


# Dispersion Properties of Explicit Finite Element Methods for Wave Propagation Modelling on Tetrahedral Meshes

S. Geevers<sup>1</sup>  · W. A. Mulder<sup>2,3</sup> ·  
J. J. W. van der Vegt<sup>1</sup>

Received: 29 August 2017 / Revised: 27 February 2018 / Accepted: 10 April 2018 /  
Published online: 20 April 2018  
© The Author(s) 2018

**Abstract** We analyse the dispersion properties of two types of explicit finite element methods for modelling acoustic and elastic wave propagation on tetrahedral meshes, namely mass-lumped finite element methods and symmetric interior penalty discontinuous Galerkin methods, both combined with a suitable Lax–Wendroff time integration scheme. The dispersion properties are obtained semi-analytically using standard Fourier analysis. Based on the dispersion analysis, we give an indication of which method is the most efficient for a given accuracy, how many elements per wavelength are required for a given accuracy, and how sensitive the accuracy of the method is to poorly shaped elements.

**Keywords** Tetrahedral mesh · Explicit finite element method · Mass lumping · Discontinuous Galerkin method · Wave equation · Dispersion analysis

**Mathematics Subject Classification** 65M60

## 1 Introduction

Realistic wave propagation problems often involve large three-dimensional domains consisting of heterogeneous materials with complex geometries and sharp interfaces. Solving such

---

This work was funded by the Shell Global Solutions International B.V. under Contract No. PT45999.

---

✉ S. Geevers  
s.geevers@utwente.nl

W. A. Mulder  
wim.mulder@shell.com

J. J. W. van der Vegt  
j.j.w.vandervegt@utwente.nl

<sup>1</sup> Department of Applied Mathematics, University of Twente, Enschede, The Netherlands

<sup>2</sup> Shell Global Solutions International BV, Amsterdam, The Netherlands

<sup>3</sup> Delft University of Technology, Delft, The Netherlands

problems requires a numerical method that is efficient in terms of computation time and is flexible enough to capture the effect of a complex geometry.

Standard finite difference methods fall short, since they rely on Cartesian grids that cannot efficiently capture the effect of complex interfaces and boundaries. Finite element methods can overcome this problem when the elements are aligned with those surfaces. However, the accuracy of the finite element method quickly deteriorates when the elements are poorly shaped or are poorly aligned with the geometry. Obtaining a high quality mesh is therefore quintessential. While both hexahedral and tetrahedral elements are commonly used for three-dimensional problems, tetrahedral elements have a big advantage in this respect, since they offer more geometric flexibility and since robust tetrahedral mesh generators based on the Delaunay criterion are available [29,35].

Apart from the construction of a high-quality mesh, finite element methods for wave propagation problems also require a (block)-diagonal mass matrix to enable explicit time-stepping. A diagonal mass matrix can be obtained with nodal basis functions and a quadrature rule, if the quadrature points coincide with the basis function nodes. This technique is known as mass-lumping. For quadrilaterals and hexahedra, mass-lumping is achieved by combining tensor-product basis functions with a Gauss–Lobatto quadrature rule, resulting in a scheme known as the spectral element method [20,30,33]. For triangles and tetrahedra, an efficient linear mass-lumped scheme is obtained by combining standard Lagrangian basis functions with a Newton–Cotes quadrature rule. For higher-degree triangles and tetrahedra, however, this approach results in an unstable, unsolvable, or inaccurate scheme. To remain accurate and stable, the space of the triangle or tetrahedron is enriched with higher-degree bubble functions. This approach has led to accurate mass-lumped triangles of degree 2 [15], 3 [7], 4 [25], 5 [5], 6 [27], 7–9 [9,24] and tetrahedra of degree 2 [25] and 3 [5].

Another way to obtain a (block)-diagonal mass matrix is by using discontinuous basis functions. The resulting schemes are known as Discontinuous Galerkin (DG) methods. The first DG methods for wave propagation problems were based on a first-order formulation of the wave equation [6,31]. In [32] and [17], DG methods were introduced that were based on the original second-order formulation of the wave problem. The advantage of finite element methods based on the second-order formulation is that they do not need to compute or store the auxiliary variables that appear in the first-order formulation. Moreover, they can be combined with a leap-frog or higher-order Lax–Wendroff time integration scheme that only requires  $K$  stages for a  $2K$ -order accuracy. We focus on the symmetric interior penalty discontinuous Galerkin (SIPDG) method, presented and analysed in [17], which is based on the second-order formulation of the wave problem and which also remains energy-conservative on the discrete level. To remain accurate and stable, face integrals and interior penalty parameters are added to the discrete operator. We consider two choices for the penalty parameter: the penalty term derived in [28], based on the trace inequality of [36], and a recently developed sharper estimate [16], based on a more involved trace inequality.

To effectively apply these methods, it is crucial to know the required mesh resolution for a given accuracy. It is also useful to know which method is the most efficient for a given accuracy and how the mesh quality and material parameters, such as the P/S-wave velocity ratio for elastic waves, affect the accuracy. A practical and common measure for the accuracy of these type of methods is the amount of numerical dispersion and dissipation. In this paper, we will focus mainly on the numerical dispersion, since the methods we consider are all energy-conservative and therefore do not suffer from numerical dissipation. We do, however, also investigate the spurious modes that appear when projecting a physical wave onto the discrete space.

The dispersion properties of DG methods based on the first-order formulation of the wave problem have already been analysed for Cartesian meshes [1, 18], triangles [18, 22], and tetrahedra [19]. For the SIPDG method, these properties have already been analysed for Cartesian meshes in [2, 13] and triangles in [3] and for the mass-lumped finite element method this has already been analysed for quadrilaterals and hexahedra in [8, 11, 26] and for triangles in [23]. However, a dispersion analysis of the mass-lumped finite element and SIPDG methods for tetrahedra is, to the best of our knowledge, still missing, even though most realistic wave problems involve three-dimensional domains for which tetrahedral elements are particularly suitable. In this paper, we therefore present an extensive dispersion analysis of these methods for tetrahedra. This analysis is based on standard Fourier analysis. We use the analysis to obtain estimates for the required number of elements per wavelength and estimate the computational cost to obtain an indication of which method is the most efficient for a given accuracy. We consider both acoustic and elastic waves and also look at the effect of poorly shaped elements and high P/S-wave velocity ratios on the accuracy of the methods.

This paper is organised as follows: in Sect. 2, we introduce the tensor notation used in this paper. The acoustic and elastic wave equations are presented in Sect. 3 and the mass-lumped and discontinuous Galerkin finite element methods are presented in Sect. 4. In Sect. 5, we explain how we analyse the dispersion properties of these methods. The results of this analysis are presented in Sect. 6 and the main conclusions are summarised in Sect. 7.

## 2 Some Tensor Notation

Before we present the acoustic and elastic wave equations, we explain the tensor notation that we use throughout this paper. We let the dot product of two tensors denote the summation over the last index of the left and first index of the right tensor. For the double dot product we also sum over the last-but-one index of the left and second index of the right tensor. A concatenation of two tensors denotes the standard tensor product. To give some examples, let  $\hat{\mathbf{n}} \in \mathbb{R}^d$ ,  $\mathbf{u} \in \mathbb{R}^m$  be two vectors,  $\boldsymbol{\sigma} \in \mathbb{R}^{d \times m}$  a second-order tensor, and  $C \in \mathbb{R}^{d \times m \times m \times d}$  a fourth-order tensor. Then

$$\begin{aligned}
 [\hat{\mathbf{n}}\mathbf{u}]_{ij} &:= \hat{n}_i u_j, & [\boldsymbol{\sigma} \cdot \mathbf{u}]_i &:= \sum_{l=1}^m \sigma_{il} u_l, \\
 [C : \boldsymbol{\sigma}]_{ij} &:= \sum_{k=1}^d \sum_{l=1}^m C_{ijkl} \sigma_{kl}, & [\hat{\mathbf{n}} \cdot C]_{qji} &:= \sum_{k=1}^d \hat{n}_k C_{kqji},
 \end{aligned}$$

for all  $i = 1, \dots, d$  and  $j, q = 1, \dots, m$ .

In the next section we will use this tensor notation to present the acoustic and isotropic elastic wave equations.

## 3 The Acoustic and Isotropic Elastic Wave Equations

Let  $\Omega \subset \mathbb{R}^3$  be a three-dimensional open domain with a Lipschitz boundary  $\partial\Omega$ , and let  $(0, T)$  be the time domain. Also, let  $\{\Gamma_d, \Gamma_n\}$  be a partition of  $\partial\Omega$ , corresponding to Dirichlet and von Neumann boundary conditions, respectively. We define the following linear hyperbolic problem:

$$\rho \partial_t^2 \mathbf{u} = \nabla \cdot C : \nabla \mathbf{u} + \mathbf{f} \quad \text{in } \Omega \times (0, T), \tag{1a}$$

$$C : \hat{\mathbf{n}}\mathbf{u} = \mathbf{0} \quad \text{on } \Gamma_d \times (0, T), \tag{1b}$$

$$\hat{\mathbf{n}} \cdot C : \nabla \mathbf{u} = \mathbf{0} \quad \text{on } \Gamma_n \times (0, T), \tag{1c}$$

$$\mathbf{u}|_{t=0} = \mathbf{u}_0 \quad \text{in } \Omega, \tag{1d}$$

$$\partial_t \mathbf{u}|_{t=0} = \mathbf{v}_0 \quad \text{in } \Omega, \tag{1e}$$

where  $\mathbf{u} : \Omega \times (0, T) \rightarrow \mathbb{R}^m$  is a vector of  $m$  variables that are to be solved,  $\nabla$  is the gradient operator,  $\rho : \Omega \rightarrow \mathbb{R}^+$  is a positive scalar field,  $C : \Omega \rightarrow \mathbb{R}^{3 \times m \times m \times 3}$  a fourth-order tensor field,  $\mathbf{f} : \Omega \times (0, T) \rightarrow \mathbb{R}^m$  the source field, and  $\hat{\mathbf{n}} : \partial\Omega \rightarrow \mathbb{R}^3$  the outward pointing normal unit vector.

By choosing the appropriate tensor and scalar field we can obtain the acoustic wave equation and the isotropic elastic wave equations.

*Case 1 (Isotropic elastic wave equations)* To obtain the isotropic elastic wave equations, set  $m = 3$  and

$$C_{ijqp} = \lambda \delta_{ij} \delta_{pq} + \mu (\delta_{ip} \delta_{jq} + \delta_{iq} \delta_{jp}),$$

for  $i, j, p, q = 1, 2, 3$ , where  $\delta$  is the Kronecker delta. Equation (1a) then becomes

$$\rho \partial_t^2 \mathbf{u} = \nabla \lambda (\nabla \cdot \mathbf{u}) + \nabla \cdot \mu (\nabla \mathbf{u} + \nabla \mathbf{u}^t) + \mathbf{f},$$

where  $\mathbf{u} : \Omega \times (0, T) \rightarrow \mathbb{R}^3$  is the displacement field,  $\rho : \Omega \rightarrow \mathbb{R}^+$  is the mass density,  $\lambda, \mu : \Omega \rightarrow \mathbb{R}^+$  are the Lamé parameters, and  $\mathbf{f} : \Omega \times (0, T) \rightarrow \mathbb{R}^3$  is the external volume force. The superscript  $t$  denotes the transposed.

*Case 2 (Acoustic wave equation)* To obtain the acoustic wave equation, set  $m = 1, u = p, \rho = (\tilde{\rho} \tilde{c}^2)^{-1}$ , and

$$C_{i11j} := \frac{1}{\tilde{\rho}} \delta_{ij},$$

for  $i, j = 1, 2, 3$ , where  $\delta$  is the Kronecker delta. Equation (1a) then becomes

$$\frac{1}{\tilde{\rho} \tilde{c}^2} \partial_t^2 p = \nabla \cdot \frac{1}{\tilde{\rho}} \nabla p + f,$$

where  $p : \Omega \times (0, T) \rightarrow \mathbb{R}$  is the pressure field,  $\tilde{\rho} : \Omega \rightarrow \mathbb{R}^+$  the mass density,  $\tilde{c} : \Omega \rightarrow \mathbb{R}^+$  the acoustic velocity field, and  $f = \nabla \cdot (\tilde{\rho}^{-1} \tilde{\mathbf{f}})$  the source term with  $\tilde{\mathbf{f}} : \Omega \times (0, T) \rightarrow \mathbb{R}^3$  the external volume force.

These equations can be solved with the finite element methods described in the next section.

## 4 The Discontinuous Galerkin and Mass-Lumped Finite Element Method

### 4.1 The Classical Finite Element Method

Let  $\mathcal{T}_h$  be a tetrahedral tessellation of  $\Omega$ , with  $h$  denoting the radius of the smallest sphere that can contain each element and let  $\mathcal{U}_h$  be the finite element space consisting of continuous element-wise polynomial basis functions satisfying boundary condition (1b). The classical

conforming finite element formulation of (1) is finding  $\mathbf{u}_h : [0, T] \rightarrow \mathcal{U}_h$  such that  $\mathbf{u}_h|_{t=0} = \Pi_h \mathbf{u}_0$ ,  $\partial_t \mathbf{u}_h|_{t=0} = \Pi_h \mathbf{v}_0$  and

$$(\rho \partial_t^2 \mathbf{u}_h, \mathbf{w}) + a(\mathbf{u}_h, \mathbf{w}) = (\mathbf{f}, \mathbf{w}), \quad \mathbf{w} \in \mathcal{U}_h, t \in [0, T], \tag{2}$$

where  $(\cdot, \cdot)$  denotes the standard  $L^2$  inner product,  $\Pi_h : L^2(\Omega) \rightarrow \mathcal{U}_h$  denotes the weighted  $L^2$  projection operator defined such that  $(\rho \Pi_h \mathbf{u}, \mathbf{w}) = (\rho \mathbf{u}, \mathbf{w})$  for all  $\mathbf{w} \in \mathcal{U}_h$ , and  $a : H^1(\Omega)^m \times H^1(\Omega)^m \rightarrow \mathbb{R}$  is the (semi)-elliptic operator given by

$$a(\mathbf{u}, \mathbf{w}) := \int_{\Omega} (\nabla \mathbf{u})^t : C : \nabla \mathbf{w} \, dx.$$

Let  $\{\mathbf{w}^{(i)}\}_{i=1}^n$  be the set of basis functions spanning  $\mathcal{U}_h$ , and let, for any  $\mathbf{u} \in L^2(\Omega)^m$ , the vector  $\underline{\mathbf{u}} \in \mathbb{R}^n$  be defined such that  $\sum_{i=1}^n \underline{u}_i \mathbf{w}^{(i)} = \Pi_h \mathbf{u}$ . Also, let  $M, A \in \mathbb{R}^{n \times n}$  be the mass matrix and stiffness matrix, respectively, defined by  $M_{ij} := (\rho \mathbf{w}^{(i)}, \mathbf{w}^{(j)})$  and  $A_{ij} := a(\mathbf{w}^{(i)}, \mathbf{w}^{(j)})$ , and let  $\underline{\mathbf{f}}^* : [0, T] \rightarrow \mathbb{R}^n$  be given by  $\underline{f}_i^* := (\mathbf{f}, \mathbf{w}^{(i)})$ . The finite element method can then be formulated as finding  $\underline{\mathbf{u}}_h : [0, T] \rightarrow \mathbb{R}^n$  such that  $\underline{\mathbf{u}}_h|_{t=0} = \underline{\mathbf{u}}_0$ ,  $\partial_t \underline{\mathbf{u}}_h|_{t=0} = \underline{\mathbf{v}}_0$ , and

$$M \partial_t^2 \underline{\mathbf{u}}_h + A \underline{\mathbf{u}}_h = \underline{\mathbf{f}}^*, \quad t \in [0, T]. \tag{3}$$

The main drawback of the classical conforming finite element approach is that when an explicit time integration scheme is applied, a system of equations of the form  $M \underline{\mathbf{x}} = \underline{\mathbf{b}}$  needs to be solved at every time step, with  $M$  not (block)-diagonal. For large-scale problems, this results in a very inefficient time stepping scheme. This problem can be circumvented by lumping the mass matrix into a diagonal matrix or by using discontinuous basis functions.

### 4.2 Mass-Lumping

When using nodal basis functions, the mass matrix can be lumped into a diagonal matrix by taking the sum over each row. This is equivalent to replacing the inner product  $(\cdot, \cdot)$  by  $(\cdot, \cdot)_h^{(L)}$ , in which the element integrals are approximated by a quadrature rule with quadrature points that coincide with the nodes of the basis functions. We can write

$$(\mathbf{u}, \mathbf{w})_h^{(L)} = \sum_{e \in \mathcal{T}_h} \sum_{\mathbf{x} \in \mathcal{Q}_e} \omega_{e,\mathbf{x}} \rho(\mathbf{x}) \mathbf{u}(\mathbf{x}) \cdot \mathbf{w}(\mathbf{x}),$$

where  $\mathcal{Q}_e$  denotes the quadrature points on element  $e$  and  $\omega_{e,\mathbf{x}}$  denote the quadrature weights. Let  $\{\mathbf{x}^{(i)}\}_{i=1}^n$  denote the global set of integration points and define  $\mathbf{w}^{(i)}$  to be the nodal basis function corresponding to  $\mathbf{x}^{(i)}$ , so  $\mathbf{w}^{(i)}(\mathbf{x}^{(j)}) = \delta_{ij}$ , with  $\delta$  the Kronecker delta. Then the mass matrix becomes diagonal with entries  $M_{ii} = \sum_{e \in \mathcal{T}_x(i)} \omega_{e,\mathbf{x}^{(i)}} \rho(\mathbf{x}^{(i)})$ , where  $\mathcal{T}_x$  denotes the set of elements containing or adjacent to  $\mathbf{x}$ .

For quadrilaterals and hexahedra, mass-lumping is achieved by using tensor-product basis functions and Gauss–Lobatto integration points. The resulting scheme is known as the spectral element method. For triangles and tetrahedra, mass-lumping is less straight-forward. Combining standard Lagrangian basis functions with a Newton–Cotes quadrature rule results in an efficient mass-lumped scheme for linear tetrahedra, but for higher-degree basis functions, this approach results either in an unstable scheme due to non-positive quadrature weights or in a scheme with a reduced order of convergence. This problem can be resolved by enriching the finite element space with higher-degree bubble functions and by adding integration points to the interior of the elements and faces. For example, by enriching the space of the quadratic tetrahedron with 3 degree-4 face bubble functions and 1 degree-4 interior bubble

function, an enriched degree-2 mass-lumped tetrahedron that remains third-order accurate can be obtained [25].

In this paper we will analyse the standard linear mass-lumped finite element method, the mass-lumped finite element method of degree 2 derived in [25], and the 2 versions of degree 3 mass-lumped finite element methods derived in [5]. We will refer to these methods as ML1, ML2, ML3a and ML3b, respectively.

### 4.3 The Symmetric Interior Penalty Discontinuous Galerkin Method

Another way to obtain a (block)-diagonal mass matrix is by allowing the finite element space  $\mathcal{U}_h$  to be discontinuous at the faces. When choosing basis functions that have support on only a single element, the mass matrix becomes block-diagonal with each block corresponding to a single element. When using orthogonal basis functions, the mass matrix even becomes strictly diagonal. In order to keep the finite element method stable and consistent with the analytic solution, the elliptic operator needs to be augmented. This can be accomplished with the symmetric interior penalty method [17], where  $a$  is replaced by the discrete (semi)-elliptic operator  $a_h^{(DG)} : \mathcal{U}_h \times \mathcal{U}_h \rightarrow \mathbb{R}$ , given by

$$a_h^{(DG)}(\mathbf{u}, \mathbf{w}) := a_h^{(C)}(\mathbf{u}, \mathbf{w}) - a_h^{(D)}(\mathbf{u}, \mathbf{w}) - a_h^{(D)}(\mathbf{w}, \mathbf{u}) + a_h^{(IP)}(\mathbf{u}, \mathbf{w})$$

with

$$\begin{aligned} a_h^{(C)}(\mathbf{u}, \mathbf{w}) &:= \sum_{e \in \mathcal{T}_h} \int_e (\nabla \mathbf{u})^t : C : \nabla \mathbf{w} \, dx, \\ a_h^{(D)}(\mathbf{u}, \mathbf{w}) &:= \sum_{f \in \mathcal{F}_{h,in} \cup \mathcal{F}_{h,d}} \int_f \llbracket \mathbf{u} \rrbracket^t : \{ \{ C : \nabla \mathbf{w} \} \} \, ds, \\ a_h^{(IP)}(\mathbf{u}, \mathbf{w}) &:= \sum_{f \in \mathcal{F}_{h,in} \cup \mathcal{F}_{h,d}} \int_f \llbracket \mathbf{u} \rrbracket^t : \{ \{ \alpha_h C \} \} : \llbracket \mathbf{w} \rrbracket \, ds, \end{aligned}$$

where  $\mathcal{F}_{h,in}$  and  $\mathcal{F}_{h,d}$  are the internal faces and boundary faces on  $\Gamma_d$ , respectively,  $\alpha_h \in \bigotimes_{e \in \mathcal{T}} L^\infty(\partial e)$  is the penalty function, and  $\{ \{ \cdot \} \}$ ,  $\llbracket \cdot \rrbracket$  are the average trace operator and jump operator, respectively, defined as

$$\{ \{ \phi \} \} |_f := \frac{1}{|\mathcal{T}_f|} \sum_{e \in \mathcal{T}_f} \phi|_{\partial e \cap f}, \quad \llbracket \mathbf{u} \rrbracket |_f := \sum_{e \in \mathcal{T}_f} (\hat{\mathbf{n}} \mathbf{u})|_{\partial e \cap f},$$

for all faces  $f \in \mathcal{F}$ , where  $\mathcal{T}_f$  denotes the set of elements adjacent to face  $f$ , and  $\hat{\mathbf{n}}|_{\partial e}$  denotes the outward pointing normal unit vector of element  $e$ . The bilinear form  $a_h^{(C)}$  is the same as the original elliptic operator  $a$  and is the part that remains when both input functions are continuous. The bilinear form  $a_h^{(D)}$  can be interpreted as the additional part that results from partial integration of the elliptic operator  $a$  when the first input function is discontinuous. Finally, the bilinear form  $a_h^{(IP)}$  is the part that contains the interior penalty function needed to ensure stability of the scheme.

The penalty term can have a significant impact on the performance of the SIPDG method, since a larger penalty term results in a more restrictive bound on the time step size, but also because it can have a significant effect on the accuracy, as we will show in Sect. 6. Several lower bounds for the penalty term are based on the trace inequality of [36], including [14, 28, 34], among which we found the bound in [28] to be the sharpest. Recently, a sharper penalty term bound was presented in [16], which is based on a more involved trace inequality.

In this paper we will consider both the penalty term of [16], given by (4a), and the one of [28], given by (4b):

$$\alpha_h|_{\partial e \cap f} := \frac{v_h|_{\partial e \cap f}}{|\mathcal{T}_f|} \sup_{\substack{\mathbf{u} \in \mathcal{P}^p(e)^m \\ \mathbf{C} : \nabla \mathbf{u} \neq \mathbf{0}}} \frac{\int_{\partial e} (\hat{\mathbf{n}} \cdot \mathbf{C} : \nabla \mathbf{u}) \cdot v_h^{-1} \mathbf{c}_{\hat{\mathbf{n}}}^{-1} \cdot (\hat{\mathbf{n}} \cdot \mathbf{C} : \nabla \mathbf{u}) \, ds}{\int_e (\nabla \mathbf{u})^t : \mathbf{C} : \nabla \mathbf{u} \, dx}, \tag{4a}$$

$$\alpha_h|_{\partial e \cap f} := \frac{p(p+2)}{\min_{e \in \mathcal{T}_f} d_e}, \tag{4b}$$

for all  $e \in \mathcal{T}_h$ ,  $f \subset \partial e$ , where  $p$  denotes the degree of the polynomial basis functions,  $\mathcal{P}^p(e)$  denotes the space of polynomial functions of degree  $p$  or less in element  $e$ ,  $v_h|_{\partial e \cap f} := |f|/|e|$  is a scaling function of order  $h^{-1}$ , with  $|e|$ ,  $|f|$  the volume of  $e$  and area of  $f$ , respectively,  $\mathbf{c}_{\hat{\mathbf{n}}}^{-1}$  denotes the (pseudo)-inverse of the second-order tensor  $\mathbf{c}_{\hat{\mathbf{n}}} := \hat{\mathbf{n}} \cdot \mathbf{C} \cdot \hat{\mathbf{n}}$ , where  $\hat{\mathbf{n}}$  is the outward pointing normal unit vector, and  $d_e$  denotes the diameter of the inscribed sphere of  $e$ . Although the first version requires more preprocessing time, it allows for an approximately 1.5 times larger time step [16].

We will refer to the SIPDG method with  $p = 1, 2, 3$  using the penalty term as defined by (4a) as DG1a, DG2a, and DG3a, respectively, and to the same methods using the penalty term as defined by (4b) as DG1b, DG2b, and DG3b.

### 4.4 The Lax–Wendroff Time Integration Scheme

To solve the resulting set of ODE’s (3) in time, we use the Lax–Wendroff method [12, 21], which is based on Taylor expansions in time and substitutes the time derivatives by matrix-vector operators using the original equations (3). For the second-order formulation, the resulting scheme is also known as Dablain’s scheme [10]. The advantage of this scheme is that it is time-reversible, energy-conservative, and only requires  $K$  stages for a  $2K$ -order of accuracy.

To introduce the scheme, let  $\Delta t > 0$  denote the time step size, and let  $\underline{\mathbf{U}}_h(t_i)$  denote the approximation of  $\mathbf{u}_h$  at time  $t_i := i \Delta t$  for  $i = 0, \dots, N_T$  with  $N_T$  the total number of time steps. The order- $2K$  Lax–Wendroff method can be written as

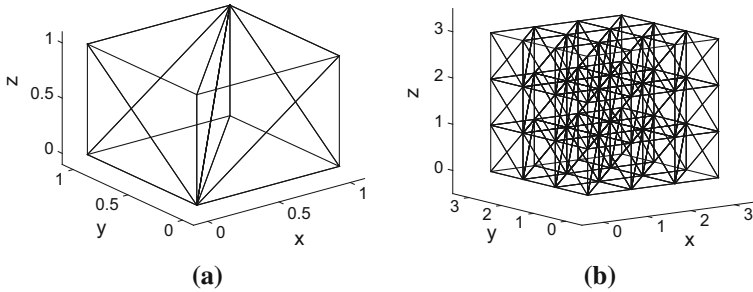
$$\underline{\mathbf{U}}_h(t_{i+1}) = -\underline{\mathbf{U}}_h(t_{i-1}) + 2 \sum_{k=0}^K \frac{1}{(2k)!} \Delta t^{2k} (\partial_t^{2k} \underline{\mathbf{U}}_h)(t_i), \quad i = 1, \dots, N_T - 1, \tag{5}$$

with  $\underline{\mathbf{U}}_h(t_0) = \underline{\mathbf{U}}_h(0) := \underline{\mathbf{u}}_0$  and  $\underline{\mathbf{U}}_h(t_1) := \sum_{k=0}^{2K-1} \frac{1}{k!} \Delta t^k (\partial_t^k \underline{\mathbf{U}}_h)(0)$ , and where  $(\partial_t^k \underline{\mathbf{U}}_h)(t_i)$  is recursively defined by

$$(\partial_t^k \underline{\mathbf{U}}_h)(0) := \begin{cases} \underline{\mathbf{u}}_0 & k = 0, \\ \underline{\mathbf{v}}_0 & k = 1, \\ -M^{-1} A (\partial_t^{k-2} \underline{\mathbf{U}}_h)(0) + \partial_t^{k-2} \underline{\mathbf{f}}(0) & k \geq 2, \end{cases}$$

and

$$(\partial_t^k \underline{\mathbf{U}}_h)(t_i) := \begin{cases} \underline{\mathbf{U}}_h(t_i) & k = 0, \\ -M^{-1} A (\partial_t^{k-2} \underline{\mathbf{U}}_h)(t_i) + \partial_t^{k-2} \underline{\mathbf{f}}(t_i) & k = 2, 4, 6, \dots, 2K, \end{cases}$$



**Fig. 1** Unit cell subdivided into tetrahedra (a), and periodic mesh made from  $3 \times 3 \times 3$  copies of this unit cell (b)

for  $i \geq 1$ , with  $\underline{\mathbf{f}} := M^{-1}\underline{\mathbf{f}}^*$ . In case  $K = 1$ , this scheme reduces to the standard leap-frog or central difference scheme. When there is no source term, (5) simplifies to

$$\underline{\mathbf{U}}_h(t_{i+1}) = -\underline{\mathbf{U}}_h(t_{i-1}) + 2 \sum_{k=0}^K \frac{1}{(2k)!} \Delta t^{2k} (-M^{-1}A)^k \underline{\mathbf{U}}_h(t_i), \tag{6}$$

for  $i = 1, \dots, N_T - 1$ .

For the dispersion analysis, we will choose  $K$  equal to the polynomial degree  $p$  of the spatial discretization, since this will result in a  $2p$ -order convergence rate of the dispersion error as shown in Sect. 6.

### 5 Dispersion Analysis

A common measure for the quality of a numerical method for wave propagation modelling is the amount of numerical dispersion and dissipation. Numerical dispersion refers in this context to the discrepancy between the numerical and physical wave propagation speed and numerical dissipation is the loss of energy in the numerical scheme. Since the schemes that we consider are all energy-conservative, they do not suffer from numerical dissipation. However, when projecting a physical wave onto the discrete space, this results in a superposition of a well-matching numerical wave and several numerical waves that have a completely different shape and frequency. We compute the number of these non-matching or spurious waves and refer to it as the eigenvector error, since it is related to the accuracy of the eigenvectors of  $M^{-1}A$ , while the dispersion error is related to the accuracy of the eigenvalues of  $M^{-1}A$ .

We analyse the dispersion and eigenvector error using standard Fourier analysis, which is also known in this context as plane wave analysis. The main idea of this analysis is to compare physical plane waves with numerical plane waves on a homogeneous periodic domain, free from external forces, using a periodic mesh. To obtain a periodic tetrahedral mesh we subdivide a small cell into tetrahedra and repeat this pattern to fill the entire domain as illustrated in Fig. 1. By using Fourier modes, we can then efficiently compute the numerical plane waves and their dispersion properties by solving eigenvalue problems on only a single cell.

Our analysis is similar to [11], but with the following extensions:

- We extend the analysis to parallelepiped cells, since this allows for a more regular tetrahedral mesh.



- We also compute the number of spurious modes that appear in the projection of the physical wave.
- In the three-dimensional elastic case, there are two distinct secondary or shear waves with the same wave vector. To compute the dispersion and eigenvector error in this case, we consider the two best matching numerical plane waves.

We explain the dispersion analysis in more detail in the following subsections. First, we show how we can derive an analytical expression for the numerical plane waves using Fourier modes. After that, we show how we use this to compute the numerical dispersion and eigenvector error. In the last subsection we explain how we estimate the computational cost for each method.

### 5.1 Analytic Expression for the Numerical Plane Waves

We first consider a periodic cubic domain of the form  $\Omega := [0, N]^3$ , with  $N$  a positive integer, and later extend the results to parallelepiped domains which allow for more regular tetrahedral meshes. The physical plane wave has the following form:

$$\mathbf{u}(\mathbf{x}, t) = \mathbf{a}e^{\hat{i}(\boldsymbol{\kappa} \cdot \mathbf{x} - \omega t)}, \quad \mathbf{x} \in \Omega, t \in [0, T], \tag{7}$$

where  $\hat{i} := \sqrt{-1}$  is the imaginary number,  $\boldsymbol{\kappa} \in \mathbb{R}^3$  is the wave vector,  $\omega \in \mathbb{R}$  is the angular velocity, and  $\mathbf{a} \in \mathbb{R}^m$  is the amplitude vector. The wave vector must be of the form  $\boldsymbol{\kappa} = \boldsymbol{\kappa}_{\mathbf{z}} = \frac{2\pi}{N}\mathbf{z}$ , with  $\mathbf{z} \in \mathbb{Z}_N^3$ , in order to satisfy the periodic boundary conditions.

The numerical plane wave can be written in a similar form when using a periodic mesh. To obtain a periodic tetrahedral mesh, we subdivide the unit cell  $\Omega_0 := [0, 1]^3$  into tetrahedra and repeat this pattern  $N \times N \times N$  times to fill the entire domain as illustrated in Fig. 1. We equip the mesh with a translation-invariant set of basis functions where each basis function has minimal support. In case of mass-lumping we use nodal basis functions and in case of DG we use basis functions that have support on only a single element. The numerical plane wave  $\underline{\mathbf{U}}_h$  of the fully discrete scheme then has the form

$$\underline{\mathbf{U}}_h(\Omega_{\mathbf{k}}, t_i) = \underline{\mathbf{U}}_{h, \Omega_0} e^{\hat{i}(\boldsymbol{\kappa} \cdot \mathbf{x}_{\mathbf{k}} - \omega_i t)}, \quad i = 0, \dots, N_T, \mathbf{k} \in \mathbb{Z}_N^3. \tag{8}$$

Here,  $\underline{\mathbf{U}}_h(\Omega_{\mathbf{k}}, t_i)$  denotes the coefficients of the basis functions corresponding to cell  $\Omega_{\mathbf{k}} := \mathbf{k} + \Omega_0$  at time  $t_i$ . In case of mass-lumping, these basis functions are the nodal basis functions corresponding to the nodes on  $\Omega_{\mathbf{k}} = \mathbf{k} + [0, 1]^3$ , while in case of DG, these are the basis functions that have support on one of the tetrahedra in  $\Omega_{\mathbf{k}}$ . The vector  $\underline{\mathbf{U}}_{h, \Omega_0} \in \mathbb{R}^{n_0}$  denotes the basis function coefficients corresponding to cell  $\Omega_0$  at time 0 and  $\mathbf{x}_{\mathbf{k}} = \mathbf{k}$  are the coordinates of the front-left-bottom vertex of cell  $\Omega_{\mathbf{k}}$ .

To show that this is indeed a numerical plane wave, let  $M^{(\Omega_{\mathbf{k}}, \Omega_{\mathbf{m}})}, A^{(\Omega_{\mathbf{k}}, \Omega_{\mathbf{m}})} \in \mathbb{R}^{n_0 \times n_0}$ , for  $\mathbf{k}, \mathbf{m} \in \mathbb{Z}_N^3$ , be submatrices of  $M$  and  $A$ , respectively, defined as follows:

$$\begin{aligned} M_{ij}^{(\Omega_{\mathbf{k}}, \Omega_{\mathbf{m}})} &:= \left( \rho \mathbf{w}^{(\Omega_{\mathbf{k}}, i)}, \mathbf{w}^{(\Omega_{\mathbf{m}}, j)} \right)_h, & i, j = 1, \dots, n_0, \\ A_{ij}^{(\Omega_{\mathbf{k}}, \Omega_{\mathbf{m}})} &:= a_h \left( \mathbf{w}^{(\Omega_{\mathbf{k}}, i)}, \mathbf{w}^{(\Omega_{\mathbf{m}}, j)} \right), & i, j = 1, \dots, n_0, \end{aligned}$$

where  $\{\mathbf{w}^{(\Omega_{\mathbf{k}}, i)}\}_{i=0}^{n_0}$  denote the basis functions corresponding to cell  $\Omega_{\mathbf{k}}$  and where  $a_h = a_h^{(DG)}$  and  $(\cdot, \cdot)_h = (\cdot, \cdot)$  in case of the DG method and  $a_h = a$  and  $(\cdot, \cdot)_h = (\cdot, \cdot)_h^{(L)}$  in case of the mass-lumped method.

Since the basis functions are translation invariant, the submatrices  $M^{(\Omega_{\mathbf{k}}, \Omega_{\mathbf{k} + \Delta \mathbf{k}})}$  and  $A^{(\Omega_{\mathbf{k}}, \Omega_{\mathbf{k} + \Delta \mathbf{k}})}$  are the same for any  $\mathbf{k} \in \mathbb{Z}_N^3$  with  $\Delta \mathbf{k} \in \mathbb{Z}_N^3$  fixed. Furthermore, the submatrices  $M^{(\Omega_{\mathbf{k}}, \Omega_{\mathbf{m}})}$  are only non-zero when  $\mathbf{k} = \mathbf{m}$ , since the mass matrix is diagonal in case

of mass-lumping and block-diagonal, with each block corresponding to an element, in case of DG. The submatrices  $A^{(\Omega_{\mathbf{k}}, \Omega_{\mathbf{k}+\Delta\mathbf{k}})}$  are only non-zero when  $\Delta\mathbf{k} \in \{-1, 0, 1\}^3$ , since the nodal basis functions for mass-lumping and the local basis functions for DG do not interact when they are two or more cells apart. This implies that we only need to consider the submatrices  $M^{(\Omega_0)} := M^{(\Omega_0, \Omega_0)}$  and  $A^{(\Omega_0, \Omega_{\Delta\mathbf{k}})}$  for  $\Delta\mathbf{k} \in \{-1, 0, 1\}^3$ .

Now let  $\kappa = \kappa_{\mathbf{z}} := \frac{2\pi}{N}\mathbf{z}$ , for some  $\mathbf{z} \in \mathbb{Z}_N^3$ , and let  $\underline{\mathbf{U}}_{h,0} \in \mathbb{R}^{N^3 \times n_0}$  be the numerical wave at time  $t = 0$ :

$$\underline{\mathbf{U}}_{h,0}(\Omega_{\mathbf{k}}) := \underline{\mathbf{U}}_{h,\Omega_0} e^{i(\kappa \cdot \mathbf{x}_{\mathbf{k}})}, \quad \mathbf{k} \in \mathbb{Z}_N^3. \tag{9}$$

Then  $M^{-1}A\underline{\mathbf{U}}_{h,0}$  satisfies

$$\begin{aligned} (M^{-1}A\underline{\mathbf{U}}_{h,0})(\Omega_{\mathbf{k}}) &= M_{inv}^{(\Omega_0)} \left( \sum_{\Delta\mathbf{k} \in \{-1,0,1\}^3} A^{(\Omega_0, \Omega_{\Delta\mathbf{k}})} \underline{\mathbf{U}}_{h,0}(\Omega_{\mathbf{k}+\Delta\mathbf{k}}) \right) \\ &= M_{inv}^{(\Omega_0)} \left( \sum_{\Delta\mathbf{k} \in \{-1,0,1\}^3} e^{i(\kappa \cdot \mathbf{x}_{\Delta\mathbf{k}})} A^{(\Omega_0, \Omega_{\Delta\mathbf{k}})} \right) \underline{\mathbf{U}}_{h,\Omega_0} e^{i(\kappa \cdot \mathbf{x}_{\mathbf{k}})} \\ &= M_{inv}^{(\Omega_0)} A^{(\kappa)} \underline{\mathbf{U}}_{h,\Omega_0} e^{i(\kappa \cdot \mathbf{x}_{\mathbf{k}})}, \end{aligned}$$

for all  $\mathbf{k} \in \mathbb{Z}_N^3$ , with  $M_{inv}^{(\Omega_0)}$  the inverse of  $M^{(\Omega_0)}$  and

$$A^{(\kappa)} := \sum_{\Delta\mathbf{k} \in \{-1,0,1\}^3} e^{i(\kappa \cdot \mathbf{x}_{\Delta\mathbf{k}})} A^{(\Omega_0, \Omega_{\Delta\mathbf{k}})}.$$

This implies that if  $(s_h, \underline{\mathbf{U}}_{h,\Omega_0})$  is an eigenpair of  $S^{(\kappa)} := M_{inv}^{(\Omega_0)} A^{(\kappa)}$ , then  $(s_h, \underline{\mathbf{U}}_{h,0})$  is an eigenpair of  $M^{-1}A$ . In other words, we can obtain eigenpairs of  $M^{-1}A$  by computing the eigenpairs of a small matrix  $S^{(\kappa)} \in \mathbb{R}^{n_0 \times n_0}$ . Note that since  $M^{(\Omega_0)}$  is symmetric positive definite, and  $A^{(\kappa)}$  is Hermitian,  $S^{(\kappa)}$  has  $n_0$  distinct eigenpairs. Since there are  $N^3$  choices for  $\mathbf{z} \in \mathbb{Z}_N^3$  and  $S^{(\kappa_{\mathbf{z}})}$  has  $n_0$  eigenpairs, we can obtain all of the  $N^3 \times n_0$  eigenpairs of  $M^{-1}A$  in this way.

Now consider the numerical plane wave in (8) with  $(s_h, \underline{\mathbf{U}}_{h,\Omega_0})$  an eigenpair of  $S^{(\kappa)}$ , so with  $(s_h, \underline{\mathbf{U}}_{h,0})$  an eigenpair of  $M^{-1}A$ . We can rewrite  $\underline{\mathbf{U}}_h$  as  $\underline{\mathbf{U}}_h(t) = \underline{\mathbf{U}}_{h,0} e^{-i(\omega t)}$ . If we then substitute this wave into (6) we obtain

$$\cos(\Delta t \omega_h) \underline{\mathbf{U}}_h(t_i) = \sum_{k=0}^K \frac{1}{(2k)!} (-\Delta t^2 s_h)^k \underline{\mathbf{U}}_h(t_i), \quad i = 1, \dots, n_T - 1.$$

From this, it follows that  $\underline{\mathbf{U}}_h$  in (8) is a discrete plane wave if  $(s_h, \underline{\mathbf{U}}_{h,\Omega_0})$  is an eigenpair of  $S^{(\kappa)}$  and if  $\omega_h$  satisfies  $\cos(\Delta t \omega_h) = \sum_{k=0}^K \frac{1}{(2k)!} (-\Delta t^2 s_h)^k$ , so if

$$\omega_h = \pm \frac{1}{\Delta t} \arccos \left( \sum_{k=0}^K \frac{1}{(2k)!} (-\Delta t^2 s_h)^k \right). \tag{10}$$

It remains to determine the time step size  $\Delta t$ . In the ‘‘Appendix’’ we show that the numerical scheme is stable, if

$$\Delta t \leq \sqrt{c_K / \sigma_{max}(M^{-1}A)}, \tag{11}$$

where  $\sigma_{max}(M^{-1}A)$  denotes the spectral radius of  $M^{-1}A$  and  $c_K$  is a constant, given by

$$c_K := \inf \left\{ x \geq 0 \mid \left| \sum_{k=0}^K \frac{1}{(2k)!} (-x)^k \right| > 1 \right\}. \tag{12}$$

To obtain a bound on the spectral radius, recall that we can write every eigenpair of  $M^{-1}A$  in the form of  $(s_h, \underline{\mathbf{U}}_{h,0})$ , with  $\underline{\mathbf{U}}_{h,0}$  given in (9) and with  $(s_h, \underline{\mathbf{U}}_{h,\Omega_0})$  an eigenpair of  $S^{(\kappa_{\mathbf{z}})}$  for some  $\mathbf{z} \in \mathbb{Z}_N^3$ . This implies that  $\sigma_{max}(M^{-1}A)$  is equal to  $\sup_{\mathbf{z} \in \mathbb{Z}_N^3} \sigma_{max}(S^{(\kappa_{\mathbf{z}})})$ . We can therefore bound  $\sigma_{max}(M^{-1}A)$  as follows:

$$\sigma_{max}(M^{-1}A) = \sup_{\mathbf{z} \in \mathbb{Z}_N^3} \sigma_{max}(S^{(\kappa_{\mathbf{z}})}) \leq \sup_{\kappa \in \mathcal{K}_0} \sigma_{max}(S^{(\kappa)}) =: s_{h,max}, \tag{13}$$

with  $\mathcal{K}_0 := [0, 2\pi]^3 \supset \{\kappa_{\mathbf{z}}\}_{\mathbf{z} \in \mathbb{Z}_N^3}$  the space of all distinct wave vectors  $\kappa$ .

The constants  $c_K$  can be computed numerically. For example,  $c_K = 4, 12, 7.57$  for  $K = 1, 2, 3$ , respectively. For higher values of  $K$ , see, for example, [27], where his  $\sigma_t$  satisfies  $c_K = 2\sigma_t$ .

We can extend the results of this section to parallelepiped cells by applying a linear transformation  $\mathbf{x} \rightarrow \mathbf{T} \cdot \mathbf{x}$ , with  $\mathbf{T} \in \mathbb{R}^{3 \times 3}$  a second-order tensor. The parallelepiped domain is then given by  $\Omega = \mathbf{T} \cdot (0, N)^3$  and the cells are given by  $\Omega_0 = \mathbf{T} \cdot [0, 1]^3$  and  $\Omega_{\mathbf{k}} = \mathbf{x}_{\mathbf{k}} + \Omega_0$ , with  $\mathbf{x}_{\mathbf{k}} = \mathbf{T} \cdot \mathbf{k}$  the front-left-bottom vertex. The wave vectors  $\kappa_{\mathbf{z}}$  are of the form  $\kappa_{\mathbf{z}} = \frac{2\pi}{N}(\mathbf{T}^{-t} \cdot \mathbf{z})$  and the wave vector space  $\mathcal{K}_0$  is given by  $\mathcal{K}_0 := \mathbf{T}^{-t} \cdot [0, 2\pi)^3$ , with  $\mathbf{T}^{-t}$  the transposed inverse of  $\mathbf{T}$ .

### 5.2 Computing the Dispersion and Eigenvector Error

To explain how we compute the dispersion error, we first consider the acoustic wave equation. Let  $\kappa$  be a given wave vector and let  $\mathbf{u}^{(\kappa)}$  be the acoustic plane wave given by  $\mathbf{u}^{(\kappa)}(\mathbf{x}, t) = e^{i(\kappa \cdot \mathbf{x} - \omega t)}$ . The angular velocity is given by  $\omega = \pm c|\kappa|$  with  $c$  the acoustic wave propagation speed. We compare this plane wave with the numerical plane waves.

To do this, we use the results from the previous subsection. There, we showed that for any eigenpair  $(s_h, \underline{\mathbf{U}}_{h,\Omega_0})$  of  $S^{(\kappa)}$  we can obtain a numerical plane wave in the form of (8) with angular velocity  $\pm \omega_h$  given by (10). Since  $S^{(\kappa)}$  has  $n_0$  eigenpairs, this means we can obtain  $n_0$  discrete plane waves  $\{\underline{\mathbf{U}}_h^{(\kappa,i)}\}_{i=1}^{n_0}$ , with angular velocities  $\{\pm \omega_h^{(\kappa,i)}\}_{i=1}^{n_0}$ , for a given wave vector  $\kappa$ . The corresponding wave propagation speeds  $\{c_h^{(\kappa,i)}\}_{i=1}^{n_0}$  can be computed by  $c_h^{(\kappa,i)} = |\omega_h^{(\kappa,i)}|/|\kappa|$  and we can order the numerical plane waves such that

$$|c - c_h^{(\kappa,1)}| \leq |c - c_h^{(\kappa,2)}| \leq \dots$$

We consider  $\underline{\mathbf{U}}_h^{(\kappa,1)}$  to be the matching numerical plane wave and  $\underline{\mathbf{U}}_h^{(\kappa,i)}$ , with  $i > 1$ , to be spurious modes. We then define the dispersion error as follows

$$e_{disp}(\kappa) = \frac{|c - c_h^{(\kappa,1)}|}{c}.$$

The complete procedure for computing  $e_{disp}(\kappa)$  in the acoustic case is given by

- a. Compute all eigenpairs  $(s_h^{(\kappa,i)}, \underline{\mathbf{U}}_{h,\Omega_0}^{(\kappa,i)})$  of  $S^{(\kappa)} := M_{inv}^{(\Omega_0)} A(\kappa)$ .
- b. Compute the angular velocities  $\omega_h^{(\kappa,i)} = \frac{1}{\Delta t} \arccos \left( \sum_{k=0}^K \frac{1}{(2k)!} (-\Delta t^2 s_h^{(\kappa,i)})^k \right)$ .

- c. Compute the wave propagation speeds  $c_h^{(\kappa,i)} = \omega_h^{(\kappa,i)} / |\kappa|$  and order everything such that  $|c - c_h^{(\kappa,1)}| \leq |c - c_h^{(\kappa,2)}| \leq \dots$
- d. Compute  $e_{disp}(\kappa) = |c - c_h^{(\kappa,1)}| / c$ .

Now let  $\mathbf{u}_0^{(\kappa)}(\mathbf{x}) := \mathbf{u}^{(\kappa)}(\mathbf{x}, 0) = e^{i(\kappa \cdot \mathbf{x})}$  be the acoustic plane wave at  $t = 0$ . Also, let  $\underline{\mathbf{u}}_0^{(\kappa)}$  be the projection of  $\mathbf{u}_0^{(\kappa)}$  onto the numerical space, and let  $\underline{\mathbf{U}}_{h,0}^{(\kappa,i)}$  be the discrete plane wave at  $t = 0$ . In the ideal case,  $\underline{\mathbf{u}}_0^{(\kappa)}$  is equal to  $\underline{\mathbf{U}}_{h,0}^{(\kappa,1)}$  up to a constant. In most cases, however, the projection  $\underline{\mathbf{u}}_0^{(\kappa)}$  is a superposition of a well-matching plane wave  $\underline{\mathbf{U}}_{h,0}^{(\kappa,1)}$  and several other plane waves  $\underline{\mathbf{U}}_{h,0}^{(\kappa,i)}$ , for  $i > 1$ , that may have a completely different shape and velocity. We can compute the number of these spurious waves by computing the projection error.

To do this, we let  $\underline{\mathbf{U}}_0^{(\kappa)} \in \text{span}\{\underline{\mathbf{U}}_{h,0}^{(\kappa,1)}\}$  denote the projection of  $\underline{\mathbf{u}}_0^{(\kappa)}$  onto  $\text{span}\{\underline{\mathbf{U}}_{h,0}^{(\kappa,1)}\}$ , such that  $(\underline{\mathbf{U}}_0^{(\kappa)}, \underline{\mathbf{U}}_{h,0}^{(\kappa,1)})_M = (\underline{\mathbf{u}}_0^{(\kappa)}, \underline{\mathbf{U}}_{h,0}^{(\kappa,1)})_M$ , with  $(\underline{\mathbf{u}}, \underline{\mathbf{v}})_M := \underline{\mathbf{u}}^T M \underline{\mathbf{v}}$ . We then define the projection error as

$$e_{vec}(\kappa) := \frac{\|\underline{\mathbf{u}}_0^{(\kappa)} - \underline{\mathbf{U}}_0^{(\kappa)}\|_M}{\|\underline{\mathbf{u}}_0^{(\kappa)}\|_M},$$

where  $\|\underline{\mathbf{u}}\|_M := \sqrt{\underline{\mathbf{u}}^T M \underline{\mathbf{u}}}$ . We refer to this as the eigenvector error, since it is related to the accuracy of  $\underline{\mathbf{U}}_{h,0}^{(\kappa,1)}$ , which is an eigenvector of  $M^{-1}A$  [26].

Since the physical plane wave, the mesh, and the set of basis functions are all translation invariant, we can efficiently compute this error by only considering  $\mathbf{u}_{\Omega_0}^{(\kappa)}$ , the part of  $\mathbf{u}_0^{(\kappa)}$  restricted to cell  $\Omega_0$ . We define  $\underline{\mathbf{u}}_{\Omega_0}^{(\kappa)}$  to be the projection of  $\mathbf{u}_{\Omega_0}^{(\kappa)}$  onto the discrete space restricted to  $\Omega_0$  and define  $\underline{\mathbf{U}}_{\Omega_0}^{(\kappa)} \in \text{span}\{\underline{\mathbf{U}}_{h,\Omega_0}^{(\kappa,1)}\}$  the projection of  $\underline{\mathbf{u}}_{\Omega_0}^{(\kappa)}$  onto  $\text{span}\{\underline{\mathbf{U}}_{h,\Omega_0}^{(\kappa,1)}\}$  such that  $(\underline{\mathbf{U}}_{\Omega_0}^{(\kappa)}, \underline{\mathbf{U}}_{h,\Omega_0}^{(\kappa,1)})_{M_0} = (\underline{\mathbf{u}}_{\Omega_0}^{(\kappa)}, \underline{\mathbf{U}}_{h,\Omega_0}^{(\kappa,1)})_{M_0}$ , with  $M_0 := M^{(\Omega_0)}$ . We can then compute  $e_{vec}(\kappa)$  by

$$e_{vec}(\kappa) = \frac{\|\underline{\mathbf{u}}_{\Omega_0}^{(\kappa)} - \underline{\mathbf{U}}_{\Omega_0}^{(\kappa)}\|_{M_0}}{\|\underline{\mathbf{u}}_{\Omega_0}^{(\kappa)}\|_{M_0}}.$$

The complete procedure for computing  $e_{vec}(\kappa)$  in the acoustic case is given by

- A. Compute all eigenpairs  $(s_h^{(\kappa,i)}, \underline{\mathbf{U}}_{h,\Omega_0}^{(\kappa,i)})$  of  $S^{(\kappa)} := M_{inv}^{(\Omega_0)} A^{(\kappa)}$ .
- B. Compute the angular velocities  $\omega_h^{(\kappa,i)} = \frac{1}{\Delta t} \arccos\left(\sum_{k=0}^K \frac{1}{(2k)!} (-\Delta t^2 s_h^{(\kappa,i)})^k\right)$ .
- C. Compute the wave propagation speeds  $c_h^{(\kappa,i)} = \omega_h^{(\kappa,i)} / |\kappa|$  and order everything such that  $|c - c_h^{(\kappa,1)}| \leq |c - c_h^{(\kappa,2)}| \leq \dots$
- D. Compute  $\underline{\mathbf{u}}_{\Omega_0}^{(\kappa)}$ : the projection of  $\mathbf{u}_{\Omega_0}^{(\kappa)}$  onto the discrete space of cell  $\Omega_0$ .
- E. Compute  $\underline{\mathbf{U}}_{\Omega_0}^{(\kappa)}$ : the projection of  $\underline{\mathbf{u}}_{\Omega_0}^{(\kappa)}$  onto  $\text{span}\{\underline{\mathbf{U}}_{h,\Omega_0}^{(\kappa,1)}\}$
- F. Compute  $e_{vec}(\kappa) = \|\underline{\mathbf{u}}_{\Omega_0}^{(\kappa)} - \underline{\mathbf{U}}_{\Omega_0}^{(\kappa)}\|_{M_0} / \|\underline{\mathbf{u}}_{\Omega_0}^{(\kappa)}\|_{M_0}$ .

For the isotropic elastic case, the procedure is very similar. Let  $\kappa$  be the wave vector and let  $\mathbf{u}^{(\kappa)}$  denote the elastic plane wave of the form  $\mathbf{u}^{(\kappa)}(\mathbf{x}, t) = \mathbf{a} e^{i(\kappa \cdot \mathbf{x} - \omega t)}$ , with  $\mathbf{a}$  the amplitude vector,  $\omega = \pm c|\kappa|$  the angular velocity, and  $c$  the elastic wave propagation speed. In the elastic isotropic case, we have to distinguish between longitudinal or primary waves, where  $\mathbf{a}$  is parallel with  $\kappa$  and the propagation speed is  $c = c_P = \sqrt{(\lambda + 2\mu)/\rho}$ , and transversal, shear or secondary waves, where  $\mathbf{a}$  is perpendicular to  $\kappa$  and the propagation speed is  $c = c_S = \sqrt{\mu/\rho}$ .

For the analysis, we will only consider the secondary wave, since the wavelength  $\lambda = 2\pi/|\kappa| = 2\pi c/\omega$  of this wave is shorter and therefore governs the required mesh resolution. In 3D, there are two linear independent amplitude vectors,  $\mathbf{a}^{(\kappa,1)}$  and  $\mathbf{a}^{(\kappa,2)}$ , that are perpendicular to  $\kappa$  and we will refer to the corresponding secondary plane waves as  $\underline{\mathbf{u}}^{(\kappa,1)}$  and  $\underline{\mathbf{u}}^{(\kappa,2)}$ . We will compare these physical plane waves with the numerical plane waves in a similar way as for the acoustic case.

Since, for a given  $\kappa$  and  $\omega = \pm c_S|\kappa|$ , there are two linearly independent secondary waves, we compare the secondary wave velocity  $c = c_S$  with the wave propagation speed of the two best matching numerical plane waves. In particular, we define the dispersion error as

$$e_{disp}(\kappa) = \frac{|c - c_h^{(\kappa,2)}|}{c}.$$

The procedures for computing this error is the same as for the acoustic case, with step d replaced by

d\*. Compute  $e_{disp}(\kappa) = |c - c_h^{(\kappa,2)}|/c$ .

The eigenvector is now computed by

$$e_{vec}(\kappa) = \sup_{\underline{\mathbf{u}}_{\Omega_0}^{(\kappa)} \in \text{span}\{\underline{\mathbf{u}}_{\Omega_0}^{(\kappa,1)}, \underline{\mathbf{u}}_{\Omega_0}^{(\kappa,2)}\}} \frac{\|\underline{\mathbf{u}}_{\Omega_0}^{(\kappa)} - \underline{\mathbf{U}}_{\Omega_0}^{(\kappa)}\|_{M_0}}{\|\underline{\mathbf{u}}_{\Omega_0}^{(\kappa)}\|_{M_0}},$$

where  $\underline{\mathbf{u}}_{\Omega_0}^{(\kappa,i)}$  is the projection of  $\mathbf{u}_{\Omega_0}^{(\kappa,i)}$  onto the discrete space of cell  $\Omega_0$ , and  $\underline{\mathbf{U}}_{\Omega_0}^{(\kappa)} \in \text{span}\{\underline{\mathbf{U}}_{h,\Omega_0}^{(\kappa,1)}, \underline{\mathbf{U}}_{h,\Omega_0}^{(\kappa,2)}\}$  is the projection of  $\mathbf{u}_{\Omega_0}^{(\kappa)} \in \text{span}\{\mathbf{u}_{\Omega_0}^{(\kappa,1)}, \mathbf{u}_{\Omega_0}^{(\kappa,2)}\}$  onto  $\text{span}\{\underline{\mathbf{U}}_{h,\Omega_0}^{(\kappa,1)}, \underline{\mathbf{U}}_{h,\Omega_0}^{(\kappa,2)}\}$ . In other words, we compute the worst possible projection error for a linear combination of  $\underline{\mathbf{u}}_{\Omega_0}^{(\kappa,1)}$  and  $\underline{\mathbf{u}}_{\Omega_0}^{(\kappa,2)}$  projected onto the span of the two best-matching numerical plane waves  $\underline{\mathbf{U}}_{h,\Omega_0}^{(\kappa,1)}$  and  $\underline{\mathbf{U}}_{h,\Omega_0}^{(\kappa,2)}$ . We can efficiently compute this by

$$e_{vec}(\kappa) = \sqrt{\sigma_{max}(B^{-1}R)},$$

where  $\sigma_{max}(B^{-1}R)$  denotes the largest eigenvalue of  $B^{-1}R$  and  $B, R \in \mathbb{R}^{2 \times 2}$  are matrices given by  $B_{ij} := (\underline{\mathbf{u}}_{\Omega_0}^{(\kappa,i)}, \underline{\mathbf{u}}_{\Omega_0}^{(\kappa,j)})_{M_0}$  and  $R_{ij} := (\underline{\mathbf{u}}_{\Omega_0}^{(\kappa,i)} - \underline{\mathbf{U}}_{\Omega_0}^{(\kappa,i)}, \underline{\mathbf{u}}_{\Omega_0}^{(\kappa,j)} - \underline{\mathbf{U}}_{\Omega_0}^{(\kappa,j)})_{M_0}$ , with  $\underline{\mathbf{U}}_{\Omega_0}^{(\kappa,i)}$  the projection of  $\mathbf{u}_{\Omega_0}^{(\kappa,i)}$  onto  $\text{span}\{\underline{\mathbf{U}}_{h,\Omega_0}^{(\kappa,1)}, \underline{\mathbf{U}}_{h,\Omega_0}^{(\kappa,2)}\}$ .

The procedure for computing  $e_{vec}(\kappa)$  is the same as for the acoustic case, with steps D-F replaced by

- D\*. Compute  $\underline{\mathbf{u}}_{\Omega_0}^{(\kappa,i)}$ : the projection of  $\mathbf{u}_{\Omega_0}^{(\kappa,i)}$  onto the discrete space of cell  $\Omega_0$ , for  $i = 1, 2$ .
- E\*. Compute  $\underline{\mathbf{U}}_{\Omega_0}^{(\kappa,i)}$ : the projection of  $\mathbf{u}_{\Omega_0}^{(\kappa,i)}$  onto  $\text{span}\{\underline{\mathbf{U}}_{h,\Omega_0}^{(\kappa,1)}, \underline{\mathbf{U}}_{h,\Omega_0}^{(\kappa,2)}\}$ , for  $i = 1, 2$ .
- F\*. Compute  $B_{ij} := (\underline{\mathbf{u}}_{\Omega_0}^{(\kappa,i)}, \underline{\mathbf{u}}_{\Omega_0}^{(\kappa,j)})_{M_0}$  and  $R_{ij} := (\underline{\mathbf{u}}_{\Omega_0}^{(\kappa,i)} - \underline{\mathbf{U}}_{\Omega_0}^{(\kappa,i)}, \underline{\mathbf{u}}_{\Omega_0}^{(\kappa,j)} - \underline{\mathbf{U}}_{\Omega_0}^{(\kappa,j)})_{M_0}$ , for  $i, j = 1, 2$ , and use this to compute  $e_{vec}(\kappa) = \sqrt{\sigma_{max}(B^{-1}R)}$ .

So far, we only considered the dispersion error and eigenvector error for a given wave vector  $\kappa$ . For a given wavelength  $\lambda = 2\pi/|\kappa|$ , we define the dispersion and eigenvector error as the worst case among all wave vectors of length  $|\kappa| = \lambda/(2\pi)$ , so among wave vectors in all possible directions:

$$e_{disp}(\lambda) := \sup_{\kappa \in \mathbb{R}^3, |\kappa| = \lambda/(2\pi)} e_{disp}(\kappa), \tag{14a}$$

$$e_{vec}(\lambda) := \sup_{\kappa \in \mathbb{R}^3, |\kappa| = \lambda/(2\pi)} e_{vec}(\kappa). \tag{14b}$$

We can use these errors to determine the required number of elements per wavelength. To compute these errors, we use a search algorithm, which requires the computation of  $e_{disp}(\kappa)$  and  $e_{vec}(\kappa)$  for a large number of wave vectors  $\kappa$ . The complete procedure for computing the dispersion and eigenvector error is given by:

1. Construct a cell  $\Omega_0$  and subdivide it into tetrahedra.
2. Compute the submatrices  $M^{(\Omega_0)}$  and  $A^{(\Omega_0, \Omega_{\Delta\mathbf{k}})}$  for  $\Delta\mathbf{k} \in \{-1, 0, 1\}^3$ .
3. Compute  $s_{h,max}$ , given by (13). This is done with a search algorithm which requires the computation of  $\sigma_{max}(S^{(\kappa)})$ , with  $S^{(\kappa)} := M_{inv}^{(\Omega_0)} A^{(\kappa)}$ , for a large number of wave vectors  $\kappa$ .
4. Compute  $\Delta t \leq \sqrt{c_K / s_{h,max}}$ , with  $c_K$  given by (12).
5. For a given wavelength  $\lambda$ , compute the errors  $e_{disp}(\lambda)$  and  $e_{vec}(\lambda)$  given in (14). For each  $\lambda$ , this requires the computation of  $e_{disp}(\kappa)$  and  $e_{vec}(\kappa)$ , using steps a-d and A-F, for a large number of wave vectors  $\kappa$ .

### 5.3 Estimating the Computational Cost

To compare the efficiency of the different methods, we also compute the number of degrees of freedom  $n_{vec}$ , the number of non-zero entries of the stiffness matrix  $n_{mat}$ , and the estimated computational cost  $n_{comp}$ , for each wavelength  $\lambda$ .

We define  $n_{vec}$  to be the number of degrees of freedom per  $\lambda^3$ -volume. This is computed by

$$n_{vec} = n_0 \frac{\lambda^3}{|\Omega_0|},$$

where  $n_0$  is the number of basis functions corresponding to cell  $\Omega_0$ , and  $|\Omega_0|$  is the volume of  $\Omega_0$ .

We define  $n_{mat}$  to be the number of non-zero entries of the stiffness matrix per  $\lambda^3$ -volume. In case of mass-lumping, we estimate this number by

$$n_{mat}^{(ML)} = \left( \sum_{q \in \mathcal{Q}_{\Omega_0}} \sum_{q' \in \mathcal{N}(q)} |\mathcal{U}_q| |\mathcal{U}_{q'}| \right) \frac{\lambda^3}{|\Omega_0|},$$

where  $|\mathcal{U}_q|$  is the number of degrees of freedom per node ( $|\mathcal{U}_q| = 1$  in the acoustic and  $|\mathcal{U}_q| = 3$  in the elastic case),  $\mathcal{Q}_{\Omega_0}$  is the set of nodes on  $\Omega_0$ , and  $\mathcal{N}(q)$  are the neighbouring nodes of  $q$  that are connected with  $q$  through an element.

In case of the SIPDG method, we estimate this number by

$$n_{mat}^{(DG)} = \left( \sum_{e \in \mathcal{T}_{\Omega_0}} \sum_{e' \in \mathcal{N}(e)} |\mathcal{U}_e| |\mathcal{U}_{e'}| \right) \frac{\lambda^3}{|\Omega_0|},$$

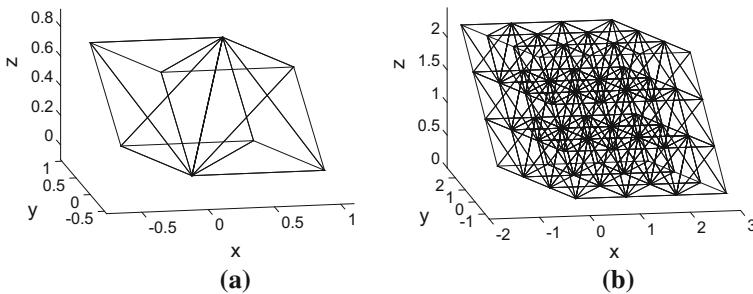
where  $|\mathcal{U}_e|$  is the number of basis functions with support on element  $e$ ,  $\mathcal{T}_{\Omega_0}$  are the elements in  $\Omega_0$ , and  $\mathcal{N}(e)$  are the neighbouring elements of  $e$  that are connected with  $e$  through a face.

To estimate the computational cost we look at the size of the matrix times the number of matrix-vector products. The resulting estimates gives a rough estimate of the relative CPU time of the different methods, since it estimates the number of computations when using a globally assembled matrix.

We define the computational cost  $n_{comp}$  as the number of non-zero matrix entries per  $\lambda^3$ -volume times the number of matrix-vector products during one oscillation in time.

**Table 1** Analysed finite element methods

Method	Description
ML1	Linear mass-lumped finite element method
ML2	Degree-2 mass-lumped finite element method [25]
ML3a, ML3b	Degree-3 mass-lumped finite element methods [5]
DGX	Symmetric Interior Penalty Discontinuous Galerkin method [17] of degree $X = 1, 2, 3$
DGXa	DGX with penalty term derived in [16] and given by (4a)
DGXb	DGX with penalty term derived in [28] and given by (4b)



**Fig. 2** Tetragonal disphenoid honeycomb restricted to cell  $\Omega_0$  (a) and restricted to  $3 \times 3 \times 3$  cells (b)

The duration of one oscillation is  $T_0 = \lambda/c$ , with  $c$  the wave propagation speed. The number of matrix-vector products during one oscillation is the number of stages of the Lax–Wendroff scheme  $K$  times the number of time steps  $N_{\Delta t} = T_0/\Delta t = \lambda/(c\Delta t)$ , where  $\Delta t = \sqrt{c_K/s_{h,max}}$ , with  $c_K$  given by (12) and  $s_{h,max}$  given by (13). We use this to compute  $n_{comp}$  as follows:

$$n_{comp} = n_{mat} K N_{\Delta t}.$$

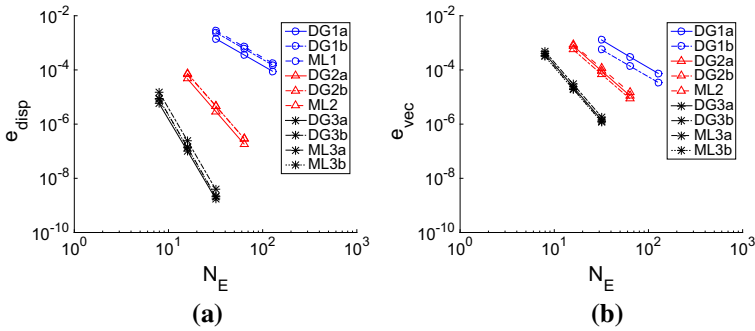
## 6 Results and Comparisons

An overview of the different finite element methods that we analyse is given in Table 1. Each method is combined with an order- $2p$  Lax–Wendroff time integration scheme, where  $p$  denotes the degree of the spatial discretization.

To analyse the dispersion properties of these methods, we use standard Fourier analysis, as explained in Sect. 5. We consider a periodic mesh of congruent nearly-regular equifacial tetrahedra, known as the tetragonal disphenoid honeycomb. To obtain this mesh, we slice the unit cell  $\Omega_0 := [0, 1]^3$  into 6 tetrahedra with the planes  $x = y$ ,  $x = z$ , and  $y = z$  and then apply the linear transformation  $\mathbf{x} \rightarrow \mathbf{T} \cdot \mathbf{x}$ , with

$$\mathbf{T} := \begin{bmatrix} 1 & -1/3 & -1/3 \\ 0 & \sqrt{8/9} & -\sqrt{2/9} \\ 0 & 0 & \sqrt{2/3} \end{bmatrix}. \tag{15}$$

An illustration of this mesh is given in Fig. 2.



**Fig. 3** Dispersion error (a) and eigenvector error (b) for the acoustic wave model

**Table 2** Approximation of the dispersion and eigenvector error for the acoustic case

Method	$e_{disp}$	$e_{vec}$
DG1a	$1.45(N_E)^{-2}$	$1.20(N_E)^{-2}$
DG1b	$2.46(N_E)^{-2}$	$0.56(N_E)^{-2}$
ML1	$2.87(N_E)^{-2}$	0
DG2a	$3.00(N_E)^{-4}$	$2.89(N_E)^{-3}$
DG2b	$4.83(N_E)^{-4}$	$2.22(N_E)^{-3}$
ML2	$4.82(N_E)^{-4}$	$3.78(N_E)^{-3}$
DG3a	$1.77(N_E)^{-6}$	$1.46(N_E)^{-4}$
DG3b	$3.98(N_E)^{-6}$	$1.88(N_E)^{-4}$
ML3a	$2.25(N_E)^{-6}$	$1.26(N_E)^{-4}$
ML3b	$2.15(N_E)^{-6}$	$1.22(N_E)^{-4}$

### 6.1 Acoustic Waves on a Regular Mesh

We first consider the acoustic wave model with  $c = \rho = 1$ . Figure 3 illustrates the dispersion and eigenvector error with respect to the number of elements per wavelength  $N_E := \sqrt[3]{\lambda^3/|e|_{av}}$ , with  $\lambda$  the wavelength and  $|e|_{av}$  the average element volume. The eigenvector error for ML1 is always zero, since it has only one degree of freedom per cell  $\Omega_{\mathbf{k}}$  and therefore allows only one numerical plane wave for a given wave vector. From this figure we can obtain the order of convergence, which is  $2p$  for the dispersion error and  $p + 1$  for the eigenvector error. These convergence rates are typical for symmetric finite element methods for eigenvalue problems, see, for example, [4] and the references therein. The  $2p$ -order superconvergence of the dispersion error is also in accordance with the results of [2, 13, 26].

By extrapolating the results shown in Fig. 3 we can obtain approximations of the errors of the form  $e = \alpha(N_E)^{-\beta}$ , where  $\alpha$  is the leading constant and  $\beta$  is the order of convergence. The approximations are given in Table 2.

We can use these results to obtain estimates for the number of elements per wavelength required for a given accuracy, but we can also use them to obtain other properties, such as the number of time steps or the computational cost required for a given accuracy. An overview for a dispersion error of 0.01 and 0.001 is given in Tables 3 and 4, respectively, and the relation between the accuracy and the computational cost is illustrated in Fig. 4.



**Table 3** Number of elements per wavelength  $N_E$ , number of degrees of freedom  $n_{vec}$ , size of the global matrix  $n_{mat}$ , number of time steps  $N_{\Delta t}$ , computational cost  $n_{comp}$  and eigenvector error  $e_{vec}$  for a dispersion error of 0.01 for different finite element methods for the acoustic wave model

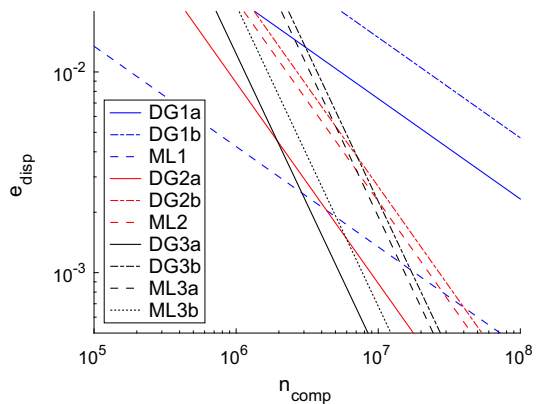
Method	$e_{disp} = 0.01$					
	$N_E$	$n_{vec}$	$n_{mat}$	$N_{\Delta t}$	$n_{comp}$	$e_{vec}$
DG1a	12	7000	$140 \times 10^3$	39	$5.4 \times 10^6$	0.0083
DG1b	16	15,000	$310 \times 10^3$	72	$22 \times 10^6$	0.0023
ML1	17	810	$12 \times 10^3$	15	$0.18 \times 10^6$	0
DG2a	4.2	720	$36 \times 10^3$	12	$0.88 \times 10^6$	0.040
DG2b	4.7	1000	$51 \times 10^3$	26	$2.7 \times 10^6$	0.022
ML2	4.7	860	$39 \times 10^3$	29	$2.3 \times 10^6$	0.037
DG3a	2.4	270	$27 \times 10^3$	14	$1.1 \times 10^6$	0.046
DG3b	2.7	400	$40 \times 10^3$	31	$3.7 \times 10^6$	0.035
ML3a	2.5	370	$31 \times 10^3$	36	$3.3 \times 10^6$	0.034
ML3b	2.4	360	$30 \times 10^3$	18	$1.7 \times 10^6$	0.034

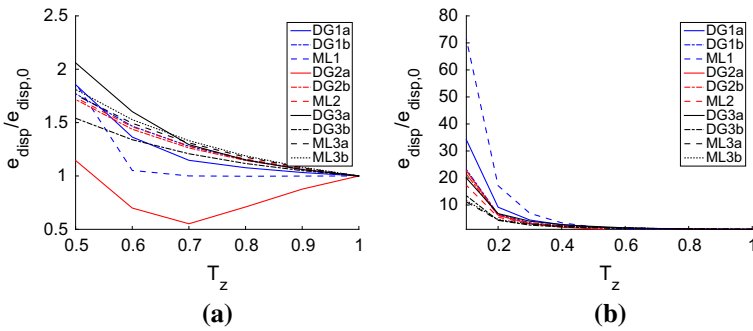
The numbers are accurate up to two decimal places

**Table 4** Same as Table 3, but for a dispersion error of 0.001

Method	$e_{disp} = 0.001$					
	$N_E$	$n_{vec}$	$n_{mat}$	$N_{\Delta t}$	$n_{comp}$	$e_{vec}$
DG1a	38	220,000	$4400 \times 10^3$	120	$540 \times 10^6$	0.00083
DG1b	50	490,000	$9700 \times 10^3$	230	$2200 \times 10^6$	0.00023
ML1	54	26,000	$390 \times 10^3$	47	$18 \times 10^6$	0
DG2a	7.4	4100	$200 \times 10^3$	22	$8.8 \times 10^6$	0.0071
DG2b	8.3	5800	$290 \times 10^3$	46	$27 \times 10^6$	0.0038
ML2	8.3	4800	$220 \times 10^3$	52	$23 \times 10^6$	0.0065
DG3a	3.5	840	$84 \times 10^3$	21	$5.3 \times 10^6$	0.010
DG3b	4.0	1300	$130 \times 10^3$	46	$17 \times 10^6$	0.0075
ML3a	3.6	1200	$98 \times 10^3$	52	$15 \times 10^6$	0.0074
ML3b	3.6	1100	$96 \times 10^3$	27	$7.7 \times 10^6$	0.0073

**Fig. 4** Dispersion error of different finite element methods for the acoustic wave model plotted against the estimated computational cost





**Fig. 5** Relative dispersion error for the disphenoid mesh scaled in the  $z$ -direction by a factor  $T_z$  for the acoustic wave model. Here,  $e_{disp,0}$  denotes the error for the original mesh

Figure 4 shows that for linear elements, the mass-lumped method ML1 is significantly more efficient than the DG methods DG1a and DG1b, while for quadratic elements, DG2a is significantly more efficient than ML2 and DG2b, and for cubic functions, DG3a is slightly more efficient than ML3b and significantly more efficient than DG3b and ML3a. In all cases, the DG methods using the sharper penalty term given by (4a) are significantly more efficient than those using the penalty term given by (4b). For a dispersion error of around 0.01 and higher, the linear mass-lumped method ML1 performs best in terms of computational cost, while for a dispersion error below 0.001 the best method is the DG method with cubic basis functions DG3a or the second degree-3 mass-lumped finite element method ML3b.

Tables 3 and 4 also show that for the case  $p = 1$ , the eigenvector error is always smaller than the dispersion error, but that for higher-order elements, the eigenvector error can become almost 5 times as large when the dispersion error is 0.01 and 10 times as large when the dispersion error is 0.001. This is due to the fact that the dispersion error converges with a faster rate (order  $2p$ ) than the eigenvector error (order  $p + 1$ ) for higher-degree methods.

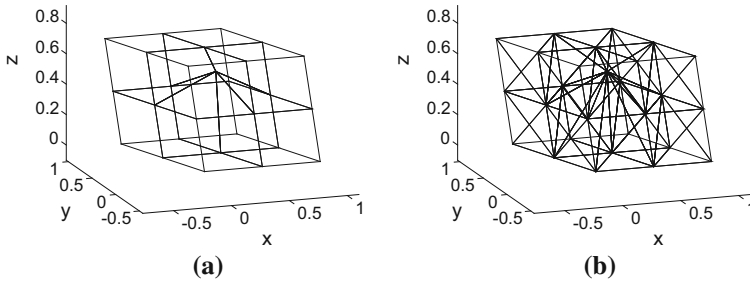
### 6.2 The Effect of Mesh Distortions

We also investigate the effect of the mesh quality on the dispersion error. To do this, we first create meshes of very flat elements by scaling the regular disphenoid mesh in the  $z$ -direction. After that, we create distorted meshes by displacing some of the vertices of the disphenoid honeycomb.

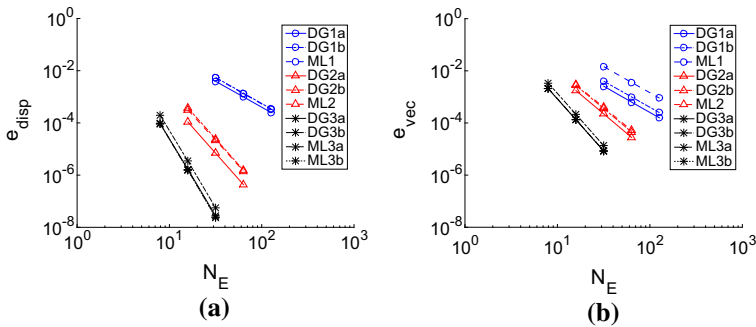
To create flat elements, we scale the disphenoid mesh in the  $z$ -direction by a factor  $T_z$ . The effect on the dispersion error is illustrated in Fig. 5. For a mesh flattened by a factor 2, the dispersion error does not grow more than a factor 2.5, but flattening the mesh by a factor 10 increases the error by a factor between 10 and 100. In all cases, the mesh resolution remains the same and even becomes smaller in the  $z$ -direction. This means that the mesh quality can have a strong effect on the accuracy of the method and that using flat tetrahedra can significantly reduce the accuracy. The methods using lower-order elements are more sensitive to the mesh quality than the higher-order methods.

To create distorted meshes, we displace some of the vertices of the disphenoid mesh. In particular, we create a distorted mesh using the following steps:

1. Slice the cube  $[0, 0.5]^3$  into 6 tetrahedra with the planes  $x = y, x = z, y = z$ .
2. Repeat this pattern  $2 \times 2 \times 2$  times to pack the unit cell  $[0, 1)$  with 48 tetrahedra.



**Fig. 6** Repeated subcells with a small distortion (a) and corresponding tetrahedral mesh (b)



**Fig. 7** Dispersion error (a) and eigenvector error (b) for the acoustic wave model for a distorted mesh with distortion  $\delta = 0.9$

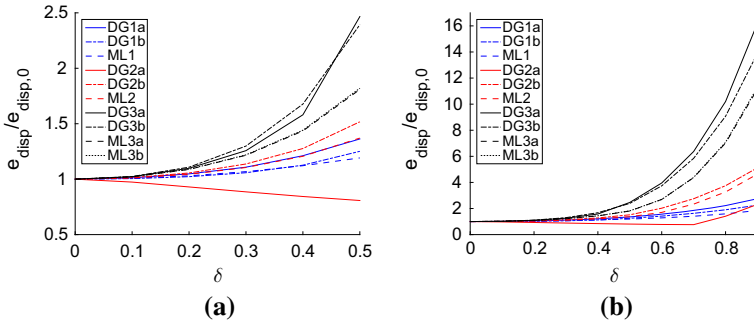
3. Displace the central node by moving it from  $(0.5, 0.5, 0.5)$  to  $(0.5(1 + \delta), 0.5(1 + \delta), 0.5(1 + \delta))$ , where  $\delta \in [0, 1)$  denotes the size of the distortion.
4. Apply the transformation  $\mathbf{x} \rightarrow \mathbf{T} \cdot \mathbf{x}$ , with  $\mathbf{T}$  defined as in (15).

In case of zero distortion,  $\delta = 0$ , we obtain the original disphenoid honeycomb, scaled by a factor 0.5. When the distortion  $\delta$  approaches 1, some of the elements become completely flat with zero volume.

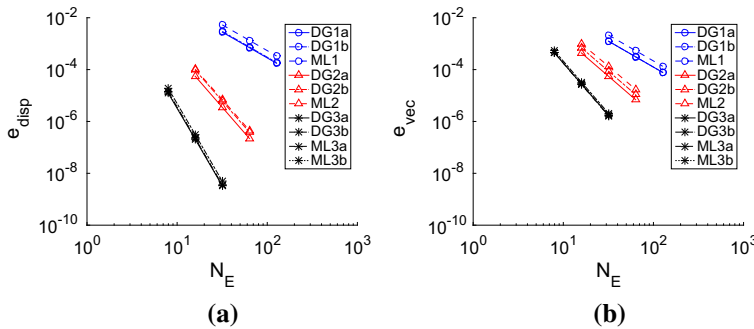
An illustration of the mesh with distortion  $\delta = 0.4$  is given in Fig. 6. In Fig. 7, the dispersion and eigenvector error are plotted against the number of elements per wavelength for a heavily distorted mesh with  $\delta = 0.9$ . These results show that the order of convergence remains  $2p$  for the dispersion and  $p + 1$  for the eigenvector error, even though the mesh is distorted. The distortion does, however, affect the leading constant of the errors. The effect of the mesh distortion on the dispersion error is illustrated in Fig. 8. Again, the accuracy is not significantly affected by small distortions, but large distortions can reduce the accuracy by an order of magnitude.

### 6.3 Elastic Waves and the Effect of the P/S-Wave Velocity Ratio

Besides the acoustic wave model, we also consider the isotropic elastic wave model. Figure 9 illustrates the dispersion and eigenvector error with respect to the number of elements per wavelength for the isotropic elastic wave model with  $\mu = \rho = 1$  and  $\lambda = 2$ , so with a P/S-wave velocity ratio of 2. Again, the order of convergence is  $2p$  for the dispersion error and  $p + 1$  for the eigenvector error.



**Fig. 8** Relative dispersion error for meshes with a distortion  $\delta$ . Here,  $e_{disp,0}$  denotes the error of the regular mesh with  $\delta = 0$



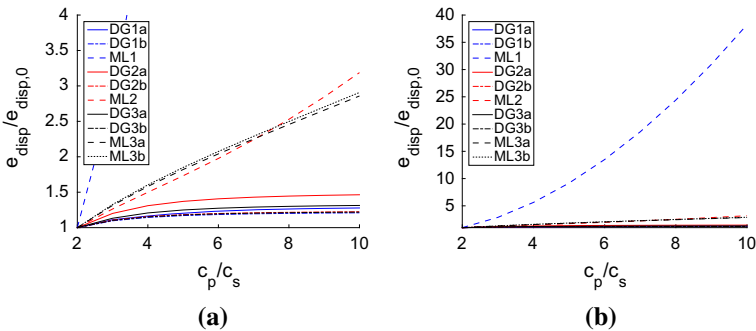
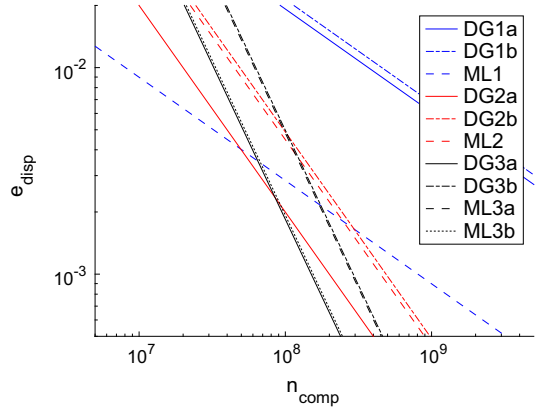
**Fig. 9** Dispersion error **(a)** and eigenvector error **(b)** for the isotropic elastic wave model with a P/S-wave velocity ratio of 2

**Table 5** Approximation of the dispersion and eigenvector error for the elastic wave model with a P/S-wave velocity ratio of 2

Method	$e_{disp}$	$e_{vec}$
DG1a	$2.81(N_E)^{-2}$	$1.25(N_E)^{-2}$
DG1b	$3.00(N_E)^{-2}$	$1.25(N_E)^{-2}$
ML1	$5.39(N_E)^{-2}$	$2.16(N_E)^{-2}$
DG2a	$3.55(N_E)^{-4}$	$1.76(N_E)^{-3}$
DG2b	$6.20(N_E)^{-4}$	$2.77(N_E)^{-3}$
ML2	$7.29(N_E)^{-4}$	$4.39(N_E)^{-3}$
DG3a	$3.32(N_E)^{-6}$	$1.79(N_E)^{-4}$
DG3b	$5.04(N_E)^{-6}$	$2.11(N_E)^{-4}$
ML3a	$3.63(N_E)^{-6}$	$1.66(N_E)^{-4}$
ML3b	$3.58(N_E)^{-6}$	$1.69(N_E)^{-4}$

By extrapolating these results we can again obtain approximations of the errors of the form  $e = \alpha(N_E)^{-\beta}$ , which are given in Table 5. Figure 10 illustrates the relation between the dispersion error and the computational cost, based on these results. The relative performance of the different methods is similar to the acoustic case.

**Fig. 10** Dispersion error of different finite element methods for the isotropic elastic wave model with a P/S-wave velocity ratio of 2, plotted against the estimated computational cost. The graphs of DG3a and ML3b and of DG3b and ML3a are almost identical



**Fig. 11** Relative dispersion error for the isotropic elastic wave model with different  $c_p/c_s$  ratios. Here,  $e_{disp,0}$  denotes the error for the original mesh with  $c_p/c_s = 2$

We also look at the influence of the P/S-wave velocity ratio  $c_p/c_s$  on the dispersion error, where  $c_s = \sqrt{\mu}$  denotes the S-wave velocity and  $c_p = \sqrt{\lambda + 2\mu}$  denotes the P-wave velocity. This relation is illustrated in Fig. 11. This figure shows that the DG methods are not really sensitive to the  $c_p/c_s$  ratio, since the dispersion error never grows more than a factor 1.5. The higher-order mass-lumped methods are slightly more sensitive, with a dispersion error becoming around 3 times as large for  $c_p/c_s = 10$ , compared to  $c_p/c_s = 2$ , while the linear mass-lumped method is very sensitive, with a dispersion error becoming almost 40 times as large in this case.

## 7 Conclusions

We analysed the dispersion properties of two types of explicit finite element methods for modelling wave propagation on tetrahedral meshes, namely mass-lumped finite elements methods and symmetric interior penalty discontinuous Galerkin (SIPDG) methods, both for degrees  $p = 1, 2, 3$  and combined with an order- $2p$  Lax–Wendroff time integration method. The analysed methods are listed in Table 1.

The dispersion properties are obtained semi-analytically using standard Fourier analysis. We used this to give an indication of which method is the most efficient for a given accuracy,

how many elements per wavelength are required for a given accuracy, and how sensitive the accuracy of the method is to poorly shaped elements and high P/S-wave velocity ratios.

Based on the results we draw the following conclusions with regard to efficiency:

- The linear mass-lumped method is the most efficient method for a dispersion error of around 1% when using approximately regular tetrahedra. Heavily distorted elements, however, can significantly reduce its accuracy.
- The degree-3 SIPDG method, with the penalty term derived in [16] and given by (4a), and the second degree-3 mass-lumped finite element method of [5] are the most efficient methods for a dispersion error of around 0.1% and less.
- The SIPDG methods using the sharper penalty term bound derived in [16] are significantly more efficient than those using the penalty term of [28], which is based on the trace inequality of [36].

The required number of elements for a given accuracy can be obtained from the approximations given in Tables 2 and 5. We also draw the following conclusions with regard to accuracy:

- Higher-order methods suffer more from spurious modes for the same dispersion error. This is due to the fact that for higher-order methods, the convergence rate of the dispersion error,  $2p$ , is larger than the convergence rate of the eigenvector,  $p + 1$ .
- All methods are significantly affected by a poor mesh quality, although lower-order methods are more sensitive to this than higher-order methods. Flattening the tetrahedra by a factor 10 reduces the accuracy of the methods by 1–2 orders of magnitude, even though the mesh resolution remains the same and even improves in one direction.
- The SIPDG methods are not really sensitive to high P/S-wave velocity ratios, while the accuracy of the higher-order mass-lumped methods reduces slightly when the P/S-wave velocity ratio is increased. The accuracy of the linear mass-lumped method, however, reduces by an order of magnitude when the P/S-wave velocity ratio is raised from 2 to 10.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## A Stability of the Lax–Wendroff Method

**Theorem A1** Consider the following time integration scheme:

$$U(t_{i+1}) = -U(t_{i-1}) + 2\beta U(t_i), \quad i = 1, 2, \dots,$$

where  $\beta \in \mathbb{R}$  is a constant and  $\{U(t_i)\}_{i \geq 0}$  is a sequence of scalars representing a scalar variable  $u(t)$  at time slots  $t_i = i \Delta t$ , with  $\Delta t$  the time step size. This scheme is stable, by which we mean that the solution grows at most linearly in time, iff  $|\beta| \leq 1$ .

*Proof* If  $\beta \in (-1, 1)$ , then the two independent solutions of the time integration scheme are given by  $U(t_n) = e^{\pm i(t_n \omega)}$ , where  $\omega$  satisfies  $\cos(\omega \Delta t) = \beta$  and  $i = \sqrt{-1}$  is the imaginary number. Otherwise, if  $\beta = 1$  (or  $\beta = -1$ ), then the two independent solutions are given by  $U(t_n) = 1, n$  (or  $U(t_n) = (-1)^n, n(-1)^n$ ). Finally, if  $\beta \geq 1$  (or  $\beta < -1$ ), then the two

independent solutions are given by  $U(t_n) = e^{\pm t_n \omega}$  (or  $U(t_n) = -e^{\pm t_n \omega}$ ), where  $\omega$  satisfies  $\cosh(\omega \Delta t) = \beta$  (or  $-\cosh(\omega \Delta t) = \beta$ ). Therefore, the scheme grows at most linearly in time iff  $\beta \in [-1, 1]$ . □

**Theorem A2** Consider the order-2K Lax–Wendroff time integration method given by

$$\underline{U}(t_{i+1}) = -\underline{U}(t_{i-1}) + 2 \sum_{k=0}^K \frac{1}{(2k)!} \Delta t^{2k} (-M^{-1}A)^k \underline{U}(t_i), \quad i = 1, 2, \dots,$$

where  $M$  and  $A$  are symmetric positive definite matrices, and  $\{\underline{U}(t_i)\}_{i \geq 0}$  is a sequence of vectors representing a vector variable  $\underline{u}(t)$  at time slots  $t_i = i \Delta t$ , with  $\Delta t$  the time step size. This scheme is stable, by which we mean that the solution grows at most linearly in time, if  $\Delta t \leq \sqrt{c_K / \sigma_{\max}(M^{-1}A)}$ , where  $\sigma_{\max}(M^{-1}A)$  denotes the spectral radius of  $M^{-1}A$  and  $c_K$  is defined as

$$c_K := \inf \left\{ x \geq 0 \mid \left| \sum_{k=0}^K \frac{1}{(2k)!} (-x)^k \right| > 1 \right\}.$$

*Proof* We can rewrite the time integration scheme as

$$\underline{U}(t_{i+1}) = -\underline{U}(t_{i-1}) + 2B\underline{U}(t_i),$$

where  $B := \sum_{k=0}^K \frac{1}{(2k)!} \Delta t^{2k} (-M^{-1}A)^k$ . Since  $M$  and  $A$  are symmetric positive definite, we can diagonalise  $M^{-1}A$  as  $V D V^{-1}$ , with  $D$  a diagonal matrix with only positive real values on the diagonal. We can then diagonalise  $B$  as  $B = V \left( \sum_{k=0}^K \frac{1}{(2k)!} \Delta t^{2k} (-D)^k \right) V^{-1}$ . Using this diagonalisation we can decouple the matrix-vector equations into scalar equations of the form

$$U(t_{i+1}) = -U(t_{i-1}) + 2\beta U(t_i), \quad i = 1, 2, \dots,$$

with

$$\beta = \sum_{k=0}^K \frac{1}{(2k)!} (-s \Delta t^2)^k \quad \text{for some eigenvalue } s \text{ of } M^{-1}A.$$

From the definition of  $c_K$ , it follows that  $|\beta| \leq 1$  for all possible  $\beta$ , if  $\Delta t^2 \sigma_{\max}(M^{-1}A) \leq c_K$ , so if  $\Delta t \leq \sqrt{c_K / \sigma_{\max}(M^{-1}A)}$ . From Theorem A1 it then follows that this scheme is stable. □

*Remark A3* The values of  $c_K$  can be computed numerically. For example,  $c_K = 4, 12, 7.57$  for  $K = 1, 2, 3$ , respectively.

### References

1. Ainsworth, M.: Dispersive and dissipative behaviour of high order discontinuous Galerkin finite element methods. *J. Comput. Phys.* **198**(1), 106–130 (2004)
2. Ainsworth, M., Monk, P., Muniz, W.: Dispersive and dissipative properties of discontinuous Galerkin finite element methods for the second-order wave equation. *J. Sci. Comput.* **27**(1–3), 5–40 (2006)
3. Antonietti, P.F., Marcati, C., Mazzieri, I., Quarteroni, A.: High order discontinuous Galerkin methods on simplicial elements for the elastodynamics equation. *Numer. Algorithms* **71**(1), 181–206 (2016)
4. Boffi, D.: Finite element approximation of eigenvalue problems. *Acta Numer.* **19**, 1–120 (2010)

5. Chin-Joe-Kong, M.J.S., Mulder, W.A., Van Veldhuizen, M.: Higher-order triangular and tetrahedral finite elements with mass lumping for solving the wave equation. *J. Eng. Math.* **35**(4), 405–426 (1999)
6. Cockburn, B., Shu, C.W.: TVB Runge–Kutta local projection discontinuous Galerkin finite element method for conservation laws. II. General framework. *Math. Comput.* **52**(186), 411–435 (1989)
7. Cohen, G., Joly, P., Tordjman, N.: Higher order triangular finite elements with mass lumping for the wave equation. In: *Proceedings of the Third International Conference on Mathematical and Numerical Aspects of Wave Propagation*, pp. 270–279. SIAM Philadelphia (1995)
8. Cohen, G.C.: *Higher-Order Numerical Methods for Transient Wave Equations*. Springer, New York (2002)
9. Cui, T., Leng, W., Lin, D., Ma, S., Zhang, L.: High order mass-lumping finite elements on simplexes. *Numer. Math. Theory Methods Appl.* **10**(2), 331–350 (2017)
10. Dablain, M.: The application of high-order differencing to the scalar wave equation. *Geophysics* **51**(1), 54–66 (1986)
11. De Basabe, J.D., Sen, M.K.: Grid dispersion and stability criteria of some common finite-element methods for acoustic and elastic wave equations. *Geophysics* **72**(6), T81–T95 (2007)
12. De Basabe, J.D., Sen, M.K.: Stability of the high-order finite elements for acoustic or elastic wave propagation with high-order time stepping. *Geophys. J. Int.* **181**(1), 577–590 (2010)
13. De Basabe, J.D., Sen, M.K., Wheeler, M.F.: The interior penalty discontinuous Galerkin method for elastic wave propagation: grid dispersion. *Geophys. J. Int.* **175**(1), 83–93 (2008)
14. Epshteyn, Y., Rivière, B.: Estimation of penalty parameters for symmetric interior penalty Galerkin methods. *J. Comput. Appl. Math.* **206**(2), 843–872 (2007)
15. Fried, I., Malkus, D.S.: Finite element mass matrix lumping by numerical integration with no convergence rate loss. *Int. J. Solids Struct.* **11**(4), 461–466 (1975)
16. Geevers, S., van der Vegt, J.J.W.: Sharp penalty term and time step bounds for the interior penalty discontinuous Galerkin method for linear hyperbolic problems. *SIAM J. Sci. Comput.* **39**(5), A1851–A1878 (2017)
17. Grote, M.J., Schneebeli, A., Schötzau, D.: Discontinuous Galerkin finite element method for the wave equation. *SIAM J. Numer. Anal.* **44**(6), 2408–2431 (2006)
18. Hu, F.Q., Hussaini, M., Rasetarinera, P.: An analysis of the discontinuous Galerkin method for wave propagation problems. *J. Comput. Phys.* **151**(2), 921–946 (1999)
19. Käser, M., Hermann, V., de la Puente, J.: Quantitative accuracy analysis of the discontinuous Galerkin method for seismic wave propagation. *Geophys. J. Int.* **173**(3), 990–999 (2008)
20. Komatitsch, D., Vilotte, J.P.: The spectral element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures. *Bull. Seismol. Soc. Am.* **88**(2), 368–392 (1998)
21. Lax, P.D., Wendroff, B.: Difference schemes for hyperbolic equations with high order of accuracy. *Commun. Pure Appl. Math.* **17**(3), 381–398 (1964)
22. Lisitsa, V.: Dispersion analysis of discontinuous Galerkin method on triangular mesh for elastic wave equation. *Appl. Math. Model.* **40**(7–8), 5077–5095 (2016)
23. Liu, T., Sen, M.K., Hu, T., De Basabe, J.D., Li, L.: Dispersion analysis of the spectral element method using a triangular mesh. *Wave Motion* **49**(4), 474–483 (2012)
24. Liu, Y., Teng, J., Xu, T., Badal, J.: Higher-order triangular spectral element method with optimized cubature points for seismic wavefield modeling. *J. Comput. Phys.* **336**, 458–480 (2017)
25. Mulder, W.A.: A comparison between higher-order finite elements and finite differences for solving the wave equation. In: *Proceedings of the Second ECCOMAS Conference on Numerical Methods in Engineering*, pp. 344–350. Wiley, New York (1996)
26. Mulder, W.A.: Spurious modes in finite-element discretizations of the wave equation may not be all that bad. *Appl. Numer. Math.* **30**(4), 425–445 (1999)
27. Mulder, W.A.: New triangular mass-lumped finite elements of degree six for wave propagation. *Prog. Electromagn. Res.* **141**, 671–692 (2013)
28. Mulder, W.A., Zhebel, E., Minisini, S.: Time-stepping stability of continuous and discontinuous finite-element methods for 3-D wave propagation. *Geophys. J. Int.* **196**(2), 1123–1133 (2014)
29. Owen, S.J.: A survey of unstructured mesh generation technology. In: *IMR*, pp. 239–267 (1998)
30. Patera, A.T.: A spectral element method for fluid dynamics: laminar flow in a channel expansion. *J. Comput. Phys.* **54**(3), 468–488 (1984)
31. Reed, W.H., Hill, T.: *Triangular mesh methods for the neutron transport equation*. Tech. rep., Los Alamos Scientific Lab., N. Mex. (USA) (1973)
32. Riviere, B., Wheeler, M.F.: Discontinuous finite element methods for acoustic and elastic wave problems. *Contemp. Math.* **329**, 271–282 (2003)
33. Seriani, G., Priolo, E.: Spectral element method for acoustic wave simulation in heterogeneous media. *Finite Elem. Anal. Des.* **16**(3–4), 337–348 (1994)



34. Shahbazi, K.: An explicit expression for the penalty parameter of the interior penalty method. *J. Comput. Phys.* **205**(2), 401–407 (2005)
35. Si, H.: Tetgen, a delaunay-based quality tetrahedral mesh generator. *ACM Trans. Math. Softw.* **41**(2), 11 (2015)
36. Warburton, T., Hesthaven, J.S.: On the constants in hp-finite element trace inverse inequalities. *Comput. Methods Appl. Mech. Eng.* **192**(25), 2765–2773 (2003)