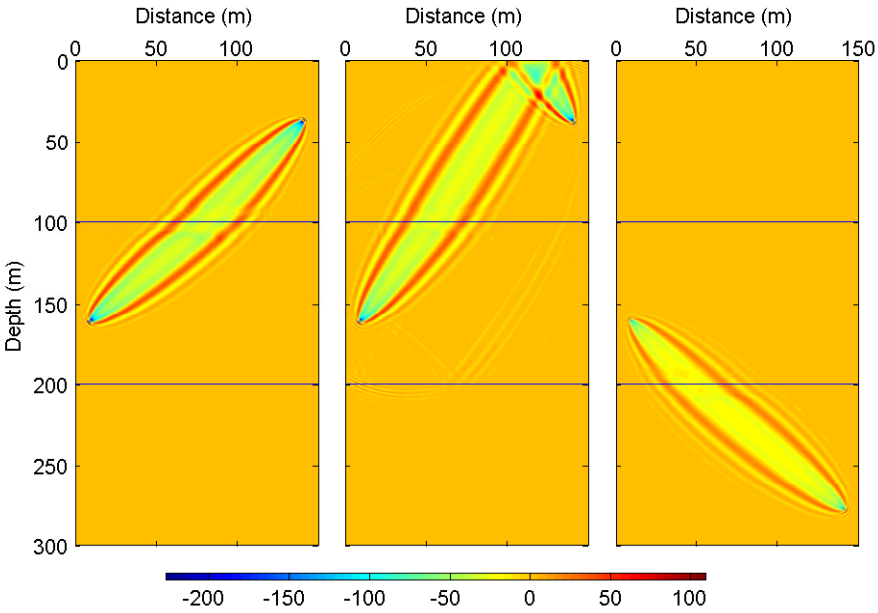

Figure 3: Banana-doughnut kernels computed by FGA for the initial velocity model (30). Left: Kernel computed from the first arrival signal in top of Figure 2(b); Middle: Kernel computed from the second arrival signal in top of Figure 2(b); Right: Kernel computed from the first arrival signal in bottom of Figure 2(b);

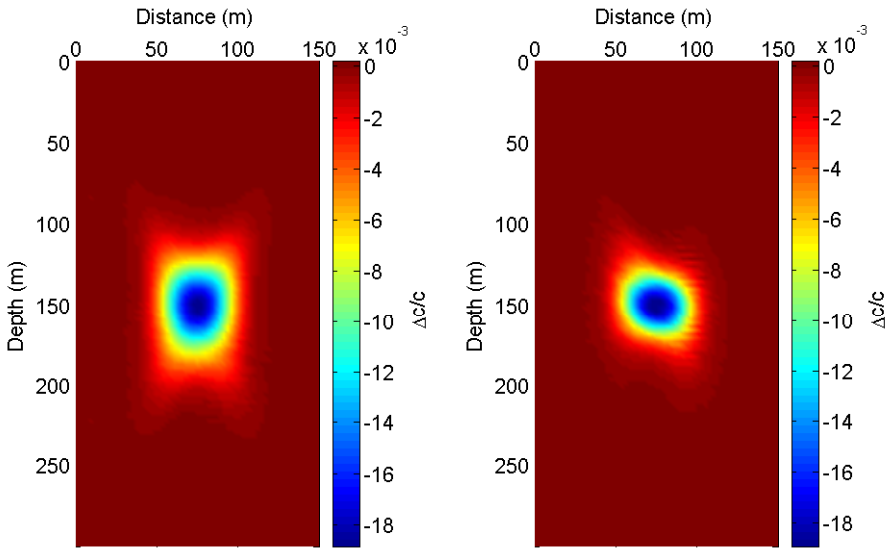

Figure 4: One-step iteration of travel-time tomography (left) and FWI (right) for $\alpha = 2\%$ using the initial model by equation (30). The least squares problem (31) is solved with smoothed parameters $\epsilon = 0.076$ and $\eta = 0.005$ for travel-time tomography, and $\epsilon = 0.001$ and $\eta = 0.00003$. Both travel-time tomography and FWI converges in the correct direction but FWI produces less artifacts than travel-time tomography.

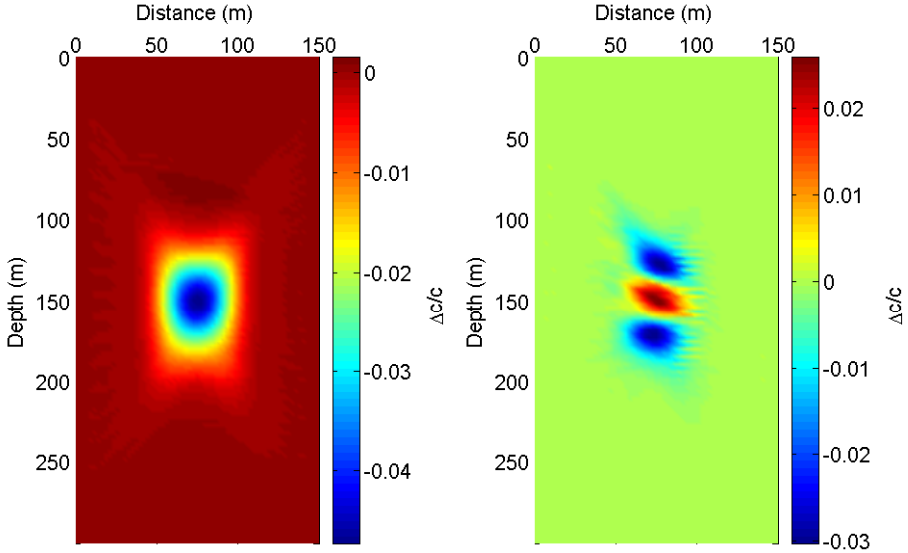

Figure 5: One-step iteration of travel-time tomography (left) and FWI (right) for $\alpha = 5\%$. The least squares problem (31) is solved with smoothed parameters $\epsilon = 0.076$ and $\eta = 0.005$ for travel-time tomography, and $\epsilon = 0.001$ and $\eta = 0.00003$. FWI does not converge in the correct direction while travel-time tomography shows convergence for the initial model (30).

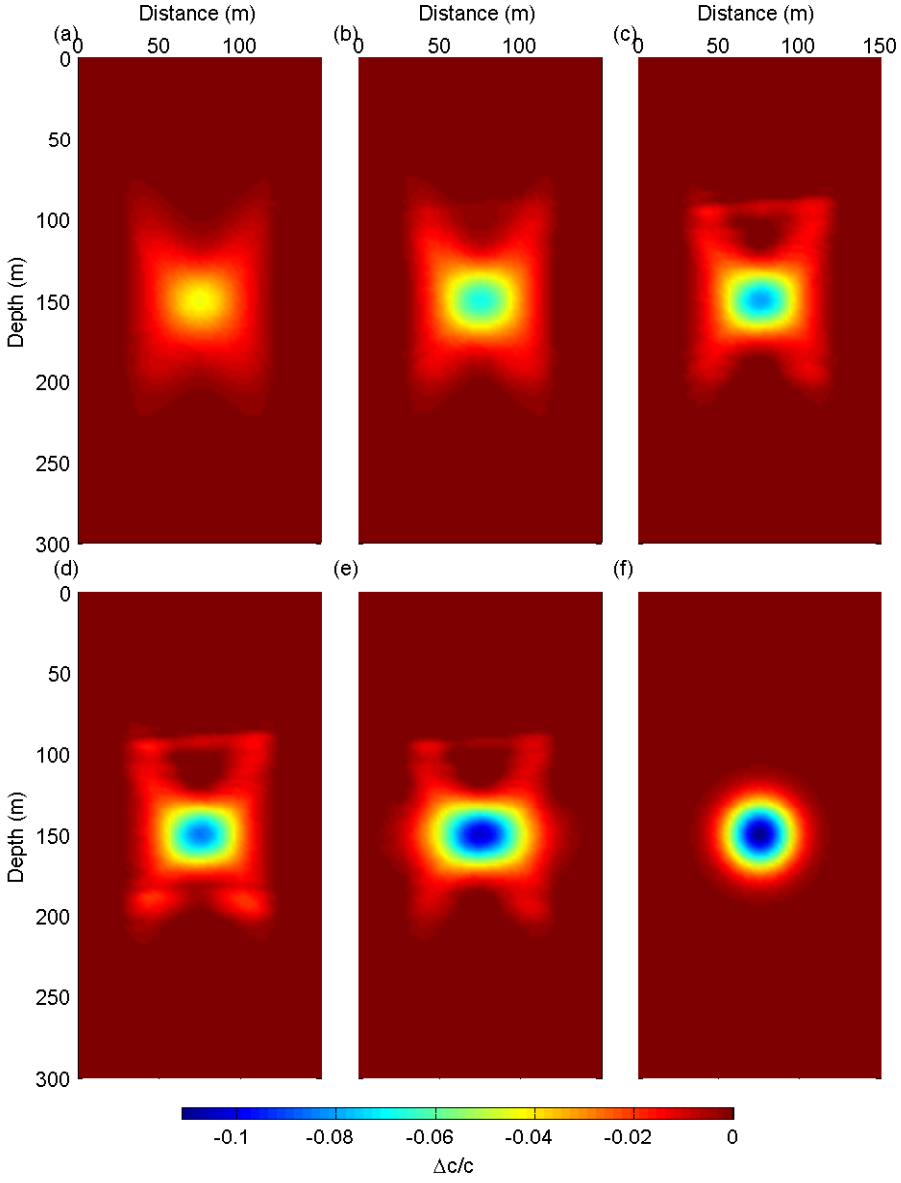

Figure 6: Inversion results for the 10%−model. (a),(b), and (c) show the velocity models provided by the first three iterations by travel-time tomography; (d) plots the velocity model of the fourth iteration by travel-time tomography; (e) gives the velocity model of the fourth iteration by FWI which removes the artifacts in travel-time tomography; (f) is true velocity model (29) with $\alpha = 10\%$.

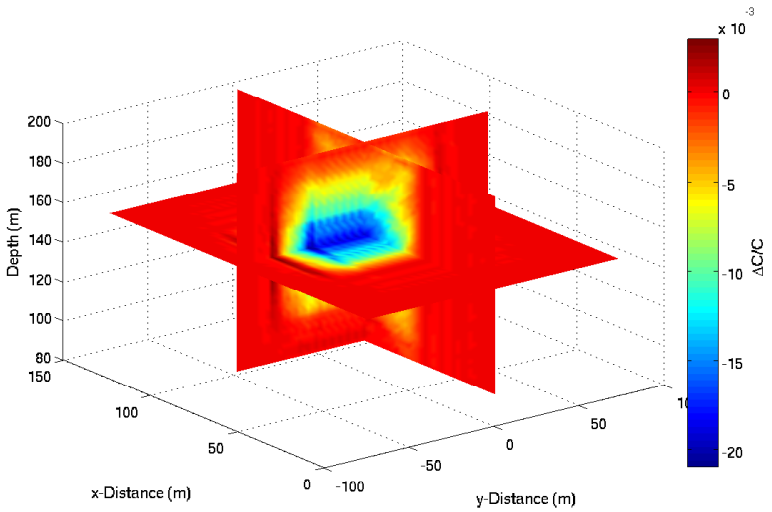

Figure 7: The 3D velocity model produced by travel-time tomography by solving the least squares problem (32) without homogeneity constraint in y. The wiggling structures in y-direction indicate the effect of homogeneity constraint on y when replaced by source-receiver pairs in y-direction.

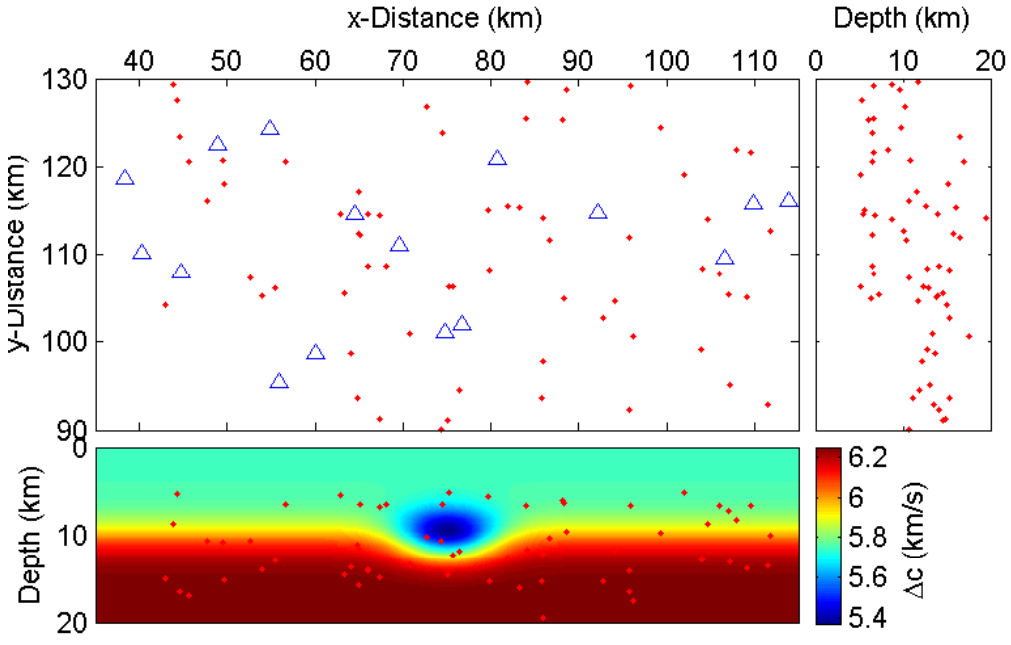

Figure 8: Distributions of the 64 earthquakes (dots) and 16 receivers (triangles).

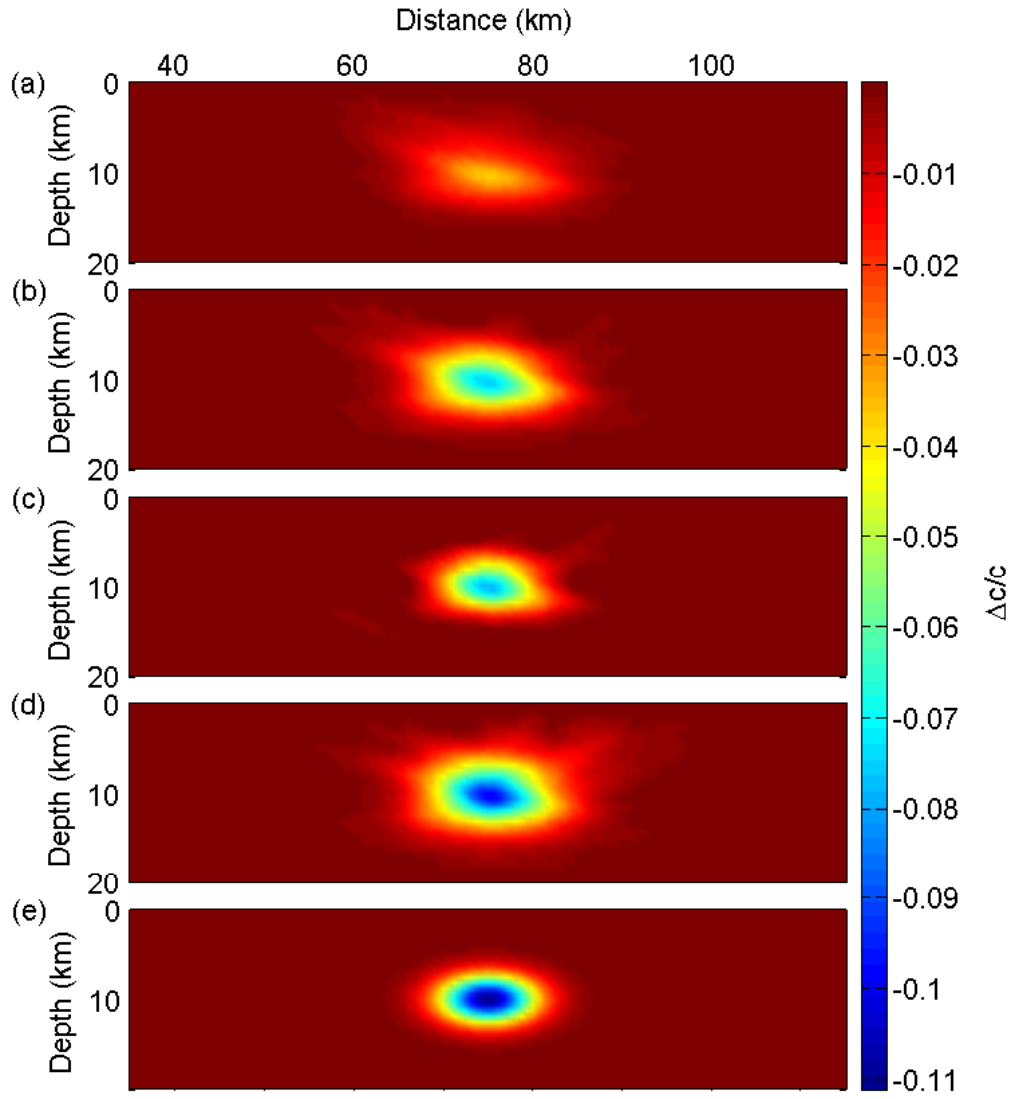

Figure 9: Inversion results for the Landers. (a) and (b) show the velocity models provided by the first two iterations by travel-time tomography; (c) plots the velocity model of the third iteration by travel-time tomography; (d) gives the velocity model of the third iteration by FWI; (e) is true velocity model given by (33).