On-the-fly Full Hessian Kernel Calculations Based upon Seismic Wave Simulations

Yujiang Xie[1], Catherine Rychert[1], Nicholas Harmon[1], Qinya Liu[2], and Dirk Gajewski[3]

[1] Ocean and Earth Science, University of Southampton, UK

[2] Department of Physics & Department of Earth Sciences, University of Toronto, Canada

[3] Institute of Geophysics, University of Hamburg, Germany

*Corresponding author; E-mail:yujiang.xie@soton.ac.uk

Abstract

Full waveform inversion (FWI) or adjoint tomography has routinely been performed to image the internal structure of the Earth at high resolution. This is typically done using the Fréchet kernels and the approximate Hessian or the approximate inverse Hessian because of the high computational cost of computing and storing the full Hessian. Alternatively, the full Hessian kernels can be used to improve inversion resolutions, convergence rates, and possibly mitigate inter-parameter tradeoffs. The storage requirements of the full Hessian kernel calculations can be reduced by compression methods, but often at a price of accuracy depending on the compression factor. Here we present open-source codes to compute both Fréchet and full Hessian kernels on the fly (in the computer RAM) through simultaneously solving four wave equations, which we call QuadSEM. By recomputing two forward fields at the same time that two adjoint fields are calculated during the adjoint simulation, QuadSEM constructs the full Hessian kernels using the exact forward and adjoint fields. In addition, we also implement an alternative approach based on the wavefield storage method (WSM), which stores forward wavefields every kth ($k \geq 1$) time step during the forward simulation and reads them back into memory during the adjoint simulation for kernel construction. Both Fréchet and full Hessian kernels can be computed simultaneously through the QuadSEM or the WSM code, only doubling the computational cost compared with the computation of Fréchet kernels alone. Compared with WSM, QuadSEM can reduce the disk space and I/O cost by three orders of magnitude in the presented examples while using 15,000 time steps. Numerical examples are presented to demonstrate the functionality of the methods, and the computer codes are provided with this contribution.

Introduction

In the past thirty years, full waveform inversion (FWI), or sometimes interchangeably known as adjoint tomography in regional or global seismology, has become popular and widely used for imaging the Earth's internal structures at multiple scales (e.g., Bamberger et al., 1982; Lailly, 1983; Tarantola, 1984, 1988; Gauthier et al., 1986; Igel et al., 1996; Pratt et al., 1998; Tape et al., 2009; Fichtner et al., 2009; Virieux and Operto, 2009; Liu and Gu, 2012; Zhu et al., 2012; French and Romanowicz, 2015; Bozdağ et al., 2016; Tromp, 2020). Sensitivity kernels which indicate the sensitivity of seismograms to model parameters are a key component of full-waveform inversion or more generally the adjoint tomography. Typically, two types of sensitivity kernels are discussed in exploration, regional and global seismology context, the first-order derivatives of the seismological data functionals, Fréchet kernels (e.g., Dahlen et al., 2000; Tromp et al., 2005), and the second-order partial derivatives of functionals applied to an arbitrary model update, known as Hessian vector products or Hessian kernels (e.g., Pratt et al., 1998; Epanomeritakis et al., 2008; Fichtner and Trampert, 2011; Métivier et al., 2013).

Fréchet kernels are widely used for waveform inversions or adjoint tomography via the scattering-integral methods (e.g., Chen et al., 2007a,b) or the adjoint methods (e.g., Lions, 1968; Lailly, 1983; Tarantola, 1984, 1988; Tromp et al., 2005; Fichtner et al., 2006; Plessix, 2006). The kernels can be computed via the correlations of incident forward wavefields with adjoint fields (e.g., Bamberger et al., 1982; Tromp et al., 2005). There are mainly two strategies to obtain the forward fields during the adjoint simulation for correlation. One is to write the forward fields onto a disk often with compression during the forward simulation and then during the adjoint simulation for the adjoint field, read the required forward fields back into the temporary memory. Compression schemes include the temporal-spatial compression (e.g., Fichtner et al., 2009) and the lossless or lossy compression (e.g., Hanzich et al., 2013; Lindstrom et al., 2016; Boehm et al., 2016), for instance, based on lossless or lossy compression techniques (e.g., Unat et al., 2009; Weiser and Götschel, 2012; Götschel and Weiser, 2015). The other approach is to only store the forward wavefield at selected time steps, called checkpoints, and during the adjoint simulation, re-solve the forward problem based on these selected time steps (e.g., Symes, 2007; Anderson et al., 2012; Komatitsch et al., 2016), for instance, based on the checkpointing algorithms (e.g., Griewank and Walther, 2000; Charpentier, 2001; Walther and Griewank, 2004). For the elastic cases or the anelastic cases with the consideration of physical dispersion only, the forward fields can be completely reconstructed via the boundary values and the last snapshot of the forward fields during adjoint simulation (e.g., Gauthier et al., 1986; Tromp et al., 2005; Liu and Tromp, 2006).

The use of Hessian in FWI may increase convergence rates, improve model resolutions, and possibly mitigate inter-parameter tradeoffs. However, the exact computation of the full Hessian matrix is computationally prohibitive. Instead, the approximate Hessian or the approximate inverse Hessian has been computed based on the Gauss-Newton or the quasi-Newton approaches (e.g., Pratt et al., 1998; Virieux and Operto, 2009, and among others). In contrast to the computation of Fréchet kernels, which uses two fields, the forward and the adjoint fields, the computation of full Hessian kernels involves four fields (e.g., Fichtner and Trampert, 2011), two forward fields and two adjoint fields. The utility the full or approximate Hessian kernels have been demonstrated in truncated-Newton FWI for exploration seismology (e.g., Métivier et al., 2013, 2014; Pan et al., 2017; Yang et al., 2018; Matharu and Sacchi, 2019), for instance, based on classical wavefield storage, checkpointing, and/or the finite-difference wave simulations. For large models, calculating the full Hessian kernels in the classical storage approach is challenging given the large space required to store the multiple wavefields and the associated I/O expense during simulations. Therefore, approximate Hessian kernels are used instead. For instance, for resolution analysis, they are estimated by a finite-difference approximation using the gradients from two nearby iterations (e.g., Zhu et al., 2012; Bozdağ et al., 2016). Luo et al. (2014) also derived the Hessian kernel formulas and then used the diagonal terms of the Hessian to construct four preconditioners for FWI and resolution analysis. Similar to those for Fréchet kernels, the formulas for these preconditioner operators involve the correlations of forward and adjoint fields and therefore can be computed based on the adjoint method (Tromp et al., 2005).

Compression methods have also been applied to the forward and adjoint wavefield storage for the computation of the Hessian kernels (Boehm and Ulbrich, 2015). In this case, as the decompressed wavefields also appear as a distributed source term in the two auxiliary wave equations and errors are propagated to the two perturbed wavefields resulting in less accurate Hessian kernel construction (Boehm et al., 2016). For instance, using decompressed fields and the trust-region Newton PCG method only resulted in a small improvement compared with LBFGS updates for FWI (Boehm and Ulbrich, 2015). This is mainly due to the trade-off between high compression factors and low storage requirements with the accuracy of conjugate-gradient (CG) update. For this reason, resolving small and/or weakly perturbed scatterers may still be challenging since the perturbed wavefields and the compressed/decompressed errors may be of the same order of magnitude for a high compression factor. Alternative Hessian kernels, for instance, the reduced Hessian kernels (e.g., Epanomeritakis et al., 2008) have also computed and used in FWI for efficiency purposes.

In this paper, we present a computationally efficient method to construct the full Hessian kernels on the fly based on the second-order adjoint state methods (e.g., Fichtner and Trampert, 2011) and the spectral element method (SEM, e.g., Seriani and Priolo, 1994; Faccioli et al., 1996, 1997; Komatitsch and Vilotte, 1998). We first review the theory on Fréchet and Hessian kernels. We then implement both the classical wavefield storage method (the full or adaptive time iteration) and the on-the-fly approach for the full Hessian kernel calculations. The latter approach, namely QuadSEM, is conducted by simultaneously solving four wave equations based on spectral-element simulations (utilizing open-source SPECFEM community codes). The on-the-fly approach is possible because only the last-state forward fields and the absorbing boundary fields need to be stored in the forward simulation for reconstructing the forward fields during the adjoint simulation for kernel reconstruction. We then present and discuss the results of the Fréchet and Hessian kernels for 2-D synthetic models. The related codes are published in the public domain for dissemination.

Theory

Fréchet kernels

Fréchet kernels, gradients or first-order derivatives of the seismic data functional, χ, can be used to update the structural model from a chosen initial model via local optimization rather than a costly global search. When the initial model is chosen sufficiently close to the global minimum and when the source term is relatively accurate, the final model from the local optimizations may be used to explain the observed data. By perturbing the data functional as $\delta\chi$ with respect to an isotropic model $\mathbf{m}$, we have (also see Tromp et al., 2005; Fichtner and Trampert, 2011)

$$\delta\chi = \int_V K_m \delta\mathbf{m}\, d^3\mathbf{x}, \tag{1}$$

where K_m denotes the Fréchet kernels and V denotes the model volume. Here we omit the spatial and temporal dependencies of the kernels for simplicity unless stated otherwise. In principle, the generic K_m can be expressed by different components depending on the choice of model parameterization. For simplicity, we only show the case for a model parameterization given by $\mathbf{m} = (\rho,\ \alpha,\ \beta)$, where ρ denotes the density and α and β denote the P- and S-wave speed, respectively. The kernels applied to the model perturbation in eq. (1) can be further expressed as

$$K_m \delta\mathbf{m} = \begin{pmatrix} K_\rho & K_\alpha & K_\beta \end{pmatrix} \begin{pmatrix} \delta\rho \\ \delta\alpha \\ \delta\beta \end{pmatrix}, \tag{2}$$

where $\delta\mathbf{m} = (\delta\rho, \delta\alpha, \delta\beta)^{\mathrm{T}}$. As the computation of Fréchet kernels relies on the forward and the adjoint fields computed from a given model, we rewrite the Fréchet kernels as a function of the forward and adjoint fields

$$\begin{pmatrix} K_\rho \\ K_\alpha \\ K_\beta \end{pmatrix} = \begin{pmatrix} K_\rho(\mathbf{s}^\dagger, \ddot{\mathbf{s}}, \mathbf{s}) \\ K_\alpha(\mathbf{s}^\dagger, \mathbf{s}) \\ K_\beta(\mathbf{s}^\dagger, \mathbf{s}) \end{pmatrix}, \tag{3}$$

where $\mathbf{s}$ denotes the forward displacement field and $\mathbf{s}^\dagger$ denotes the adjoint field in this given model. The $\ddot{\mathbf{s}}$ is the second-order time derivative of $\mathbf{s}$, i.e., the forward acceleration field.

Hessian kernels

Components of Hessian kernels

Similar to the first-order form of the Fréchet kernels as shown in eq. (1), the second-order form or the Hessian operator can be written as (see Fichtner and Trampert, 2011)

$$H(\delta \mathbf{m}_1, \delta \mathbf{m}_2) = \int_V K_m^1 \delta \mathbf{m}_2 \; d^3\mathbf{x} = \int_V (\mathrm{H}_a + \mathrm{H}_b + \mathrm{H}_c)\, \delta \mathbf{m}_2 \; d^3\mathbf{x}, \tag{4}$$

where $K_m^1 = \mathrm{H}_a + \mathrm{H}_b + \mathrm{H}_c$ denotes the full Hessian kernels. Based upon the work of Fichtner and Trampert (2011), we rewrite each part of the product as

$$\mathrm{H}_a(\rho, \alpha, \beta) = \begin{pmatrix} K_\rho(\mathbf{s}^\dagger, \delta\ddot{\mathbf{s}}, \delta\mathbf{s}) \\ K_\alpha(\mathbf{s}^\dagger, \delta\mathbf{s}) \\ K_\beta(\mathbf{s}^\dagger, \delta\mathbf{s}) \end{pmatrix}, \tag{5}$$

$$\mathrm{H}_b(\rho, \alpha, \beta) = \begin{pmatrix} K_\rho(\delta\mathbf{s}^\dagger, \ddot{\mathbf{s}}, \mathbf{s}) \\ K_\alpha(\delta\mathbf{s}^\dagger, \mathbf{s}) \\ K_\beta(\delta\mathbf{s}^\dagger, \mathbf{s}) \end{pmatrix}, \tag{6}$$

$$\mathrm{H}_c(\rho, \alpha, \beta) = \begin{pmatrix} \rho^{-1} K_\alpha(\mathbf{s}^\dagger, \mathbf{s})\delta\alpha + \rho^{-1} K_\beta(\mathbf{s}^\dagger, \mathbf{s})\delta\beta \\ \rho^{-1} K_\alpha(\mathbf{s}^\dagger, \mathbf{s})\delta\rho + \alpha^{-1} K_\alpha(\mathbf{s}^\dagger, \mathbf{s})\delta\alpha \\ \rho^{-1} K_\beta(\mathbf{s}^\dagger, \mathbf{s})\delta\rho + \beta^{-1} K_\beta(\mathbf{s}^\dagger, \mathbf{s})\delta\beta \end{pmatrix}, \tag{7}$$

where $\delta\mathbf{s}$, $\delta\ddot{\mathbf{s}}$, and $\delta\mathbf{s}^\dagger$ denote the perturbed wavefields based upon the model perturbation $\delta\mathbf{m}_1 = \delta\mathbf{m} = (\delta\rho, \delta\alpha, \delta\beta)^{\mathrm{T}}$. For simplicity, we use $\delta\mathbf{m}$ as the model perturbation from this point on wards. Eqs. (5)-(7) show a link between the Hessian kernels (e.g., Fichtner and Trampert, 2011) and the Fréchet kernels (e.g., Tromp et al., 2005). It implies that the implementation framework for computing the Fréchet kernels can be used to compute the Hessian kernels by replacing the regular field with its associated perturbed field. The H_a can be computed with the implementation of eq. (3) by replacing the forward fields with the perturbed forward fields. H_b includes two contributions, i.e.,

$$\mathrm{H}_b = \mathrm{H}_b^{\langle m \rangle} + \mathrm{H}_b^{\langle s \rangle}, \tag{8}$$

where

$$\mathrm{H}_{\mathrm{b}}^{\langle m \rangle}(\rho, \alpha, \beta) = \begin{pmatrix} K_\rho(\delta \mathbf{s}_m^\dagger, \ddot{\mathbf{s}}, \mathbf{s}) \\ K_\alpha(\delta \mathbf{s}_m^\dagger, \mathbf{s}) \\ K_\beta(\delta \mathbf{s}_m^\dagger, \mathbf{s}) \end{pmatrix}, \tag{9}$$

$$\mathrm{H}_b^{\langle s \rangle}(\rho, \alpha, \beta) = \begin{pmatrix} K_\rho(\delta \mathbf{s}_s^\dagger, \ddot{\mathbf{s}}, \mathbf{s}) \\ K_\alpha(\delta \mathbf{s}_s^\dagger, \mathbf{s}) \\ K_\beta(\delta \mathbf{s}_s^\dagger, \mathbf{s}) \end{pmatrix}. \tag{10}$$

The former is due to the perturbation of the model, and the latter is due to the perturbation of the adjoint source which is defined as the approximate Hessian kernels in Fichtner and Trampert (2011). Both the $\mathrm{H}_b^{\langle m \rangle}$ and $\mathrm{H}_b^{\langle s \rangle}$ can be computed with the implementation of eq. (3) by replacing the adjoint fields with the associated perturbed adjoint fields. The $\mathrm{H}_b^{\langle m \rangle}$ and $\mathrm{H}_b^{\langle s \rangle}$ are computed to determine the H_b component for the full Hessian kernel calculations, i.e., considering the model perturbation and the adjoint source perturbation as well in one or two adjoint simulations. The $\mathrm{H}_b^{\langle s \rangle}$ can be computed separately to obtain the approximate Hessian kernels, i.e., only accounting for the adjoint source perturbation. The construction of H_c is straightforward based upon the Fréchet kernels K_m and the model perturbation $\delta \mathbf{m}$.

Perturbed wavefields

Eqs. (5)-(10) show that the Hessian kernels can be computed with the same implementation framework as that for the Fréchet kernel calculations, e.g., by eq. (3) using the adjoint methods. Any other packages for wave simulations and Fréchet kernel computation can be redesigned and adapted to compute the Hessian kernels with additional effort to compute the perturbed forward fields $\delta\mathbf{s}$ and $\delta\ddot{\mathbf{s}}$, and the perturbed adjoint field $\delta\mathbf{s}^\dagger$ due to the model perturbation $\delta\mathbf{m}$ and the perturbation of the adjoint source.

The H_a component accounts for the perturbation of the forward fields (Fichtner and Trampert, 2011), e.g.,

$$\delta\mathbf{s} = \lim_{v\to 0}\frac{1}{v}[\mathbf{s}(\mathbf{m}_r + v\delta\mathbf{m};\mathbf{x},t) - \mathbf{s}(\mathbf{m}_r;\mathbf{x},t)], \tag{11}$$

where $\mathbf{m}_r$ denotes the reference model, and $r = 0, 1, 2, ..., N$ represents the iteration number. The initial model is set to $\mathbf{m}_0$. The same consideration applies to the perturbed acceleration field $\delta\ddot{\mathbf{s}}$ for density kernel calculations.

The H_b component consists of two contributions: $\mathrm{H}_b^{\langle s\rangle}$ and $\mathrm{H}_b^{\langle m\rangle}$, where the former $\mathrm{H}_b^{\langle s\rangle}$ relies on the approximate perturbed adjoint field

$$\delta\mathbf{s}_s^\dagger = \mathbf{s}_s^\dagger(\mathbf{m}_r;\mathbf{x},T-t) - \mathbf{s}^\dagger(\mathbf{m}_r;\mathbf{x},T-t). \tag{12}$$

In the equation, the $\mathbf{s}_s^\dagger(\mathbf{m}_r;\mathbf{x},T-t)$ field is generated by the adjoint source $\mathbf{f}^\dagger(\mathbf{m}_r + v\delta\mathbf{m};\mathbf{x},T-t)$, and $\mathbf{s}^\dagger(\mathbf{m}_r;\mathbf{x},T-t)$ is generated by the adjoint source $\mathbf{f}^\dagger(\mathbf{m}_r;\mathbf{x},T-t)$. The adjoint sources could be the traveltime adjoint source, the waveform adjoint source, or any other adjoint source based on the choices of seismic data functional χ as discussed in Tromp et al. (2005). The $\mathrm{H}_b^{\langle s\rangle}$ and $\mathrm{H}_b^{\langle m\rangle}$ need to be computed for the H_b determination in order to compute the full Hessian kernels: $\mathrm{H}_a + \mathrm{H}_b^{\langle s\rangle} + \mathrm{H}_b^{\langle m\rangle} + \mathrm{H}_c$. The perturbed adjoint field for the $\mathrm{H}_b^{\langle m\rangle}$ may be given by

$$\delta\mathbf{s}_m^\dagger = \lim_{v\to 0}\frac{1}{v}[\mathbf{s}_m^\dagger(\mathbf{m}_r + v\delta\mathbf{m};\mathbf{x},T-t) - \mathbf{s}^\dagger(\mathbf{m}_r;\mathbf{x},T-t)], \tag{13}$$

where the two adjoint fields $\mathbf{s}_m^\dagger(\mathbf{m}_r + v\delta\mathbf{m},\mathbf{x},T-t)$ and $\mathbf{s}^\dagger(\mathbf{m}_r,\mathbf{x},T-t)$ are generated through the perturbed and unperturbed model from the same adjoint source $\mathbf{f}^\dagger(\mathbf{m}_r;\mathbf{x},T-t)$. Thereafter, the total perturbed adjoint field is

$$\delta\mathbf{s}^\dagger = \delta\mathbf{s}_s^\dagger + \delta\mathbf{s}_m^\dagger. \tag{14}$$

The $\delta\mathbf{s}_s^\dagger$ and $\delta\mathbf{s}_m^\dagger$ may be computed in one adjoint simulation with the perturbation accounting for both the model and the adjoint source simultaneously, or computed separately in two adjoint simulations considering the two perturbations individually.

The computation of H_c relies on the Fréchet kernels and model perturbation, see eq. (7). It has also been shown that H_c is non-zero when the model is parametrized as ρ, α, and β but zero when the model is given in terms of density and elastic moduli (see Fichtner and Trampert, 2011).

Implementation

The computation of the full Hessian kernels relies on the regular and perturbed fields as mentioned above. Its implementation is straightforward based on the wavefield storage method (WSM), which saves the forward fields at full or adaptive time steps and reads each saved time step of the forward fields back into temporary memory during the adjoint simulation for kernel constructions (See Appendix B for the WSM method based on the full or adaptive time integration scheme). Or only two adjoint wave equations need to be solved simultaneously during the adjoint simulation. However, in light of the large storage requirement by the WSM, here we focus on showing how the full Hessian kernels are computed by the on-the-fly approach implemented in QuadSEM, which involves simultaneously solving four wave equations during the adjoint simulation. For the following examples, we only consider cases with purely elastic models.

Forward simulation

Figure 1 shows the comparison between the classical SEM and the QuadSEM during the forward simulations. In comparison to one model used by the classical SEM, QuadSEM carries wavefield simulations for two models simultaneously, e.g., $\mathbf{m}_1$ and $\mathbf{m}_2$, where $\mathbf{m}_2 = \mathbf{m}_1 + v\delta\mathbf{m} = \mathbf{m}_1 + \Delta\mathbf{m}$. In this case, the wavefields, including displacement $\mathbf{s}$, velocity $\mathbf{v}$, acceleration $\ddot{\mathbf{s}}$, and the boundary contribution $\mathbf{b}$ (a general symbol used to refer to velocity or traction fields on the coupled boundary with an external model) are computed for the two models at each time step. The displacement seismograms $\mathbf{s}(\mathbf{x}_r, t)$ are computed by spatial interpolation of fields at the receiver $\mathbf{x}_r$ at each time step. The grid-point locations and mesh topology database files are shared by the two models used simultaneously in the forward simulation with QuadSEM, and only arrays/files related to model material properties such as ρ, α, and β need to be defined separately for the two models. The CPU and memory requirements for QuadSEM are about twice the cost of the classical SEM simulation. The forward simulations for either the classical SEM or the QuadSEM are designed to provide the absorbing boundary fields, the last state of the forward field, and the seismograms at receivers, for subsequent simulations.

Simultaneous backward and adjoint simulations

The strategy of using simultaneous backward and adjoint simulations was adopted for instance in the SPECFEM2D (`https://geodynamics.org/cig/software/specfem2d/`) and the SPECFEM3D (`https://geodynamics.org/cig/software/specfem3d/`) packages to compute the Fréchet kernels on the fly. A workflow for computing the Fréchet kernels using the classical SEM method is shown in Figure A1. For purely elastic models, the backward simulation is a time-reversed reconstruction of the forward field using the last state of the forward field as a starting point. The absorbing boundary contributions saved in the forward simulation are re-injected into the backward simulation when the forward field is reconstructed backward in time. The simulations for the backward reconstruction and the adjoint wavefield are performed simultaneously so that the corresponding time slices of the forward and adjoint wavefields can be accessed both in memory to construct the Fréchet kernels. A similar idea is adopted in the QuadSEM, and as shown in Figure 2 and Figure A2, where both the regular and perturbed forward wavefields, as well as the regular and perturbed adjoint wavefields for the two models are simultaneously computed every time step, so that the Fréchet and full Hessian kernels can be constructed on the fly as wavefield products are computed and integrated over time steps. As indicated by Figure 2, the calculations of Fréchet kernels (by Arrows 1 and 2) and the full Hessian kernels (by Arrows 1, 2, 4, and 5, or Arrows 1, 2, 3, 4, and 5 depending on the adjoint source of 5) are simultaneously performed on the fly since all required wavefields are computed for each time step. Alternatively, the approximate Hessian kernels can be computed by the solutions indicated by Arrows 1, 2, 3, and 4. The QuadSEM degenerates to classical SEM when solutions indicated by 4 and 5 or 3 are not used.

Although the forward and adjoint wavefields for $\mathbf{m}_1$ and $\mathbf{m}_2$ are combined to compute kernels in the QuadSEM, the same mesher database is used except for those variables or matrices that define $\mathbf{m}_1$ and $\mathbf{m}_2$. The memory cost is small since only one time step of the various fields and the integrated kernels are kept in memory. The Fréchet kernels need 3 (1 in the forward and 2 in the adjoint) simulations, while the QuadSEM needs 6 (2 in the forward and 4 in the adjoint) simulations for the simultaneous computation of both Fréchet and full Hessian kernels. A simultaneous computation of both Fréchet and approximate Hessian kernels also requires 6 (2 in the forward and 4 in the adjoint) simulations. Therefore, roughly QuadSEM doubles the memory and CPU time required for the simultaneous computation for both Fréchet and full Hessian kernels compared to the requirement for Fréchet kernels alone.

Numerical examples

Models

To test the numerical implementation of QuadSEM, three models are considered in this study. First, a homogeneous 2D model (*Model 1*) that is 800 km in the horizontal direction and 360 km in the vertical direction and with density ρ=2900 kg/m^3, P-wave speed α=8000 m/s, and S-wave speed β=4800 m/s, is used as a starting background model to generate initial wavefields and waveforms. The second and the third models are perturbed versions of the homogeneous model. The second model (*Model 2*) has an additional +10% perturbation in α and β over a 10 km $\times$ 10 km area centered at the horizontal location of 335 km and depth of 135 km. The third model (*Model 3*) includes three anomalies that are 8 km $\times$ 10 km, centered at a depth of 115 km at horizontal locations of 120 km, 180 km, and 240 km, respectively, with +10% perturbations in α and β. No density perturbation is considered for the second and third model.

For all three models, we use 400 elements in the horizontal direction and 360 elements in the depth direction with 5 $\times$ 5 Gauss-Lobatto-Legendre (GLL) points for each element, which leads to about 500 m and 250 m grid point spacing in the horizontal and vertical direction, respectively. The mesher and related databases for these models are built by the internal mesher tool of the SPECFEM2D package with slightly changing for the two models. Figure 3 shows the locations of these model perturbations (Part II, blue boxes in the last column) and the source-receiver geometry, together with the kernel images discussed in the section of Single source-receiver combination. These models are chosen to illustrate the differences in the calculation of Hessian kernels between the single source-receiver pair and single-source multiple-receiver case.

Single source-receiver combination

We first examine the kernel calculation for a single source-receiver combination based on *Model 1* and *Model 2*. We place a point source at (x, z)=(100 km, -260 km), and a standard Ricker wavelet with the dominant frequency of 0.5 Hz is used. A single receiver is placed on the top surface of the model at (x, z)=(600 km, 0 km). The simulations use $dt = 0.01$ s and run for a total of $7,000$ time steps. Adjoint sources for cross-correlation traveltime (Tromp et al., 2005) are first calculated based on the first P-wave arrival recorded by the two-component seismograms. The P-wave is primarily sensitive to P-wave speed, α, so only α kernels are shown in the examples below. As discussed in the section of Simultaneous backward and adjoint simulations, QuadSEM computes the Fréchet kernels using the same solutions of the forward and adjoint equations as the classical SEM (see Figure 2).

First, we examine the Fréchet kernel for *Model 2* (i.e., $\mathbf{m}_1 + \mathrm{v}\delta\mathbf{m}$), and the full Hessian kernel for the model perturbation from *Model 1* (i.e., $\mathbf{m}_1$) to *Model 2*. The approximate Hessian kernels are computed for *Model 1* but using the adjoint source computed from the seismograms/measurements of *Model 2*. We show the Fréchet kernel, the approximate Hessian kernel, and the full Hessian kernel for P wavespeed α in the first row of Figure 3 (Part I) with a zoomed-in version around the perturbations given in the first row of Figure 3 (Part II). The H_c is restricted to the perturbation indicated by the black box (see Figure 3c and its expression of eq. (7)). Note that the black box here is the H_c Hessian kernel with a negative value of 10^{-9}, not the model perturbation although they are located in the same position. The $\mathrm{H}_b^{\langle s \rangle}$ kernel is mostly invisible in Figure 3c except those around the black box due to its relatively small amplitude. The H_a and $\mathrm{H}_b^{\langle m \rangle}$ are separated by the black box. Substantial differences are observed between the approximate and full Hessian kernels.

It takes the QuadSEM about a total of 37.57 mins with a maximum memory usage of $\sim$2.48 GB to simultaneously compute both the Fréchet and full Hessian kernels using 4 cores on a standard laptop (with 2.3 GHz Dual-Core Intel Core i5 processor and 8GB 2133 MHz LPDDR3 memory). The computation of Fréchet kernels alone by the classical adjoint method takes about 18.76 mins with a maximum memory usage of 1.5 GB using the same computer. Therefore, in this case, a simultaneous computation of both the Fréchet and full Hessian kernels via QuadSEM roughly takes about 2 times the CPU time and $\sim$ 1.53 times the memory when compared to the computation of Fréchet kernels alone. The memory cost is slightly less than 2 because we use the same variables and matrices and these are independent of the wavefields generated from the two input models. The storage required for the QuadSEM is small due to the on-the-fly nature of the calculations, which takes 751 MB disk space to store the absorbing boundary fields, the last-state forward fields as well as the seismograms. The wavefield storage method (WSM) stores the fields at all time steps and requires about 513 GB disk space to store these fields.

One source and three receivers

We also show an example with one source and three receivers for the calculation of Hessian kernels, where *Model 1* is used as the background model (i.e., $\mathbf{m}_1$) and Hessian kernels are computed with respect to the perturbation in *Model 3* (i.e., $\mathbf{m}_1 + v\delta\mathbf{m}$, and the $v\delta\mathbf{m}$ here indicates a new perturbation for the three scatterers). The source is placed at $(x, z) = (150\ \text{km}, -260\ \text{km})$ with the same source time function as in the section of Single source-receiver combination. Three receivers are placed on the top surface of the model located at horizontal locations of 100 km, 200 km, and 300 km, respectively. The total number of time steps and time intervals are the same as the example in the section of Single source-receiver combination.

The second row of Figure 3 (Part I) shows the Fréchet kernel, the approximate Hessian kernel, and the full Hessian kernel computed for α, again for a traveltime adjoint source measured from the first-arrival of the P-wave. A zoomed-in version around the perturbations is given in the second row of Figure 3 (Part II). More detailed descriptions about the Fréchet and Hessian kernels are given in the figure caption. The computational cost for this example is almost the same as for that in the section of Single source-receiver combination since the forward and adjoint simulation time is almost independent of the number of receivers. There is one additional step in the window picking and computation of the adjoint source, which the cost is negligible compared to the field calculations. A few selected time steps of the regular wavefields and their perturbations computed by the QuadSEM are shown in Figure 4 and Figure 5. For the on-the-fly implementation, QuadSEM, the key output files used in the forward simulation and in the simultaneous backward and adjoint simulation are presented in Figure 6.

Kernel comparisions

For both the one- and multi-receiver cases shown in Figure 3, we found substantial differences between the approximate Hessian kernels and the full Hessian kernels, in agreement with previous work (Fichtner and Trampert, 2011). Most notably, the amplitudes of the full Hessian kernels can be up to 100% stronger than those of the approximate Hessian kernels within the first and second Fresnel zones. These areas are covered by H_a, $\mathrm{H}_b^{\langle m \rangle}$, and H_c in the full Hessian kernels and usually omitted in the calculation of the approximate Hessian kernels. The greater positive values of the Hessian kernels in the vicinity of the perturbation suggest that the inversion using the full Hessian kernels will result in better illumination in the region of the model perturbation. In comparison, when only the approximate Hessian kernel is used, the model updates tend to be distributed along the entire kernel.

In the multi-receiver case, we observe similarly higher amplitude in the Hessian kernels near the three model perturbations (Figure 3f) (Part I and II); whereas, for the approximate Hessian kernels, the sensitivity has high amplitudes around the middle anomaly only. This again suggests that using the full Hessian kernels in the inversion will focus model perturbations closer to the actual anomalies and hence provide better resolution for small anomalies within the model.

Discussions

The Hessian kernels are typically used with the Fréchet kernels for computing the model update or search direction based upon truncated Newton optimization (e.g., Nash, 1985; Grippo et al., 1989; Nash, 2000). This may potentially generate more accurate results and quicker convergence compared with L-BFGS based optimization for multi-parameter full-waveform inversion (FWI) (e.g., Métivier et al., 2013, 2014; Pan et al., 2017; Yang et al., 2018; Matharu and Sacchi, 2019).

The QuadSEM implementation can compute both the Fréchet and full Hessian kernels simultaneously, requiring only double the computational cost of Fréchet kernels alone. To further reduce the computational costs, the source encoding techniques (e.g., Tromp and Bachmann, 2019) or a reverse propagation of a superposition of forward and residual wavefields (e.g., Robertsson et al., 2021) may be considered as well.

An important question remains as to whether the additional costs of the simultaneous computation of the Fréchet and full Hessian kernels at twice the computational cost can be offset by the more rapid convergence of the non-linear inversion. As high-performance computing becomes more accessible and efficient, this may become much less of a concern.

In addition to the model expressed in terms of density, P-wave and S-wave velocities, the approximate Hessian kernels and the full Hessian kernels can be expressed by different model components with similar derivations. In this study, we only show kernel examples for elastic models to demonstrate the on-the-fly approach. For anelastic models, a checkpointing method (e.g., Komatitsch et al., 2016) may be needed, and we are in the process of developing such method to compute the anelastic full Hessian kernels by applying similar idea as QuadSEM but storing the two forward fields at some checkpoints in order to reconstruct them accurately during the adjoint simulation. The idea of QuadSEM is not limited to the SEM, and other numerical solvers that simulate seismic wave propagation can be adapted to compute the Frécket kernels as well the full Hessian kernels.

Conclusions

We present QuadSEM, a package for on-the-fly full Hessian kernel calculation through simultaneously solving four wave equations, which is designed to simultaneously compute the Fréchet and full Hessian kernels on the fly with only about double of the computational cost for the calculation of Fréchet kernels alone. In the QuadSEM, the WSM is also implemented as it is complementary to the on-the-fly approach. While the examples presented in this paper are rather specific to three elastic models, the underlying idea is very general. The QuadSEM trades off the computational cost with storage and I/O, and improves the accuracy of full Hessian kernel calculations by combining the exact forward and adjoint wavefields on the fly in temporary memory. This makes it possible to use the accurate full Hessian information for multi-parameter FWI based upon the spectral-element and adjoint methods. It potentially provides a step forward in improving FWI to better image and understand Earth structure, particularly in regions characterized by weak and/or small scale heterogeneities.

Data and Resources

No field data were used in this work. Models and the QuadSEM codes can be freely downloaded via `https://github.com/yujiangxie/QuadSEM` or requested from the authors.

Acknowledgments

We acknowledge funding from the Natural Environment Research Council (NE/M003507/1) and the European Research Council (GA 638665) for supporting this work. Q. Liu is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant No. 487237. Some funding is also provided by the Wave Inversion Technology (WIT) consortium. We thank all the developers of the SPECFEM2D/3D packages for their continued community support. We thank Fichtner Andreas for his helpful comments which significantly improved our methods after revision. We would also like to thank the Editor-in-Chief, Allison Bent, the Associate Editor and one anonymous referee for their valuable suggestions that greatly helped to improve the paper. Y. Xie would like to thank Daniel Peter for continued discussions on the Hessian kernels.

Author contributions: methodology and software, Y. Xie; validation, Y. Xie, C. Rychert, N. Harmon, Q. Liu and D. Gajewski; writing-original draft preparation, Y. Xie, C. Rychert, N. Harmon; writing-review and editing, Q. Liu and D. Gajewski; supervision, C. Rychert, N. Harmon, Q. Liu and D. Gajewski. All authors have read and agreed to the published version of the manuscript and the authors declare no conflict of interest.

References

Anderson, J. E., L. Tan, and D. Wang, 2012. Time-reversal checkpointing methods for RTM and FWI, Geophysics, **77**, S93–S103.

Bamberger, A., G. Chavent, C. Hemons, and P. Lailly, 1982. Inversion of normal incidence seismograms, Geophysics, **47**, 757–770.

Boehm, C., M. Hanzich, J. de la Puente, and A. Fichtner, 2016. Wavefield compression for adjoint methods in full-waveform inversion, Geophysics, **81(6)**, R385–R397.

Boehm, C., and M. Ulbrich, 2015. A semismooth Newton-CG method for constrained parameter identification in seismic tomography, SIAM Journal on Scientific Computing, **37**, S334–S364.

Bozdağ, E., D. Peter, M. Lefebvre, D. Komatitsch, J. Tromp, J. Hill, N. Podhorszki, and D. Pugmire, 2016. Global adjoint tomography: first-generation model, Geophys. J. Int., **207**, 1739–1766.

Charpentier, I., 2001. Checkpointing schemes for adjoint codes: application to the meteorological model Meso-NH, SIAM J. Sci. Comp., **22**, 2135–2151.

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

Chen, P., L. Zhao, and T. H. Jordan, 2007b. Full 3D Tomography for the crustal structure of the Los Angeles region, Bulletin of the Seismological Society of America, **97**, 1094–1120.

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

Epanomeritakis, I., V. Akçelik, O. Ghattas, and J. Bielak, 2008. A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion, Inverse Problems, **24**, 034015.

Faccioli, E., F. Maggio, R. Paolucci, and A. Quarteroni, 1997. 2D and 3D elastic wave propagation by a pseudospectral domain decomposition method, J. Seismology, **1**, 237–

251.

Faccioli, E., F. Maggio, A. Quarteroni, and A. Tagliani, 1996. Spectral-domain decomposition methods for the solution of acoustic and elastic wave equations, Geophysics, **61**, 1160–1174.

Fichtner, A., H. Bunge, and H. Igel, 2006a. The adjoint method in seismology — I. Theory, Physics of the Earth and Planetary Interiors, **157**, 86–104.

Fichtner, A., B. L. N. Kennett, H. Igel, and H. P. Bunge, 2009. Full waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, Geophys. J. Int., **179**, 1703–1725.

Fichtner, A., and J. Trampert, 2011. Hessian kernels of seismic data functionals based upon adjoint techniques, Geophys. J. Int., **185**, 775–798.

French, S. W., and B. Romanowicz, 2015. Broad plumes rooted at the base of the Earth's mantle beneath major hotspots, Nature, **525**, 95–99.

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

Götschel, S., and M. Weiser, 2015. Lossy compression for PDE-constrained optimization: Adaptive error control, Computational Optimization and Applications, **62**, 131–155.

Griewank, A., and A. Walther, 2000. Algorithm 799: revolve: an implementation of checkpointing for the reverse or adjoint mode of computational differentiation, ACM Trans. Math. Softw., **26**, 19–45.

Grippo, L., F. Lampariello, and S. Lucidi, 1989. A Truncated Newton Method with Nonmonotone Line Search for Unconstrained Optimization, J. Optim. Theory Appl, **60**, 401–419.

Hanzich, M., F. Rubio, G. Aguilar, and N. Gutierrez, 2013. Efficient lossy compression for seismic processing, 75th EAGE Conference and Exhibition, EAGE, Expanded Abstracts, Th–16–08.

Igel, H., H. Djikpesse, and A. Tarantola, 1996. Waveform inversion of marine reflection seismograms for P impedance and Poisson's ratio, Geophys. J. Int., **124**, 363–371.

Komatitsch, D., and J. P. Vilotte, 1998. The spectral element method: an effective tool to simulate the seismic response of 2D and 3D geological structures, Bull. Seism. Soc. Am., **88**, 368–392.

Komatitsch, D., Z. Xie, E. Bozdağ, E. S. d. Andrade, D. Peter, Q. Liu, and J. Tromp, 2016. Anelastic sensitivity kernels with parsimonious storage for adjoint tomography and full waveform inversion, Geophys. J. Int., **206**, 1467–1478.

Lailly, P., 1983. The seismic inverse problem as a sequence of before stack migrations: Conference on Inverse Scattering, Theory and Application, Society for Industrial and Applied Mathematics, Expanded Abstracts, 206–220.

Lindstrom, P., P. Chen, and E. Lee, 2016. Reducing disk storage of full-3D seismic waveform tomography (F3DT) through lossy online compression, Comput. Geosci., **93**, 45–54.

Lions, J., 1968. Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles, Dunod Gauthier-Villars.

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

Liu, Q., and J. Tromp, 2006. Finite-Frequency Kernels Based on Adjoint Methods, Bulletin of the Seismological Society of America, **96**, 2383–2397.

Luo, Y., R. Modrak, and J. Tromp, 2014. Strategies in Adjoint Tomography. In: Freeden W., Nashed M., Sonar T. (eds), Handbook of Geomathematics, Springer.

Matharu, G., and M. Sacchi, 2019. A subsampled truncated-Newton method for multiparameter full-waveform inversion, Geophysics, **84(3)**, R333–R340.

Métivier, L., F. Bretaudeau, R. Brossier, S. Operto, and J. Virieux, 2014. Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures, Geophysical Prospecting, **62**, 1353–1375.

Métivier, L., R. Brossier, J. Virieux, and S. Operto, 2013. Full Waveform Inversion and the Truncated Newton Method, SIAM Journal on Scientific Computing, Society for Industrial and Applied Mathematics, **35**, B401–B437.

Nash, S. G., 1985. Preconditioning of truncated-Newton methods, SIAM J. Sci. Stat. Comput., **6**, 599–616.

Nash, S. G., 2000. A survey of truncated-Newton methods, Journal of Computational and Applied Mathematics, **124**, 45–59.

Pan, W., K. A. Innanen, and W. Liao, 2017. Hessian-free Gauss-Newton full-waveform inversion via l-BFGS preconditioned conjugate-gradient algorithm, Geophysics, **82(2)**, R49–R64.

Plessix, R.-E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, Geophys. J. Int., **167**, 495–503.

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

Robertsson, J. O., F. Andersson, and R.-É. Plessix, 2021. Efficient snapshot-free reverse time migration and computation of multiparameter gradients in full waveform inversion, Geophysics, **0**, 1–79.

Seriani, G., and E. Priolo, 1994. Spectral element method for acoustic wave simulation in heterogeneous media, Finite Elements in Analysis and Design, **16**, 337–348.

Symes, W. W., 2007. Reverse time migration with optimal checkpointing, Geophysics, **72**, SM213–SM221.

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

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

Tarantola, A., 1988. Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation, Pure appl. Geophys., **128**, 365–399.

Tromp, J., 2020. Seismic wavefield imaging of Earth's interior across scales, Nature Reviews Earth & Environment, **1**, 40–53.

Tromp, J., and E. Bachmann, 2019. Source encoding for adjoint tomography, Geophys. J. Int., **218**, 2019–2044.

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

Unat, D., T. Hromadka, and S. Baden, 2009. An adaptive sub-sampling method for in-memory compression of scientific data, Data Compression Conference, IEEE, DCC'09, 262–271.

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

Walther, A., and A. Griewank, 2004. Advantages of Binomial Checkpointing for Memory-reduced Adjoint Calculations. In: Feistauer M., DolejšíV., Knobloch P., Najzar K. (eds) Numerical Mathematics and Advanced Applications, Springer.

Weiser, M., and S. Götschel, 2012. State trajectory compression for optimal control with parabolic PDEs, SIAM J. Sci. Comput., **34**, A161–A184.

Yang, P., R. Brossier, L. Métivier, J. Virieux, and W. Zhou, 2018. A Time-Domain Preconditioned Truncated Newton Approach to Visco-acoustic Multiparameter Full Waveform Inversion, SIAM Journal on Scientific Computing, **40**, B1101–B1130.

Zhu, H., E. Bozdağ, D. Peter, and J. Tromp, 2012. Structure of the European upper mantle revealed by adjoint tomography, Nat. Geosci., **5**, 493–498.

Full mailing address for each author

Dr. Yujiang Xie

Ocean and Earth Science, National Oceanography Centre Southampton (NOCS)
University of Southampton Waterfront Campus
European Way, Southampton SO14 3ZH, UK

Prof. Catherine Rychert

Ocean and Earth Science, National Oceanography Centre Southampton (NOCS)
University of Southampton Waterfront Campus
European Way, Southampton SO14 3ZH, UK

Prof. Nicholas Harmon

Ocean and Earth Science, National Oceanography Centre Southampton (NOCS)
University of Southampton Waterfront Campus
European Way, Southampton SO14 3ZH, UK

Prof. Qinya Liu

Department of Physics & Department of Earth Sciences, University of Toronto
22 Ursula Franklin Street, Toronto, ON, M5S 3B1, Canada

Prof. Dirk Gajewski

Department of Earth Sciences, Universität Hamburg
Room: 1432, Geomatikum, Bundesstrasse 55
20146 Hamburg, Germany

List of Figure Captions

Figure 1. Sketch illustrating the workflow of forward simulation for classical SEM vs. QuadSEM.

Figure 2. Sketch illustrating the workflows for the simultaneous backward and adjoint simulations for classical SEM vs. QuadSEM.

Figure 3. Fréchet and Hessian kernels computed for the investigated models.

Figure 4. Four selected time steps of the five wavefields computed by the QuadSEM.

Figure 5. A few time steps of selected perturbed fields computed on the fly using the first-order finite-difference approximation.

Figure 6. Key files output from the forward simulation and the simultaneous backward and adjoint simulation in the QuadSEM.

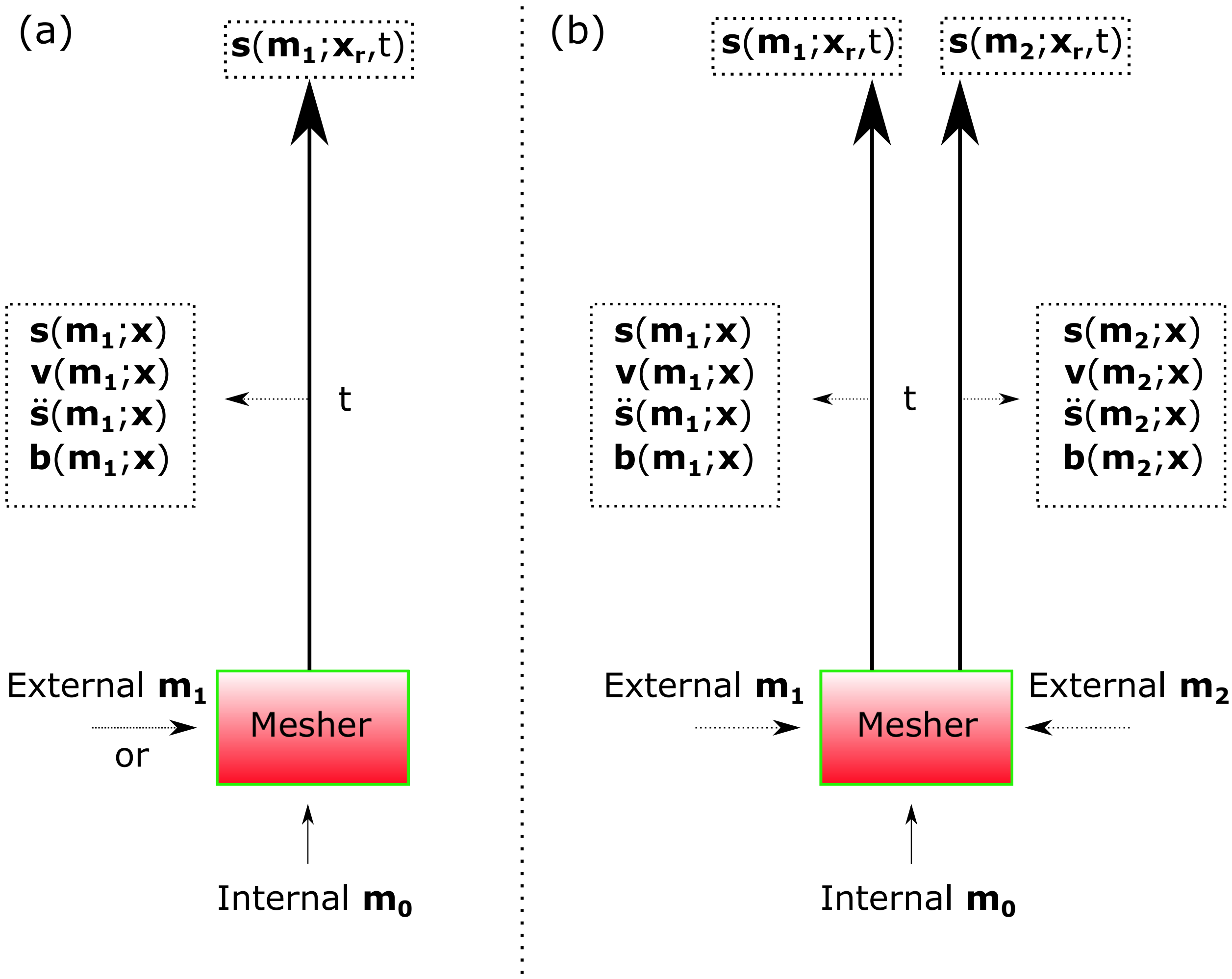

Figure 1: Sketch illustrating the workflow of forward simulation for classical SEM vs. QuadSEM. (a) In classical SEM forward simulation, a single model is used and it is set either by the internal mesher (e.g., $\mathbf{m}_0$) or importing from external file ($\mathbf{m}_1$) after the mesher is set up. (b) In the QuadSEM forward simulation, two models ($\mathbf{m}_1$ and $\mathbf{m}_2$) are imported into the internal mesher, where $\mathbf{m}_2 = \mathbf{m}_1 + v\delta\mathbf{m} = \mathbf{m}_1 + \Delta\mathbf{m}$, and $\mathbf{m}_0$ will be omitted when external models are loaded.

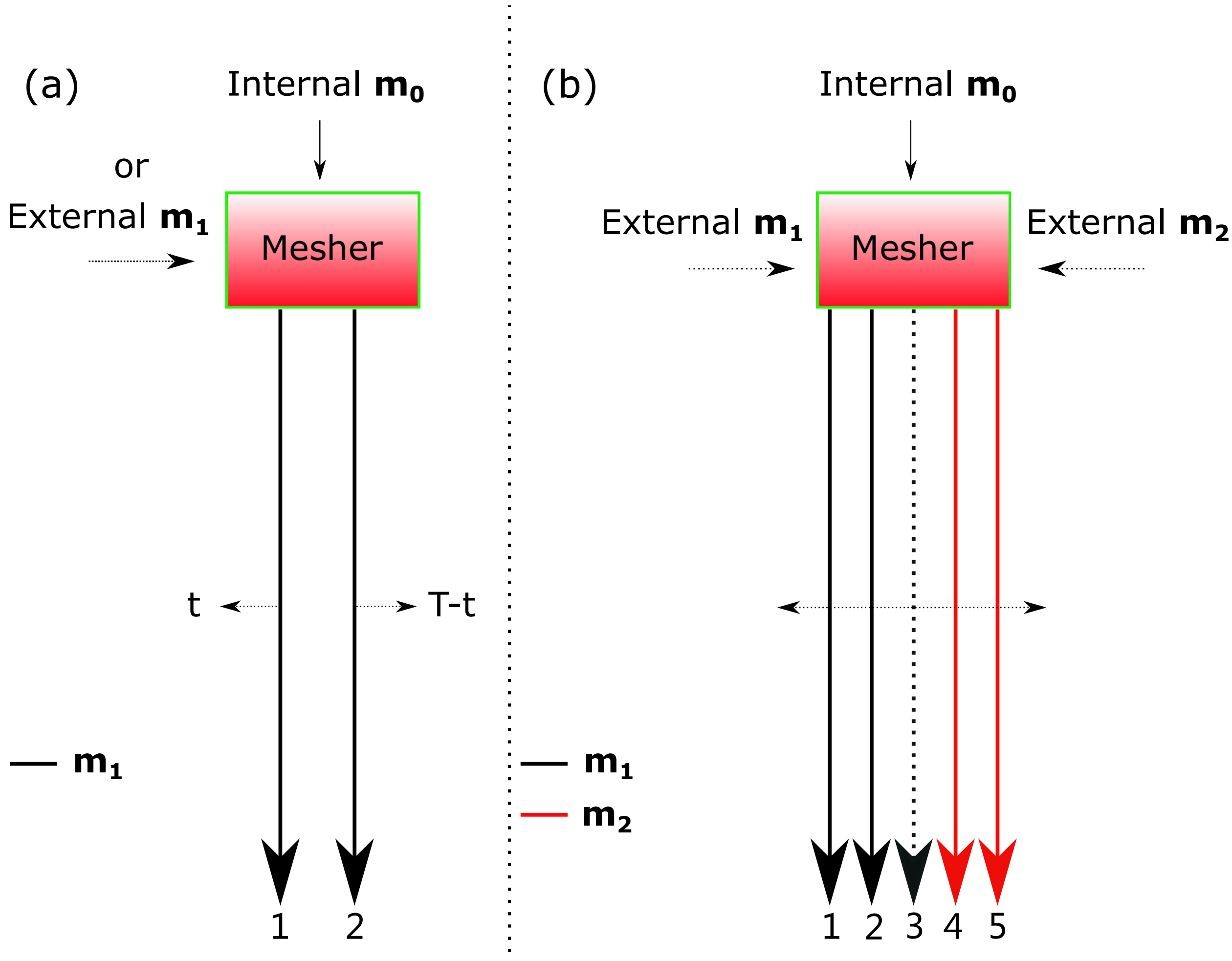

Figure 2: Sketch illustrating the workflows for the simultaneous backward and adjoint simulations for classical SEM vs. QuadSEM. (a) In the simultaneous backward and adjoint simulation of the classical SEM. Each arrow represents the solution for one wave equation with Arrow 1 indicating the backward simulation (i.e. the reconstruction of the forward field) and Arrow 2 indicating the adjoint simulation. (b) In the simultaneous backward and adjoint simulation of the QuadSEM, Arrows 1, 2, and 3 indicate the solutions of the wave equations for model $\mathbf{m}_1$, and Arrows 1 and 2 perform the same as in (a), and Arrow 3 performs the same as Arrow 2 but its adjoint source is computed by using the measurements of $\mathbf{m}_2$ to account for the perturbation of the adjoint source. The red Arrows 4 and 5 indicate the computation of the backward and adjoint fields for model $\mathbf{m}_2$.

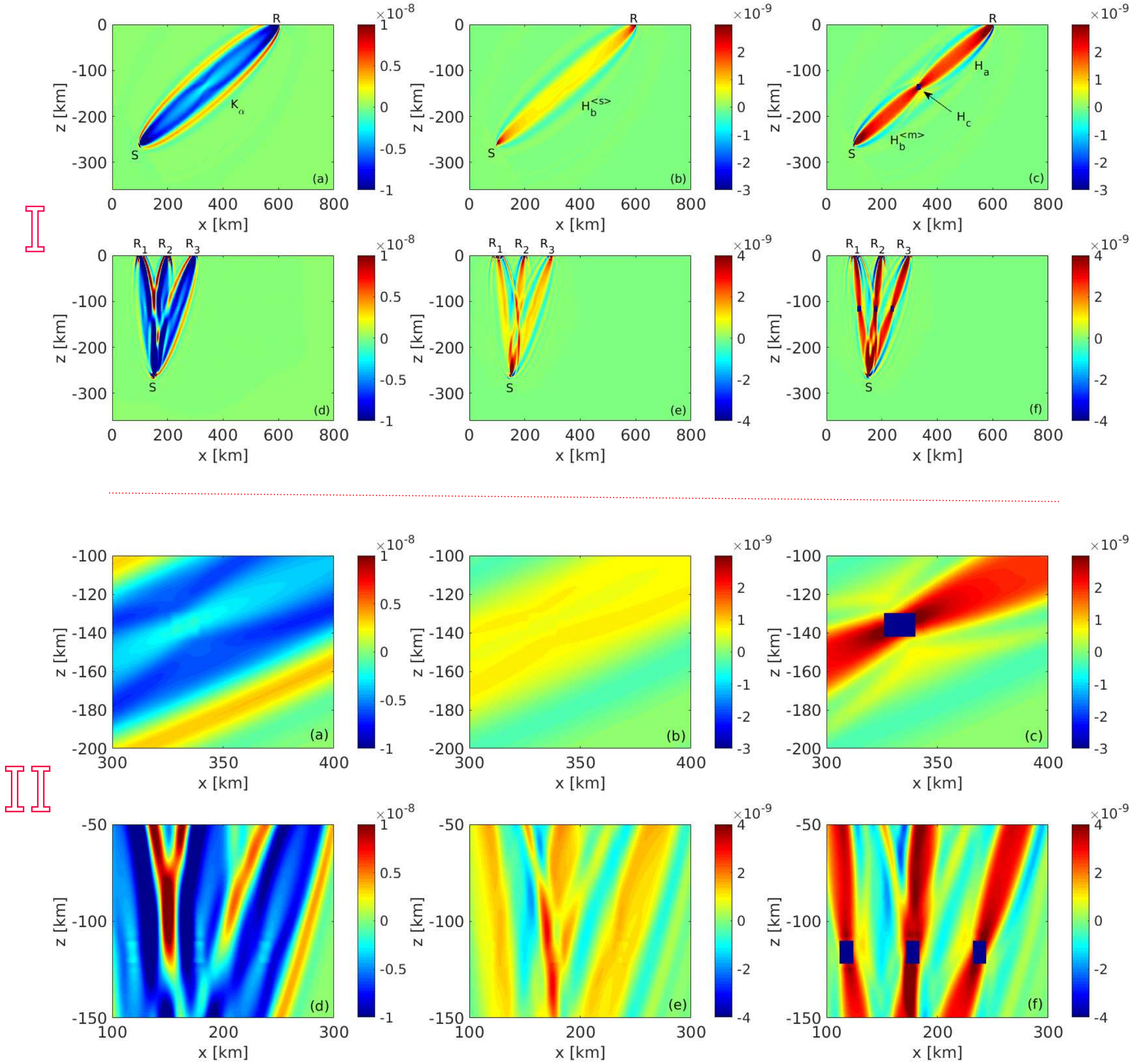

Figure 3: Part I: Fréchet and Hessian kernels computed for *Model 2* (top row) and *Model 3* (bottom row) as discussed in section of Numerical examples. In the top row we show (a) the Fréchet kernel K_α, (b) the approximate Hessian kernel $\mathrm{H}_b^{\langle s\rangle}$, and (c) the full Hessian kernel for α for the single source single station case with a single scattering object. Similarly, Panels (d), (e), and (f) in the bottom row show the various kernels for the case of a single source and three stations with three scattering objects. The kernel unit for all sub-figures is $[s\ m^{-2}]$. A zoomed view of the perturbations within Part I is shown in Part II.

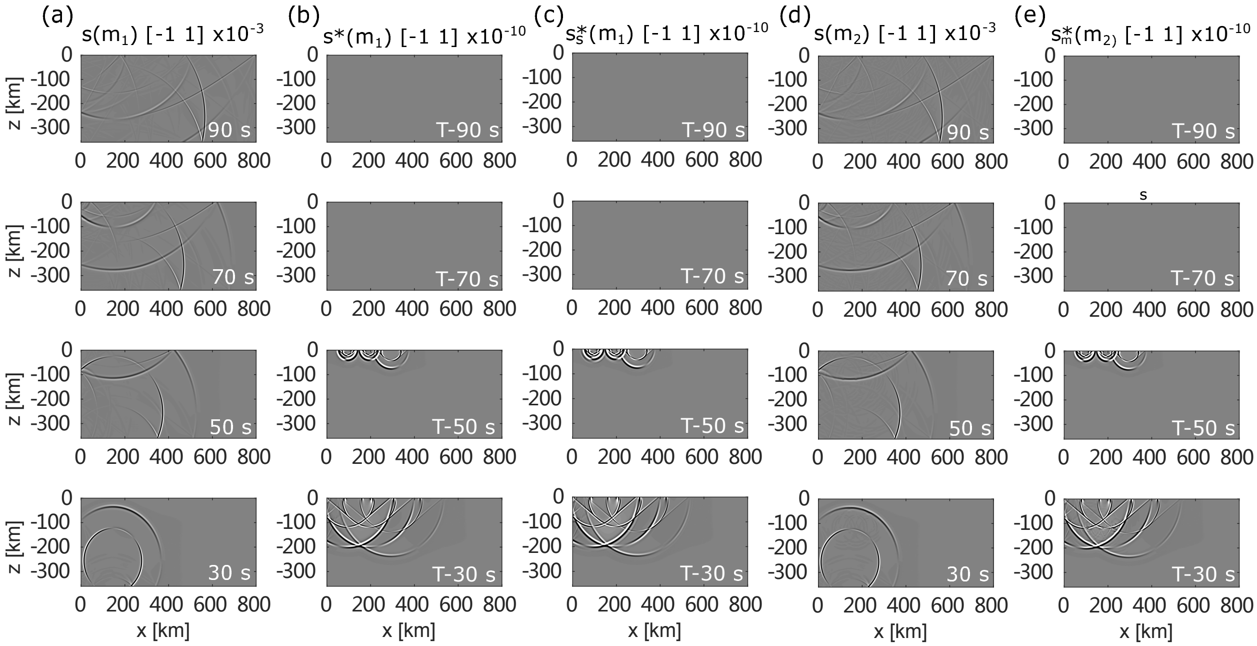

Figure 4: Four selected time steps of the five wavefields computed by the QuadSEM using the on-the-fly approach. (a) The forward fields recorded at times 30 s, 50 s, 70 s, and 90 s for model $\mathbf{m}_1$. (b) The adjoint fields for the same model but recorded at reversed times of T-90 s, T-70 s, T-50 s, and T-30 s, where T = 100 s in this test. (c) The adjoint fields generated by the adjoint source computed from the measurements for $\mathbf{m}_2$. (d) and (e) show the similar simulation as (a) and (b) but for model $\mathbf{m}_2$, instead of $\mathbf{m}_1$. (b) and (e) look similar due to the use of the same adjoint source, but they are different after the adjoint fields travel through the scatterers.

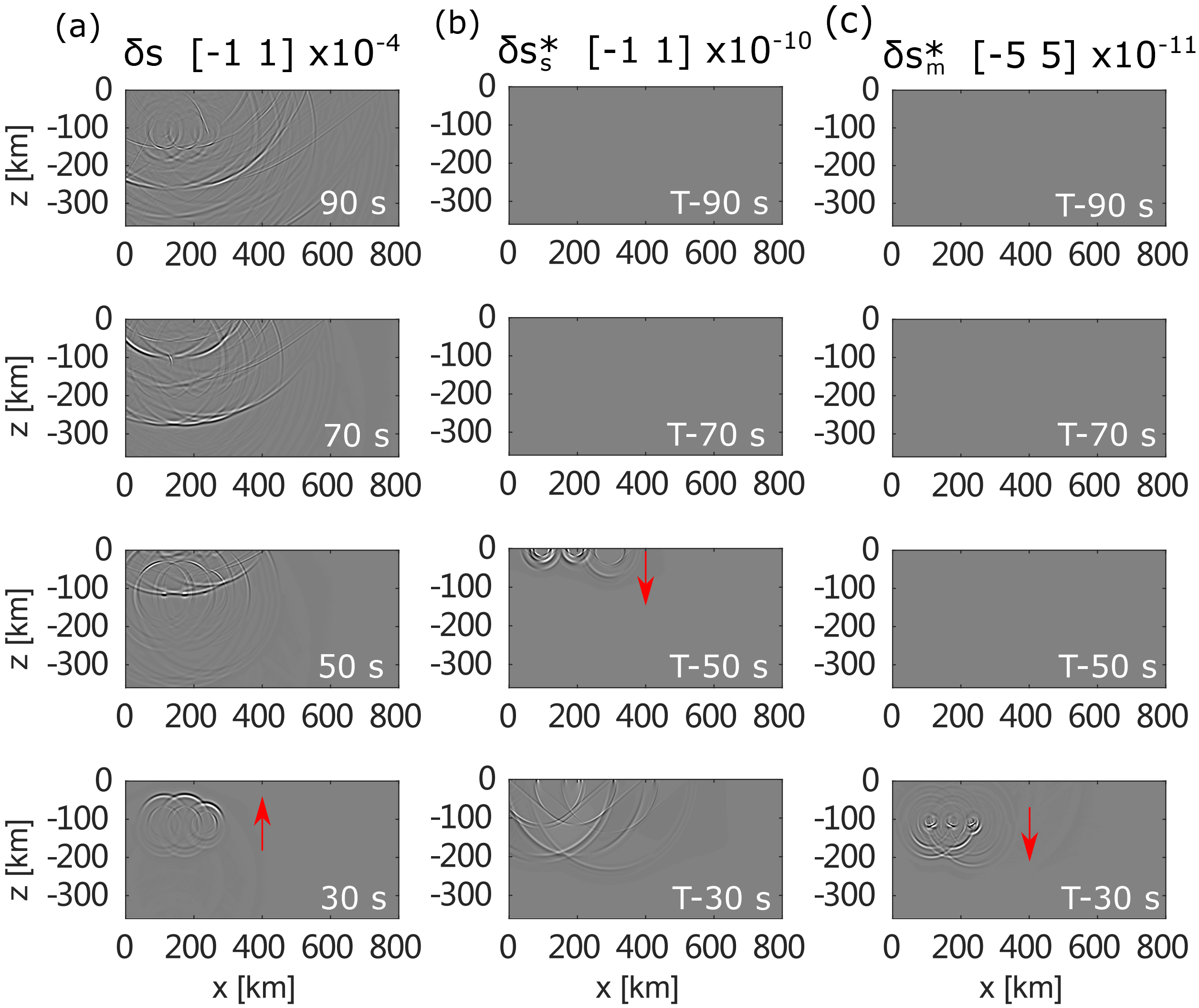

Figure 5: A few time steps of selected perturbed fields computed on the fly using the first-order finite-difference approximation. (a) shows the perturbed forward fields. (b) shows the perturbed adjoint fields due to the perturbation of the adjoint source. (c) shows the perturbed adjoint fields due to the perturbation of the model. The perturbed fields, e.g., generated around the red arrows, are due to the perturbations either from the model or from the adjoint source.

Forward output	Backward and adjoint output
Database******.bin	proc******_rho_kappa_mu_kernel.dat proc******_rho_kappa_mu_kernel_Ha.dat proc******_rho_kappa_mu_kernel_Hbm.dat proc******_rho_kappa_mu_kernel_Hbs.dat proc******_rho_kappa_mu_kernel_Hc.dat proc******_rho_kappa_mu_kernel_Habc.dat
absorb_elastic_bottom******.bin absorb_elastic_left******.bin absorb_elastic_right******.bin absorb_elastic_bottom_m2_******.bin absorb_elastic_left_m2_******.bin absorb_elastic_right_m2******.bin	proc******_rhop_alpha_beta_kernel.dat proc******_rhop_alpha_beta_kernel_Ha.dat proc******_rhop_alpha_beta_kernel_Hbm.dat proc******_rhop_alpha_beta_kernel_Hbs.dat proc******_rhop_alpha_beta_kernel_Hc.dat proc******_rhop_alpha_beta_kernel_Habc.dat
AA.S****.BXX.semd AA.S****.BXZ.semd AA.S****.BXX.semd_m2 AA.S****.BXZ.semd_m2	
lastframe_elastic******.bin lastframe_elastic_m2_******.bin	

Figure 6: Key files output from the forward simulation and the simultaneous backward and adjoint simulation in the QuadSEM. The left column shows the files output from the forward simulation. The first row shows the meshing database which includes the internal model to be replaced by the two external models before the main time loop in the simultaneous backward and adjoint simulation. The second row shows the absorbing boundary fields, where the shadow part indicates files output for model $\mathbf{m}_2$. The third and fourth rows show the seismograms registered at the receivers and the last state of the forward field. These files output in the forward simulation will be used in the simultaneous backward and adjoint simulation. The right column shows the key files output in the simultaneous backward and adjoint simulation, including the Fréchet kernels, the approximate Hessian kernels ('Hbs'), and the full Hessian kernels ('Habc'), etc. In the right column, the top part shows kernels for the ρ, κ, and μ parameter set and the bottom part shows kernels for the ρ, α, and β parameter set.

Simultaneous backward and adjoint simulations for the computation of Fréchet and Hessian kernels on the fly

Figure A1 shows the simultaneous backward and adjoint simulations for the computation of Fréchet kernels on the fly, where the backward simulation is designed to reconstruct the forward fields backward in time during the adjoint simulation. In this way, the Fréchet kernels can be constructed on the fly since the forward fields for time t and the adjoint fields for time $(T - t)$ or vice versa can be simultaneously accessed. The T indicates the total simulation time. This on-the-fly strategy can be extended to compute the Hessian kernels but the solutions of two forward and two adjoint equations are combined. Figure A2 shows the simultaneous computation of several forward and adjoint fields for constructing the Hessian kernels on the fly.

Computation of Hessian kernels by wavefield storage method

Hessian kernels can be computed when the required wavefields are available. To get the required fields, we use one forward simulation and two adjoint simulations (see Figure A3). The forward simulation is to compute and save four forward fields, that is $\mathbf{s}(\mathbf{m}_1)$, $\mathbf{s}(\mathbf{m}_2)$, $\ddot{\mathbf{s}}(\mathbf{m}_1)$, $\ddot{\mathbf{s}}(\mathbf{m}_2)$, where $\mathbf{m}_2 = \mathbf{m}_1 + v\delta\mathbf{m}$. The first adjoint simulation (Adjoint simulation I) is designed to compute and save the adjoint field $\mathbf{s}^{\dagger}(\mathbf{m}_2)$. The second adjoint simulation (Adjoint simulation II) is a simultaneous adjoint simulation and the Hessian calculation. Figure A3 shows the computation of full Hessian kernels. It can be similarly changed for the computation of approximate Hessian kernels, where one needs to store $\mathbf{s}^{\dagger}_{s}(\mathbf{m}_1)$ in the Adjoint simulation I. Figure A4 shows each component of the Hessian kernels with respect to the three model parameters. For this example, the *Model 2* and the single source-receiver pair (as the section of Single source-receiver combination) are used. Only the first P-wave arrival is used for the calculation of the traveltime adjoint sources.

List of Figure Captions for Appendix

Figure A1. Forward simulation and the simultaneous backward and adjoint simulation for computing the Fréchet kernels.

Figure A2. Simultaneous backward and adjoint simulation in the QuadSEM for the computation of Hessian kernels on the fly.

Figure A3. A workflow illuminating the computation of full Hessian kernels by the wavefield storage method (WSM).

Figure A4. Four components of the full Hessian kernels with respect to the model given in ρ, α, and β.

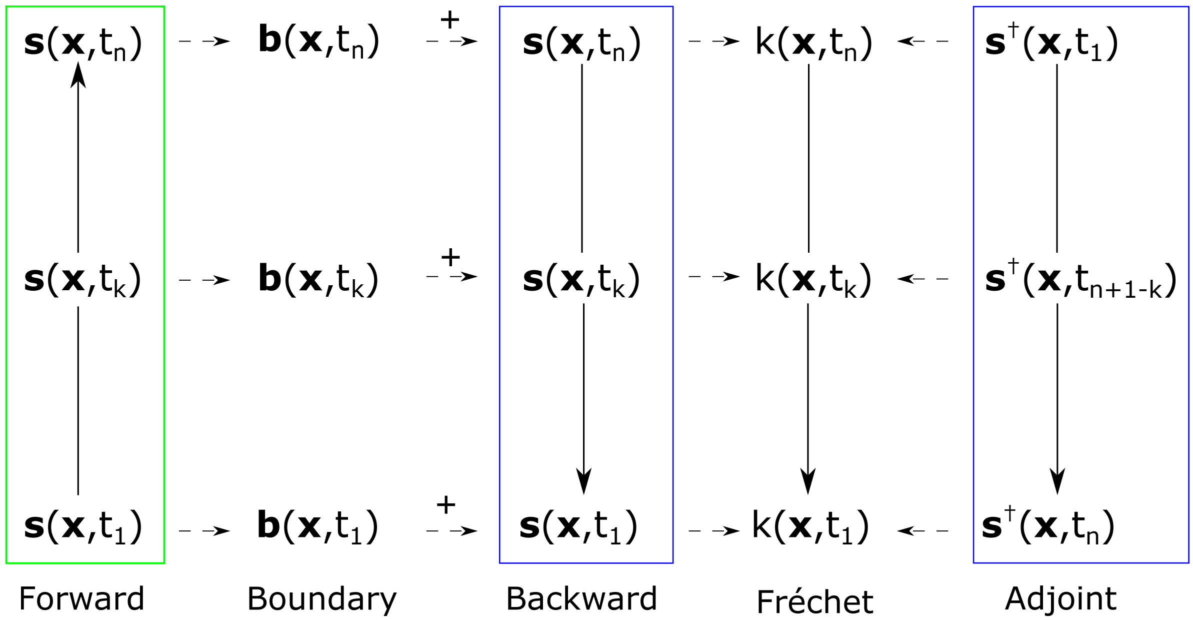

Figure A1: Forward simulation (green rectangle) and the simultaneous backward and adjoint simulation (two blue rectangles) for computing the Fréchet kernels. The forward simulation is started from the first time step t_1 and ended at the last time step t_n. The absorbing boundary field $\mathbf{b}(\mathbf{x}, \mathrm{t_k})$ of each time step t_k and the last state field $\mathbf{s}(\mathbf{x}, \mathrm{t_n})$ are stored in the forward simulation. The backward simulation takes the last state field as a start point and reconstructs the forward field backward in time. In each time step, the absorbing boundary field $\mathbf{b}(\mathbf{x}, \mathrm{t_k})$ is re-injected into the backward simulation to reconstruct the forward fields (called backward fields here). The adjoint simulation is started from the time-reversed adjoint source from the receivers. The Fréchet kernels at each time step or at a sub-sampled time step are constructed on the fly based upon the backward and adjoint fields. If each time step is used, the kernels are summed at each time step until the final step as $K_m = \sum_{k=1}^{n} \mathrm{K}(\mathbf{x}, \mathrm{t_k})\delta t$, where δt is time interval in the simulation.

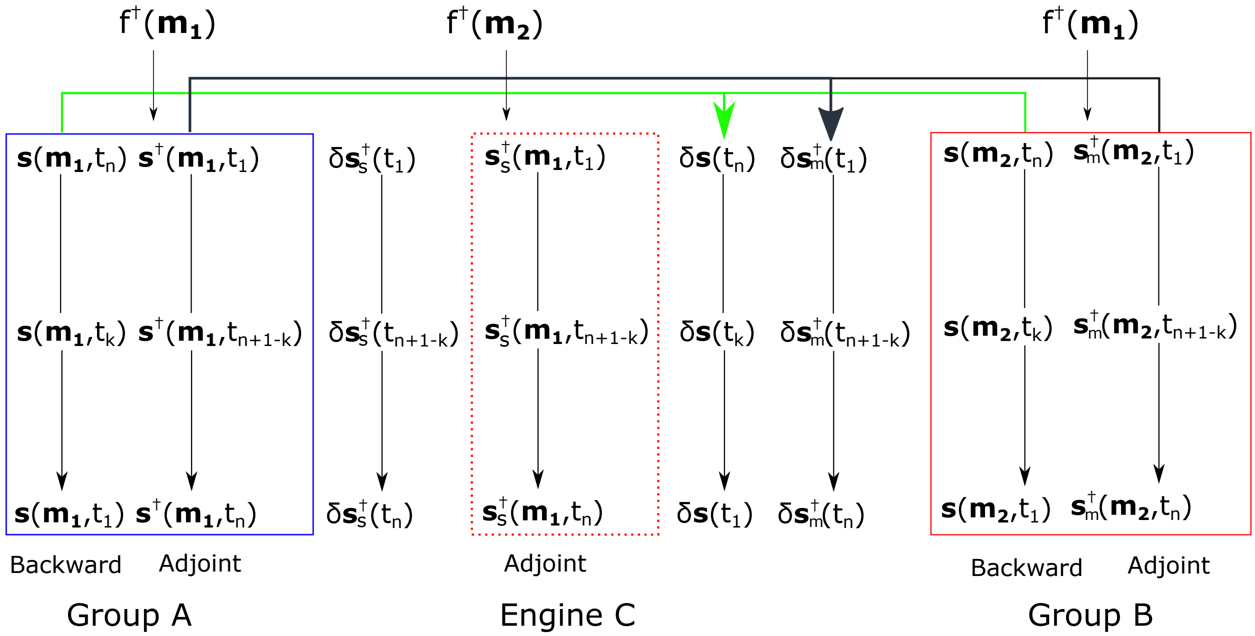

Figure A2: Simultaneous backward and adjoint simulation in the QuadSEM for the computation of the Hessian kernels on the fly. Group A: the solutions of forward and adjoint equations are combined and used for model $\mathbf{m}_1$, where one solution is solved for the backward simulation and the other is solved for the adjoint simulation. Group A is designed to compute the backward and adjoint fields for model $\mathbf{m}_1$. On the right side, Group B combines the solutions of the forward and adjoint equations for model $\mathbf{m}_2$. Engine C represents one solution of the adjoint equation designed to compute the adjoint field due to the perturbation of the adjoint source $f^{\dagger}(\mathbf{m}_2)$. The simulation in Engine C is the same as the adjoint simulation in Group A except the source term is different. We design the workflow to show the computation of each required wavefield. In the full Hessian kernel calculations, we use Group A and Group B, just changing the adjoint source of Group B from $f^{\dagger}(\mathbf{m}_1)$ to $f^{\dagger}(\mathbf{m}_2)$. Since all the fields are computed on the fly for each designed time step (each time step or adaptive time step), the perturbed fields to be used in the calculation of Hessian kernels can be computed also on the fly.

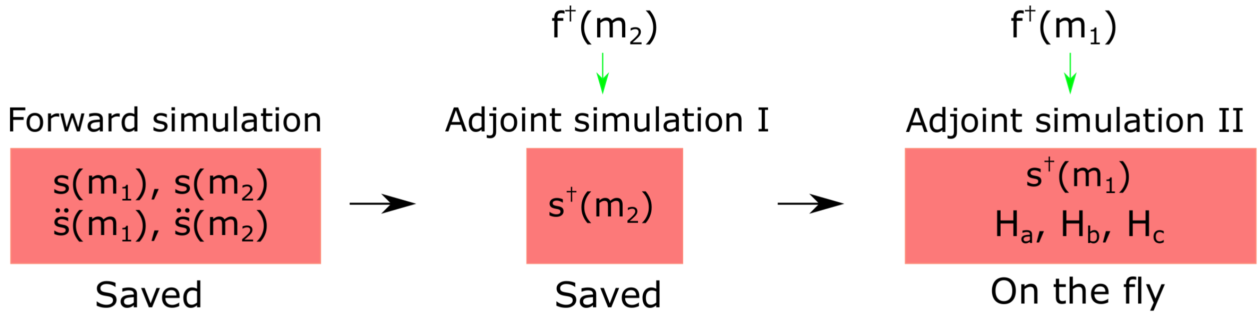

Figure A3: A workflow illuminating the computation of full Hessian kernels by the wavefield storage method (WSM) using the adaptive time integration or storing the fields at all time steps. The first step (Forward simulation) is designed to compute and save the forward fields and the second step (Adjoint simulation I) is to compute and save one adjoint field due to the perturbations of model and adjoint source. The last step (Adjoint simulation II) is to compute one adjoint field $\mathbf{s}^{\dagger}(\mathbf{m}_1)$ on the fly, and read one time step of the saved five fields into the temporary memory for the computation of the full Hessian kernels. The $\mathbf{f}^{\dagger}(\mathbf{m}_1)$ and $\mathbf{f}^{\dagger}(\mathbf{m}_2)$ denote the two adjoint sources computed from the measurements of the two models.

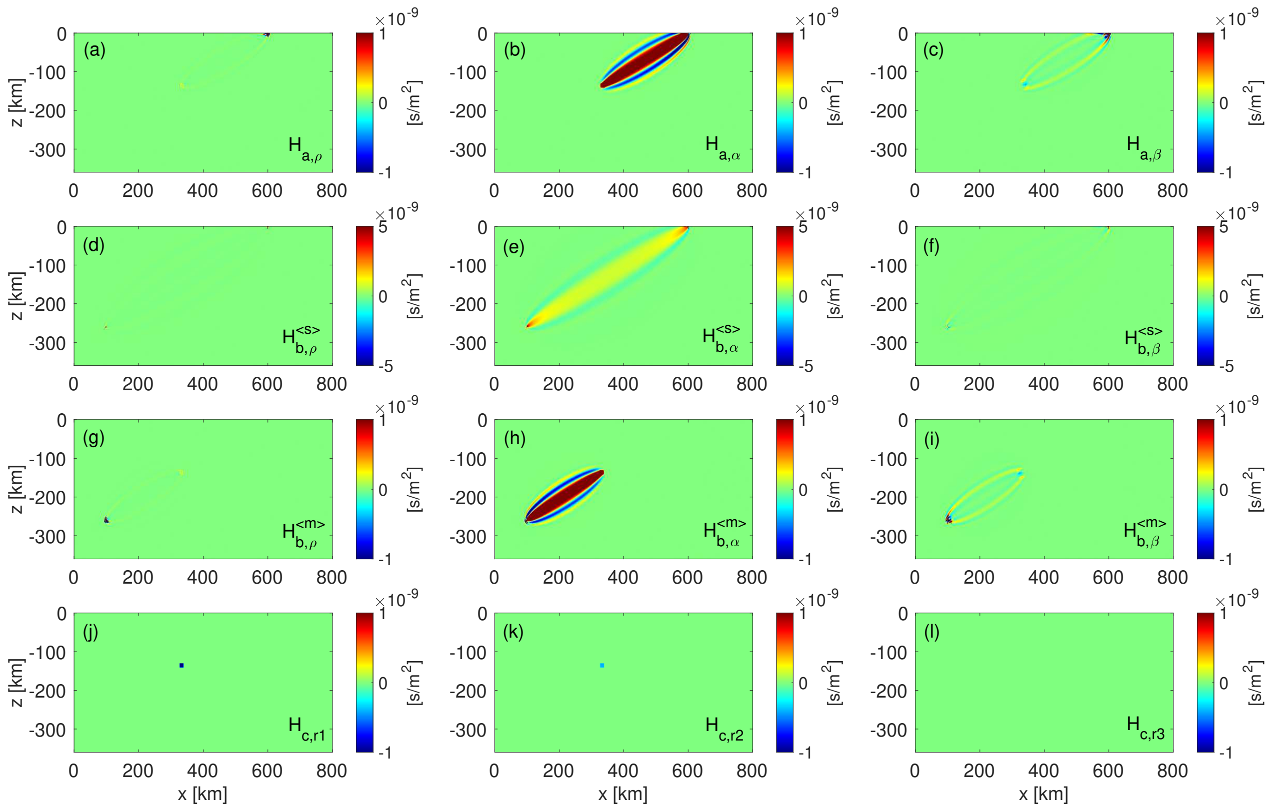

Figure A4: Four components of the full Hessian kernels with respect to the model given in ρ, α, and β. The top first row shows the H_a component with respect to the three model parameters. Only the $\mathrm{H}_{a,\alpha}$ is well observed since only the first P-wave arrival is used for the adjoint source calculation. The second row shows the $\mathrm{H}_b^{\langle s\rangle}$ component, which is approximate Hessian kernels due to the perturbation of the adjoint source to the adjoint field. The third row shows the $\mathrm{H}_b^{\langle m\rangle}$ component which is due to the perturbation of the model for the adjoint field. The bottom row shows the H_c component. Only the kernels for $\mathrm{H}_{c,r1}$ and $\mathrm{H}_{c,r2}$ are observed since the K_β is very small due to the use of the first P-wave arrival only. The ri (where $i = 1, 2, 3$) indicates the three rows in the H_c expression. The full Hessian kernels are obtained by summing the four components together.