Frozen Gaussian Approximation for 3-D Elastic Wave Equation and Seismic Tomography

J. C. Hateley[1], L. Chai[2], P. Tong[3], X. Yang[1]

[1] *Department of Mathematics, University of California, Santa Barbara, CA 93106, USA*

[2] *School of Mathematics, Sun Yat-sen University, Guangzhou, 510275, China*

[3] *School of Physical and Mathematical Sciences & Asian School of the Environment,*

Nanyang Technological University, 637371, Singapore

SUMMARY

The purpose of this work is to generalize the frozen Gaussian approximation (FGA) theory to solve the 3-D elastic wave equation and use it as the forward modeling tool for seismic tomography with high-frequency data. FGA has been previously developed and verified as an efficient solver for high-frequency acoustic wave propagation (P-wave). The main contribution of this paper consists of three aspects: 1. We derive the FGA formulation for the 3-D elastic wave equation. Rather than standard ray-based methods (e.g. geometric optics and Gaussian beam method), the derivation requires to do asymptotic expansion in the week sense (integral form) so that one is able to perform integration by parts. Compared to the FGA theory for acoustic wave equation, the calculations in the derivation are much more technically involved due to the existence of both P- and S-waves, and the coupling of the polarized directions for SH- and SV-waves. In particular, we obtain the diabatic coupling terms for SH- and SV-waves, with the form closely connecting to the concept of Berry phase that is intensively studied in quantum mechanics and topology (Chern number). The accuracy and parallelizability of the FGA algorithm is illustrated by comparing to the spectral element method for 3-D elastic wave equa-

tion in homogeneous media; 2. We derive the interface conditions of FGA for 3-D elastic wave equation based on an Eulerian formulation and the Snell's law. We verify these conditions by simulating high-frequency elastic wave propagation in a 1-D layered Earth model. In this example, we also show that it is natural to apply the FGA algorithm to geometries with non-Cartesian coordinates; 3. We apply the developed FGA algorithm for 3-D seismic wave-equation-based traveltime tomography and full waveform inversion, respectively.

Key words: Frozen Gaussian approximation; elastic wave equation; seismic tomography; Wasserstein distance; high-frequency wavefield; weak asymptotic analysis.

1   INTRODUCTION

Images computed by seismic tomography can provide crucial information for the subsurface structures of Earth at different scales, and the understanding of tectonics, volcanism, and geodynamics (e.g. Aki & Lee 1976; Romanowicz 1991; Rawlinson et al. 2010; Zhao 2012). Wave-equation-based seismic tomography solves the nonlinear optimization problem iteratively for velocity models by computing seismograms and sensitivity kernels in 3-D complex models (Tromp et al. 2005; Liu & Tromp 2008; Liu & Gu 2012; Tong et al. 2014). This leads to successful applications including imaging the velocity models of the southern California crust (Tape et al. 2009, 2010), the European upper mantle (Zhu et al. 2012), the North Atlantic region (Rickers et al. 2013), and the Japan islands (Simute et al. 2016). The performance of seismic tomography is restricted by how accurate one can solve the 3-D wave equation for synthetic seismograms and sensitivity kernels. The dominant frequency of a low-frequency earthquake is around 1 Hz, while the one of a typical earthquake is around 5 Hz (e.g. Nakamichi et al. 2003), which leads to demanding and even unaffordable computational cost. Real applications usually chose significantly low-frequency data in order to be computationally feasible (e.g. Tape et al. 2009; Zhu et al. 2012; Simute et al. 2016), since simulating high-frequency seismic waves requires much more powerful computational resources on both memories and CPU time than the low-frequency waves. This imposes a demand on improving the efficiency and accuracy of numerical methods for computing high-frequency waves in order to make use of real seismic data around their dominant frequencies.

In previous works (Chai et al. 2017, 2018), the authors have developed and verified frozen Gaussian approximation (FGA) as an efficient solver for computing high-frequency acoustic wave propagation (P-wave), wave-equation-based traveltime tomography and full waveform inversion (FWI). The

core idea of FGA is to approximate seismic wavefields by fixed-width Gaussian wave packets, whose dynamics follow ray paths with the prefactor amplitude equation derived delicately from an asymptotic expansion on phase plane. With a multicore processors computer station, the property that the FGA algorithm is embarrassingly parallel makes possible the application of FGA to compute 3-D high frequency sensitivity kernels, and further used for 3-D traveltime tomography and FWI. Compared to other ray-based methods including WKBJ (e.g. Chapman 1976; Chapman & Drummond 1982), WKM (e.g. Ni et al. 2000; Helmberger & Ni 2005), generalized ray theory (e.g. Helmberger 1968; Vidale & Helmberger 1988), seismic traveltime tomography (e.g. Aki et al. 1977; Tong et al. 2017), Kirchhoff migration (e.g. Gray 1986; Keho & Beydoun 1988) and Gaussian beam migration (e.g. Hill 1990, 2001; Nowack et al. 2003; Gray 2005; Gray & Bleistein 2009; Popov et al. 2010), FGA does not need to solve ray paths by shooting to reach the receivers, and provides accurate solutions at the presence of caustics and multipathing, with no requirement on tuning beam width parameters to achieve a good resolution (Cerveny et al. 1982; Hill 1990; Fomel & Tanushev 2009; Qian & Ying 2010; Lu & Yang 2011).

In this paper, we first generalize the FGA theory to solve the 3-D elastic wave propagation. Different from the WKBJ theory or Gaussian beam method which can be derived by direct asymptotic expansion, the derivation of FGA formulation requires to do the asymptotic expansion in an integral form (weak sense) so that one is able to perform integration by parts to eliminate the extra constraints yielded by direct asymptotic expansion. Compared to the previous works on FGA (Lu & Yang 2012a; Chai et al. 2017, 2018), the calculations for the derivation are much more technically involved due to the existence of both P- and S-waves. Particularly, the diabatic coupling of the polarized directions for SH- and SV-waves leads to a term closely connected to the concept of Berry phase which is intensively studied in quantum mechanics and topology (Chern number) (e.g. Berry 1984; Simon 1983). We prove the accuracy and parallelizability of the FGA algorithm by comparing to the spectral element method (e.g. Komatitsch & Tromp 1999; Komatitsch et al. 2005; Tromp et al. 2008) for the 3-D elastic wave equation in homogeneous media, where one has the analytical solution as a benchmark. We derive the interface conditions of FGA for the 3-D elastic wave equation based on an Eulerian formulation and the Snell's law. These interface conditions are verified by simulating high-frequency elastic wave propagation in a 1-D layered Earth model. We also show how natural to apply the FGA algorithm to geometries with non-Cartesian coordinates in the simulation of this model.

The second part of this paper contributes to the study of seismic tomography with high-frequency data. Wave-equation-based traveltime tomography and FWI are able to generate higher-resolution images of the Earth's interior than the conventional ray-based tomography method (Virieux & Operto 2009). Especially, some recent novel studies on FWI using optimal transport distance (e.g. Wasserstein

metric) are able to capture traveltime differences between seismic signals, and thus overcomes the cycling effect in standard FWI method (e.g. Engquist & Froese 2014; Métivier et al. 2016; Yang et al. 2018; Chen et al. 2017). However, seismic tomography using high-frequency data increases computational cost drastically, which restricts the application of these wave-equation-based inversion methods, and only low-frequency data are modeled and inverted in some real applications (e.g. Tape et al. 2007; Zhu et al. 2012; Simute et al. 2016). To address the computation challenge, we use the developed FGA algorithm to compute the 3-D high-frequency sensitivity kernels for wave-equation-based traveltime tomography and FWI, respectively. We apply FGA to both wave-equation-based traveltime tomography and FWI on synthetic crosswell seismic data with dominant frequencies as high as those of real crosswell data, and use a hierarchical approach which first uses traveltime tomography to create a macro-scale model and then adopts FWI to generate a high-resolution micro-scale model.

2   FROZEN GAUSSIAN APPROXIMATION

We derive the frozen Gaussian approximation (FGA) formulation for the following inhomogeneous, isotropic elastic wave equation:

$$\rho(\boldsymbol{x})\partial_t^2\boldsymbol{u} = \big(\lambda(\boldsymbol{x}) + \mu(\boldsymbol{x})\big)\nabla(\nabla \cdot \boldsymbol{u}) + \mu(\boldsymbol{x})\Delta\boldsymbol{u} + \mathbf{F}(t,\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^3, \tag{1}$$

where $\boldsymbol{u}(t,\boldsymbol{x})$ is the wavefield, $\rho(\boldsymbol{x})$ is the density, $\lambda(\boldsymbol{x})$ and $\mu(\boldsymbol{x})$ are the Lamé parameters, and $\mathbf{F}(t,\boldsymbol{x})$ is the body force. We shall assume $\mathbf{F} = 0$ in this section as a sake of simplicity for presenting the FGA formula. Remark that eq. (1) can be rewritten as,

$$\partial_t^2\boldsymbol{u} = c_{\mathrm{p}}^2\nabla(\nabla \cdot \boldsymbol{u}) - c_{\mathrm{s}}^2\nabla \times \nabla \times \boldsymbol{u} + \frac{\boldsymbol{F}}{\rho}, \tag{2}$$

with the P-wave, and S-wave speeds defined as,

$$c_{\mathrm{p}}^2(\boldsymbol{x}) = \frac{\lambda(\boldsymbol{x}) + 2\mu(\boldsymbol{x})}{\rho(\boldsymbol{x})} \qquad c_{\mathrm{s}}^2(\boldsymbol{x}) = \frac{\mu(\boldsymbol{x})}{\rho(\boldsymbol{x})}. \tag{3}$$

We consider the elastic wave eq. (1), with the following initial conditions

$$\begin{cases} \boldsymbol{u}(0,\boldsymbol{x}) = \boldsymbol{f}^k(\boldsymbol{x}), \\ \partial_t\boldsymbol{u}(0,\boldsymbol{x}) = \boldsymbol{g}^k(\boldsymbol{x}). \end{cases} \tag{4}$$

Note that all the bolded variables and parameters will be referred to vectors in $\mathbb{R}^3$ without any further clarification.

2.1   Solution Ansatz

The FGA approximates the wavefield $\boldsymbol{u}(t, \boldsymbol{x})$ in eq. (1) by a summation of dynamic frozen Gaussian wave packets,

$$
\begin{aligned}
u_{\mathrm{F}}^k(t, \boldsymbol{x}) \approx & \sum_{(\boldsymbol{q}, \boldsymbol{p}) \in G_{\pm}^{\mathrm{p}}} \frac{a_{\mathrm{p}} \hat{\boldsymbol{N}}_{\mathrm{p}} \psi_{\mathrm{p}}^k}{(2\pi/k)^{9/2}} e^{ik\boldsymbol{P}_{\mathrm{p}} \cdot (\boldsymbol{x} - \boldsymbol{Q}_{\mathrm{p}}) - \frac{k}{2}|\boldsymbol{x} - \boldsymbol{Q}_{\mathrm{p}}|^2} \delta \boldsymbol{q} \delta \boldsymbol{p} \\
& + \sum_{(\boldsymbol{q}, \boldsymbol{p}) \in G_{\pm}^{\mathrm{s}}} \frac{a_{\mathrm{s}} \hat{\boldsymbol{N}}_{\mathrm{s}} \psi_{\mathrm{s}}^k}{(2\pi/k)^{9/2}} e^{ik\boldsymbol{P}_{\mathrm{s}} \cdot (\boldsymbol{x} - \boldsymbol{Q}_{\mathrm{s}}) - \frac{k}{2}|\boldsymbol{x} - \boldsymbol{Q}_{\mathrm{s}}|^2} \delta \boldsymbol{q} \delta \boldsymbol{p},
\end{aligned}
\tag{5}
$$

with the weight functions

$$
\psi_{\mathrm{p,s}}^k(\boldsymbol{q}, \boldsymbol{p}) = \int \alpha_{\mathrm{p,s}}^k(\boldsymbol{y}, \boldsymbol{q}, \boldsymbol{p}) e^{-ik\boldsymbol{p} \cdot (\boldsymbol{y} - \boldsymbol{q}) - \frac{k}{2}|\boldsymbol{y} - \boldsymbol{q}|^2} \, \mathrm{d}\boldsymbol{y},
\tag{6}
$$

$$
\alpha_{\mathrm{p,s}}^k(\boldsymbol{y}, \boldsymbol{q}, \boldsymbol{p}) = \frac{1}{2kc_{\mathrm{p,s}}|\boldsymbol{p}|^3} \left(k\boldsymbol{f}^k(\boldsymbol{y}) c_{\mathrm{p,s}}|\boldsymbol{p}| \pm i\boldsymbol{g}^k(\boldsymbol{y}) \right) \cdot \hat{\boldsymbol{n}}_{\mathrm{p,s}}.
\tag{7}
$$

In eq. (5), $i = \sqrt{-1}$ is the imaginary unit, and we use unbolded subscripts/superscripts "p" and "s" to indicate P- and S-waves respectively, and superscripts k to indicate quantities depending on wave number k. The quantities that have both p and s as subscripts/superscripts mean that they can be referred to both P- and S-waves. Here $G_{\pm}^{\mathrm{p,s}}$ refers to the initial sets of Gaussian center $\boldsymbol{q}$ and propagation vector $\boldsymbol{p}$ for P- and S-waves respectively, and $\pm$ indicates the two-way wave propagation directions correspondingly. $\hat{\boldsymbol{N}}_{\mathrm{p,s}}(t)$ are unit vectors indicating the polarized directions of P- and S-waves, i.e. if $(\boldsymbol{q}, \boldsymbol{p}) \in G_{\pm}^{\mathrm{p}}$, then $\hat{\boldsymbol{N}}_{\mathrm{p}} \parallel \boldsymbol{P}_{\mathrm{p}}$; and if $(\boldsymbol{q}, \boldsymbol{p}) \in G_{\pm}^{\mathrm{s}}$, then $\hat{\boldsymbol{N}}_{\mathrm{s}} \perp \boldsymbol{P}_{\mathrm{s}}$. In (7), $\hat{\boldsymbol{n}}_{\mathrm{p,s}}$ are the initial directions of P- and S-waves, i.e. $\hat{\boldsymbol{n}}_{\mathrm{p,s}} = \hat{\boldsymbol{N}}_{\mathrm{p,s}}(0)$, and the "$\pm$" on the right-hand-side of the equation indicate that the $\alpha_{\mathrm{p,s}}^k$ correspond to $(\boldsymbol{q}, \boldsymbol{p}) \in G_{\pm}^{\mathrm{p,s}}$. Remark that $e^{ik\boldsymbol{p} \cdot (\boldsymbol{y} - \boldsymbol{q}) - \frac{k}{2}|\boldsymbol{y} - \boldsymbol{q}|^2}$ can be understood as complex localized wave-packet centered at $\boldsymbol{q}$ with propagation vector $\boldsymbol{p}$, and $\psi_{\mathrm{p,s}}^k$ is the projection of the initial wavefield onto each wave-packet computed by eq. (6) with $\boldsymbol{y}$ serving as the dummy variable in the integration. Associated with each frozen Gaussian wave-packet, the *time-dependent* quantities are the position center $\boldsymbol{Q}_{\mathrm{p,s}}(t, \boldsymbol{q}, \boldsymbol{p})$, momentum center $\boldsymbol{P}_{\mathrm{p,s}}(t, \boldsymbol{q}, \boldsymbol{p})$, amplitude $a_{\mathrm{p,s}}(t, \boldsymbol{q}, \boldsymbol{p})$ and unit direction vectors $\hat{\boldsymbol{N}}_{\mathrm{p,s}}(t)$. Note that all the S-waves discussed in the formulation will include both SH- and SV-waves.

2.2   Formulation and Algorithm

The derivation of the FGA formulation involves with the asymptotic expansion in the integral form (weak sense), with proper integration by parts performed to convert powers of distance to the Gaussian center $\boldsymbol{x} - \boldsymbol{Q}$ to orders of the wavelength k^{-1}. It is quite lengthy and technically involved, and thus we leave it to Appendix A for the readers who are interested in the mathematical details, and only present the FGA formulation as below.

For the "+" wave propagation, i.e., $(\boldsymbol{q}, \boldsymbol{p}) \in G_+^{\mathrm{p,s}}$, the Gaussian center $\boldsymbol{Q}_{\mathrm{p,s}}(t, \boldsymbol{q}, \boldsymbol{p})$ and propagation vector $\boldsymbol{P}_{\mathrm{p,s}}(t, \boldsymbol{q}, \boldsymbol{p})$ follow the ray dynamics

$$
\begin{cases}
\dfrac{\mathrm{d}\boldsymbol{Q}_{\mathrm{p,s}}}{\mathrm{d}t} = c_{\mathrm{p,s}}(\boldsymbol{Q}_{\mathrm{p,s}}) \dfrac{\boldsymbol{P}_{\mathrm{p,s}}}{|\boldsymbol{P}_{\mathrm{p,s}}|}, \\
\dfrac{\mathrm{d}\boldsymbol{P}_{\mathrm{p,s}}}{\mathrm{d}t} = -\partial_{\boldsymbol{Q}} c_{\mathrm{p,s}}(\boldsymbol{Q}_{\mathrm{p,s}})|\boldsymbol{P}_{\mathrm{p,s}}|,
\end{cases}
\tag{8}
$$

with initial conditions

$$
\boldsymbol{Q}_{\mathrm{p,s}}(0, \boldsymbol{q}, \boldsymbol{p}) = \boldsymbol{q} \quad \text{and} \quad \boldsymbol{P}_{\mathrm{p,s}}(0, \boldsymbol{q}, \boldsymbol{p}) = \boldsymbol{p}.
\tag{9}
$$

For the "-" wave propagation, i.e., $(\boldsymbol{q}, \boldsymbol{p}) \in G_-^{\mathrm{p,s}}$, the Gaussian center $\boldsymbol{Q}_{\mathrm{p,s}}(t, \boldsymbol{q}, \boldsymbol{p})$ and propagation vector $\boldsymbol{P}_{\mathrm{p,s}}(t, \boldsymbol{q}, \boldsymbol{p})$ follow the ray dynamics

$$
\begin{cases}
\dfrac{\mathrm{d}\boldsymbol{Q}_{\mathrm{p,s}}}{\mathrm{d}t} = -c_{\mathrm{p,s}}(\boldsymbol{Q}_{\mathrm{p,s}}) \dfrac{\boldsymbol{P}_{\mathrm{p,s}}}{|\boldsymbol{P}_{\mathrm{p,s}}|}, \\
\dfrac{\mathrm{d}\boldsymbol{P}_{\mathrm{p,s}}}{\mathrm{d}t} = \partial_{\boldsymbol{Q}} c_{\mathrm{p,s}}(\boldsymbol{Q}_{\mathrm{p,s}})|\boldsymbol{P}_{\mathrm{p,s}}|,
\end{cases}
\tag{10}
$$

with initial conditions

$$
\boldsymbol{Q}_{\mathrm{p,s}}(0, \boldsymbol{q}, \boldsymbol{p}) = \boldsymbol{q} \quad \text{and} \quad \boldsymbol{P}_{\mathrm{p,s}}(0, \boldsymbol{q}, \boldsymbol{p}) = \boldsymbol{p}.
\tag{11}
$$

Remark that, the equations for "-" wave propagation (10) have opposite signs of the right-hand side to the equations for "+" wave propagation (8), and other than that, they are the same. Actually they can be both viewed as the Hamiltonian system with $H(\boldsymbol{Q}, \boldsymbol{P}) = \pm c_{\mathrm{p,s}}(\boldsymbol{Q})|\boldsymbol{P}|$ as the Hamiltonian function for "+" and "-" wave propagation, respectively.

The prefactor amplitudes $\boldsymbol{a}_{\mathrm{p,s}}(t, \boldsymbol{q}, \boldsymbol{p})$ satisfy the following equations, where S-waves have been decomposed into SH- and SV-waves,

$$
\frac{\mathrm{d}a_{\mathrm{p}}}{\mathrm{d}t} = a_{\mathrm{p}} \left(\pm \frac{\partial_{\boldsymbol{Q}_{\mathrm{p}}} c_{\mathrm{p}} \cdot \boldsymbol{P}_{\mathrm{p}}}{|\boldsymbol{P}_{\mathrm{p}}|} + \frac{1}{2}\mathrm{Tr}\left(Z_{\mathrm{p}}^{-1} \frac{\mathrm{d}Z_{\mathrm{p}}}{\mathrm{d}t} \right) \right),
\tag{12}
$$

$$
\frac{\mathrm{d}a_{\mathrm{sv}}}{\mathrm{d}t} = a_{\mathrm{sv}} \left(\pm \frac{\partial_{\boldsymbol{Q}_{\mathrm{s}}} c_{\mathrm{s}} \cdot \boldsymbol{P}_{\mathrm{s}}}{|\boldsymbol{P}_{\mathrm{s}}|} + \frac{1}{2}\mathrm{Tr}\left(Z_{\mathrm{s}}^{-1} \frac{\mathrm{d}Z_{\mathrm{s}}}{\mathrm{d}t} \right) \right) - a_{\mathrm{sh}} \frac{\mathrm{d}\hat{\boldsymbol{N}}_{\mathrm{sh}}}{\mathrm{d}t} \cdot \hat{\boldsymbol{N}}_{\mathrm{sv}},
\tag{13}
$$

$$
\frac{\mathrm{d}a_{\mathrm{sh}}}{\mathrm{d}t} = a_{\mathrm{sh}} \left(\pm \frac{\partial_{\boldsymbol{Q}_{\mathrm{s}}} c_{\mathrm{s}} \cdot \boldsymbol{P}_{\mathrm{s}}}{|\boldsymbol{P}_{\mathrm{s}}|} + \frac{1}{2}\mathrm{Tr}\left(Z_{\mathrm{s}}^{-1} \frac{\mathrm{d}Z_{\mathrm{s}}}{\mathrm{d}t} \right) \right) + a_{\mathrm{sv}} \frac{\mathrm{d}\hat{\boldsymbol{N}}_{\mathrm{sh}}}{\mathrm{d}t} \cdot \hat{\boldsymbol{N}}_{\mathrm{sv}},
\tag{14}
$$

with the initial conditions $a_{\mathrm{p,sv,sh}} = 2^{3/2}$, and $\hat{\boldsymbol{N}}_{\mathrm{sv}}$ and $\hat{\boldsymbol{N}}_{\mathrm{sh}}$ are the two unit directions perpendicular to $\boldsymbol{P}_{\mathrm{s}}$, referring to the polarized directions of SV- and SH-waves, respectively. Here "$\pm$" corresponds to the two-way wave propagation directions, and we have used the short-hand notations,

$$
\partial_{\boldsymbol{z}} = \partial_{\boldsymbol{q}} - \mathrm{i}\partial_{\boldsymbol{p}}, \qquad Z_{\mathrm{p,s}} = \partial_{\boldsymbol{z}}(\boldsymbol{Q}_{\mathrm{p,s}} + \mathrm{i}\boldsymbol{P}_{\mathrm{p,s}}).
\tag{15}
$$

Note that, the prefactor equation (12) is consistent with the one for acoustic wave equation with $c^2 = (\lambda + 2\mu)/\rho$ (Chai et al. 2017) , and the last terms on the right-hand-side of (13)-(14) indicate the diabatic coupling of the polarized directions for SH- and SV-waves, which are closely connected to
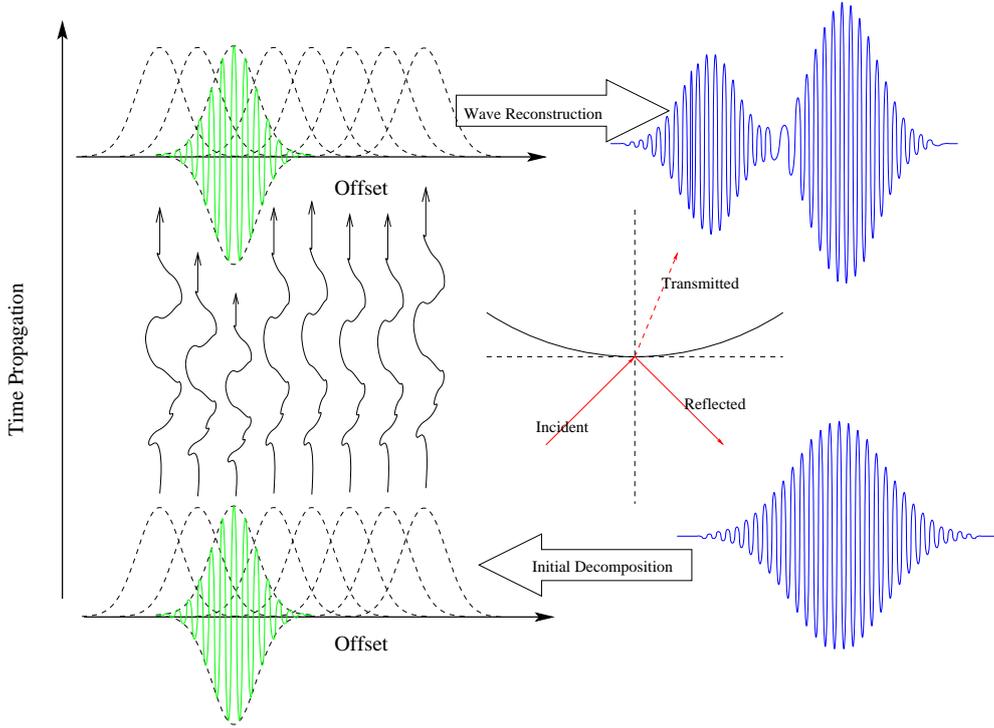
Figure 1. This cartoon figure illustrates the main three steps of the FGA algorithm: Step 1, Initial decomposition, i.e., to choose the proper sets $G_{\pm}^{p,s}$ by the fast FBI transform algorithm (Yang et al. 2013, Page 2) to decompose the initial wavefields into Gaussian wave packets, and calculate the corresponding weight function $\psi_{p,s}^{k}$ defined in eq. (6); Step 2, Time propagation, i.e., to solve by some numerical solver eqs. (8), (10), (12), (13) and (14) for the dynamics of Gassian center, propagation vector and prefactor amplitude. Transmission and reflection conditions are needed at the presence of interfaces as given in Section 3.1; Step 3, Wave reconstruction, i.e., to compute eq. (5) for the elastic wavefield at time T. The fast Gaussian summation algorithm proposed in Chai et al. (2018, Section 3.3) is highly recommended if a large number of reconstruction grid points are needed.

the concept of Berry phase intensively studied in quantum mechanics and topology (Chern number) (e.g. Berry 1984; Simon 1983).

Algorithm. A flowchart is shown in Fig. 1 to describe the FGA algorithm, with the technique details discussed in the figure caption.

2.3 Accuracy and Parallelizability

We check the accuracy and parallelizability of the FGA method by simulating the elastic wave propagation at a homogeneous media with absorbing boundary conditions, for which one can solve the analytical solution as a benchmark. Specifically, we consider

$$\rho\partial_t^2 \boldsymbol{u} = (\lambda + \mu)\nabla(\nabla \cdot \boldsymbol{u}) + \mu\Delta\boldsymbol{u} + \boldsymbol{F}(t)\delta(\boldsymbol{x} - \boldsymbol{x}_0), \tag{16}$$

with $\boldsymbol{x}_0 = (x_{1,0}, x_{2,0}, x_{3,0})$ as the source location, $\boldsymbol{F} = \big(F_1(t), F_2(t), F_3(t)\big)^{\mathrm{T}}$ as the source time function at $\boldsymbol{x}_0$, and ρ, λ and μ as constants. Eq. (16) has the following analytical solution

$$
\begin{aligned}
u_i = {} & \frac{(x_i - x_{i,0})(x_j - x_{j,0})}{4\pi\rho c_{\mathrm{p}}^2 r^3} F_j(t - r/c_{\mathrm{p}}) + \frac{r^2\delta_{ij} - (x_i - x_{i,0})(x_j - x_{j,0})}{4\pi\rho c_{\mathrm{s}}^2 r^3} F_j(t - r/c_{\mathrm{s}}) \\
& + \frac{3(x_i - x_{i,0})(x_j - x_{j,0}) - r^2\delta_{ij}}{4\pi\rho r^5} \int_{r/c_{\mathrm{p}}}^{r/c_{\mathrm{s}}} s F_j(t - s)\ \mathrm{d}s,
\end{aligned}
\tag{17}
$$

where we have used the Einstein's index summation convention, $\boldsymbol{u} = (u_1, u_2, u_3)^{\mathrm{T}}$, $r = \|\boldsymbol{x} - \boldsymbol{x}_0\|_2$ and δ_{ij} is the Kronecker delta function. Here we use $\boldsymbol{x} = (x_1, x_2, x_3)^{\mathrm{T}}$ for a simple formulation of the analytical solution, which is actually $\boldsymbol{x} = (x, y, z)^{\mathrm{T}}$ in standard notations.

In the numerical tests, we simulate the elastic wave propagation in a domain of the size 128 km $\times$ 128 km $\times$ 128 km, and take the source time function as

$$
F_j(t) = \cos\left(2\pi f(t + T_0)\right) \exp\left(-(t + T_0)^2/\sigma^2\right),
$$

with f as the frequency, $T_0 = 0.1768$ s and $\sigma = 0.8660$ s. We choose P- and S-wave speeds as $c_{\mathrm{p}} = 8$ km/s, $c_{\mathrm{s}} = 4.619$ km/s, density $\rho = 1$ kg/km^3, final propagation time $T = 13.86$ s and the source location $\boldsymbol{x}_0 = (64, 64, 64)$ km. We use a fourth order Runge-Kutta (RK4) with fixed time step as the numerical solver for eqs. (8), (10), (12), (13) and (14). To compare the performance of the FGA method to the spectral element software package SPECFEM3D*, simluations are done on a cluster of each node equipped with an Intel(R) Xeon(R) E5-2670 (2.60GHz) and a total of 64GB RAM. The accuracy and parallelizability of FGA are tested for $f = 1.4702$ Hz, and the comparison of computational time to SPECFEM3D is done for a range of f from 0.3676 Hz to 11.7617 Hz.

The dynamics of the elastic wave propagation is shown by the FGA method in Fig. 2, which is consistent with the analytical solution (17) as an expanding ball. The component-wise wavefields, including both P- and S-wave components, are shown in Fig. 3 for $t = 6.93$ s. With the comparable accuracy of FGA to SPECFEM3D as illustrated in Fig. 4 for $t = 6.93$ s and the source frequency $f = 1.4702$ Hz. The relative error for Fig. 4(d), FGA versus the analytical solution is computed to be 3.84%; while for Fig. 4(e), the relative error is computed to be 3.57%. With a comparable accuracy, the FGA shows a much faster computational speed and better parallelizablity than SPECFEM3D for high-frequency (≥ 1.4702 Hz) elastic wave propagation, with details described in Figs. 5 and 6. Particularly, one can see in Fig. 5 that the computational time of SPECFEM3D has nearly cubic growth in the frequency of elastic waves, while FGA has about 1.5-times growth in the frequency. Fig. 6 shows that the speed-up ratio for FGA is almost 2 when one doubles the number of processors, which indicates that the FGA algorithm is embarrassingly parallel. Remark that, in the simulation for the

---

* https://geodynamics.org/cig/software/specfem3d/

(a) Initial wavefield   (b) $t = 6.93$ s   (c) $t = 13.86$ s
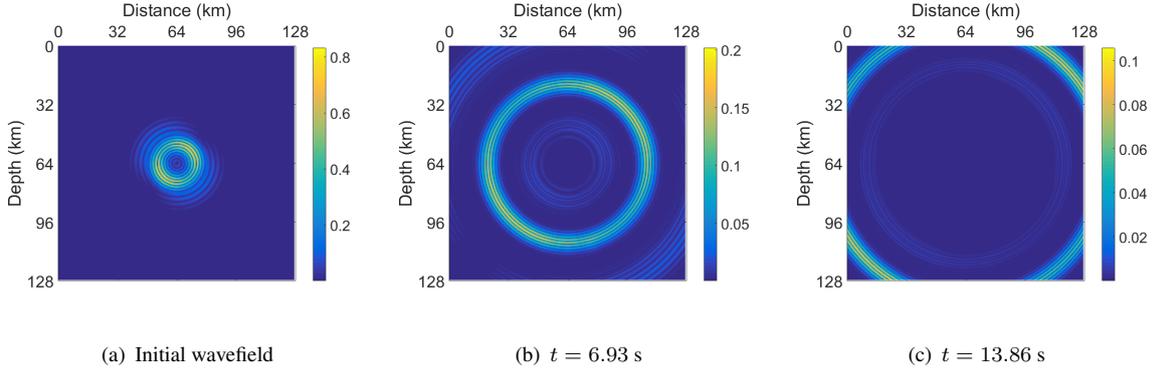
Figure 2. Modulus of the elastic wavefield computed by the FGA method in homogeneous media, with the analytical solution given by eq. (17). The subfigures from left to right show the slices of $\|\boldsymbol{u}\|_2 = \sqrt{u_1^2 + u_2^2 + u_3^2}$ at $y = 64$km for $t = 0, 6.93, 13.86$ s. The frequency of the source time function is $f = 1.4702$ Hz.

source frequency $f = 1.4702$ Hz where FGA and SPECFEM3D produce comparable accuracy, we use 1504436 Gaussians for computing P-wave of both "$\pm$" propagation directions, and 2120482 Gaussians for computing S-wave of both "$\pm$" propagation directions, which needs to roughly compute a total number of $8 \times (1504436 + 2120482) \approx 30$ millions of variables in the simulation. Note that, the equations for both P- and S-waves have the same number of variables, and the prefactor 8 comes by counting the number of variables needed in eqs. (8) and (12) where $\boldsymbol{Q}$ and $\boldsymbol{P}$ are 3-D real vectors and the prefactor amplitude a is a complex number; while in SPECFEM3D, we use 128 elements in each direction with 5^3 nodes in each element. One needs to roughly to compute a total number of $3 \times 128^3 \times 5^3 \approx 800$ millions of variables in the simulation, where the prefactor 3 is due to $\boldsymbol{u}$ is a 3-D real vector in eq. (1). In addition, the stability conditions on time step for solving the ODE systems (8) and (12) by RK4 is better than solving the elastic wave equation (1) by SPECFEM, since the CFL condition is restricted by small wavelength in solving eq. (1).

3   INTERFACE CONDITIONS AND SIMULATION OF A 1-D LAYERED EARTH MODEL

In this section, we derive the interface conditions of FGA for 3-D elastic wave propagation in heterogeneous media with strong discontinuities, e.g., the Moho surface and Core-mantle boundary. We verify these interface conditions by simulating a 1-D layered Earth model, which actually requires to revise the FGA solution in spherical coordinates (non-Cartesian coordinates). This can be easily done by an observation that the only step of the FGA algorithm depending on the $\boldsymbol{x}$-coordinate is the reconstruction part given by eq. (5) and all the other steps including the initial wavefield decomposition by eqs. (6) and (7) and propagation of ray equations eqs. (8), (10), (12), (13) and (14) are $\boldsymbol{x}$-coordinate-free. Therefore, to generalize the FGA method in spherical coordinates, one simply needs to change

(a) u_1

(b) u_1: P-wave component

(c) u_1: S-wave component

(d) u_2

(e) u_2: P-wave component

(f) u_2: S-wave component

(g) u_3

(h) u_3: P-wave component

(i) u_3: S-wave component
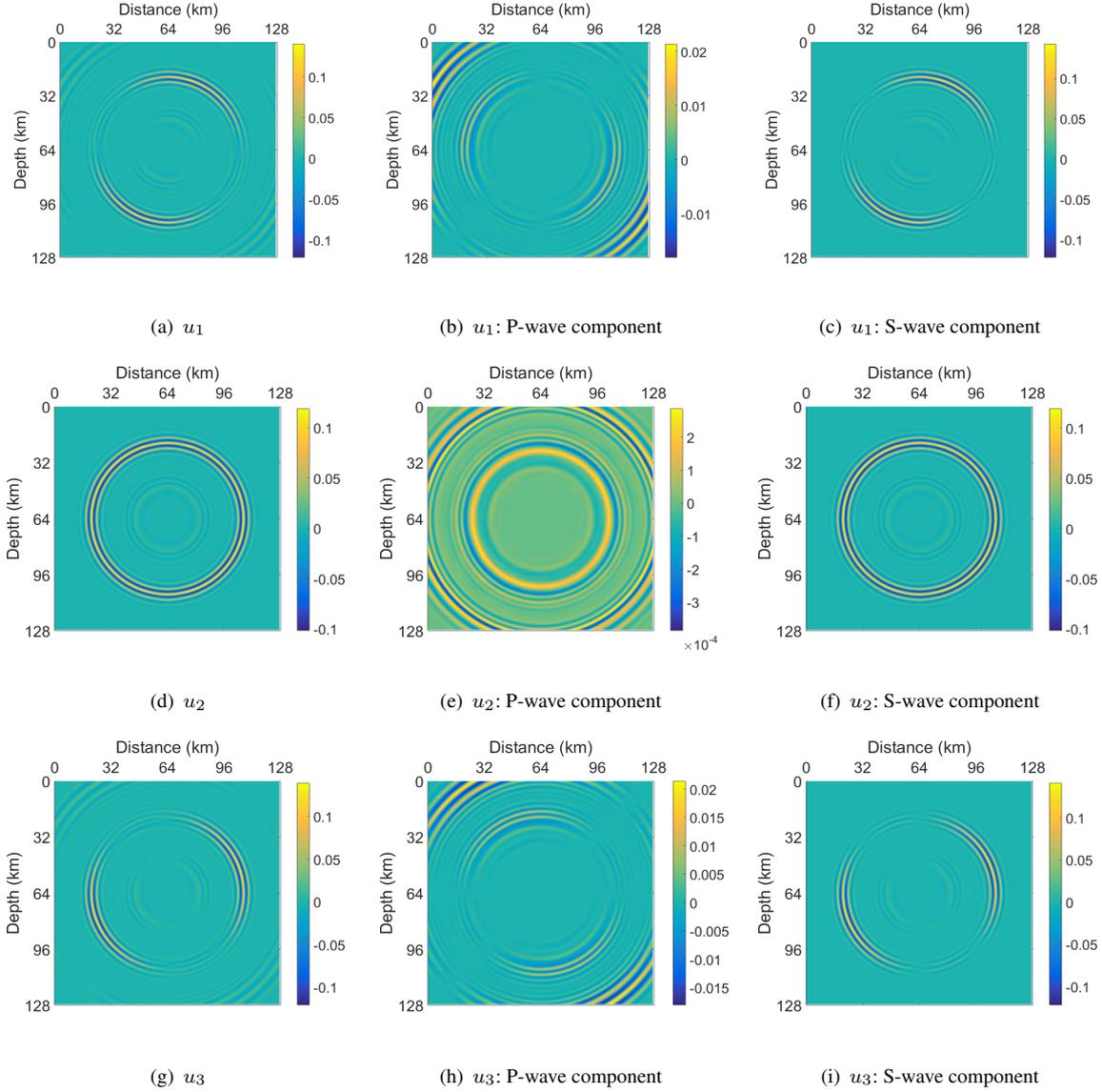
Figure 3. Modulus of the componentwise elastic wavefield computed by the FGA method in homogeneous media, with the analytical solution given by eq. (17). The subfigures show the components of $\boldsymbol{u} = (u_1, u_2, u_3)^{\mathrm{T}}$ for $t = 6.93$ s and the source frequency $f = 1.4702$ Hz. The second and third columns show the computed P- and S-wavefields by the FGA method, respectively.

$\boldsymbol{x} = (x, y, z)^{\mathrm{T}}$ in eq. (5) to $\boldsymbol{x} = (r \cos\theta \cos\phi, r \sin\theta \cos\phi, r \sin\phi)$ with r as the distance to earth center, θ and ϕ as the longitude and latitude degrees.

3.1   Transmission and Reflection Interface Conditions

For a sake of clarity and simplicity, we only give the interface conditions for a flat interface located at $z = z_0$, and for a reflecting geometry of general shape, one needs to apply the formulation in the local

(a) FGA (b) SEM (c) Solution

(d) FGA vs Solution (e) SEM vs Solution
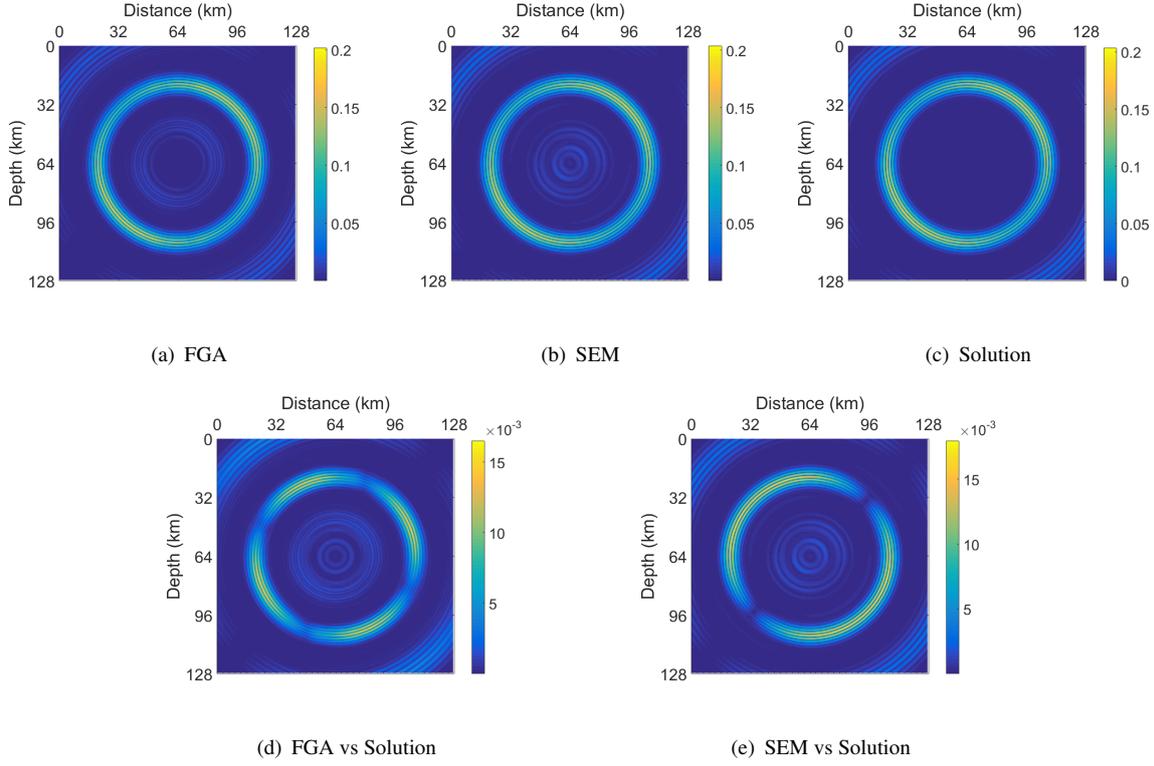
Figure 4. Comparison of accuracy for FGA and SPECFEM3D to the analytical solution (17) at $t = 6.93$ s with the source frequency $f = 1.4702$ Hz. The top subfigures are the modulus of wavefields given by FGA, SPECFEM and analytical solution. The bottom subfigures are the modulus differences between the FGA and analytical solutions (bottom left) and between the SPECFEM and analytical solutions (bottom right). This shows that FGA and SPECFEM3 produce comparable accuracy in this case.

tangent-normal coordinates by treating the tangential direction as the local flat horizontal interface. The wave speeds of the two layers are assumed to be,

$$c_p(\boldsymbol{x}) = \begin{cases} c_p^\vee(\boldsymbol{x}) & z > z_0 \\ c_p^\wedge(\boldsymbol{x}) & z < z_0 \end{cases}, \quad c_s(\boldsymbol{x}) = \begin{cases} c_s^\vee(\boldsymbol{x}) & z > z_0 \\ c_s^\wedge(\boldsymbol{x}) & z < z_0 \end{cases}. \tag{18}$$

In Fig. 7, we only consider an incident Gaussian wave packet for P-wave hitting the interface at $z = z_0$, and then reflected and transmitted as Gaussian wave packets for P- and SV-waves, respectively. The other cases including an incident Gaussian wave packet for SV- and SH-waves can be handled similarly, although there is no interaction between P- and S-waves for the case of SH-wave.

Associated with each Gaussian wave packet, one needs to provide the reflection and transmission conditions for $a_{p,s}$, $\boldsymbol{Q}_{p,s}$, and $\boldsymbol{P}_{p,s}$, which follow the Snell's Law and the Zoeppritz equations (Yilmaz 2001). However, for FGA, what is different from standard ray theory on the flat interface is that, one also needs to derive the interface conditions for $Z_{p,s}$ which will change after the Gaussian wave packet hits the interface and affect the dynamics of $a_{p,s}$ given by eqs. (12)-(14). Eq. (15) implies that to derive
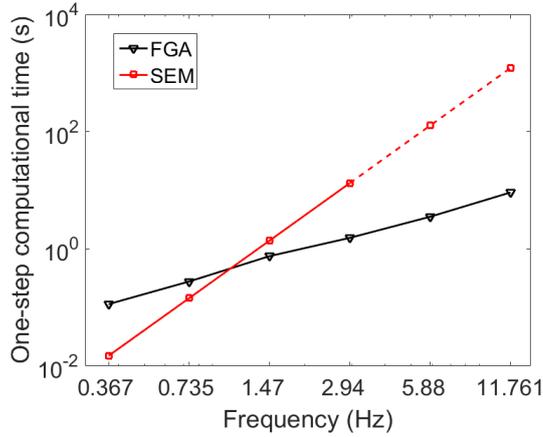
Figure 5. Dependence of one-step computational time on frequency for both FGA and SPECFEM3D in homogeneous media. The horizontal axis is the frequency f (Hz) and the vertical axis is the one-step computational time (s) of the solvers. The triangle line stands for the FGA simulations and the square line stands for the SPECFEM3D simulations. Due to the limitation of memory, SPEFEM3D can not run for $f \geq 5.88084$ Hz, and the dashed square line is obtained by extrapolation.
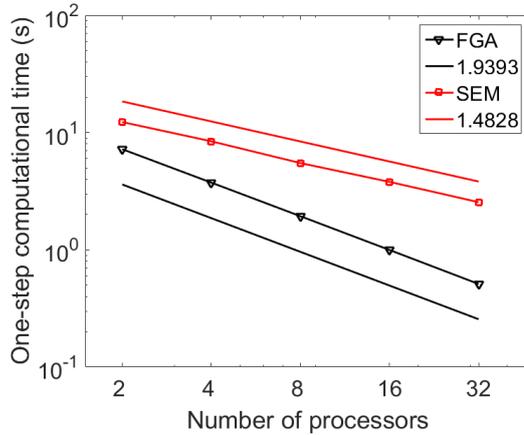
Figure 6. Dependence of time on the number of processors for the source frequency $f = 1.4702$ Hz in homogeneous media. The ideal speed-up ratio is 2, while one can see that the speed-up ratio for FGA is approximately 1.9393, which is slightly smaller than 2 indicating an almost perfectly parallel efficiency. On the other hand, as a comparison, SPECFEM3D is used with 128 elements in each spatial direction to achieve a comparable accuracy to FGA. The speed-up ratio for the SPECFEM3D solver is around 1.4828, which is smaller that those of FGA. This is because SPECFEM3D solves eq. (1) on a parallel computer with N processors by partitioning the whole domain into N slabs with each processor solving the equation in each slab. Therefore, for each time step, each processor needs to communicate with its neighbors to get necessary boundary information, which decreases the speed-up ratio.
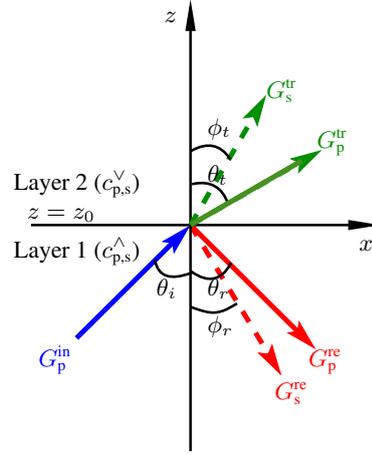
**Figure 7.** Cartoon illustration of an incident Gaussian wave packet for P-wave hitting the interface at $z = z_0$, and then reflected and transmitted as Gaussian wave packets for P- and SV-waves. Here the $G_{\mathrm{p,s}}^{\mathrm{in,re,tr}}$ stands for the Gaussian wave packet for the incident, reflected and transmitted P- and SV-waves, respectively. We denote $\theta_i, \theta_r, \theta_t$ to be the incident, reflection and transmission angles of P-waves, and ϕ_r, ϕ_t to be the reflection and transmission angles of SV-waves, respectively.

the interface conditions of $Z_{\mathrm{p,s}}$ will be equivalent to derive the interface conditions for $\partial_z \boldsymbol{Q}_{\mathrm{p,s}}$ and $\partial_z \boldsymbol{P}_{\mathrm{p,s}}$, which requires to use the conservation of level set functions designed in the Eulerian frozen Gaussian approximation formula (Lu & Yang 2012b; Wei & Yang 2012). The mathematical details of the derivation are lengthy and technical, and thus we leave them to Appendix B for the interested readers, and only present the results here:

$$\partial_z \boldsymbol{Q}^{\mathrm{re,tr}} = \partial_z \boldsymbol{Q}^{\mathrm{in}} F,$$
$$\partial_z \boldsymbol{P}^{\mathrm{re,tr}} = \partial_z \boldsymbol{P}^{\mathrm{in}} W - \frac{|\boldsymbol{P}^{\mathrm{re,tr}}|}{c(\boldsymbol{Q}^{\mathrm{re,tr}}) p_z^{\mathrm{re,tr}}} \left(\partial_z \boldsymbol{Q}^{\mathrm{re,tr}} \cdot \nabla c(\boldsymbol{Q}^{\mathrm{re,tr}}) - \partial_z \boldsymbol{Q}^{\mathrm{in}} \cdot \nabla c(\boldsymbol{Q}^{\mathrm{in}}) \right) \boldsymbol{e}_3, \tag{19}$$

where $(\boldsymbol{Q}^{\mathrm{in,re,tr}}, \boldsymbol{P}^{\mathrm{in,re,tr}})$ corresponds to the center and propagation vector of incident, reflected and transmitted Gaussian wave packet for either P- or S-waves, respectively, and $\boldsymbol{P}^{\mathrm{in,re,tr}} = (p_x, p_y, p_z^{\mathrm{in,re,tr}})$. Here $\boldsymbol{e}_3 = (0,0,1)$ is a row vector, F and W are two 3×3 matrices, $F^T = W^{-1}$, and

$$F = \begin{bmatrix} 1 & & \\ & 1 & \\ (\kappa - 1) \frac{p_x}{p_z^{\mathrm{in}}} & (\kappa - 1) \frac{p_y}{p_z^{\mathrm{in}}} & \kappa \frac{p_z^{\mathrm{re,tr}}}{p_z^{\mathrm{in}}} \end{bmatrix}, \quad \text{with} \quad \kappa = \left(\frac{c(\boldsymbol{Q}^{\mathrm{re,tr}})}{c(\boldsymbol{Q}^{\mathrm{in}})} \right)^2.$$

Let us explain the formulation of reflection and transmission interface conditions with more detials for the incident P-wave as illustrated in Fig. 7. In this case, if one denotes $(\boldsymbol{Q}_{\mathrm{p,s}}^{\mathrm{in,re,tr}}, \boldsymbol{P}_{\mathrm{p,s}}^{\mathrm{in,re,tr}}, a_{\mathrm{p,s}}^{\mathrm{in,re,tr}})$ to be the center, propagation vector and prefactor amplitude of the corresponding incident, reflected and transmitted Gaussian wave packets for P- and S-waves respectively, and $\boldsymbol{P}_{\mathrm{p}}^{\mathrm{in}} = (p_x, p_y, p_z)$, then

for $\boldsymbol{P}_{\mathrm{p}}^{\mathrm{re,tr}} = (p_x, p_y, p_{z,\mathrm{p}}^{\mathrm{re,tr}})$ and $\boldsymbol{P}_{\mathrm{s}}^{\mathrm{re,tr}} = (p_x, p_y, p_{z,\mathrm{s}}^{\mathrm{re,tr}})$,

$$
\begin{aligned}
p_{z,\mathrm{p}}^{\mathrm{re}} &= -p_z, \\
p_{z,\mathrm{p}}^{\mathrm{tr}} &= \sqrt{(p_z n_1)^2 + (n_1^2 - 1)(p_x^2 + p_y^2)}, \\
p_{z,\mathrm{s}}^{\mathrm{re}} &= -\sqrt{(p_z n_2)^2 + (n_2^2 - 1)(p_x^2 + p_y^2)}, \\
p_{z,\mathrm{s}}^{\mathrm{tr}} &= \sqrt{(p_z n_3)^2 + (n_3^2 - 1)(p_x^2 + p_y^2)},
\end{aligned}
\tag{20}
$$

where $n_1 = c_{\mathrm{p}}^{\wedge}/c_{\mathrm{p}}^{\vee}$, $n_2 = c_{\mathrm{p}}^{\wedge}/c_{\mathrm{s}}^{\wedge}$ and $n_3 = c_{\mathrm{p}}^{\wedge}/c_{\mathrm{s}}^{\vee}$ with $c_{\mathrm{p,s}}^{\vee} = c_{\mathrm{p,s}}^{\vee}(\boldsymbol{x})|_{z=z_0}$ and $c_{\mathrm{p,s}}^{\wedge} = c_{\mathrm{p,s}}^{\wedge}(\boldsymbol{x})|_{z=z_0}$.

Moreover, if one denotes $\theta_i, \theta_r, \theta_t$ to be the P-wave incident, reflection and transmission angles, and ϕ_r, ϕ_t to be the SV-wave reflection and transmission angles, respectively, then the Zoeppritz equations (Yilmaz 2001) read as

$$
M \begin{pmatrix} a_{\mathrm{p}}^{\mathrm{re}} \\ a_{\mathrm{s}}^{\mathrm{re}} \\ a_{\mathrm{p}}^{\mathrm{tr}} \\ a_{\mathrm{s}}^{\mathrm{tr}} \end{pmatrix} = \begin{pmatrix} \cos(\theta_r) \\ \sin(\theta_r) \\ \cos(2\phi_r) \\ \cos(2\theta_r) \end{pmatrix} a_{\mathrm{p}}^{\mathrm{in}},
\tag{21}
$$

with the matrix M as

$$
M = \begin{pmatrix}
\cos(\theta_r) & \frac{c_{\mathrm{p}}^{\wedge}}{c_{\mathrm{s}}^{\wedge}}\sin(\phi_r) & \frac{c_{\mathrm{p}}^{\wedge}}{c_{\mathrm{p}}^{\vee}}\cos(\theta_t) & -\frac{c_{\mathrm{p}}^{\wedge}}{c_{\mathrm{s}}^{\vee}}\sin(\phi_t) \\
-\sin(\theta_r) & \frac{c_{\mathrm{p}}^{\wedge}}{c_{\mathrm{s}}^{\wedge}}\cos(\phi_r) & \frac{c_{\mathrm{p}}^{\wedge}}{c_{\mathrm{p}}^{\vee}}\sin(\theta_t) & \frac{c_{\mathrm{p}}^{\wedge}}{c_{\mathrm{s}}^{\vee}}\cos(\phi_t) \\
-\cos(2\phi_r) & -\sin(2\phi_r) & \frac{\rho_2}{\rho_1}\cos(2\phi_t) & -\frac{\rho_2}{\rho_1}\sin(2\phi_t) \\
\sin(2\theta_r) & -(\frac{c_{\mathrm{p}}^{\wedge}}{c_{\mathrm{s}}^{\wedge}})^2\cos(2\phi_r) & \frac{\rho_2(c_{\mathrm{p}}^{\wedge}c_{\mathrm{s}}^{\vee})^2}{\rho_1(c_{\mathrm{p}}^{\vee}c_{\mathrm{s}}^{\wedge})^2}\sin(2\theta_t) & \frac{\rho_2(c_{\mathrm{p}}^{\wedge})^2}{\rho_1(c_{\mathrm{s}}^{\wedge})^2}\cos(2\phi_t)
\end{pmatrix},
\tag{22}
$$

where $\rho_{1,2}$ are the densities for the layers 1 and 2, respectively.

More explicitly in eq. (19) corresponding to the case in Fig. 7, $c(\boldsymbol{Q}_{\mathrm{p}}^{\mathrm{in}}) = c_{\mathrm{p}}^{\wedge}(\boldsymbol{Q}_{\mathrm{p}}^{\mathrm{in}})$, $c(\boldsymbol{Q}_{\mathrm{p}}^{\mathrm{re}}) = c_{\mathrm{p}}^{\wedge}(\boldsymbol{Q}_{\mathrm{p}}^{\mathrm{re}})$, $c(\boldsymbol{Q}_{\mathrm{p}}^{\mathrm{tr}}) = c_{\mathrm{p}}^{\vee}(\boldsymbol{Q}_{\mathrm{p}}^{\mathrm{tr}})$, $c(\boldsymbol{Q}_{\mathrm{s}}^{\mathrm{re}}) = c_{\mathrm{s}}^{\wedge}(\boldsymbol{Q}_{\mathrm{s}}^{\mathrm{re}})$, and $c(\boldsymbol{Q}_{\mathrm{s}}^{\mathrm{tr}}) = c_{\mathrm{s}}^{\vee}(\boldsymbol{Q}_{\mathrm{s}}^{\mathrm{tr}})$.

3.2 Waveguide example in a 1-D layered Earth model

We verify the interface conditions (19) and (21) by simulating a waveguide example in a 1-D layered Earth model, with the layered P-wave velocity given in Fig. 8(a) following the data in the IASP91 model (Kennett 1991; IRIS DMC 2010). We are particularly interested in choosing the 410-km discontinuity for the numerical proof of the conditions (19) and (21), which presents to a $5-6\%$ increase on P-wave velocity and calibrates the mantle transition zone. We consider a radially symmetric surface source as shown in Fig. 8(b), so that the elastic wave equation (1) has a solution of P-waves in the form of $\boldsymbol{u}(t, \boldsymbol{x}) = \nabla_{\boldsymbol{x}}\psi(t, |\boldsymbol{x}|)$ where $\psi(t, r)$ is the radially symmetric solution to the scalar wave equation,

(a)

(b)
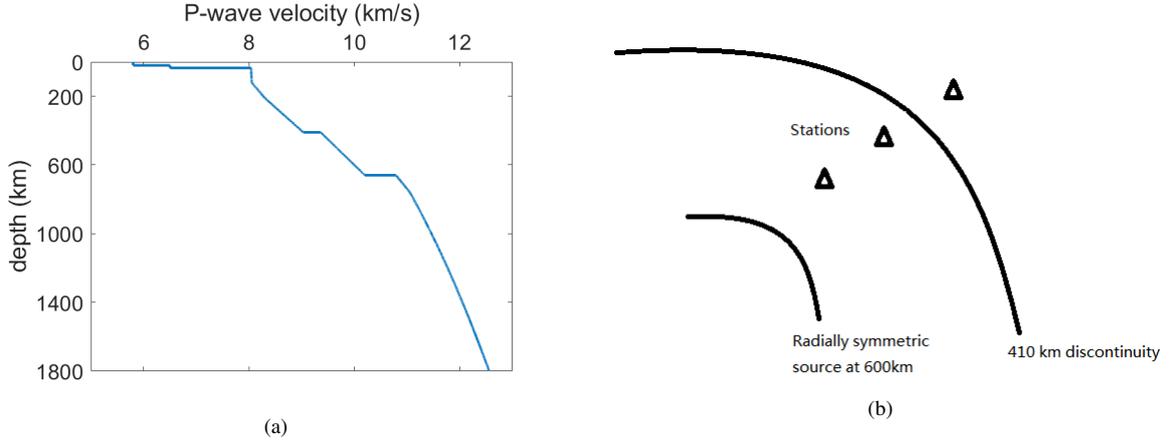
Figure 8. (a): The layered P-wave velocity follows the data in the IASP91 model (Kennett 1991; IRIS DMC 2010). (b): Cartoon plot of the 410-km discontinuity, and distributions of source and stations.

i.e.,

$$\partial_t^2 \psi - c_{\mathrm{p}}^2(\boldsymbol{x})\left(\partial_r^2 \psi + \frac{2}{r}\partial_r \psi\right) = 0. \tag{23}$$

We choose the initial condition for the elastic wave equation (1) as

$$\boldsymbol{u}(t=0, \boldsymbol{x}) = \nabla_{\boldsymbol{x}}\phi_0(r - r_0), \quad \text{and} \quad \partial_t \boldsymbol{u}(t=0, \boldsymbol{x}) = \boldsymbol{0},$$

where $r = |\boldsymbol{x}|$, $r_0 = 600$ km, and

$$\phi_0(r) = \exp\left(-\frac{r^2}{2\sigma^2}\right)\cos\left(\frac{2\pi r}{\ell}\right),$$

with $\sigma = 3.3146$ km and $\ell = 7.3631$ km. The dominant frequency is around 1.36 Hz. We solve eq. (23) using 1-D finite difference method with fine enough grid points as the reference solution, to which we compare the full elastic wave solution of (1) computed by the FGA algorithm. Fig. 9 shows the seismic signals of P-waves received at stations of depth 480 km, 420 km and 360 km, respectively, where one can see a good agreement of FGA simulation with the reference solution for the P-waves.

4 WAVE-EQUATION-BASED TRAVELTIME TOMOGRAPHY AND FWI

In this section, we apply the developed FGA algorithm for 3-D wave-equation-based traveltime tomography (e.g. Tromp et al. 2005; Liu & Gu 2012) and FWI (e.g. Pratt & Shipp 1999; Virieux & Operto 2009), respectively. In particular, we consider the following crosswell seismic tomography, where the three-layered velocity model is set up as follows, with a low-velocity region located at the

(a) P-wave at depth 480 km   (b) P-wave at depth 420 km   (c) P-wave at depth 360 km
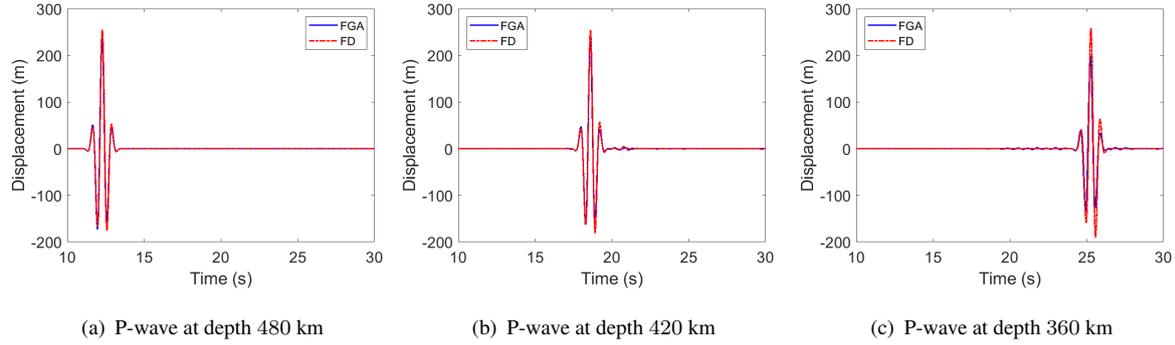
Figure 9. The seismic signals of P-waves received at stations of depth 480 km, 420 km and 360 km, respectively, simulated by the finite difference (FD) and FGA methods, with the source locating at the depth of 600 km. FGA shows a good agreement with the reference signals computed by the FD method for the P-waves.

second layer and homogeneity in the y-direction,

$$c_{\mathrm{p}}(x,y,z) = \sqrt{3}c_{\mathrm{s}}(x,y,z) = \begin{cases} C_1, & \text{if } z_0 < z < z_1, \\ C_2\left(1 - \alpha e^{-\beta\left((x-x_c)^2 + (z-z_c)^2\right)}\right), & \text{if } z_1 < z < z_2, \\ C_3, & \text{if } z > z_2, \end{cases} \quad (24)$$

where the layered velocities 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 center of the low-velocity region is set at $x_c = 75$ m, $z_c = 150$ m. We choose $\alpha = 10\%$ and $\beta = 1/450$ m^{-2} to indicate the largest magnitude of the low-velocity perturbation from the background velocity and the area of low velocity region; see Fig. 10 for an illustration of the P-wave velocity. In Fig. 10, 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.

We start with the following initial velocity model

$$c_0(x,y,z) = \begin{cases} C_1, & \text{if } z_0 < z < z_1, \\ C_2, & \text{if } z_1 < z < z_2, \\ C_3, & \text{if } z > z_2, \end{cases} \quad (25)$$

then use FGA to compute the forward and adjoint wavefields and construct 3-D kernels of different phases by the methods of wave-equation-based traveltime tomography (Tromp et al. 2005; Liu & Gu 2012) and FWI (Pratt & Shipp 1999; Virieux & Operto 2009). Since FWI requires a more sophisticated initial velocity model for the convergence than wave-equation-based traveltime tomography, we use a hierarchical strategy as in our previous work (Chai et al. 2018), which first uses wave-equation-based traveltime tomography to create a macro-scale model and then adopts FWI to generate a high-resolution micro-scale model. For the signals received at the top station (Fig. 11), we show the

(a) Three-layered crosswell model

(b) Relative velocity $(c_{\mathrm{p}} - c_0)/c_0$
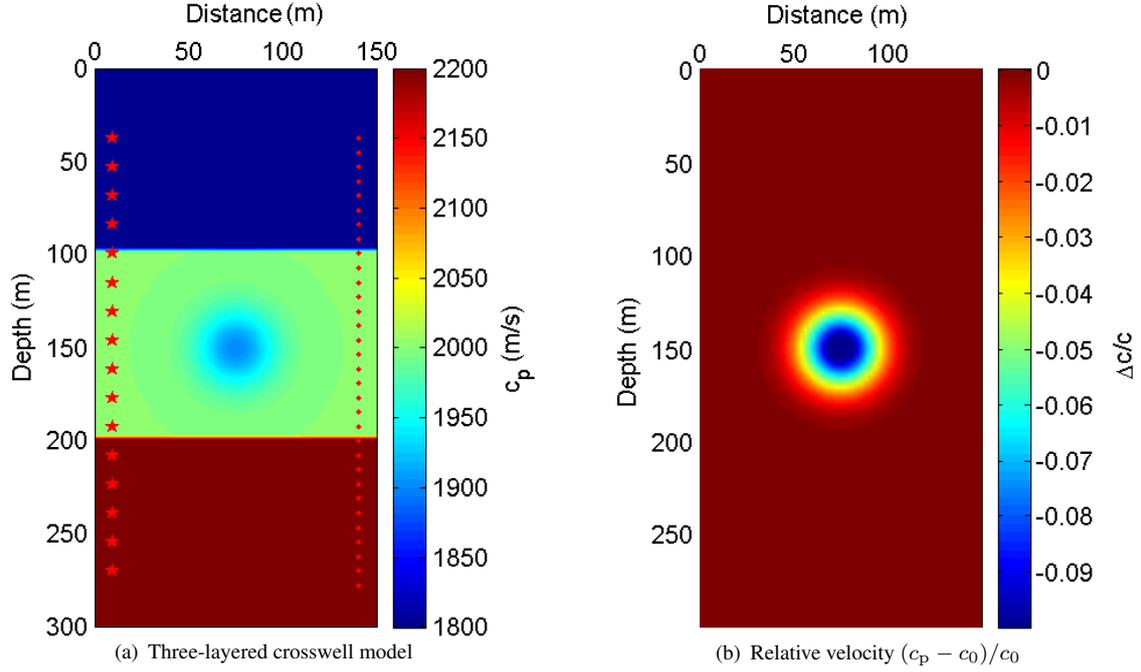
Figure 10. (a): The three-layered crosswell P-wave velocity model with homogeneity in horizontal y-direction. The background velocities from top to bottom are $C_1 = 1800$ m/s, $C_2 = 2000$ m/s, $C_3 = 2200$ m/s, and the interfaces locate at $z_0 = 0$ m, $z_1 = 100$ m, $z_2 = 200$ m. The low-velocity 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. (b): we use the relative velocity $\Delta c/c = (c_{\mathrm{p}} - c_0)/c_0$ to indicate the largest magnitude of the low-velocity perturbation from the background velocity and the area of low velocity region.

corresponding kernels of different phases in Fig. 12. The inversion results are shown in Fig. 13. The damping and smoothing parameters 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.

5   DISCUSSION AND CONCLUSIONS

This is the last part of our theoretical study on the application of the frozen Gaussian approximation (FGA) method to seismic tomography using high-frequency seismic data. Together with Chai et al. (2017, 2018), we systematically derive the FGA formulation for acoustic and elastic wave equations, derive the transmission and reflection conditions of FGA for sharp interfaces (e.g., Moho surface and Core-mantle boundary), reformulate the equations of the forward and adjoint wavefields for a convenient application of the FGA algorithm, propose a fast Gaussian summation algorithm for the
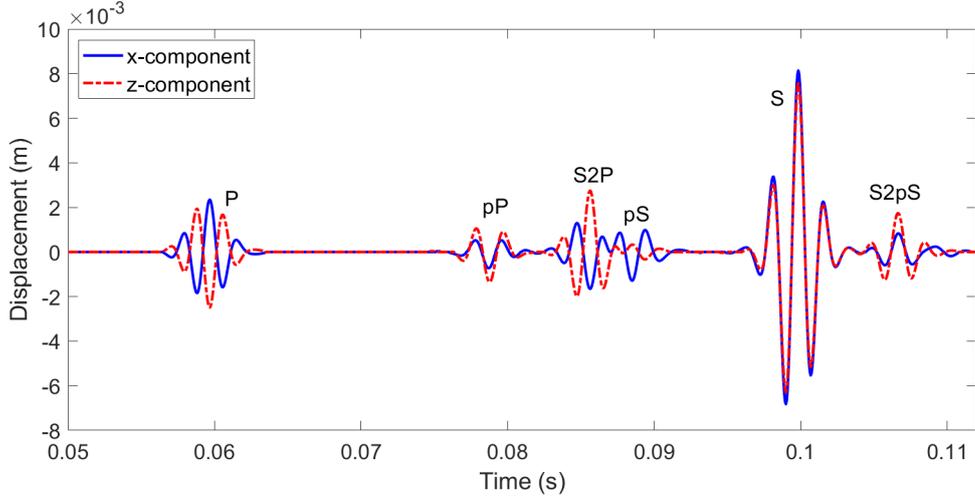
Figure 11. The signals of direct P, pP, S2P, pS, direct S, and S2pS received at the top station with the source as the top ninth source in Fig. 10. Here 2 means the second interface locates at $z_2 = 200$ m, e.g. S2P means S-wave gets reflected at the interface at $z_2 = 200$ m and arrives at the station as P-wave.

reconstruction of FGA, and conduct 3-D high-frequency seismic inversion by the methods of wave-equation-based traveltime tomography and FWI. As a proof of methodology, we compare FGA to SPECFEM3D for acoustic and elastic wave propagation in homogeneous media where an analytical benchmark solution is available. We also test the performance of FGA in a 1-D layered Earth model for checking the accuracy of the interface conditions, and apply FGA to the cross-well example for the 3-D seismic tomography using high-frequency data. We remark that, in the original mathematical work on FGA for strictly linear hyperbolic systems (Lu & Yang 2012a), the FGA formulation is rather complicated, and done for a purpose of rigorously proving its accuracy in the high-frequency regime. In this paper, we weakly expand the solution ansatz of FGA with a consideration of decomposition into P- and S-waves, and reach simple prefactor equations (12), (13) and (14) after lengthy and tricky simplifications. This also generalizes the results in Lu & Yang (2012a) to some extent in the sense that the elastic wave equation is actually not strictly hyperbolic, which has an eigenfunction space of multiplicity-2 (corresponding to S-waves). By these efforts, we hope to bridge a gap between the semiclassical approximation theory in mathematics and the application of 3-D seismic tomography in geophysics. We also remark that, it will be interesting to study the performance of the FGA method in the optimal transport theory-based seismic tomography, with the normalization strategies proposed in, e.g., Engquist & Froese (2014); Métivier et al. (2016); Qiu et al. (2017); Yang et al. (2018); Chen et al. (2017). Our future plans include: 1. to apply 3-D seismic tomography with the FGA method into real data around their dominant frequencies; 2. to use the fast computation performance of FGA to

(a) Kernel of dirct P -- 3D view

(b) Kernel of dirct P

(c) Kernel of pP

(d) Kernel of S2P and pS
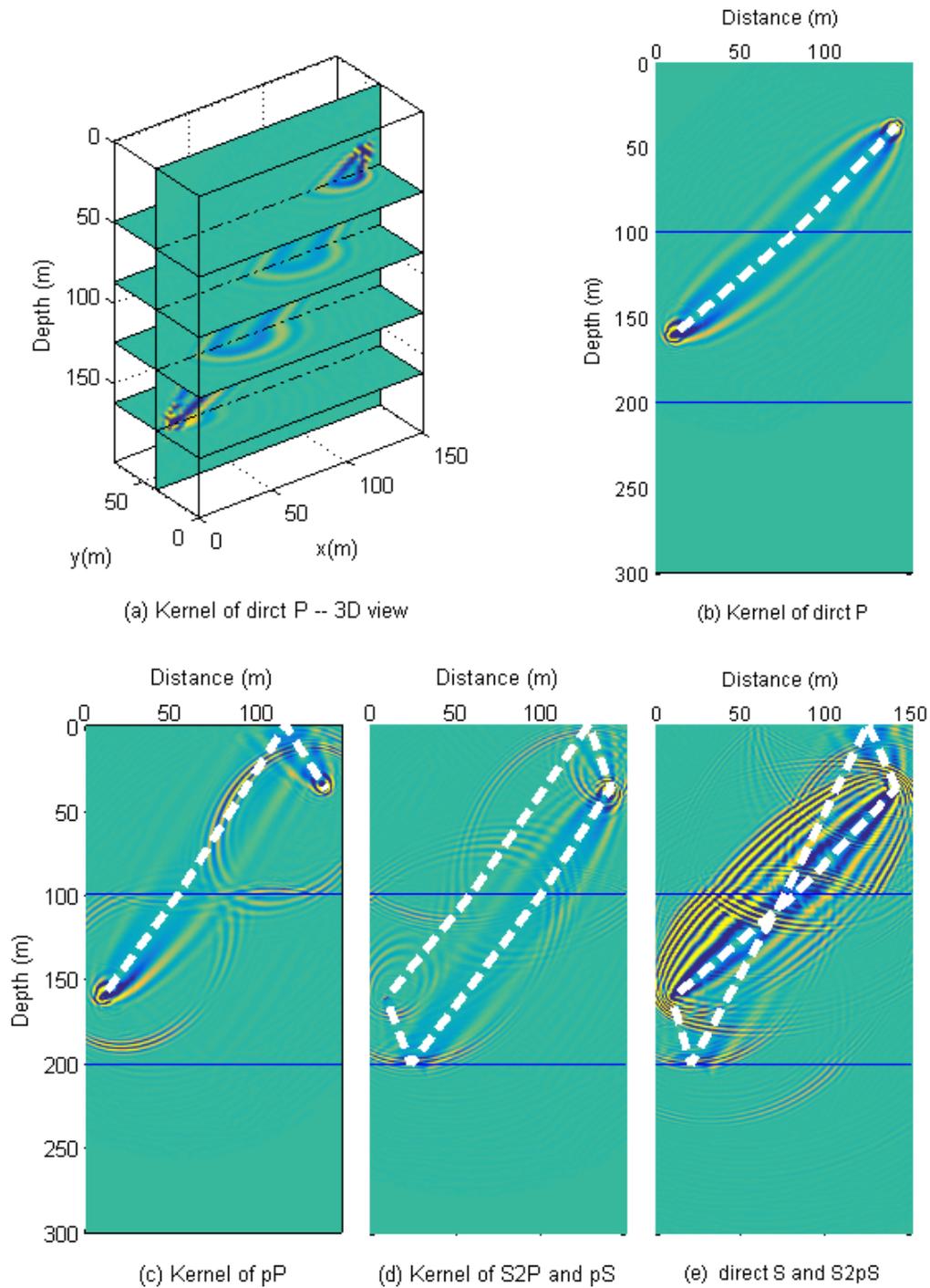
(e) direct S and S2pS

Figure 12. Seismic tomography kernels computed by FGA for the initial velocity model (25), with the thick dashed lines as the actual ray paths of direct P, pP, S2P and pS, direct S, and S2pS signals in Fig. 11, respectively. (a): Kernel computed from the direct P signal – 3D slices view; (b): Kernel computed from the direct P signal; (c): Kernel computed from the pP signal; (d): Kernel computed from the S2P and pS signals; (e): Kernel computed from the direct S and S2pS signals.
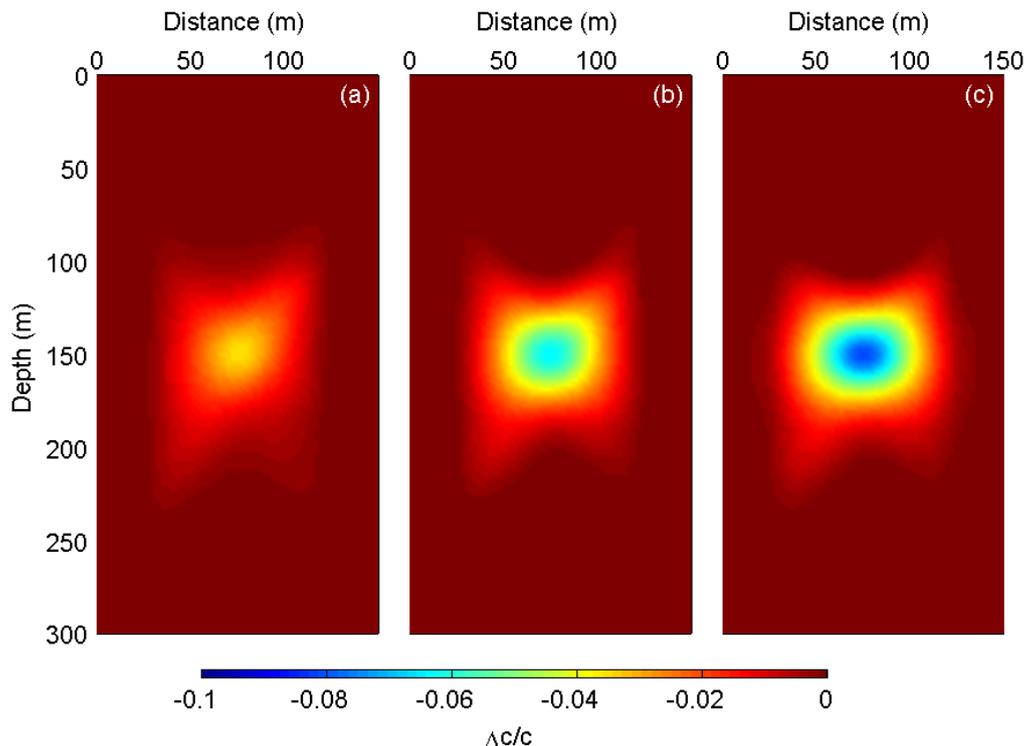
Figure 13. For $\alpha = 10\%$ and $\beta = 1/450$ m^{-2} in eq. (24), the subfigures (a) and (b) are the first two iterations using travel-time tomography, and the subfigure (c) is the third iteration using FWI which removes the artifacts in travel-time tomography and is close to the true velocity profile given in Fig. 10(b).

train neural networks for seismic inversion; 3. to develop the FGA algorithm for the optimal transport theory-based seismic tomography using high-frequency data.

ACKNOWLEDGMENTS

APPENDIX A: DERIVATION OF THE FGA FORMULATION

This appendix is devoted to the mathematical details of deriving the FGA formulation. It is lengthy and technically involved, making use of a few mathematical concepts including weak asymptotic expansion, Schwartz class functions, canonical transform and symplecticity. We shall do our best to make the details as straightforward as possible so that they are accessible for interested readers with different background.

For a sake of clarity and simplicity, we only consider the case of λ and μ being constants in eq. (1) with only "+" wave propagation direction, and the derivation of FGA for general $\lambda(\boldsymbol{x})$ and $\mu(\boldsymbol{x})$ with two-way wave propagation directions will be essentially the same but just with more calculations. We plug eq. (6) into eq. (5), combine the phase functions, and define the total phase $\Phi_{\mathrm{p,s}}$ as

$$\Phi_{\mathrm{p,s}}(t, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{q}, \boldsymbol{p}) = \boldsymbol{P}_{\mathrm{p,s}} \cdot (\boldsymbol{x} - \boldsymbol{Q}_{\mathrm{p,s}}) - \frac{1}{2}|\boldsymbol{x} - \boldsymbol{Q}_{\mathrm{p,s}}|^2 - \boldsymbol{p} \cdot (\boldsymbol{y} - \boldsymbol{q}) - \frac{1}{2}|\boldsymbol{y} - \boldsymbol{q}|^2. \qquad (A.1)$$

We call two functions $f(\boldsymbol{y}, \boldsymbol{q}, \boldsymbol{p})$ and $g(\boldsymbol{y}, \boldsymbol{q}, \boldsymbol{p})$ are weakly equal to each other, denoted by $f \sim g$, if

$$\iiint f(\boldsymbol{y}, \boldsymbol{q}, \boldsymbol{p}) \exp(\mathrm{i}k\Phi_{\mathrm{p,s}}) \, d\boldsymbol{y} \, d\boldsymbol{q} \, d\boldsymbol{p} = \iiint g(\boldsymbol{y}, \boldsymbol{q}, \boldsymbol{p}) \exp(\mathrm{i}k\Phi_{\mathrm{p,s}}) \, d\boldsymbol{y} \, d\boldsymbol{q} \, d\boldsymbol{p}. \qquad (A.2)$$

Note that both f and g can be scalar-, vector-, matrix- or tensor-valued functions.

A map: $(\boldsymbol{q}, \boldsymbol{p}) \rightarrow \left(\boldsymbol{Q}_{\mathrm{p,s}}(\boldsymbol{q}, \boldsymbol{p}), \boldsymbol{P}_{\mathrm{p,s}}(\boldsymbol{q}, \boldsymbol{p})\right)$ is called canonical transformation if the associated Jacobian matrix

$$J_{\mathrm{p,s}}(\boldsymbol{q}, \boldsymbol{p}) = \begin{pmatrix} (\partial_q \boldsymbol{Q}_{\mathrm{p,s}})^{\mathrm{T}}(\boldsymbol{q}, \boldsymbol{p}) & (\partial_p \boldsymbol{Q}_{\mathrm{p,s}})^{\mathrm{T}}(\boldsymbol{q}, \boldsymbol{p}) \\ (\partial_q \boldsymbol{P}_{\mathrm{p,s}})^{\mathrm{T}}(\boldsymbol{q}, \boldsymbol{p}) & (\partial_p \boldsymbol{P}_{\mathrm{p,s}})^{\mathrm{T}}(\boldsymbol{q}, \boldsymbol{p}) \end{pmatrix},$$

is symplectic, i.e., for any $(\boldsymbol{q}, \boldsymbol{p})$,

$$J_{\mathrm{p,s}}^{\mathrm{T}} \begin{pmatrix} 0 & \mathrm{Id}_3 \\ -\mathrm{Id}_3 & 0 \end{pmatrix} J_{\mathrm{p,s}} = \begin{pmatrix} 0 & \mathrm{Id}_3 \\ -\mathrm{Id}_3 & 0 \end{pmatrix}, \qquad (A.3)$$

where Id_3 is a 3×3 identity matrix. It is easy to verify that the Hamiltonian flow given by either eq. (8) or (10) is a canonical transform by observing that

$$\frac{\mathrm{d}}{\mathrm{d}t} J_{\mathrm{p,s}}(\boldsymbol{q}, \boldsymbol{p}) = \begin{pmatrix} \partial_{\boldsymbol{P}}\partial_{\boldsymbol{Q}} H_{\mathrm{p,s}} & \partial_{\boldsymbol{P}}\partial_{\boldsymbol{P}} H_{\mathrm{p,s}} \\ -\partial_{\boldsymbol{Q}} p \partial_{\boldsymbol{Q}} H_{\mathrm{p,s}} & -\partial_{\boldsymbol{Q}}\partial_{\boldsymbol{P}} H_{\mathrm{p,s}} \end{pmatrix} J_{\mathrm{p,s}}(\boldsymbol{q}, \boldsymbol{p}),$$

with the Hamiltonian function $H_{\mathrm{p,s}}(\boldsymbol{Q}, \boldsymbol{P}) = c_{\mathrm{p,s}}(\boldsymbol{Q}_{\mathrm{p,s}})|\boldsymbol{P}_{\mathrm{p,s}}|$, thus $J_{\mathrm{p,s}}(\boldsymbol{q}, \boldsymbol{p})$ satisfies eq. (A.3). This property guarantees that the matrix $Z_{\mathrm{p,s}}$ defined in eq. (15) is always invertible for all $t > 0$ by the following argument, with the subscripts p, s omitted for convenience,

$$Z(\boldsymbol{q}, \boldsymbol{p}) = \partial_z \big(Q(\boldsymbol{q}, \boldsymbol{p}) + \mathrm{i}P(\boldsymbol{q}, \boldsymbol{p})\big) = \begin{pmatrix} \mathrm{i}\mathrm{Id}_3 & \mathrm{Id}_3 \end{pmatrix} J^{\mathrm{T}}(\boldsymbol{q}, \boldsymbol{p}) \begin{pmatrix} -\mathrm{i}\mathrm{Id}_3 \\ \mathrm{Id}_3 \end{pmatrix},$$

which implies

$$
\begin{aligned}
(ZZ^*)(\boldsymbol{q},\boldsymbol{p}) &= \begin{pmatrix} \mathrm{iId}_3 & \mathrm{Id}_3 \end{pmatrix} J^{\mathrm{T}}(\boldsymbol{q},\boldsymbol{p}) \begin{pmatrix} \mathrm{Id}_3 & -\mathrm{iId}_3 \\ \mathrm{iId}_3 & \mathrm{Id}_3 \end{pmatrix} J(\boldsymbol{q},\boldsymbol{p}) \begin{pmatrix} -\mathrm{iId}_3 \\ \mathrm{Id}_3 \end{pmatrix} \\
&= \begin{pmatrix} \mathrm{iId}_3 & \mathrm{Id}_3 \end{pmatrix} (J^{\mathrm{T}}J)(\boldsymbol{q},\boldsymbol{p}) \begin{pmatrix} -\mathrm{iId}_3 \\ \mathrm{Id}_3 \end{pmatrix} \\
&\quad + \begin{pmatrix} \mathrm{iId}_3 & \mathrm{Id}_3 \end{pmatrix} J^{\mathrm{T}}(\boldsymbol{q},\boldsymbol{p}) \begin{pmatrix} 0 & -\mathrm{iId}_3 \\ \mathrm{iId}_3 & 0 \end{pmatrix} J(\boldsymbol{q},\boldsymbol{p}) \begin{pmatrix} -\mathrm{iId}_3 \\ \mathrm{Id}_3 \end{pmatrix} \\
&= \begin{pmatrix} \mathrm{iId}_3 & \mathrm{Id}_3 \end{pmatrix} (J^{\mathrm{T}}J)(\boldsymbol{q},\boldsymbol{p}) \begin{pmatrix} -\mathrm{iId}_3 \\ \mathrm{Id}_3 \end{pmatrix} + 2\mathrm{Id}_3,
\end{aligned}
$$

where Z^* is the conjugate transpose of Z, and in the last equality, we have used eq. (A.3). Therefore, the $Z_{\mathrm{p,s}}^{-1}$ used in eqs. (12), (13) and (14) is a well behaved quantity.

With the theoretical preparation above, we are ready to derive the formulation of FGA. We only consider the case $\boldsymbol{F} = 0$ in eq. (1), and refer to Chai et al. (2018, Section 3.1) for the treatment of a general source time function. We start with the following form of Gaussian wave packet as the solution ansatz,

$$
\boldsymbol{u}_{\mathrm{p,s}}(t,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p}) = \boldsymbol{A}_{\mathrm{p,s}}(t,\boldsymbol{q},\boldsymbol{p}) \exp\big(\mathrm{i}k\Phi_{\mathrm{p,s}}(t,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})\big), \tag{A.4}
$$

with Φ given by eq. (A.1). Remark that, for the readers who are familiar with Gaussian beam method (e.g. Hill 1990, 2001; Nowack et al. 2003; Gray 2005; Gray & Bleistein 2009; Popov et al. 2010), it is easy to see that, a direct plugin of eq. (A.4) into eq. (1) will yield more equations than the number of variables due to the missing of a time-dependent Hessian function in the total phase Φ. Therefore, one can only derive the FGA formulation using the weak asymptotic expansion with a sense of equality defined in eq. (A.2), where the following integration by parts lemma will help to eliminate the extra constraints by converting the powers of $\boldsymbol{x} - \boldsymbol{Q}$ to the powers of k^{-1}.

Lemma of integration by parts: For any vector $\boldsymbol{a}(\boldsymbol{y},\boldsymbol{q},\boldsymbol{p}) = (a_j)$ and matrix $M(\boldsymbol{y},\boldsymbol{q},\boldsymbol{p}) = (M_{ij})$ in Schwartz class, i.e., with decay properties at infinity so that the integrals in eq. (A.2) are well defined, one has the following integration by parts formula in the componentwise form, with $\partial_{\boldsymbol{z}} = (\partial_{z_1}, \partial_{z_2}, \partial_{z_3})$,

$$
\begin{aligned}
a_j(x-Q)_j &\sim -k^{-1}\partial_{z_m}\big(a_j Z_{jm}^{-1}\big), \\
(x-Q)_j M_{jl}(x-Q)_l &\sim k^{-1}\partial_{z_m} Q_j M_{jl} Z_{lm}^{-1} + \mathcal{O}(k^{-2}),
\end{aligned} \tag{A.5}
$$

where the Einstein's index summation convention is used, the matrix Z is given in eq. (15), and we refer to Lu & Yang (2011, 2012a) for the detailed proofs.

Note that eq. (1) with $\boldsymbol{F} = 0$ can be rewritten as the following "curl" form,

$$\rho(\boldsymbol{x})\partial_t^2 \boldsymbol{u} = (\lambda + 2\mu)\nabla(\nabla \cdot \boldsymbol{u}) - \mu\nabla \times \nabla \times \boldsymbol{u}. \tag{A.6}$$

Notice that eq. (A.6) is linear, and thus one can derive the prefactor equations for P- and S-waves individually by assuming $\boldsymbol{A}_{\mathrm{p}} \parallel \boldsymbol{P}$ or $\boldsymbol{A}_{\mathrm{s}} \perp \boldsymbol{P}$ in eq. (A.4), respectively. Without loss of generality, we first consider the prefactor equation for the P-wave, with the governing equations for $\boldsymbol{Q}_{\mathrm{p}}$ and $\boldsymbol{P}_{\mathrm{p}}$ given by eq. (8). For an ease of notations, we shall omit the subscript p in calculating the prefactor equation for the P-wave.

Plugging eq. (A.4) into eq. (A.6) and expanding the asymptotics in the weak sense of (A.2) yield

$$\rho\left(\boldsymbol{A}_{tt} + 2\mathrm{i}k\boldsymbol{A}_t\Phi_t + \mathrm{i}k\boldsymbol{A}\Phi_{tt} - k^2\boldsymbol{A}\Phi_t^2\right) \sim (\lambda + 2\mu)\left(\mathrm{i}k\nabla(\boldsymbol{A} \cdot \nabla\Phi) - k^2(\boldsymbol{A} \cdot \nabla\Phi)\nabla\Phi\right)$$
$$- \mu\left(\mathrm{i}k\,\nabla\times(\nabla\Phi \times \boldsymbol{A}) - k^2\nabla\Phi \times (\nabla\Phi \times \boldsymbol{A})\right). \tag{A.7}$$

The spatial and temporal derivatives of Φ are given by

$$\begin{aligned}
&\nabla\Phi = \boldsymbol{P} + \mathrm{i}(\boldsymbol{x} - \boldsymbol{Q}), \ \Delta\Phi = 3\mathrm{i}, \ \nabla^2\Phi = \mathrm{i}\mathrm{Id}_3, \\
&|\nabla\Phi|^2 = |\boldsymbol{P}|^2 + 2\mathrm{i}\boldsymbol{P} \cdot (\boldsymbol{x} - \boldsymbol{Q}) - |\boldsymbol{x} - \boldsymbol{Q}|^2, \\
&\Phi_t = (\boldsymbol{P}_t - \mathrm{i}\boldsymbol{Q}_t) \cdot (\boldsymbol{x} - \boldsymbol{Q}) - \boldsymbol{P} \cdot \boldsymbol{Q}_t, \\
&\Phi_t^2 = \left[(\boldsymbol{P}_t - \mathrm{i}\boldsymbol{Q}_t) \cdot (\boldsymbol{x} - \boldsymbol{Q})\right]^2 + (\boldsymbol{P} \cdot \boldsymbol{Q}_t)^2 - 2(\boldsymbol{P} \cdot \boldsymbol{Q}_t)\left[(\boldsymbol{P}_t - \mathrm{i}\boldsymbol{Q}_t) \cdot (\boldsymbol{x} - \boldsymbol{Q})\right], \\
&\Phi_{tt} = (\boldsymbol{P}_{tt} - \mathrm{i}\boldsymbol{Q}_{tt}) \cdot (\boldsymbol{x} - \boldsymbol{Q}) - (\boldsymbol{P}_t - \mathrm{i}\boldsymbol{Q}_t) \cdot \boldsymbol{Q}_t - \boldsymbol{P}_t \cdot \boldsymbol{Q}_t - \boldsymbol{P} \cdot \boldsymbol{Q}_{tt}.
\end{aligned} \tag{A.8}$$

Notice that the terms containing $k(\boldsymbol{x} - \boldsymbol{Q})$ will be of $\mathcal{O}(1)$ by the lemma of integration by parts, and for P-waves, $\boldsymbol{P} \times \boldsymbol{A} = 0$ and $\nabla\times\left((\boldsymbol{x} - \boldsymbol{Q}) \times \boldsymbol{A}\right) = -2\boldsymbol{A}$. Plugging the derivatives of Φ in eq. (A.8) into eq. (A.7) produces, after neglecting the $\mathcal{O}(1)$ and lower order terms,

$$\begin{aligned}
2k\rho\boldsymbol{A}_t(\boldsymbol{P} \cdot \boldsymbol{Q}_t) \sim\ & k\rho\boldsymbol{A}\Big(-(\boldsymbol{P}_t - \mathrm{i}\boldsymbol{Q}_t) \cdot \boldsymbol{Q}_t - \boldsymbol{P}_t \cdot \boldsymbol{Q}_t - \boldsymbol{P} \cdot \boldsymbol{Q}_{tt}\Big) \\
&+ \mathrm{i}k\rho\boldsymbol{A}\Big((\boldsymbol{x} - \boldsymbol{Q}) \cdot \big((\boldsymbol{P}_t - \mathrm{i}\boldsymbol{Q}_t) \otimes (\boldsymbol{P}_t - \mathrm{i}\boldsymbol{Q}_t)\big)(\boldsymbol{x} - \boldsymbol{Q})\Big) \\
&+ \mathrm{i}k^2\rho\boldsymbol{A}\Big(-2(\boldsymbol{P} \cdot \boldsymbol{Q}_t)\big((\boldsymbol{P}_t - \mathrm{i}\boldsymbol{Q}_t) \cdot (\boldsymbol{x} - \boldsymbol{Q})\big) + (\boldsymbol{P} \cdot \boldsymbol{Q}_t)^2\Big) \\
&+ (\lambda + 2\mu)\Big(-\mathrm{i}k\boldsymbol{A} + k^2\big((\boldsymbol{A} \cdot \boldsymbol{P})(\boldsymbol{x} - \boldsymbol{Q}) \\
&\quad + (\boldsymbol{A} \otimes \boldsymbol{P})(\boldsymbol{x} - \boldsymbol{Q}) + \mathrm{i}\big((\boldsymbol{x} - \boldsymbol{Q}) \otimes (\boldsymbol{x} - \boldsymbol{Q})\big)\boldsymbol{A}\Big) \\
&- 2\mathrm{i}\mu k\boldsymbol{A} - \mathrm{i}\mu k^2\Big(((\boldsymbol{x} - \boldsymbol{Q}) \cdot \boldsymbol{A})(\boldsymbol{x} - \boldsymbol{Q}) - |\boldsymbol{x} - \boldsymbol{Q}|^2\boldsymbol{A}\Big) \\
&- \mu k^2\Big((\boldsymbol{P} \cdot \boldsymbol{A})(\boldsymbol{x} - \boldsymbol{Q}) - (\boldsymbol{P} \cdot (\boldsymbol{x} - \boldsymbol{Q}))\boldsymbol{A}\Big),
\end{aligned} \tag{A.9}$$

where $\otimes$ means the tensor product, e.g., $(\boldsymbol{A} \otimes \boldsymbol{P})_{jl} = A_j P_l$.

Expanding $\rho(\boldsymbol{x})$ around $\boldsymbol{Q}$ and truncating at order third order,

$$\rho(\boldsymbol{x}) = \rho + \partial_{\boldsymbol{Q}}\rho \cdot (\boldsymbol{x} - \boldsymbol{Q}) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{Q}) \cdot \partial^2_{\boldsymbol{QQ}}\rho(\boldsymbol{x} - \boldsymbol{Q}), \tag{A.10}$$

and noticing that $\rho(\boldsymbol{Q})c^2(\boldsymbol{Q}) = \lambda + 2\mu$ is constant, one has

$$\partial_{\boldsymbol{Q}}\rho c^2 + 2c\partial_{\boldsymbol{Q}}c\rho = 0, \quad \partial_{\boldsymbol{Q}}\rho = -\frac{2c\partial_{\boldsymbol{Q}}c\rho}{c^2}. \tag{A.11}$$

Taking the second derivative for $\partial_{\boldsymbol{QQ}}\rho$ and substituting eq. (A.11) bring

$$\rho(\boldsymbol{x}) = p\rho - \frac{2\partial_{\boldsymbol{Q}}c\rho}{c} \cdot (\boldsymbol{x} - \boldsymbol{Q}) + (\boldsymbol{x} - \boldsymbol{Q}) \cdot \left(3\frac{\partial_{\boldsymbol{Q}}c\partial_{\boldsymbol{Q}}c^T\rho}{c^2} - \frac{\partial_{\boldsymbol{QQ}}c\rho}{c}\right)(\boldsymbol{x} - \boldsymbol{Q}). \tag{A.12}$$

Plugging eqs, (8) and (A.12) into eq. (A.9), and dividing by ρ yield, in componentwise form,

$$
\begin{aligned}
2k\partial_t A_i c|\boldsymbol{P}| \sim & kA_i\left(c(\partial_{\boldsymbol{Q}}c)_j P_j + \mathrm{i}c^2\right) - \mathrm{i}c^2 kA_i \\
& + k^2 A_i(x-Q)_j M_{jl}(x-Q)_l - 2k^2 A_i c^2 P_j(x-Q)_j + c^2 N_i \\
& - \frac{2\mathrm{i}k\mu}{\rho}A_i - \frac{k^2\mu}{\rho}\left(P_j A_j(x-Q)_i - P_j(x-Q)_j A_i\right) \\
& - \frac{\mathrm{i}k^2\mu}{\rho}\left((x-Q)_j A_j(x-Q)_i - |\boldsymbol{x} - \boldsymbol{Q}|^2 A_i\right),
\end{aligned} \tag{A.13}
$$

where M_{jl} and N_i are given as follows,

$$
\begin{aligned}
M_{jl} = & -\mathrm{i}|\boldsymbol{P}|^2 c(\partial^2_{\boldsymbol{Q}}c)_{jk} + c(P_l + P_j)(\partial_{\boldsymbol{Q}}c)_j - \mathrm{i}c^2\frac{P_j P_l}{|\boldsymbol{P}|^2}, \\
N_i = & A_j P_j(x-Q)_i + A_i P_j(x-Q)_j + \mathrm{i}(x-Q)_i(x-Q)_j A_j.
\end{aligned} \tag{A.14}
$$

Assuming that $\boldsymbol{A} = a_{\mathrm{p}}\hat{\boldsymbol{N}}_{\mathrm{p}}$, with $\hat{\boldsymbol{N}}_{\mathrm{p}} = \dfrac{\boldsymbol{P}}{|\boldsymbol{P}|}$, then

$$\partial_t \boldsymbol{A} = \partial_t a_{\mathrm{p}}\hat{\boldsymbol{N}}_{\mathrm{p}} + a_{\mathrm{p}}\partial_t\hat{\boldsymbol{N}}_{\mathrm{p}} = \partial_t a_{\mathrm{p}}\frac{\boldsymbol{P}}{|\boldsymbol{P}|} + a_{\mathrm{p}}\partial_t\left(\frac{\boldsymbol{P}}{|\boldsymbol{P}|}\right). \tag{A.15}$$

Plugging this into eq. (A.13) with the ray equations (8), using the P-wave velocity eq. (3) and grouping in powers of $(\boldsymbol{x} - \boldsymbol{Q})$ produce

$$
\begin{aligned}
2k\partial_t a_{\mathrm{p}} c|\boldsymbol{P}|P_i \sim & -2ka_{\mathrm{p}}c|\boldsymbol{P}|^2\partial_t\left(\frac{P_i}{|\boldsymbol{P}|}\right) + ka_{\mathrm{p}}c(\partial_{\boldsymbol{Q}}c)_j P_j P_i \\
& - \frac{2\mathrm{i}k\mu P_i}{\rho}a_{\mathrm{p}} + k^2 a_{\mathrm{p}}|\boldsymbol{P}|^2\left(c^2 - \frac{\mu}{\rho}\right)(x-Q)_i \\
& + k^2 a_{\mathrm{p}}\left(\frac{\mu}{\rho} - c^2\right)P_j P_i(x-Q)_j + k^2\frac{\mathrm{i}\mu a_{\mathrm{p}}}{\rho}|x-Q|^2 P_i \\
& + k^2\mathrm{i}a_{\mathrm{p}}P_j\left(c^2 - \frac{\mu}{\rho}\right)(x-Q)_j(x-Q)_i + k^2 a_{\mathrm{p}}P_i(x-Q)_j M_{jl}(x-Q)_l.
\end{aligned}
$$

Denoting $\partial_{\boldsymbol{z}} = (\partial_1, \partial_2, \partial_3)$ for an ease of notations, applying eq. (A.5) and dropping the lower order

terms yield

$$2\partial_t a_{\mathrm{p}} c|\boldsymbol{P}|P_i \sim -2k a_{\mathrm{p}} c|\boldsymbol{P}|^2 \partial_t\left(\frac{P_i}{|\boldsymbol{P}|}\right) + a_{\mathrm{p}} c(\partial_{\boldsymbol{Q}} c)_j P_j P_i - \frac{2\mathrm{i}\mu a_{\mathrm{p}}}{\rho} P_i$$
$$- \partial_l\left(\left(\left(|\boldsymbol{P}|^2 - P_j P_i\right) a_{\mathrm{p}}\left(\frac{\mu}{\rho} - c^2\right) Z_{jl}^{-1}\right)\right)$$
$$+ \mathrm{i}a_{\mathrm{p}} P_j\left(c^2 - \frac{\mu}{\rho}\right) Z_{il}^{-1}\partial_l Q_j + \frac{\mathrm{i}\mu a_{\mathrm{p}}}{\rho} P_i Z_{jl}^{-1}\partial_l Q_j \qquad \text{(A.16)}$$
$$+ a_{\mathrm{p}} P_i\left(c(\partial_{\boldsymbol{Q}} c)_j P_r + c P_j(\partial_{\boldsymbol{Q}} c)_j - \mathrm{i}c^2\frac{P_j P_r}{|\boldsymbol{P}|^2}\right) c Z_{rl}^{-1}\partial_l Q_j$$
$$- a_{\mathrm{p}}\mathrm{i}P_i|\boldsymbol{P}|^2 c(\partial_{\boldsymbol{Q}}^2 c)_{jr} Z_{rl}^{-1}\partial_l Q_j.$$

To derive an ordinary differential equation (ODE) instead of a partial differential equation (PDE) for a_{p}, one needs to simplify the terms containing $\partial_k a_{\mathrm{p}}$ as

$$\left(|\boldsymbol{P}|^2 - P_j P_i\right)\partial_l\left(a_{\mathrm{p}}\left(\frac{\mu}{\rho} - c^2\right) Z_{jl}^{-1}\right). \qquad \text{(A.17)}$$

Recall that eq. (A.16) holds in the sense of integral form (A.2), and now we shall consider a strong form of eq. (A.16), i.e., equate the integrands of the integrals on both sides. After taking the dot product of integrands with $\boldsymbol{P}$, one has

$$2\frac{\partial_t a_{\mathrm{p}}}{a_{\mathrm{p}}} = 2\frac{(\partial_{\boldsymbol{Q}} c)_j P_j}{|\boldsymbol{P}|} - \frac{2\mu\mathrm{i}}{\rho c|\boldsymbol{P}|} + \frac{\mathrm{i}\mu}{\rho c|\boldsymbol{P}|}\left(\delta_{ij} - \frac{P_j P_i}{|\boldsymbol{P}|^2}\right) Z_{il}^{-1}\partial_l Q_j$$
$$+ \frac{1}{c}\left(c^2 - \frac{\mu}{\rho}\right)\frac{P_i}{|\boldsymbol{P}|}\partial_l\left(\frac{P_j P_i}{|\boldsymbol{P}|^2}\right) Z_{jl}^{-1} + \frac{1}{|\boldsymbol{P}|}(\partial_{\boldsymbol{Q}} c)_j P_k Z_{kl}^{-1}\partial_l Q_j$$
$$+ \frac{1}{|\boldsymbol{P}|} P_j(\partial_{\boldsymbol{Q}} c)_j Z_{kl}^{-1}\partial_l Q_j - \mathrm{i}|\boldsymbol{P}|(\partial_{\boldsymbol{Q}}^2 c)_{jk} Z_{kl}^{-1}\partial_l Q_j,$$

where the terms (A.17) actually become zero since $\boldsymbol{P}\cdot(|\boldsymbol{P}|^2 - \boldsymbol{P}\otimes\boldsymbol{P}) = 0$, and we have used the fact that

$$\frac{P_i}{|\boldsymbol{P}|}\partial_t\left(\frac{P_i}{|\boldsymbol{P}|}\right) = \frac{1}{2}\left(\frac{P_i}{|\boldsymbol{P}|}\partial_t\left(\frac{P_i}{|\boldsymbol{P}|}\right) + \partial_t\left(\frac{P_i}{|\boldsymbol{P}|}\right)\frac{P_i}{|\boldsymbol{P}|}\right) = \partial_t\left(\frac{P_i}{|\boldsymbol{P}|}\frac{P_i}{|\boldsymbol{P}|}\right) = 0,$$

which implies $\partial_t\left(\frac{P_i}{|\boldsymbol{P}|}\right)P_i = 0$.

Since $Z = \partial_{\boldsymbol{z}}(\boldsymbol{Q} + \mathrm{i}\boldsymbol{P})$ by eq. (15), $\partial_t Z = \partial_t\partial_{\boldsymbol{z}}\boldsymbol{Q} + \mathrm{i}\partial_t\partial_{\boldsymbol{z}}\boldsymbol{P}$. Then eq. (8) implies

$$\partial_t\partial_{\boldsymbol{z}}\boldsymbol{Q} = \partial_{\boldsymbol{z}}\boldsymbol{Q}\frac{\partial_{\boldsymbol{Q}} c\otimes\boldsymbol{P}}{|\boldsymbol{P}|} + c\partial_{\boldsymbol{z}}\boldsymbol{P}\left(\frac{\mathrm{Id}_3}{|\boldsymbol{P}|} - \frac{\boldsymbol{P}\otimes\boldsymbol{P}}{|\boldsymbol{P}|^3}\right),$$
$$\partial_t\partial_{\boldsymbol{z}}\boldsymbol{P} = -|\boldsymbol{P}|\partial_{\boldsymbol{z}}\boldsymbol{Q}\partial_{\boldsymbol{Q}}^2 c - \partial_{\boldsymbol{z}}\boldsymbol{P}\frac{\boldsymbol{P}\otimes\partial_{\boldsymbol{Q}} c}{|\boldsymbol{P}|}. \qquad \text{(A.18)}$$

Using eq. (A.18) for further simplifications give

$$2\frac{\partial_t a_{\mathrm{p}}}{a_{\mathrm{p}}} = 2\frac{(\partial_{\boldsymbol{Q}} c)_i P_i}{|\boldsymbol{P}|} + \mathrm{Tr}\left(Z^{-1}\partial_t Z\right)$$
$$+ \frac{1}{c}\left(c^2 - \frac{\mu}{\rho}\right)\frac{P_i}{|\boldsymbol{P}|}\partial_l\left(\frac{P_j P_i}{|\boldsymbol{P}|^2}\right) Z_{jl}^{-1} - \frac{1}{c|\boldsymbol{P}|}\left(c^2 - \frac{\mu}{\rho}\right)\left(\delta_{ij} - \frac{P_j P_i}{|\boldsymbol{P}|^2}\right) Z_{il}^{-1}\partial_l P_j,$$
$$\text{(A.19)}$$

where the last two terms can be grouped as

$$\frac{1}{c|\boldsymbol{P}|}\left(c^2 - \frac{\mu}{\rho}\right)\partial_l\left(P_i\left(\frac{P_jP_i}{|\boldsymbol{P}|^2} - \delta_{ij}\right)\right)Z_{jl}^{-1} = 0,\tag{A.20}$$

which implies a clean ODE for a_{p} as in eq. (12),

$$\frac{\mathrm{d}a_{\mathrm{p}}}{\mathrm{d}t} = a_{\mathrm{p}}\left(\frac{\partial_{\boldsymbol{Q}}c\cdot\boldsymbol{P}}{|\boldsymbol{P}|} + \frac{1}{2}\mathrm{Tr}\left(Z^{-1}\frac{\mathrm{d}Z}{\mathrm{d}t}\right)\right).\tag{A.21}$$

Similarly, one can derive the prefactor equations for SV- and SH-waves by assuming $\boldsymbol{A} = a_{\mathrm{sv}}\hat{\boldsymbol{N}}_{\mathrm{sv}} + a_{\mathrm{sh}}\hat{\boldsymbol{N}}_{\mathrm{sh}}$ with $\hat{\boldsymbol{N}}_{\mathrm{sv}} \perp \boldsymbol{P}$, $\hat{\boldsymbol{N}}_{\mathrm{sh}} \perp \boldsymbol{P}$ and $\hat{\boldsymbol{N}}_{\mathrm{sv}} \perp \hat{\boldsymbol{N}}_{\mathrm{sh}}$ in eq. (A.4). The calculations will be essentially the same as the prefactor equation for P-waves except that one will have the diabatic coupling terms of $\hat{\boldsymbol{N}}_{\mathrm{sv}}$ and $\hat{\boldsymbol{N}}_{\mathrm{sh}}$ as shown below,

$$\begin{aligned}\frac{\mathrm{d}a_{\mathrm{sv}}}{\mathrm{d}t} &= a_{\mathrm{sv}}\left(\frac{\partial_{\boldsymbol{Q}_{\mathrm{s}}}c_{\mathrm{s}}\cdot\boldsymbol{P}_{\mathrm{s}}}{|\boldsymbol{P}_{\mathrm{s}}|} + \frac{1}{2}\mathrm{Tr}\left(Z_{\mathrm{s}}^{-1}\frac{\mathrm{d}Z_{\mathrm{s}}}{\mathrm{d}t}\right)\right) - a_{\mathrm{sh}}\left(\frac{\mathrm{d}\hat{\boldsymbol{N}}_{\mathrm{sh}}}{\mathrm{d}t}\cdot\hat{\boldsymbol{N}}_{\mathrm{sv}} + m_{\mathrm{sh}\to\mathrm{sv}}\right),\\\frac{\mathrm{d}a_{\mathrm{sh}}}{\mathrm{d}t} &= a_{\mathrm{sh}}\left(\frac{\partial_{\boldsymbol{Q}_{\mathrm{s}}}c_{\mathrm{s}}\cdot\boldsymbol{P}_{\mathrm{s}}}{|\boldsymbol{P}_{\mathrm{s}}|} + \frac{1}{2}\mathrm{Tr}\left(Z_{\mathrm{s}}^{-1}\frac{\mathrm{d}Z_{\mathrm{s}}}{\mathrm{d}t}\right)\right) - a_{\mathrm{sv}}\left(\frac{\mathrm{d}\hat{\boldsymbol{N}}_{\mathrm{sv}}}{\mathrm{d}t}\cdot\hat{\boldsymbol{N}}_{\mathrm{sh}} + m_{\mathrm{sv}\to\mathrm{sh}}\right),\end{aligned}\tag{A.22}$$

where the interaction terms are given by

$$m_{\mathrm{sh}\to\mathrm{sv}} = \mathrm{i}\frac{\lambda+\mu}{\rho c_{\mathrm{s}}|P_{\mathrm{s}}|}\left(\hat{\boldsymbol{N}}_{\mathrm{sv}}\cdot(Z_{\mathrm{s}}^{-1}\partial_{\boldsymbol{z}}\boldsymbol{Q}_s)\hat{\boldsymbol{N}}_{\mathrm{sh}} - \hat{\boldsymbol{N}}_{\mathrm{sh}}\cdot(Z_{\mathrm{s}}^{-1}\partial_{\boldsymbol{z}}\boldsymbol{Q}_s)\hat{\boldsymbol{N}}_{\mathrm{sv}}\right),$$

$$m_{\mathrm{sv}\to\mathrm{sh}} = -m_{\mathrm{sh}\to\mathrm{sv}}.$$

Also, note that by $\hat{\boldsymbol{N}}_{\mathrm{sv}} \perp \hat{\boldsymbol{N}}_{\mathrm{sh}}$, one has that $\dfrac{\mathrm{d}\hat{\boldsymbol{N}}_{\mathrm{sh}}}{\mathrm{d}t}\cdot\hat{\boldsymbol{N}}_{\mathrm{sv}} + \dfrac{\mathrm{d}\hat{\boldsymbol{N}}_{\mathrm{sv}}}{\mathrm{d}t}\cdot\hat{\boldsymbol{N}}_{\mathrm{sh}} = 0$.

Next, we shall show that $m_{\mathrm{sh}\to\mathrm{sv}} = m_{\mathrm{sv}\to\mathrm{sh}} = 0$ by proving that $Z_{\mathrm{s}}^{-1}\partial_{\boldsymbol{z}}\boldsymbol{Q}_s$ is symmetric using the following argument. Eq. (A.3) implies, with the subscript s omitted for convenience,

$$\partial_{\boldsymbol{q}}\boldsymbol{Q}(\partial_{\boldsymbol{q}}\boldsymbol{P})^{\mathrm{T}} - \partial_{\boldsymbol{q}}\boldsymbol{P}(\partial_{\boldsymbol{q}}\boldsymbol{Q})^{\mathrm{T}} = 0_{3\times3},\tag{A.23}$$

$$\partial_{\boldsymbol{q}}\boldsymbol{Q}(\partial_{\boldsymbol{p}}\boldsymbol{P})^{\mathrm{T}} - \partial_{\boldsymbol{q}}\boldsymbol{P}(\partial_{\boldsymbol{p}}\boldsymbol{Q})^{\mathrm{T}} = \mathrm{Id}_3,\tag{A.24}$$

$$\partial_{\boldsymbol{p}}\boldsymbol{Q}(\partial_{\boldsymbol{q}}\boldsymbol{P})^{\mathrm{T}} - \partial_{\boldsymbol{p}}\boldsymbol{P}(\partial_{\boldsymbol{q}}\boldsymbol{Q})^{\mathrm{T}} = -\mathrm{Id}_3,\tag{A.25}$$

$$\partial_{\boldsymbol{p}}\boldsymbol{Q}(\partial_{\boldsymbol{p}}\boldsymbol{P})^{\mathrm{T}} - \partial_{\boldsymbol{p}}\boldsymbol{P}(\partial_{\boldsymbol{p}}\boldsymbol{Q})^{\mathrm{T}} = 0_{3\times3},\tag{A.26}$$

where $0_{3\times3}$ is 3-by-3 zero matrix.

Eq. (A.23)$-\mathrm{i}\times$ Eq. (A.25) gives

$$\partial_{\boldsymbol{z}}\boldsymbol{Q}(\partial_{\boldsymbol{q}}\boldsymbol{P})^{\mathrm{T}} - \partial_{\boldsymbol{z}}\boldsymbol{P}(\partial_{\boldsymbol{q}}\boldsymbol{Q})^{\mathrm{T}} = \mathrm{i}\mathrm{Id}_3.\tag{A.27}$$

Eq. (A.24)$-\mathrm{i}\times$ Eq. (A.26) gives

$$\partial_{\boldsymbol{z}}\boldsymbol{Q}(\partial_{\boldsymbol{p}}\boldsymbol{P})^{\mathrm{T}} - \partial_{\boldsymbol{z}}\boldsymbol{P}(\partial_{\boldsymbol{p}}\boldsymbol{Q})^{\mathrm{T}} = \mathrm{Id}_3.\tag{A.28}$$

Eq. (A.27)$-\mathrm{i}\times$ Eq. (A.28) gives

$$\partial_{\boldsymbol{z}}\boldsymbol{Q}(\partial_{\boldsymbol{z}}\boldsymbol{P})^{\mathrm{T}} - \partial_{\boldsymbol{z}}\boldsymbol{P}(\partial_{\boldsymbol{z}}\boldsymbol{Q})^{\mathrm{T}} = 0_{3\times3}.$$

Combined with $\partial_z \boldsymbol{Q}(\partial_z \boldsymbol{Q})^{\mathrm{T}} - \partial_z \boldsymbol{Q}(\partial_z \boldsymbol{Q})^{\mathrm{T}} = 0_{3\times3}$, one has

$$\partial_z Q Z^{\mathrm{T}} - Z(\partial_z Q)^{\mathrm{T}} = 0,$$

which implies $Z^{-1}\partial_z Q = (\partial_z Q)^{\mathrm{T}}(Z^{\mathrm{T}})^{-1} = (\partial_z Q)^{\mathrm{T}}(Z^{-1})^{\mathrm{T}} = (Z^{-1}\partial_z Q)^{\mathrm{T}}$. Therefore, $Z^{-1}\partial_z Q$ is symmetric, and then $m_{\mathrm{sh}\to\mathrm{sv}} = m_{\mathrm{sv}\to\mathrm{sh}} = 0$, which brings eqs. (13) and (14) by eq. (A.22).

APPENDIX B: DERIVATION OF THE INTERFACE CONDITIONS FOR $\partial_Z Q$ AND $\partial_Z P$

The derivation of the interface conditions for $\partial_z \boldsymbol{Q}_{\mathrm{p,s}}$ and $\partial_z \boldsymbol{P}_{\mathrm{p,s}}$ requires to use the conservation of level set functions designed in the Eulerian frozen Gaussian approximation formula (Lu & Yang 2012b; Wei & Yang 2012), where the idea is to use the following Liouville operator to describe the dyanmics of Gaussian wave packet on phase plane,

$$\mathcal{L}_{\mathrm{p,s}} = \partial_t + \partial_{\boldsymbol{p}} H_{\mathrm{p,s}} \cdot \partial_{\boldsymbol{q}} - \partial_{\boldsymbol{q}} H_{\mathrm{p,s}} \cdot \partial_{\boldsymbol{p}} \tag{B.1}$$

whose corresponding characteristic equations are given by the Hamiltonian systems eqs. (8) and (10) with $H_{\mathrm{p,s}} = \pm c_{\mathrm{p,s}}(\boldsymbol{q})|\boldsymbol{p}|$. For example, the prefactor equation of the P-wave for the "+" wave propagation direction is given by, in the Eulerian formulation,

$$\mathcal{L}a_{\mathrm{p}} = a_{\mathrm{p}}\left(\frac{\partial_{\boldsymbol{Q}_{\mathrm{p}}} c_{\mathrm{p}}(\boldsymbol{Q}_{\mathrm{p}}) \cdot \boldsymbol{P}_{\mathrm{p}}}{|\boldsymbol{P}_{\mathrm{p}}|} + \frac{1}{2}\mathrm{Tr}\left(Z_{\mathrm{p}}^{-1}\partial_t Z_{\mathrm{p}}\right)\right).$$

We shall derive the interface conditions of $\partial_z \boldsymbol{Q}$ and $\partial_z \boldsymbol{P}$ for the transmitted P-wave and the "+" wave propagation direction, and omit the subscript "p" in the following derivation for a sake of simplicity. Consider a level set function $\boldsymbol{\phi}(t, \boldsymbol{q}, \boldsymbol{p}) = (\phi_1, \phi_2, \phi_3)$ which satisfies

$$\mathcal{L}\boldsymbol{\phi} = 0, \quad \text{with} \quad \boldsymbol{\phi}(0, \boldsymbol{q}, \boldsymbol{p}) = \boldsymbol{p} + \mathrm{i}\boldsymbol{q}, \tag{B.2}$$

then the Eulerian formulation of FGA (Lu & Yang 2012b) shows that

$$\partial_z \boldsymbol{Q} = (\partial_{\boldsymbol{p}} \boldsymbol{\phi})^{\mathrm{T}}, \qquad \partial_z \boldsymbol{P} = -(\partial_{\boldsymbol{q}} \boldsymbol{\phi})^{\mathrm{T}}. \tag{B.3}$$

We will follow the strategy described in Chai et al. (2018); Wei & Yang (2012), and consider the case illustrated in Fig. 7, where the level set functions $\boldsymbol{\phi}^{\mathrm{re,tr}}$ for the transmitted P-waves satisfy the same evolution as $\boldsymbol{\phi}$ in eq. (B.2) with the following interface conditions

$$\boldsymbol{\phi}^{\mathrm{tr}}(t, \boldsymbol{q}^{\mathrm{tr}}, \boldsymbol{p}^{\mathrm{tr}}) = \boldsymbol{\phi}^{\mathrm{tr}}(t, \boldsymbol{q}^{\mathrm{in}}, \boldsymbol{p}^{\mathrm{in}}). \tag{B.4}$$

Differentiating eq. (B.4) by the definition of partial derivatives and chain rule, and making use of

eq. (20) yield

$$\partial_{p_x}\phi^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{tr}},\boldsymbol{p}^{\mathrm{tr}}) = \left(n_1^{-2}-1\right)\frac{p_x}{p_z}\partial_{p_z}\phi^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{in}},\boldsymbol{p}^{\mathrm{in}}) + \partial_{p_x}\phi^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{in}},\boldsymbol{p}^{\mathrm{in}}),$$

$$\partial_{p_y}\phi^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{tr}},\boldsymbol{p}^{\mathrm{tr}}) = \left(n_1^{-2}-1\right)\frac{p_y}{p_z}\partial_{p_z}\phi^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{in}},\boldsymbol{p}^{\mathrm{in}}) + \partial_{p_y}\phi^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{in}},\boldsymbol{p}^{\mathrm{in}}), \tag{B.5}$$

$$\partial_{p_z}\phi^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{tr}},\boldsymbol{p}^{\mathrm{tr}}) = n_1^{-2}\frac{p_z^{\mathrm{tr}}}{p_z}\partial_{p_z}\phi^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{in}},\boldsymbol{p}^{\mathrm{in}}).$$

Moreover, differentiating eq. (B.4) with respect to t gives

$$\partial_t\phi^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{tr}},\boldsymbol{p}^{\mathrm{tr}}) = \partial_t\phi^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{in}},\boldsymbol{p}^{\mathrm{in}}), \tag{B.6}$$

which implies by eq. (B.1), at the interface,

$$\left[\nabla_{\boldsymbol{p}}H\cdot\nabla_{\boldsymbol{q}}\phi^{\mathrm{tr}}\right]_{q_z=z_0} = \left[\nabla_{\boldsymbol{q}}H\cdot\nabla_{\boldsymbol{p}}\phi^{\mathrm{tr}}\right]_{q_z=z_0}, \tag{B.7}$$

with $[\cdot]$ denoting the jump function. Therefore,

$$\partial_{q_x}\phi^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{tr}},\boldsymbol{p}^{\mathrm{tr}}) = \partial_{q_x}\phi^t(t,\boldsymbol{q}^{\mathrm{in}},\boldsymbol{p}^{\mathrm{in}}),$$

$$\partial_{q_y}\phi^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{tr}},\boldsymbol{p}^{\mathrm{tr}}) = \partial_{q_y}\phi^t(t,\boldsymbol{q}^{\mathrm{in}},\boldsymbol{p}^{\mathrm{in}}),$$

$$\partial_{q_z}\phi^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{tr}},\boldsymbol{p}^{\mathrm{tr}}) = n_1^2\frac{p_z}{p_z^{\mathrm{tr}}}\partial_{q_z}\phi^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{in}},\boldsymbol{p}^{\mathrm{in}}) + (n_1^2-1)\left(\frac{p_x}{p_z^{\mathrm{tr}}}\partial_{q_x} + \frac{p_y}{p_z^{\mathrm{tr}}}\partial_{q_y}\right)\phi^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{in}},\boldsymbol{p}^{\mathrm{in}})$$

$$+ \frac{|\boldsymbol{p}^{\mathrm{tr}}|}{c^\vee p_z^{\mathrm{tr}}}\left[|\boldsymbol{p}|\nabla_{\boldsymbol{q}}c(\boldsymbol{q})\cdot\nabla_{\boldsymbol{p}}\phi^{\mathrm{tr}}(t,\boldsymbol{q},\boldsymbol{p})\right]_{q_z=z_0}. \tag{B.8}$$

Then eqs. (B.3) and (B.5) imply the interface condition for $\partial_z\boldsymbol{Q}$ in eq. (19), while eqs. (B.3) and (B.8) imply the interface condition for $\partial_z\boldsymbol{P}$ in eq. (19) for the transmitted P-wave. The other cases including the interface conditions for the reflected P-wave, transmitted and reflected S-wave can be derived in an essentially same way.

REFERENCES

Aki, K. & 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.

Aki, K., Christofferson, A., & Husebye, E., 1977. Determination of the three-dimensional seismic structure of the lithosphere, *J. Geophys. Res.*, **82**, 277–296.

Berry, M., 1984. Quantal phase factors accompanying adiabatic changes, *Proc. Roy. Soc. Lond. A*, **392**, 45–57.

Cerveny, V., Popov, M. M., & Psencik, I., 1982. Computation of wave fields in inhomogeneous media – Gaussian beam approach, *Geophys. J. Roy. Astr. Soc.*, **70**, 109–128.

Chai, L., Tong, P., & Yang, X., 2017. Frozen Gaussian approximation for 3-D seismic wave propagation, *Geophysical Journal International*, **208**(1), 59–74.

Chai, L., Tong, P., & Yang, X., 2018. Frozen Gaussian approximation for 3-D seismic tomography, *Inverse Problems*, **34**, 055004.

Chapman, C., 1976. Exact and approximate ray theory in vertically inhomogeneous media, *Geophys. J. R. astr. Soc.*, **6**, 201–233.

Chapman, C. & Drummond, R., 1982. Body-wave seismograms in inhomogeneous media using Maslov asymptotic theory, *Bull. Seism. Soc. Am.*, **72**, 277–317.

Chen, J., Chen, Y., Wu, H., & Yang, D., 2017. The quadratic Wasserstein metric for earthquake location, arXiv:1710.10447.

Engquist, B. & Froese, B., 2014. Application of the wasserstein metric to seismic signals, *Commun. Math. Sci.*, **12**, 979–988.

Fomel, S. & Tanushev, N., 2009. Time-domain seismic imaging using beams, in *79th Ann. Internat. Mtg*, pp. 2747–2752, Soc. of Expl. Geophys.

Gray, S. & Bleistein, N., 2009. True-amplitude gaussian-beam migration, *Geophysics*, **74**(2), S11–S23.

Gray, S. H., 1986. Efficient traveltime calculations for Kirchhoff migration, *Geophysics*, **51**, 1685–1688.

Gray, S. H., 2005. Gaussian beam migration of common-shot records, *Geophysics*, **70**, S71–S77.

Helmberger, D., 1968. The crust-mantle transition in the Bering sea, *Bull. Seism. Soc. Am.*, **58**, 179–214.

Helmberger, D. & Ni, S., 2005. Approximate 3D Body-Wave Synthetics for Tomographic Models, *Bull. Seism. Soc. Am.*, **95**, 212–224.

Hill, N. R., 1990. Gaussian beam migration, *Geophysics*, **55**, 1416–1428.

Hill, N. R., 2001. Prestack Gaussian-beam depth migration, *Geophysics*, **66**, 1240–1250.

IRIS DMC, 2010. Data Services Products: iasp91, P & S velocity reference Earth model.

Keho, T. H. & Beydoun, W. B., 1988. Paraxial ray Kirchhoff migration, *Geophysics*, **53**, 1540–1546.

Kennett, B., 1991. (complier and editor). IASPEI 1991 seismological tables, Bibliotech, Canberra, Australia.

Komatitsch, D. & Tromp, J., 1999. Introduction to the spectral element method for three-dimensional seismic wave propagation, *Geophys. J. Int.*, **139**(3), 806–822.

Komatitsch, D., Tsuboi, S., & Tromp, J., 2005. The spectral-element method in seismology, in *Seismic Earth:*

Array Analysis of Broadband Seismograms: Geophysical monograph. 157, AGU, pp. 205–228, eds Levander, A. & Nolet, G., Washington DC, USA.

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

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

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

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

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

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J., 2016. An optimal transport approach for seismic tomography: application to 3D full waveform inversion, *Inverse Problems*, **32**, 115008.

Nakamichi, H., Hamaguchi, H., Tanaka, S., Ueki, S., Nishimura, T., & Hasegawa, A., 2003. Source mechanisms of deep and intermediate-depth low-frequency earthquakes beneath iwate volcano, northeastern japan, *Geophysical Journal International*, **154**(3), 811–828.

Ni, S., Ding, X., & Helmberger, D., 2000. Constructing synthetics from deep earth tomographic models, *Geophys. J. Int.*, **140**, 71–82.

Nowack, R. L., Sen, M. K., & Stoffa, P. L., 2003. Gaussian beam migration for sparse common-shot and common-receiver data, in *SEG Technical Program Expanded Abstracts 2003*, pp. 1114–1117, Society of Exploration Geophysicists.

Popov, M. M., Semtchenok, N. M., Popov, P. M., & Verdel, A. R., 2010. Depth migration by the Gaussian beam summation method, *Geophysics*, **75**, S81–S93.

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

Qian, J. & Ying, L., 2010. Fast Gaussian wavepacket transforms and Gaussian beams for the Schrödinger equation, *J. Comput. Phys.*, **229**, 7848–7873.

Qiu, L., Ramos-Martínez, J., Valenciano, A., Yang, Y., & Engquist, B., 2017. Full-waveform inversion with an exponentially encoded optimal-transport norm, in *SEG Technical Program Expanded Abstracts 2017*, pp. 1286–1290, Society of Exploration Geophysicists.

Rawlinson, N., Pozgay, S., & Fishwick, S., 2010. Seismic tomography: A window into deep Earth, *Phys. Earth Planet. Inter.*, **178**(3-4), 101–135.

Rickers, F., Fichtner, A., & 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.

Simon, B., 1983. Holonomy, the quantum adiabatic theorem, and Berrys phase, *Phys. Rev. Lett.*, **51**, 2167.

Simute, S., Steptoe, H., Cobden, L., & Fichtner, A., 2016. Full-waveform inversion of the Japanese Islands region, *Journal of Geophysical Research: Solid Earth*, **121**, 3722–3741.

Tape, C., Liu, Q., & Tromp, J., 2007. Finite-frequency tomography using adjoint methods-Methodology and examples using membrane surface waves, *Geophys. J. Int.*, **168**(3), 1105–1129.

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

Tape, C., Liu, Q., Maggi, A., & Tromp, J., 2010. Seismic tomography of the southern California crust based on spectral-element and adjoint methods, *Geophys. J. Int.*, **180**(1), 433–462.

Tong, P., Chen, C.-W., Komatitsch, D., Basini, P., & Liu, Q., 2014. High-resolution seismic array imaging based on an SEM-FK hybrid method, *Geophys. J. Int.*, **197**(1), 369–395.

Tong, P., Yang, D., Li, D., & 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**(7), 3165–3175.

Tromp, J., Tape, C., & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, **160**(1), 195–216.

Tromp, J., Komatitsch, J., & Liu, Q., 2008. Spectral-element and adjoint methods in seismology, *Comm. in Comput. Phys*, **3**, 1–32.

Vidale, J. & Helmberger, D., 1988. Elastic finite-difference modeling of the 1971 San-Fernando, California earthquake, *Bull. Seism. Soc. Am.*, **78**, 122–141.

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

Wei, D. & Yang, X., 2012. Eulerian gaussian beam method for high frequency wave propagation in heterogeneous media with discontinuities in one direction, *Commun. Math. Sci.*, **10**, 1287–1299.

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

Yang, Y., Engquist, B., Sun, J., & Hamfeldt, B., 2018. Application of optimal transport and the quadratic wasserstein metric to full-waveform inversion, *Geophysics*, **83**, R43–R62.

Yilmaz, O., 2001. *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data*, Society of Exploration Geophysicists.

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

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