Fourier analyses of continuous and discontinuous Galerkin methods of arbitrary degree of approximation

Daniel Le Roux, Christopher Eldred, Mark Taylor

Fourier analyses of continuous and discontinuous Galerkin methods of arbitrary degree of approximation

Daniel Y. Le Roux[a,*], Christopher Eldred[b], Mark Taylor[c]

[a] *Université de Lyon, CNRS, Université Lyon 1, Institut Camille Jordan, 43, blvd du 11 novembre 1918, 69622 Villeurbanne Cedex, France*
[b] *Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France*
[c] *Sandia National Laboratories, P.O. Box 5800, Albuquerque, NM 87185, USA*

Abstract

We present a Fourier analysis of the first order wave equation in a periodic domain subject to a class of high-order continuous and discontinuous discretizations with either centered or upwind flux. This allows us to analytically derive the dispersion relation, group velocity and identify the emergence of gaps in the dispersion relation at specific wavenumbers. Wave packets with energy at these wavenumbers will fail to propagate correctly, and there will be significant numerical dispersion and other undesirable artifacts. Through our analysis we provide analytic formulas for the dispersion relation when approximation spaces of polynomial functions of degree n are considered. The formulas have been checked for polynomial degrees up to degree 20 for the continuous Galerkin method and up to degree 10 for the discontinuous case. We conjecture that our results hold for arbitrary polynomial degree. Such a Fourier analysis provides an alternative proof of existing results (the eventual presence of a stationary erratic mode, order of convergence, etc.). Finally, for the first time to our knowledge, the existence of gaps is characterized analytically and their specific locations are computed for both the continuous and centered discontinuous Galerkin methods. Conversely, the upwind discontinuous Galerkin method is shown to have neither spectral gaps or an erratic stationary mode.

Keywords: finite-element method, discontinuous Galerkin method, numerical dispersion, dispersion analysis, erratic numerical modes, computational modes

*Corresponding Author: Daniel Y. Le Roux (dleroux@math.univ-lyon1.fr)

1. Introduction

We analyze the linear first order wave equation in a periodic domain subject to a class of high-order collocated finite element discretizations: continuous and discontinuous Galerkin with exact integration with either centered or upwind flux.

In the continuum, the solutions are well known, given by Fourier modes with the dispersion relation given by constant phase velocity. There is a long history of evaluating numerical discretizations by their ability to capture this behavior. For finite difference approximations, see [21, chapter 5.3]. For high-order finite element methods, the initial work was done numerically, where the spatial discretization and/or time discretization is used to convert the equations into a linear system for which one can numerically compute eigenvalues and eigenmodes. The eigenmodes form a basis for the discrete solution space. Numerical solutions for these eigenmodes and eigenvalues have been obtained for several variants of high-order finite element methods [10, 11, 12, 17, 18, 20]. Based on these numerical results, many aspects of the discrete solutions could be well understood, such as the discrete dispersion relation, group velocity and convergence rates of these properties.

These numerical results were first confirmed analytically in [1, 4, 13], based on a Bloch wave approach also used for analysis of the second-order wave equation [2, 3]. In the Bloch wave approach, the solutions are shown to naturally split into two categories, a family of eigenmodes which closely match the low frequency solutions (up to $4h$) and a family of eigenmodes associated with high frequencies ranging from $2h$ up to $4h$ which behave erratically, where h is the meshlength parameter.

The Bloch wave approach is a powerful analytic tool, but there are several reasons to extend these results to a Fourier mode interpretation. These reasons include the ability to derive the dispersion relation (relation between phase speed and wavenumber), group velocity, and to identify the potentially damaging emergence of gaps in the dispersion relation [6, 7, 17, 22]. Convergence properties of the discontinuous Galerkin method has been investigated via Fourier analysis [9, 24]. However, such an approach is difficult to perform for higher polynomial degree since it relies explicitly on the structures of algorithm matrices.

In this work, contrary to the previous Fourier approaches, we provide explicit analytical formulas for the Fourier expansion of each eigenmode as well as the dispersion relation when approximation spaces of polynomial functions of degree n are considered. The formulas have been checked for polynomial degrees up to degree 20 for the continuous Galerkin method in section 3, and up to degree 10 for the discontinuous Galerkin method in section 4. It is conjectured that the results hold for any n. For both methods, the analytic results are derived with the aid of a computer algebra system (Maple). Because the Fourier analysis can give rise to mathematical artifacts, these are removed through a careful branch selection procedure driven by analysis of eigenvectors and associated reconstructed solutions. Such a "cleaning" procedure was missing in most of previous studies.

Let N and m represent the degrees of freedom in our discretization and the number of intervals or elements of the periodic domain, respectively. The discrete equations will contain N eigenmodes (corresponding to N Bloch waves) which can be identified with N Fourier modes to define the dispersion relation. In the continuous Galerkin case, $N = mn$ and each of the N eigenmodes can be expressed as the sum of n Fourier modes (shown in [17]). In the discontinuous Galerkin case, $N = m(n + 1)$ and each eigenmode can be expressed as the sum of $n + 1$ Fourier modes. This fact has caused some confusion in the literature, with several works associating this single solution with n different solutions, one physical and $n - 1$ unphysical spurious or parasitic modes [11, 10]. This interpretation results in a multiple valued dispersion relation $\omega(k)$, where ω is the frequency and k is the wavenumber, with n different branches for each k. We find this interpretation misleading and prefer the approach taken in [17], where each eigenmode is associated with a single k rather than treated as n different solutions with n different values of k. This interpretation is also more consistent with the Bloch wave approach, where the phase velocity is a single valued function of each Bloch mode.

For the dispersion relation, each eigenmode must be identified with a wavenumber k. The purely numerical approach used in [17] expresses the eigenmode in n Fourier modes and takes the mode with the largest amplitude. This sensible approach leads to a dispersion relation for large k that is multiple valued (for large n) and the domain for k contains large gaps near $4h$ [17, Figure 11], making it challenging to compute the group velocity at high frequencies. Here we propose a modified k identification procedure that results in a single valued dispersion relation over the entire domain of

k. Our analytical derivation of the dispersion relation allows us to also analytically compute the discrete group velocity. We show that even in the high wavenumber part of the dispersion relationship, where the solutions are quite erratic and have minimal correspondence to the analytical solutions, the group velocity of correctly predicts the wave packet behavior, as we demonstrate with numerical examples.

With this k identification procedure, there is a one-to-one correspondence between each discrete eigenmode and an associated propagating wave solution to the continuum equations. Thus the class of methods analysed here do not contain any spurious modes, where spurious modes are defined as eigenmodes representing spurious numerical solutions with no corresponding solution in the continuum equations. Despite the lack of spurious modes using this definition, it is well known that many of the higher frequency eigenmodes have spuriously large phase velocity errors and we refer to these modes as erratic. In particular, as with many collocated discretizations, the highest frequency eigenmode is in the null space of the derivative operator and is thus stationary.

We also retrieve the order of convergence originally proved in [4], and show that for both the continuous and centered discontinuous Galerkin methods the discrete schemes admit a unique erratic stationary mode. Finally, in line with previous results, these methods were found to have spectral gaps. For the first time to our knowledge, the existence of gaps has been characterized analytically and their specific locations computed. Conversely, the upwind discontinuous Galerkin method was shown to have neither spectral gaps or an erratic stationary mode.

2. Model problem

For an enclosed domain of length L, where $0 \leqslant x \leqslant L$, we introduce the one-dimensional, inviscid, linearized system of equations

$$\frac{\partial w_1}{\partial t}(x,t) + \rho_1 \frac{\partial w_2}{\partial x}(x,t) = 0, \tag{1}$$

$$\frac{\partial w_2}{\partial t}(x,t) + \rho_2 \frac{\partial w_1}{\partial x}(x,t) = 0. \tag{2}$$

Equations (1) and (2) may lead to a simplified form of

(*i*) The shallow-water equations [14], where w_1 and w_2 play the role of the surface-elevation and velocity, respectively, ρ_1 is the mean depth and ρ_2 is the gravity.

(*ii*) The electromagnetic Maxwell system.

(*iii*) The acoustic sound wave equations.

Periodic boundary conditions and initial data complete the specification of (1) and (2). In the following ρ_1 and ρ_2 are considered as constant parameters.

Letting $\mathbf{w} = (w_1, w_2)$, equations (1) and (2) can be conveniently expressed as

$$\frac{\partial \mathbf{w}}{\partial t} + \mathcal{A}\frac{\partial \mathbf{w}}{\partial x} = 0, \qquad \text{where } \mathcal{A} = \begin{pmatrix} 0 & \rho_1 \\ \rho_2 & 0 \end{pmatrix}. \tag{3}$$

The eigenvalues and eigenvectors of $\mathcal{A}$ are $\pm\sqrt{\rho_1\rho_2}$ and $(\rho_1, \pm\sqrt{\rho_1\rho_2})$, respectively. Since the latter are linearly independent, the system (3) is diagonalizable and we obtain $\mathcal{A} = \mathcal{Q}\,\Lambda\,\mathcal{Q}^{-1}$, with

$$\mathcal{Q} = \begin{pmatrix} \rho_1 & \rho_1 \\ -\sqrt{\rho_1\rho_2} & \sqrt{\rho_1\rho_2} \end{pmatrix} \text{ and } \Lambda = \begin{pmatrix} -\sqrt{\rho_1\rho_2} & 0 \\ 0 & \sqrt{\rho_1\rho_2} \end{pmatrix}. \tag{4}$$

We now define a new set of dependent variables $\mathbf{u} = (u_1, u_2)^T$, the characteristic variables, via the transformation $\mathbf{u} = \mathcal{Q}^{-1}\,\mathbf{w}$. The original system (3) then becomes a simple set of decoupled equations

$$\frac{\partial \mathbf{u}}{\partial t} + \Lambda\frac{\partial \mathbf{u}}{\partial x} = 0. \tag{5}$$

In this paper we let $\sqrt{\rho_1\rho_2} = c$, and only consider the equation for u_2 in (5). The solution u_1 can be deduced by symmetry arguments. Further, assuming time is continuous, we seek periodic solutions of the form $u_2(x,t) = u(x)\,e^{-i\omega t}$, where ω is the angular frequency, and we obtain

$$i\omega u - c\frac{\partial u}{\partial x} = 0. \tag{6}$$

If we examine the free mode of (6) by perturbing about the basic state $u = 0$, and substituting a periodic solution of the form $u(x) = \widehat{u}\,e^{ikx}$ into (6), where

$\widehat{u}$ is the amplitude and k is the wavenumber in the x- direction, we obtain the relations

$$\omega = c\,k := \omega^{AN} \quad \text{and} \quad \frac{\partial \omega}{\partial k} = c := \frac{\partial \omega^{AN}}{\partial k}, \tag{7}$$

for the phase speed and group velocity, respectively. Hence, the wave propagates independently of the wavenumber k and there is no dispersion.

In section 3 and section 4, equation (6) is spatially discretized using the continuous (CG) and discontinuous (DG) Galerkin methods, respectively.

3. The continuous Galerkin discretization

3.1. Continuous Galerkin formulations

We first introduce the weak formulation and then describe the CG spaces that are employed in this section. Finally, the CG discretization is presented.

The space $H^1(\Omega)$ will denote the usual Sobolev space of functions in the square-integrable space $L^2(\Omega)$, whose first derivatives belong to $L^2(\Omega)$. Let u be in a subspace V of $H^1(\Omega)$. Multiplying (6) by a function v belonging to V, and integrating over the domain Ω we obtain

$$i\omega \int_\Omega uv\,dx - c\int_\Omega \frac{\partial u}{\partial x}\,v\,dx = 0, \quad \forall v \in V. \tag{8}$$

Let ε_h denote a partition of the model domain $\Omega = (0, L)$, where h denotes the meshlength parameter, namely ε_h is a finite collection of m elements e_j, $j = 1, 2, ..., m$, of the real line, such that $\bar{\Omega} = \bigcup_{e_j \in \varepsilon_h} \bar{e}_j$. Consider a uniform mesh of m intervals on $(0, L)$ with elements $e_j = (x_j, x_{j+1})$ for $j = 1, 2, ..., m$, and nodes x_j for $j = 1, 2, ..., m + 1$, as shown in Figure 1. For the sake of performing the subsequent Fourier analyses in section 3.2, periodicity of the solution is imposed at end nodes $j = 1$ and $j = m + 1$, and we have $x_1 = x_{m+1}$.

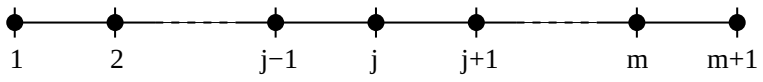

Figure 1: The position of nodes $j, j + 1, ...,\ j = 1, 2, 3, ..., m + 1$.

For $n \in \mathbb{N}$, we consider finite-element spaces V_h^n of polynomial functions, continuous at the element interfaces, such that $V_h^n \subset V$, with

$$V_h^n = \{\, u_h \in H^1(\Omega);\ u_h|_e = \widetilde{u}_h \circ F_e^{-1},\ \widetilde{u}_h \in \mathcal{P}_n(\widetilde{e}),\ \forall\, e \in \varepsilon_h \}\,,$$

where F_e is the affine mapping from the master element $\widetilde{e}$ to the element e in the partition ε_h, and $\mathcal{P}_n(\widetilde{e})$ is the space of polynomial functions of degree at most n on $\widetilde{e}$. In order to match the requirements of the Fourier analysis, the mesh is uniform and Lagrange test functions are employed. The basis $\mathcal{V}_h^n$ of V_h^n is chosen such that on e_j, $\mathcal{V}_h^n$ is a basis of the $n+1$ Lagrange interpolating functions of degree n with $\mathcal{V}_h^n = \{v_s\}$, for $s \in \mathcal{I}_j^n$, $j = 1, 2, ..., m$, where $\mathcal{I}_j^n = \left\{j, j+\frac{1}{n}, j+\frac{2}{n}, j+\frac{3}{n}, ..., j+\frac{n-1}{n}, j+1\right\}$ is the set of indices of the local degrees of freedom on each element e_j of ε_h, $j = 1, 2, ..., m$, as shown in Figure 2 for elements e_{j-1} and e_j. Further, since the mesh is uniform we let $x_{j+\frac{q+1}{n}} - x_{j+\frac{q}{n}} = h$, for $j = 1, 2, ..., m$ and $q = 0, 1, 2, ..., n-1$.

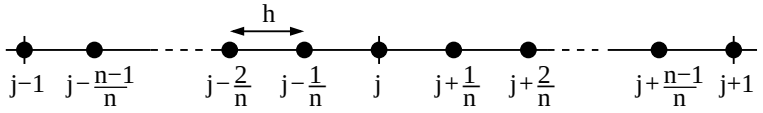

Figure 2: Indices of the local degrees of freedom on element e_j of ε_h, $j = 1, 2, ..., m$.

Note that the Lagrange basis function $v_{j+\frac{q}{n}}$ of degree n over e_j reads

$$v_{j+\frac{q}{n}}|_{e_j} = \frac{1}{(-1)^{n-q}\, q!\, (n-q)!\, h^n} \prod_{\substack{r=0 \\ r \neq q}}^{n} (x - x_{j+\frac{r}{n}}), \quad \text{for } q = 0, 1, 2, ..., n. \tag{9}$$

Introducing the finite-element basis leads to a finite-element discretization of (8), and we search for u_h belonging to V_h^n for the selected basis, at node s, such that

$$i\omega \sum_{j=1}^{m} \int_{e_j} u_h v_s \, dx - c \sum_{j=1}^{m} \int_{e_j} \frac{\partial u_h}{\partial x}\, v_s \, dx = 0, \quad \forall v_s \in V_h^n. \tag{10}$$

In order to perform the analyses of section 3.2, we need to perform the discretization of (10) at nodes $s = j \pm \frac{q}{n}$, for $q = 0, 1, 2, ..., n-1$ and $j = 1, 2, ..., m$.

First, we compute (10) at nodes $s = j+\frac{q}{n}$, for $q = 1, 2, ..., n-1$. Expanding u_h over e_j in the basis $\mathcal{V}_h^n$, namely $u_h|_{e_j} = \sum_{r=0}^{n} u_{j+\frac{r}{n}} v_{j+\frac{r}{n}}$, and taking into account that the compact support of the basis functions v_s is zero outside e_j for $q = 1, 2, ..., n-1$, lead to the following $n-1$ equations over e_j, $j = 1, 2, ..., m$,

$$i\omega \sum_{r=0}^{n} u_{j+\frac{r}{n}} \int_{e_j} v_{j+\frac{r}{n}}\, v_{j+\frac{q}{n}}\, dx - c \sum_{r=0}^{n} u_{j+\frac{r}{n}} \int_{e_j} \frac{\partial v_{j+\frac{r}{n}}}{\partial x}\, v_{j+\frac{q}{n}}\, dx = 0. \tag{11}$$

Similarly, equation (10) is evaluated at nodes $s = j-\frac{q}{n}$, for $q = 1, 2, ..., n-1$, with $u_h|_{e_{j-1}} = \sum_{r=0}^{n} u_{j-\frac{r}{n}} v_{j-\frac{r}{n}}$ over e_{j-1} (see Figure 2), and we obtain a second set of $n-1$ equations over e_{j-1}, $j = 1, 2, ..., m$,

$$i\omega \sum_{r=0}^{n} u_{j-\frac{r}{n}} \int_{e_{j-1}} v_{j-\frac{r}{n}}\, v_{j-\frac{q}{n}}\, dx - c \sum_{r=0}^{n} u_{j-\frac{r}{n}} \int_{e_{j-1}} \frac{\partial v_{j-\frac{r}{n}}}{\partial x}\, v_{j-\frac{q}{n}}\, dx = 0. \tag{12}$$

Equation (10) is finally computed at node $s = j$, for $j = 1, 2, ..., m$. The compact support of the basis function v_j is only non-zero over $e_{j-1} \cup e_j$. Expanding u_h over e_{j-1} and e_j in $\mathcal{V}_h^n$, yields

$$\begin{aligned} i\omega \sum_{r=0}^{n} u_{j-\frac{r}{n}} \int_{e_{j-1}} v_{j-\frac{r}{n}}\, v_j\, dx + i\omega \sum_{r=0}^{n} u_{j+\frac{r}{n}} \int_{e_j} v_{j+\frac{r}{n}}\, v_j\, dx \\ - c \sum_{r=0}^{n} u_{j-\frac{r}{n}} \int_{e_{j-1}} \frac{\partial v_{j-\frac{r}{n}}}{\partial x}\, v_j\, dx - c \sum_{r=0}^{n} u_{j+\frac{r}{n}} \int_{e_j} \frac{\partial v_{j+\frac{r}{n}}}{\partial x}\, v_j\, dx = 0. \end{aligned} \tag{13}$$

Due to periodicity at end nodes, with $u_0 = u_m$ and $u_{m+1} = u_1$, and for $j = 1, 2, ..., m$, equation (13) leads to m equations and (11) or (12) provide $m(n-1)$ additional equations at internal element nodes, namely a total of mn discrete equations.

Example 1. *For $n = 1$, equation (13) leads to a system of m equations*

$$i\omega \frac{h}{6} (u_{j-1} + 4u_j + u_{j+1}) - \frac{c}{2} (u_{j+1} - u_{j-1}) = 0, \quad j = 1, 2, ..., m. \tag{14}$$

In the case $n = 2$, when quadratic basis functions v are used, equation (13) at nodes $j = 1, 2, ..., m$, for $v = v_j$, yields a system of m equations

$$i\omega \frac{h}{15} (-u_{j-1} + 2u_{j-\frac{1}{2}} + 8u_j + 2u_{j+\frac{1}{2}} - u_{j+1}) \tag{15}$$

$$-\frac{c}{6} (u_{j-1} - 4u_{j-\frac{1}{2}} + 4u_{j+\frac{1}{2}} - u_{j+1}) = 0, \quad j = 1, 2, ..., m. \tag{16}$$

and additional m equations are obtained from (11) at mid-nodes $j+\frac{1}{2}$ for $v = v_{j+\frac{1}{2}}$

$$i\omega\frac{2h}{15}\left(u_j + 8u_{j+\frac{1}{2}} + u_{j+1}\right) - \frac{2c}{3}\left(u_{j+1} - u_j\right) = 0, \quad j = 1, 2, ..., m. \tag{17}$$

3.2. The Fourier analysis

A stability/dispersion analysis is now constructed for the CG schemes based on the different choices for n. Periodic solutions of (11) and (13) are sought. Due to symmetry reasons, only selected discrete equations are considered, namely, $n-1$ equations at interior nodes $j+\frac{q}{n}$, for $q = 1, 2, ..., n-1$, obtained from (11) and one equation at a typical boundary element node j obtained from (13). It is crucial to note that the results of the present section are independent of the choice of the typical node j, $j = 1, 2, ..., m$.

The periodic solutions read

(*i*) $u_{j-\frac{r}{n}} = \widehat{u}_{n-r}\, e^{ikx_{j-\frac{r}{n}}}$ at the $n-1$ interior nodes $j-\frac{r}{n}$ of e_{j-1}, $r = 1, 2, ..., n-1$,

(*ii*) $u_{j+\frac{r}{n}} = \widehat{u}_r\, e^{ikx_{j+\frac{r}{n}}}$ at the $n-1$ interior nodes $j+\frac{r}{n}$ of e_j, $r = 1, 2, ..., n-1$,

(*iii*) $u_j = \widehat{u}_n\, e^{ikx_j}$ and $u_{j\pm1} = \widehat{u}_n\, e^{ikx_{j\pm1}}$ at the boundary nodes of e_{j-1} and e_j.

For symmetry reasons, the same Fourier amplitudes $\widehat{u}_r$ are considered at nodes $j-\frac{n-r}{n}$ of e_{j-1} and at nodes $j+\frac{r}{n}$ of e_j, $r = 1, 2, ..., n-1$. Similarly, the amplitudes at end nodes $j-1$, j and $j+1$ of e_{j-1} and e_j are identical and they are denoted by $\widehat{u}_n$. A total of n amplitudes are hence involved as shown in Figure 3.

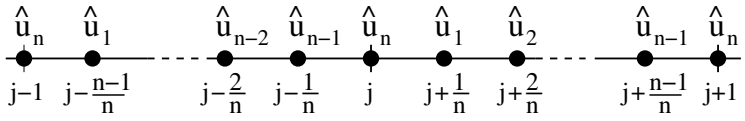

Figure 3: Indices of the Fourier amplitudes on element e_j of ε_h, $j = 1, 2, ..., m$.

For a typical node j, $j = 1, 2, ..., m$, we let

$$M_{q,r} = \frac{1}{h}\int_{e_j} v_{j+\frac{q}{n}}\, v_{j+\frac{r}{n}}\, dx, \quad G_{q,r} = \int_{e_j} v_{j+\frac{q}{n}}\, \frac{\partial v_{j+\frac{r}{n}}}{\partial x}\, dx, \quad \widetilde{\omega} = \frac{\omega h}{c}, \tag{18}$$

for $0 \leqslant q, r \leqslant n$, where $M_{q,r}$ and $G_{q,r}$ are the elements of the elementary mass and gradient matrices of size $(n+1) \times (n+1)$, denoted by $\mathcal{M}$ and $\mathcal{G}$, respectively, with

$$\mathcal{M} = (M_{q,r}), \qquad \mathcal{G} = (G_{q,r}), \quad \text{for } 0 \leqslant q, r \leqslant n, \tag{19}$$

and $\widetilde{\omega}$ is the normalized frequency.

Due to the nature of the Lagrange basis function and particularly the fact that $\sum_{q=0}^{n} v_{j+\frac{q}{n}} = 1$, we have the following properties

$$(i)\ \sum_{r=0}^{n} G_{q,r} = 0, \quad \text{for } 0 \leqslant q \leqslant n, \tag{20}$$

$$(ii)\ \sum_{q=0}^{n} G_{q,r} = 0, \quad \text{for } 1 \leqslant r \leqslant n-1, \qquad \sum_{q=0}^{n} G_{q,0} = -1, \qquad \sum_{q=0}^{n} G_{q,n} = 1, \tag{21}$$

$$(ii)\ G_{q,r} = -G_{n-q,n-r}, \quad \text{for } 0 \leqslant q, r \leqslant n, \tag{22}$$

$$(iv)\ G_{q,r} = -G_{r,q}, \quad \text{except for } (q,r) = (0,0) \text{ and } (q,r) = (n,n). \tag{23}$$

Example 2. *For $n = 2$ and 3, we obtain the gradient matrices of size 3×3 and 4×4, respectively*

$$\mathcal{G} = \begin{pmatrix} -\frac{1}{2} & \frac{2}{3} & -\frac{1}{6} \\ -\frac{2}{3} & 0 & \frac{2}{3} \\ \frac{1}{6} & -\frac{2}{3} & \frac{1}{2} \end{pmatrix}, \qquad \mathcal{G} = \begin{pmatrix} -\frac{1}{2} & \frac{57}{80} & -\frac{3}{10} & \frac{7}{80} \\ -\frac{57}{80} & 0 & \frac{81}{80} & -\frac{3}{10} \\ \frac{3}{10} & -\frac{81}{80} & 0 & \frac{57}{80} \\ -\frac{7}{80} & \frac{3}{10} & -\frac{57}{80} & \frac{1}{2} \end{pmatrix}. \tag{24}$$

For $q = 1, 2, ..., n-1$, substituting $u_{j+\frac{r}{n}}$, u_j and u_{j+1} in (11) leads to

$$\sum_{r=1}^{n-1} \widehat{u}_r\, e^{irkh}\, (G_{q,r} - i\widetilde{\omega} M_{q,r}) + \widehat{u}_n \left((G_{q,0} - i\widetilde{\omega} M_{q,0}) + e^{inkh}\, (G_{q,n} - i\widetilde{\omega} M_{q,n}) \right) = 0, \tag{25}$$

and substituting $u_{j\pm\frac{r}{n}}, u_j$ and $u_{j\pm1}$ in (13) by using the properties (57) yields

$$\sum_{r=1}^{n-1} \widehat{u}_r \left(e^{irkh}\left(G_{0,r} - i\widetilde{\omega}M_{0,r}\right) + e^{-i(n-r)kh}\left(G_{n,r} - i\widetilde{\omega}M_{n,r}\right)\right) + \widehat{u}_n \left(-2i\widetilde{\omega}M_{0,0} + e^{inkh}\left(G_{0,n} - i\widetilde{\omega}M_{0,n}\right) + e^{-inkh}\left(G_{n,0} - i\widetilde{\omega}M_{n,0}\right)\right) = 0. \tag{26}$$

Note that the properties arising from (18), namely, $M_{0,n-r} = M_{n-r,0} = M_{n,r}$ and $G_{0,n-r} = -G_{n-r,0} = -G_{n,r}$, for $1 \leqslant r \leqslant n-1$, have been employed to obtain (26).

Equations (25) and (26) then lead to the matrix system

$$(\widehat{\mathcal{G}} - i\widetilde{\omega}\widehat{\mathcal{M}})\,\mathcal{D}\,\widehat{\mathbf{U}} = 0, \tag{27}$$

for the amplitudes $\widehat{\mathbf{U}} = (\widehat{u}_1, \widehat{u}_2, ..., \widehat{u}_n)$, where $\mathcal{D}$ is the $n \times n$ diagonal matrix whose diagonal entries starting in the upper left corner are $(e^{ikh}, e^{2ikh}, ..., e^{i(n-1)kh}, 1)$ and $\widehat{\mathcal{G}}$ and $\widehat{\mathcal{M}}$ are the $n \times n$ matrices

$$\widehat{\mathcal{G}} = \begin{pmatrix} G_{1,1} & \cdots & G_{1,n-1} & G_{1,0} + e^{inkh}\,G_{1,n} \\ \vdots & \ddots & \vdots & \vdots \\ G_{n-1,1} & \cdots & G_{n-1,n-1} & G_{n-1,0} + e^{inkh}\,G_{n-1,n} \\ G_{0,1} + e^{-inkh}\,G_{n,1} & \cdots & G_{0,n-1} + e^{-inkh}\,G_{n,n-1} & 2i\sin(nkh)G_{0,n} \end{pmatrix},$$

$$\widehat{\mathcal{M}} = \begin{pmatrix} M_{1,1} & \cdots & M_{1,n-1} & M_{1,0} + e^{inkh}\,M_{1,n} \\ \vdots & \ddots & \vdots & \vdots \\ M_{n-1,1} & \cdots & M_{n-1,n-1} & M_{n-1,0} + e^{inkh}\,M_{n-1,n} \\ M_{0,1} + e^{-inkh}\,M_{n,1} & \cdots & M_{0,n-1} + e^{-inkh}\,M_{n,n-1} & 2(M_{0,0} + \cos(nkh)M_{0,n}) \end{pmatrix}.$$

The properties $G_{0,n} = -G_{n,0}$ and $M_{0,n} = M_{n,0}$, arising from (18), have been employed to simplify the entry in the n-th row and n-th column of both $\widehat{\mathcal{G}}$ and $\widehat{\mathcal{M}}$. Because $\mathcal{D}$ is a nonsingular matrix, indeed $\det\mathcal{D} = e^{in(n-1)kh/2} \neq 0$, solving the matrix system (27) for the amplitudes is equivalent to solving the generalized eigenvalue problem

$$\widehat{\mathcal{G}}\,\widehat{\mathbf{U}} = i\widetilde{\omega}\widehat{\mathcal{M}}\,\widehat{\mathbf{U}}. \tag{28}$$

For a nontrivial solution $\widehat{\mathbf{U}}$ to exist, the possible values of $\widetilde{\omega}$ need to satisfy the characteristic equation

$$\det(\widehat{\mathcal{G}} - i\widetilde{\omega}\widehat{\mathcal{M}}) = 0, \tag{29}$$

namely, a polynomial equation of degree n in the unknown $\widetilde{\omega}$.

Theorem 1. *The following properties hold*

(i) $\widehat{\mathcal{G}}$ *is skew-Hermitian,*

(ii) $\widehat{\mathcal{M}}$ *is Hermitian,*

(iii) $\widehat{\mathcal{M}}$ *is a positive definite matrix,*

(iv) The roots $\widetilde{\omega}$ *of the characteristic equation* $\det(\widehat{\mathcal{G}} - i\widetilde{\omega}\widehat{\mathcal{M}}) = 0$ *are real.*

Proof. Let $\mathcal{X}^*$ be the conjugate transpose of a matrix $\mathcal{X}$. Properties (i) and (ii), namely $\widehat{\mathcal{G}} = -\widehat{\mathcal{G}}^*$ and $\widehat{\mathcal{M}} = \widehat{\mathcal{M}}^*$, respectively, are a direct consequence of (18).

To prove that $\widehat{\mathcal{M}}$ is a positive definite matrix, let $\mathcal{P}$ be the $(n+1)\times(n+1)$ matrix

$$\mathcal{P} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & \cdots & 0 & e^{inkh} \end{pmatrix}.$$

Since the element mass matrix $\mathcal{M}$ is positive definite [8] and $\mathcal{P}$ is a nonsingular matrix ($\det \mathcal{P} = e^{inkh} \neq 0$), the matrix $\mathcal{P}^*\mathcal{M}\mathcal{P}$ is also positive definite. Indeed, $\forall \mathbf{v} \in \mathbb{C}^{n+1}$ we have $\mathbf{v}^*\mathcal{P}^*\mathcal{M}\mathcal{P}\mathbf{v} = (\mathcal{P}\mathbf{v})^*\mathcal{M}(\mathcal{P}\mathbf{v})$, and because $\mathcal{M}$ is a real matrix we obtain $\mathbf{v}^*\mathcal{P}^*\mathcal{M}\mathcal{P}\mathbf{v} \in \mathbb{R}$. Elementary computations then show that $\widehat{\mathcal{M}}$ is nothing else than a leading principal submatrix of $\mathcal{P}^*\mathcal{M}\mathcal{P}$ obtained from $\mathcal{P}^*\mathcal{M}\mathcal{P}$ by removing the first column and the first row. Consequently, $\widehat{\mathcal{M}}$ is also a positive definite matrix.

Finally, the generalized eigenvalue problem (28) may be rewritten as $i\widehat{\mathcal{G}}\,\widehat{\mathbf{U}} = -\widetilde{\omega}\widehat{\mathcal{M}}\,\widehat{\mathbf{U}}$, where $i\widehat{\mathcal{G}}$ and $\widehat{\mathcal{M}}$ are both Hermitian matrices due to properties (i) and (ii). Since $\widehat{\mathcal{M}}$ is definite positive the roots $\widetilde{\omega}$ of the characteristic equation $\det(i\widehat{\mathcal{G}} - \widetilde{\omega}\widehat{\mathcal{M}}) = 0$ are real and two eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ with distinct eigenvalues are $\widehat{\mathcal{M}}$-orthogonal with $\mathbf{v}_1^*\widehat{\mathcal{M}}\mathbf{v}_2 = 0$. Further, there exists a basis of generalized eigenvectors. □

Theorem 2. *For continuous finite-element approximations of degree n, with $1 < n \leqslant 20$, the characteristic equation or dispersion relation obtained from the generalized eigenvalue problem (28) is the polynomial of degree n*

$$P_n^{CG}(\widetilde{\omega}) = \widetilde{\omega}^n + \sum_{j=1}^{\lfloor \frac{n+1}{2} \rfloor} \widetilde{a}_{n-2j+1}\, \widetilde{\omega}^{n-2j+1} + \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} \widetilde{a}_{n-2j}\, \widetilde{\omega}^{n-2j} = 0, \tag{30}$$

where floor$(x) = \lfloor x \rfloor$ and $\binom{q}{r} = \frac{q!}{(q-r)!r!}$ with

$$a_{n-2j+1} = \frac{(-1)^j}{n^{2j-1}} \binom{n}{2j-1} \frac{(n+2j)!}{(n+1)!}, \quad a_{n-2j} = \frac{(-1)^j}{n^{2j}} \binom{n}{2j} \frac{(n+2j+1)!}{(n+1)!}, \tag{31}$$

$$\frac{\widetilde{a}_{n-2j+1}}{a_{n-2j+1}} = \frac{\sin(nkh)}{\cos(nkh) - (-1)^n(n+1)}, \quad \frac{\widetilde{a}_{n-2j}}{a_{n-2j}} = \frac{\cos(nkh) - (-1)^n\left(\frac{n+1}{2j+1}\right)}{\cos(nkh) - (-1)^n(n+1)}. \tag{32}$$

These results have been obtained using a computer algebra system (Maple). It is conjectured that (30)-(32) holds $\forall n > 20$. Note that the coefficient $\widetilde{a}_j$, $j = 1, 2, 3, \cdots$, has two different meanings depending n is even or odd.

Example 3. *The case $n = 1$ yields the classical characteristic equation result [15, 23]*

$$\widetilde{\omega}\Big(\cos(kh) + 2\Big) - 3\sin(kh) = 0,$$

and $\widetilde{\omega}$ is graphed in Figure 4 for $kh \in [-\pi, \pi]$, with the continuous frequency $\widetilde{\omega}^{AN} := kh$, obtained from (7) and (18).

Such a discrepancy between the analytical and computed solutions are discussed later in this section.

The following corollaries aim to make precise a few properties of the frequency solutions $\widetilde{\omega}$ of (30).

Corollary 1. *Let $\widetilde{\omega}_j$, $j = 1, 2, ..., n$, be the n roots of $P_n^{CG}(\widetilde{\omega})$ in (30). We have the following properties:*

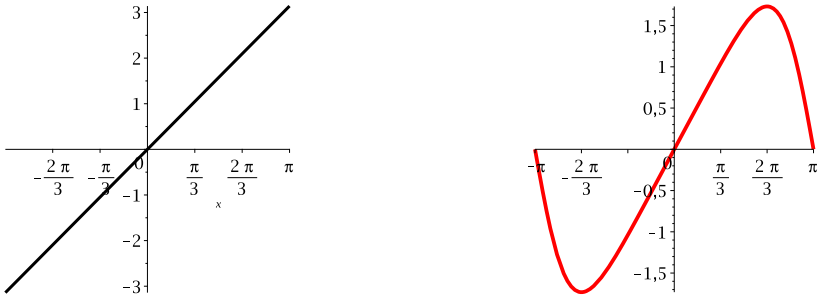

Figure 4: The continuous frequency $\widetilde{\omega}^{AN}$ (left) and the computed frequency $\widetilde{\omega}$ (right) for the CG scheme in the case $n = 1$.

The case $n = 2p+1, \quad p = 1,2,3,...$	*The case* $n = 2p, \quad p = 1,2,3,...$
(i) For $j = 1,2,...,n$: $\widetilde{\omega}_j(kh) = \widetilde{\omega}_j\big(kh + \frac{2\pi}{n}\big)$	*(ii) For* $j = 1,2,...,n$: $\widetilde{\omega}_j(kh) = \widetilde{\omega}_j\big(kh + \frac{2\pi}{n}\big)$
(iii) For $j = 1,2,...,p$, *we have* $2p$ *solutions satisfying:* $\widetilde{\omega}_j(kh) = -\widetilde{\omega}_{2(p+1)-j}(-kh)$ *and* 1 *solution such that:* $\widetilde{\omega}_{p+1}(kh) = -\widetilde{\omega}_{p+1}(-kh)$	*(iv) For* $j = 1,2,...,p$, *we have* $2p$ *solutions satisfying:* $\widetilde{\omega}_j(kh) = -\widetilde{\omega}_{2p+1-j}(-kh)$
(v) At $kh = 0$, *only* $\widetilde{\omega}_{p+1}$ *is zero*	*(vi) At* $kh = 0$, *both* $\widetilde{\omega}_p$ *and* $\widetilde{\omega}_{p+1}$ *vanish*

Proof. Properties (i) and (ii) are obtained by substituting kh by $kh + \frac{2\pi}{n}$ in (32) leading to $P_n^{CG}(\widetilde{\omega}(kh)) = P_n^{CG}(\widetilde{\omega}(kh + \frac{2\pi}{n}))$ in (30). Further, equations (30)-(32) yield $P_n^{CG}(-\widetilde{\omega}(-kh)) = -P_n^{CG}(\widetilde{\omega}(kh))$ in the case n is odd, and $P_n^{CG}(-\widetilde{\omega}(-kh)) = P_n^{CG}(\widetilde{\omega}(kh))$ in the case n is even, leading to properties (iii) and (iv). Finally, we let $kh = 0$ in (30) and (32) and consider the coefficients of the lowest degree terms in $\widetilde{\omega}$ in $P_n^{CG}(\widetilde{\omega})$. The case $n = 2p+1$ yields $\widetilde{a}_0 = 0$ and $\widetilde{a}_1 = \frac{(-1)^p(2n+1)!}{n^{2p}(n+2)!} \neq 0$, while when $n = 2p$ we obtain $\widetilde{a}_0 = \widetilde{a}_1 = 0$ and $\widetilde{a}_2 = \frac{(-1)^{p-1}(2n-1)!}{n^{2p-2}(n+1)!} \neq 0$. □

Corollary 2. *In the limit as mesh spacing $kh \to 0$, the group velocities $\frac{\partial \widetilde{\omega}_j}{\partial kh}$ $j = 1, 2, ..., n$, obtained from the solutions $\widetilde{\omega}_j$ of $P_n^{CG}(\widetilde{\omega})$ in (30), satisfy:*

The case $n = 2p+1, \quad p = 1, 2, 3, ...$	*The case* $n = 2p, \quad p = 1, 2, 3, ...$
For $j = 1, 2, ..., n,$ $\lim_{kh\to 0} \frac{\partial \widetilde{\omega}_j}{\partial kh} = 1,$	*For* $j = 1, 2, ..., p-1, p+2, ..., n,$ $\lim_{kh\to 0} \frac{\partial \widetilde{\omega}_j}{\partial kh} = 1,$ *and for* $j = p$ *and* $j = p+1,$ $\lim_{kh\to 0} \frac{\partial \widetilde{\omega}_p}{\partial kh}$ *and* $\lim_{kh\to 0} \frac{\partial \widetilde{\omega}_{p+1}}{\partial kh}$ *are equal to* $(-1)^{p+1}(n+1) - n$ *or* $(-1)^p(n+1) - n$, *namely* 1 *or* $-(2n+1)$ *depending on whether* p *is even or odd.*

Proof. Since $\widetilde{\omega}(kh)$ is a function of kh we have

$$\frac{dP_n^{CG}}{d\,kh} = \frac{\partial P_n^{CG}}{\partial kh} + \frac{\partial P_n^{CG}}{\partial \widetilde{\omega}} \frac{\partial \widetilde{\omega}}{\partial kh} = 0, \tag{33}$$

and hence, by letting $\widetilde{a}_n = a_n := 1$, and taking account that

$$n\widetilde{\omega}^{n-1} + \sum_{j=1}^{\lfloor \frac{n-1}{2} \rfloor} (n-2j)\widetilde{a}_{n-2j}\, \widetilde{\omega}^{n-2j-1} = \sum_{j=1}^{\lfloor \frac{n+1}{2} \rfloor} (n-2j+2)\widetilde{a}_{n-2j+2}\, \widetilde{\omega}^{n-2j+1},$$

we obtain from (30) and (33)

$$\frac{dP_n^{CG}}{d\,kh} = \sum_{j=1}^{\lfloor \frac{n+1}{2} \rfloor} \frac{\partial \widetilde{a}_{n-2j+1}}{\partial kh}\, \widetilde{\omega}^{n-2j+1} + \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} \frac{\partial \widetilde{a}_{n-2j}}{\partial kh}\, \widetilde{\omega}^{n-2j}$$
$$+ \frac{\partial \widetilde{\omega}}{\partial kh} \left(\sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} (n-2j+1)\widetilde{a}_{n-2j+1}\, \widetilde{\omega}^{n-2j} + \sum_{j=1}^{\lfloor \frac{n+1}{2} \rfloor} (n-2j+2)\widetilde{a}_{n-2j+2}\, \widetilde{\omega}^{n-2j+1} \right) = 0. \tag{34}$$

From (32) we deduce at $kh = 0$

$$\widetilde{a}_{n-2j+1}\big|_{kh=0} = 0, \qquad \widetilde{a}_{n-2j+2}\big|_{kh=0} = \frac{(-1)^n(n+1)-2j+1}{(2j-1)\left((-1)^n(n+1)-1\right)}\, a_{n-2j+2},$$
$$\frac{\partial \widetilde{a}_{n-2j}}{\partial kh}\Big|_{kh=0} = 0, \qquad \frac{\partial \widetilde{a}_{n-2j+1}}{\partial kh}\Big|_{kh=0} = \frac{n}{1-(-1)^n(n+1)}\, a_{n-2j+1},$$

further, equation (31) yields $\frac{a_{n-2j+1}}{a_{n-2j+2}} = -\frac{(n-2j+2)(n+2j)}{n(2j-1)}$, and letting

$$\Psi_1 = \sum_{j=1}^{\lfloor \frac{n+1}{2} \rfloor} \frac{(n-2j+2)(n+2j)}{(2j-1)} a_{n-2j+2} \widetilde{\omega}^{n-2j+1}, \tag{35}$$

$$\Psi_2 = \sum_{j=1}^{\lfloor \frac{n+1}{2} \rfloor} \frac{(n-2j+2)\Big((-1)^{n+1}(n+1)+2j-1\Big)}{(2j-1)} a_{n-2j+2} \widetilde{\omega}^{n-2j+1}, \tag{36}$$

equation (34) then leads to

$$\lim_{kh\to 0} \frac{\partial \widetilde{\omega}}{\partial kh} = \lim_{kh\to 0} \frac{\Psi_1}{\Psi_2}.$$

When n is odd, with $n = 2p+1$, for $p = 1, 2, 3, ...$, we deduce from (35) and (36) that

$$\Psi_1 = \Psi_2 = (2p+1)(2p+3)\,\widetilde{\omega}^{2p} + \frac{(2p-1)(2p+5)}{3} a_{2p-1}\,\widetilde{\omega}^{2p-2} + \cdots + \frac{4p+3}{2p+1} a_1,$$

with $a_1 \neq 0$, namely $\lim_{kh\to 0} \frac{\partial \widetilde{\omega}}{\partial kh} = 1$, and the group velocity is continuous at $kh = 0$.

In the case where n is even, with $n = 2p$, for $p = 1, 2, 3, ...$, equations (35) and (36) yield

$$\Psi_1 - \Psi_2 = 2(n+1) \sum_{j=1}^{p} \frac{(n-2j+2)}{(2j-1)} a_{n-2j+2} \widetilde{\omega}^{n-2j+1}, \tag{37}$$

further, from (30)-(32), we obtain

$$0 = \lim_{kh\to 0} P_n^{CG}(\widetilde{\omega}) = \lim_{kh\to 0} \left(\widetilde{\omega}^n + \frac{1}{n} \sum_{j=1}^{p} \frac{(n-2j)}{(2j+1)} a_{n-2j}\, \widetilde{\omega}^{n-2j} \right). \tag{38}$$

The lowest degree term with respect to $\widetilde{\omega}$ in the right hand side of (38) is $\frac{2}{n(n-1)} a_2 \widetilde{\omega}^2$, with $a_2 \neq 0$, and two cases need to be considered: $\widetilde{\omega} \neq 0$ and $\widetilde{\omega} = 0$.

We first examine the case $\widetilde{\omega} \neq 0$ (with $n = 2p, p = 1, 2, 3, ...$). In such a case dividing (38) by $\widetilde{\omega}$ leads to

$$0 = \lim_{kh\to 0} P_n^{CG}(\widetilde{\omega}) = \lim_{kh\to 0} \frac{1}{n} \sum_{j=1}^{p} \frac{(n-2j+2)}{(2j-1)} a_{n-2j+2}\, \widetilde{\omega}^{n-2j+1}, \tag{39}$$

and hence we deduce from (37) and (39) that $\lim_{h\to 0}(\Psi_1 - \Psi_2) = 0$. Since $a_2 \neq 0$ we have $\lim_{kh\to 0}\Psi_2 \neq 0$ from (36), and we obtain $\lim_{h\to 0}\frac{\partial\widetilde{\omega}_j}{\partial kh} = 1$, for $j = 1, 2, ..., p-1, p+2, ..., 2p$, namely the solutions $\widetilde{\omega}_j$ which do not vanish at $kh = 0$ when n is even, and the group velocity is continuous at $kh = 0$.

We now examine the case $\widetilde{\omega} = 0$ at $kh = 0$, and n is still even. We obtain from (31) and (32)

$$a_0 = \frac{(-1)^p}{n^n}\frac{(2n+1)!}{(n+1)!}\frac{\cos(nkh)-1}{\cos(nkh)-n-1}, \quad a_2 = \frac{(-1)^{p-1}}{2n^{n-3}}\frac{(2n-1)!}{(n+1)!}\frac{(n-1)\cos(nkh)-n-1}{\cos(nkh)-n-1},$$
$$a_1 = \frac{(-1)^p}{n^{n-2}}\frac{(2n)!}{(n+1)!}\frac{\sin(nkh)}{\cos(nkh)-n-1},$$

and hence, $a_0 = 0$ and $a_1 = 0$ at $kh = 0$. Consequently, when $\widetilde{\omega} = 0$ at $kh = 0$, namely for $\widetilde{\omega}_p$ and $\widetilde{\omega}_{p+1}$, equation (33) is not usable and we need to consider

$$\frac{d^2P_n^{CG}}{d\,(kh)^2} = \frac{\partial^2 P_n^{CG}}{\partial\widetilde{\omega}^2}\left(\frac{\partial\widetilde{\omega}}{\partial kh}\right)^2 + 2\,\frac{\partial^2 P_n^{CG}}{\partial\widetilde{\omega}\partial kh}\left(\frac{\partial\widetilde{\omega}}{\partial kh}\right) + \frac{\partial^2 P_n^{CG}}{\partial(kh)^2} + \frac{\partial P_n^{CG}}{\partial\widetilde{\omega}}\frac{\partial^2\widetilde{\omega}}{\partial(kh)^2} \quad = \quad 0.$$

At $kh = 0$ we obtain

$$\frac{\partial P_n^{CG}}{\partial\widetilde{\omega}} = 0, \qquad \frac{\partial^2 P_n^{CG}}{\partial(kh)^2} = \frac{\partial^2 a_0}{\partial(kh)^2},$$
$$\frac{\partial^2 P_n^{CG}}{\partial\widetilde{\omega}\partial kh} = \frac{\partial a_1}{\partial kh}, \qquad \frac{\partial^2 P_n^{CG}}{\partial\widetilde{\omega}^2} = 2a_2,$$

which leads to

$$\lim_{kh\to 0}\frac{\partial\widetilde{\omega}}{\partial kh} = \lim_{kh\to 0}\frac{1}{2a_2}\left(-\frac{\partial a_1}{\partial kh} \pm \sqrt{\left(\frac{\partial a_1}{\partial kh}\right)^2 - 2a_2\frac{\partial^2 a_0}{\partial(kh)^2}}\right). \qquad (40)$$

The computations of (40) end the proof. In the case n is even, the group velocity corresponding to the frequencies $\widetilde{\omega}_p$ and $\widetilde{\omega}_{p+1}$ is thus discontinuous at $kh = 0$. □

For further discussions and analyses, similar evaluations at $kh = \pi$ for $n = 2p+1$ ($p = 0, 1, 2, \dots$) leads to

$$\lim_{kh\to\pi}\frac{\partial\widetilde{\omega}_{p+1}}{\partial kh} = -(2n+1),$$

and for $n = 2p$ ($p = 1, 2, 3, \dots$) as it is the case at $kh = 0$, we obtain

$$\lim_{kh\to\pi} \frac{\partial\widetilde{\omega}_p}{\partial kh} \text{ and } \lim_{kh\to\pi} \frac{\partial\widetilde{\omega}_{p+1}}{\partial kh} \text{ are equal to } 1 \text{ or } -(2n+1).$$

Note that above we are considering only the solutions that are equal to zero at $kh = \pi$. The reason for this will be apparent later on, when they are shown to be the cause of the appearance of the stationary erratic mode.

Corollary 3. *In the limit as mesh spacing $h \to 0$, the roots $\widetilde{\omega}_j$, $j = 1, 2, ..., n$, of $P_n^{CG}(\widetilde{\omega})$ in (30), at least for $n \leqslant 20$, satisfy the following asymptotics obtained by using a computer algebra system (Maple):*

In the case $n = 2p+1$, $p = 0, 1, 2, ...$	*In the case $n = 2p$, $p = 1, 2, 3, ...$*
For $j = 1, 2, ..., p$,	*For $j = 1, 2, ..., p-1$,*
(i) we have $2p$ solutions:	*(ii) we have $2p-2$ solutions:*
$\widetilde{\omega}_{j,2(p+1)-j}(kh) = \pm\sigma_j(n) + kh \mp O(h^2)$,	$\widetilde{\omega}_{j,2p+1-j}(kh) = \pm\varsigma_j(n) + kh \pm O(h^2)$,
(iii) and 1 solution:	*(iv) and 2 solutions:*
$\widetilde{\omega}_{p+1} = kh - \kappa_{p+1}\,(kh)^{2n+3} + O(h^{2n+5})$,	$\widetilde{\omega}_p = kh + \kappa_p\,(kh)^{2n+1} + O(h^{2n+3})$,
	$\widetilde{\omega}_{p+1} = -(2n+1)\,kh + O(h^3)$,

with

$$\kappa_{p+1} = \frac{n^{2n+2}(n+1)(n!)^2}{2(2n+3)\big((2n+1)!\big)^2} \qquad \textit{and} \qquad \kappa_p = \frac{n^{2n}(2n+1)(n!)^2}{2(n+1)\big((2n+1)!\big)^2}. \tag{41}$$

Note that the results of corollary 2 could be retrieved from the asymptotic expansions obtained in corollary 3. Further, the order of convergence, namely $2n+3$ in (*iii*) and $2n+1$ in (*iv*), have been formally proved in [4, equation (14)] for all n by using Bloch waves instead of the present Fourier approach. Since the methodology is different in both studies, the leading terms κ_{p+1} and κ_p are mentioned in corollary 3 for the asymptotic expansions of $\widetilde{\omega}$ in (*iii*) and (*iv*), even if there is a certain similarity with the results presented in [4, equation (14)]. It is worthwhile to mention that is is proved in [19] that for piecewise polynomials of even degree superconvergence occurs in function values, while for piecewise polynomials of odd degree, it occurs in derivatives.

Finally, the solutions corresponding to the frequencies in (i) and (ii) and $\widetilde{\omega}_{p+1}$ in (iv) are artifacts of the Fourier analysis, as we will discuss later in this section.

By rearranging terms in (30) we obtain

$$Q_n(\widetilde{\omega})\,\cos(nkh) + R_{n-1}(\widetilde{\omega})\,\sin(nkh) + S_n(\widetilde{\omega}) = 0, \tag{42}$$

where Q_n, R_{n-1} and S_n are polynomials of degree n or $n-1$ with respect to $\widetilde{\omega}$ with

$$Q_n(\widetilde{\omega}) = \sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} a_{n-2j}\,\widetilde{\omega}^{n-2j}, \quad R_{n-1}(\widetilde{\omega}) = \sum_{j=1}^{\lfloor \frac{n+1}{2} \rfloor} a_{n-2j+1}\,\widetilde{\omega}^{n-2j+1}, \tag{43}$$

$$S_n(\widetilde{\omega}) = (-1)^{n+1}(n+1)\Big(\sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} \frac{a_{n-2j}}{2j+1}\,\widetilde{\omega}^{n-2j}\Big). \tag{44}$$

where we let $a_n := 1$ in the definition of $Q_n(\widetilde{\omega})$ and $S_n(\widetilde{\omega})$.

Example 4. *For $n = 2$ we obtain*

$$Q_2(\widetilde{\omega}) = \widetilde{\omega}^2 - 5, \qquad R_1(\widetilde{\omega}) = -4\,\widetilde{\omega}, \qquad S_2(\widetilde{\omega}) = -3\,\widetilde{\omega}^2 + 5,$$

and the case $n = 3$ leads to

$$Q_3(\widetilde{\omega}) = \widetilde{\omega}^3 - 10\,\widetilde{\omega}, \qquad R_2(\widetilde{\omega}) = -5\,\widetilde{\omega}^2 + \tfrac{70}{9}, \qquad S_3(\widetilde{\omega}) = 4\,\widetilde{\omega}^3 - \tfrac{40}{3}\,\widetilde{\omega}.$$

Corollary 4. *The roots $\widetilde{\omega}_j$, $j = 1, 2, ..., n$, of $P_n^{CG}(\widetilde{\omega})$ in (30) do not exist when*

$$|S_n(\widetilde{\omega})| > \sqrt{Q_n^2(\widetilde{\omega}) + R_{n-1}^2(\widetilde{\omega})}. \tag{45}$$

Proof. After long and tedious algebra we obtain for all n

$$Q_n^2(\widetilde{\omega}) + R_{n-1}^2(\widetilde{\omega}) = \sum_{j=0}^{n} \binom{2j+1}{j} \frac{(n+j+1)!}{n^{2j}(n+1)(n-j)!}\,\widetilde{\omega}^{2(n-j)}, \tag{46}$$

and hence, $Q_n^2(\widetilde{\omega}) + R_{n-1}^2(\widetilde{\omega}) > 0$ since the lowest degree term in (46) for $j = n$, with respect to $\widetilde{\omega}$, is different from zero. Consequently, by letting

$$(\sin\varphi, \cos\varphi) = \frac{1}{\sqrt{Q_n^2(\widetilde{\omega}) + R_{n-1}^2(\widetilde{\omega})}}\,(Q_n(\widetilde{\omega}), R_{n-1}(\widetilde{\omega})), \tag{47}$$

equation (42) is rewritten in the form

$$S_n(\widetilde{\omega}) = -\sqrt{Q_n^2(\widetilde{\omega}) + R_{n-1}^2(\widetilde{\omega})}\,\sin(nkh + \varphi), \tag{48}$$

which ends the proof. □

We deduced that if (45) is satisfied, $\widetilde{\omega}$ is no longer continuous in kh and hence, the frequency exhibits discontinuities or spectral gaps. In such a case, a few intervals exist where $\widetilde{\omega}$ is not defined. Such intervals are given for $n = 3, 4, 5, ..., 10$, in Table 1. These gaps are well-known features of high-order finite element discretizations. However, to our knowledge, this is the first time the existence of such gaps and their exact location has been proven analytically.

Table 1: Intervals of values of $\widetilde{\omega}$, in the case $n = 1, 2, ..., 10$, for which the condition (45) is satisfied, namely when $\widetilde{\omega}$ cannot be computed.

n	$\mathcal{I}_{\widetilde{\omega}}(n)$
1	$\emptyset$ (i.e., $\widetilde{\omega}(kh)$ is continuous in kh)
2	$\emptyset$
3	$[0.8820\,,\,0.9481]$
4	$[1.323\,,\,1.525]$
5	$[0.57463\,,\,0.57574] \cup [1.597\,,\,1.952]$
6	$[0.96357\,,\,0.97461] \cup [1.794\,,\,2.311]$
7	$[0.42053\,,\,0.42054] \cup [1.239\,,\,1.273] \cup [1.953\,,\,2.641]$
8	$[0.74050\,,\,0.74085] \cup [1.442\,,\,1.512] \cup [2.094\,,\,2.959]$
9	$[0.33153114\,,\,0.33153119] \cup [0.9912\,,\,0.9934] \cup [1.599\,,\,1.714] \cup [2.227\,,\,3.273]$
10	$[0.59930435\,,\,0.59931207] \cup [1.1916\,,\,1.1989] \cup [1.725\,,\,1.893] \cup [2.357\,,\,3.586]$

Theorem 3. *The following properties hold*

(i) For the element-wise gradient matrix $\mathcal{G}$ of size $(n+1)\times(n+1)$ defined in (19), we have $\mathrm{rank}(\mathcal{G}) = n$,

(ii) A leading principal submatrix of $\mathcal{G}$ of size $n \times n$, obtained by removing the first column and the last row of $\mathcal{G}$, namely

$$\mathcal{G}^{\diamond} = (G_{q,r}), \quad \textit{for } 0 \leqslant q \leqslant n-1 \textit{ and } 1 \leqslant r \leqslant n, \tag{49}$$

is nonsingular.

(iii) For all $n \in \mathbb{N}$, the discrete scheme admits a unique erratic stationary mode.

Proof. We first prove property (i). A vector $(\zeta_1, \zeta_2, \zeta_3, ..., \zeta_{n+1})^T$ is in the kernel (also known as null space) of $\mathcal{G}$, denoted by $\ker(\mathcal{G})$, if and only if, for all r such that $0 \leqslant r \leqslant n$, we have on element e_j

$$\sum_{q=0}^{n} \int_{e_j} \zeta_{q+1}\, v_{j+\frac{q}{n}}\, \frac{\partial v_{j+\frac{r}{n}}}{\partial x}\, dx = 0,$$

and thus if and only if, the polynomial $\sum_{q=0}^{n} \zeta_{q+1}\, v_{j+\frac{q}{n}}$ belongs to the orthogonal of the space Υ, denoted by $\Upsilon^{\perp}$, generated by the vectors $\left(\frac{\partial v_{j+\frac{r}{n}}}{\partial x}\right)\Big|_{0 \leqslant r \leqslant n}$, for the scalar product on $\mathbb{R}_n[x]$ defined as $< v_k, v_l >= \int_{e_j} v_k v_l\, dx$. Since $\left(v_{j+\frac{q}{n}}\right)_{0 \leqslant q \leqslant n}$ is a basis of $\mathbb{R}_n[x]$, we have $\dim\left(\ker(\mathcal{G})\right) = \dim(\Upsilon^{\perp})$ in $\mathbb{R}_n[x]$, and thus we need to show that $\dim(\Upsilon) = n$.

To finish the proof, let Θ be a linear application such that for $\Xi = (\xi_0, \xi_1, \xi_2, ..., \xi_n)^T$ belonging to $\mathbb{R}^{n+1}$ we have $\Theta(\Xi) = \sum_{r=0}^{n} \xi_r \frac{\partial v_{j+\frac{r}{n}}}{\partial x}$. Then, Ξ belongs to the kernel of Θ if and only if $\sum_{r=0}^{n} \xi_r v_{j+\frac{r}{n}}$ is a constant. The fact that $\left(v_{j+\frac{q}{n}}\right)_{0 \leqslant q \leqslant n}$ is a basis of $\mathbb{R}_n[x]$ ends the proof of property (i). It is interesting to note that the proof is true for any basis of $\mathbb{R}_n[x]$ and not only for the Lagrange polynomials employed in this study.

We now prove property (ii). As a consequence of property (i) of the present theorem we have $\operatorname{rank}(\mathcal{G}) = \operatorname{rank}(\mathcal{G}^T) = n$, and hence we deduce that $\dim\ker(\mathcal{G}) = \dim\ker(\mathcal{G}^T) = 1$. Let $\mathbf{u}^{\ker\mathcal{G}}$ and $\mathbf{u}^{\ker\mathcal{G}^T}$ be the vectors with $n+1$ components belonging to $\ker(\mathcal{G})$ and $\ker(\mathcal{G}^T)$, respectively. On element e_j of ε_h, $j = 1, 2, ..., m$, we have

$$\mathbf{u}^{\ker\mathcal{G}} = (1, u_1^{\ker\mathcal{G}}, ..., u_{n-1}^{\ker\mathcal{G}}, u_n^{\ker\mathcal{G}})^T, \tag{50}$$

and

$$\mathbf{u}^{\ker\mathcal{G}^T} = (u_0^{\ker\mathcal{G}^T}, u_1^{\ker\mathcal{G}^T}, ..., u_{n-1}^{\ker\mathcal{G}^T}, 1)^T, \tag{51}$$

where $\mathbf{u}^{\ker \mathcal{G}}$ and $\mathbf{u}^{\ker \mathcal{G}^T}$ have been normalized by setting $u_0^{\ker \mathcal{G}} = 1$ and $u_n^{\ker \mathcal{G}^T} = 1$, respectively. We let

$$\mathcal{G}^L = \begin{pmatrix} & & \mathcal{I}_{n\times n} & & \\ & & & & \\ -u_0^{\ker \mathcal{G}^T} & -u_1^{\ker \mathcal{G}^T} & \cdots & -u_{n-2}^{\ker \mathcal{G}^T} & -u_{n-1}^{\ker \mathcal{G}^T} \end{pmatrix}, \quad \mathcal{G}^R = \begin{pmatrix} -u_1^{\ker \mathcal{G}} & & \\ -u_2^{\ker \mathcal{G}} & & \\ & \mathcal{I}_{n\times n} & \\ -u_{n-1}^{\ker \mathcal{G}} & & \\ -u_n^{\ker \mathcal{G}} & & \end{pmatrix}, \tag{52}$$

where $\mathcal{G}^L$ and $\mathcal{G}^R$ are matrices of size $(n+1)\times n$ and $n\times(n+1)$, respectively, and $\mathcal{I}_{n\times n}$ is the $n \times n$ identity matrix.

We have the following property

$$\mathcal{G} = \mathcal{G}^L\, \mathcal{G}^\diamond\, \mathcal{G}^R. \tag{53}$$

Indeed, the computation $\mathcal{G}^\diamond\, \mathcal{G}^R$ gives a $n \times (n+1)$ matrix coinciding with the n first lines of $\mathcal{G}$ due to (20). Then, multiplying the resulting matrix on the left by $\mathcal{G}^L$ gives the full $(n+1) \times (n+1)$ matrix $\mathcal{G}$ taking into account that $\mathcal{G}^T\, \mathbf{u}^{\ker \mathcal{G}^T} = \mathbf{0}$ and $u_n^{\ker \mathcal{G}^T} = 1$.

The definition of $\mathcal{G}^L$ and $\mathcal{G}^R$ in (52) yields $\mathrm{rank}(\mathcal{G}^L) = \mathrm{rank}(\mathcal{G}^R) = n$. Further, $\mathrm{rank}(\mathcal{G}) = n$ as shown in property (i). Since $\mathrm{rank}(\mathcal{G}^L\, \mathcal{G}^\diamond\, \mathcal{G}^R) \leqslant \min(n, \mathrm{rank}(\mathcal{G}^\diamond))$, we deduced from (53) that $\mathrm{rank}(\mathcal{G}^\diamond) = n$. Consequently, the $n \times n$ matrix $\mathcal{G}^\diamond$ is full rank with $\det(\mathcal{G}^\diamond) \neq 0$.

Finally, we now prove property (iii). Searching for stationary erratic modes leads to let $\omega = 0$ in (10), and hence to compute the kernel of the discrete gradient operator in (11) - (13). Due to symmetry reasons, only selected discrete equations are considered, namely,

- (i) $n-1$ equations at interior nodes $j + \frac{q}{n}$ over e_j,
- (ii) $n-1$ equations at interior nodes $j - \frac{q}{n}$ over e_{j-1}, for $1 \leqslant q \leqslant n-1$,
- (iii) one equation at a typical boundary element node j.

From (11) - (13) we then obtain the following system of $2n-1$ equations

with the $2n+1$ unknowns: $u_{j-1}, u_{j-\frac{n-1}{n}}, ..., u_{j-\frac{1}{n}}, u_j, u_{j+\frac{1}{n}}, ..., u_{j+\frac{n-1}{n}}, u_{j+1}$,

$$\sum_{r=0}^{n} u_{j+\frac{r}{n}} \int_{e_j} \frac{\partial v_{j+\frac{r}{n}}}{\partial x} v_{j+\frac{q}{n}}\, dx = 0, \quad \text{at node } j+\frac{q}{n}, \quad \text{for } 1 \leqslant q \leqslant n-1, \tag{54}$$

$$\sum_{r=0}^{n} u_{j-\frac{r}{n}} \int_{e_{j-1}} \frac{\partial v_{j-\frac{r}{n}}}{\partial x} v_{j-\frac{q}{n}}\, dx = 0, \quad \text{at node } j-\frac{q}{n}, \quad \text{for } 1 \leqslant q \leqslant n-1, \tag{55}$$

$$\sum_{r=0}^{n} u_{j-\frac{r}{n}} \int_{e_{j-1}} \frac{\partial v_{j-\frac{r}{n}}}{\partial x} v_j\, dx + \sum_{r=0}^{n} u_{j+\frac{r}{n}} \int_{e_j} \frac{\partial v_{j+\frac{r}{n}}}{\partial x} v_j\, dx = 0, \quad \text{at node } j. \tag{56}$$

By employing (18) and (22), and the following property

$$\int_{e_{j-1}} \frac{\partial v_{j-\frac{r}{n}}}{\partial x} v_{j-\frac{q}{n}}\, dx = \int_{e_j} \frac{\partial v_{j+\frac{n-r}{n}}}{\partial x} v_{j+\frac{n-q}{n}}\, dx, \tag{57}$$

obtained from (9), for $0 \leqslant q, r \leqslant n$, equations (54) - (56) are rewritten on the form

$$\sum_{r=0}^{n} u_{j+\frac{r}{n}} G_{q,r} = 0, \quad \text{for } 1 \leqslant q \leqslant n-1, \tag{58}$$

$$-\sum_{r=0}^{n} u_{j-\frac{r}{n}} G_{q,r} = 0, \quad \text{for } 1 \leqslant q \leqslant n-1, \tag{59}$$

$$\sum_{r=0}^{n} \left(u_{j+\frac{r}{n}} - u_{j-\frac{r}{n}} \right) G_{0,r} = 0. \tag{60}$$

By adding (58) and (59), and employing (60), the following system of n equations is obtained

$$\sum_{r=1}^{n} \left(u_{j+\frac{r}{n}} - u_{j-\frac{r}{n}} \right) G_{q,r} = 0, \quad \text{for } 0 \leqslant q \leqslant n-1. \tag{61}$$

The matrix of the homogeneous linear system in (61) is nothing else than the matrix $\mathcal{G}^{\diamond}$ defined in (49), with $\det(\mathcal{G}^{\diamond}) \neq 0$ obtained from property (ii). Consequently, equation (61) leads to $u_{j+\frac{r}{n}} = u_{j-\frac{r}{n}}$, for $1 \leqslant r \leqslant n$.

It remains to solve (58), a homogeneous linear system of $n-1$ equations with the $n+1$ unknowns: $u_j, u_{j+\frac{1}{n}}, u_{j+\frac{2}{n}}, ..., u_{j+\frac{n-1}{n}}, u_{j+1}$. Let $\mathcal{G}^{\square}$ be the $(n-1)\times(n+1)$ matrix in the left hand side of (58), obtained by removing the first and the last row of $\mathcal{G}$, namely

$$\mathcal{G}^{\square} = (G_{q,r}), \quad \text{for } 1 \leqslant q \leqslant n-1 \text{ and } 0 \leqslant r \leqslant n. \tag{62}$$

Further, let $\mathbf{g}_r^{\square}$ be the $n+1$ column vectors of $\mathcal{G}^{\square}$, for $0 \leqslant r \leqslant n$, and $\mu_r \in \mathbb{R}$ for $0 \leqslant r \leqslant n$. Due to (20) we obtain

$$\sum_{r=0}^{n} \mu_r \, \mathbf{g}_r^{\square} = \sum_{r=1}^{n} (\mu_r - \mu_0) \, \mathbf{g}_r^{\square}, \tag{63}$$

Since the n column vectors in the right hand side of (63) also coincide with the $n-1$ row vectors of $\mathcal{G}^{\diamond}$, with $\text{rank}(\mathcal{G}^{\diamond}) = n$, we deduce that $\text{rank}(\mathcal{G}^{\square}) = n-1$, and consequently, $\ker(\mathcal{G}^{\square}) = 2$.

Two independent vectors thus lie in the kernel of $\mathcal{G}^{\square}$. The first vector is nothing else than the geostrophic mode of the form $(1,1,1,...,1)^T$, which reflects the fact that u is defined up to a constant in the continuous equations. It corresponds to $\mathbf{u}^{\ker \mathcal{G}}$ in (50) and its form is deduced here at the discrete level by invoking the property (20). The second vector takes the form of a unique erratic stationary mode. It is the highest frequency mode made up of alternating positive and negative values and has the largest phase speed error. It corresponds to $\mathbf{u}^{\ker \mathcal{G}^T}$ in (51). Examples of a such erratic mode are computed from (58), for $n = 1, 2, 3, 4$, and they are displayed in Figure 5. The alternating of positive and negative values is typical for this type of mode. $\square$

The n solutions of $P_n^{CG}(\widetilde{\omega})$ are shown in Figures 7 and 8 for $n = 2, \ldots, 7$, on the left side, distinguishing between odd (Figure 7) and even n (Figure 8) as discussed in Corollary 1 (v) and (vi). At a given spatial wavenumber kh there are n solutions, and each solution is given its own color. Each of these n solutions represents the n Fourier modes present in each eigenmode solution of the discrete wave equation. Only one of these solutions corresponds to the physical solution, while the remaining solutions are mathematical artifacts that arise from symmetries in the Fourier analysis. Each solution is valid over only a limited wavenumber range, this is termed a branch and the union of all such branches gives the complete dispersion relationship, named $\widetilde{\omega}_S$ in

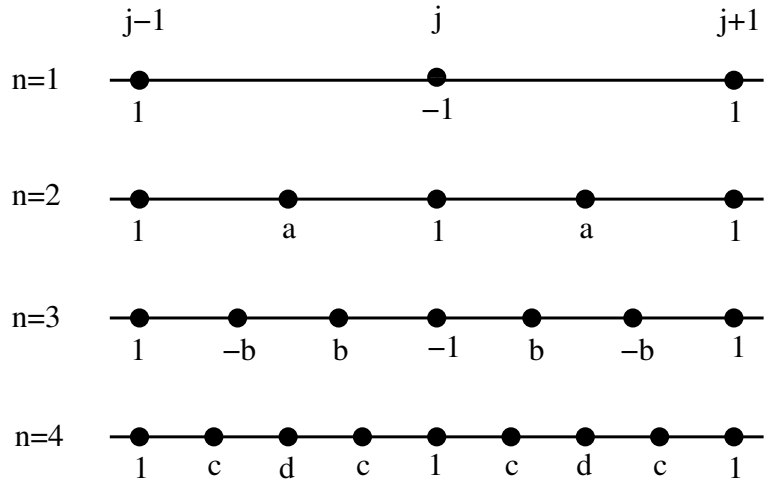

Figure 5: Examples of erratic modes computed from (58) on the intervals $e_{j-1} \cup e_j$ of ε_h, for $n = 1, 2, 3, 4$, with $a = -1/2$, $b = 11/27$, $c = -37/128$ and $d = 3/8$.

the right hand side of the figures. The ranges of validity for $n = 1, \ldots, 10$ are given in table 1, noting that they ensure that for a given $\widetilde{\omega}_S$ there are only two associated spatial wavenumbers.

The physical solution and the mathematical artifacts can be distinguished by inspecting the spatial structure of u_h corresponding to each branch, obtained by using the eigenvectors. We will now give some examples of this for $n = 3$ and $n = 4$ for several kh. The spatial structure is computed as follows: First numerically solve the generalized eigenvalue problem (28) for a given kh, yielding a set of n eigenvalues and eigenvectors, which are the n Fourier amplitudes. Then the solution can be reconstructed on any mesh using the definition of nodal values in terms of Fourier amplitudes at the beginning of section 3.2. In Figures 9 - 12 we have chosen to use a mesh with 12 elements of unit width, which corresponds to $h = 1/n$.

Consider first the case $n = 4$ computed at $kh = \pi/4$, found in Figure 9, where the colors correspond to those used in Figure 8. The red reconstructed solution corresponds to a spatial wavenumber of $kh = -3\pi/4$, not $kh = \pi/4$. The magenta reconstructed solution corresponds to a spatial wavenumber of $kh = -\pi/4$, not $kh = \pi/4$. The green reconstructed solution corresponds to a spatial wavenumber of $kh = -\pi/4$, as expected. The blue reconstructed solution corresponds to a spatial wavenumber of $kh = 3\pi/4$, not $kh = \pi/4$. It is clear that the solutions with negative kh represent a mirror symmetry that

arises when both x and t are inverted. Identical results are obtained if we instead compute the eigenvectors associated with $kh = -3\pi/4$; $kh = -\pi/4$ or $kh = 3\pi/4$ (not shown).

It is also useful to consider the behavior at the largest representable wavelength on the mesh, since this captures the behavior in the limit as h goes to 0. This is done for $n = 4$ at $kh = \pm\pi/24$ in Figures 10 and 11, and for $n = 3$ at $kh = \pi/18$ in Figure 12 (the results for $kh = -\pi/18$ are identical and not shown). For $n = 4$ the green and magenta solutions exchange, and the red and blue solutions exchange between $kh = \pi/24$ and $kh = -\pi/24$. This is a demonstration of the consequences of Corollary 2 for n even. The stationary mode, in the form of a wave packet consisting of a superposition of a high-frequency mode near the grid scale and a low-frequency mode near $kh = 0$, is clearly demonstrated in the magenta solution in Figure 10 at $kh = \pi/24$ and the green solution at $kh = -\pi/24$ in Figure 11. For $n = 3$ the three solutions do not change between $kh = \pi/18$ and $kh = -\pi/18$. This is a demonstration of the consequences of Corollary 2 for n odd. Although this is not a proof, these results are a strong argument for the validity of the branch selection procedure used to construct the dispersion relationship $\widetilde{\omega}_S$ from $P_n^{CG}(\widetilde{\omega})$. The final dispersion relationship obtained after doing so is shown on the right side of Figures 7 and 8.

A final line of evidence supporting the branch selection procedure is found in Figure 13. Here, the numerical eigenvalues and eigenvectors for the discrete linear operator with $n = 1$ and $n = 2$ on a mesh with 32 element were computed. From these, spatial wavenumbers were associated to each eigenvalue/eigenvector pair by performing a discrete Fourier transform (DFFT) on the eigenvector and letting the discrete Fourier component with the largest power set the spatial wavenumber. The exact same results are obtained as in Figures 4 and 8. This maximal amplitude procedure is in fact the same approach as used in [17], which can be seen by comparing [17, Figure 11] with Figures 4, 8 and 13. For $n \geqslant 4$, this maximum amplitude approach produces less optimal results than our branch selection procedure, leading to a multiple valued dispersion relation and missing wave numbers in k.

The spectral gaps discussed in Corollary 4 that occur for $n \geqslant 3$ are clearly visible in Figures 7 and 8. Spectral gaps are spatial wavenumbers where the range of $\omega(kh)$ contains gaps and it is hence discontinuous. Wave packets with energy at this wavenumber will fail to propagate correctly, and there will be significant numerical dispersion and other undesirable artifacts. In fact, the gaps will always come in pairs and there are $\text{floor}((n-1)/2)$ pairs.

It is interesting to note that although the number of pairs increases as n increases, the gaps in the low-frequency part of the spectrum decrease in size, while only the last pair in the highest-frequency part of the spectrum increases in size.

It is also interesting to consider the behaviour at the end of the spectrum. For both even and odd n the slope of the dispersion relationship at $kh = \pi$ is $-(2n+1)$ and $\widetilde{\omega} = 0$ (corresponding to the erratic stationary mode). Such an incorrect slope is key to understanding the behaviour of discrete simulations. Consider, for example, the propagation of a Gaussian as shown in Figure 6 at intial time and at time 2. Here we consider a mesh of 11 elements of width 2 with $n = 2, 3, 4, 5$. In all cases there is an anomalous wave packet with a group velocity of $-(2n+1)$ (as predicted). This can be understood as follows. The Gaussian initial condition projects onto a range of wavenumbers, and for a given $\widetilde{\omega}$ (when it is representable), there are two spatial wavenumbers, giving rise to two distinct wavepackets. One wavepacket will have a group velocity close to 1, and the other group velocity close to $-(2n + 1)$. Both of these wavepackets are clearly visible on Figure 6. The (anomalous) wave packet with a group velocity of $-(2n + 1)$ is a manifestation of the slope of $-(2n + 1)$ at the end of the spectrum and the erratic stationary mode.

Additionally, the maximal frequency increases as n increases, which is a likely cause of the observation that the CFL limit gets progressively stricter with higher n.

In the next section, we now consider the discontinuous approximation of (6).

4. Discontinuous Galerkin discretizations

4.1. Discontinuous Galerkin formulations

The setting of Section 3.1 is still used, except that ε_h is now a finite collection of m *open* elements e_j, $j = 1, 2, ..., m$, of the real line, such that $\bar{\Omega} = \bigcup_{e_j \in \varepsilon_h} \bar{e}_j$ and $e_i \cap e_j = \varnothing$ for $i \neq j$. The so-called *broken space* $H^1(\varepsilon_h)$

is defined as $H^1(\varepsilon_h) = \{\, v \in L^2(\Omega);\ v|_e \in H^1(e),\ \forall\, e \in \varepsilon_h\}$, where e simply denotes an element e_j of ε_h, $j = 1, 2, ..., m$.

Let u be a sufficiently smooth function. Multiplying (6) by a function v belonging to $H^1(\varepsilon_h)$, and integrating over the domain Ω we obtain

$$i\omega \sum_{j=1}^{m} \int_{e_j} u\, v\, dx - c \sum_{j=1}^{m} \int_{e_j} \frac{\partial u}{\partial x}\, v\, dx = 0, \tag{64}$$

an equation similar to (8), except that u and v are now discontinuous at the element boundaries of e_j, $j = 1, 2, ..., m$, and the integrals are computed over e_j and not Ω. To obtain the DG formulation the integrals in (64) are integrated by parts, yielding

$$i\omega \sum_{j=1}^{m} \int_{e_j} u\, v\, dx + c \sum_{j=1}^{m} \left(\int_{e_j} u\, \frac{\partial v}{\partial x}\, dx - u^* v \,|_{j^+}^{(j+1)^-} \right) = 0, \tag{65}$$

where u^* denotes the numerical trace of u and j^- and j^+ being the nodal positions of adjacent elements corresponding to a typical node j. That is, node j corresponds to the coincident node pair (j^-/j^+), $j = 1, 2, 3, ..., m+1$, as shown in Figure 14.

The numerical trace u^* being uniquely defined at the element boudary, we obtain

$$\sum_{j=1}^{m} u^* v \,|_{j^+}^{(j+1)^-} = \sum_{j=1}^{m} \left(u^*_{j+1} v(x_{(j+1)^-}) - u^*_j v(x_{j^+}) \right). \tag{66}$$

Developing the sum, rearranging the terms and applying the periodic boundary condition, equation (66) leads to

$$\sum_{j=1}^{m} u^* v \,|_{j^+}^{(j+1)^-} = \sum_{j=1}^{m} u^*_j \left(v(x_{j^-}) - v(x_{j^+}) \right). \tag{67}$$

Let $[\, v(x_j)\,] = v(x_{j^-}) - v(x_{j^+})$ and $\{u(x_j)\} = \frac{1}{2}\big(u(x_{j^-}) + u(x_{j^+})\big)$ be the jump of v and the mean of u, respectively, at node j. For a real parameter λ, the numerical trace is chosen as

$$u^*_j := u^*_j(\lambda) = \{u(x_j)\} + \Big(\frac{1}{2} - \lambda\Big)[\, u(x_j)\,] = (1-\lambda)\, u(x_{j^-}) + \lambda\, u(x_{j^+}), \tag{68}$$

i.e. the weighted averages and jump of u, respectively, at node j, for $j = 1, 2, ..., m$. For example, the choice $\lambda = 1$ corresponds to the upwind case, as the wave is progressing from the right part of the domain to the left one, while the case $\lambda = 1/2$ corresponds to the centered flux case. The variational formulation for the DG method then reads

$$i\omega \sum_{j=1}^{m} \int_{e_j} u\, v\, dx + c \sum_{j=1}^{m} \int_{e_j} u\, \frac{\partial v}{\partial x}\, dx - c \sum_{j=1}^{m} u_j^*(\lambda)\, [\, v(x_j)\,] = 0. \tag{69}$$

In the DG approximation, we consider finite-element spaces V_h^n of polynomial functions, discontinuous at the element interfaces, such that

$$V_h^n = \{\, u_h \in L^2(\Omega);\ u_h|_e = \widetilde{u}_h \circ F_e^{-1},\ \widetilde{u}_h \in \mathcal{P}_n(\widetilde{e}),\ \forall\, e \in \varepsilon_h \}\,,$$

where F_e and $\mathcal{P}_n(\widetilde{e})$ are defined as in Section 3.1. The basis $\mathcal{V}_h^n$ of V_h^n is a basis of the $n+1$ Lagrange interpolating functions of degree n given in (9), with $\mathcal{V}_h^n = \{v_s\}$, for $s \in \mathcal{J}_j^{\,n}$, $j = 1, 2, ..., m$, where

$$\mathcal{J}_j^{\,n} = \left\{ j^+, j + \frac{1}{n}, j + \frac{2}{n}, j + \frac{3}{n}, ..., j + \frac{n-1}{n}, (j+1)^- \right\}$$

is the set of indices of the local degrees of freedom on each element e_j of ε_h, as shown in Figure 15. The mesh is still uniform with $h = x_{j+\frac{q+1}{n}} - x_{j+\frac{q}{n}}$, for $j = 1, 2, ..., m$, and $q = 0, 1, 2, ..., n-1$.

Introducing the basis $\mathcal{V}_h^n$ leads to a discretization of (69), that consists of finding u_h belonging to V_h^n for the selected bases, at node $s \in \mathcal{J}_j^{\,n}$, such that

$$i\omega \sum_{j=1}^{m} \int_{e_j} u_h v_s\, dx + c \sum_{j=1}^{m} \int_{e_j} u_h\, \frac{\partial v_s}{\partial x}\, dx - c \sum_{j=1}^{m} u_j^*(\lambda)\, [\, v_s(x_j)\,] = 0, \quad \forall v_s \in V_h^n. \tag{70}$$

The discretization of (70) at nodes $s = j + \frac{q}{n}$, for $q = 1, 2, ..., n-1$, $s = j^+$ and $s = (j+1)^-$, $j = 1, 2, ..., m$, is now performed in the purpose of the Fourier analysis.

First, equation (70) is computed at nodes $s = j + \frac{q}{n}$, for $q = 1, 2, ..., n-1$. Expanding u_h over e_j in the basis $\mathcal{V}_h^n$, namely $u_h|_{e_j} = \sum_{r \in \mathcal{J}_j^{\,n}} u_r v_r$, for $v_r \in \mathcal{V}_h^n$, and taking into account that the basis function v_s is zero outside e_j and

$[\, v_s(x_j)\,] = 0$, for $j = 1, 2, ..., m$, yields

$$i\omega \left(u_{j^+} \int_{e_j} v_{j^+}\, v_{j+\frac{q}{n}}\, dx + \sum_{r=1}^{n-1} u_{j+\frac{r}{n}} \int_{e_j} v_{j+\frac{r}{n}}\, v_{j+\frac{q}{n}}\, dx + u_{(j+1)^-} \int_{e_j} v_{(j+1)^-}\, v_{j+\frac{q}{n}}\, dx \right)$$
$$+ c \left(u_{j^+} \int_{e_j} v_{j^+}\, \frac{\partial v_{j+\frac{q}{n}}}{\partial x}\, dx + \sum_{r=1}^{n-1} u_{j+\frac{r}{n}} \int_{e_j} v_{j+\frac{r}{n}}\, \frac{\partial v_{j+\frac{q}{n}}}{\partial x}\, dx \right.$$
$$\left. + u_{(j+1)^-} \int_{e_j} v_{(j+1)^-}\, \frac{\partial v_{j+\frac{q}{n}}}{\partial x}\, dx \right) = 0. \qquad (71)$$

Equation (70) is then computed at node $s = j^+$, $j = 1, 2, ..., m$,

$$i\omega \left(u_{j^+} \int_{e_j} v_{j^+}^2\, dx + \sum_{r=1}^{n-1} u_{j+\frac{r}{n}} \int_{e_j} v_{j+\frac{r}{n}}\, v_{j^+}\, dx + u_{(j+1)^-} \int_{e_j} v_{(j+1)^-}\, v_{j^+}\, dx \right)$$
$$+ c \left((1-\lambda) u_{j^-} + u_{j^+} \Big(\lambda + \int_{e_j} v_{j^+}\, \frac{\partial v_{j^+}}{\partial x}\, dx \Big) \right.$$
$$\left. + \sum_{r=1}^{n-1} u_{j+\frac{r}{n}} \int_{e_j} v_{j+\frac{r}{n}}\, \frac{\partial v_{j^+}}{\partial x}\, dx + u_{(j+1)^-} \int_{e_j} v_{(j+1)^-}\, \frac{\partial v_{j^+}}{\partial x}\, dx \right) = 0, \quad (72)$$

where u_h has been expanded over element e_j and $[\, v_{j^+}(x_j)\,] = -1$ since $v_{j^+}(x_{j^-}) = 0$ and $v_{j^+}(x_{j^+}) = 1$. Further, $[\, v_{j^+}(x_{j_0})\,] = 0$ for $j_0 \neq j$. Following the same procedure, but for $s = (j+1)^-$, $j = 1, 2, ..., m$, equation (70) leads to

$$i\omega \left(u_{j^+} \int_{e_j} v_{j^+}\, v_{(j+1)^-}\, dx + \sum_{r=1}^{n-1} u_{j+\frac{r}{n}} \int_{e_j} v_{j+\frac{r}{n}}\, v_{(j+1)^-}\, dx + u_{(j+1)^-} \int_{e_j} v_{(j+1)^-}^2\, dx \right)$$
$$+ c \left(u_{j^+} \int_{e_j} v_{j^+}\, \frac{\partial v_{(j+1)^-}}{\partial x}\, dx + \sum_{r=1}^{n-1} u_{j+\frac{r}{n}} \int_{e_j} v_{j+\frac{r}{n}}\, \frac{\partial v_{(j+1)^-}}{\partial x}\, dx \right.$$
$$\left. + u_{(j+1)^-} \Big(-(1-\lambda) + \int_{e_j} v_{(j+1)^-}\, \frac{\partial v_{(j+1)^-}}{\partial x}\, dx \Big) - \lambda u_{(j+1)^+} \right) = 0, \quad (73)$$

where $[\, v_{(j+1)^-}(x_{j+1})\,] = 1$ since $v_{(j+1)^-}(x_{(j+1)^-}) = 1$ and $v_{(j+1)^-}(x_{(j+1)^+}) = 0$, and $[\, v_{(j+1)^-}(x_{j_0})\,] = 0$ for $j_0 \neq j+1$.

For $j = 1, 2, ..., m$, and due to periodicity at end nodes, $m(n-1)$ equations are obtained from (71) at internal element nodes and (72) and (73) provide m

additional equations each, leading to a total of $m(n+1)$ discrete equations, instead of mn equations for the CG scheme.

Example 5. *In the case $n = 1$, equation (72) at node j^+ for $v = v_{j^+}$ leads to*

$$i\omega\frac{h}{6}\left(2u_{j^+} + u_{(j+1)^-}\right) + c\left((1-\lambda)\,u_{j^-} + (\lambda - \frac{1}{2})\,u_{j^+} - \frac{1}{2}\,u_{(j+1)^-}\right) = 0, \tag{74}$$

while (73) at node $(j+1)^-$ for $v = v_{(j+1)^-}$ yields

$$i\omega\frac{h}{6}\left(u_{j^+} + 2u_{(j+1)^-}\right) + c\left(\frac{1}{2}\,u_{j^+} + (\lambda - \frac{1}{2})\,u_{(j+1)^-} - \lambda\,u_{(j+1)^+}\right) = 0, \tag{75}$$

for $j = 1, 2, ..., m$.

4.2. The Fourier analysis

As for the CG scheme, a dispersion analysis is now performed for the DG schemes and again, due to symmetry reasons, only selected discrete equations are considered, Equation (71) provides $n-1$ equations at interior nodes $j+\frac{r}{n}$, $r = 1, 2, ..., n-1$, and two equations are obtained from (72) and (73) at the typical nodes j^+ and $(j+1)^-$, respectively. The periodic solutions read

(i) $u_{j+\frac{r}{n}} = \widehat{u}_r\, e^{ikx_{j+\frac{r}{n}}}$ at the $n-1$ interior nodes $j+\frac{r}{n}$ of e_j, $r = 1, 2, ..., n-1$,

(ii) $u_{j^+} = \widehat{u}_n\, e^{ikx_j}$ and $u_{(j+1)^-} = \widehat{u}_{n+1}\, e^{ikx_{j+1}}$ at the boundary nodes of e_j,

and hence, $n+1$ amplitudes are involved, instead of n for the CG scheme.

Substituting $u_{j+\frac{r}{n}}$, u_{j^+} and $u_{(j+1)^-}$ in (71) leads to

$$\sum_{r=1}^{n-1} \widehat{u}_r\, e^{i(r-q)kh}\left(i\widetilde{\omega}M_{j+\frac{q}{n},j+\frac{r}{n}} + G_{j+\frac{q}{n},j+\frac{r}{n}}\right) + \widehat{u}_n e^{-iqkh}\left(i\widetilde{\omega}M_{j+\frac{q}{n},j^+} + G_{j+\frac{q}{n},j^+}\right)$$
$$+ \widehat{u}_{n+1}\, e^{i(n-q)kh}\left(i\widetilde{\omega}M_{j+\frac{q}{n},(j+1)^-} + G_{j+\frac{q}{n},(j+1)^-}\right) = 0, \tag{76}$$

for $q = 1, 2, ..., n-1$, and substituting in (72) yields

$$\sum_{r=1}^{n-1} \widehat{u}_r\, e^{irkh}\left(i\widetilde{\omega}M_{j^+,j+\frac{r}{n}} + G_{j^+,j+\frac{r}{n}}\right) + \widehat{u}_n\left(i\widetilde{\omega}M_{j^+,j^+} + G_{j^+,j^+} + \lambda\right)$$
$$+ \widehat{u}_{n+1}\left(e^{inkh}\left(i\widetilde{\omega}M_{j^+,(j+1)^-} + G_{j^+,(j+1)^-}\right) + 1 - \lambda\right) = 0. \tag{77}$$

Finally, substituting $u_{j+\frac{r}{n}}$, u_{j^+} and $u_{(j+1)^-}$ in (73), we obtain

$$\sum_{r=1}^{n-1} \widehat{u}_r\, e^{i(r-n)kh} \left(i\widetilde{\omega} M_{(j+1)^-,j+\frac{r}{n}} + G_{(j+1)^-,j+\frac{r}{n}}\right) + \widehat{u}_n \left(e^{-inkh}\left(i\widetilde{\omega} M_{(j+1)^-,j^+} + G_{(j+1)^-,j^+}\right) - \lambda\right) + \widehat{u}_{n+1} \left(i\widetilde{\omega} M_{(j+1)^-,(j+1)^-} + G_{(j+1)^-,(j+1)^-} + \lambda - 1\right) = 0. \qquad (78)$$

Equations (76) – (78) lead to a $(n+1)\times(n+1)$ matrix system

$$(\widehat{\mathcal{G}} - i\widetilde{\omega}\widehat{\mathcal{M}})\,\widehat{\mathbf{U}} = 0, \qquad (79)$$

for the amplitudes, with $\widehat{\mathbf{U}} = (\widehat{u}_1, \widehat{u}_2, ..., \widehat{u}_{n+1})$ where $\widehat{\mathcal{G}}$ and $\widehat{\mathcal{M}}$ are the $(n+1)\times(n+1)$ matrices:

$$\widehat{\mathcal{G}} = \begin{pmatrix} G_{1,1} & e^{ikh}G_{1,2} & \cdots & e^{i(n-2)kh}G_{1,n-1} & e^{-ikh}G_{1,0} & e^{i(n-1)kh}\,G_{1,n} \\ e^{-ikh}G_{2,1} & G_{2,2} & \cdots & e^{i(n-3)kh}G_{2,n-1} & e^{-2ikh}G_{2,0} & e^{i(n-2)kh}\,G_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ e^{-(n-2)ikh}G_{n-1,1} & e^{-(n-3)ikh}G_{n-1,2} & \cdots & G_{n-1,n-1} & e^{-(n-1)ikh}G_{n-1,0} & e^{ikh}\,G_{n-1,n} \\ e^{ikh}G_{0,1} & e^{2ikh}G_{0,2} & \cdots & e^{(n-1)ikh}G_{0,n-1} & G_{0,0}+\lambda & e^{inkh}\,G_{0,n}+1-\lambda \\ e^{-i(n-1)kh}G_{n,1} & e^{-(n-2)ikh}G_{n,2} & \cdots & e^{-ikh}G_{n,n-1} & e^{-inkh}G_{n,0}-\lambda & G_{n,n}+\lambda-1 \end{pmatrix}$$

$$\widehat{\mathcal{M}} = \begin{pmatrix} M_{1,1} & e^{ikh}M_{1,2} & \cdots & e^{i(n-2)kh}M_{1,n-1} & e^{-ikh}M_{1,0} & e^{i(n-1)kh}\,M_{1,n} \\ e^{-ikh}M_{2,1} & M_{2,2} & \cdots & e^{i(n-3)kh}M_{2,n-1} & e^{-2ikh}M_{2,0} & e^{i(n-2)kh}\,M_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ e^{-(n-2)ikh}M_{n-1,1} & e^{-(n-3)ikh}M_{n-1,2} & \cdots & M_{n-1,n-1} & e^{-(n-1)ikh}M_{n-1,0} & e^{ikh}\,M_{n-1,n} \\ e^{ikh}M_{0,1} & e^{2ikh}M_{0,2} & \cdots & e^{(n-1)ikh}M_{0,n-1} & M_{0,0} & e^{inkh}\,M_{0,n} \\ e^{-i(n-1)kh}M_{n,1} & e^{-(n-2)ikh}M_{n,2} & \cdots & e^{-ikh}M_{n,n-1} & e^{-inkh}M_{n,0} & M_{n,n} \end{pmatrix}$$

Recalling that $M_{0,n-r} = M_{n-r,0} = M_{n,r}$ and $G_{0,n-r} = -G_{n-r,0} = -G_{n,r}$ due to (18), it is clear that $\widehat{\mathcal{G}}$ is skew-Hermitian if and only if $\lambda = 1/2$; and $\widehat{\mathcal{M}}$ is Hermitian.

A nontrivial solution $\widehat{\mathbf{U}}$ is obtained if the determinant of $(\widehat{\mathcal{G}} - i\widetilde{\omega}\widehat{\mathcal{M}})$ vanishes, leading to the dispersion relation, namely a polynomial in $\widetilde{\omega}$ of degree $n+1$. Note that for the CG scheme, the dispersion relation is a polynomial of degree n. The discrepancy is due to an extra amplitude in the DG scheme, reflecting the discontinuous aspect of the method.

Theorem 4. *For discontinuous finite-element approximations of degree $n \leqslant 10$, the characteristic equation or dispersion relation obtained from the generalized eigenvalue problem (79) is the polynomial of degree $n+1$*

$$P_{n+1}^{DG}(\widetilde{\omega}) = \widetilde{\omega}^{n+1} + \sum_{j=1}^{\lfloor \frac{n+1}{2} \rfloor} \tilde{b}_{n-2j+1}\, \widetilde{\omega}^{n-2j+1} + \sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} \tilde{b}_{n-2j}\, \widetilde{\omega}^{n-2j} = 0, \tag{80}$$

with

$$\begin{aligned}
\frac{\tilde{b}_{n-2j+1}}{a_{n-2j+1}} &= \frac{(n+1)^2}{2jn} + (-1)^n \frac{(n+1)}{n}\Big(\cos(nkh) + i\,(2\lambda-1)\sin(nkh)\Big), \\
\frac{\tilde{b}_{n-2j}}{a_{n-2j}} &= -i(2\lambda-1)\frac{(n+1)^2}{(2j+1)n} + (-1)^n \frac{(n+1)}{n}\Big(i(2\lambda-1)\cos(nkh) - \sin(nkh)\Big),
\end{aligned}$$

where λ is the parameter defined in (68). These results have been computed using a computer algebra system (Maple). It is conjectured that (80) *holds for any $n > 10$.*

By rearranging terms in (80), we obtain a simplified form

$$\begin{aligned}
&\cos(nkh)\Big(R_{n-1}(\widetilde{\omega}) + i(2\lambda-1)Q_n(\widetilde{\omega})\Big) + \sin(nkh)\Big(i(2\lambda-1)R_{n-1}(\widetilde{\omega}) - Q_n(\widetilde{\omega})\Big) \\
&+ T_{n+1}(\widetilde{\omega}) + i(2\lambda-1)S_n(\widetilde{\omega}) = 0,
\end{aligned} \tag{81}$$

where

$$T_{n+1}(\widetilde{\omega}) = (-1)^n \frac{n}{n+1}\widetilde{\omega}^{n+1} + (-1)^n (n+1) \sum_{j=1}^{\lfloor \frac{n+1}{2} \rfloor} \frac{a_{n-2j+1}}{2j}\, \widetilde{\omega}^{n-2j+1}.$$

Example 6. *For $n = 1$, the dispersion relation reads*

$$\widetilde{\omega}^2 + 2\widetilde{\omega}\Big(\sin(kh) - i\big(2\lambda - 1\big)\big(\cos(kh) + 2\big)\Big) \\ +6\Big(\cos(kh) - 1 + i(2\lambda - 1)\sin(kh)\Big) = 0, \tag{82}$$

and we obtain

$$T_2(\widetilde{\omega}) = -\tfrac{1}{2}\,\widetilde{\omega}^2 + 3, \qquad T_3(\widetilde{\omega}) = \tfrac{2}{3}\,\widetilde{\omega}^3 - 6\,\widetilde{\omega}, \qquad T_4(\widetilde{\omega}) = -\tfrac{3}{4}\,\widetilde{\omega}^4 + 10\,\widetilde{\omega}^2 - \tfrac{70}{9}.$$

Higher values for n are considered in the following.

For subsequent need we let

$$\mathcal{G}_\lambda := (G_{q,r,\lambda}) = \mathcal{G} + \Lambda, \quad \text{for } 0 \leqslant q, r \leqslant n, \tag{83}$$

where the matrix Λ in the right hand side of (83) is of size $(n+1) \times (n+1)$, with

$$\Lambda := \begin{pmatrix} \lambda & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \lambda - 1 \end{pmatrix}, \tag{84}$$

and we deduce from (20) and (21) the following properties

$$(i)\ \sum_{r=0}^{n} G_{q,r,\lambda} = 0, \quad \text{for } 1 \leqslant q \leqslant n-1, \quad \sum_{r=0}^{n} G_{0,r,\lambda} = \lambda, \quad \sum_{q=0}^{n} G_{q,0,\lambda} = \lambda - 1,$$

$$(ii)\ \sum_{q=0}^{n} G_{q,r,\lambda} = 0, \quad \text{for } 1 \leqslant r \leqslant n-1, \quad \sum_{r=0}^{n} G_{n,r,\lambda} = \lambda - 1, \quad \sum_{q=0}^{n} G_{q,n,\lambda} = \lambda,$$

(iii) $\mathcal{G}_{\frac{1}{2}}$ is a skew-symmetric matrix, namely, $\mathcal{G}_{\frac{1}{2}}^T = -\mathcal{G}_{\frac{1}{2}}$.

Example 7. *For $n = 3$, we obtain the gradient matrices of size 4×4*

$$\mathcal{G}_{\frac{1}{2}} = \begin{pmatrix} 0 & \frac{57}{80} & -\frac{3}{10} & \frac{7}{80} \\ -\frac{57}{80} & 0 & \frac{81}{80} & -\frac{3}{10} \\ \frac{3}{10} & -\frac{81}{80} & 0 & \frac{57}{80} \\ -\frac{7}{80} & \frac{3}{10} & -\frac{57}{80} & 0 \end{pmatrix}, \quad \mathcal{G}_1 = \begin{pmatrix} \frac{1}{2} & \frac{57}{80} & -\frac{3}{10} & \frac{7}{80} \\ -\frac{57}{80} & 0 & \frac{81}{80} & -\frac{3}{10} \\ \frac{3}{10} & -\frac{81}{80} & 0 & \frac{57}{80} \\ -\frac{7}{80} & \frac{3}{10} & -\frac{57}{80} & \frac{1}{2} \end{pmatrix}. \tag{85}$$

We now examine two important cases: $\lambda = 1/2$ (centered DG) and $\lambda = 1$ (upwind DG).

4.3. The centered Discontinuous Galerkin case ($\lambda = 1/2$)

In the case $\lambda = 1/2$, the polynomials $P_n^{CG}(\widetilde{\omega})$ and $P_{n+1}^{DG}(\widetilde{\omega})$ do not coincide. However, their expressions are very similar and hence several results obtained in section 3 are still valid with odd and even n exchanged, due to the fact that $\text{degree}(P_{n+1}^{DG}(\widetilde{\omega})) = \text{degree}(P_n^{CG}(\widetilde{\omega})) + 1$.

Corollary 5. *The dispersion relation when $\lambda = 1/2$ behaves similarly to the relation obtained for the CG scheme. In particular:*

(i) theorem 1 holds with $\widehat{\mathcal{G}}$ and $\widehat{\mathcal{M}}$ defined by (79).

(ii) corollary 1 and corollary 2 hold with odd and even n exchanged.

(iii) The convergence order in corollary 3 also holds with odd and even n exchanged, and for the modes of interest we obtain

For $n = 2p+1$, $p = 0, 1, 2, ...$	*For $n = 2p$, $p = 1, 2, 3, ...$*
$\widetilde{\omega}_{p+1} = kh + \kappa_p\,(kh)^{2n+1} + O(h^{2n+3})$,	$\widetilde{\omega}_p = kh - \kappa_{p+1}\,(kh)^{2n+3} + O(h^{2n+5})$,
$\widetilde{\omega}_{p+1} = -(2n+1)\,kh + O(h^3)$,	

where κ_p and κ_{p+1} are defined in (41).

(iv) A similar result to corollary 4 exists. When $\lambda = 1/2$, (81) *becomes*

$$\cos(nkh)R_{n-1}(\widetilde{\omega}) - \sin(nkh)Q_n(\widetilde{\omega}) + T_{n+1}(\widetilde{\omega}) = 0,$$

and therefore, employing the same reasoning than in the proof of corollary 4, we obtain

$$T_{n+1}(\widetilde{\omega}) = -\sqrt{Q_n^2(\widetilde{\omega}) + R_{n-1}^2(\widetilde{\omega})}\,\cos(nkh + \varphi),$$

and the following inequality must hold

$$|T_{n+1}(\widetilde{\omega})| < \sqrt{Q_n^2(\widetilde{\omega}) + R_{n-1}^2(\widetilde{\omega})}. \tag{86}$$

to avoid spectral gaps. In fact, floor(n/2) pairs of gaps occur when $n \geqslant 2$, and table 2 gives the intervals on which they exist.

(v) theorem 3 (iii) holds, namely for all $n \in \mathbb{N}$, the centered DG scheme admits a unique erratic stationary mode. Searching for stationary erratic modes leads to let $\omega = 0$ in (70) and hence to compute the kernel of the discrete gradient operator. As for the CG case, due to symmetry reasons, only the selected discrete equations are considered over interval e_j, namely,

a) $n-1$ equations at interior nodes $j + \frac{q}{n}$, $q = 1, 2, ..., n-1$,

b) one equation at a typical boundary element node j^+,

c) one equation at a typical boundary element node $(j+1)^-$.

From (71) - (73), we then obtain the following system of $n+1$ equations with the $n+3$ unknowns: $u_{j^-}, u_{j^+}, u_{j+\frac{1}{n}}, u_{j+\frac{2}{n}}, ..., u_{j+\frac{n-1}{n}}, u_{(j+1)^-}, u_{(j+1)^+}$, written in the form

$$\mathcal{G}_{\frac{1}{2}} \left(u_{j^+}, u_{j+\frac{1}{n}}, u_{j+\frac{2}{n}}, ..., u_{j+\frac{n-1}{n}}, u_{(j+1)^-}\right)^T = \left(-\tfrac{1}{2}u_{j^-}, 0, ..., 0, \tfrac{1}{2}u_{(j+1)^+}\right)^T. \tag{87}$$

As is the case for $\mathcal{G}$, the matrix $\mathcal{G}^\diamond$ is a leading principal submatrix of $\mathcal{G}_{\frac{1}{2}}$, and we have already established in theorem 3 (iii) that $\operatorname{rank}(\mathcal{G}^\diamond) = n$. Consequently, $\operatorname{rank}(\mathcal{G}_{\frac{1}{2}}) \geqslant n$. Further, since $\mathcal{G}_{\frac{1}{2}}$ is a skew-symmetric matrix, $\operatorname{rank}\left(\mathcal{G}_{\frac{1}{2}}\right)$ is an even integer.

For $n = 2p+1$, $p = 0, 1, 2, ...$, we hence deduce that $\operatorname{rank}(\mathcal{G}_{\frac{1}{2}}) = n+1$. For $n = 2p$, $p = 1, 2, 3, ...$, we have $\det(\mathcal{G}_{\frac{1}{2}}) = (-1)^{n+1}\det(\mathcal{G}_{\frac{1}{2}})$, since $\mathcal{G}_{\frac{1}{2}}$ is a skew-symmetric matrix, leading to $\det(\mathcal{G}_{\frac{1}{2}}) = 0$. We thus deduce that $\operatorname{rank}(\mathcal{G}_{\frac{1}{2}}) = n$. In both cases two vectors are solutions of (87). The first one is the geostrophic mode of the form $(1, 1, 1, ..., 1)^T$, as for the CG scheme. The second vector is the erratic stationary mode. As for the CG case, it is the highest frequency mode made up of alternating positive and negative values and has the largest phase speed error. From (87) we obtain over interval e_j

$$\begin{array}{lll} \text{for } n=1, & (u_{j^+}, u_{(j+1)^-}) & = (-1, 1), \\ \text{for } n=2, & (u_{j^+}, u_{j+\frac{1}{2}}, u_{(j+1)^-}) & = (1, -\frac{1}{2}, 1), \\ \text{for } n=3, & (u_{j^+}, u_{j+\frac{1}{3}}, u_{j+\frac{2}{3}}, u_{(j+1)^-}) & = (-1, \frac{11}{27}, -\frac{11}{27}, 1). \end{array}$$

Table 2: Intervals of values of $\widetilde{\omega}$, in the case $n = 1, 2, ..., 9$, for which the condition (86) is satisfied, namely when $\widetilde{\omega}$ cannot be computed.

n	$\mathcal{I}_{\widetilde{\omega}}(n)$
1	$\emptyset$ (i.e., $\widetilde{\omega}(kh)$ is continuous in kh)
2	$[1.152\,,\,1.611]$
3	$[1.601\,,\,2.509]$
4	$[0.7005\,,\,0.7098] \cup [1.877\,,\,3.217]$
5	$[1.1222\,,\,1.1722] \cup [2.086\,,\,3.871]$
6	$[0.48575\,,\,0.48587] \cup [1.399\,,\,1.513] \cup [2.270\,,\,4.510]$
7	$[0.83858\,,\,0.84071] \cup [1.597\,,\,1.788] \cup [2.445\,,\,5.145]$
8	$[0.370754\,,\,0.370755] \cup [1.104\,,\,1.113] \cup [1.751\,,\,2.027] \cup [2.621\,,\,5.779]$
9	$[0.662515\,,\,0.662571] \cup [1.308\,,\,1.332] \cup [1.879\,,\,2.248] \cup [2.802\,,\,6.412]$

The $n + 1$ solutions of $P_{n+1}^{DG}(\widetilde{\omega})$ are shown in Figures 17 and 18 for $n = 1, \ldots, 6$ on the left side, distinguishing between even (Figure 17) and odd n (Figure 18). Each solution is given its own color, and at a given spatial wavenumber kh there are $n + 1$ solutions but only one of these solutions corresponds to the physical solution. As for the CG scheme, each of these $n+1$ solutions represents the $n+1$ Fourier modes present in each eigenmode solution of the discrete wave equation. The remaining solutions are mathematical artifacts that arise from symmetries in the Fourier analysis, as was the case in section 3 for CG. Therefore, each solution is valid over only a limited wavenumber range; this is termed a branch. The union of all such branches gives the complete dispersion relationship. The physical solution and the mathematical artifacts can be distinguished by inspecting the spatial structure of u_h corresponding to each branch. The same branch selection procedure based on eigenvectors can be employed in the DG centered case as in the CG case. In fact, extremely similar reconstructed u_h are obtained (not shown). The final dispersion relationship obtained after doing so, namely $\widetilde{\omega}_S$, is shown on the right side of Figures 17 and 18.

In Figures 17 and 18, we consider $kh \in [-\frac{n+1}{n}\,\pi, \frac{n+1}{n}\,\pi]$ instead of $kh \in [-\pi, \pi]$ as for CG. This is due to the definition of the meshlength parameter as $h = L/m(n + 1) = h^*/(n + 1)$, which is given graphically in Figure 16.

When $n \geqslant 2$ the branch selection procedure will give rise to spectral gaps, namely, to a discontinuous representation for $\omega(kh)$. These gaps always come in pairs, and in fact there will be floor(n/2) pairs. Wave packets with energy at this wavenumber will fail to propagate correctly, and there will be significant numerical dispersion and other undesirable artifacts. It is interesting to note that although the number of pairs increases as n increases, the gaps in the low-frequency part of the spectrum decrease in size, while only the last pair in the highest-frequency part of the spectrum increases in size. Additionally, the maximal frequency increases with increasing n; this is likely the root cause of the increase in CFL condition noted as n increases. The spectral gaps that occur for $n \geqslant 2$ are clearly visible in Figures 17 and 18, as is the increasing maximal frequency and slope at the end of the spectrum.

4.4. The upwind Discontinuous Galerkin case ($\lambda = 1$)

Corollary 6. *When $\lambda = 1$, we obtain $n+1$ solutions $\widetilde{\omega}_j$ of $P_{n+1}^{DG}(\widetilde{\omega})$ with the following properties, which differ significantly from those found for CG and DG centered.*

(i) Only parts (ii) and (iii) of theorem 1 hold with $\widehat{\mathcal{M}}$ defined by (79). Crucially, (i) no longer holds, i.e. $\widehat{\mathcal{G}}$ is not skew-Hermitian due to the presence of the λ terms, and therefore (iv) also does not hold.

(ii) The solutions are periodic with

$$\widetilde{\omega}_j(kh) = \widetilde{\omega}_j(kh + \frac{2\pi}{n}),\ j = 1, 2, 3, ..., n+1,$$

(iii) The slope of the phase at $kh = \frac{(n+1)\pi}{n}$ (the end of the spectrum) is $-(2n+1)$, as for the CG and DG centered cases. The proof and calculations are similar to those found in corollary 2 for the CG case. However, as will be shown, this does not give rise to anomalous dispersion and a erratic stationary mode since the imaginary component of $\widetilde{\omega}$ acts to damp high-frequency components.

(iv) For all n (even or odd), in the limit as mesh spacing $h \to 0$, the roots $\widetilde{\omega}_j$, $j = 1, 2, ..., n+1$, of $P_{n+1}^{DG}(\widetilde{\omega})$ in (80), at least for $n \leqslant 10$, satisfy the following asymptotics obtained by using a computer algebra system

$$\Re(\widetilde{\omega}) = kh + \frac{2}{2n+1}\,\kappa_{p+1}\,h^{2n+3} + O(h^{2n+5}), \tag{88}$$

$$\Im(\widetilde{\omega}) = \frac{n(n+1)}{2n+1}\,\kappa_{p}\,h^{2n+2} + O(h^{2n+4}). \tag{89}$$

Similar comments to those following (41) *apply. In particular, the order of convergence have been formally proved in [1, equations (12) - (14)] for all n by using Bloch waves instead of the present Fourier approach. However, the leading terms are given in* (88) *and* (89) *for the dispersion and dissipation errors, respectively, in order to reflect the results of the Fourier approach.*

(v) When $\lambda = 1$, (81) *can be rewritten in the form*

$$e^{inkh}\Big(R_{n-1}(\widetilde{\omega}) + iQ_n(\widetilde{\omega})\Big) = -\Big(T_{n+1}(\widetilde{\omega}) + iS_n(\widetilde{\omega})\Big). \tag{90}$$

Since $\widetilde{\omega} \in \mathbb{C}$, equation (90) *is not easy to solve. However, for $n \leqslant 10$ we do not observe any gaps when $\lambda = 1$. In fact we observe graphically the disappearance of gaps when $\lambda \geqslant C > 1/2$, where C ranges from approximately* 0.65 *for $n = 2$ to approximately* 0.75 *for $n = 4$, for example.*

Finally, we have the following result.

Theorem 5. *For $\lambda = 1$ and $\forall n \in \mathbb{N}$, the upwind DG scheme does not admit any erratic stationary mode.*

Proof. Following the same procedure as in corollary 5 (v), but for $\lambda = 1$, equations (71) - (73) now lead to the following system of $n+1$ equations with the $n+2$ unknowns: $u_{j^+}, u_{j+\frac{1}{n}}, u_{j+\frac{2}{n}}, ..., u_{j+\frac{n-1}{n}}, u_{(j+1)^-}, u_{(j+1)^+}$, written in the form

$$\mathcal{G}_1\left(u_{j^+}, u_{j+\frac{1}{n}}, u_{j+\frac{2}{n}}, ..., u_{j+\frac{n-1}{n}}, u_{(j+1)^-}\right)^T = \left(0, 0, ..., 0, u_{(j+1)^+}\right)^T, \tag{91}$$

where the vector in the right hand side of (91) has $n+1$ components.

We first establish that $\mathcal{G}_1$ is non singular. From [16] we obtain

$$\det(\mathcal{G}_\lambda) = \det(\mathcal{G} + \Lambda) = \sum_{r=0}^{n+1} \sum_{\phi,\chi=0}^{n+1} (-1)^{s(\phi)+s(\chi)} \det(\mathcal{G}[\phi|\chi]) \det(\Lambda[\phi|\chi]), \quad (92)$$

where for a particular r, the inner sum is over strictly increasing integer sequences ϕ and χ of length r chosen from 1 to $n+1$; $\mathcal{G}[\phi|\chi]$ is the $r-$square submatrix of $\mathcal{G}$ lying in rows ϕ and columns χ; $\Lambda[\phi|\chi]$ is the $(n+1-r)-$square submatrix of Λ lying in rows complementary to ϕ and columns complementary to χ; and $s(\phi)$ is the sum of the integers in ϕ. Further, for $r = 0$ we have $\det(\mathcal{G}_\lambda) = \det(\Lambda)$, and for $r = n+1$ it comes $\det(\mathcal{G}_\lambda) = \det(\mathcal{G})$. The proof of (92) is nothing else than a consequence of the linearity of the determinant in each row of the matrix, and the Laplace expansion theorem.

We have $\det(\mathcal{G}) = 0$ from theorem 3 (i) and $\det(\Lambda) = 0$ for $n > 1$ from (84), with $\det(\Lambda) = \lambda(\lambda - 1)$ for $n = 1$. Further, due to the particular form of Λ, with only two non zero entries, the only summands that survive in the right hand side of (92) are those corresponding to $n + 1 - r = 1$ and $n + 1 - r = 2$.

- (i) For $n + 1 - r = 1$, two cases need to be considered. Firstly, the case $\phi = \chi = 1$ leading to $\lambda \det(\mathcal{G}^a)$, where $\mathcal{G}^a$ is a $n \times n$ matrix obtained from $\mathcal{G}$ by removing the first row and the first column, and we have $\det(\mathcal{G}^a) = \det(\mathcal{G}^\diamond)$. Secondly, the case $\phi = \chi = n + 1$, leading to $(\lambda - 1) \det(\mathcal{G}^b)$, where $\mathcal{G}^b$ is a $n \times n$ matrix obtained from $\mathcal{G}$ by removing the last row and the last column, and we obtain $\det(\mathcal{G}^b) = (-1)^n \det(\mathcal{G}^\diamond)$.

- (ii) For $n + 1 - r = 2$, only the case $\phi, \chi = \{2, n\}$ needs to be considered and it yields $\lambda(\lambda - 1) \det \mathcal{G}^c$, where $\mathcal{G}^c$ is a $(n-1) \times (n-1)$ matrix obtained from $\mathcal{G}$ by removing the first and last rows and the first and last columns.

For $n > 1$, we thus obtain

$$\det(\mathcal{G}_\lambda) = \lambda \det(\mathcal{G}^\diamond) + (-1)^n(\lambda - 1) \det(\mathcal{G}^\diamond) + \lambda(\lambda - 1) \det \mathcal{G}^c, \quad (93)$$

while the case $n = 1$ leads to $\det(\mathcal{G}_\lambda) = \lambda^2 - \lambda + \frac{1}{2} \neq 0$ as $\lambda \in \mathbb{R}$.

Letting $\lambda = 1$ in (93) yields $\det(\mathcal{G}_1) = \det(\mathcal{G}^\diamond)$, where $\mathcal{G}^\diamond$ is defined in (49) and $\text{rank}(\mathcal{G}^\diamond) = n$, as shown in the proof of theorem 3. Since $\mathcal{G}^\diamond$ is full rank, the matrix $\mathcal{G}_1$ is hence invertible and because the right hand side of (91) only contains $u_{(j+1)^+}$, the case $\lambda = 1$ cannot permit the existence of a stationary erratic mode. □

Remark 1. *When $n = 2p$, $p = 1, 2, 3, ...$, we have $\det \mathcal{G}^c = 0$ since $\mathcal{G}^c$ is a skew-symmetric matrix and hence, $\forall n \in \mathbb{N}$, equation (93) leads to*

$$\det(\mathcal{G}_\lambda) = (2\lambda - 1) \det(\mathcal{G}^\diamond). \tag{94}$$

In the case $n = 2p+1$, $p = 1, 2, 3, ...$, and by using a computer algebra system (Maple), we obtain $\det \mathcal{G}^c = 2 \det(\mathcal{G}^\diamond)$, at least for $n \leqslant 10$, and (93) yields

$$\det(\mathcal{G}_\lambda) = \left(\lambda^2 + (\lambda - 1)^2\right) \det(\mathcal{G}^\diamond). \tag{95}$$

It is conjectured that such a result holds $\forall n > 10$.

The $n + 1$ solutions of $P^{DG}_{n+1}(\widetilde{\omega})$ are now shown in the case $\lambda = 1$ in Figures 19 and 20, for $n = 1, ..., 4$. As for the CG and centered DG schemes we distinguish between n odd (Figure 19) and n even (Figure 20). Further, since $\widetilde{\omega}$ is now a complex number both the real (phase) and imaginary (damping) parts of $\widetilde{\omega}$ are shown. The correct computation of the roots is quite delicate due to the need to take square roots of complex trigonometric functions. Therefore, we give an example in the case $n = 1$ (and $\lambda = 1$) where (82) leads to the following pair of solutions for $\widetilde{\omega}$

$$\widetilde{\omega}_{1,2} = i\left(e^{ikh} + 2 \mp \sqrt{e^{2ikh} + 10e^{ikh} - 2}\right). \tag{96}$$

We let

$$\begin{aligned} &\alpha_0 = \cos 2kh + 10 \cos kh - 2, && \beta_0 = \sin 2kh + 10 \sin kh, \\ &\alpha = \tfrac{1}{\sqrt{2}}\sqrt{\alpha_0 + \sqrt{\alpha_0^2 + \beta_0^2}}, && \beta = \tfrac{1}{\sqrt{2}}\sqrt{-\alpha_0 + \sqrt{\alpha_0^2 + \beta_0^2}}, \end{aligned}$$

and we obtain

$$\sqrt{e^{2ikh} + 10e^{ikh} - 2} \equiv z \;\; = \;\; \begin{cases} \alpha + i\beta & \text{if} \quad kh \in [-2\pi, -\pi] \cup [0, \pi], \\ \alpha - i\beta & \text{if} \quad kh \in [-\pi, 0] \cup [\pi, 2\pi], \end{cases} \tag{97}$$

since the product of the real and imaginary parts of z has the same sign as β_0. Note that if $-z$ is considered in the left hand side of (97) instead, $\widetilde{\omega}_1$ and $\widetilde{\omega}_2$ are simply exchanged in (96). We then rewrite $\widetilde{\omega}_{1,2}$ in the form

$$\widetilde{\omega}_{1,2} \;\; = \;\; \begin{cases} -\sin kh \pm \beta + i\big(\cos kh + 2 \mp \alpha\big) & \text{if} \quad kh \in [-2\pi, -\pi] \cup [0, \pi], \\ -\sin kh \mp \beta + i\big(\cos kh + 2 \mp \alpha\big) & \text{if} \quad kh \in [-\pi, 0] \cup [\pi, 2\pi]. \end{cases}$$

In Figure 19 the real (phase) and imaginary (damping) parts of $\widetilde{\omega}_1$ and $\widetilde{\omega}_2$ are shown in blue ($\widetilde{\omega}_1$) and red ($\widetilde{\omega}_2$), respectively, on $[-2\pi, 2\pi]$. Both quantities coincide at $kh = \pi$ and $kh = -\pi$ with the numerical value $\pm\sqrt{11} + i$, and hence, there is no jump in the representation of the frequency.

As in the CG and centered DG cases, each solution is given its own color, and at a given spatial wavenumber kh there are $n+1$ solutions but only one of these solutions corresponds to the physical solution. The remaining solutions are mathematical artifacts that arise from symmetries in the Fourier analysis. Again, each solution is valid over only a limited wavenumber range or branch and the union of all such branches gives the complete dispersion relationship $\widetilde{\omega}_S$. The solution $\widetilde{\omega}_S$ can be constructed from $P^{CG}_{n+1}(\widetilde{\omega})$ using a branch selection procedure based on the reconstructed solutions and eigenvectors. However, the resulting dispersion relationship and determination of branches is structurally quite different than found in the CG and centered DG cases. We have illustrated the real part of the reconstructed solutions for $n = 3$ at $kh = \pi/24$ in Figure 22. Contrary to what happens for the CG and centered DG cases, the two solutions that cross at $kh = 0$ for odd n do not exchange roles for $kh = -\pi/24$, instead we obtain identical reconstructions. Note that the reconstructed solution is unique only up to multiplication by a complex constant. This explains the shift in phase seen compared to Figure 10 for the CG case.

The real and imaginary parts of $P^{DG}_{n+1}(\widetilde{\omega})$ and $\widetilde{\omega}_S$ are found in Figures 19 - 21 for $n = 1, 2, 3, 4, 5, 7, 10$. Unlike the CG and DG centered cases, there are no spectral gaps for any of these n, and over $kh \in [0, \frac{(n+1)\pi}{n}]$ the $(n+1)$ branches are connecting at $kh = \frac{p\pi}{n}$, $p = 1, 2, 3, ..., n$. Additionally, there is now damping since $\widetilde{\omega}$ is complex. As n increases the damping becomes increasingly scale-selective and localized to the high-frequency part of the spectrum, and also increases in strength. Although the phase of the mode at $kh = \frac{n+1}{n}\pi$ is equal to zero, the damping is not and therefore this is not a erratic stationary mode, unlike the CG and DG centered cases. However the maximal frequency (phase) still shows a strong increase as n increases.

5. Concluding remarks

In this paper we have studied continuous and discontinuous Galerkin discretizations of the first order linear wave equation in 1D, when approximation spaces of polynomial functions of arbitrary degree n are considered.

These include both centered and upwind fluxes for the DG case. This study was performed by using Fourier analysis, which is powerful but can give rise to many mathematical artifacts. These artifacts were removed through a careful branch selection procedure driven by analysis of eigenvectors and associated reconstructed solutions. This procedure was validated by comparing the computed dispersion relationship to numerical eigenvalues of the linear operator for CG with $n = 1$ and $n = 2$, where a perfect match was found.

In line with previous results, CG and centered DG were found to have spectral gaps and an erratic/spurious stationary mode. For the first time, the existence of gaps has been characterized analytically and their specific locations computed. Additionally, the order of convergence for CG (centered DG) was confirmed to be $O(h^{2n+1})$ for n even (odd) and $O(h^{2n+3})$ for n odd (even). Conversely, upwind DG was shown to have neither spectral gaps or an erratic stationary mode, with an order of convergence equal to $O(h^{2n+3})$ for all n. Both CG and DG were shown to have increasing maximal frequency with higher n. This is a possible explanation of the observation of an increasingly restrictive CFL limit as n increases.

This paper is the first step towards a detailed study of the dispersion properties of DG methods for the linear shallow water equations in 2D for both triangles and quadrilaterals, which will be reported in forthcoming papers.

In this paper only two possibilities for DG flux were studied: centered and upwind. For more complicated equations, such as the nonlinear shallow water equations in 2D or the compressible Euler equations in 3D more sophisticated fluxes such as PVM [5] must be employed. However, the limit cases of pure upwind and pure centered are still useful as a benchmark for understanding the linearized behavior.

6. Acknowledgements

The first and third authors gratefully acknowledge the support of the Isaac Newton Institute for Mathematical Sciences, Cambridge, UK, during the programme on Multiscale Numerics for the Atmosphere and Ocean in which this work was initiated. It is with pleasure and gratitude that the authors acknowledge a number of stimulating discussions with Mark Ainsworth. Christopher Eldred was supported by the French National Research Agency through contract ANR-14-CE23-0010 (HEAT).

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a

wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energys National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

7. References

[1] M. Ainsworth, Dispersive and dissipative behaviour of high order discontinuous Galerkin finite element methods, J. Comput. Phys. 198 (2004) 106-130. https://doi.org/10.1016/j.jcp.2004.01.004

[2] M. Ainsworth, Discrete dispersion relation for hp-version finite element approximation at high wave number, SIAM J. Numer. Anal. 42(2) (2004) 553575. https://doi.org/10.1137/S0036142903423460

[3] M. Ainsworth, P. Monk, W. Muniz, Dispersive and dissipative properties of discontinuous Galerkin finite element methods for the second-order wave equation, Journal of Scientific Computing. 27 (2006) 5-40. https://doi.org/10.1007/s10915-005-9044-x

[4] M. Ainsworth, Dispersive behaviour of high order finite element schemes for the one-way wave equation, J. Comput. Phys. 259 (2014) 1-10. https://doi.org/10.1016/j.jcp.2013.11.003

[5] M.J. Castro-Díaz, E.D. Fernández-Nieto, A class of computationally fast first order finite volume solvers: PVM methods, SIAM J. Sci. Comput. 34 (2012) A2173-A2196.

[6] C. Eldred, D.Y. Le Roux, Dispersion analysis of compatible Galerkin schemes for the 1d shallow water model, J. Comput. Phys. 371 (2018) 779 - 800. DOI:10.1016/j.jcp.2018.06.007

[7] C. Eldred, D.Y. Le Roux, Dispersion analysis of compatible Galerkin schemes on quadrilaterals for shallow water model, J. Comput. Phys. 387 (2019) 539-568. DOI:10.1016/j.jcp.2019.02.009

[8] A. Ern, J.L. Guermond, Theory and practice of finite elements, Sringer, 2004.

[9] W. Guo, X. Zhong, J.-M. Qiu, Superconvergence of discontinuous Galerkin and local discontinuous Galerkin methods: Eigen-structure analysis based on Fourier approach, J. Comput. Phys. 235 (2013) 458-485. https://doi.org/10.1016/j.jcp.2012.10.020

[10] J.S. Hesthaven, T. Warburton, Insight through theory. In: Nodal Discontinuous Galerkin Methods. Texts in Applied Mathematics, vol 54. Springer, New York, NY (2008). https://link.springer.com/chapter/10.1007%2F978-0-387-72067-8_4

[11] F.Q. Hu, M.Y. Hussaini, P. Rasetarinera, An analysis of the discontinuous Galerkin method for wave propagation problems, J. Comput. Phys. 151 (1999) 921-946. https://doi.org/10.1006/jcph.1999.6227

[12] F.Q. Hu, H. L. Atkins, Eigensolution analysis of the discontinuous Galerkin method with nonuniform grids, J. Comput. Phys. 182 (2002) 516-545. https://doi.org/10.1006/jcph.2002.7184

[13] L. Krivodonova, R. Qin, An analysis of the spectrum of the discontinuous Galerkin method, Applied Numerical Mathematics 64 (2013) 1-18. https://doi.org/10.1016/j.apnum.2012.07.008

[14] P.H. LeBlond and L.A. Mysak, Waves in the Ocean, Elsevier, Amsterdam, 1978.

[15] D.Y. Le Roux, G.F. Carey, Stability/dispersion analysis of the discontinuous Galerkin linearized shallow-water system, Int. J. Numer. Meth. Fluid. 48 (2005) 325-347. https://doi.org/10.1002/fld.893

[16] M. Marcus, Determinants of Sums, The College Mathematics Journal, vol. 21, no. 2, 1990, pp. 130-135. doi:10.2307/2686755

[17] T. Melvin, A. Staniforth, J. Thuburn, Dispersion analysis of the spectral element method, Q. J.R. Meteorol. Soc. 138 (2012) 1934-1947. DOI:10.1002/qj.1906

[18] G. Mengaldo, R.C. Moura, B. Giralda, J. Peiró, S.J. Shervin, Spatial eigensolution analysis of discontinuous Galerkin schemes with practical insights for under-resolved computations and implicit LES, Computers & Fluids, 169 (2018) 349-364. https://doi.org/10.1016/j.compfluid.2017.09.016

[19] A.H. Schatz, I.H. Sloan, L.B. Wahlbin, Superconvergence in finite element methods and meshes that are locally symmetric with respect to a point, SIAM J. Numer. Anal. 33(2) (1996) 505521. http://www.jstor.org/stable/2158385

[20] S. Sherwin, Dispersion analysis of the continuous and discontinuous Galerkin formulations. In Discontinuous Galerkin Methods: Theory, Computation and Applications, B. Cockburn, G.E. Karniadakis, and C.W. Shu (Eds.), Lecture Notes in Computational Science and Engineering 11, Springer-Verlag, Berlin, 1999, pp. 425-431.

[21] J. Strikwerda, Finite Difference Schemes and Partial Differential Equations, second ed., SIAM, 2004.

[22] P.A. Ullrich, D.R. Reynolds, J.E. Guerra, M.A. Taylor, Impact and importance of hyperdiffusion on the spectral element method: A linear dispersion analysis. J. Comput. Phys. 375 (2018) 427-446. DOI:10.1016/j.jcp.2018.06.035

[23] R.A. Walters, G.F. Carey, Analysis of spurious oscillation modes for the shallow water and Navier-Stokes equations, Computers and Fluids 11 (1983) 51–68.

[24] X. Zhong, C.-W. Shu, Numerical resolution of discontinuous Galerkin methods for time dependent wave equations, Comput. Methods Appl. Mech. Engrg. 200 (2011) 2814-2827. DOI:10.1016/j.cma.2011.05.010

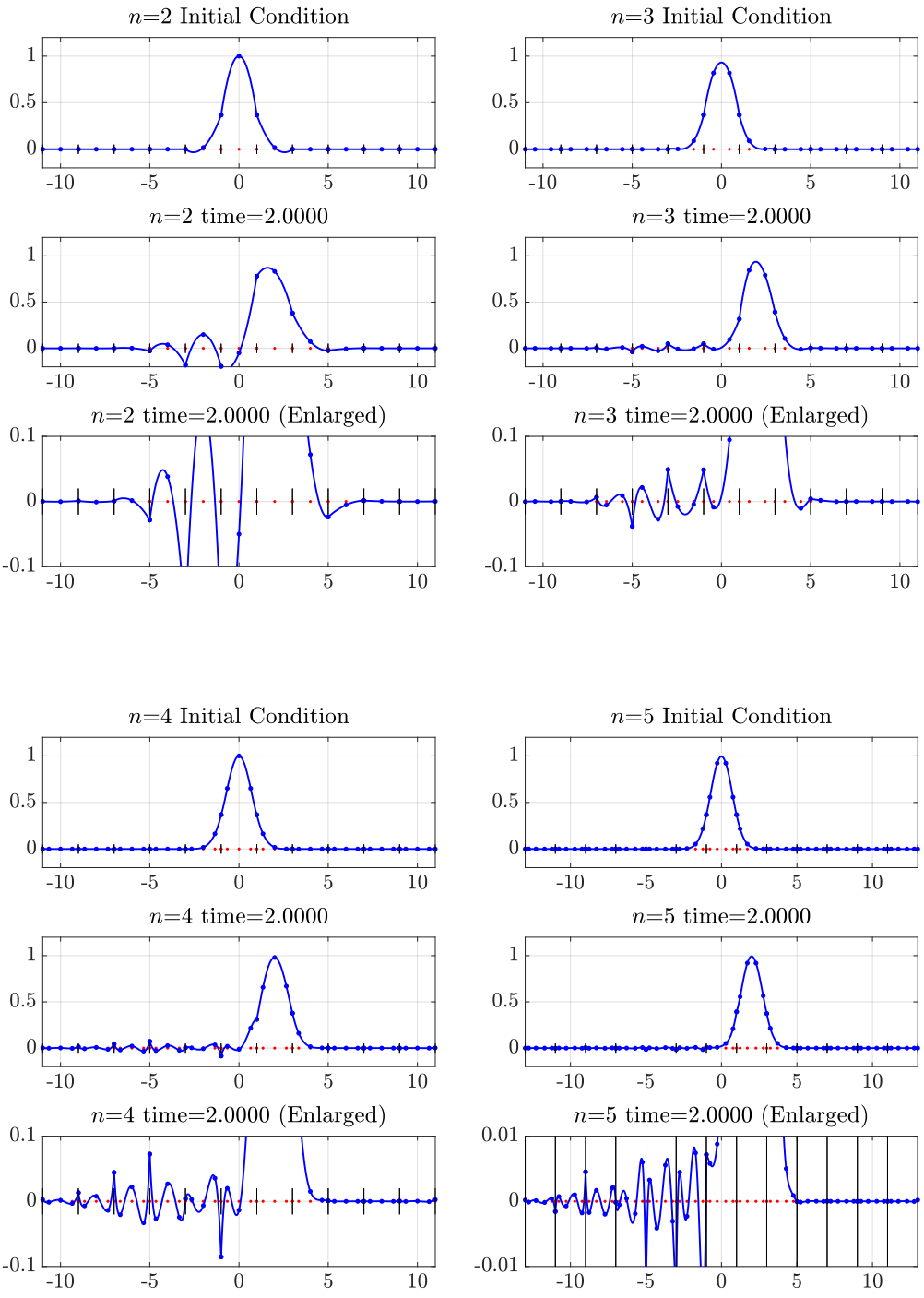

Figure 6: Propagation of an initial Gaussian at time 2, on a mesh of 11 elements of width 2 with $n = 2, 3, 4, 5$. In all cases there is an anomalous wave packet with a predicted group velocity of $-(2n + 1)$ clearly visible in the enlarged figure.

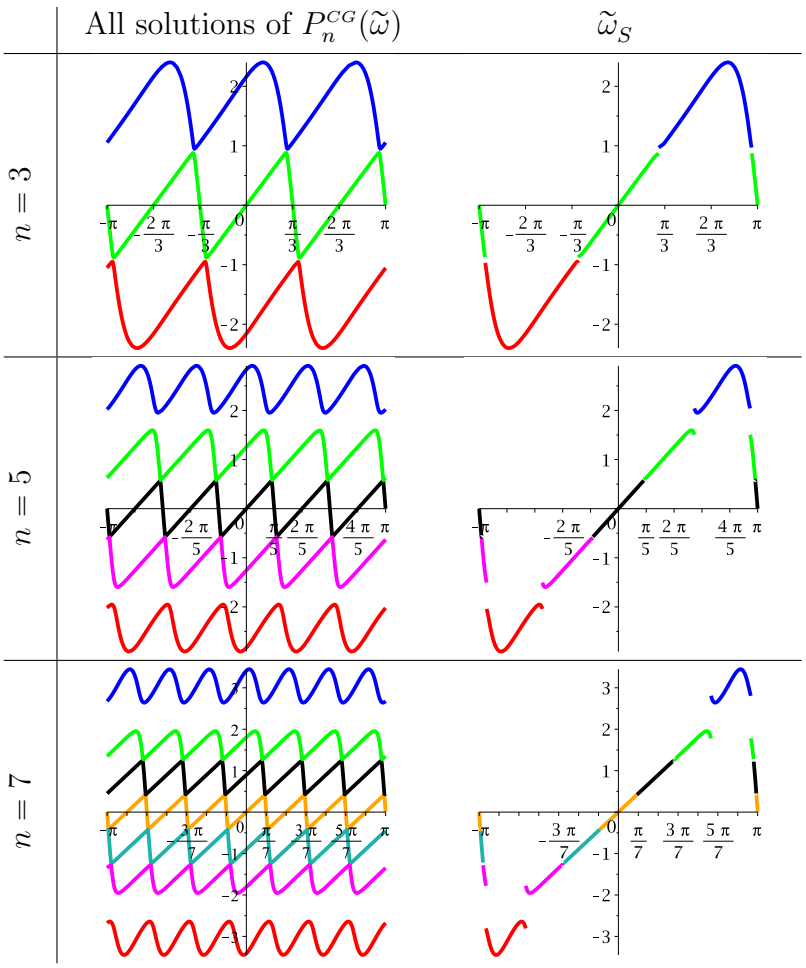

Figure 7: Phase $\widetilde{\omega} = \omega h/c$ for the CG scheme in the case $n = 3, 5$ and $n = 7$. For a given n, the left column corresponds to all solutions of $P_n^{CG}(\widetilde{\omega})$ (the dispersion relation), while the dispersion relationship $\widetilde{\omega}_S$ is given in the right column. Each root is colored differently on the graphics.

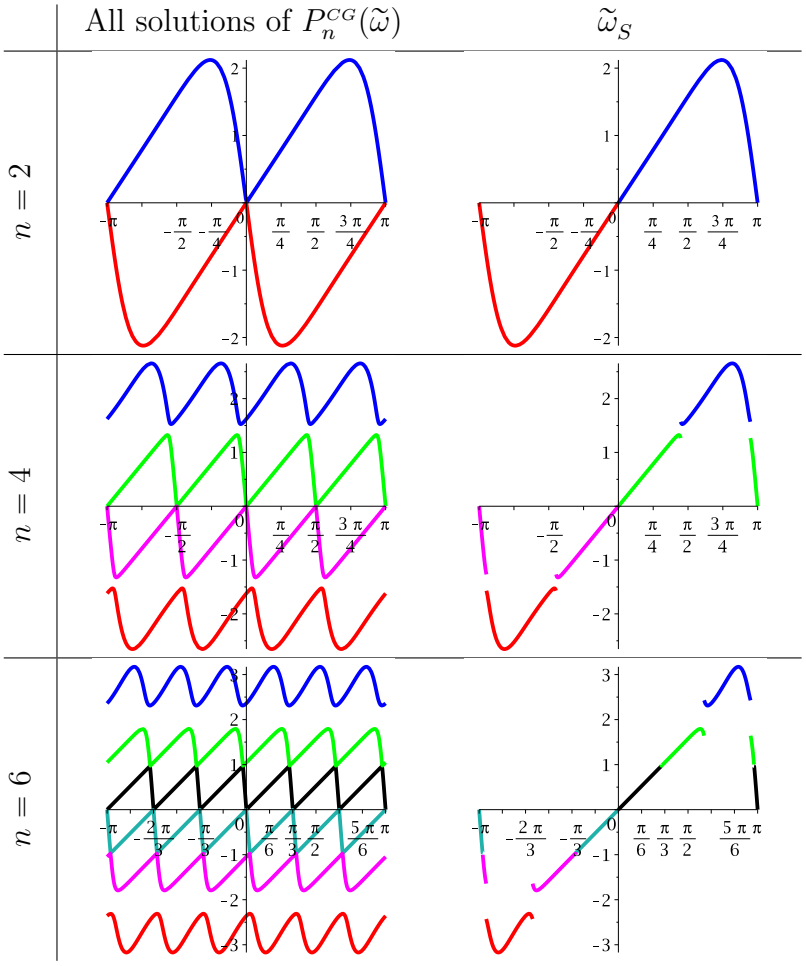

Figure 8: As for Figure 7 but in the case $n = 2, 4$ and $n = 6$.

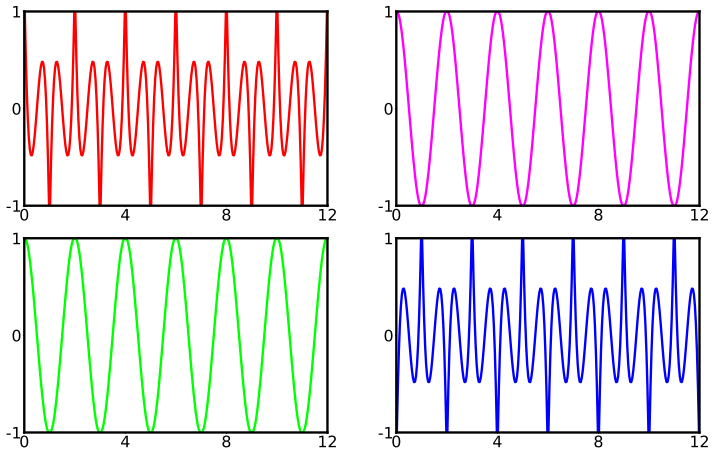

Figure 9: Reconstructed solutions for the CG method with $n = 4$ at $kh = \pi/4$. The colors correspond with the associated branches in Figure 8.

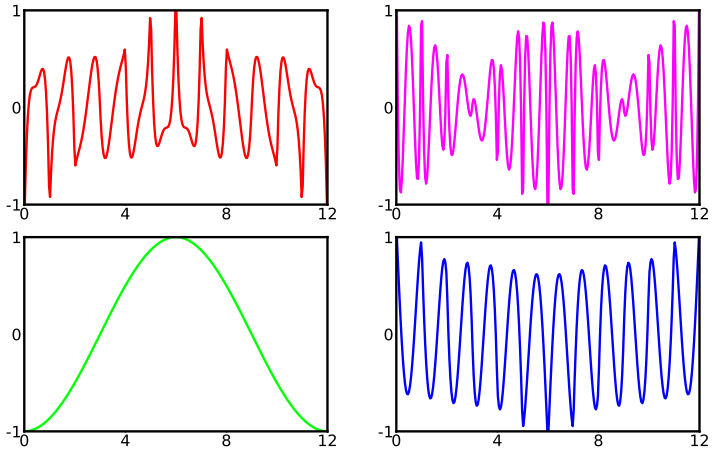

Figure 10: Reconstructed solutions for the CG method with $n = 4$ at $kh = \pi/24$. The colors correspond with the associated branches in Figure 8.

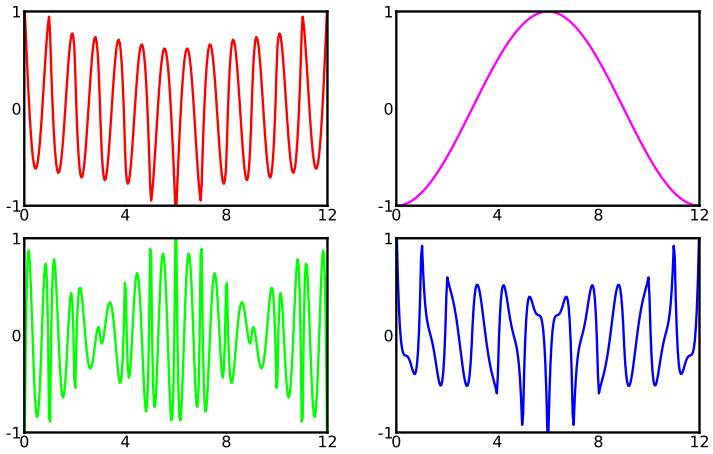

Figure 11: Reconstructed solutions for the CG method with $n = 4$ at $kh = -\pi/24$. The colors correspond with the associated branches in Figure 8.

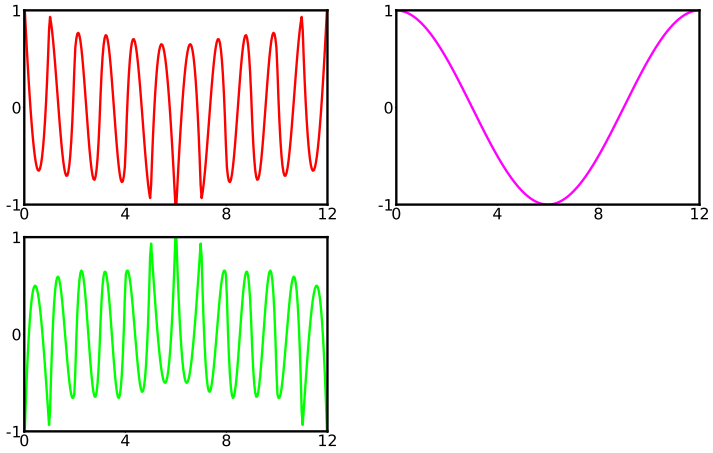

Figure 12: Reconstructed solutions for the CG method with $n = 3$ at $kh = \pi/18$. The colors correspond with the associated branches in Figure 7.

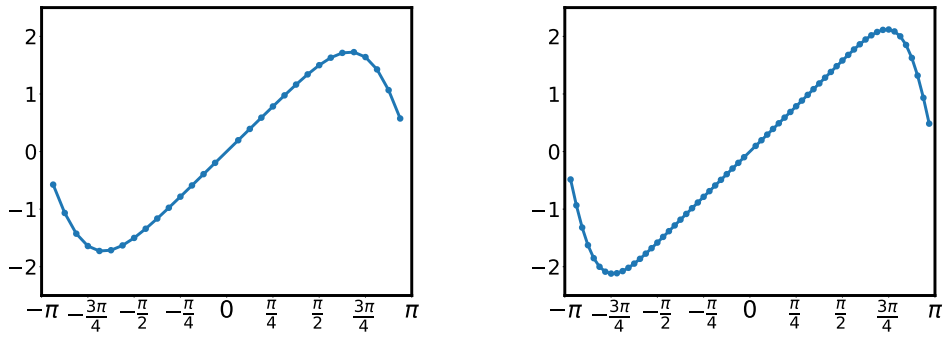

Figure 13: The numerical eigenvalues computed from a mesh of 32 elements, with $n = 1$ on the left and $n = 2$ on the right. Note the exact correspondence to Figures 4 and 8.

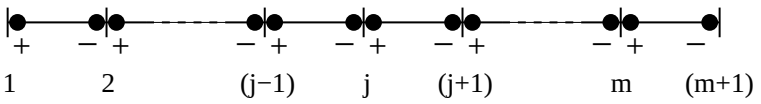

Figure 14: Node j corresponds to the coincident node pair (j^-/j^+), $j = 1, 2, 3, ..., m + 1$.

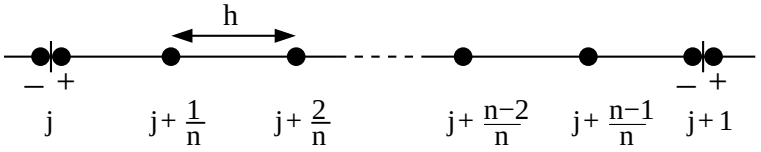

Figure 15: Indices of the local degrees of freedom on element e_j of ε_h, $j = 1, 2, ..., m$.

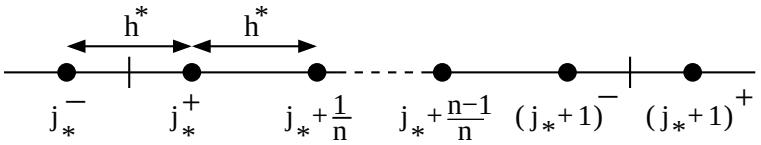

Figure 16: Definition of the meshlength parameter h by redistributing the indices of the local degrees of freedom of Figure 15 on a regular and uniform grid, and using j_* instead of j.

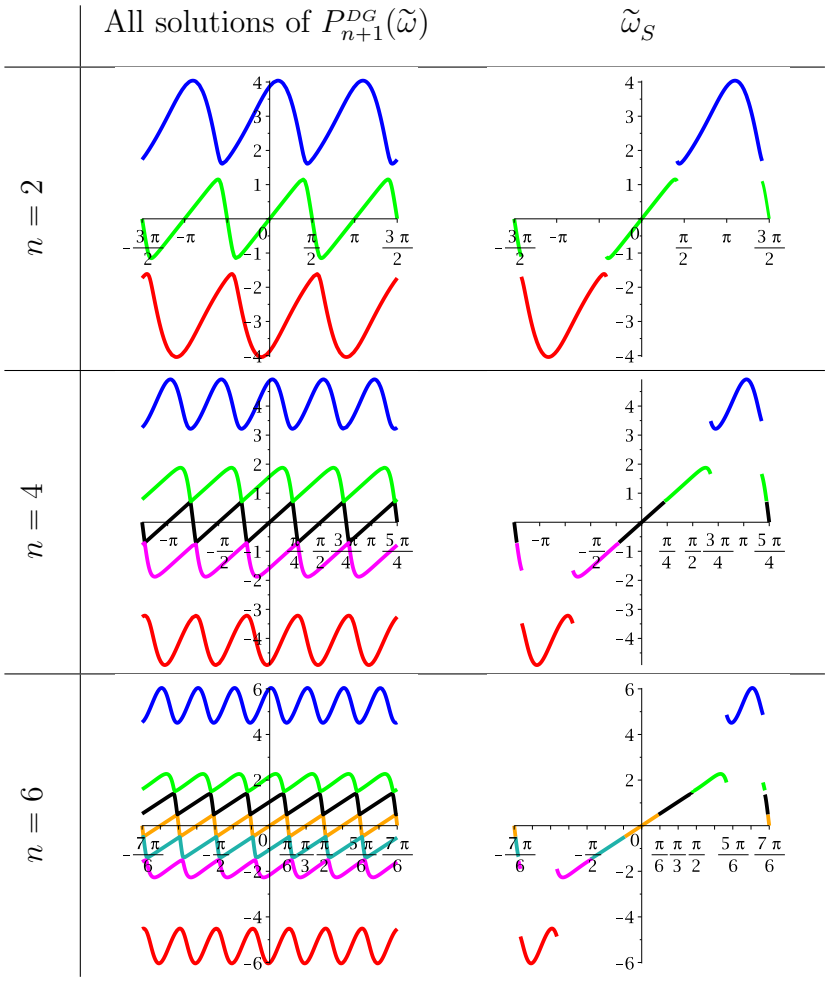

Figure 17: Phase $\widetilde{\omega} = \omega h/c$ for the DG centered scheme ($\lambda = 1/2$) in the case $n = 2, 4, 6, 8$. For a given n, the left column corresponds to all solutions of $P_{n+1}^{DG}(\widetilde{\omega})$ (the dispersion relation), while the dispersion relationship $\widetilde{\omega}_S$ is given in the right column. Each root is colored differently on the graphics.

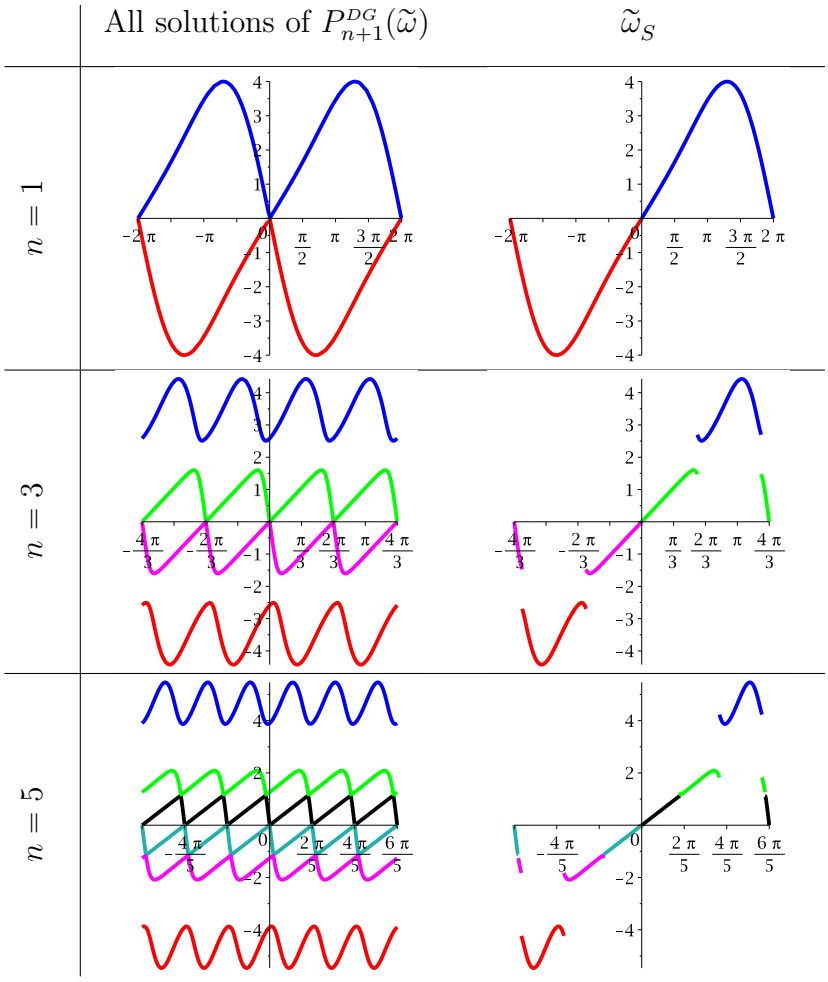

Figure 18: As for Figure 17 but in the case $n = 1, 3, 5$.

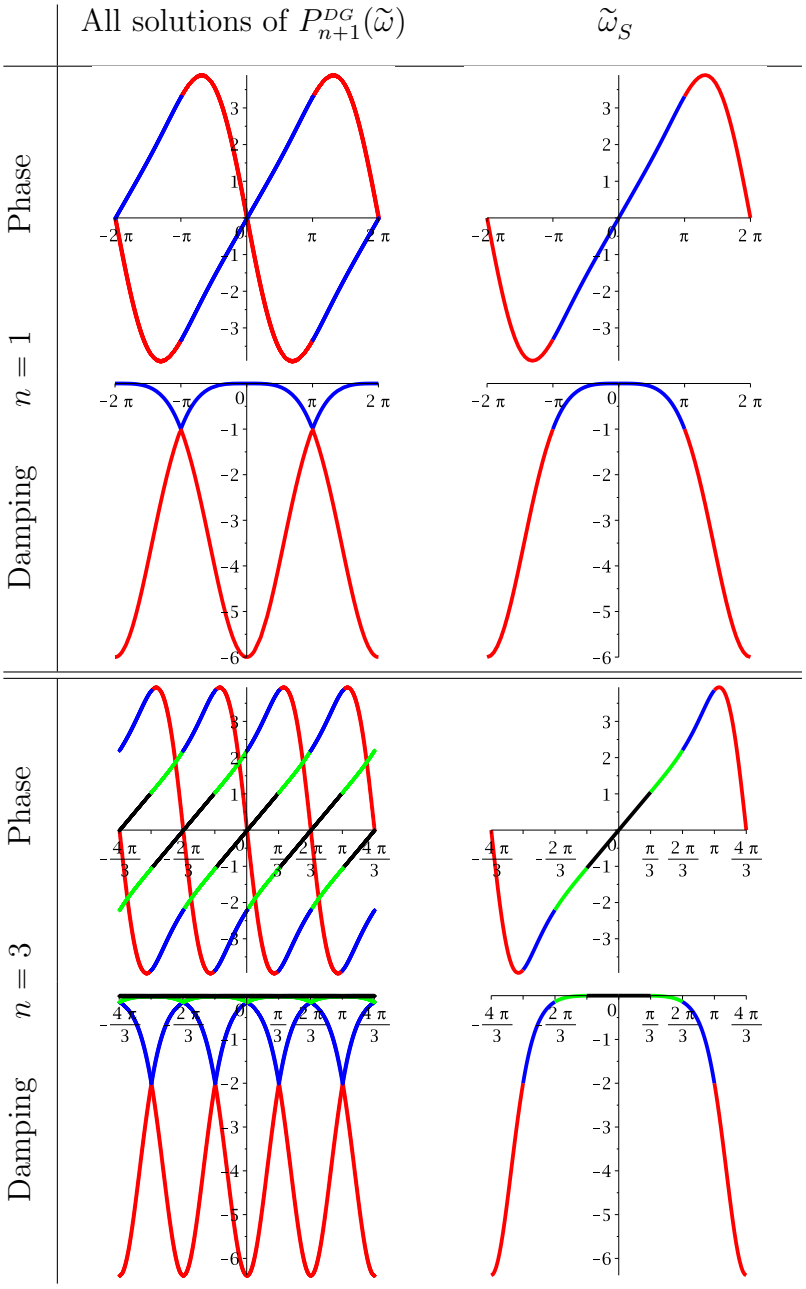

Figure 19: Phase $\Re(\widetilde{\omega})$ and damping $-\Im(\widetilde{\omega})$ of $\widetilde{\omega} = \omega h/c$ for the DG upwind scheme ($\lambda = 1$) in the case $n = 1$ and $n = 3$. For a given n, the left column corresponds (for both phase and damping) to all solutions of $P_{n+1}^{DG}(\widetilde{\omega})$ (the dispersion relation), while the dispersion relationship $\widetilde{\omega}_S$ is given in the middle column. Note that each root is colored differently on the graphics.

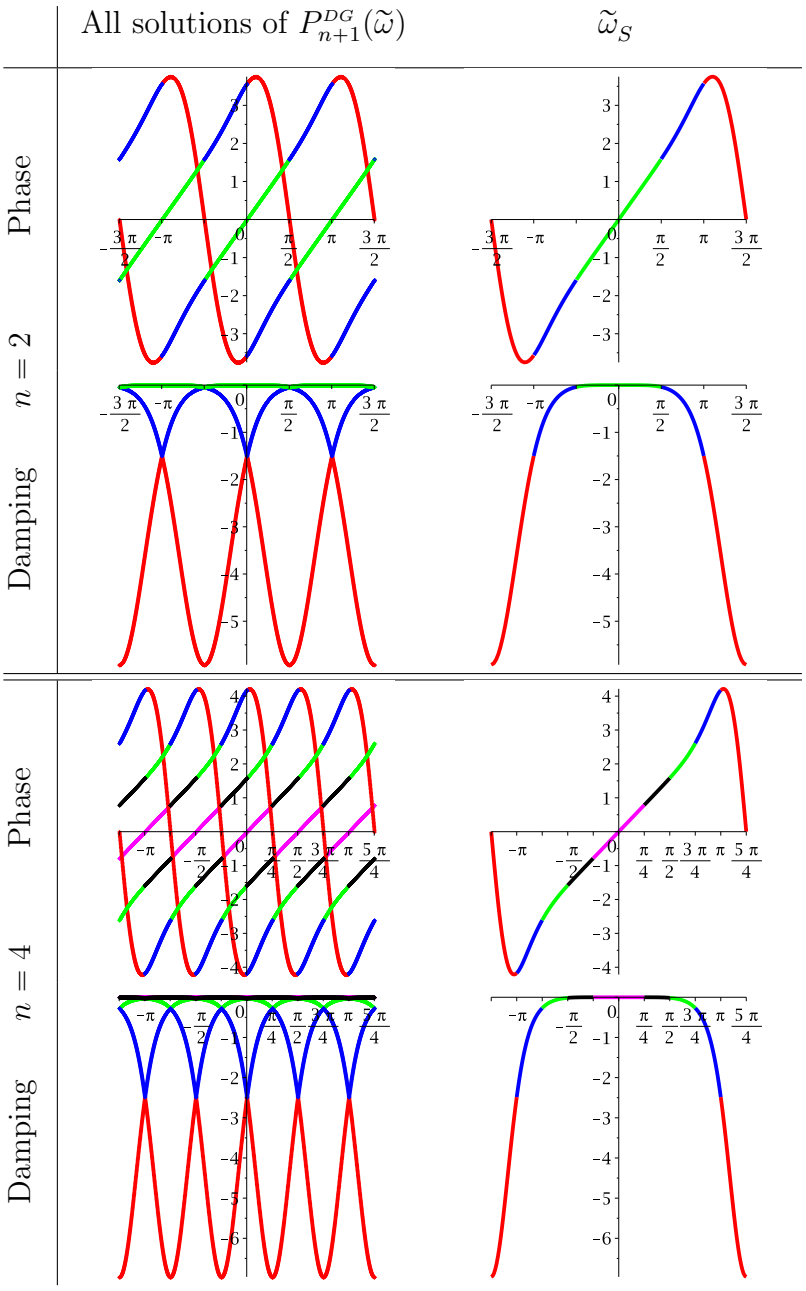

Figure 20: As for Figure 19 but in the case $n = 2$ and $n = 4$.

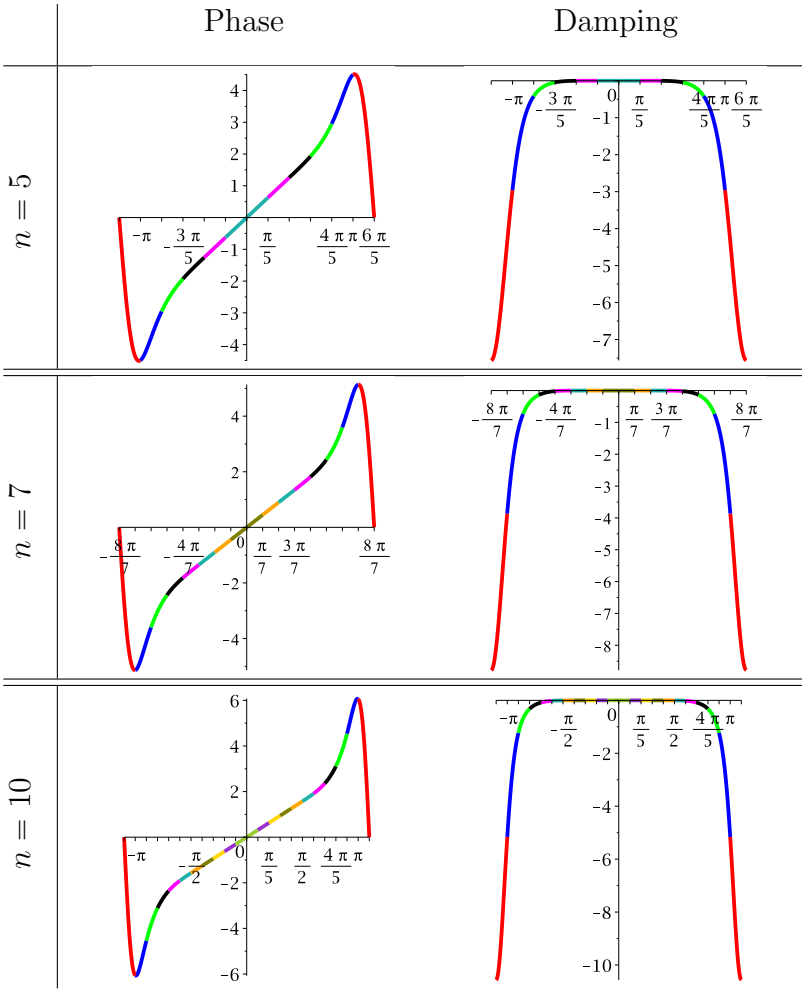

Figure 21: As for Figure 19 but for $\widetilde{\omega}_S$ in the case $n = 5, 7$ and $n = 10$.

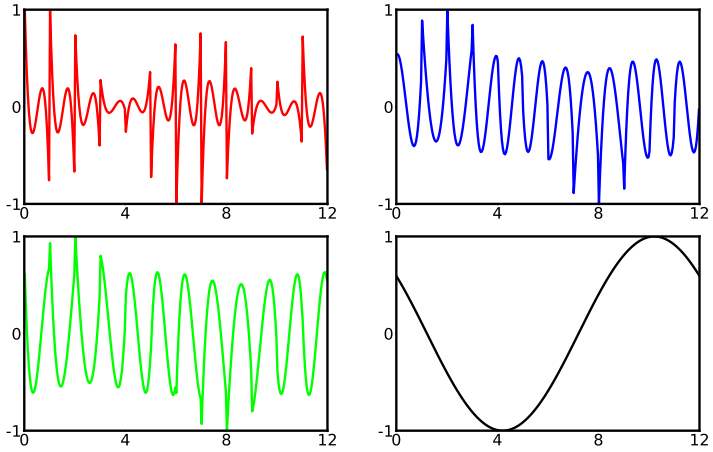

Figure 22: Reconstructed solution for DG upwind with $n = 3$ at $kh = \pi/24$. The same solution is found at $kh = -\pi/24$.