Multi-solver spectral-element and adjoint methods

Yujiang Xie[1], Catherine Rychert[1], Nicholas Harmon[1], Qinya Liu[2], 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

Key Points:

- Simultaneous construction of Fréchet and Hessian kernels on the fly based upon spectral-element and adjoint methods.
- Only about a 2-fold computational cost required for the simultaneous computation when compared to the computation of Fréchet kernels.
- Truncated-Newton full-waveform inversion can be performed efficiently based upon the multi-solver spectral-element and adjoint methods.

Corresponding author: Yujiang Xie and Catherine Rychert, `yujiang.xie@soton.ac.uk;C.Rychert@soton.ac.uk`

Abstract

The spectral-element method (SEM) for simulating wave propagation is widely used with adjoint methods for full-waveform inversion. Typically, SEM is used to compute forward and adjoint wavefields, which is then applied to evaluate the Fréchet derivatives for updating the seismic structural model. The Hessian is rarely computed as the high computational and storage costs, although it can improve the accuracy of the model update and model convergence. Instead the approximate Hessian is determined, which is obtained with less computational effort. We present a method for simultaneously constructing Fréchet and Hessian kernels on the fly, which we call Multi-solver spectral-element and adjoint methods (Multi-SEM). Rather than storing all the wavefields, Multi-SEM is computed on the fly and requires only about a 2-fold computational cost when compared to the computation of Fréchet kernels. Numerical examples demonstrate the functionality of the method and the computer codes are provided with this contribution.

Plain Language Summary

Recent advances in high-performance computing and quantum computing mean that full-waveform inversions (FWIs) are now routinely performed to achieve high-resolution imaging of the interior structure of the Earth. Typically, these are done using first-order derivatives, known as Fréchet kernels. Second-order derivatives, known as Hessian kernels, can be used to speed up convergence and to determine higher resolution of small-scale features. However, the Hessian is not commonly computed due to computational challenges such as high storage needs and long run times related to reading and writing. We present the Multi-solver spectral-element and adjoint methods (Multi-SEM), which generalizes

the conventional spectral-element and adjoint methods from the computation of Fréchet kernels into the simultaneous computation of Fréchet and Hessian kernels. The kernels are computed on the fly, which means that only a double computational cost is required in comparison to the computation of Fréchet kernels only without the need to store several 4-D wavefields, saving several TB of memory. We present the Hessian Kernels for two different models to demonstrate their potential for achieving higher accuracy. Multi-SEM improves the capability of FWI to image Earth structure, particularly in regions characterized by small scale heterogeneities such as subductions zones.

1 Introduction

During the past twenty years the spectral-element method (SEM) (e.g., Patera, 1984; Maday & Patera, 1989) has been widely used in the seismology community for simulating the propagation of surface and body waves in the Earth (e.g., Komatitsch & Tromp, 1999, 2002a, 2002b; Komatitsch et al., 2002c; Chaljub & Valette, 2004; Tromp et al., 2005; Liu & Tromp, 2006; Chen et al., 2007; Tape et al., 2007; Chaljub et al., 2007; Liu & Tromp, 2008; Tromp et al., 2008; Fichtner et al., 2009; Tape et al., 2009; Peter et al., 2011; Liu & Gu, 2012; Afanasiev et al., 2019), see Tromp (2020) for a review. Compared to other solvers, the SEM is popular in seismology due to its great ability in handling complex geometries and simulating surface waves with low numerical dispersion. Since 2005, the adjoint method (e.g., Tarantola, 1984; Talagrand & Courtier, 1987) was successfully connected with the SEM by Tromp et al. (2005), and has been used to compute the sensitivity kernels with the forward and adjoint fields. For the elastic case, an implementation of little storage cost requires two simulations per event: a forward simulation of the earthquake to the receivers, and another simulation carrying both the forward wavefield and the adjoint wavefield simultaneously. In the latter simulation, the forward field is reconstructed backward in time and the adjoint simulation is triggered by time-reversed adjoint sources simultaneously at receivers. The computation of Fréchet kernels is achieved via correlation of the reconstructed forward fields with the adjoint fields (e.g., Tromp et al., 2008; Liu & Gu, 2012).

Computation and use of event-based Fréchet kernels from SEM and adjoint methods have been performed in many studies. However, due to the high computational cost, the use of Hessian kernels for one source and multiple receivers is not common even though the theory was presented (e.g., Fichtner & Trampert, 2011). In practice, authors may use the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm (e.g., Nocedal, 1989; Liu & Nocedal, 1989; Zou et al., 1993; Nocedal & Wright, 1999), which computes the product of the inverse approximate Hessian and the gradient to estimate model update using gradients and models from previous iterations. This solution is popular due to its numerical efficiency. One competitive algorithm called truncated-Newton optimization (e.g., Nash, 1985; Grippo et al., 1989; Nash & Nocedal, 1991; Nash, 2000) has been well-documented in exploration seismology for full-waveform inversion (see e.g., Métivier et al., 2014, 2017; Pan et al., 2017; Yang et al., 2018; Matharu & Sacchi, 2019), and it has been demonstrated that it produces better results than the L-BFGS algorithm in multi-parameter full-waveform inversion due to its mitigation in inter-parameter trade-off, such as inversions for vp, vs, density, attenuation, and anisotropy or some of them. Significant differences between the approximate Hessian and the full Hessian were observed (Fichtner & Trampert, 2011). The truncated-Newton method is rarely used in earthquake seismology due to the computational issue to construct the Hessian kernels. However, efficient solutions constructing the Hessian kernels may make the truncated-Newton method more appealing for full-waveform inversion (e.g., Tromp, 2020) or adjoint tomography (e.g., Tape et al., 2007, 2009).

The Hessian kernels can be computed by the method of Fichtner and Trampert (2011) using pre-existing implementations of the adjoint tomography. One such approach involves storing the forward and adjoint wavefields at all or sub-sampled time steps for later determination of the Fréchet and Hessian kernels. This practically leads to big challenges for the Hessian construction because of huge disk storage requirements in saving forward and adjoint fields as well as their perturbations. Practical simulations may involve tens to hundreds of millions of grid points and tens of thousands of time steps for each wavefield. For computing the Hessian kernels, at least four sets of such wavefields are required (Fichtner & Trampert, 2011). The disk storage may become a daunting issue even after sub-sampling schemes are introduced.

Another type of method to compute the Hessian is the scattering integral (SI) method (e.g., Chen, Zhao, & Jordan, 2007; Chen, Jordan, & Li, 2007; Chen, 2011; Lee et al., 2014), which is closed related to the adjoint methods (Tromp et al., 2005, 2008). The relative computational efficiency of the two types of methods for the kernel calculation and inversion depends on the overall problem geometry, in particular the ratio of the number of sources to receivers (see Chen, Jordan, & Li, 2007; Lee et al., 2014). The SI method may be more computationally efficient when the number of sources is comparable or larger than the number of receivers. But when the number of receivers is large or the computation domain is expansive or shorter periods seismic waves are inverted, the computation and storage demand for the SI may become a daunting issue, in particular when the updated structure is far away from the reference model where the Hessian for individual measurement needs to be recomputed in each iteration of the inversion. The disk storage can be another challenging issue. For example in the Southern California crustal inversion presented by Lee et al. (2014), the peak disk storage during the SI inversion was about 39 Tb in addition to the huge input/output (I/O) overhead.

We present a numerically efficient method to compute Hessian kernels for one event, which we call Multi-solver spectral-element and adjoint method (Multi-SEM). It is different from the aforementioned wavefield storage techniques. Further developed from the adjoint methods in Tromp et al. (2005); Liu and Tromp (2006, 2008) where sensitivity kernels are calculated from the simultaneous computation of adjoint wavefield and back-reconstructed forward field, the Multi-SEM resolves the storage issue by constructing the Fréchet and Hessian kernels on the fly for each or incremental time step through five SEM solvers. Since only one time-step of both wavefields and the integrated kernels are kept in memory, the Multi-SEM is cheap in memory and easy to realize on present-day hardware with only limited storage required as that of adjoint methods (Tromp et al., 2005; Liu & Tromp, 2006; Tromp et al., 2008), e.g., storing for the last frame of the forward fields. The computation of the Hessian kernels by Multi-SEM requires only about two times the CPU time compared to the computation of the Fréchet kernels alone. The Multi-SEM method can be implemented on pre-existing spectral-element solvers such as the SPECFEM2D (`https://github.com/geodynamics/specfem2d`), where one just slightly rearranges the coding structure by coupling two solvers simultaneously for the forward simulation and coupling five solvers simultaneously for the simultaneous backward and adjoint simulation. Although five solvers are coupled and used, memory requirement could be designed to be as small as possible since only one time-step of both wavefields and the integrated kernels are kept in the temporary memory. The computational cost is slightly reduced over individual five solver runs as all solvers share the same mesher database files except those describing model material properties for the model and its update as discussed in Section 3.

In this paper, we first review the theory on Fréchet and Hessian kernels and then present the Multi-SEM method. Results for Fréchet and Hessian kernels are presented and discussed for 2-D synthetic models. The related codes are published in the public domain for dissemination.

2 Theory

2.1 Fréchet kernels

Fréchet kernels, gradient 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 also approach the true model. By perturbing the measurements as $\delta\chi$ with respect to an isotropic model $\mathbf{m}$, we have (also see Tromp et al., 2005)

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

where $\overline{K}_m = K_m\,\mathbf{m}$. The $\overline{K}_m$ or 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 into different components depending on the choice of model parameterization (See Section 1 of the Supporting Information). For simplicity, we only show the case for model parameterization given by $\mathbf{m} = (\rho,\, \alpha,\, \beta)$, where ρ denotes the density and α and β denote the compressional and shear wave speeds. The kernel 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, we rewrite the Fréchet kernels as a function of the forward and adjoint fields

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

where $\mathbf{s}$ and $\mathbf{s}^\dagger$ are the forward and adjoint displacement fields, and $\ddot{\mathbf{s}}$ is the second-order time derivative of $\mathbf{s}$, i.e., the forward acceleration field. In practice, the field storage method and/or the forward-field back-reconstruction method may be used to compute the Fréchet kernels (see Section 1 of the Supporting Information).

2.2 Hessian kernels

2.2.1 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 & 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 Hessian kernels. Based upon the work of Fichtner and Trampert (2011), we rewrite each part of the product as

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

$$\mathrm{H}_c(\rho,\alpha,\beta) = \begin{pmatrix} \rho^{-1} K_\alpha(\mathbf{s}^\dagger, \mathbf{s})\delta\alpha + \rho^{-1} K_\beta(\mathbf{s}^\dagger, \mathbf{s})\delta\beta \\ \rho^{-1} K_\alpha(\mathbf{s}^\dagger, \mathbf{s})\delta\rho + \alpha^{-1} K_\alpha(\mathbf{s}^\dagger, \mathbf{s})\delta\alpha \\ \rho^{-1} K_\beta(\mathbf{s}^\dagger, \mathbf{s})\delta\rho + \beta^{-1} K_\beta(\mathbf{s}^\dagger, \mathbf{s})\delta\beta \end{pmatrix}. \tag{6}$$

where $\delta\mathbf{s}$ and $\delta\mathbf{s}^\dagger$ denote the perturbed forward and adjoint field due to 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. Eq. (5)-(6) show a link between the Hessian kernels (e.g., Fichtner & Trampert, 2011) and the Fréchet kernels (e.g., Tromp et al., 2005). It implies that the implementation framework for computing the Fréchet kernel can be used to compute the

Hessian kernels by replacing the regular field with its associated perturbed field. H_a can be computed with the implementation of eq. (3) by replacing the forward fields with the perturbed forward fields. H_b practically includes two contributions, i.e.,

$$\mathrm{H}_b = \mathrm{H}_b^{\langle m \rangle} + \mathrm{H}_b^{\langle s \rangle}, \tag{7}$$

where

$$\mathrm{H}_\mathrm{b}^{\langle m \rangle}(\rho, \alpha, \beta) = \begin{pmatrix} K'_\rho(\delta \mathbf{s}_m^\dagger, \ddot{\mathbf{s}}) \\ K_\alpha(\delta \mathbf{s}_m^\dagger, \mathbf{s}) \\ K_\beta(\delta \mathbf{s}_m^\dagger, \mathbf{s}) \end{pmatrix}, \mathrm{H}_b^{\langle s \rangle}(\rho, \alpha, \beta) = \begin{pmatrix} K'_\rho(\delta \mathbf{s}_s^\dagger, \ddot{\mathbf{s}}) \\ K_\alpha(\delta \mathbf{s}_s^\dagger, \mathbf{s}) \\ K_\beta(\delta \mathbf{s}_s^\dagger, \mathbf{s}) \end{pmatrix}. \tag{8}$$

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 *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 construction for H_c is straightforward based upon the Fréchet kernel K_m and the perturbation of the model $\delta \mathbf{m}$.

2.3 Perturbed fields and perturbed model

As eq. (5)-(8) show that the Hessian kernels can be computed with the same implementation framework as that for the Fréchet kernels by adjoint methods in eq. (3), any spectral-element package for wavefield generation can be redesigned and adapted to compute the Hessian kernels just with additional efforts to compute the perturbed forward fields $\delta \mathbf{s}$ and the perturbed adjoint field $\delta \mathbf{s}^\dagger$ due to a model perturbation $\delta \mathbf{m}$ and the perturbed adjoint source.

2.3.1 Perturbed fields for H_a *component*

The H_a component of the Hessian kernels accounts for the perturbation of the forward field, $\delta\mathbf{s}$. If we denote the wavefield generated due to the perturbed model $\mathbf{m}_r + v\delta\mathbf{m}$ as $\mathbf{s}(\mathbf{m}_r{+}v\delta\mathbf{m};\mathbf{x},t)$, we may obtain the perturbed forward field due to $v\delta\mathbf{m}$ as (see also Fichtner & Trampert, 2011)

$$\delta\mathbf{s} = \lim_{v\to 0}\frac{1}{v}[\mathbf{s}(\mathbf{m}_r + v\delta\mathbf{m};\mathbf{x},t) - \mathbf{s}(\mathbf{m}_r;\mathbf{x},t)], \tag{9}$$

where $\mathbf{m}_r$ denotes the reference model, $r = 0, 1, 2, ..., N$ represents the iteration number, and $\mathbf{m}_0$ means the initial model. The same consideration applies to the perturbed acceleration field $\delta\ddot{\mathbf{s}}$ for density kernel computation. In practical application such as full-waveform inversion, the model perturbation can be estimated by using truncated Newton optimization (see e.g., Métivier et al., 2014, 2017; Pan et al., 2017; Yang et al., 2018; Matharu & Sacchi, 2019). In the first iteration, the steepest descent method may be used to compute the model update. For more details of the $v\delta\mathbf{m}$ determination, please refer to Fichtner and Trampert (2011). The computation of H_a is straightforward if we use the field storage method. However, storage and I/O demands may be quite significant when the model size or the number of sources is large.

2.3.2 Perturbed fields for H_b *component*

The H_b component consists of two contributions. One is from the approximate Hessian kernels $\mathrm{H}_b^{\langle s\rangle}$ due to the perturbation of the adjoint source, and the other is from the $\mathrm{H}_b^{\langle m\rangle}$ due to the perturbation of the model. To compute $\mathrm{H}_b^{\langle s\rangle}$, the approximate perturbed adjoint field may be calculated as

$$\delta \mathbf{s}_s^\dagger = \mathbf{s}_s^\dagger(\mathbf{m}_r; \mathbf{x}, T-t) - \mathbf{s}^\dagger(\mathbf{m}_r; \mathbf{x}, T-t). \tag{10}$$

where 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 only difference between the two adjoint fields is the adjoint sources used since the former accounts for the perturbation of the adjoint source as a result of $v\delta\mathbf{m}$.

The perturbed adjoint field for the $\mathrm{H}_b^{\langle m \rangle}$ calculation may be given by

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

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 $\mathbf{f}^\dagger(\mathbf{m}_r; \mathbf{x}, T-t)$. The adjoint sources may be different based on the choices of seismic data functional χ as discussed in Tromp et al. (2005). Thereafter, the total perturbed adjoint field is

$$\delta \mathbf{s}^\dagger = \delta \mathbf{s}_s^\dagger + \delta \mathbf{s}_m^\dagger. \tag{12}$$

2.3.3 Perturbed model for H_c *component*

From eq. (6), it is clear that the computation of H_c relies on the Fréchet kernels and model perturbation. 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 another two sets of parameterization (Fichtner & Trampert, 2011). See also Section 2 of the Supporting Information.

3 Implementation

The computation of Hessian kernels relies on the regular and perturbed forward and adjoint fields. Its implementation is relatively straightforward based on the wavefield storage method (WSM) (see Section 3 of the Supporting Information), where for each time step or incremental time step, the associated stored fields are read into temporary memory for the kernel calculation, and this process is repeated until the end of simulation. In this section, we show how the Hessian kernels is computed on the fly by the Multi-SEM. For the following examples we only consider cases with purely elastic models.

3.1 Forward simulation

Figure 1 shows the comparison between the single-solver SEM and the Multi-SEM for forward simulations. The Multi-SEM carries wavefield simulations for two models simultaneously, e.g., $\mathbf{m}_1$ and $\mathbf{m}_2$, instead of one model used by the single-solver SEM, where $\mathbf{m}_2 = \mathbf{m}_1 + v\delta\mathbf{m}$. In this case, the wavefields, including displacement $\mathbf{s}$, velocity $\mathbf{v}$, acceleration $\ddot{\mathbf{s}}$, and the boundary contribution $\mathbf{b}$ (we use $\mathbf{b}$ for generality since it is typically the velocity fields or the velocity and force fields when the SEM domain is coupled 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 a 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 Multi-SEM, 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 Multi-SEM are about twice the cost in the single-solver SEM simulation. The forward simulations either for the single-solver SEM or the Multi-SEM are designed to provide the absorbing boundary fields, the last state of the forward field, and the seismograms at receivers, for the subsequent simulations.

3.2 Simultaneous backward and adjoint simulations

Simultaneous backward and adjoint simulations are widely used in many SPECFEM packages (`https://geodynamics.org/cig/software/`) to construct the Fréchet kernels on the fly. A workflow for computing the Fréchet kernels by conventional single-solver SEM method is shown in Figure S1 of the Supporting Information. 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 as the forward field is reconstructed backward in time. The simulations for backward reconstruction and adjoint wavefield are performed simultaneously so that the corresponding time slices of forward and adjoint wavefield can be accessed both in memory in order to calculate Fréchet kernels. The same course is used in the Multi-SEM with five SEM solvers instead of two (see Figure 2 and Figure S2). In this case, the regular, perturbed forward fields and the regular, perturbed adjoint fields for the two models are simultaneously reconstructed and computed for a time step, so that the Fréchet and Hessian kernels can be calculated on the fly as wavefield products are computed and integrated over time steps (see Figure 2 and Figure S2). Although the five SEM solver engines are coupled and use the same mesh database excluding $\mathbf{m}_1$ and $\mathbf{m}_2$ loaded externally. The memory cost is small since only one time step of the various fields and the integrated kernels are kept in memory compared to the wavefield storage methods. Each Fréchet kernel needs 3 (1 in forward and 2 in adjoint) simulations, while the Multi-SEM carries 7 (2 in forward and 5 in adjoint) simulations for the simultaneous computation of Fréchet and Hessian kernels. During the adjoint simulation, the memory is not 5/2 times that of a regular kernel simulation due to the shared memory for the same mesh database (excluding the two models' material properties). The CPU hours will be less than 2.5 (5/2) times due to the shared mesher for all SEM solver. Most of the CPU time is spent computing the strain and stress calculations.

4 Numerical Examples

4.1 Models

To test the numerical implementation of Multi-SEM, three models are considered in this study. First, a homogeneous 2D model (*Model 1*) of the size of 800 km in the horizontal direction and 360 km in the vertical direction and with density ρ=2900 kg/m^3, compressional wave speed α=8000 m/s, and shear wave speed β=4800 m/s, is used as a starting background model to generate initial wavefields and waveforms. We use the internal mesher of the SPECFEM2D package to mesh the model with 400 elements in the horizontal direction and 360 elements in the depth direction. With 5×5 Gauss-Lobatto-Legendre (GLL) points used for each element in 2D, this leads to $\sim$ 500 m/250 m horizontal/vertical grid-point spacing for the model. The second and the third model 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×10 km squared area centered at the horizontal location of 335 km and depth of 135 km (see Figure 3c for the perturbation location indicated by H_c). The third model (*Model 3*) comprises three anomalies of the size of 8 km×10 km, centered at the same depth of 115 km and horizontally at 120 km, 180 km, and 240 km, respectively, with +10% perturbations in α and β (see Figure 3f for the three perturbation locations indicated by H_c). No density perturbation is considered for the second and third model. 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. The locations of the perturbations are indicated by the H_c kernels in Figure 3.

4.2 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) with the standard Ricker wavelet source-time function of dominant frequency of 0.5 Hz. A single receiver is placed on the surface of the model at (x, z)=(600 km, 0 km). The simulations use $dt = 0.01$ s and run for a total of $10,000$ time steps.

To see the kernels over the model perturbation, we show here the Fréchet kernels for *Model 2*, and the Hessian kernels for *Model 1* and *Model 2*. The Fréchet kernels computed for *Model 1* are shown in Figure S16 of the Supporting Information. The Multi-SEM computes the Fréchet kernels shared the same solvers with conventional SEM (see Figure 2). The first row of Figure 3 (Part I) shows the Fréchet kernel, the approximate Hessian kernel, and the full Hessian kernel. A zoomed-in version around the perturbations is given in the first row of Part II. Detailed descriptions about the kernels are given in the figure caption for Figure 3.

For the adjoint field calculations we use traveltime adjoint sources with waveform window selected for the P phase, and the same procedure can be applied to the full waveforms. It takes the Multi-SEM method about a total of 31 mins with maximum memory usage of $\sim$3.1 GB to simultaneously compute the Fréchet and Hessian kernels on a standard laptop (with 2.3 GHz Dual-Core Intel Core i5 processor and 8GB 2133 MHz LPDDR3 memory). In comparison, the computation of Fréchet kernel alone by the conventional SEM and adjoint method takes about 13.5 mins with maximum memory usage of 1.5 GB. Therefore in this case, all the quantities computed by Multi-SEM takes $\sim$2.29 times the CPU time and $\sim$2.06 times the memory compared to the computation of Fréchet kernels. The storage required for the Multi-SEM is small due to the on-the-fly nature of the calculations, which takes about 1 GB disk space to store the absorbing boundary fields, the last-state forward fields as well as the seismograms, while for the wavefield storage method (WSM, see Section 3 of the Supporting Information), it requires about 400 GB disk space to store these fields even without considering the density kernels.

4.3 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 and Hessian kernels are computed with respect to the perturbation in *Model 3*. The source is placed at $(x, z) =$ (150 km, -260 km) with the same source time function as in Section 4.2. Three receivers are placed on the top surface of the model horizontally located at 100 km, 200 km, and 300 km, respectively. The total number of time steps and time interval are the same as the example in Section 4.2.

The second row of Figure 3 (Part I) shows the Fréchet kernel, the approximate Hessian kernels, and the full Hessian kernels computed for P phase on the seismograms. A zoomed-in version of Figure 3 (Part I) around the perturbations is given in 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 section 4.2 since the simulation cost is almost independent of the number of receivers. There is one additional step in the window picking and computation of adjoint source, which is much cheaper than the field calculations. A few selected time steps of the regular wavefields and their perturbations are shown in Figure S3 and Figure S4 in the Supporting Information. The adjoint sources computed from the seismograms for $\mathbf{m}_1$ and $\mathbf{m}_2$ are also provided there in Figure S5-S7. The key output files for the Multi-SEM package in the forward simulation and in the simultaneous backward and adjoint simulation are presented in Figure S8.

5 Discussions

We found significant differences between the approximate Hessian kernels and the full Hessian kernels for both the one- and multi-receiver case (Figure 3), as also noted in Fichtner and Trampert (2011). Most notably, the amplitudes of the Hessian kernels can be up to 100% stronger than those of the approximate Hessian kernels within the red areas, as areas also 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 in the vicinity of the perturbation suggest that the inversion using the Hessian instead of the approximate Hessian will result in better illumination in the region of the model perturbation, in addition to distributing them along the kernel.

In the multi-receiver case, we observe a similar 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 the use of full Hessian kernels would provide better resolution for smaller anomalies within the earth model.

The Hessian kernels are typically used with the Fréchet kernels for computing the model updating based upon truncated Newton optimization (Nash, 1985; Grippo et al., 1989; Nash & Nocedal, 1991; Nash, 2000), which has demonstrated better results over the L-BFGS based optimization for multi-parameter full-waveform inversion (FWI) in exploration seismology (e.g., Métivier et al., 2014, 2017; Pan et al., 2017; Yang et al., 2018; Matharu & Sacchi, 2019). The truncated-Newton FWI, however, is rarely reported based upon the spectral-element and adjoint methods in earthquake seismology due to the computational and storage issues. We leave this for further investigation with the on-the-fly Multi-SEM presented.

An important question remains as to whether the additional costs of the simultaneous computation of the Fréchet and Hessian kernels at twice the computational cost can be offset by more rapid convergence of the non-linear inversion. As high performance computing becomes more accessible and efficient, this may not necessarily be as much of a concern.

In addition to the expressions shown here, the approximate Hessian kernels and the full Hessian kernels can be expressed in different model components as given in Section 2 of the Supporting Information. For the anelastic case, the parsimonious storage method (see Komatitsch et al., 2016) can be used which first performs forward simulation with full attenuation to compute predictions to the seismic measurements and construct the proper adjoint sources. The forward field is stored at selected checkpoints and reconstructed back during the adjoint simulation to calculate the kernels for attenuating medium.

The ideas of Multi-SEM is not limited to the SEM and it can be also implemented in solvers based on other methods such as finite difference. The Multi-SEM so far is designed to compute Fréchet and Hessian kernels for single event. The Hessian kernels for all events can be summed together as that of the misfit Fréchet kernels (Tromp et al., 2005). The Multi-SEM method computes the Fréchet kernels, the approximate and the full Hessian kernels simultaneously on the fly with only about a 2 fold computational cost when compared to the computation for Fréchet kernels alone. The Multi-SEM also supports the simultaneous computation of Fréchet and approximate Hessian kernels as selected function of the Multi-SEM, which is more computationally efficient since only three SEM solvers need to be switched on in the simultaneous backward and adjoint simulation. To further reduce the computational cost for multiple sources, one may use the source encoding techniques (Tromp & Bachmann, 2019).

6 Conclusions

Considering the fast advance in high-performance computing in recent years and the increasing demands in high-resolution multi-parameter imaging, we present the Multi-solver spectral-element and adjoint methods (Multi-SEM) for simultaneously computing the Fréchet and the Hessian kernels on the fly. The simultaneous access to Fréchet and Hessian kernels may potentially provide better images and convergence properties for FWI iterations than those in gradient-only-based FWI. In contrast to the wavefield storage methods that require saving the wavefields for the duration of the simulation, Multi-SEM constructs the Fréchet and Hessian kernels on the fly. The memory requirement for the Multi-SEM is reasonably small since only a single time step of the wavefields and the integrated kernels are kept in memory. The simultaneous computation by the Multi-SEM requires only about a 2-fold computational time when compared to the computation of Fréchet kernels.

The on the fly feature resolves the challenging storage and I/O issues for the Hessian kernel calculation, and makes the use of full Hessian possible for multi-parameter full-waveform inversion (FWI) based upon the spectral-element and adjoint methods. It potentially provides a step forward for improving FWI to better image and understand earth structure, particularly in regions characterised by small scale heterogeneities such as subductions zones.

Acknowledgments

We acknowledge funding from the Natural Environment Research Council (NE/M003507/1 and NE/K010654/1) 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. The Muiti-SEM package will be freely available via `https://zenodo.org/badge/latestdoi/296587481` after the work is accepted and can be requested from the authors after the submission. 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

Afanasiev, M., Boehm, C., Driel, M. v., Krischer, L., Rietmann, M., May, D. A., ... Fichtner, A. (2019). Modular and flexible spectral-element waveform modelling in two and three dimensions. *Geophys. J. Int.*, *216*, 1675–1692.

Chaljub, E., Komatitsch, D., Vilotte, J. P., Capdeville, Y., Valette, B., & Festa, G. (2007). Spectral element analysis in seismology, in: R. s. wu and v. maupin (eds.). *IASPEI Monograph on Advances in Wave Propagation in Heterogeneous Media, Elsevier*, *48*, 365–419.

Chaljub, E., & Valette, B. (2004). Spectral-element modeling of three-dimensional wave propagation in a self-gravitating earth with an arbitrarily stratified outer core. *Geophys. J. Int.*, *158*, 131–141.

Chen, M., Tromp, J., Helmberger, D., & Kanamori, H. (2007). Waveform modeling of the slab beneath japan. *J. geophys. Res.*, *112*, B02305.

Chen, P. (2011). Full-wave seismic data assimilation: theoretical background and recent advances. *Pure and Applied Geophysics*, *168*, 1527–1552.

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

Chen, P., Zhao, L., & Jordan, T. H. (2007). Full 3d tomography for the crustal structure of the los angeles region. *Bulletin of the Seismological Society of America*, *97*, 1094–1120.

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

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

Grippo, L., Lampariello, F., & Lucidi, S. (1989). A truncated newton method with nonmonotone line search for unconstrained optimization. *J. Optim. Theory Appl*, *60*, 401-419.

Komatitsch, D., Ritsema, J., & Tromp, J. (2002c). The spectral-element method, beowulf computing, and global seismology. *Science*, *298*, 1737–1742.

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

Komatitsch, D., & Tromp, J. (2002a). Spectral-element simulations of global seismic wave propagation–i. validation. *Geophys. J. Int.*, *149*, 390–412.

Komatitsch, D., & Tromp, J. (2002b). Spectral-element simulations of global seismic wave propagation–ii. 3-d models, oceans, rotation, and self-gravitation. *Geophys. J. Int.*, *150*, 303–318.

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

Lee, E. J., Chen, P., Jordan, T. H., Maechling, P. B., Denolle, M. A. M., & Beroza, G. C. (2014). Full-3-d tomography for crustal structure in southern california based on the scattering-integral and the adjoint-wavefield methods. *J. Geophys. Res. Solid Earth*, *119*, 6421–6451.

Liu, D. C., & Nocedal, J. (1989). On the limited memory bfgs method for large scale optimization. *Math. Program*, *45*, 504–528.

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

Liu, Q., & Tromp, J. (2006). Finite-frequency kernels based on adjoint methods. *Bulletin of the Seismological Society of America*, *96*, 2383–2397.

Liu, Q., & Tromp, J. (2008). Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods. *Geophys. J. Int.*, *174*, 265–286.

Maday, Y., & Patera, A. T. (1989). Spectral element methods for the incompressible navier-stokes equations. *in State of the Art Survey in Computational Mechanics, eds Noor, A.K. Oden, J.T.*, 71–143.

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

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

Métivier, L., Brossier, R., Operto, S., & Virieux, J. (2017). Full waveform inversion and the truncated newton method. *SIAM Review*, *59*, 153-195.

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.

Nash, S. G., & Nocedal, J. (1991). A numerical study of the limited memory bfgs method and the truncated-newton method for large scale optimization. *SIAM J. Optim.*, *1*, 358-372.

Nocedal, J. (1989). Updating quasi-newton matrices with limited storage. *Mathematics of computation*, *35*(151), 773–782.

Nocedal, J., & Wright, S. J. (1999). Numerical optimisation. *Springer series in oper-*

ations research, Springer-Verlag New York.

Pan, W., Innanen, K. A., & Liao, W. (2017). Hessian-free gauss-newton full-waveform inversion via l-bfgs preconditioned conjugate-gradient algorithm. *Geophysics*, *82(2)*, R49-R64.

Patera, A. T. (1984). A spectral element method for fluid dynamics: laminar flow in a channel expansion. *Journal of Computational Physics*, *54*, 468–488.

Peter, D., Komatitsch, D., Luo, Y., Martin, R., Goff, N. L., Casarotti, E., ... Tromp, J. (2011). Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes. *Geophys. J. Int.*, *186*, 721–739.

Talagrand, O., & Courtier, P. (1987). Variational assimilation of meteorological observations with the adjoint vorticity equation. i: Theory. *Q. J. R. Meteorol. Soc.*, *113*, 1311–1328.

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

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

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

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

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

Tromp, J., Komatitsch, D., & Liu, Q. (2008). Spectral-element and adjoint methods in seismology. *Communications in Computational Physics*, *3*, 1–32.

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

Yang, P., Brossier, R., Métivier, L., Virieux, J., & Zhou, W. (2018). A time-domain preconditioned truncated newton approach to visco-acoustic multiparameter full waveform inversion. *SIAM Journal on Scientific Computing*, *40*, B1101-B1130.

Zou, X., Navon, I. M., Berger, M., Phua, K. H., Schlick, T., & Dimet, F. X. L. (1993). Numerical experience with limited-memory quasi-newton and truncated newton methods. *SIAM J. Optim.*, *3*, 582–608.

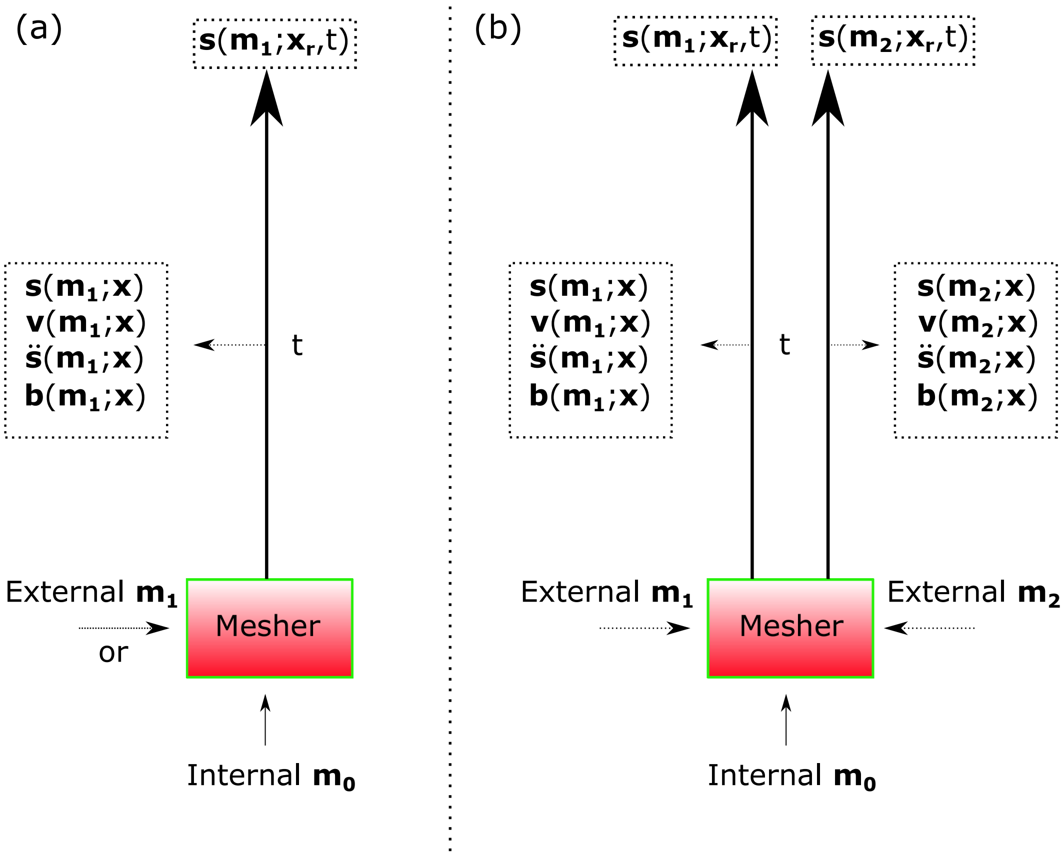

Figure 1. Sketch illustrating the workflow of forward simulation for Conventional SEM vs. Multi-SEM. (a) In Conventional SEM forward simulation, a single model is used and it is set either by the internal mesher (e.g., $\mathbf{m}_0$) or importing from external file ($\mathbf{m}_1$) after the mesher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

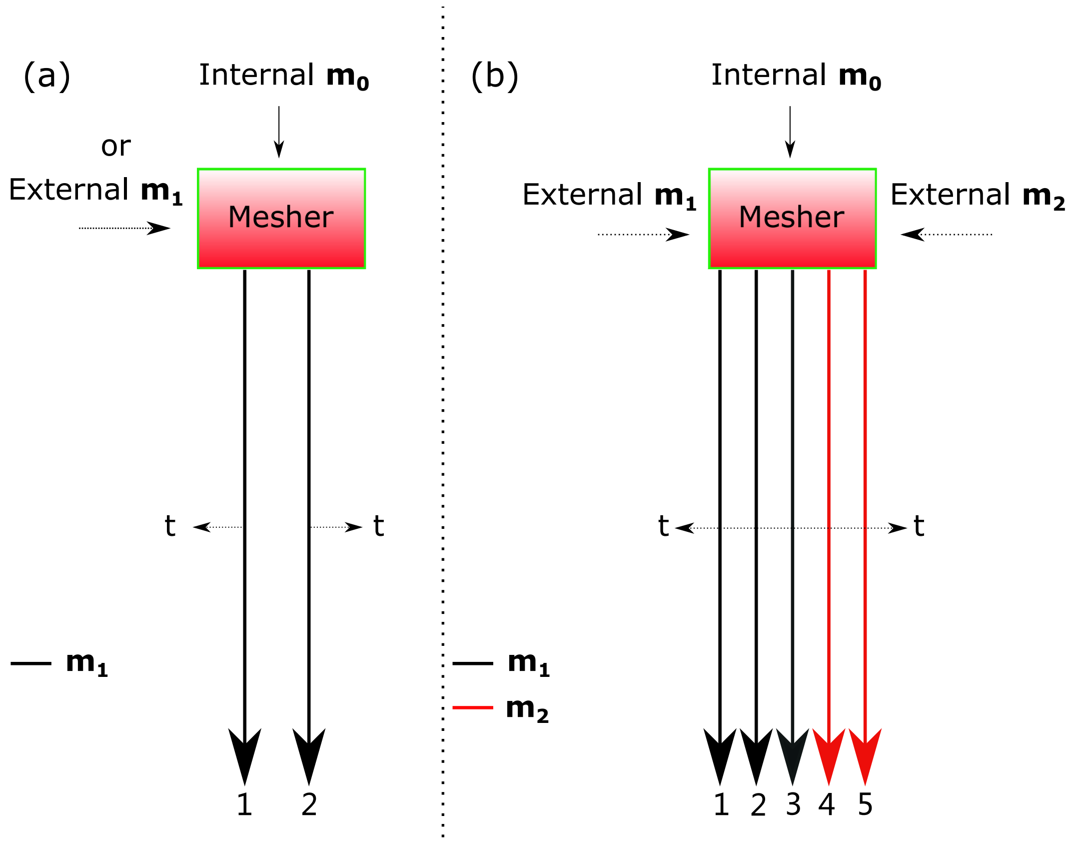

Figure 2. Sketch illustrating the workflows for the simultaneous backward and adjoint simulations for Conventional SEM vs. Multi-SEM. (a) In the simultaneous backward and adjoint simulation of the Conventional SEM, a single model is used. Each arrow represents one solver engine with Arrow 1 indicating the backward simulation (i.e. the reconstruction of the forward field) and Arrow 2 indicating the adjoint simulation which is started from the time-reversed adjoint sources at the receivers. The Fréchet kernel contributions of each time step or an incremental time step are calculated on the fly. (b) In the simultaneous backward and adjoint simulation of the Multi-SEM, Arrows 1, 2, and 3 indicate the solver engines for model $\mathbf{m}_1$, where Arrow 1 and 2 performs the same as in (a) and Arrow 3 performs the same as Arrow 2 except with the perturbation of the adjoint source is taken into account. The red Arrows 4 and 5 indicate the computation of the backward and adjoint fields for the perturbed model $\mathbf{m}_2$. The calculations of Fréchet kernels (by Arrows 1 and 2), approximate Hessian kernels (by Arrows 1, 2, and 3), and the full Hessian kernels (by Arrows 1, 2, 3, 4, and 5) are simultaneously performed on the fly since the required wavefields are computed for each time step. Some solvers can be switched off for computational efficiency if necessary for instance in the computation of approximate Hessian kernels. The Multi-SEM reduces to Conventional SEM when switched off solvers indicated by Arrows 3, 4, and 5.

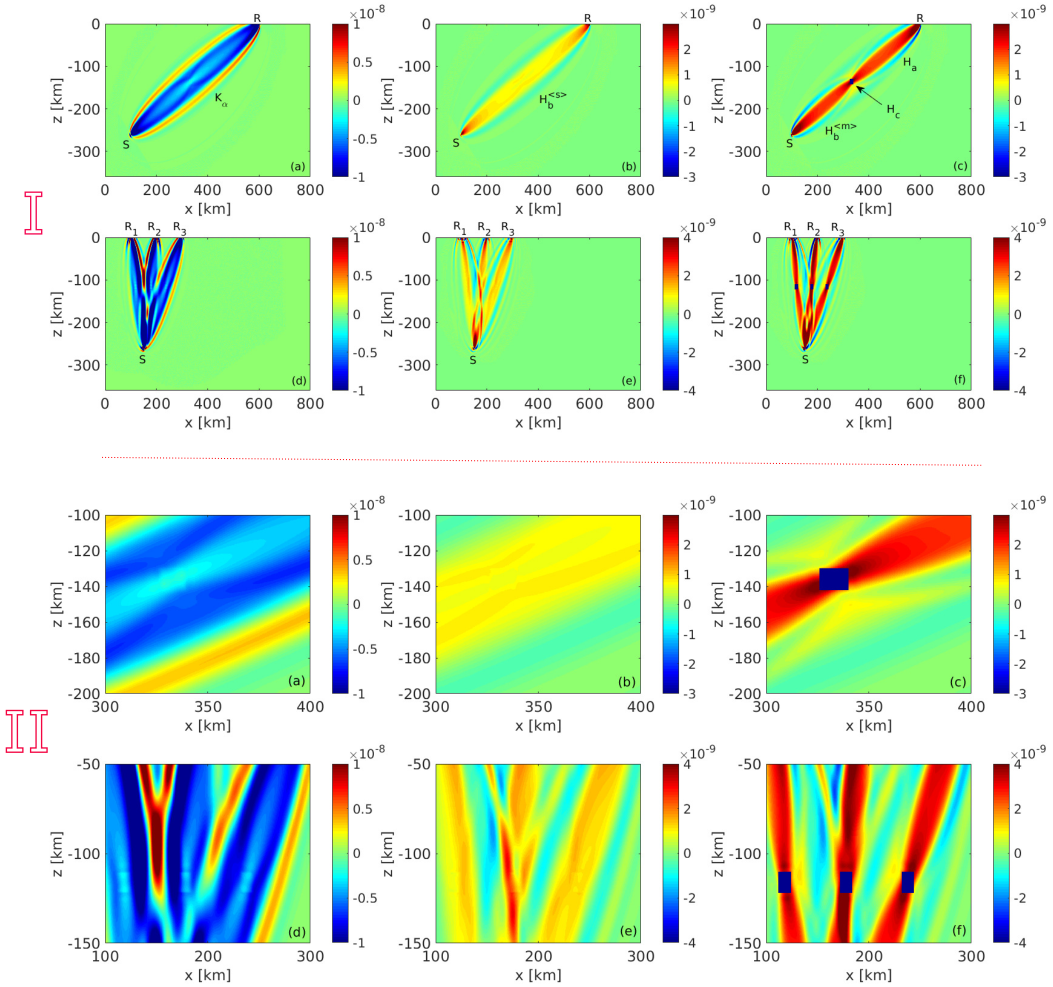

Figure 3. Part I: Fréchet and Hessian kernels computed for *Model 2* (top row) and *Model 3* (bottom row) as discussed in section 4. In the top row we show (a) the Fréhcet kernel K_α, (b) the approximate Hessian kernels $\mathrm{H}_b^{\langle s\rangle}$, and (c) the full Hessian kernels for the single source single station case with a single scattering object, where the full Hessian kernels is a summation of H_a, $\mathrm{H}_b^{\langle s\rangle}$, $\mathrm{H}_b^{\langle m\rangle}$ and H_c. The H_c is restricted to the perturbation indicated by the black box in (c) dictated by its expression, eq. (6). Note that the black box here is the H_c Hessian kernels with a negative value of 10^{-9} scale, not the model perturbation although they are located in the same position. The $\mathrm{H}_b^{\langle s\rangle}$ kernel is mostly invisible in (c) except those around the black box due to its relative small amplitude. The H_a and $\mathrm{H}_b^{\langle m\rangle}$ are separated by the black box. 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 correspondingly shown in Part II. Significant differences are observed between the approximate and full Hessian kernels.