Accelerating numerical wave propagation by wavefield adapted meshes, Part II: Full-waveform inversion

Journal Article

Author(s):
Thrastarson, Sölvi; van Driel, Martin; Krischer, Lion; Boehm, Christian; Afanasiev, Michael; van Herwaarden, Dirk-Philip; Fichtner, Andreas

Publication date:
2020-06

Permanent link:
https://doi.org/10.3929/ethz-b-000396951

Rights / license:
Creative Commons Attribution 4.0 International

Originally published in:
Geophysical Journal International 221(3), https://doi.org/10.1093/gji/ggaa065

Funding acknowledgement:
823844 - Centre of Excellence for Exascale in Solid Earth (EC)
Accelerating numerical wave propagation by wavefield adapted meshes. Part II: full-waveform inversion

Solvi Thrastarson, Martin van Driel, Lion Krischer, Christian Boehm, Michael Afanasiev, Dirk-Philip van Herwaarden and Andreas Fichtner

Department of Earth Sciences, Institute of Geophysics, ETH Zürich, Sonneggstrasse 5, 8092 Zürich, Switzerland. E-mail: soelvi.thrastarson@erdw.ethz.ch

Accepted 2020 February 1. Received 2019 July 18; in original form 2020 January 30

201SUMMARY
We present a novel full-waveform inversion (FWI) approach which can reduce the computational cost by up to an order of magnitude compared to conventional approaches, provided that variations in medium properties are sufficiently smooth. Our method is based on the usage of wavefield adapted meshes which accelerate the forward and adjoint wavefield simulations. By adapting the mesh to the expected complexity and smoothness of the wavefield, the number of elements needed to discretize the wave equation can be greatly reduced. This leads to spectral-element meshes which are optimally tailored to source locations and medium complexity. We demonstrate a workflow which opens up the possibility to use these meshes in FWI and show the computational advantages of the approach. We provide examples in 2-D and 3-D to illustrate the concept, describe how the new workflow deviates from the standard FWI workflow, and explain the additional steps in detail.

Key words: Waveform inversion; Computational seismology; Seismic tomography.

1 INTRODUCTION
Since the first applications of seismic tomography (e.g. Aki & Lee 1976; Dziewoński et al. 1977; Aki et al. 1977) the field has advanced through an increase in quantity and quality of data, which in turn has motivated methodological advancements. Today it is technologically possible to utilize the full information from waveforms recorded at seismic stations distributed around the globe in a broad frequency range. To exploit this information, the physical equations describing how a wavefield propagates away from a source and through a potentially complex medium need to be solved accurately.

Relying on numerical wavefield simulations, full-waveform inversion (FWI) is a method to extract information from complete seismic waveforms, which has become practical only comparatively recently, despite its much earlier conceptualization (Bamberger et al. 1977, 1982; Lailly & Bednar 1983; Tarantola 1984, 1988). Meanwhile, FWI has been applied in global seismology (e.g. Lekić & Romanowicz 2011; French & Romanowicz 2014; Bozza et al. 2016; Afanasiev et al. 2016; Fichtner et al. 2018), in studies of regional Earth structure (e.g. Chen et al. 2007b; Tape et al. 2009; Fichtner et al. 2009; Zhu et al. 2012; Simutė et al. 2016; Krischer et al. 2018), at the exploration scale (e.g. Igel et al. 1996; Virieux & Operto 2009; Sirgue et al. 2010; Warner et al. 2013), and in medical ultrasound tomography (Pratt et al. 2007; Boehm et al. 2018).

FWI requires the computation of synthetic waveforms which are compared to observed data. Starting from an initial model (medium), FWI, in a tomographic application, iteratively minimizes waveform misfits by modifying the model in each iteration. The modifications applied to the model are computed using the gradient of the misfit with respect to the model parameters. The gradient is computed either using the adjoint method (e.g. Tarantola 1988; Tromp et al. 2005; Fichtner et al. 2006; Plessix 2006) or the scattering-integral method (e.g. Chen et al. 2007b, a). Both methods compute a gradient with one additional wavefield simulation.

1.1 Waveform modelling
Simulating wave propagation through arbitrary heterogeneous media can only be done numerically. This is a computationally challenging task which has been studied extensively in recent decades, leading to a diverse family of methods (e.g. Igel 2016). The spectral-element method (SEM, Patra 1984; Seriani & Priolo 1994; Facioli et al. 1996, 1997; Komatitsch & Vilotte 1998) has become a standard method for numerical wave propagation simulations in the community of passive seismology. Several wave propagation solvers based on SEM currently exist (e.g. Komatitsch & Tromp 2002; Peter et al. 2011; Gokhberg & Fichtner 2016; Cupillard et al. 2012; Afanasiev et al. 2019). These solvers have the capability to accurately model seismic wave propagation through complex 3-D media. As with any other numerical wave propagation method, the simulations swiftly become computationally expensive, either
through an increase in modelled frequency or spatial extent of the model. In an $N$-dimensional medium, the computational cost of the simulation scales with frequency to the power of $N+1$. The $N$ is due to an increase in the number of elements needed to discretize the medium in each dimension and the $+1$ results from the reduced time-step needed to meet the Courant–Friedrichs–Lewy (CFL) criterion (Courant et al. 1928) in explicit time stepping schemes.

The computational cost of FWI scales with simulation cost, the number of sources used in the inversion and the number of iterations performed. In an ideal case, it would be beneficial to include as much data as possible (many sources) and iterate until observational error tolerance is reached. It is thus of great importance to reduce the cost of the wave propagation simulations, ideally while still resolving the relevant waveforms.

1.2 Reducing the computational cost

Improvements in hardware and algorithm design have contributed to reduce the computational burden of wave propagation simulations. For instance, moving the SEM to a GPU architecture can result in significant speed-ups depending on the setup (Komatitsch et al. 2010; Rietmann et al. 2012; Gokhberg & Fichtner 2016). The implementation of local time-stepping (Dumbser et al. 2007; Rietmann et al. 2015) can result in similar speed-ups for specific problems where a relatively low number of small elements limit the global time-step of the simulation.

Over the past decades, a lot of research has been devoted to the design of adaptive mesh refinement techniques for the wave equation and other time-dependent partial differential equations. Significant speed-ups have been reported for several applications using time-dependent meshes that refine and coarsen the mesh dynamically based on error estimators (Arney & Flaherty 1989; Berger & LeVeque 1998; Bangerth & Rannacher 2001; Bangerth et al. 2010; Kröner 2011; Schoeder et al. 2019). This includes simulations on massively parallel compute architectures (Burstedde et al. 2011; Bangerth et al. 2012), and strategies that make use of the adjoint equation to steer the mesh refinement (Davis & LeVeque 2018).

For FWI on continental to global scales, the SEM on conforming unstructured hexahedral meshes using an explicit time-stepping scheme has emerged as the most widely used method (Tramp et al. 2008; Fichtner 2010; Ferroni et al. 2017). While there is no inherent limitation in the mesh refinement techniques mentioned above, the authors are not aware of fully automated and stable refinement schemes for these types of meshes.

Simulation cost has also been reduced with simplifications of the full 3-D problem. AxiSEM, for instance, exploits the approximate axial symmetry of global wavefields by using 2-D meridian slice computations and analytically extending to the full globe (Nissen-Meyer et al. 2007, 2008, 2014). However, as the medium complexity increases, both in terms of shape and interior structure, the requisite assumptions on symmetry begin to break down.

As a remedy to that limitation, Leng et al. (2016, 2019) propose the coupling of SEM and the pseudospectral method (Gazdag 1981; Kosloff et al. 1984), which extends the AxiSEM principle to handle more complex 3-D structure (AxiSEM3D). The method can result in a speedup of more than one order of magnitude, depending on model complexity and modelled frequencies.

Conceptually related to AxiSEM, Part I of this paper by van Driel et al. (2020) propose to use anisotropic adaptive mesh refinements (aAMR) to construct wavefield adapted meshes for SEM simulations. The novelty of the wavefield adapted meshes is to make use of prior knowledge on the approximate shape of wave fronts. In media with smooth deviations from a potentially discontinuous background, the total number of elements needed to represent the wavefield may be reduced significantly, thereby leading to a decrease in computational cost. In contrast to dynamic mesh refinements during the simulation, the construction of an aAMR mesh happens only once, prior to the simulation, thereby avoiding costly dynamic repartitioning and load rebalancing on massively parallel machines.

1.3 Motivation and outline

The principal motivation of this contribution is a first proof of concept, showing that wavefield adapted meshes combined with the discrete adjoint technique may lead to an FWI implementation that requires significantly lower computational resources, in cases where variations in medium properties are sufficiently smooth. Indeed, we demonstrate, that under favourable conditions, the approach can result in up to an order of magnitude speed-up in both waveform and adjoint modelling, while effectively reducing the frequency ($f$) scaling from $f^{N+1}$ to $f^N$, where $N$ is the dimension. Though this pilot study is focused on 2-D, it serves the important purpose of algorithmic developments, especially in preparation of extensions to 3-D. Furthermore, it may, as it stands, be applied in scenarios where 2-D wave propagation can serve as an analogue of surface waves in a 3-D Earth.

The manuscript is organized as follows. In Section 2, we begin by briefly explaining the underlying idea behind the new meshing technique. Section 3 is dedicated to a detailed description of how this fits into the FWI workflow, which additional steps are required, and how we implement them. In Sections 4 and 5, we show examples in 2-D and 3-D, respectively, and demonstrate how this can potentially be used for large scale seismic tomography. Finally, we discuss the most important advantages and limitations of the method, as well as potential future applications.

2 WAVEFIELD ADAPTED MESHES

In order to solve a differential equation (e.g. the wave equation) numerically, the continuous fields have to be approximated in a discrete representation. To properly represent a wavefield in a discrete domain, it has to be sampled with a certain number of gridpoints per wavelength, typically around 5–10 gridpoints depending on the method (Fichtner 2010; Igel 2016). The number of gridpoints in a spectral-element mesh is controlled by the number of elements in the mesh and the polynomial order of the basis functions inside the elements. From here on we will assume the most widely used 4th-order polynomal basis functions and focus on the number of elements.

The computational cost of spectral-element simulations directly scales with the number of elements in the discrete mesh. It is thus of great interest to use the minimum number of elements required in each simulation.

In the 2-D example in Fig. 1(a), using a polar coordinate system with the pole at the source location, one can compare the wavelengths of the wavefield in the radial and the azimuthal dimensions. Note that by a wavelength in a certain dimension we refer to the length of the wavefield in the radial and the azimuthal dimensions.

In smooth media, the azimuthal wavelength (perpendicular to the propagation direction) is mostly controlled by medium complexity,
and the radial one (parallel to the propagation direction) mostly by the period of the injected source–time–function. Except close to the source, the azimuthal wavelength is much larger than the radial one. This suggests to elongate the elements in azimuthal direction while still keeping an appropriate number of gridpoints per wavelength, thereby greatly reducing the total number of elements in the mesh.

Adapting the mesh according to the expected wavefield complexity is called anisotropic adaptive mesh refinement. The following is a summary showing how AMR meshes work in forward and adjoint wavefield modelling; for more details the reader is referred to van Driel et al. (2020).

Figs 1(a) and (b) show a wavefield propagating through a homogeneous medium on an AMR mesh and on a rectilinear mesh, respectively. The elements in the AMR mesh are aligned with the wavefield propagating from the source location, and the wavefields are approximately identical on both meshes. In the simplistic case...
of a homogeneous medium, the wavefront and the elements of the aAMR mesh are perfectly aligned. With increasing medium complexity, the wavefield becomes more complex, and so the number of elements in azimuthal direction will need to be increased accordingly, as detailed in Section 4.

The aAMR meshes are designed for a single source location. This apparently poses a problem in adjoint simulations where adjoint sources are located at the actual receiver positions. As can be seen in Fig. 1(c), the aAMR adjoint wavefield does not resemble a physical wavefield. The adjoint wavefield computed using the rectilinear mesh (Fig. 1d), in contrast, resembles a physical wavefield propagating from the 25 adjoint source locations.

Fortunately, the problem is only an apparent one. In fact, the discrete adjoint state does not have to resemble a physical wavefield that is radiating away from point sources emitting energy at the receiver locations. Despite the poor approximation of near-source effects,
the relative error between the directional derivative and its finite difference approximation, that is,
\[ \varepsilon = \frac{1}{h} \left( \chi(m + h \delta m) - \chi(m) - (\nabla \chi(m) \cdot \delta m) \right) \]  
with the relative error expressed as,
\[ \text{Rel}(\varepsilon) = \frac{\varepsilon}{\nabla \chi(m) \cdot \delta m}, \]
where \( \chi(m) \) is the \( L_2 \) waveform misfit, eq. (3) using model \( m, h \) is the relative model perturbation or finite-difference step length, and \( m = (v_P, v_S, \rho) \) are the physical model parameters \( P \) velocity, \( S \) velocity and density, respectively. The relative error plotted as a function of \( h \) has a classic hockey stick shape, where larger values of \( h \) correspond to linearization errors, and small values of \( h \) are related to errors due to numerical precision. This is the expected behaviour of an accurate discrete gradient and demonstrates the correctness of our implementation.

Gradients are discrete representations of continuous sensitivity kernels and the way a gradient looks depends on which basis it is represented in. The gradient test shows that the \( aAMR \) gradients are accurate in the \( aAMR \) discretization and that the differences to the rectilinear gradients are simply the result of different model parametrization. In this work, we demonstrate where this approach is applicable to FWIs and where it starts to break down.

### 3 Full-Waveform Inversion Workflow

The summary in the previous section and the more detailed analysis in van Driel et al. (2020) demonstrate that \( aAMR \) is applicable to both forward and adjoint wave equation modelling. In the following section we will detail how to construct a complete FWI workflow using these developments.

Complications arise when \( aAMR \) meshes are used in FWI. Without the \( aAMR \) meshes the model parametrization can be kept on a single inversion mesh, which can be used in both forward and adjoint modelling for all the (adjoint) sources. Hence, there is no distinction between the discrete representation of the model space for the inversion and the simulations. This makes it trivial to sum different gradients and to update the model.

This procedure is not an option for the \( aAMR \) approach since it requires a unique simulation mesh for each source. It is thus a necessity to discretize the model independently of the wavefields which has been done in tomography studies (e.g. Aki & Lee 1976; Li & Romanowicz 1995; Wang & Dahlen 1995; Ritsema et al. 2011; French & Romanowicz 2014). There are multiple options available to store the model parametrization. Using an inversion mesh is a subjective choice of parametrization often used for convenience. The approach used in this study is to apply a rectilinear spectral-element mesh to store the velocity model. In each iteration, the new model is interpolated onto the simulation meshes for the forward and adjoint simulations. Subsequently, the resulting gradients are interpolated back to the inversion mesh. The complete workflow is illustrated in a flowchart in Fig. 3. The additional steps in the workflow which are required to use the \( aAMR \) meshes are explained in more detail in the coming sections.

#### 3.1 Building meshes for individual sources

One of the main extensions to standard FWI is that each source requires a unique mesh. The meshing only needs to be carried out once
per source, and so it does not impose a significant computational burden.

In 2-D, each mesh is created by meshing the domain with a regular grid in polar coordinates. The origin of the coordinate system is placed at the source location. To avoid singular elements, a rectilinear grid is built around the source location, replacing the polar grid in a small region. The radial dimension of the polar grid is discretized based on the minimum modelled period. The discretization in the azimuthal dimension is, however, a free parameter governed by (i) the complexity of the medium and (ii) how well a circular wavefield needs to be approximated. The latter can be achieved with fewer elements if the shapes of the elements can follow higher-order polynomials, rather than being restricted to straight lines. The number of elements in the azimuthal direction is kept constant, as a function of radius, as seen in Fig. 1. This is, however, not a restriction, as we will demonstrate in Section 5. As a circular domain around a source location is not a realistic domain of application, absorbing boundaries are often required. The implementation of absorbing boundary conditions are no different for aAMR meshes compared to other meshes. In this study we implement a combination of the absorbing boundaries described in Clayton & Engquist (1977) and Kosloff & Kosloff (1986).

### 3.2 Interpolation

When an inversion discretization, that is an inversion mesh or a Fourier basis, is used for storage and representation of the medium, interpolation to and from the various aAMR meshes becomes necessary. The main objective when defining the operator which interpolates from the inversion discretization onto the simulation meshes is to preserve the properties of the medium which the wavefield is sensitive to, that is, the effective medium. Once the interpolation operator is defined, the adjoint interpolation operator is used to map the computed gradients back to the inversion discretization. In the numerical examples studied in Section 4, a pointwise evaluation of the GLL-basis of the model on the inversion mesh is sufficient to produce an aAMR representation for numerical simulations. However, this may not generally be sufficient, as we will further discuss in Section 6.3.

### 4 NUMERICAL EXPERIMENTS

To test the method, we constructed a 2-D synthetic example with a homogeneous, isotropic starting model and attempted to reconstruct a random medium with up to ±8 per cent Lamé parameter (μ and λ) deviations from the initial model. The random target medium was computed using the Fourier method (Igel & Gudmundsson 1997; Meschede & Romanowicz 2015).

#### 4.1 Experimental setup

The domain is 1400 km by 1400 km wide, with 9 sources distributed regularly between 400 and 1000 km in each direction. Receivers are placed every 100 km in an 800 km by 800 km region. Source–receiver distribution used in the experiment is illustrated in Fig. 4. The sources are all moment tensor sources, and artificial data is calculated using a regular grid rectilinear mesh designed to resolve 5 s period elastic waves. The source time function is a Ricker wavelet, and each source was simulated for 250 s using the spectral-element wave propagation solver Salvus (Afanasiev et al. 2019). We conducted the synthetic inversion experiment using the proposed workflow with aAMR meshes which have a varying number of elements in azimuthal direction.

#### 4.2 Forward modelling validation

In any waveform inversion one knows the experimental setup prior to simulating and thus has the option of running benchmark simulations for that setup. This step should be performed before the first iteration and additionally whenever there are significant changes to the model or modelled frequency content. Given the limited number of required simulations for the benchmarking, the cost of it is negligible compared to the inversion.

We analysed forward modelling errors through the true medium and compared the results from the various meshes used in the experiment. The wavefield was sampled at 80 regularly spaced receivers. At each receiver, the absolute maximum amplitude calculated on the rectilinear mesh was used as a normalization factor prior to misfit calculation. The normalization was done to make each receiver and each source equally important, removing the effect of source–receiver distance and magnitude on the misfit. The $L_2$ waveform error was computed component wise,\[ \chi(d_{obs},d_{pred}) = \frac{1}{2} \int_T \left[ \frac{d_{obs}(t) - d_{pred}(t)}{\max |d_{obs}|} \right]^2 dt, \] (3)

where $d_{obs}$ is the artificial data computed on a fine rectilinear mesh, $d_{pred}$ is the synthetic data and $T$ is the time window, the error was averaged across the receivers. To test the effect of medium complexity, we simplified the true medium through filtering. For this, we performed a 2-D Fourier transform to obtain the wavenumber domain medium, and then multiplied the spectrum with a centred flat circular filter with a Gaussian taper at the edges. An inverse 2-D Fourier transform was used to acquire the filtered medium.
and the misfit had been reduced by 88 per cent compared to the initial model. The maximum wavenumber (complexity) existing in the new mesh was controlled by varying the extent of the circular filter in the azimuthal elements in the aAMR meshes. The reconstructed models along with the respective meshes can be seen in van Driel et al. (2020).

4.3 Inversion

As a baseline, we inverted for the random medium starting from a homogeneous model using the conventional approach with the same rectilinear grid as was used to calculate the artificial data (noise free). The mesh contains 20,736 elements, and we did 30 trust-region L-BFGS (Liu & Nocedal 1989; Nocedal & Wright 2006) iterations, reducing the waveform misfit in each iteration. We simultaneously inverted for shear modulus, \( \mu \), and the Lamé parameter \( \lambda \), keeping density, \( \rho \), constant. We compare the shear modulus models using the \( L_2 \) model misfit

\[
\chi(M_1, M_2) = \int_\Omega \left[ M_1(x) - M_2(x) \right]^2 \, dx,
\]

where \( M_1 \) and \( M_2 \) are two separate models, \( x \) is a point in the models and \( \Omega \) is the area of the models. After the last iteration, the \( L_2 \) model misfit had been reduced by 88 per cent compared to the initial model and the \( L_2 \) waveform misfits, eq. (3) compared to the artificial data had been reduced by 99.6 per cent from the starting model.

To test the new method we attempted the same type of recovery for various versions of the aAMR meshes. We varied the number of elements used to cover the azimuthal circumference around the source from 8 to 80. The number of elements in the respective meshes varied from 820 to 5782. We analysed the quality of model recoveries as a function of simulation cost and iterations (Fig. 6). The actual value of the computational speed-up is problem-dependent and correlates with the size of the domain and the maximum frequency, while it is anticorrelated with medium complexity.

As Fig. 6 shows, increasing the number of azimuthal elements above 32 does not lead to significant improvements in model reconstruction and mostly leads to an increase in computational load (see Table 1). All the meshes with 32 azimuthal elements or above manage to reduce the model misfit by roughly the same amount as the regular FWI. The waveform misfit reductions vary from 99.0 to 99.6 per cent, and the model misfit reductions vary from 84 to 88 per cent. The computational gain achieved by using the aAMR meshes is visualized on the left-hand panel of Fig. 6, where the evolution of the model misfit is plotted as a function of simulation time.

The reconstructed models along with the respective meshes can be seen in Fig. 7. The differences between the rectilinear reconstruction and the aAMR reconstructions, which use 48 or more elements, are visually indistinguishable. For 24–32 azimuthal elements, the reconstruction looks similar to the rectilinear one. For fewer azimuthal elements the final images start to deviate significantly from the rectilinear recovery. The quantitative analysis of the model differences in Fig. 8 supports that observation and reveals that the 16 azimuthal element reconstruction mostly catches the large-scale features of the model but not the fine-scale details. The other models, however, capture both the fine- and the large-scale features.

Fig. 6 reveals that the errors in model reconstruction, seen in Figs 7 and 8, correspond to the forward modelling errors for the different aAMR meshes through the true medium. This indicates that the errors in the reconstructions relate to the errors in the forward wavefield and interpolations between grids (steps 3 and 4 in the wavefield modelling box and step 3 in the gradient calculation box in Fig. 3).

5 TOWARDS 3-D

While this proof of concept is focused on 2-D, the longer-term goal is a 3-D global-scale application, already hinted at by the 3-D forward modelling experiments of van Driel et al. (2020). Here, we provide a first step in this direction in the form of global-scale kernel calculations.

The extension of the aAMR meshing to 3-D is conceptually related to the AxiSEM approach (Nissen-Meyer et al. 2008). Given that we are designing a mesh to represent a wavefield propagating away from a source at an arbitrary depth below the North Pole, a D-shaped AxiSEM mesh is extruded, rotating around a vertical axis through the poles. To avoid singular elements at the symmetry axis, a number of elements around the axis are removed and replaced by a cylinder which is meshed as a disc which is extruded between the poles. The discretization in the radial and polar directions is controlled by the frequency which needs to be resolved, while the discretization in the azimuthal direction is, similar to the 2-D case, controlled by the medium/source complexity and the need to approximate a sphere accurately. If the source is located at some other location than one of the poles, the discretization is rotated in order to make the cylinder go through the source location. Fig. 9c) shows a 3-D global aAMR mesh with 4th-order shape maps of the elements (van Driel et al. 2020, section 4.2). The surface elements are refined 30° from the source location to resolve surface waves better. This has the additional benefit of better point approximations for receiver locations.

Table 1. A comparison of model misfit (MM) between final model and true model, normalized by misfit between homogeneous starting model and final model, waveform misfit reduction (WMR), the simulation time (ST) used to achieve these misfit reductions and the relative speed-up achieved by using the aAMR meshes (SU) with the same time-step. The Rect mesh is the classic method while the numbers in the Mesh column represent the azimuthal elements in the aAMR meshes.

| Mesh | MM [per cent] | WMR [per cent] | ST [CPUhrs] | SU |
|------|--------------|----------------|-------------|----|
| Rect | 12           | 99.66          | 66.67       | –  |
| 8    | 1318         | 27.64          | 2.62        | 25.45 |
| 16   | 125          | 88.24          | 4.35        | 15.33 |
| 24   | 26           | 97.87          | 6.20        | 10.75 |
| 32   | 16           | 99.03          | 8.07        | 8.26  |
| 48   | 14           | 99.45          | 11.63       | 5.73  |
| 64   | 12           | 99.57          | 15.23       | 4.38  |
| 80   | 12           | 99.58          | 18.57       | 3.59  |

The maximum wavenumber (complexity) existing in the new medium was controlled by varying the extent of the circular filter in the azimuthal direction. The complexity and compared the computed errors for the varying aAMR meshes. The results are displayed in Fig. 5. It shows how the meshes with 48 or more azimuthal elements perform similarly well for the unfiltered medium but as the medium becomes less complex, the mesh with 32 elements achieves similar accuracy as the others. The mesh with 24 azimuthal elements gets close to the others as the medium becomes simpler, but the 16 azimuthal element mesh does not get below an order of magnitude higher errors than the others in spite of the increasing simplification of the medium. That indicates that the dominating factor in the errors associated with the 16 azimuthal element mesh is the limitation of the linear element shapes and how well they can approximate a circular wavefield. An overview on the effect of higher order polynomial element shapes can be seen in van Driel et al. (2020).
We computed a single source–receiver pair gradient on a standard global cubed sphere mesh, as well as on a 3-D $aAMR$ mesh. These are shown in Figs 9(a) and (d), respectively. The meshes are designed to resolve a minimum period of 50 s at 2 elements per minimum wavelength, and they both have a 1-D isotropic velocity model (Dziewonski & Anderson 1981). The cubed sphere has 1.699 840 elements, while the $aAMR$ mesh has 120 416. Representative slices through the $P$-wave velocity gradients of the first-arrival $P$-wave traveltime are displayed in Figs 9(b) and (c). Due to the differences in model parametrization, the gradients, that is, the projection of sensitivity kernels onto the different basis functions, do not look identical. However, as discussed in Section 2, both are correct.
Figure 8. Left-hand panel: mean of pointwise absolute model misfits between rectilinear recovery and various aAMR recoveries. The misfits are normalized by the shear modulus of the homogeneous medium to make the misfit relate to deviations from the homogeneous medium. The true model and the homogeneous model would have an average absolute misfit of about 2 per cent while the perturbations range from –8 to 8 per cent (in order of expected global tomography deviations). Models are lowpass filtered in the wavenumber domain, the horizontal axis displays the minimum spatial wavelength in the filtered model. The shaded regions represent half a standard deviation. Right-hand panel: the histogram of the pointwise model misfit measurements for the unfiltered rectilinear model.

Figure 9. (a) A typical global 3-D cubed sphere hexahedral mesh (Ronchi et al. 1996) where element dimensions are controlled by the S-wave velocity. (b) A global 3-D aAMR mesh with the radial and the polar dimensions identical to the cubed sphere mesh, but elongated elements in azimuthal direction. Also note the cylindrical mesh along the axis that avoids ill-shaped elements (similarly to the 2-D examples) and the surface refinements for better resolution of surface waves which can be seen at 30° latitude. (c, d) Bodywave gradients calculated from a single source–receiver pair on meshes (a) and (b), respectively. Note that, as in the 2-D case, the main difference is close to the receiver location (triangle).
discrete gradients corresponding to their respective parametrizations. Similar to the 2-D case, the main difference is near the receiver locations. As demonstrated in the numerical example, combining many of these gradients should conclude the iteration with approximately the same model update at a greatly reduced computational cost. Computing the aAMR gradient was over 14 times cheaper computationally, meaning that one can perform roughly 14 iterations at the cost of one, by using aAMR meshes.

6 DISCUSSION AND CONCLUSIONS

We presented a first proof of concept, showing that wavefield adapted meshes can be used in an FWI workflow, potentially leading to significant computational savings. In the following paragraphs, we discuss limitations of the method, current and future domains of applicability, as well as further technical issues such as homogenization for improved interpolation, and the effective frequency scaling.

6.1 Limitations of this approach

The usage of aAMR meshes is an add-on to spectral-element methods. The applicability and potential speed-ups depend on the specific problem and whether it meets the assumptions of the aAMR meshes. It is thus essential to clearly outline the niche where this approach is likely to be beneficial. For this, we recall that the concept of wavefield adapted meshes rests on the assumption that our
prior knowledge on the geometry of wave fronts and complexity of the wavefield can be approximated by the mesh.

In 2-D, this is the case, for instance, when the medium is sufficiently smooth to avoid significant scattering. A comparison of the usage of 2-D aAMR meshes in a smooth medium versus a medium with a sharp discontinuity is shown in Fig. 10. A more quantitative analysis was performed by inserting a circular inclusion with radius \( R \) (100 km) into the centre of a homogeneous background medium \( H \), where the magnitude \( S \) of the perturbation \( P \) varies between 0.05 and 0.7. Furthermore, we apply a Gaussian taper \( T \) to vary the smoothness of the interface. The model is then given by

\[
P = (1 + T(r))SH,
\]

where \( r \) is defined as the distance from the centre of the inclusion and \( T \) is defined as

\[
T(r) = \begin{cases} 
\frac{e^{-r^2}}{2\sigma^2}, & \text{if } r > R \\
1, & \text{if } r \leq R,
\end{cases}
\]

where the standard deviation, \( \sigma \), of the taper is what defines the smoothness of perturbation \( P \). Marginals from that analysis are shown in Fig. 11. It demonstrates that the approach requires smooth media because scattering breaks the approximate azimuthal symmetry of the wavefield. The approach is thus not directly applicable for an exploration scale study where reflected waves are the dominant quantity of interest.

In 3-D and at regional to global scales, the Earth is roughly spherically symmetric. The reflected waves from the global internal discontinuities preserve the azimuthal symmetry; a fact that is exploited by the AxiSEM approach (Nissen-Meyer et al. 2007, 2008, 2014; Leng et al. 2016, 2019). Thus, approximately spherical discontinuities do not pose a problem for aAMR meshes in 3-D. Similar to 2-D, the deviations from the approximately spherically symmetric background must, however, be smooth enough to avoid significant off-great-circle scattering. The precise meaning of ‘significant’ depends on the actual data one wishes to exploit, and therefore must be assessed on a case-by-case basis, for example using a careful forward modelling study along the lines of Fig. 5.

6.3 Homogenization prior to interpolation

The simulation meshes have, like the wavefields, variable spatial resolution. Smaller-scale structure in some regions may thus be well represented on the fine inversion mesh but less so on a potentially coarser simulation mesh. This undersampling may result in spatial aliasing when the model is interpolated from the inversion onto the simulation meshes. To ensure that the wavefield ‘sees’ the effective medium in both discretizations, the model should ideally be homogenized for each simulation mesh prior to interpolation.

Several homogenization approaches for wave propagation have been developed in recent years (e.g. Capdeville & Marigo 2007; Capdeville et al. 2013; Fichtner & Hanasoge 2017; Cupillard & Capdeville 2018). An approximate alternative to homogenization is space-dependent anisotropic smoothing, where, in our specific case, the medium is smoothed primarily in azimuthal direction. This is visualized as the second step in the wavefield modelling box in Fig. 3. The reasoning behind space-dependence and anisotropy of the smoothing is that the element edge lengths vary both as a function of space and dimension, and the risk of spatial aliasing is a function of model complexity and element edge lengths. Though we expect homogenization or smoothing to become relevant in future applications, we empirically found that straightforward sampling of the medium on the Gauss–Lobatto–Legendre points was sufficient for the examples presented in Section 4.
6.4 Frequency scaling

An important feature of the aAMR meshes is how the number of required elements scales with frequency. As mentioned in the introduction, the number of elements required in the current standard meshes, scales with frequency to the power of their respective dimension (N). The aAMR meshes, however, have one dimension which is quasi-independent of frequency. For a homogeneous medium, the number of elements thus scales with frequency to the power of N − 1. Fig. 12 illustrates this relation where a meshing algorithm is run to generate meshes for various frequencies and compared to theoretical scaling laws. This relation, however, only strictly applies to a homogeneous medium and as the medium gets more complex, the aAMR curve on Fig. 12 approaches the conventional mesh curve. To quantify how fast it approaches it would require a general theoretical relation between frequency and azimuthal complexity of the wavefield, which unfortunately does not exist. It has, however, been investigated numerically by Leng et al. (2016, 2019) where it was found that in a period range where global seismic tomographic models currently exist, 34 to 5 s, the frequency has a minimal effect on the azimuthal complexity of the wavefield.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to editors Louise Alexander and Carl Tape, as well as two anonymous reviewers for their fruitful comments and help with the publication of this paper. This work was supported by the European Unions Horizon 2020 research and innovation programme through an ERC Starting Grant (The Collaborative Seismic Earth Model, grant No. 714069) and a Centre of Excellence (CheESE, grant No. 823844), and by the Platform for Advanced Scientific Computing (PASC, project Salvis). We gratefully acknowledge support from the Swiss National Supercomputing Center (CSCS) in the form of computing time grants c13 and s868. Data visualizations were done using Paraview (Ahrens et al. 2005 and Matplotlib (Hunter 2007), data processing was done using Numpy (Oliphant 2006), Obspy (Beyreuther et al. 2010; Krischer et al. 2015) and ASDF (Krischer et al. 2016). The inversion workflow was developed using Jupyter Notebooks (Kluyver et al. 2016). The inversion algorithms as well as meshes used in this study are available in a Github repository (https://github.com/solvthurstar/multimtse_inversion_2d).

REFERENCES

Afanasiev, M., Boehm, C., van Driel, M., Krischer, L., Rietmann, M., May, D.A., Knepley, M.G. & Fichtner, A., 2019. Modular and flexible spectral-element waveform modelling in two and three dimensions, Geophys. J. Int. 216(3), 1675–1692.

Afanasiev, M.V., Peter, D.B., Sager, K., Simute, S., Ermert, L., Krischer, L. & Fichtner, A., 2016. Foundations for a multiscale collaborative Earth model, Geophys. J. Int., 204, 39–58.

Ahrens, J., Geveci, B. & Law, C., 2005. Paraview: an end-user tool for large data visualization, Vis. Handb., 717–731.

Aki, K. & Lee, W.H.K., 1976. Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes - 1. A homogeneous initial model, J. geophys. Res., 81, 4381–4399.

Aki, K., Christoffersson, A. & Husebye, E.S., 1977. Determination of the three-dimensional seismic structure of the lithosphere, J. geophys. Res., 82(2), 277–296.

Arney, D.C. & Flaherty, J.E., 1989. An adaptive local mesh refinement method for time-dependent partial differential equations, Appl. Numer. Math., 5(4), 257–274.
Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**, 1259–1266.

Tarantola, A., 1988. Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation, in *Scattering and Attenuations of Seismic Waves, Part I*, pp. 365–399, Springer.

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

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

van Driel, M., Boehm, C., Krischer, L. & Afanasiev, M., 2020. Accelerating numerical wave propagation by wavefield-adapted meshes, Part I: forward and adjoint modelling, *Geophys. J. Int.*, doi:10.1093/gji/ggaa058.

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

Wang, Z. & Dahlen, F.A., 1995. Spherical-spline parameterisation of three-dimensional Earth models, *Geophys. Res. Lett.*, **22**, 3099–3102.

Warner, M. et al., 2013. Anisotropic 3D full-waveform inversion, *Geophysics*, **78**, R59–R80.

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