Efficiency of the spectral element method with very high polynomial degree to solve the elastic wave equation

Chao Lyu, Yann Capdeville, Liang Zhao

Efficiency of the spectral element method with very high polynomial degree to solve the elastic wave equation

Lyu Chao[1,2], *Capdeville Yann*[1], *Zhao Liang*[2]

[1] *Laboratoire de Planétologie et Géodynamique de Nantes, CNRS, Université de Nantes*

[2]*State Key Laboratory of Lithospheric Evolution, Institute of Geology and Geophysics, Chinese Academy of Sciences, 100029 Beijing, China*

email : lyuchao1988@gmail.com

right-running head: Efficient high degree spectral element

ABSTRACT

The Spectral Element Method (SEM) has gained tremendous popularity within the seismological community to solve the wave equation at all scales. Classical SEM applications mostly rely on degrees 4 to 8 elements in each tensorial direction. Higher degrees are usually not considered due to two main reasons. First, high degrees imply large elements, which makes the meshing of mechanical discontinuities difficult. Second, the SEM's collocation points cluster toward the edge of the elements with the degree, degrading the time marching stability criteria, imposing a small time step and a high numerical cost. Recently, the homogenization method has been introduced in seismology. This method can be seen as a pre-processing step before solving the wave equation which smooths out the internal mechanical discontinuities of the elastic model. It releases the meshing constraint and makes the use of very high degree elements more attractive. Thus, we address the question of memory and computing time efficiency of very high degree elements in SEM, up to degree 40. Numerical analyses reveal that, for a fixed accuracy, very high degree elements require less computer memory than low degree elements. With minimum sampling points per minimum wavelength of 2.5, the memory needed for a degree 20 is about a quarter that of the one necessary for a degree 4 in 2-D and about one eighth in 3-D. Moreover, for the SEM codes tested in this work, the computation time with degrees 12 to 24 can be up to twice faster than the classical degree 4. This makes SEM with very high degrees attractive and competitive for solving the wave equation in many situations.

INTRODUCTION

Solving the acoustic and elastic wave equations numerically is a critical step for many research based on seismic data. It is especially important for most seismic imaging methods, from the exploration scale to the global Earth scale. Even though the Finite Difference (FD)

method still dominates in the seismic exploration community, the spectral element method (SEM) (Maday and Patera, 1989; Seriani and Priolo, 1994; Komatitsch and Vilotte, 1998; Komatitsch and Tromp, 1999; Chaljub et al., 2007) has been gaining more and more popularity, especially in the academic community. It is often the chosen method for global or regional Earth scale seismic imaging developments based on adjoint methods and full waveform inversion method (Capdeville et al., 2005; Tromp et al., 2005; Fichtner et al., 2009; Zhu et al., 2012; Monteiller et al., 2015; Wang et al., 2016; Beller et al., 2018; Trinh et al., 2019).

SEM is a finite element type method and, as such, is based on the weak form of the wave equation. Two versions of the SEM exist, one based on Chebyshev polynomials and one based on Legendre polynomials. In the following, we only use the Legendre version, which is the basis for many available programs. Compared to classical low-degree finite elements, the Legendre SEM is based on a tensorised high-degree polynomial approximation per element combined with a precise numerical quadrature associated with the so-called Gauss-Lobatto-Legendre (GLL) points. It has a spectral convergence with the element polynomial degree, thus leading to a low spatial dispersion. Moreover, its tensorial formulation leads to a diagonal mass matrix, which is a strong advantage for explicit time schemes. It can naturally and accurately handle free surface and material discontinuities. This capability to accurately model interface waves such as surface waves is one of the main reasons for its popularity.

Despite its quality, the SEM has one critical drawback: to warrant the accuracy of the method, each material discontinuity interface has to be explicitly meshed. Moreover, because this method is based on a tensorial formulation, the mesh needs to be based on quadrilateral (in 2-D) or hexahedron (in 3-D) elements. The meshing difficulty can strongly limit the applicability of the method, especially in 3-D. Note that some examples of SEM based on triangles exist but their efficiency is lower and with a limited polynomial degree range (Komatitsch et al., 2001; Mercerat et al., 2006; Afanasiev et al., 2018).

One of the key parameters when using SEM is the polynomial degree N used in each tensorial direction for each element. In seismology, $N = 4$ is often chosen (Komatitsch and Vilotte, 1998; Komatitsch and Tromp, 1999) and sometimes $N = 8$ (Chaljub et al., 2003; Capdeville et al., 2003). Although it depends on the desired accuracy, the propagated distance and on how the minimum wavelength is estimated, the commonly admitted number of grid points per minimum wavelength (G) to obtain sufficient accuracy in a constant velocity medium is approximately $G = 5 \sim 6$ for $N = 4$ and $G = 4 \sim 5$ for $N = 8$ (Priolo and Seriani, 1991; De Basabe and Sen, 2007; Seriani and Oliveira, 2008). A low G is obviously an advantage as it lowers the required computer memory and maybe the computing cost of a modeling. In particular, a low G can be important in the full waveform inversion (FWI) context. Indeed, FWI schemes operate on the forward and adjoint wavefield. To do so, some methods rely on the partial or compressed storage (Komatitsch et al., 2016; Boehm et al., 2016) or even full storage to avoid backward propagation (Fichtner et al., 2009). In 3-D, dividing G by 2 implies a storage requirement divided by 8, which is significant. In the 1-D case, it has been shown, for $N = 60$, G can be as low as about $2 \sim 2.5$ (Priolo and Seriani, 1991). Nevertheless, the most widely used degree N is still 4 not 60, which is for two reasons:

- Taking advantage of a large N and a low G imposes being able to use very large elements. For example, an element for $N = 60$ needs to be about 15 times larger than an element for $N = 4$ (assuming a constant $G(N)$ for the sake of simplicity). In most realistic situations, this is not possible because of the necessity to mesh mechanical discontinuities and the domain geometry;

- The GLL points are not evenly spaced. They cluster toward the edges of the elements and this clustering is stronger with large N as shown in Figure 1 (see Appendix A for the detailed formula). Indeed, in a d dimensional space, each element e of the mesh is a deformed version through a transformation $\mathcal{F}_e$ of the reference element Λ^d

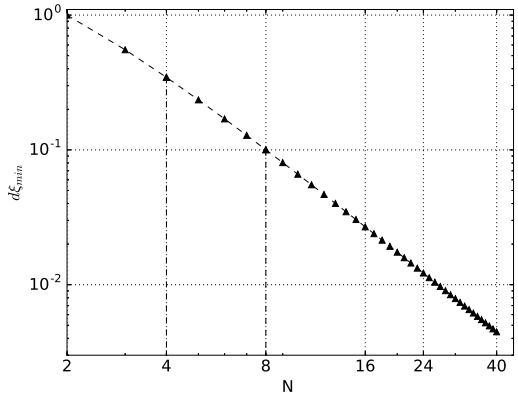

Figure 1: Minimum distance $d\xi^N_{\text{min}}$ between two GLL points for the 1-D reference element $\Lambda = [-1, 1]$ as a function of the degree N with logarithmic scales.

where $\Lambda = [-1, 1]$. The GLL points ξ^N_i for $i \in \{0...N\}$ are non-uniformly spaced along the reference segment Λ and cluster near the edges. When N becomes larger and larger, the smallest grid distance ($d\xi^N_{\text{min}} = \xi^N_1 - \xi^N_0$) between the two nearest GLL points becomes smaller and smaller. Once mapped into the physical elements, this $d\xi^N_{\text{min}}$ becomes a $dx_{\text{min}} = \min_e\{\mathcal{F}_e(d\xi^N_{\text{min}})\}$, which imposes a smaller and smaller time step dt to fulfill the stability condition of the explicit time scheme classically used (e.g. Komatitsch and Vilotte (1998)). The Courant–Friedrichs–Lewy (CFL) stability condition is governed by the minimum value of the ratio between the size of the grid cells and the P-wave velocity expressed as:

$$dt \leq C\frac{dx_{\text{min}}}{V_P}\,, \tag{1}$$

where C is the courant constant and V_P is P-wave velocity. In the following, we refer to dt_{CFL} as the largest stable time step of the Newmark scheme for a given test case.

As a consequence, the possibility to use very high degree spectral elements has not been considered as a valid option and has not been studied yet unlike for finite differences method (Liang et al., 2015). Nevertheless, in the forward modeling context, the recent

introduction of the non-periodic homogenization (Capdeville et al., 2010; Cupillard and Capdeville, 2018) has solved many of the meshing difficulties and opens the door to the use of very high degree N. In the inverse problem context, models are generally smooth and using very high degree elements has always been an option.

The non-periodic homogenization is an asymptotic method designed to compute effective media in the case of deterministic media with no scale separation such as geological media. For a given maximum source frequency and a complex "true" medium, this method computes a smoothly varying effective medium in which the computed waveform is the same as a reference solution computed in the true medium, up to a controllable error. This is true for all type of waves, including surface, refracted, backscattered, etc, waves. In the forward modeling context, homogenization can be seen as a preprocessing step upscaling original media to the wavelength scale.

Regarding SEM, homogenization solves some aspects of the meshing issue: indeed, after homogenization, there is no more mechanical discontinuity to honor, except for the free surface. This makes the option of using very large elements with a high degree N possible, opening the door to a lower G than for the classical degree $N = 4$ and therefore allowing a lower memory. Nevertheless, about the computing cost, the fact that GLL points cluster toward the edges of the elements is still true and will go against the advantages of a lower G. The objective of this work is, therefore, to determine if using very high degree N can be an advantage, for both memory and computing time aspects for SEM and if so, to find if there is an optimal degree N.

The paper is organized as follows: we first give information about the different SEM codes, the criteria for the different test configurations and the methodology we use to address the paper's objective. We then perform our tests in 2-D and 3-D homogeneous media. Finally, we perform the same test in simple heterogeneous media before concluding our work. For each program, we limit ourselves to the standard order 2 Newmark scheme for

the explicit time marching.

SEM CONSIDERATION AND NUMERICAL EXPERIMENTS SETUP

SEM complexity consideration

As shown in the Appendix A, for a given accuracy, if the maximum time step allowed by the CFL condition given in equation 1 can be reached, the complexity $C(N)$, that is the number of operations of any SEM program, as a function of the degree N, is scaling as

$$C(N) \propto G^{d+1}(N)N^2\,, \tag{2}$$

where d is the dimension of the problem. If the maximum time steps cannot be reached (typically for long time series for which the time step error can be large), the complexity scales as

$$C(N) \propto G^{d}(N)N\,. \tag{3}$$

In both cases, it shows that the decreasing G as a function of N is competing with the increasing N. Knowing that the decrease of G as a function of N is non-linear and cannot go below 2, it is already clear that there is a limit after which increasing N will not be efficient. But before that, an optimal N exists and needs to be determined numerically, which is the purpose of the next sections.

SEM codes

For a given source maximum frequency f_{max}, geometry, elastic properties, and signal duration t_{max}, there are two aspects in assessing whether using the very high degree in SEM

can be a good idea. First, we consider the amount of computer memory required to perform the simulation. This is directly related to the number of points per wavelength G and does not depend much on the particular code implementation of SEM. Second, we consider the computing time needed to perform the simulation. This is strongly related to code implementation. To mitigate a possible bias in our results, we use three different programs:

- SPEC2DY. This program has originally been written by G. Festa (Festa and Vilotte, 2005) and has been widely modified over the years, but the philosophy and the core of the original program are still the same. In particular, its efficiency has not been seriously optimized.
- SEM3D: This program is a 3-D version of the above 2-D program. It has been originally written by G. Festa and E. Delavaud (Delavaud, 2007). In contrast to its 2-D version, its efficiency has then been optimized by a team of the CEA (Commissariat ï£¡ l'ï£¡nergie atomique, France);
- SPECFEM (2D and 3D). These two programs are from the popular SPECFEM program suites, both the 2-D and 3-D versions are well maintained and optimized (Komatitsch and Vilotte, 1998; Komatitsch and Tromp, 1999).

Each of these three codes uses an order 2 Newmark time marching scheme.

Numerical experiments setup and models

In order to measure the numerical efficiency of SEM with degree N, we set up a series of 2-D and 3-D numerical experiments in homogeneous and heterogeneous elastic models. To make the efficiency measurements meaningful, we need to perform our experiments for a fixed accuracy. There are many ways to measure accuracy and we choose one that is close to many realistic situations: a rectangular domain Ω with a free surface on the top, shallow sources and shallow receivers with recorded waveforms in the time domain.

We perform our tests in three different models: 2-D homogeneous, 3-D homogeneous and 2-D heterogeneous. We did not test the 3-D heterogeneous case because of large computing resources necessary to obtain an accurate reference solution in that case, but we believe that this does not detract from the main conclusion of this work. For the 2-D and 3-D homogeneous cases, we use $V_P = 3.4$ km/s and $V_S = 2.0$ km/s for the P and S wave velocities and $\rho = 2000$ kg/m^3 for the density. For the heterogeneous case, a monochromatic oscillatory heterogeneity is added on top of the homogeneous model. The detailed heterogeneity is defined in the 2-D heterogeneous experiment section below.

In the homogeneous case, the maximum frequency and S-wave velocity makes it possible to define the minimum wavelength of the propagating wavefield in the far-field as

$$\lambda_{\text{min}} = \frac{V_S}{f_{\text{max}}}. \tag{4}$$

In the following, we measure each spatial distances as a function of λ_{min} and time as a function of

$$t_{\text{min}} = \frac{1}{f_{\text{max}}}. \tag{5}$$

The source is a vertical point force located at a $3\ \lambda_{\text{min}}$ distance below the free surface and its time wavelet is a Ricker function (second derivative of a Gaussian) of central frequency $f_0 = 10$ Hz with maximum frequency $f_{\text{max}} \simeq 3f_0$. This estimation of the maximum frequency is important because it determines the minimum wavelength and therefore $G(N)$. This estimate can be changed and it will change the obtained $G(N)$ and it implies the absolute value of G must be taken with caution. Nevertheless, this estimate does not affect the relative results between different degrees.

To measure the error, we rely on two bins of receivers at two different epicentral distances, a short one ($20\lambda_{\text{min}}$) and a long one ($200\lambda_{\text{min}}$). Each bin is wide enough to contain at least one element, even for the largest degree N that we tested. The main reason for doing

so is that error is not constant within an element and depends on where exactly the receiver is located (Moczo et al., 2011). To average out this effect, we use many receivers within at least one element.

The boundary conditions are free normal stress conditions all around the domain to fully exclude the influence of absorbing boundaries. The chosen domain is wide enough to ensure that no reflecting waves from boundaries affect the results. The time duration of the signal is long enough to ensure that the full waveforms include P, S and the Rayleigh wave phases.

2-D homogeneous experiments

For the 2-D tests, $\mathbf{\Omega}$ is a $900 \times 450\ \lambda_{\text{min}}^2$ rectangular domain. Each of the different epicentral $20\ \lambda_{\text{min}}$ and $200\ \lambda_{\text{min}}$ receiver bin has a size of $15 \times 15\ \lambda_{\text{min}}^2$ and contains 2601 receivers. $\mathbf{\Omega}$ and an example of energy snapshot of the wavefield at $t = 270\ t_{\text{min}}$ is displayed in Figure 2. The recorded signal at each receiver last for $60\ t_{\text{min}}$ for the $20\ \lambda_{\text{min}}$ distance receiver bin and $300\ t_{\text{min}}$ for the $200\ \lambda_{\text{min}}$ distance receiver bin.

3-D homogeneous experiment

For the 3-D test, we use a homogeneous media with the same elastic properties as in the 2-D homogeneous case above. Here, because of computing resources limitation, we only use a relatively small model and a short distance receiver bins (20 λ_{min}). For this test, Ω is $300 \times 300 \times 150\ \lambda_{\text{min}}^3$ parallelogram. The receiver bin is a $15^3\ \lambda_{\text{min}}^3$ cube, just below the surface, $20\ \lambda_{\text{min}}$ away from the source (see Figure 3), containing 1331 receivers. The 200 λ_{min} receiver bin case has not been studied in 3D because the computing resources required to do so are beyond our capacity (indeed, to avoid domain border reflections, for the 200 λ_{min} case an even larger domain is required).

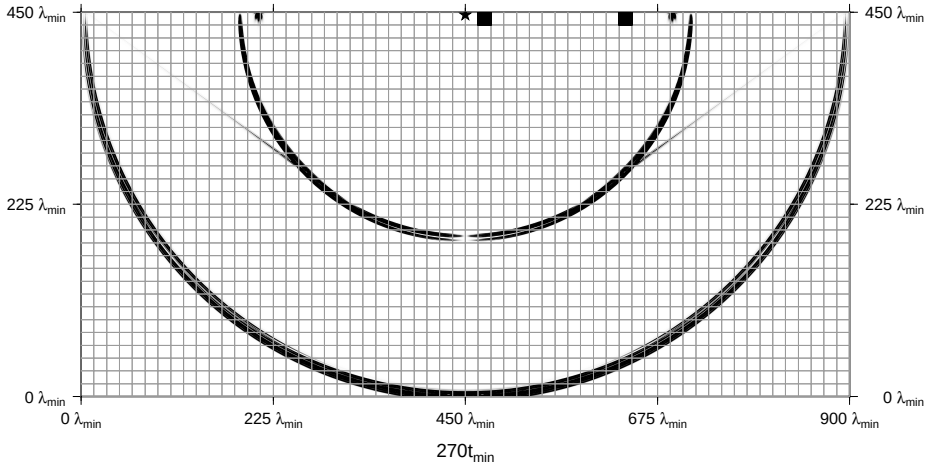

Figure 2: The 2-D homogeneous domain used in this work. The source (black star) is located about $3\lambda_{\min}$ below the surface. The short and long distance receiver bins (black square) are represented. The kinetic energy snapshot for $t = 270\, t_{\min}$ and a 60×30 elements mesh, associated to a polynomial degree $N = 40$, are also displayed.

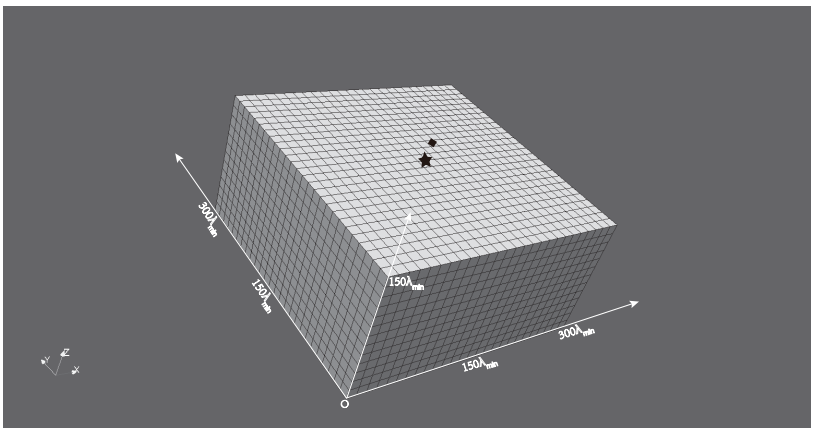

Figure 3: The 3-D homogeneous domain and mesh used in this work. The source position (black star) and the receiver bin (black square) are represented. The represented mesh has $26\times26\times13$ elements and is associated with a polynomial degree $N = 40$.

2-D heterogeneous experiment

One of the main arguments promoting the use of high-degree SEM is the fact that homogenization can remove the meshing difficulty by smoothing out all the internal mechanical discontinuities. The homogenized model is indeed smooth but its properties are oscillating with space. These spatial variation needs to be accurately integrated which usually implies a larger spatial sampling G than for homogeneous models. The smallest oscillations λ_{h} of the homogenized model depend on the heterogeneity spectrum of the original medium and on the desired accuracy. Typically, λ_{h} lies in $[\lambda_{\text{min}}/2, \lambda_{\text{min}}]$ (Capdeville et al., 2010).

In order to further analyze the efficiency of SEM as a function of the degree N in the heterogeneous 2-D case, we use the same geometry as for the 2-D homogeneous test and only focus on the long epicentral distance bin (200 λ_{min}). To keep our analysis simple, the heterogeneous mechanical properties have been chosen with the following form:

$$\begin{cases} \lambda(\mathbf{x}) = \lambda_0 f(\mathbf{x}) \\ \mu(\mathbf{x}) = \mu_0 f(\mathbf{x}) \\ f(\mathbf{x}) = 1 + a\big(cos(\frac{2\pi}{\lambda_h}\mathbf{k}_x \cdot \mathbf{x}) + cos(\frac{2\pi}{\lambda_h}\mathbf{k}_z \cdot \mathbf{x})\big) \\ \mathbf{k}_x \quad = (cos(t_1\pi/180^\circ), sin(t_1\pi/180^\circ)) \\ \mathbf{k}_z \quad = (cos(t_2\pi/180^\circ), sin(t_2\pi/180^\circ)) \end{cases} \tag{6}$$

where λ and μ are the Lamé elastic coefficients for the heterogeneous models, λ_0 and μ_0 the constant Lamé coefficients corresponding to $V_P = 3.4$ km/s, $V_S = 2.0$ km/s, $\rho = 2000$ kg/m^3. The density is kept constant. We use $a = 0.05$, $t_1 = 45^\circ$, and $t_2 = 135^\circ$, corresponding to two orthogonal directions shown in Figure 4. In the following, only three values for λ_{h} are tested, $\lambda_{\text{h}} = 2\lambda_{\text{min}}$, λ_{min} and $\lambda_{\text{min}}/2$. Note that a real model, once homogenized, has a continuous spectrum of heterogeneity, not just a single wavelength as here. In general, geological media lead to amplitude spectrum of heterogeneities that decreases as $1/\lambda$. $\lambda_{\text{h}} = \lambda_{\text{min}}/2$ in equation 6 correspond to a strongly heterogeneous case,

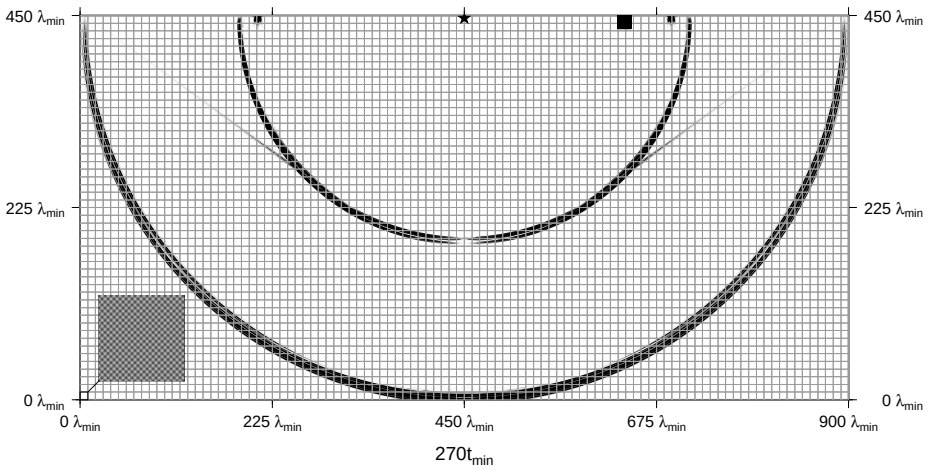

Figure 4: Same as Figure 2 but for the 2-D heterogeneous media with $\lambda_{\text{h}} = \lambda_{\text{min}}$. The lower left zoom displays a $14.1\lambda_{\text{min}} \times 14.1\lambda_{\text{min}}$ area of V_S. The 100×50 elements mesh associated to the degree $N = 40$ is displayed.

$\lambda_{\text{h}} = \lambda_{\text{min}}$ is representative of a typical geological model whereas $\lambda_{\text{h}} = 2\lambda_{\text{min}}$ corresponds to a full waveform inversion model.

SEM mesh

We base our experiments on a trivial regular mesh made of $n_x \times n_z$ square elements of size L_e in each direction. We keep $n_x = 2n_z$ and the domain size constant. The element size L_e can only be tuned by changing the number of horizontal elements n_x. In each element, the polynomial expansion of degree N is used in each direction. The mesh is fully characterized by the (L_e, N) quantities. For a given mesh (L_e, N), we can compute dx_{min},

the minimum distance between two GLL points:

$$dx_{min}(L_e, N) = \frac{L_e}{2} d\xi^N_{\text{min}} \,. \tag{7}$$

The number of grid points per minimum wavelength G in one direction is defined as follows:

$$G(L_e, N) = (N+1)\frac{\lambda_{\text{min}}}{L_e} \,. \tag{8}$$

Error measurement

There are several ways to measure the error of a given modeling with respect to a reference solution. One could for example compute the L^2 norm of the difference between computed and reference wavefield all over the domain for the final time step. Here, our choice is more closely related to situations faced in geophysics: receivers are located on the free surface or at shallow depths (in boreholes), with relatively long time series and for different epicentral distances. For a set of N_r receivers, located in $\{\mathbf{x}_r, r \in \{1..N_r\}\}$, the error E is computed as

$$E^2 = \frac{\sum_{r=1,N_r} \int_0^{t_{\text{max}}} (\mathbf{u} - \mathbf{u}^{\text{ref}})^2(\mathbf{x}_r, t)\, dt}{\sum_{r=1,N_r} \int_0^{t_{\text{max}}} (\mathbf{u}^{\text{ref}})^2(\mathbf{x}_r, t)\, dt} \,, \tag{9}$$

where $\mathbf{u}$ is the computed displacement including all the components, $\mathbf{u}^{\text{ref}}$ is the reference solution.

The numerical error for a given model and receiver bin is jointly affected by the mesh design (L_e, N) and the time step dt:

$$E = E(L_e, N, dt) \,. \tag{10}$$

Reference Waveforms

The reference solution $\mathbf{u}^{\text{ref}}$ should ideally be an analytical solution. Unfortunately, such analytical solutions are only available for simple models. In many cases, we need to rely on "converged" numerical solutions computed with a very small time step and spatially over-sampled numerical solution. Here, for the sake of completeness, we first show the equivalence of analytical and "converged" numerical solutions for 2-D and 3-D homogeneous half spaces. Afterward, we assume this equivalence holds for all our tests and all the reference waveforms are "converged" numerical solutions.

We first perform a 2-D comparison between a converged SEM solution and an analytical solution for the short distance receiver bin. The analytical solution was obtained using the program "EX2DDIR" (the source code can be found in `www.spice-thn.org`), which is based on the Caniard-de Hoop technique (De Hoop, 1960; Johnson, 1974). For the SEM solution, a 180×90 λ^2_{min} domain with a structural 100×50 spectral elements with degree 12 is designed. The resulting G value is about twice that recommended by Priolo and Seriani (1991). We finally use a very small time step, $dt \approx \frac{1}{250} dt_{\text{CFL}}$. The agreement between the two solutions is displayed in Figure 5 for a representative receiver. The total misfit, computed according to equation 9, is $E = 1.5 \times 10^{-6}$.

We then perform a similar test, but in 3-D. For the analytical solution, we use the program "CANHFS" (personal communication) which calculates the Green function and is also based on the Caniard-de Hoop technique. For the SEM solution, a 180×180×90 λ^3_{min} domain with a structural 100×100×50 spectral elements mesh with degree 12 is designed. The resulting G value is once again about twice that recommended by Priolo and Seriani (1991). We finally use a very small time step, $dt \approx \frac{1}{100} dt_{\text{CFL}}$. The agreement between the two solutions is displayed in Figure 6 for a representative receiver. The total misfit is here, is $E = 2.2 \times 10^{-3}$.

From these two tests, we conclude that we can replace the analytical solutions by con-

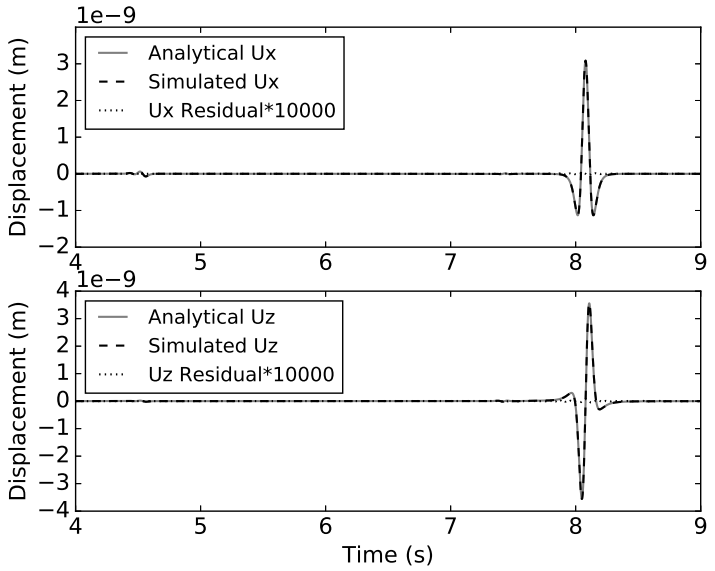

Figure 5: 2-D homogeneous model waveform displacement comparison between the analytical solution (gray line), the spectral element simulated solution (dashed line) for the horizontal (top plot) and vertical components (bottom plot). The residual ($\times 10^4$) is displayed in dotted line. The receiver is a typical receiver from the short distance bin ($20\ \lambda_{\text{min}}$ distance).

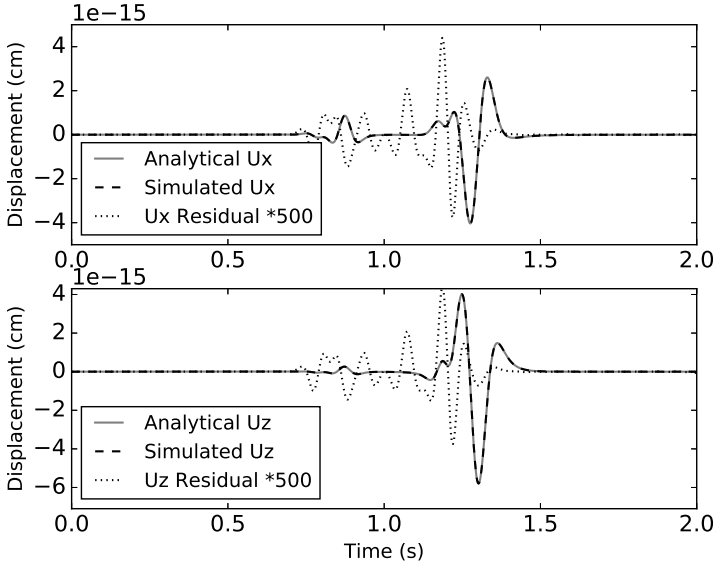

Figure 6: 3-D homogeneous model waveform displacement comparison between the analytical solution (gray line) and the spectral element simulated solution (dashed line) for the horizontal (top plot) and vertical component (bottom plot). The residual ($\times 500$) is displayed in dotted line. The receiver is a typical receiver from the short distance bin ($20\ \lambda_{\text{min}}$ distance).

verged numerical solution in our analysis. Indeed, our misfit threshold is 1%, and the misfit between converged and analytical solutions is much smaller than 1%. For the rest of the paper, we assume that we can use converged numerical solutions as reference solutions. Finally, note the fact that the agreement between numerical and analytical solutions is not as good in 3-D than in 2-D is more related to our difficulty to accurately use the analytical solution, rather than related to a larger error in the 3-D numerical modeling. The 3-D misfit is nevertheless still much smaller than 1% and we did not investigate the code problems any further.

G AS A FUNCTION OF THE DEGREE N

Here, we study the minimum number of GLL points per wavelength G required to reach the 1% error threshold. To do so, we start by choosing a time step dt small enough so that the error due to the time marching scheme can be ignored. The error is then dominated by the spatial error and only depends on L_e and N ($E = E(L_e, N)$). Knowing that a fixed size domain is used, the elements size L_e is controlled by the number of elements n_z in the z direction (the number of elements in the x direction is tied to n_z by the relation $n_x = 2n_z$). For given degree N, receiver bin ($20\lambda_{\text{min}}$ or $200\lambda_{\text{min}}$) and model, we proceed as follow to determine G so that the error $E = 1\%$:

- we start with a large n_z so that the error is below 1%;
- we then gradually decrease n_z until we find its values so that the error is just below and just above 1%;
- we compute G for each n_z available (discrete) values and we finally obtain the G to reach exactly 1% using a quadratic interpolation.

An example of the procedure for $N = 40$ is given in Figure 7.

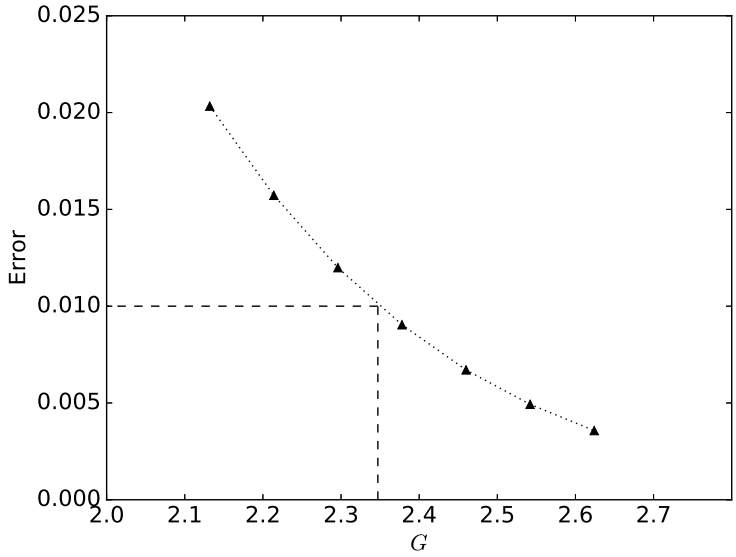

Figure 7: Error (see equation 9) as a function of G (dotted line) for the short distance receiver bin and for degree $N = 40$ and a very small dt. The triangles are measured values and the dashed line shows the quadratic interpolation estimation of G to obtain a 1% error. Here, $G \simeq 2.35$.

Following this procedure, we compute the 1% error G for a set of degree N ($4, 8, 12, 16, 20, 24, 30$ and 40), in our four 2-D models (homogeneous, heterogeneous with $\lambda_{\text{h}} = 2\lambda_{\text{min}}$, λ_{min} and $0.5\lambda_{\text{min}}$) and the 3-D homogeneous model for the two receiver bins. The results are given in Figure 8, Figure 9 and Table 1.

For the homogeneous case, it can be noted that, for $N = 40$, $G \simeq 2.5$ is reached in 2-D as well as in 3-D and for both short and long-distance bins. For the long-distance

Information for $200\lambda_{\text{min}}$ distance								
N	4	8	12	16	20	24	30	40
nz	505.9	186.2	111.7	79.9	61.9	50.3	39.5	28.9
$dx^{ref}_{min,4}/dx^{ref}_{min,N}$	1	3.45	7.40	12.85	19.82	28.28	43.81	77.20
$G_N/G_{N=4}$	1	0.662	0.574	0.537	0.514	0.497	0.484	0.469
$dt_{N=4}/dt$	1	1.270	1.634	2.028	2.426	2.812	3.419	4.413

Table 1: The number of Ns, nz (number of elements in z direction), minimum dx ratio and maximum dt ratio in 2-D homogeneous half space case for the 200 λ_{min} epicentral distances case.

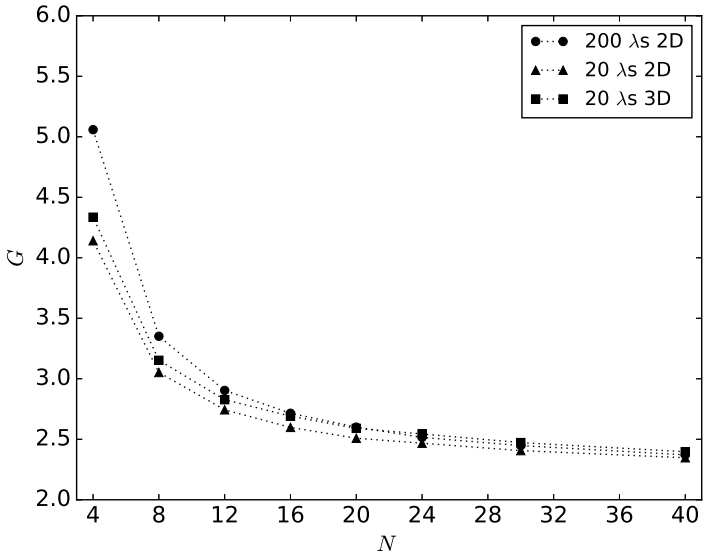

Figure 8: G as a function of the degree N for the 2-D and 3-D homogeneous models and for the 20 and 200 λ_{min} epicentral distance receiver bins.

bin, it corresponds to twice fewer points per wavelength for $N = 40$ than $N = 4$. This implies that factor of 4 in 2-D and 8 in 3-D of computer memory can be saved using high degrees compared to low degrees. As expected, G is lower for short distance than for long, but this difference is significant only for low degree. For the heterogeneous case, Figure 9 displays a comparison of the G obtained in the three heterogeneous models versus the one obtained in the homogeneous case. It is worth noting that the sampling rule of thumbs used in homogenization (Capdeville and Cance, 2014) is

$$G^{\varepsilon_0} \simeq G_{\text{homo}}(1 + \frac{1}{2\varepsilon_0}) \,, \tag{11}$$

where G_{homo} is G in the homogeneous case, G^{ε_0} is G in the heterogeneous case and $\varepsilon_0 = \lambda_{\text{h}}/\lambda_{\text{min}}$. Here we have $\varepsilon_0 = 2,\ 1$ and 0.5 , which leads to a $G^{\varepsilon_0}/G_{\text{homo}}$ ratio of 1.2, 1.5 and 2.0 respectively, which is roughly what is observed in Figure 9, at least for large N. The fact that this rule of thumb is less accurate for low degrees is not understood yet.

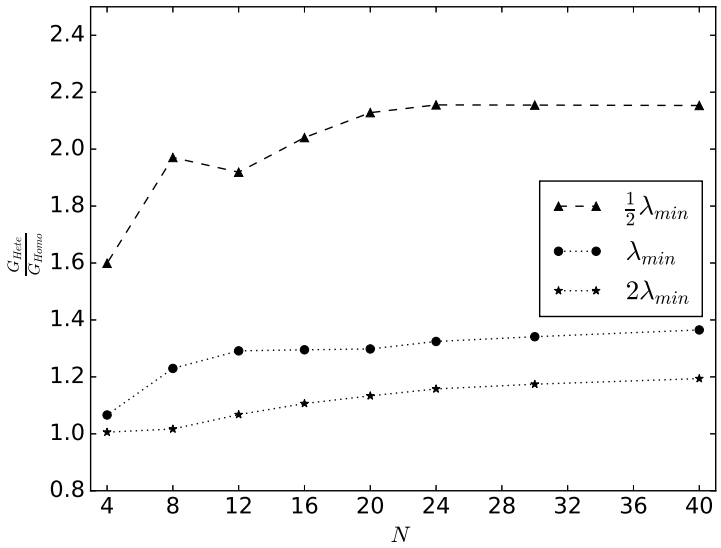

Figure 9: Heterogeneous models G ratio with the corresponding homogeneous model G as a function of the degree N. Three different heterogeneity roughness are displayed ($\lambda_h = \{0.5, 1, 2\}\lambda_{\text{min}}$).

OPTIMAL TIME STEP AS A FUNCTION OF THE DEGREE N

In this work, we only use the standard second-order explicit Newmark scheme. The error is therefore a quadratic function of dt. The time step is always subject to the CFL condition equation 1: $dt \leq dt_{\text{CFL}}$.

For a given degree N, we proceed in a similar way to the previous section to find the maximum time step dt_c such that the error remains just below 1%. We first choose a large G (twice the optimal G) so that the spatial error is much smaller than 1%. In that case, the error is only dependent upon dt. We then find dt_c by the method of trial and error and a quadratic interpolation. We finally check that the obtained dt_c leads to an error indeed just below 1%. An example is given in Figure 10 for $N = 4$. However, this procedure is limited by the CFL condition. Indeed, for high degrees, the error cannot reach 1% even for $dt = dt_{\text{CFL}}$. In that case, dt_c cannot be determined and is set to dt_{CFL}. For low degrees, it appears that dt_c is almost independent of the degree and that it only depends on the receiver

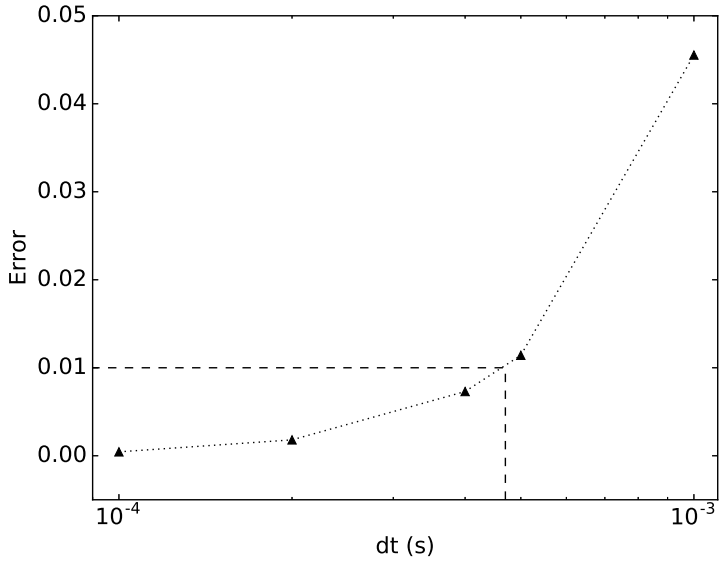

Figure 10: Error (see equation 9) as a function of the time step dt (dotted line), for a large G (spatially over sampled), N=4 in the 2-D homogeneous model and for the long distance receiver bin (200λ_{min}). The 1% error threshold and its corresponding time step (here $dt_c = 4.7 \times 10^{-4} s$) are displayed (dashed line).

bin distance. Finally, combining the optimal G and dt_c, we obtain an error lying between 1% and 2% as shown in Figure 11 for the 2-D homogeneous model and the long distance bin.

It can be seen from Figure 10 that, for $N = 4$, $dt = dt_{\text{CFL}}$ leads to an error of 4.5%, which is large. In such a case, a higher order time scheme would be necessary to take advantage of a large time step.

From this section, we can conclude that the error in time and space are almost independent.

GLOBAL COMPUTING TIME AS A FUNCTION OF DEGREE N

We finally evaluate the computing time as a function of the degree N for the optimal sampling G and time step dt determined in the two previous sections. If the gain in memory

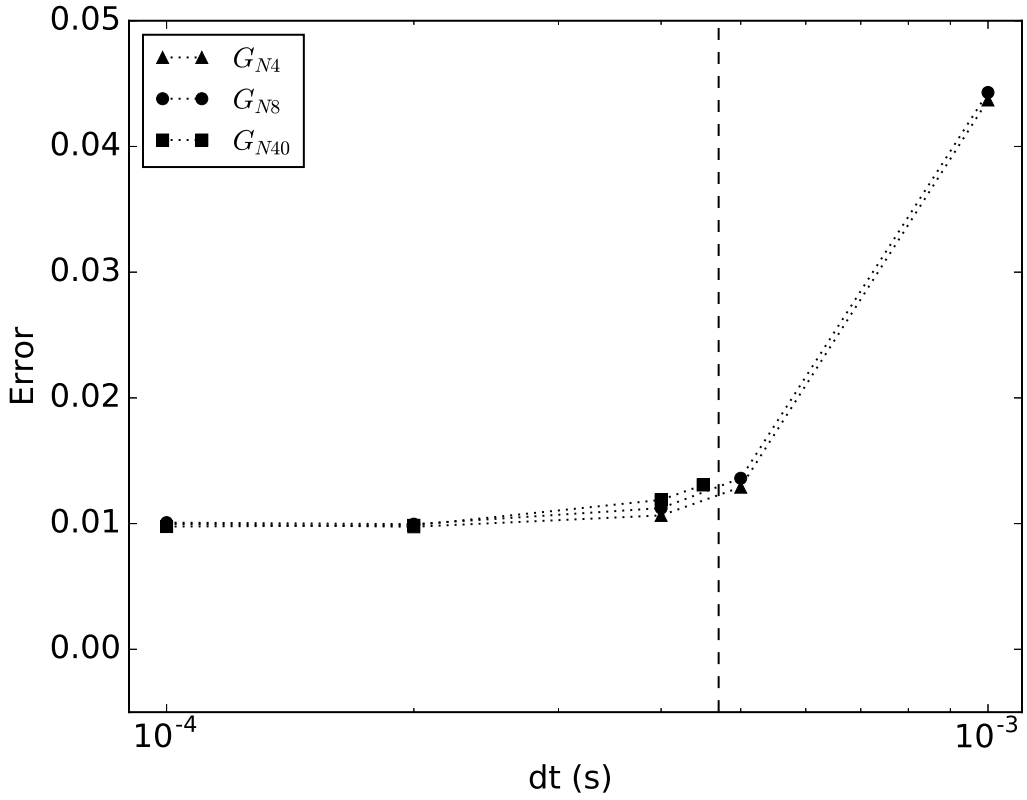

Figure 11: Error as a function of the time step dt with the optimal $G(N)$ (in Figure 7). Three different degrees are displayed ($N = 4, 8$ and 40). The optimal $dt_{\mathrm{c}} = 4.7 \times 10^{-4} s$ is displayed (dashed line). For $N = 40$, the dt_{CFL} is reached before the optimal dt_c.

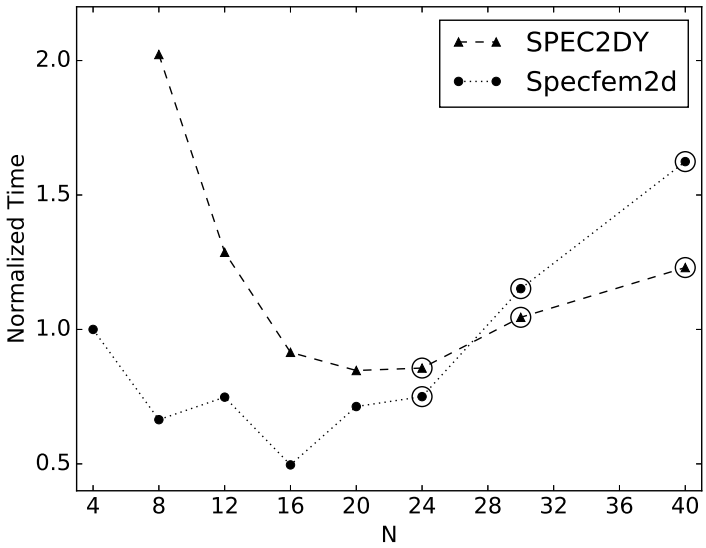

Figure 12: Computing time as a function of the degree N for the 2-D homogeneous case, for the short distance bin (20 λ_{min}) and for the SPEC2DY and SPECFEM2D programs. Symbols with an extra black circle correspond to measurements done with $dt = dt_{\text{CFL}}$.

of high degree N with respect to low degrees is only determined by G, the computing time strongly depends on particular code implementation. Therefore, the results presented here only give an idea of what can be done at the present time and could be changed with different implementation or hardware. It nevertheless gives an indication.

For each case, we choose as reference the computing time obtained using SPECFEM2D or 3D programs with $N = 4$, normalized to one. In this section, the gains or losses in computing time are always in comparison to this reference. Each computing time measurement is obtained using the optimal G and dt, performing five runs and averaging the obtained elapsed computing time. We used 40, 280 and 100 computing cores for the 2D homogeneous, 3D homogeneous model and 2D heterogeneous models respectively.

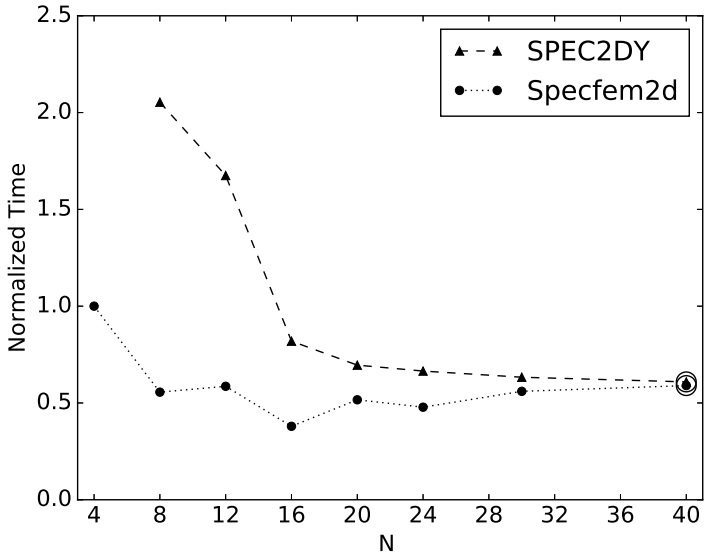

Figure 13: Computing time as a function of the degree N for the 2-D homogeneous case, for the long distance bin (200 λ_{min}) and for the SPEC2DY and SPECFEM2D programs. Symbols with an extra black circle correspond to measurements done with $dt = dt_{\text{CFL}}$.

2-D homogeneous model case results

Figures 12 and 13 show the results for the homogeneous 2-D case, for the short and long distance bins respectively. For the short distance receiver bin, it can be seen that the computing time decreases with the increasing degree for both programs until degree 16 or 20 and then increases again. To its maximum, the high degree computing time gain compared to the low degree one is about a factor of 2. For the long distance bin, a similar pattern can be observed. Nevertheless, the computing time does not increase much after a degree 16. The pattern is slightly different for each program, but the overall result is similar: the computing time almost constantly decreases with the degree and a maximum gain in the computing time by a factor of 2 can be achieved.

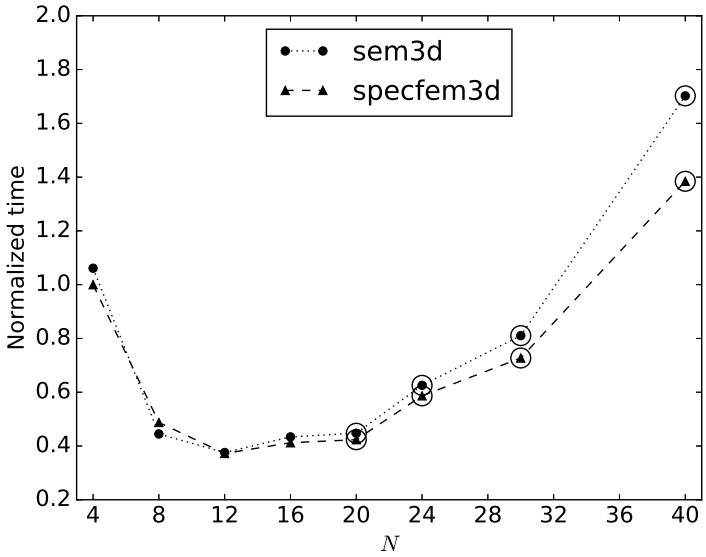

Figure 14: Computing time as a function of the degree N for the 3-D homogeneous case, for the short distance bin (20 λ_{min}) and for the SEM3D and SPECFEM3D programs. Symbols with an extra black circle correspond to measurements done with $dt = dt_{\text{CFL}}$.

3-D homogeneous model case results

For the 3-D case, we performed only the short distance bin case, because of computing resources limitation. The computing time as a function of N obtained for the two programs is displayed in Figure 14. First, it can be noted that there is no fundamental difference in computing time between the two programs. Next, the trend is very similar to the 2-D homogeneous short distance receiver case: first a decrease of the computing time with the degree and then an increase after a minimum. The minimum computing time is obtained in the 12-16 degree range.

For the 3-D long range case, we can only speculate that the results would be similar to the 2-D results. The short distance bin results are similar, and there is no specific reason to expect a difference for the long distance. It is nevertheless unfortunate that we can not check this point.

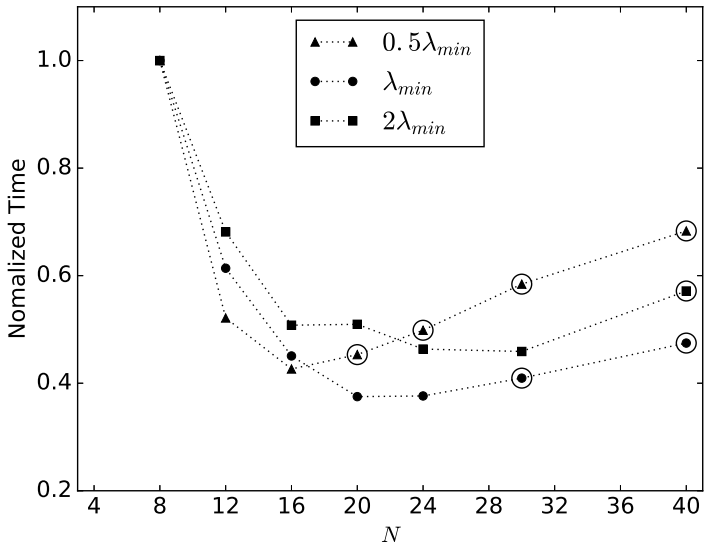

Figure 15: Computing time as a function of the degree N for the 2-D heterogeneous cases for the long distance bin (200 λ_{min}) and for the SPEC2DY program. Symbols with an extra black circle correspond to measurements done with $dt = dt_{\text{CFL}}$.

2-D heterogenous case results

We finally perform a test in the heterogeneous models described earlier. This test is only performed in 2-D for the long distance bin using the SPEC2DY program. We did not test the 3-D model because of computing resource limitation. We also did not test SPECFEM2D in that test because there is no simple way to input oscillating models in this program.

In Figure 15 the computing times as a function of the degree N for three 2-D heterogeneous models are shown (for $\lambda_{\text{h}} = 0.5\lambda_{\text{min}}, \lambda_{\text{min}}$and $2\lambda_{\text{min}}$) for the long distance bin. In Figure 13, it can be noted that the cases $\lambda_{\text{h}} = \lambda_{\text{min}}$ and $\lambda_{\text{h}} = 2\lambda_{\text{min}}$ display little differences compared to the homogeneous long distance case. The remaining differences are mostly due to the fact that the heterogeneous models have a smaller dt_{CFL} (because of a denser mesh) compared to the long receiver homogeneous case. The case $\lambda_{\text{h}} = 0.5\lambda_{\text{min}}$ is similar, but the minimum computing time occurs for lower degrees, 16-20, compared to 20-24 for the previous cases.

DISCUSSION

In this work, to assess if using very high degree ($N > 8$) in SEM is of any interest, we had to make several choices. This first one is the error threshold that we chose to 1%. Obviously, one could have made a different choice and this would have affected the optimal G, dt and measured computing time results. For example, for the short distance bin, using a threshold to 5% moves the optimal G for $N = 40$ from 2.4 to 1.97. Nevertheless, it does not change either the observed trend or the conclusions of this work. The way of estimating the source maximum frequency f_{max} has the same effect on G, but does not change the conclusions either. The second one is about the chosen programs to perform the computing time measurements. This aspect is problematic as different programs and different hardware can potentially significantly affect the computing time results. For example, we did not try the GPU spectral element implementations for which the behavior can be different. This work shows that very high degree SEM can be interesting, but one should keep in mind that, for a significantly new or different SEM implementation, or significantly different hardware, a benchmark should be run to determine which degree is the best from the point of view of computing time. Nevertheless, the memory gain results would remain unchanged.

Using very high degree SEM implies using very large elements, much larger than for low degrees. In a forward modeling context, the interest of very high degree SEM therefore strongly relies on the homogenization technique. It is the key to remove mechanical discontinuities: it allows to release the mesh constraints and thus allows the use of large elements. In the inversion context, the models are smooth anyway and it is usually simple to use very large elements. In both cases, the elastic properties are not constant per pieces but continuous and oscillatory with space. In this work, we did not test any realistic homogenized model such as the Marmousi model in Capdeville et al. (2010). Such models make it difficult to draw any general conclusion because results are then strongly dependent on the source and receiver locations. Instead, we tested single wavenumber heterogeneity

models in order to obtain a reference G as a function of the degree N for three different model roughness. Because realistic geological elastic models amplitude spectra decrease with the wavenumber, these periodic tests can be seen as what can be expected at worst. The test model $\lambda_{\text{h}} = \lambda_{\text{min}}/2$ is an upper bound and a realistic geological model is expected to behave more like the $\lambda_{\text{h}} = \lambda_{\text{min}}$ case. In the context of inversion, models are rather smooth and we expect they behave like the $\lambda_{\text{h}} = 2\lambda_{\text{min}}$ case. Finally, let us mention that homogenization cannot remove free-surface or solid-fluid interface topographies, even if they can be homogenized (Capdeville and Marigo, 2012). This can be a complication, but it can be handled generally by deforming the elements in the vertical direction to match the topography. This comes at the price of a high degree element transformation (where a linear or quadratic element are classically used).

Regarding the number of points per wavelength G, our work confirms that it can be as low as 2.5 using $N = 40$ for both 2-D and 3-D homogeneous or smooth media. Note that our G is an average number meaning that, for example, it is possible to cast about 16 λ_{min} within one degree 40 element direction. For rough heterogeneous media, G increases more or less following the rule of thumb of equation 11. Compared to a degree 4, a high degree G can be twice smaller. This implies that a factor up to 4 in 2-D and 8 in 3-D in memory can be gained by using a very high degree. This aspect can be particularly interesting in the adjoint inversion context where storing the wavefield on hard disks can be necessary either partially (Komatitsch et al., 2016) or fully to avoid backward propagation (Fichtner et al., 2009).

It is well known that the GLL points cluster near the edge of the elements for high degrees. This fact a priori implies a dramatically small time step to respect the CFL stability condition. Nevertheless, this effect is not that dramatic because G also decreases for high degrees and this partly compensates the clustering of the GLL points. For example, even if the closest distance between two GLL points is 77 times smaller in the reference element for $N = 40$ compared to $N = 4$, the time step is only 4.4 times smaller. Then, a low G

also implies a lower number of degrees of freedom and therefore less computation.

Finally, it is important to note that $G(N)$ strongly depends on the Poisson's ratio (Seriani and Oliveira, 2008). For a fixed degree N, G is degraded (larger) for Poisson's ratio close to 0.5 (in this work, we used a Poisson's ratio of 0.24). Seriani and Oliveira (2008) also have shown that G is less degraded for large degrees than for low degrees. It implies that high degree element is even more interesting compared to low degrees for Poisson's ratio close to 0.5.

CONCLUSIONS

In the end, high-degree SEM is often more effective than lower degrees and when it is not, the losses in computing time is not dramatic whereas the gain in memory is significant. For example, the memory needed for a degree 20 is about a quarter that necessary for a degree 4 in 2-D and about one eighth in 3-D in our numerical experiments. The computation time with degree 12 to 24 can be up to twice faster than the classical degree 4. To conclude, considering using SEM with a significantly higher degree than usual is a good idea in some situations such as full waveform inversion.

ACKNOWLEDGEMENTS

This research was supported by the National Natural Science Fund for Distinguished Young Scholars in China (Grant No.41625016,2017-2021) , the French National Research Agency (ANR-16-CE31-0022-01, 2016-2021) and the China Scholarship Council (File No. 201804910289). We are very grateful to the computing center CCIPL in Nantes and the server "WUKONG" in Beijing. We thank the two reviewers, E. Chaljub and A. Fichtner for their comments which greatly helped to improve the manuscript. E. Chaljub is at the origin of the appendix of this article. We express our sincere gratitude to their authors for the SPECFEM2D and 3D

suits in `https://geodynamics.org/cig/software/` and program EX2DDIR in `www.spice-thn.org`.

APPENDIX A

COMPUTATIONAL COMPLEXITY ANALYSIS OF CODE-INDEPENDENT FEATURES FOR SEM

In this appendix, we estimate the theoretical scaling of the numerical complexity of a SEM simulation as a function of the element degree N. The overall numerical complexity of such a simulation is the product of the cost at the element level times the number of elements times the number of time steps. At the element level, the numerical complexity is dominated by the stiffness matrix multiplication, which scales as N^{d+1}. This is usually where the main optimization effort is mainly put into, following matrix-matrix multiplication strategy (Deville et al., 2002). We have numerically checked that the calculation of the internal forces (the product between elemental stiffness matrix and displacement vector) indeed behaves in N^{d+1} for the SEM code used in this work. Regarding the number of elements N_e, assuming a domain size of L^d, we have $L^d = (N_\lambda \lambda_{\text{min}})^d$, where N_λ is the number of the minimum wavelength. We also have $\lambda_{\text{min}} = G(N)\overline{\Delta_x}$ where $\overline{\Delta_x}$ the average distance between GLL points within each spectral element. The element size in one direction is $L_e = N\overline{\Delta_x}$ and therefore

$$N_e \propto \left(N_\lambda \frac{G(N)}{N} \right)^d. \tag{A-1}$$

The number of time steps is controlled by equation 1 and therefore by $1/dx_{\text{min}}$. Because of the GLL points cluster near the element edges, dx_{min} scales as $\overline{\Delta_x}/N$ (see Figure 1) and therefore as $(NG(N))^{-1}$. Gathering those estimates, one finds the SEM numerical

complexity $C(N)$ scales as

$$C(N) \propto G^{d+1}(N)N^2\,. \tag{A-2}$$

This formula has a limit: it doesn't account for the fact that, in order to reach a given accuracy, it may happen that the maximum time steps allowed by the CFL cannot be reached. In that case, SEM numerical complexity $C(N)$ scales as

$$C(N) \propto G^{d}(N)N\,. \tag{A-3}$$

REFERENCES

Afanasiev, M., C. Boehm, M. van Driel, L. Krischer, M. Rietmann, D. A. May, M. G. Knepley, and A. Fichtner, 2018, Modular and flexible spectral-element waveform modelling in two and three dimensions: Geophysical Journal International, **216**, 1675–1692.

Beller, S., V. Monteiller, S. Operto, G. Nolet, A. Paul, and L. Zhao, 2018, Lithospheric architecture of the south-western alps revealed by multiparameter teleseismic full-waveform inversion: Geophysical Journal International, **212**, 1369–1388.

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

Capdeville, Y., and P. Cance, 2014, Residual homogenization for elastic wave propagation in complex media: Geophysical Journal International, **200**, 986–999.

Capdeville, Y., E. Chaljub, J. P. Vilotte, and J. P. Montagner, 2003, Coupling the spectral element method with a modal solution for elastic wave propgation in global earth models: Geophysical Journal International, **152**, 34–66.

Capdeville, Y., L. Guillot, and J. Marigo, 2010, 2-D non-periodic homogenization to upscale elastic media for P-SV waves: Geophysical Journal International, **182**, 903–922.

Capdeville, Y., Y. Gung, and B. Romanowicz, 2005, Towards global earth tomography using the spectral element method: a technique based on source stacking: Geophysical

Journal International, **162**, 541–554.

Capdeville, Y., and J.-J. Marigo, 2012, A non-periodic two scale asymptotic method to take account of rough topographies for 2-D elastic wave propagation: Geophysical Journal International, **192**, 163–189.

Chaljub, E., Y. Capdeville, and J. Vilotte, 2003, Solving elastodynamics in a solid heterogeneous 3-Sphere: a spectral element approximation on geometrically non-conforming grids: Journal of Computational Physics, **183**, 457–491.

Chaljub, E., D. Komatitsch, Y. Capdeville, J.-P. Vilotte, B. Valette, and G. Festa, 2007, Spectral element analysis in seismology: Advances in Wave Propagation in Heterogeneous Media, **48**, 365–419.

Cupillard, P., and Y. Capdeville, 2018, Non-periodic homogenization of 3-D elastic media for the seismic wave equation: Geophysical Journal International, **213**, 983–1001.

De Basabe, J. D., and M. K. Sen, 2007, Grid dispersion and stability criteria of some common finite-element methods for acoustic and elastic wave equations: Geophysics, **72(6)**, T81–T95.

De Hoop, A. T., 1960, A modification of cagniard's method for solving seismic pulse problems: Applied Scientific Research, Section B, **8**, 349–356.

Delavaud, E., 2007, Simulation numérique de la propagation d'ondes en milieu géologique complexe: application à l'évaluation de la réponse sismique du bassin de caracas (venezuela): PhD thesis, Paris, Institut de physique du globe.

Deville, M. O., P. F. Fischer, and E. H. Mund, 2002, High-order methods for incompressible fluid flow: Cambridge University Press. Cambridge Monographs on Applied and Computational Mathematics.

Festa, G., and J.-P. Vilotte, 2005, The newmark scheme as velocity-stress time-staggering: an efficient implementation for spectral element simulations of elastodynamics: Geophysical Journal International, **161**, 789–812.

Fichtner, A., B. L. N. Kennett, H. Igel, and H.-P. Bunge, 2009, Full seismic waveform

tomography for upper-mantle structure in the australasian region using adjoint methods: Geophysical Journal International, **179**, 1703–1725.

Johnson, L. R., 1974, Green's function for Lamb's problem: Geophysical Journal International, **37**, 99–131.

Komatitsch, D., R. Martin, J. Tromp, M. A. Taylor, and B. A. Wingate, 2001, Wave propagation in 2-D elastic media using a spectral element method with triangles and quadrangles: Journal of Computational Acoustics, **9**, 703–718.

Komatitsch, D., and J. Tromp, 1999, Introduction to the spectral-element method for 3-D seismic wave propagation: Geophysical Journal International, **139**, 806–822.

Komatitsch, D., and J.-P. Vilotte, 1998, The spectral element method: an effective tool to simulate the seismic response of 2D and 3D geological structures: Bulletin of the Seismological Society of America, **88**, 368–392.

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

Liang, W., Y. Wang, and C. Yang, 2015, Determining finite difference weights for the acoustic wave equation by a new dispersion-relationship-preserving method: Geophysical Prospecting, **63**, 11–22.

Maday, Y., and A. T. Patera, 1989, Spectral element methods for the incompressible navier-stokes equations: State-of-the-art surveys on computational mechanics, American Society of Mechanical Engineers, 71–143.

Mercerat, E. D., J. P. Vilotte, and F. J. Sánchez-Sesma, 2006, Triangular spectral element simulation of two-dimensional elastic wave propagation using unstructured triangular grids: Geophysical Journal International, **166**, 679–698.

Moczo, P., J. Kristek, M. Galis, E. Chaljub, and V. Etienne, 2011, 3-D finite-difference, finite-element, discontinuous-Galerkin and spectral-element schemes analysed for their accuracy with respect to P-wave to S-wave speed ratio: Geophysical Journal Interna-

tional, **187**, 1645–1667.

Monteiller, V., S. Chevrot, D. Komatitsch, and Y. Wang, 2015, Three-dimensional full waveform inversion of short-period teleseismic wavefields based upon the SEM-DSM hybrid method: Geophysical Journal International, **202**, 811–827.

Priolo, E., and G. Seriani, 1991, A numerical investigation of Chebyshev spectral element method for acoustic wave propagation: Proc. 13th world congress on computation and applied mathematics, 551–556.

Seriani, G., and S. Oliveira, 2008, Dispersion analysis of spectral element methods for elastic wave propagation: Wave Motion, **45**, 729 – 744.

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

Trinh, P.-T., R. Brossier, L. Métivier, L. Tavard, and J. Virieux, 2019, Efficient time-domain 3D elastic and viscoelastic full-waveform inversion using a spectral-element method on flexible cartesian-based mesh: Geophysics, **84(1)**, R61–R83.

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

Wang, Y., S. Chevrot, V. Monteiller, D. Komatitsch, F. Mouthereau, G. Manatschal, M. Sylvander, J. Diaz, M. Ruiz, F. Grimaud, S. Benahmed, H. Pauchet, and R. Martin, 2016, The deep roots of the western pyrenees revealed by full waveform inversion of teleseismic p waves: Geology, **44**, 475.

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

Table Caption

Table 1: The number of Ns, nz (number of elements in z direction), minimum dx ratio and maximum dt ratio in 2-D homogeneous half space case for the 200 λ_{min} epicentral distances case.

Figure Captions

- Figure 1: Minimum distance $d\xi_{\text{min}}^N$ between two GLL points for the 1-D reference element $\Lambda = [-1, 1]$ as a function of the degree N with logarithmic scales.

- Figure 2: The 2-D homogeneous domain used in this work. The source (black star) is located about $3\lambda_{\text{min}}$ below the surface. The short and long distance receiver bins (black square) are represented. The kinetic energy snapshot for $t = 270\ t_{\text{min}}$ and a 60×30 elements mesh, associated to a polynomial degree $N = 40$, are also displayed.

- Figure 3: The 3-D homogeneous domain and mesh used in this work. The source position (black star) and the receiver bin (black square) are represented. The represented mesh has 26×26×13 elements and is associated with a polynomial degree $N = 40$.

- Figure 4: Same as Figure 2 but for the 2-D heterogeneous media with $\lambda_{\text{h}} = \lambda_{\text{min}}$. The lower left zoom displays a $14.1\lambda_{\text{min}}$×$14.1\lambda_{\text{min}}$ area of V_S. The 100×50 elements mesh associated to the degree $N = 40$ is displayed.

- Figure 5: 2-D homogeneous model waveform displacement comparison between the analytical solution (gray line), the spectral element simulated solution (dashed line) for the horizontal (top plot) and vertical components (bottom plot). The residual ($\times 10^4$) is displayed in dotted line. The receiver is a typical receiver from the short distance bin ($20\ \lambda_{\text{min}}$ distance).

- Figure 6: 3-D homogeneous model waveform displacement comparison between the analytical solution (gray line) and the spectral element simulated solution (dashed line) for the horizontal (top plot) and vertical component (bottom plot). The residual ($\times 500$) is displayed in dotted line. The receiver is a typical receiver from the short distance bin ($20\ \lambda_{\text{min}}$ distance).

- Figure 7: Error (see equation 9) as a function of G (dotted line) for the short distance receiver bin and for degree $N = 40$ and a very small dt. The triangles are measured values and the dashed line shows the quadratic interpolation estimation of G to obtain a 1% error. Here, $G \simeq 2.35$.

- Figure 8: G as a function of the degree N for the 2-D and 3-D homogeneous models and for the 20 and 200 λ_{min} epicentral distance receiver bins.

- Figure 9: Heterogeneous models G ratio with the corresponding homogeneous model G as a function of the degree N. Three different heterogeneity roughness are displayed ($\lambda_h = \{0.5, 1, 2\}\lambda_{\text{min}}$).

- Figure 10: Error (see equation 9) as a function of the time step dt (dotted line), for a large G (spatially over sampled), N=4 in the 2-D homogeneous model and for the long distance receiver bin (200λ_{min}). The 1% error threshold and its corresponding time step (here $dt_c = 4.7 \times 10^{-4} s$) are displayed (dashed line).

- Figure 11: Error as a function of the time step dt with the optimal $G(N)$ (in Figure 7). Three different degrees are displayed ($N = 4, 8$ and 40). The optimal $dt_{\text{c}} = 4.7 \times 10^{-4} s$ is displayed (dashed line). For $N = 40$, the dt_{CFL} is reached before the optimal dt_c.

- Figure 12: Computing time as a function of the degree N for the 2-D homogeneous case, for the short distance bin (20 λ_{min}) and for the SPEC2DY and SPECFEM2D programs. Symbols with an extra black circle correspond to measurements done with $dt = dt_{\text{CFL}}$.

- Figure 13: Computing time as a function of the degree N for the 2-D homogeneous case, for the long distance bin (200 λ_{min}) and for the SPEC2DY and SPECFEM2D programs. Symbols with an extra black circle correspond to measurements done with $dt = dt_{\text{CFL}}$.

- Figure 14: Computing time as a function of the degree N for the 3-D homogeneous case, for the short distance bin (20 λ_{min}) and for the SEM3D and SPECFEM3D programs. Symbols with an extra black circle correspond to measurements done with $dt = dt_{\text{CFL}}$.

- Figure 15: Computing time as a function of the degree N for the 2-D heterogeneous cases for the long distance bin (200 λ_{min}) and for the SPEC2DY program. Symbols with an extra black circle correspond to measurements done with $dt = dt_{\text{CFL}}$.