Geophys. J. Int. (0000) **000**, 000–000

A complexity-driven framework for waveform tomography with discrete adjoints

Alexandre Szenicer[1], Kuangdai Leng[1], Tarje Nissen-Meyer[1]

[1] Department of Earth Sciences, University of Oxford, South Parks Road, Oxford OX1 3AN, United Kingdom.

27 July 2020

SUMMARY

We develop a new approach for computing Fréchet sensitivity kernels in full waveform inversion by using the discrete adjoint approach in addition to the widely used continuous adjoint approach for seismic waveform inversion. This method is particularly well suited for the forward solver AxiSEM3D, a combination of the spectral-element method (SEM) and a Fourier pseudo-spectral method, which allows for a sparse azimuthal wavefield parametrization adaptive to wavefield complexity, leading to lower computational costs and better frequency scaling than conventional 3D solvers. We implement the continuous adjoint method to serve as a benchmark, additionally allowing for simulating off-axis sources in axisymmetric or 3D models. The kernels generated by both methods are compared to each other, and benchmarked against theoretical predictions based on linearized Born theory, providing an excellent fit to this independent reference solution. Our verification benchmarks show that the discrete adjoint method can produce exact kernels, largely identical to continuous kernels. While using the continuous adjoint method we lose the computational advantage and fall back on a full-3D frequency scaling, using the discrete adjoint retains the speedup offered by AxiSEM3D. We also discuss the creation of a data-coverage based mesh to run the simulations on during the inversion process, which would allow to exploit the flexibility of the Fourier parametrization and thus the speedup offered by our method.

Keywords: Inverse theory, Computational seismology, Seismic tomography, Theoretical seismology.

1 INTRODUCTION

In the beginning of the twentieth century, observations of seismic signals from teleseismic and regional earthquakes have provided a first glimpse into the structure of our planet (Oldham 1906), with fundamental discoveries such as the Moho (Mohorovičić 1909), and the outer and inner cores (Jeffreys 1926; Lehmann 1936). These early models are spherically symmetric with only radial variations in the structure of the planet, and yet account for the bulk of the observed seismic phases. Explaining the remaining discrepancies in seismic data by introducing 3D heterogeneities is the purpose of global seismic tomography. It might at first seem inconsequential to try and model the details of the internal structure of the planet in order to explain some relatively minor discrepancies in arrival times of seismic phases, and yet these 3D structures are a cornerstone input for geological and geophysical studies that aim to explain the creation, history, and dynamics of our planet; such as mantle convection (van der Hilst et al. 1997; Ritsema and van Heijst 2000), uprising mantle plumes (Bijwaard and Spakman 1999; Montelli et al. 2004b; Zhao 2004; French and Romanowicz 2015), continental evolution (Ritsema et al. 1999), reconstruction of tectonic plate configurations (Van Der Voo et al. 1999a; van der Meer et al. 2010; Sigloch and MihalnyuK 2013) and thermal/compositional components of structural heterogeneities (Schubert et al. 2009a,b).

The development of global seismic tomography started in the 1970s with methods based on arrival times of different seismic phases modelled with ray theory (Dziewonski et al. 1977), and opened doors for a plethora of global P and S wave velocity models of the Earth (van der Hilst et al. 1997; Grand et al. 1997; Kustowski et al. 2008; Ritsema et al. 2011; Schaeffer and Lebedev 2013; Auer et al. 2014; Hosseini et al. 2020). The theoretical framework used for these models assumes that the traveltimes of seismic waves only depend on the structure seen by an infinitely thin geometric ray path connecting the source and receiver. There are considerable benefits to using these methods, such as availability of large datasets of hand-picked arrival times and earthquake locations from organisations such as the International Seismological Centre (Willeman and StorchaK 2001), as well as computational efficiency. However, the main limitation is that they are only valid when the wavelength of waves used for the measurements is very small compared to the size of structural heterogeneities, and for those wave types which can be represented by ray theory, for instance neglecting diffracted (Hosseini et al. 2015) and triplicated waves (Stähler et al. 2013).

Global seismic tomography transitioned to more physically accurate, yet computationally demanding methods such as finite frequency theory (Dahlen et al. 2000; Sigloch and Nolet 2006; Nissen-Meyer et al. 2007), leading to tomography efforts whereby kernels are calculated either by ray theory (Montelli et al. 2004; Hosseini et al. 2020), or by wavefields in spherically symmetric structures (Colombi et al. 2014). Finally, full waveform inversion (FWI) based on adjoint methods (Fichtner et al. 2009; Bozdag et al. 2016) provides a framework for iteratively improv-

ing 3D models with gradient methods, even for mildly non-linear regimes. All these methods aim to resolve several limitations from ray-theoretical approaches by incorporating information from the whole frequency-dependent waveform. This allows to obtain advantages such as quantifying the quality of observations, for instance extracting new observables from the seismograms such as their frequency-dependent amplitudes (Sigloch et al. 2008), and correctly modelling multiscale effects from structural heterogeneities that are unaccounted for in ray-based methods, such as diffraction and wavefront healing. Finite frequency tomography assumes that the perturbation of the measurement is a linear function of a perturbation in the model parameters, and these two perturbations are connected by a so-called sensitivity kernel. Though such linearization brings many benefits such as a drastically lower computational cost and the ability to include large waveform datasets, it inevitably limits the capability of resolving structures with a higher perturbation strength or a finer scale (Tromp et al. 2008). Similarly to finite-frequency tomography, FWI compares synthetic seismograms to observed ones through a misfit function, and the objective is to minimize that function. One of the main differences from finite-frequency tomography is that in FWI the objective function is minimized with a gradient-based iterative procedure. In this way, the non-linear relation between observables and model parameters is at least partially accounted for as long as the nonlinearity does not cause cycle skipping leading to local minima in the misfit function. FWI is based on full 3D simulations, and allows for 3D starting models to be then improved iteratively. In certain settings such as exploration geophysics or basin scales, this may be inevitable (Virieux and Operto 2009; Tape et al. 2009).

In the mathematical literature, FWI is called a PDE-constrained optimization problem, where we aim to minimize a function χ (the misfit), subject to constraints imposed by a partial differential equation $\mathcal{L}$ (the wave equation in our case) and associated initial and boundary conditions. To solve this minimization problem, a standard way is to compute $d_{\mathbf{m}}\chi$, the gradient of χ in the model space $\mathbf{m}$. To do this, one possibility is to evaluate $\mathcal{L}$ for each model parameter and compute the gradient by finite differences, which would require solving $\mathcal{L}$ for $O(n_m)$ times, where n_m is the dimension of the parameter vector $\mathbf{m}$. In global scale tomography n_m can be millions. A more elegant (but more importantly, computationally efficient) way is using the continuous adjoint method, through which the gradient $d_{\mathbf{m}}\chi$ (called Fréchet sensitivity kernels in seismic tomography) can be obtained by just two solutions of $\mathcal{L}$. Continuous adjoint methods have been used for quite some time in optimal control theory (Lions 1971) and in computational fluid dynamics (CFD) (Pironneau 1974), but they were first introduced in seismology by Tarantola (1984) for seismic reflection data. While FWI is now well developed (Bozdag et al. 2016), we still lack the computational power to run it at short simulation periods on a global scale (Nissen-Meyer et al. 2014), which would be desirable to obtain a finer picture of the interior of the planet. As a crude example of the challenging computational cost of using 3D solvers, a global scale FWI tomography at 7s ($\approx 10^4$ CPU hours to produce an hour of seismograms (Leng et al. 2016)) with 3000 earthquakes and 30 iterations of gradient descent would have a computational cost on the order of one billion CPU hours. While this could be in reach with the fastest supercomputers on Earth and moving towards exascale (Bozdag et al. 2016), keeping a supercomputer with 10^4 cores busy for more than ten years is still a formidable task and any means of significantly reducing this cost are welcome, especially if such resources are out of reach.

In summary, global seismic tomography can, for the purpose of the argument in this paper, be crudely divided into methods based on (1) ray theory but using large amounts of traveltime data for great data coverage, (2) finite frequency theory acknowledging the finite sensitivity of a wave due to its frequency-dependent character leading to physically more accurate representation of the inverse problem and a wider class of data measurements, and (3) full-waveform inversion allowing to iteratively update 3D models, thereby accounting for mildly nonlinear relations between seismic measurements and underlying structure. Each method has advantages and disadvantages, but the continually increasing computational efficiency in hardware adhering to Moore's Law increases the interest in computationally more intensive methods.

In this paper, we propose a framework for the computation of 3D Fréchet kernels in any background model based on the discrete adjoint approach, leading to a significant speedup compared to conventional waveform inversion. The discrete adjoint approach has been developed and used extensively in computational fluid dynamics for complex shape optimization problems (Elliott and Peraire 1997) because continuous methods had poor convergence rates (Giles and Duta 2003), but until now it has not become sufficiently relevant in seismology compared to the continuous adjoint approach. This however changed when the azimuthal smoothness of a global wavefield was discovered by Leng et al. (2016), which naturally gave rise to a spectral-element pseudo-spectral method AxiSEM3D (Leng et al. 2016, 2019) adapted to wavefield complexity: As the wavefield is sparsely represented on a Fourier basis, it leads to a drastic speedup in computational cost. Building on Leng et al.'s idea on azimuthal smoothness, this has also been identified by van Driel et al. (2019), who showed that the discrete adjoint approach can obtain exact 2D Fréchet kernels at a much lower computational cost. They achieved such adaptability by an alternative implementation of exploiting azimuthal smoothness by azimuthally adaptive meshing instead of Fourier series, thereby retaining the conventional 3D SEM solver strategy at the expense of having to mesh in 3D. Thrastarson et al. (2020) present an FWI workflow compatible with the discrete adjoint method for 3D global applications, but thus far using 50s discrete adjoint kernels in spherically symmetric models.

Our new approach computes 3D kernels in 3D Earth models and at high frequencies for global scale applications using AxiSEM3D, in which the 3D globe is discretized by a 2D SEM mesh coupled with a Fourier expansion over the azimuthal direction. The azimuthal adaptivity is achieved in AxiSEM3D by varying the number of terms in the Fourier expansion. This discretization is irregular and for the full inversion process can be adapted to the resolvable complexity of the structures based on data coverage, which is what makes the method computationally cheaper. We thus require the resolution of the mesh (and thus model parameters) to be fixed prior to the inversion. Such a priori inversion meshes with spatially varying element sizes based on data coverage are standardly used to reflect that not all regions have the same resolution (Nolet and Montelli 2005; Auer et al. 2014; Zaroli et al. 2015; Hosseini et al. 2020)

2 THEORY

We set ourselves the task of imaging the interior of the Earth, based on a dataset of seismograms recorded after earthquakes on receivers across the globe, $\mathbf{d}(\mathbf{x}_r, t)$, where $\mathbf{x}_r$ are the receiver positions. In full waveform inversion (FWI), we simulate an earthquake by solving the elastodynamic momentum equation in a given

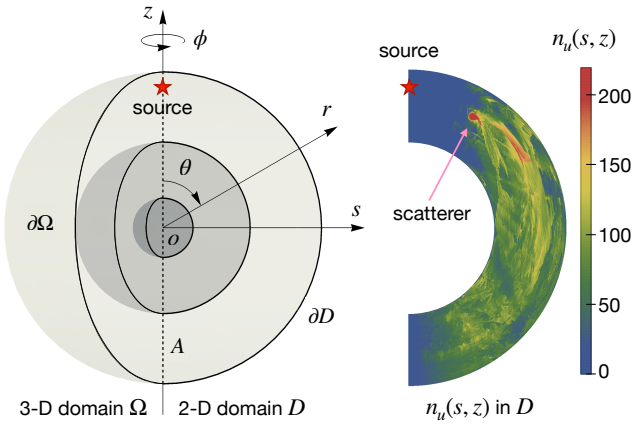

Figure 1. Figure from Leng et al. (2020) illustrating the coordinate system in AxiSEM3D, and the Fourier expansion. The mesh is source-centered so that the source lies on the axis, the slice plane contains an SEM mesh, and the azimuthal direction is expanded in Fourier series. In the source-centred coordinate system (s, ϕ, z), we reduce the 3-D computational domain Ω to a 2-D meridian domain D through a Fourier series characterisation of the ϕ-dimension. At different locations in D, the Fourier series may contain different number of terms, i.e., $n_u = n_u(s, z)$, determined by the azimuthal complexity of the wavefield. In this example, we show $n_u(s, z)$ for a PREM model including a spherical heterogeneity (with a radius of 160 km and a 50% velocity reduction, located at 30°in distance and 1200 km in depth), obtained by wavefield scanning at a 5 s period and with a 1-hr record length.

Earth model parametrised by a vector $\mathbf{m}$ containing various physical properties (P and S wave velocities, density, etc), and record the resulting synthetic seismograms $\mathbf{u}(\mathbf{x}_r, \mathbf{m}, t)$. We then define a misfit $\chi(\mathbf{m})$ that compares synthetic to observed seismograms ($\mathbf{u}$ and $\mathbf{d}$ respectively), e.g. using the Euclidian norm $\chi(\mathbf{m}) = \int_T dt||\mathbf{u}(\mathbf{m}, t) - \mathbf{d}(t)||_2^2$ for waveform tomography, but there is a plethora of other norms one could use. We want the synthetic seismograms to match the data as closely as possible, and hence we want to minimize our misfit by updating the model parameters $\mathbf{m}$ by using gradients of the misfit obtained through the (continuous or discrete) adjoint method. Comprehensive treatments of the adjoint approach in seismology can be found in (Tromp et al. 2005; Fichtner 2011). The computational effort in FWI is largely determined by the efficiency of its underlying wave solver which computes forward and adjoint wavefields.

The crux of adjoint tomography is thus to have a powerful PDE-solver, which one can leverage to efficiently compute gradients of the misfit function. This is particularly true in our case, since we exploit the unique combination of spectral-element/pseudo spectral methods in AxiSEM3D to compute gradients differently than in other seismic tomography software, and thus cut computational cost. Therefore, to better understand our implementation of the adjoint method, we will first give an overview of the functionality of AxiSEM3D.

2.1 Forward problem: AxiSEM3D

We rely on the computational speedup offered by AxiSEM3D for solving the wave equation. This speedup is achieved by a combination of spectral element/pseudo spectral methods. The wave equation is solved on a D-shaped spectral elements (SEM) mesh, and the azimuthal direction is solved in Fourier space, see figure 1.

Consider the wave equation written in condensed form

$$L(\mathbf{u}, \mathbf{m}) = \mathbf{f} \tag{1}$$

where $L()$ is the wave operator, $\mathbf{u}$ is the displacement, $\mathbf{m}$ the model parameters and $\mathbf{f}$ the source term.

AxiSEM3D takes advantage of the observation that in the azimuthal direction, the wavefields propagating through the Earth are relatively smooth, and that complexity is localized. Therefore, the quantities $\mathbf{u}$, $\mathbf{m}$, $\mathbf{f}$ are expanded in Fourier series in the azimuthal direction. Furthermore, the earthquake source is positioned on the axis of the mesh, rendering the expansion of the force term analytical.

$$\mathbf{u}(s, z, \phi) = \sum_{|\alpha| \leq n_u} \mathbf{u}^{\alpha}(s, z) e^{i\alpha\phi} \tag{2}$$

$$\mathbf{m}(s, z, \phi) = \sum_{|\beta| \leq n_m} \mathbf{m}^{\beta}(s, z) e^{i\beta\phi} \tag{3}$$

$$\mathbf{f}(s, z, \phi) = \sum_{|\gamma| \leq n_f} \mathbf{f}^{\gamma}(s, z) e^{i\gamma\phi} \tag{4}$$

One can think of this Fourier expansion as corresponding to the number of grid points in the azimuthal direction at a given point (s, z). If at a point (s, z) we have n_u Fourier coefficients, it is equivalent to $2 \times n_u + 1$ physical grid points in the azimuthal direction. The power of this approach is that usually n_u is much smaller than the number of azimuthal grid points in a standard 3D SEM mesh, and that one can locally choose n_u depending on the complexity of the model. What's more, AxiSEM3D contains a feature to derive a Fourier expansion that is sufficient to accurately model the wavefield in a given Earth model, called wavefield learning (Leng et al. 2019): Starting from a high Fourier expansion, the solver runs a simulation and progressively reduces the Fourier expansion while conserving the accuracy of the wavefield. The resulting Fourier field is called the wisdom Fourier expansion $n_u^{\text{wisdom}}(s, z)$. These aspects deliver a tremendous speedup when compared to classic 3D solvers (Leng et al. 2016).

At a given location, n_u can be bounded from above such that the wavelength resolution in the ϕ-direction matches that in D, namely (Leng et al. 2020)

$$n_u^{\text{sup}}(s, z) = \frac{2\pi s}{\lambda(s, z)} \times N_D \times \frac{1}{2} = \frac{\pi N_D s}{\lambda(s, z)} \tag{5}$$

where $\lambda(s, z)$ denotes the spatial resolution taken as the local wavelength (S-wavelength in solid or P-wavelength in fluid), and thus $2\pi s/\lambda$ the number of wavelengths along the circle generated by revolving the point (s, z), and N_D the number of grid points in D to resolve one wavelength in the s- or z-direction, a constant normally ranging from 6 to 8 in SEM, and the rearward $1/2$ stems from the equivalence of one complex term in the Fourier series and two grid points along the circle. Given the Earth model or $\lambda(s, z)$, n_u^{sup} scales with the distance s to the axis.

As mentioned above, a requirement of this approach is that a source necessarily has to be placed on the axis of the D-shaped mesh. Indeed, because in the azimuthal direction the source term is a Dirac delta, it would theoretically require an infinite Fourier expansion to model accurately. However, in the next section we show how in practice we can accurately simulate off-axis sources by setting the Fourier expansion close to n_u^{sup}, and how we can

exploit that to compute exact sensitivity kernels using the discrete adjoint approach. In summary, AxiSEM3D efficiently delivers accurate 3D wavefields for 3D Earth models with internal and external boundary undulations, including arbitrarily anisotropic (van Driel and Nissen-Meyer 2014; Tesoniero et al. 2020) viscoelastic rheologies.

Now that we understand in more detail the framework of AxiSEM3D, as well as its strengths and limitations, we can look at the adjoint method for computing sensitivity kernels, and how we can leverage AxiSEM3D to exploit smoothness and speed up the computations.

2.2 Inverse problem

We provide a summary of the theory behind seismic waveform inversion, which consists of determining the 3D model parameters based on measured seismograms. As mentioned in the introduction, we will focus on full waveform inversion (FWI), which exploits the entire waveform (or a time window within it), and allows to obtain Fréchet derivatives using the adjoint approach. Adjoint methods have been used for a long time in optimal control theory and in fluid dynamics. They were first introduced for FWI by Tarantola (1984) and thorough theoretical derivations can be found in Tromp et al. (2005) and Fichtner (2011). In particular, we will detail two ways of looking at the calculation of the gradient, namely the continuous adjoint and the discrete adjoint approaches. Note that the approach used will not affect the outcome of the inverse solution as the kernels (gradients) used to update each iteration are identical. However, and this is the crucial point, the discrete adjoint in our AxiSEM3D context provides a much more efficient time to solution as shown further below, especially at higher frequencies and thus opens doors to new parameter regimes such as using more data or higher frequencies.

Let us define a misfit $\chi(\mathbf{u})$ that compares synthetic to observed seismograms ($\mathbf{u}$ and $\mathbf{d}$ respectively) at a number of receiver locations $\mathbf{x}_r$, e.g. using the Euclidian norm $\chi(\mathbf{u}) = \frac{1}{2}\sum_{\mathrm{r}=1}^{N}\int_T dt||\mathbf{u}(\mathbf{x}_r,\mathbf{m},t)-\mathbf{d}(\mathbf{x}_r,t)||^2$, which we want to minimize. The perturbation of the misfit induced by a perturbation in the model parameters is the total derivative:

$$\delta\chi = \frac{d\chi}{d\mathbf{m}}\cdot\delta\mathbf{m} \tag{6}$$

Adjoint methods allow us to obtain gradients $d\chi/d\mathbf{m}$ by solving a second equation

$$L^\dagger(\mathbf{u}^\dagger,\mathbf{m}) = \mathbf{f}^\dagger \tag{7}$$

where $L^\dagger$ is the adjoint wave operator, $\mathbf{u}^\dagger$ the adjoint wavefield and $\mathbf{f}^\dagger = -\partial\chi/\partial\mathbf{u}$ is the adjoint source term (Tromp et al. 2005; Fichtner 2011). The continuous wave operator is self-adjoint, meaning we can solve the adjoint equations with the same solver used in the forward simulation, simply with different initial conditions.

In conventional derivations of kernels in seismology (Tromp et al. 2005; Fichtner 2011), running the adjoint simulation delivers the adjoint wavefield $\mathbf{u}^\dagger$. From there, we can compute the directional derivative (called sensitivity kernel) for various model parameters. For instance with respect to P wave velocity v_P, the sensitivity kernels are (Tromp et al. 2005; Fichtner 2011) :

$$K_{v_P} = 2\rho v_P\int_T \nabla\cdot\mathbf{u}^\dagger(t)\nabla\cdot\mathbf{u}(T-t)dt \tag{8}$$

and for a given model parameter $\mathbf{m}$, we have the misfit variation

$$\frac{d\chi}{d\mathbf{m}}\cdot\delta\mathbf{m} = \int_\Omega K_m(\mathbf{r})\delta\mathbf{m}(\mathbf{r})d\mathbf{r}. \tag{9}$$

This equality provides the essential, locally linearized relation between measurement and model for FWI. The solution to the inverse problem is the final model $\mathbf{m}$ for the case where its corresponding residual misfit $\delta\chi$ is deemed acceptably small. Note that we deliberately keep $\mathbf{m}$ generic; the theory below and our corresponding methodology allow for computing sensitivity kernel for any anisotropic element of the elasticity tensor, as well as attenuation, irrespective of whether the background model used in the forward problem has the same model complexity.

2.3 Continuous adjoint

In the derivation above, which is the common approach in seismology, we can see that we haven't specified any simulation grid: the derivation of the adjoint system is performed on the continuous equations, and we can then discretize and solve our system.

From the previous section, the adjoint equation as derived in the continuous adjoint framework (Tromp et al. 2005; Fichtner 2011) is given by eq. (7). Let us now discretise this equation in space, to obtain the semi-discrete representation of the continuous adjoint equation

$$\ddot{\mathbf{u}}^\dagger = \mathsf{M}^{-1}(\mathbf{m})\mathsf{K}(\mathbf{m})\mathbf{u}^\dagger + \mathsf{M}^{-1}\mathbf{f}^\dagger \tag{10}$$

with initial conditions $\mathbf{u}^\dagger(\mathsf{T}) = \mathbf{0}$ and $\dot{\mathbf{u}}^\dagger(\mathsf{T}) = \mathbf{0}$, and sans-serif font always denoting discretised quantities. The dots denote derivative in time, $\mathbf{u}^\dagger$ is the discretised adjoint displacement wavefield, $\mathbf{m}$ are the discretised model parameters, M is the discretised mass matrix (which is diagonal in AxiSEM3D), K is the discretised stiffness matrix, and $\mathbf{f}^\dagger$ the discretised adjoint source term (which is a derivative of the misfit function). Note that in this approach, we have to choose the discretisation such that the adjoint wavefield $\mathbf{u}^\dagger$ is physical. In the case of AxiSEM3D this means a very high and expensive Fourier expansion to accurately express an off-axis source. This approach of optimize-then-discretize is called the continuous adjoint approach. There is one bottleneck to this approach with AxiSEM3D. Indeed, if we want to compute the adjoint wavefield originating from a number of globally distributed receivers simultaneously as is customary in adjoint-based tomography, we need to place sources off-axis. As we have discussed above, in AxiSEM3D's context, this requires modeling a point force source in Fourier space, which has an infinite Fourier expansion. In practice, this means that we need to use a rather large Fourier field, sufficient to accurately model in azimuth the wavelength of the dominant period of the simulation. We investigate this in the next section, by plotting seismograms for increasingly high Fourier fields, and we show that they converge to a physically correct seismogram.

2.4 Discrete adjoint

An alternative approach to optimize-then-discretize is the discretize-then-optimize paradigm. In the former, we first optimise the continuous system and then discretise the resulting equations, and in the latter we first discretise the wave equation on the forward grid, and then optimise with respect to the discrete model parameters. In this section, we will focus on showing that the two approaches are equivalent, in that they give rise to the same adjoint equations. However, we'll see that the discrete adjoint approach imposes fewer constraints on the discretisation of the adjoint equation, which is very beneficial in AxiSEM3D.

Starting directly from the discrete wave equation with the forward model discretisation and optimising with respect to those discrete model parameters, we can arrive at eq. (10) but without the expensive Fourier expansion for the source term. We demonstrate that we can re-use the same expansion as the forward simulation, which in AxiSEM3D is very advantageous because it exploits the azimuthal smoothness of the wavefield. The adjoint wavefield $\mathbf{u}^\dagger$ will not be physical anymore, but the kernels will still be exact with respect to the model parameters of the forward simulation.

We derive the adjoint equations in the discrete adjoint framework following the Lagrange multipliers approach from (Liu and Tromp 2008), except that they start with the continuous wave equation, and we start with the discretised one. The equivalence of the two approaches may seem evident, but for clarity it shall be useful to re-derive it explicitly. As we will see, we cast the computation of the kernels as a constrained optimisation problem to minimise the misfit function. The adjoint wave equation then arises as a convenient means to remove terms that are intractable to compute, thus rendering the calculation of the kernels much simpler, by two solutions of the wave equation.

Consider again the wave operator $L(\mathbf{u}, \mathbf{m}) = \mathbf{f}$. We can discretise this operator in space on the grid of the forward simulation similar to eq (10), to obtain the semi-discrete ordinary differential equation

$$\ddot{\mathbf{u}} = \mathsf{M}^{-1}(\mathbf{m})\mathsf{K}(\mathbf{m})\mathbf{u} + \mathsf{M}^{-1}(\mathbf{m})\mathbf{f}(\mathbf{m}) \tag{11}$$

with some initial conditions $\mathbf{g}(\mathbf{u}(0), \mathbf{m}) = \mathbf{0}$ and $\mathbf{h}(\mathbf{u}(0), \mathbf{m}) = \mathbf{0}$. The dots denote derivative in time, $\mathbf{u}$ is the discretised displacement wavefield, $\mathbf{m}$ are the discretised model parameters for the forward problem, M is the discretised mass matrix, K is the discretised stiffness matrix, and $\mathbf{f}$ the discretised source term. Note that K is time-independent since attenuation can be incorporated in the Earth model parameters without time dependence, as shown by Fichtner and Van Driel. (2014). Note also that unlike eq. (10), the discretisation is that of the forward problem, so a relatively low Fourier expansion in AxiSEM3D. For convenience of manipulation we will rewrite the semi-discretised wave equation as

$$\ddot{\mathbf{u}} = \mathbf{I}(\mathbf{u}, \mathbf{m}, t) \tag{12}$$

where $\mathbf{I}(\mathbf{u}, \mathbf{m}, t) = \mathsf{M}^{-1}(\mathbf{m})\mathsf{K}(\mathbf{m})\mathbf{u}(t) + \mathsf{M}^{-1}(\mathbf{m})\mathbf{f}(\mathbf{m})$.

As previously stated, the goal of seismic tomography is to minimize a misfit function of some measurement between synthetic and observed seismograms with respect to the model parameters. Such a measurement choice could be the Euclidian norm $\chi(\mathbf{u}) = \frac{1}{2}\sum_{r=1}^{N}\int_0^T ||\mathbf{u}(\mathbf{x}_r, \mathbf{m}, t) - \mathbf{d}(\mathbf{x}_r, t)||^2 dt$ which we write as $\chi(\mathbf{u}) = \int_0^T \Xi(\mathbf{u}, t)dt$, where Ξ is a concatenated, generic notation for any chosen measurement. This minimization is constrained by the semi-discrete wave equation $\ddot{\mathbf{u}} = \mathbf{I}(\mathbf{u}, \mathbf{m}, t)$, and initial conditions $\mathbf{g}(\mathbf{u}(0), \mathbf{m}) = \mathbf{0}$ and $\mathbf{h}(\dot{\mathbf{u}}(0), \mathbf{m}) = \mathbf{0}$. Now we will solve this constrained optimisation problem to re-derive eq. (10) but starting from the semi-discrete wave equation and thus obeying the forward discretisation, by using Lagrange multipliers. To do so, we start with the Lagrangian to describe the dynamics of this system as

$$\mathcal{L} = \int_0^T [\Xi(\mathbf{u}, t) + \boldsymbol{\lambda}^T(t)\cdot(\ddot{\mathbf{u}} - \mathbf{I}(\mathbf{u}, \mathbf{m}, t))]dt + \boldsymbol{\mu}^T\cdot\mathbf{g}(\mathbf{u}(0), \mathbf{m}) + \boldsymbol{\eta}^T\cdot\mathbf{h}(\dot{\mathbf{u}}(0), \mathbf{m}), \tag{13}$$

where $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$ are the vector Lagrange multipliers. Note that because the constraints $\ddot{\mathbf{u}} = \mathbf{I}(\mathbf{u}, \mathbf{m}, t)$, $\mathbf{g}(\mathbf{u}(0), \mathbf{m}) = \mathbf{0}$, $\mathbf{h}(\dot{\mathbf{u}}(0), \mathbf{m}) = \mathbf{0}$ are (by construction) always satisfied, we have $d_\mathbf{m}\mathcal{L} = d_\mathbf{m}\chi$, and we are also free to set the multipliers $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$ to any value. This is crucial, as it will allow us to set the multipliers to a value that conveniently makes some incomputable terms vanish in the expression of the kernel, and it is the beauty of the adjoint method. We can now derive the gradient $d_\mathbf{m}\mathcal{L}$. As stated above, because the constraints are satisfied by construction and $d_\mathbf{m}\mathcal{L} = d_\mathbf{m}\chi$, we're actually computing the gradient $d_\mathbf{m}\chi$. Note again that in this section, all gradients and partial derivatives are discrete operators, with respect to the forward discretisation.

The Lagrangian variation with respect to discrete model parameters becomes

$$\begin{aligned} d_\mathbf{m}\mathcal{L} = \int_0^T & [\partial_\mathbf{u}\Xi(\mathbf{u}, t)d_\mathbf{m}\mathbf{u}(t) + \partial_\mathbf{m}\Xi(\mathbf{u}, t) \\ & + \boldsymbol{\lambda}^T(t)\cdot(d_\mathbf{m}\ddot{\mathbf{u}} - \partial_\mathbf{u}\mathbf{I}(\mathbf{u}, \mathbf{m}, t)d_\mathbf{m}\mathbf{u}(t) - \partial_\mathbf{m}\mathbf{I}(\mathbf{u}, \mathbf{m}, t))]dt \\ & + \boldsymbol{\mu}^T\cdot(\partial_{\mathbf{u}(0)}\mathbf{g}(\mathbf{u}(0), \mathbf{m})d_\mathbf{m}\mathbf{u}(0) + \partial_\mathbf{m}\mathbf{g}(\mathbf{u}(0), \mathbf{m}) \\ & + \boldsymbol{\eta}^T\cdot(\partial_{\dot{\mathbf{u}}(0)}\mathbf{h}(\dot{\mathbf{u}}(0), \mathbf{m})d_\mathbf{m}\dot{\mathbf{u}}(0) + \partial_\mathbf{m}\mathbf{h}(\dot{\mathbf{u}}(0), \mathbf{m}). \end{aligned} \tag{14}$$

Integrating the term $\int_0^T \boldsymbol{\lambda}^T(t)d_\mathbf{m}\ddot{\mathbf{u}}(t)dt$ twice by parts yields

$$\begin{aligned} \int_0^T \boldsymbol{\lambda}^T(t)d_\mathbf{m}\ddot{\mathbf{u}}(t)dt = & [\boldsymbol{\lambda}^T(t)d_\mathbf{m}\dot{\mathbf{u}}(t)]_0^T - [\dot{\boldsymbol{\lambda}}^T(t)d_\mathbf{m}\mathbf{u}(t)]_0^T \\ & + \int_0^T \ddot{\boldsymbol{\lambda}}^T(t)d_\mathbf{m}\mathbf{u}(t)dt \end{aligned} \tag{15}$$

Reinjecting this into eq. (14) and grouping terms by wavefield derivatives with respect to model parameters, we obtain

$$\begin{aligned} d_\mathbf{m}\mathcal{L} = \int_0^T & [(\partial_\mathbf{u}\Xi(\mathbf{u}, t) - \boldsymbol{\lambda}^T(t)\partial_\mathbf{u}\mathbf{I}(\mathbf{u}, \mathbf{m}, t) \\ & + \ddot{\boldsymbol{\lambda}}^T(t))d_\mathbf{m}\mathbf{u}(t) + \partial_\mathbf{m}\Xi(\mathbf{u}, t) - \boldsymbol{\lambda}^T(t)\partial_\mathbf{m}\mathbf{I}(\mathbf{u}, \mathbf{m}, t]dt \\ & + (\boldsymbol{\mu}^T\partial_{\mathbf{u}(0)}\mathbf{g}(\mathbf{u}(0), \mathbf{m}) + \dot{\boldsymbol{\lambda}}^T(0))d_\mathbf{m}\mathbf{u}(0) \\ & + (\boldsymbol{\eta}^T\partial_{\mathbf{u}(0)}\mathbf{h}(\dot{\mathbf{u}}(0), \mathbf{m}) - \boldsymbol{\lambda}^T(0))d_\mathbf{m}\dot{\mathbf{u}}(0) \\ & - \dot{\boldsymbol{\lambda}}^T(T)d_\mathbf{m}\mathbf{u}(T) \\ & + \boldsymbol{\lambda}^T(T)d_\mathbf{m}\dot{\mathbf{u}}(T) \\ & + \boldsymbol{\mu}^T\partial_\mathbf{m}\mathbf{g}(\mathbf{u}(0), \mathbf{m}) + \boldsymbol{\eta}^T\partial_\mathbf{m}\mathbf{h}(\dot{\mathbf{u}}(0), \mathbf{m}). \end{aligned} \tag{16}$$

Now we can freely set the Lagrange multiplier values to what we deem convenient because the constraints are satisfied by construction. What's more, all the terms that are derivatives of the wavefield with respect to model parameters such as $\partial_\mathbf{m}\mathbf{u}$ are not straightforward to compute because they would require many runs to do finite-differences, which is intractable with many model parameters. Therefore, we can set $\boldsymbol{\lambda}^T(T) = \mathbf{0}$, and $\dot{\boldsymbol{\lambda}}^T(T) = \mathbf{0}$, and $\boldsymbol{\eta}^T = \boldsymbol{\lambda}^T(0))\partial_{\mathbf{u}(0)}\mathbf{h}(\dot{\mathbf{u}}(0), \mathbf{m})^{-1}$, and $\boldsymbol{\mu}^T = -\dot{\boldsymbol{\lambda}}^T(0)\partial_{\mathbf{u}(0)}\mathbf{g}(\mathbf{u}(0), \mathbf{m})^{-1}$, and finally $\partial_\mathbf{u}\Xi(\mathbf{u}, t) -$

$\boldsymbol{\lambda}^T(t)\partial_{\mathbf{u}}\mathbf{l}(\mathbf{u},\mathbf{m},t)+\ddot{\boldsymbol{\lambda}}^T(t)=\mathbf{0}$, and this removes all the intractable terms. The last identity is nothing but the adjoint equation

$$\ddot{\boldsymbol{\lambda}}(t)=\partial_{\mathbf{u}}\mathbf{l}(\mathbf{u},\mathbf{m},t)^T\boldsymbol{\lambda}(t)-\partial_{\mathbf{u}}\Xi(\mathbf{u},t)^T \tag{17}$$

with initial conditions $\boldsymbol{\lambda}(T)=\mathbf{0}$, and $\dot{\boldsymbol{\lambda}}(T)=\mathbf{0}$. We can further simplify it by expanding $\mathbf{l}(\mathbf{u},\mathbf{m},t)$. Because the adjoint equation is solved backwards in time due to initial conditions, we rewrite it in terms of $T-t$. Finally, we rename $\boldsymbol{\lambda}$ as $\mathbf{u}^\dagger$ and apply Leibniz's rule to replace $\partial_{\mathbf{u}}\Xi$ with $\partial_{\mathbf{u}}\chi$, to obtain

$$\ddot{\mathbf{u}}^\dagger(T-t)=(\mathsf{M}^{-1}(\mathbf{m})\mathsf{K}(\mathbf{m}))^T\mathbf{u}^\dagger(T-t)-\partial_{\mathbf{u}}\chi(\mathbf{u},T-t)^T \tag{18}$$

with initial conditions $\mathbf{u}^\dagger(T)=\mathbf{0}$, and $\dot{\mathbf{u}}^\dagger(T)=\mathbf{0}$. We see that eq. (18) only differs from the initial semi-discrete wave equation eq. (12) by the adjoint operator, and the source term which is a partial derivative of the misfit function of choice. It also only differs from eq. (10) by the adjoint operator. A last step is then to verify whether we can use the same solver to solve the adjoint wave equation, i.e. whether the discrete operator $(\mathsf{M}^{-1}\mathsf{K})$ is self-adjoint. First, from Leng et al. (2016) we know that M is diagonal in AxiSEM3D, so we only need to determine whether K is self adjoint. Furthermore, from Leng et al. (2016), the computation of the stiffness matrix K in AxiSEM3D involves five operations: applying a differential operator D_1, a Fourier transform operator P_1 which is unitary and linear, multiplying with symmetric 3D elasticity matrix C, then applying the inverse Fourier transform $\mathsf{P}_2=\mathsf{P}_1^{-1}=\mathsf{P}_1^T$, and then applying the same differential operator $\mathsf{D}_2=\mathsf{D}_1^T$. Hence $\mathsf{K}=\mathsf{D}_2\mathsf{P}_2\mathsf{C}\mathsf{P}_1\mathsf{D}_1$ is self adjoint.

We now see that eq. (18) is the same as eq. (10), except that it is discretised identically to the forward simulation (avoiding a high Fourier expansion for off-axis sources), thus rendering it much cheaper to compute with AxiSEM3D than the continuous adjoint. As mentioned previously, this means that the wavefield $\mathbf{u}^\dagger$ does not need to be physical in the discrete adjoint approach, but the gradients are exact with respect to the discretised model parameters. We can illustrate the difference in the case of AxiSEM3D: the wave equation is discretized on a 2D SEM mesh and in Fourier series in the azimuth direction. It is important to notice that the Fourier expansion is finite and dictated by the forward simulation with wavefield learning. We thus get the following accurate representation of the wavefield:

$$\mathbf{u}(s(\xi,\eta),z(\xi,\eta),\phi,t)=\sum_{|\alpha|\leq n_u}\sum_{p,q}\mathbf{u}_{pq}^{\alpha}(t)l_p^N(\xi,\eta)l_q^N(\xi,\eta)e^{i\alpha\phi}, \tag{19}$$

where n_u is the local Fourier expansion order, $\mathbf{u}_{pq}^{\alpha}$ contains the Fourier coefficient α at the node p, q, and $l_p^N(\xi,\eta)$ is the pth Lagrange interpolating function of polynomial order N. We can then start from the discrete adjoint equation source term in eq. (18), and use our AxiSEM3D notation:

$$-\frac{\partial}{\partial\mathbf{u}_{pq}^{\alpha}}\chi(\mathbf{u}_{pq}^{\alpha})=-\frac{\partial}{\partial\mathbf{u}}\chi(\mathbf{u})\frac{\partial\mathbf{u}}{\partial\mathbf{u}_{pq}^{\alpha}}=-\frac{\partial}{\partial\mathbf{u}}\chi(\mathbf{u})l_p^Nl_q^Ne^{i\alpha\phi} \tag{20}$$

where the result is the continuous adjoint source term in eq. (7), interpolated onto the forward simulation mesh.

The implementation of the discrete adjoint source is therefore equivalent to that of the discretised continuous adjoint source, with the difference that the Fourier expansion is imposed by the discretization of the forward simulation, and thus does not need to model a physically accurate point force source by using a very high and expensive Fourier expansion.

Finally, by removing all the vanishing terms due to the appropriately chosen Lagrange multipliers, as well as the vanishing partial derivatives of the initial conditions (usually just the earthquake, so independent of $\mathbf{m}$), we obtain the expression for the gradients with respect to model parameters

$$d_{\mathbf{m}}\chi=d_{\mathbf{m}}\mathcal{L}=-\int_0^T\boldsymbol{\lambda}^T(T-t)\partial_{\mathbf{m}}(\mathsf{M}^{-1}(\mathbf{m})\mathsf{K}(\mathbf{m})\mathbf{u}(t)+\mathsf{M}^{-1}(\mathbf{m})\mathbf{f}(\mathbf{m}))dt \tag{21}$$

We now have the well known recipe to compute the gradients using the adjoint method:

1) Integrate $\ddot{\mathbf{u}}=\mathsf{M}^{-1}(\mathbf{m})\mathsf{K}(\mathbf{m})\mathbf{u}+\mathsf{M}^{-1}(\mathbf{m})\mathbf{f}(\mathbf{m})$ from 0 to T, with initial conditions $\mathbf{g}(\mathbf{u}(0),\mathbf{m})=\mathbf{0}$ and $\mathbf{h}(\mathbf{u}(0),\mathbf{m})=\mathbf{0}$.
2) Integrate $\ddot{\mathbf{u}}^\dagger(T-t)=\mathsf{M}^{-1}(\mathbf{m})\mathsf{K}(\mathbf{m})\mathbf{u}^\dagger(T-t)-\partial_{\mathbf{u}}\chi(\mathbf{u},T-t)$ from T to 0 with initial conditions $\mathbf{u}^\dagger(T)=\mathbf{0}$, and $\dot{\mathbf{u}}^\dagger(T)=\mathbf{0}$.
3) Compute the gradients with $d_{\mathbf{m}}\chi=-\int_0^T\boldsymbol{\lambda}^T(T-t)\partial_{\mathbf{m}}(\mathsf{M}^{-1}(\mathbf{m})\mathsf{K}(\mathbf{m})\mathbf{u}(t)+\mathsf{M}^{-1}(\mathbf{m})\mathbf{f}(\mathbf{m}))dt$

As we have seen for the adjoint source in eq. (20), we do not need to bother re-deriving the expressions of the gradient for any specific parameter, which involves a lot of tedious algebra, as this will be identical to previous derivations obtained in the continuous adjoint framework. More importantly, we have established the equivalence of the continuous and discrete adjoint framework, i.e. the operations of adjoint and discretisation commute in our system. Therefore, we can appropriately discretise the results obtained in the continuous adjoint framework of (Tromp et al. 2005; Fichtner 2011).

Note again that in the discrete adjoint case all the quantities are discretised on the grid of the forward simulation. For methods where the mesh is source- and complexity-independent, whether to use the discrete or continuous adjoint formulation does not matter. However in AxiSEM3D this is a crucial point, as it allows us to drastically cut down the Fourier expansion of the adjoint simulation, and thus the computational time. Very importantly, the discrete adjoint formulation highlights the fact that one can only update the discrete model parameters used in the mesh of the forward simulation. This is of course no different in the continuous adjoint case, but provides a direct and natural connection between the discretisation of the forward, adjoint and inverse problems: For instance, in the azimuthal direction we can design meshes that are much coarser than a standard 3D SEM mesh, reflecting our limited data coverage and resolution. Such meshes of spatially varying element size are already used in tomographic inversions (Nolet and Montelli 2005; Hosseini et al. 2020).

3 CONTINUOUS ADJOINT KERNELS

In this section we demonstrate how we obtain exact kernels in the continuous adjoint framework. As mentioned in the previous section, we need to model a point force source off-axis (at the receiver position), which in AxiSEM3D theoretically means an infinite Fourier expansion. First, we show that in practice, we just need a Fourier expansion that is high enough to correctly approximate the dominant wavelength in the simulation.

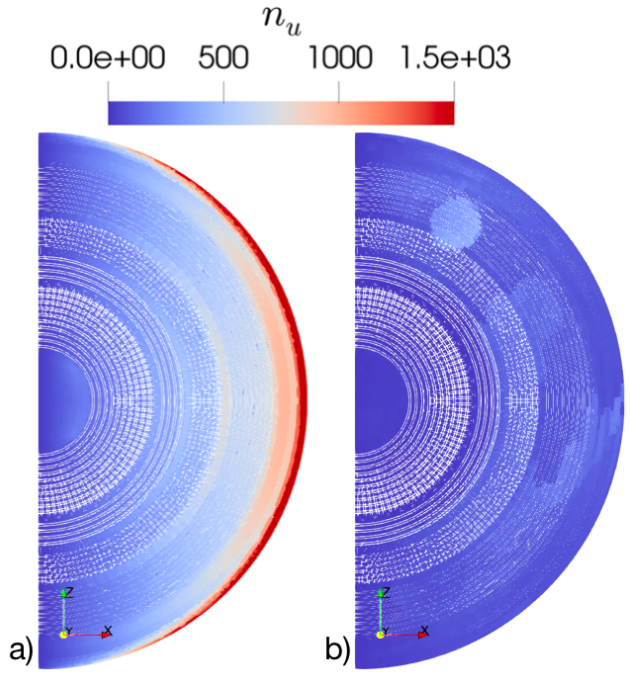

Figure 2. a): Constant Fourier expansion n_u^{const}. One can increase this constant Fourier expansion until arriving to the upper limit n_u^{sup}. b): Fourier expansion learned n_u^{wisdom} from a forward simulation on PREM (Dziewonski et al. 1981) containing a 3D bubble. We can see that the expansion needs to be higher where the 3D bubble is located to capture the wavefield complexity arising from effects such as scattering. However, overall the wisdom Fourier expansion is much lower than the constant Fourier expansion.

3.1 Physical adjoint wavefields

We model our source to be a nascent Dirac in azimuth, approximated by a Gaussian.

$$\delta_{n_u}(\phi) = \frac{2n_u}{\sqrt{\pi}} e^{-(2n_u\phi)^2} \tag{22}$$

Where n_u is the local order of Fourier expansion, and ϕ is the azimuth. We postulate that given a high enough Fourier expansion, this representation accurately models a point force. We use a constant Fourier expansion n_u^{const} pictured in figure 2 a). As we increase n_u^{const} we get to n_u^{sup}, which corresponds to the case where the azimuthal direction is sampled as densely as the D-shaped SEM mesh, and we do not exploit any structural smoothness, similarly to a traditional 3D wave equation solver.

We then put the off axis source on the equator. This is the worst case scenario where the azimuthal extent of the wavelength is the largest, thus requiring the largest Fourier expansion to model accurately. Finally, we place receivers on several locations around the Earth. We record seismograms for increasing spatially constant Fourier expansions n_u^{const}, shown in figure 3. Simulations are run at 30s dominant period, both in a spherically symmetric Earth model (PREM (Dziewonski et al. 1981)), and in a 3D model (S40RTS (Ritsema et al. 2011)), to underline the fact that the computations work in fully 3D models.

Looking at the resulting seismograms and at the mean absolute errors errors in 1, we can see that they quickly converge to the reference seismograms obtained with a source positioned on the axis, as n_u^{const} gets closer to n_u^{sup}, showing that we accurately compute

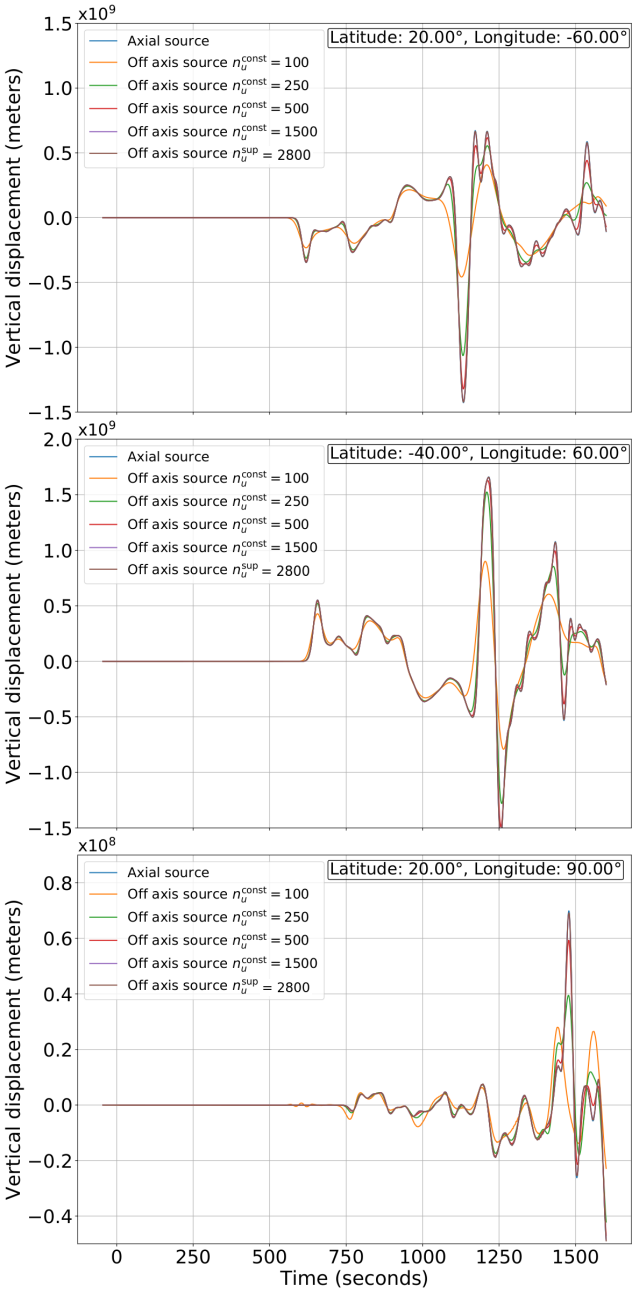

Figure 3. Converging seismograms for increasing values of the n_u^{const} field in s40rts, with a point force at the equator at 0 degrees azimuth. When the source is positioned on the axis of the SEM mesh, there is an analytical solution and the seismograms are exact, it is thus our reference. When the source is positioned off-axis, the seismograms converge to the exact ones as we increase n_u^{const} and get closer to n_u^{sup}. Amplitudes are scaled up to avoid floating point errors.

wavefields from an off-axis source by simply having a high enough Fourier expansion. This result intuitively makes sense because as we increase the Fourier expansion, we get closer and closer to a fully 3D mesh as used in other SEM solvers such as SPECFEM, except that our azimuthal dimension is not discretized with spectral elements.

n_u field \ (Lat, Lon)	$(20°, -60°)$	$(-40°, 60°)$	$(20°, 90°)$
$n_u^{\text{const}} = 100$	59.69	43.39	38.97
$n_u^{\text{const}} = 250$	30.65	18.02	14.05
$n_u^{\text{const}} = 500$	12.27	5.98	4.41
$n_u^{\text{const}} = 1500$	1.68	0.74	0.54
$n_u^{\text{sup}} = 2800$	0.51	0.23	0.17

Table 1. Normalised mean absolute error between offaxis seismogram and axial seismogram (in %) $\int_t(|s^{\text{offaxis}} - s^{\text{axial}}|)/\int_t |s^{\text{axial}}|$ with increasing n_u field for the stations displayed in figure 3.

3.2 Continuous kernels benchmark

Now that we have established the ability of AxiSEM3D to accurately model off axis sources, we can compute kernels in the continuous adjoint framework. In the following figures, our source is located at the surface on the geographic North pole. We inject an M_{rr} only moment tensor, and our simulations are at a dominant period of 30s. The receiver is placed at an epicentral distance of 60 degrees. The computation is performed at n_u^{sup}, ensuring we propagate physical adjoint wavefields. We run tests in PREM and in S40RTS. For the benchmarks we do not add a 3D crust because it makes simulations quite a bit more expensive, but in section 4.2 we run a 15s simulation to produce a collection of kernels for a given earthquake in S40RTS + crust1.0 (Laske et al. 2013).

In figure 4 a) we can see a cut through the kernel obtained using the continuous adjoint approach for PREM, and in 4 b) for S40RTS. The fact that both kernels are essentially identical (as previously noted by Liu and Tromp (2008)) raises interesting questions about how sensitive to 3D structure the wavefield really is.

In traveltime tomography, we can examine the cross-correlation traveltime update δT_{xcorr} caused by a small perturbation to the model parameters (e.g. for a P wave velocity pertubation $\delta \boldsymbol{v}_P$), computed by cross correlation between synthetic seismograms from the initial model, and perturbed seismograms from the model + perturbation (Dahlen et al. 2000). To benchmark our kernels, we can compare δT_{xcorr} to the traveltime perturbation predicted by the kernels

$$\delta T_{\text{ker}} = \int_\Omega K_{v_P}(\mathbf{r}) \delta \boldsymbol{v}_P(\mathbf{r}) d\mathbf{r} = \langle K_{v_P}, \delta \boldsymbol{v}_P \rangle_\Omega. \qquad (23)$$

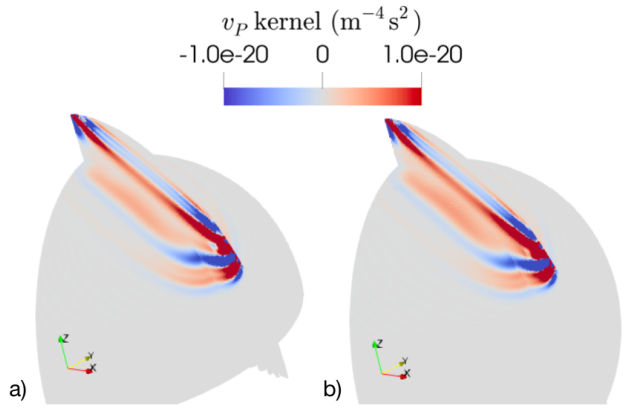

Figure 4. Kernel computed with AxiSEM3D in the continuous adjoint framework with the Fourier expansion n_u^{sup}, ensuring the adjoint wavefields are physical. a): result in PREM. b): result in s40rts. Note that the kernels are very similar.

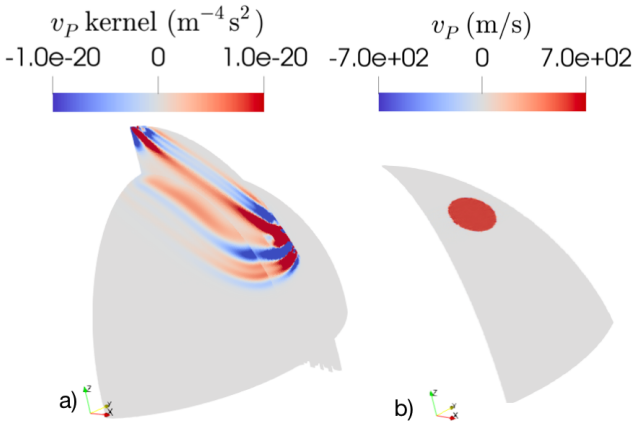

Figure 5. a) PREM kernel used in the benchmark. b) Azimuthal cut through 500km radius bubble at 5% strength used in benchmarks.

We want to emphasize that our framework is capable of computing kernels for other misfit measurements, for instance based on amplitude misfit, L2 misfit, phase and envelope etc. Indeed, this simply corresponds to a different adjoint source time function, as shown in the theoretical section. However, we choose cross-correlation traveltime delays because of their linear stability with the strength of velocity perturbations, as shown in Mercerat and Nolet. (2013). Indeed, we know that for the perturbations we choose, the linear traveltime update is correct, and hence we expect a correct kernel to reproduce it. If we were to choose a more non-linear misfit such as amplitude, any error in the benchmark would inevitably be attributable either to the kernel accuracy, or to the non-linearity of the measurement (i.e. the accuracy of the Born approximation for that measumrent), which would then not allow us to quantify the accuracy of the adjoint implementation.

Because our mesh is not a standard 3D mesh, we need to perform some manipulations to be able to compute δT_{ker} via this volume integral. From Leng et al. (2016), we know we can rewrite the integral over the 3D domain Ω in the following way:

$$\langle K_{v_P}, \delta \boldsymbol{v}_P \rangle_\Omega = 2\pi \sum_{|\beta| \leq n_u} \langle K_{v_P}^{-\beta}, \delta \boldsymbol{v}_P^{\beta} \rangle_D \qquad (24)$$

where the right hand side is now an integral over the D shaped SEM mesh and n_u is the local Fourier expansion. What's more, the sum over complex Fourier coefficients simplifies: because the azimuthal quantities are purely real, the Fourier space quantities are hermitian symmetric.

$$2\pi \sum_{|\beta| \leq n_u} \langle K_{v_P}^{-\beta}, \delta \boldsymbol{v}_P^{\beta} \rangle_D = 4\pi \sum_{\beta \leq n_u} \Re \langle K_{v_P}^{\beta}, \delta \boldsymbol{v}_P^{\beta} \rangle_D \qquad (25)$$

where $\Re$ denotes taking the real part.

The model perturbations we use are spherical (or "bubbles"), such as the one depicted in Figure 5 b). The center of the bubble is fixed, but we vary its radius and its strength relative to the background model.

To start with, we compute δT_{ker} in the continuous adjoint paradigm, i.e. by using a high Fourier expansion that ensures that our adjoint wavefields are physical, and thus reproduce kernels such as in Liu and Tromp (2008). In figure 6 we plot the traveltime updates for 3D bubbles of varying radius and strength.

We can see that the traveltime perturbations computed from the kernels match the cross correlation values while the size and strength of the model perturbation allows to use the Born approxi-

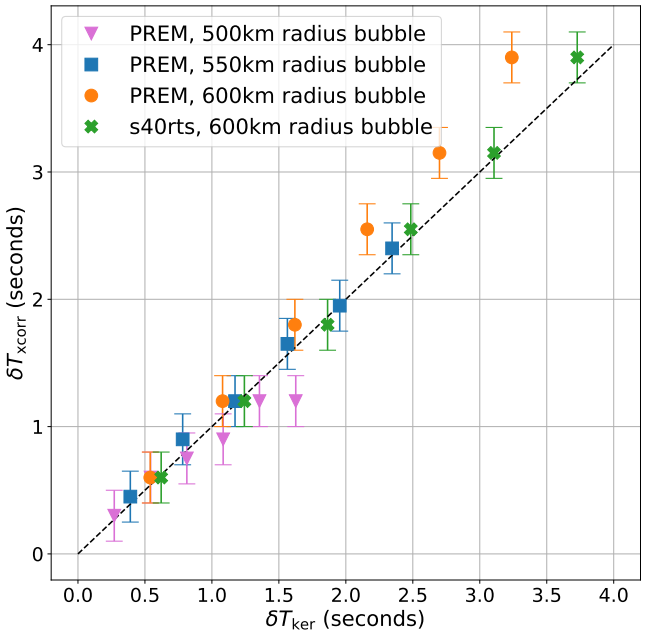

Figure 6. Traveltime perturbations predicted by kernels computed in PREM and S40RTS, compared with forward modelling prediction, for different perturbation geometries and strengths.

mation. It is noteworthy to recognize that the domain of validity of the Born approximation seems to be different for PREM and s40rts.

We have shown that we can compute accurate sensitivity kernels in the continuous adjoint framework. In the paradigm of AxiSEM3D this is not the most interesting approach because we lose most of the computational speedup offered by the flexible Fourier expansion, preserving the advantage mostly for the forward simulation.

4 DISCRETE ADJOINT KERNELS

Now that we have shown that the continuous adjoint kernels produce correct predictions, we can turn our attention to discrete adjoint kernels. In the discrete adjoint case, we do not enforce the adjoint simulation to produce physical wavefields. Therefore, we do not need a high and dense Fourier expansion, and we can simply reuse the Fourier expansion from the forward simulation. For our benchmarks, we will use a wisdom Fourier expansion that is sufficient to accurately model wavefields in our perturbed velocity model which includes the bubble. AxiSEM3D contains a feature that allows to obtain this sufficient expansion, called wavefield learning (Leng et al. 2019): Starting from a high Fourier expansion, the solver runs a simulation and progressively reduce the Fourier expansion while conserving the accuracy of the wavefield. We use the ”wisdom” (i.e. learned) Fourier expansion shown in figure 2, which is learned for the 500km radius bubble. The learned expansion is significantly lower that the initial dense Fourier expansion.

We emphasize the fact that the wavefield produced by an off-axis source with this reduced Fourier expansion is not physical, as we can see on figure 7 where a seismogram from the off-axis source with the new learned Fourier expansion is compared to the exact seismogram from the same station recorded with the source positioned on the axis of the SEM mesh. However, as we saw in theory and will now see in practice, this is not a requirement to

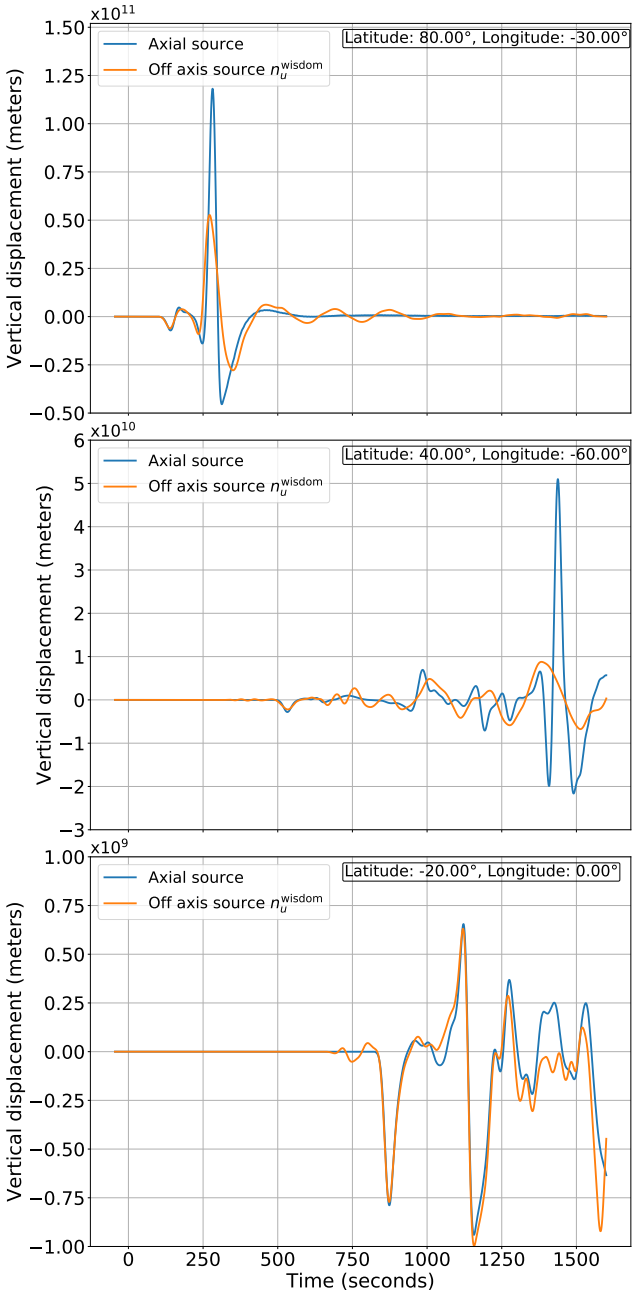

Figure 7. Seismograms obtained from an axial source, and the same seismograms with an off-axis source computed in the Fourier expansion learned for the 3D bubble, see Figure 2 b). As expected, the off-axis source with the n_u^{wisdom} expansion does not provide physically correct wavefields. This is the essence of the discrete adjoint method and the crucial point is that, as shown in figure 8 and figure 9, the kernels produced by these wavefields are exact. Amplitudes are scaled up to avoid floating point errors.

compute accurate gradients for the discretized model parameters. This is the key idea of the discrete adjoint approach.

In figure 8 b) we look at the same slice from the kernel as previously, and it appears to be very similar to the one obtained from the continuous adjoint (8 a)), barring artefacts localized around the adjoint source (8 c)). Kernels are routinely smoothed and filtered around the source and receiver areas because of near-fields effects, therefore we do not concern ourselves with a discrepancy at these specific locations.

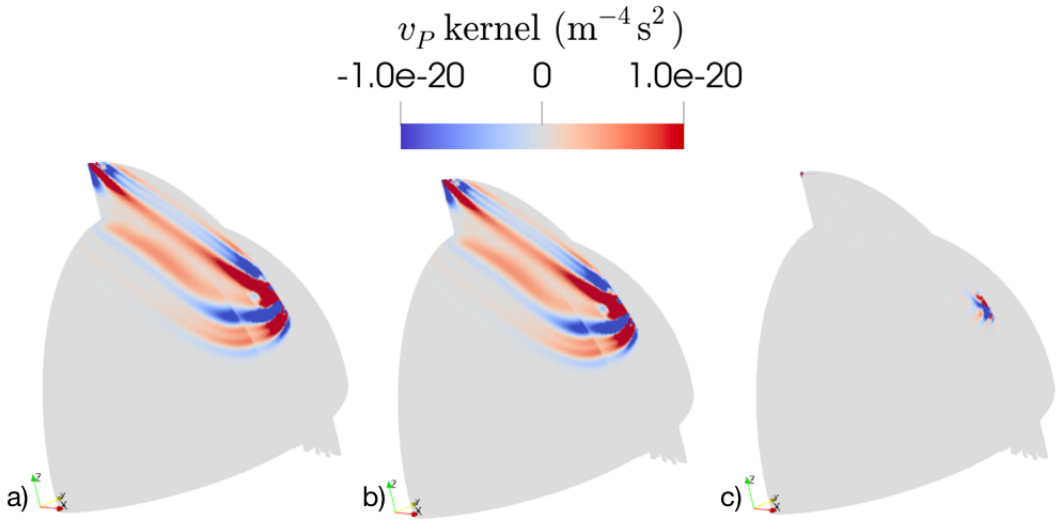

Figure 8. a): Cut from the continuous adjoint v_P kernel in PREM. b): Cut from the discrete adjoint v_P kernel in PREM. c): Difference between a) and b) kernels. The kernels are identical, except for some artefacts at the location of the adjoint source.

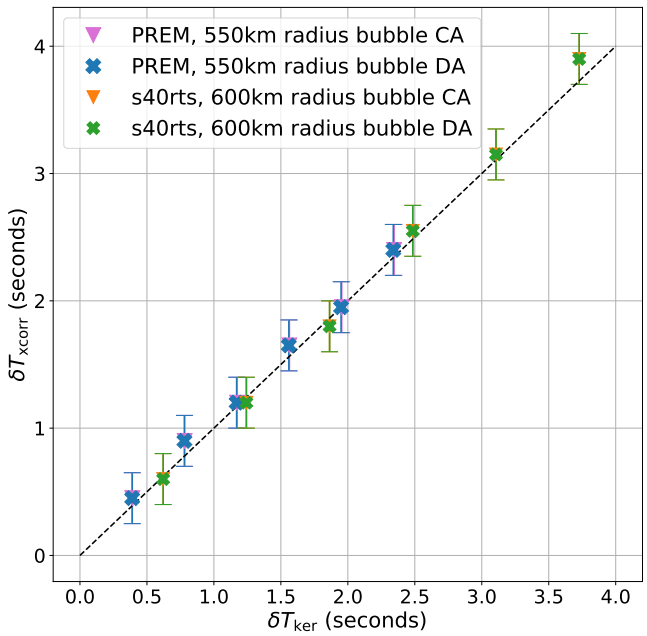

Figure 9. Traveltime perturbations predicted by discrete adjoint (DA) kernels and continuous adjoint (CA) kernels, compared to cross correlations predictions. The results from continuous and discrete adjoint kernels are identical.

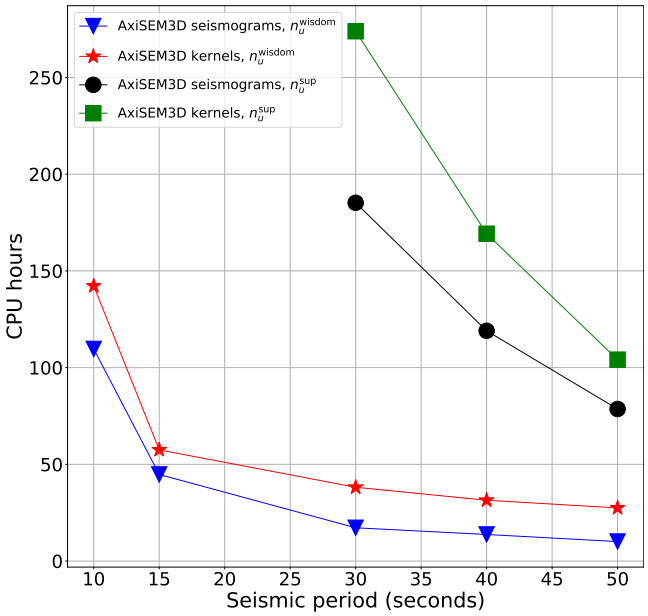

Figure 10. Computational cost of the forward ("seismograms") and the adjoint ("kernels") simulation for different seismic periods, in the continuous ("sup") and the discrete ("wisdom") adjoint approach.

We can now repeat the experiment from the previous section, in which we computed the traveltime updates obtained from the kernels for a given bubble, but this time using the discrete adjoint kernels computed in the wisdom Fourier expansion. The results are shown in figure 9, and as expected the kernels computed in the discrete adjoint framework produce the same traveltime perturbation as the kernels computed with the continuous adjoint.

4.1 Computational cost

The appealing aspect of our implementation of the discrete adjoint approach is that it decreases computational cost for kernels compared to other 3D SEM based software. Indeed, the learned wisdom expansion for a given inversion is mostly frequency independent, but instead driven by structural complexity. The computation of the kernels, besides the I/O overhead for loading and dumping wavefields, only consist of a few extra FFTs to convolve the forward and adjoint wavefields, and is thus of the same time complexity as a standard simulation, i.e. $O(\omega^3)$ as opposed to $O(\omega^4)$ for methods such as SPECFEM (Leng et al. 2016). If we want to use the continuous adjoint framework however, the complexity will rise to

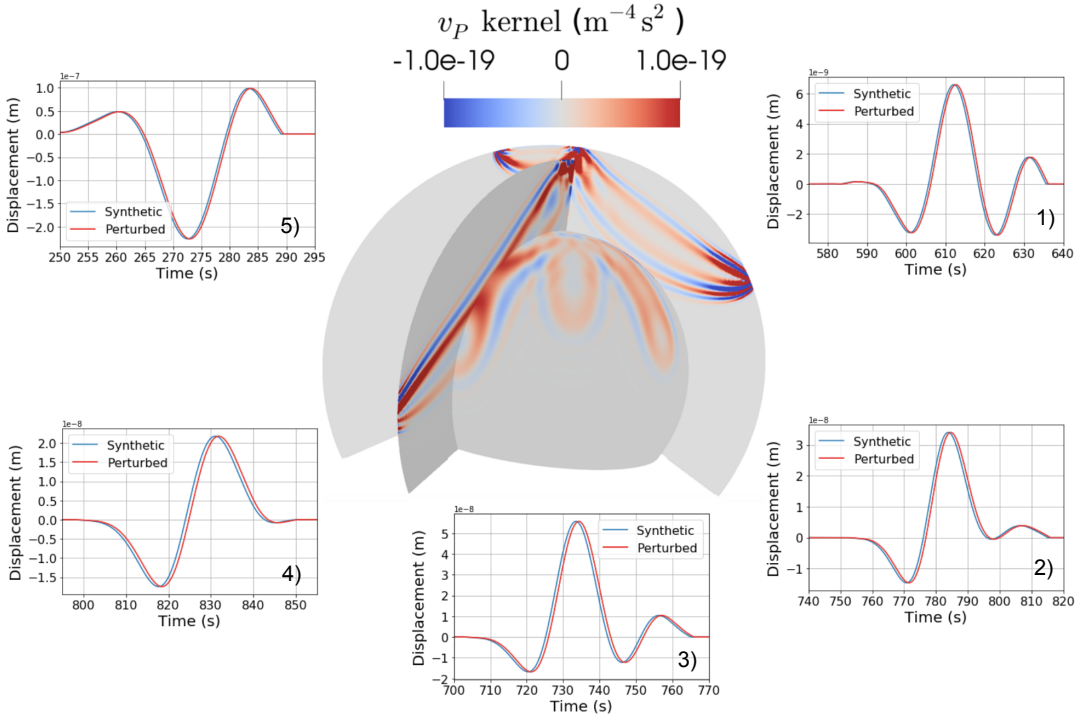

Figure 11. Event kernel at 15s with a magnitude 5.8 earthquake in Virginia, and 5 receivers, in S40RTS + crust1.0. We display three slices through the kernels, as well as a shell-like cut at 4000km radius. We also display the measurement used to compute the source time function of each kernel. "Synthetic" refers to the seismogram computed in S40RTS+crust1.0, and "Perturbed" refers to that same seismogram shifted by a small time offset, which artificially creates an observed seismogram. The time shifts used are (from seismograms 1) to 5)) 0.7s, 1s, 1s, 1s and 0.6s

$O(\omega^4)$, because the n_u^{sup} Fourier expansion is frequency dependent, just like in a standard 3D SEM method.

In figure 10 we compare the computational cost of a standard forward simulation with AxiSEM3D and a kernel computation, both in a discrete and continuous adjoint settings. The simulations are run in S40RTS and they produce 700s of seismograms. The figure reveals two main features: First, we see that for both the continuous and discrete adjoint approaches, the computational cost of the kernels scales similarly to a simple forward simulation. Furthermore, the discrete adjoint approach to kernel computation indeed scales much more advantageously with simulation period than the continuous adjoint approach.

4.2 Event kernels

In the above, we computed kernels from an individual source/receiver pair so that we can perform benchmarks. However, in FWI one typically computes kernels for all n_r receivers at once for any given earthquake owing to the efficiency of the adjoint approach. Our discrete-adjoint framework is naturally capable of this feature. We compute an event kernel at 15s and one at 5s from the 2011 Virginia earthquake at 12km depth with magnitude 5.8. For an event kernel, the individual source time functions are multiplied by a measured cross correlation time delay between synthetics and real data. Since our example is purely synthetic, we introduce time delays for each station reminiscent of having actual observations at hand, by simply perturbing the computed seismograms by a given time shift. While in the benchmarks we did not include a 3D crust, we now run the simulation at 15s in S40RTS + crust1.0 (Ritsema et al. 2011; Laske et al. 2013), and we perform the run in a Fourier expansion n_u^{wisdom} learned during the forward simulation. We only place 5 receivers to render details discernible, but the same simulation principle and cost applies to any number of receivers. We compute latitudinal slices, as well as constant radius shells, shown in figure 11. The receivers are placed at various latitudes to illustrate different kernel shapes, from a close 20° epicentral distance kernel, to a much deeper penetrating kernel at 110°. We conveniently choose artificial travel time delays so that kernels at different latitudes are clearly visible on the same color scale. We can clearly see the traditional banana-doughnut shaped sensitivity of body waves.

For the 5s run in figure 12, we place three receivers at equal epicentral distance of 50°, so that we can choose artificial time delays that highlight changes due to measurements in similar kernels. The figure illustrates how emphasis is placed on regions where there is more misfit between synthetics and observations. We perform the simulation in S40RTS.

5 DISCUSSION AND CONCLUSIONS

We are introducing a new framework for computing the sensitivity kernels using adjoint methods in the context of an accelerated approach relying on the forward solver AxiSEM3D. Our approach exploits the inherent azimuthal smoothness of a global wavefield by using AxiSEM3D for both forward and adjoint simulations, which allows to flexibly adjust the mesh according to the desired complexity and thus lends itself very well to the discrete adjoint approach. We compute kernels in the continuous adjoint and discrete adjoint frameworks in global scale tomographic models, and

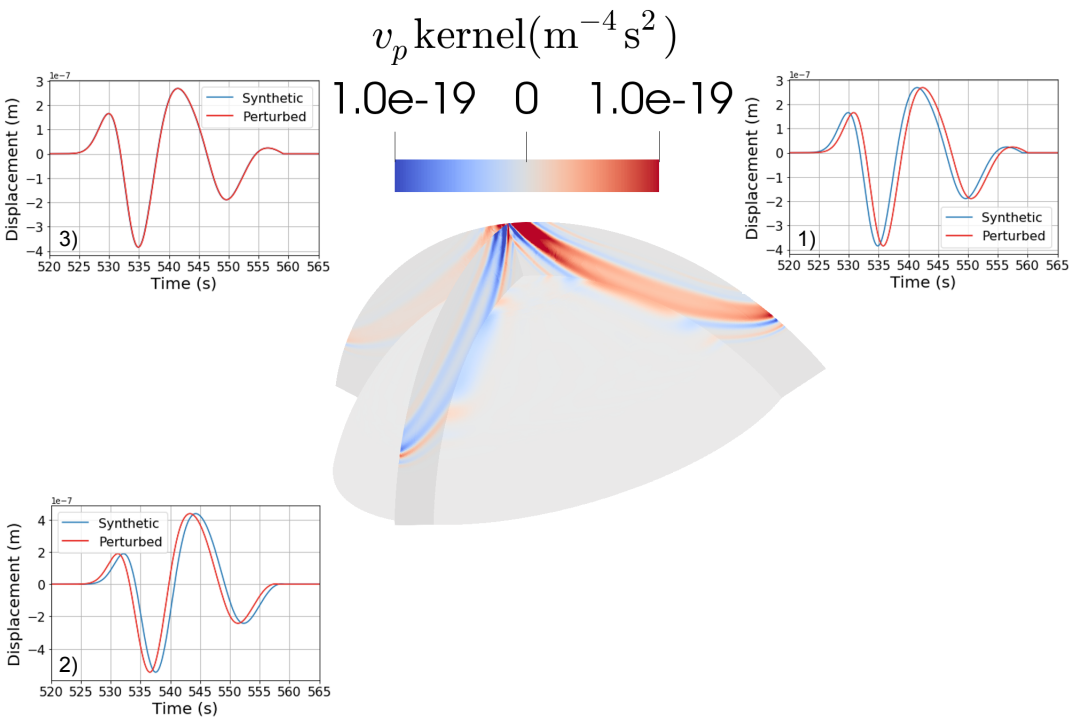

Figure 12. Event kernel at 5s with a magnitude 5.8 earthquake in Virginia, and 3 receivers, in S40RTS, at epicentral distance 50°. We display three slices through the kernels, as well as a shell-like cut at 5000km radius. We also display the measurement used to compute the source time function of each kernel. "Synthetic" refers to the seismogram computed in S40RTS, and "Perturbed" refers to that same seismogram shifted by a small time offset, which artificially creates an observed seismogram. The time shifts used are (from seismograms 1) to 3)) 1s, -1s, and 0.1s

we show that our kernels are exact by comparing them with cross correlation traveltime perturbations predicted from forward simulations with AxiSEM3D. The continuous adjoint approach is built upon the novel implementation of off-axis sources which are required to leverage the benefits of the adjoint method by igniting them simultaneously. Of course, our off-axis source implementation can also be used for other problems such as finite kinematic earthquake sources and source encoding (Tromp and Bachmann 2019).

By using the discrete adjoint method, we are able to obtain kernels at a reduced computational cost and with reduced disk storage compared to traditional spectral element methods. Indeed, our method retains the $O(\omega^3)$ scaling with simulation frequency from AxiSEM3D for tomographic mantle models, as opposed to $O(\omega^4)$ for other software. Even though disk storage in AxiSEM3D is much lower than in other methods because of the flexible mesh parametrisation, we will in future work implement well documented checkpointing algorithms (Komatitsch et al. 2016), which allow to only store a subset of the time steps of the forward wavefield. These methods are more computationally expensive because they require recomputing part of the forward wavefield during the adjoint simulation. However, this does not change the frequency scaling and our method will remain $O(\omega^3)$, which is the key benefit of our implementation.

The main idea of the discrete adjoint method, thanks to the discrete wave operator being self adjoint and therefore hermitian, is that it is not necessary for the adjoint simulation to produce physically accurate wavefields, instead the only requirement is that it is run on the same discretization as the forward simulation. This allows for the obtained kernels to be exact for the given discretized model parameters. In our case, this means we can keep the flexible Fourier expansion enabled by AxiSEM3D, and thus apply the complexity-adapted efficiency and speedup from the forward problem in the same manner for the inverse problem. However, we need to be cautious when choosing the Fourier expansion we work in. In our benchmarks we assumed a known expansion corresponding to the perturbed Earth model. In a tomographic inversion however, we have no knowledge of what is inside the Earth. If choosing an arbitrary Fourier expansion to work in, one may run the risk that the expansion is not high enough to accurately model the actual velocity model resolved by the data. This, of course, is always the case for any inversion, but the discrete adjoint approach illuminates this pre-defined fixed resolution problem clearly. This situation is identical to the problem of designing a tomographic inversion mesh, where one risks missing structure but also possibly overfitting the data. In that case, the mesh structure is determined by data coverage and resolution, with regions of dense ray paths having a fine meshing, and a coarser one for regions with less ray paths. Essentially, this means fixing the model resolution prior to running the inversion, which is exactly what we're interested in, because it will have very dense regions for which we'll have a high Fourier expansion, and very sparse regions for which the expansion will be very low. In subsequent work we will be designing such a mesh adapted to our framework, and demonstrating an example of global seismic tomography using this novel framework.

6 ACKNOWLEDGMENTS

The authors thank Christian Böhm and Martin van Driel for helpful discussions related to the discrete adjoint method during and since the TIDES summer school in Oxford (2017). Comments by Pierre Jacquet, editor Sidao Ni and an anonymous reviewer significantly improved the quality of the manuscript. AS thanks the anonymous donor of his PhD grant which allowed this work to be carried out. KL and TNM were supported by NERC grant NE/R012199/1. The numerical simulations for this work were performed on TNM's student and NERC project grants on the ARCHER supercomputer, whose support is kindly acknowledged.

REFERENCES

Auer, L., Boschi, L., Becker, T. W., Nissen-Meyer, T., and Giardini, D. (2014). Savani: A variable resultion whole-mantle model of anisotropic shear velocity variations based on multiple data sets, in *Journal of Geophysical Research: Solid Earth*. **119(4)**, 3006-3034.

Bijwaard, H. and Spakman, W., 1999. Tomographic evidence for a narrow whole mantle plume below Iceland, in *Earth and Planetary Science Letters*, **166(3)**, 121–126.

Bozdag, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N., and Pugmire, D., 2016. Global adjoint tomography: first-generation model, in *Geophysics Journal International*, **207**, 1739-1766.

Colombi, A., Nissen-Meyer, T., Boschi, L. and Giardini, D., 2014. Seismic waveform inversion for core–mantle boundary topography, in *Geophysical Journal International*, **198(1)**, 55–71.

Dahlen, F. A., Hung, S.-H., and Nolet, G. (2000). Frechet kernels for finite-frequency traveltimes - I. Theory, in *Geophysical Journal International*, **141(1)**, 157–174.

Dziewonski, A. M., Hager, B. H., and O'Connell, R. J., 1977. Large-scale heterogeneities in the lower mantle, in *Journal of Geophysical Research*, **82(2)**, 239-255.

Dziewonski, A. M.; Anderson, D. L., 1981. Preliminary reference Earth model, in *Physics of the Earth and Planetary Interiors*, **25(4)**, 297–356.

Elliott, J., and Peraire, J., 1997. Practical Three-Dimensional Aerodynamic Design and Optimization Using Unstructured Meshes, in *AIAA Journal*, **35 (9)**, 1479–1485.

Fichtner, A., 2011 *Full Waveform Inversion and Modelling*, Springer.

Fichtner, A., Kennett, B. L. N., Igel, H., and Bunge, H., 2009. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, in *Geophysical Journal International*, **179(3)**, 1703–1725.

Fichtner, A., van Driel, M., 2014. Models and Frechet kernels for frequency-(in)dependent Q, in *Geophysical Journal International*, **198**, 1878-1889.

French, S. W. and Romanowicz, B., 2015. Broad plumes rooted at the base of the Earth's mantle beneath major hotspots, in *Nature*, **525(7567)**, 95–99.

Giles, M.B. and Duta, M.C., 2003. Algorithm Developments for Discrete Adjoint Methods, in *AIAA Journal*, **41(2)**, 198–205.

Grand, S.P., Van der Hilst, R.D., Widiyantoro, S., 1997. High resolution global tomography: a snapshot of convection in the Earth, in *Geological Society of America Today*, **7(4)**.

van der Hilst, R., Widiyantoro, S., and Engdahl, E., 1997. Evidence for deep mantle circulation from global tomography, in *Nature*, **386**, 578–584.

Hosseini, K., Sigloch, 2015. Multifrequency measurements of core-diffracted P waves (Pdiff) for global waveform tomography, in *Geophysical Journal International*, **203(1)**, 506-521.

Hosseini, K., Sigloch, K., Tsekhmistrenko, M., Zaheri, A., Nissen-Meyer, T., Igel, H. 2020. Global mantle structure from multifrequency tomography using P, PP and P-diffracted waves, in *Geophysical Journal International*. **220(1)**, 96-141.

Jeffreys, H., 1926. The Rigidity of the Earth's Central Core, in *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*. **1(7)**, 371–383.

Kennett, B., Widiyantoro, S., and Van Der Hilst, R., 1998. Joint seismic tomography for bulk-sound and shear wavespeed, in *J. geophys. Res*. **103(12)**, 469–12.

Koelemeijer, P., Ritsema, J., Deuss, A., and van Heijst, H.-J., 2016. Sp12rts: a degree-12 model of shear- and compressional-wave velocity for earth's mantle, in *Geophysical Journal International*, **204(2)**, 1024–1039.

Komatitsch, D., Xie, Z., Bozdag, E., Sales de Andrade, E., Peter, D., Liu, Q., and Tromp, J., 2016. Anelastic sensitivity kernels with parsimonious storage for adjoint tomography and full waveform inversion, in *Geophysical Journal International*, **206(3)**, 1467–1478.

Kustowski, B., Ekström, G, Dziewonski, A.M., 2008. Anisotropic shear-wave velocity structure of the Earth's mantle: A global model, in *Journal*

of Geophysical Research: Solid Earth, **113(B6)**

Laske, G., Masters., G., Ma, Z. and Pasyanos, M., 2013. Update on CRUST1.0 - A 1-degree Global Model of Earth's Crust, in *Geophys. Res. Abstracts*, **15**, Abstract EGU2013-2658.

Lehmann, I., 1936. P', Publ, in *Bur. Centr. Seism. Internat. Serie A*, **14**, 87–115.

Leng, K., Nissen-Meyer, T., and van Driel, M., 2016. Efficient global wave propagation adapted to 3-D structural complexity: a pseudospectral/spectral-element approach, in *Geophysical Journal International*, **207**, 1700-1721.

Leng, K., Nissen-Meyer, T., Van Driel, M., Hosseini, K., and Al-Attar, D.,2019. AxiSEM3D: broad-band seismic wavefields in 3-D global Earthmodels with undulating discontinuities, in *Geophysical Journal International*, **217(3)**, 2125–2146.

Leng, K., Korenaga, J. and Nissen-Meyer, T. Three-dimensional scattering of elastic waves by small-scale heterogeneities in the Earth's mantle, submitted to *Geophysical Journal International*.

Lions, J.L., 1971 *Optimal Control of Systems Governed by Partial Differential Equations*, translated by S.K. Mitter, Springer-Verlag, Berlin.

Liu, Q. and Tromp, J., 2008. Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods, in *Geophysical Journal International*, **174(1)**, 265-286.

van der Meer, D. G., Spakman, W., van Hinsbergen, D. J. J., Amaru, M. L., and Torsvik, T. H., 2010. Towards absolute plate motions constrained by lowermantle slab remnants, in *Nature Geoscience*, **3(1)**, 36–40.

Mercerat, E. D. and Nolet, G., (2013). On the linearity of cross-correlation delay times in finite-frequency tomography, in *Geophysical Journal International*, **192(2)**, 681–687.

Mohorovičić, A., 1909. Earthquake of 8 October 1909 (Potres od 8. X 1909; Das Beben vom 8. X. 1909), in *Yearly Report of the Zagreb Meteorological Observatory for the Year*.

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

Montelli, R., Nolet, G., Masters, G., Dahlen, F. A., and Hung, S.-H., 2004. Global P and PP traveltime tomography: rays versus waves, in *Geophysical Journal International*, **158(2)**, 637–654.

Nissen-Meyer, T., Dahlen, F.A., Fournier, A., 2007. Spherical-earth Fréchet sensitivity kernels, in *Geophysical Journal International*, **168(3)**, 1051-1066.

Nissen-Meyer, T., van Driel, M., Stähler, S., Hosseini, K., Hempel, S., Auer, L., Colombi, A., Fournier, A., 2014. AxiSEM: broadband 3-D seismic wavefields in axisymmetric media, in *Solid Earth*, **1**, 425-445.

Nolet, G. and Montelli, R., 2005. Optimal parametrization of tomographic models, in *Geophysical Journal International*, **161**, 365–372.

Oldham, R. D., 1906. The constitution of the interior of the Earth, as revealed by earthquakes, in *Quarterly Journal of the Geological Society*, **62(1-4)**,456–475.

Pironneau, O., 1974. On Optimum Design in Fluid Mechanics, in *Journal of Fluid Mechanics*, **64**, 97-110.

Ritsema, J., Deuss, A., van Heijst, H.J., Woodhouse, J.H., 2011. S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, in *Geophysical Journal International*, **184(3)**, 1223-1236.

Ritsema, J. and Van Heijst, H. J., 2000. Seismic imaging of structural heterogeneity in Earth's mantle: Evidence for large-scale mantle flow, in *Science Progress*, **83**, 243–259.

Ritsema, J., van Heijst, H. J., and Woodhouse, J. H., 1999. Complex shear wave velocity structure imaged beneath Africa and Iceland, in *Science*, **286(5446)**, 1925–1928.

Schaeffer, A.J. and Lebedev, S., 2013. Global shear speed structure of the upper mantle and transition zone, in *Geophysical Journal International*, **194(1)**, 417-449.

Schuberth, B. S. A., Bunge, H.-P., and Ritsema, J., 2009a. Tomographic filtering of high-resolution mantle circulation models: Can seismic heterogeneity be explained by temperature alone?, in *Geochem. Geophys. Geosyst.*, **10(5)**, Q05W03.

Schuberth, B. S. A., Bunge, H.-P., Steinle-Neumann, G., Moder, C., and Oeser, J., 2009b. Thermal versus elastic heterogeneity in high-resolution mantle circulation models with pyrolite composition: High plume excess temperatures in the lowermost mantle, in *Geochem. Geophys. Geosyst.*, **10(1)**, Q01W01.

Sigloch, K. and Nolet, G., 2006. Measuring finite-frequency body-wave amplitudes and traveltimes, in *Geophysical Journal International*, **167(1)**, 271–287.

Sigloch, K., McQuarrie, N., and Nolet, G., 2008. Two-stage subduction history under North America inferred from multiple-frequency tomography, in *Nature Geoscience*, **1**, 458–462.

Sigloch, K. and Mihalynuk, M. G., 2013. Intra-oceanic subduction shaped the assembly of Cordilleran North America, in *Nature*, **496(7443)**, 50–56.

Stähler, S.C., Sigloch, K. and Nissen-Meyer, T., 2012. Triplicated P-wave measurements for waveform tomography of the mantle transition zone, in *Solid Earth*, **3**, 339–354.

Tape, C., Liu, Q., Maggi, A., and Tromp, J., 2009. Adjoint tomography of the southern California crust, in *Science*, **325(5943)**, 988–92.

Tarantola, A., 1984. Linearized Inversion of Seismic Reflection Data, in *Geophysics*, **49**, 1259-1266.

Tesoniero, A., Leng, K., Long, M.D. and Nissen-Meyer, T., 2020. Full wave sensitivity of SK(K)S phases to arbitrary anisotropy in the upper and lower mantle, in *Geophysical Journal International*, **222(1)**, 412-435.

Thrastarson, S., van Driel, M., Krischer, L., Boehm, C., Afanasiev, M., van Herwaarden, D-P. and Fichtner, A., 2020. Accelerating numerical wave propagation by wavefield adapted meshes. Part II: full-waveform inversion, in *Geophysical Journal International*, **221**, 1591-1604.

Tromp, J., Tape, C., and Liu, C., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, in *Geophysical Journal International*, **160**, 195-216.

Tromp, J., Komatitsch, D., Liu, Q., 2008. Spectral-Element and Adjoint Methods in Seismology, in *Communications in computational physics*, **3(1)**, 1-32.

Tromp, J. and Bachmann, E., 2019. Source encoding for adjoint tomography, in *Geophysical Journal International*, **218**, 2019-2044.

Van der Voo, R., Spakman, W., and Bijwaard, H., 1999a. Mesozoic subducted slabs under Siberia, in *Nature*, **397(6716)**, 246–249.

van Driel, M. and Nissen-Meyer, T., 2014. Seismic wave propagation in fully anisotropic axisymmetric media, in *Geophysical Journal International*, **199(2)**880-893.

van Driel, M., Boehm, C., Krischer, L., and Afanasiev, M., 2019. Accelerating numerical wave-propagation using wavefield adapted meshes, Part I: Forward and adjoint modelling. https://doi.org/10.31223/osf.io/43ydf

Virieux, J. and Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, in *Geophysics*, **74(6)**, WCC1-WCC26.

Willeman, R.J., Storchak, D.A., 2001. Data Collection at the International Seismological Centre, in *Seismological Research Letters*, **72(4)**, 440-453.

Zaroli, C., Lambotte, S., and L'évêque,J.-J., 2015. Joint inversion of normal-mode and finite-frequency S -wave data using an irregular tomographic grid., in *Geophysical Journal International*, **203(3)**1665-1681.

Zhao, D., 2004. Global tomographic images of mantle plumes and subducting slabs: insight into deep Earth dynamics, in *Physics of the Earth and Planetary Interiors*, **146(1)**, 3–34.