Discrete Dispersion Relation for *hp*-Version Finite Element Approximation at High Wave Number

Mark Ainsworth*

February 24, 2003

Abstract

The dispersive properties of high order finite element schemes are analysed in the setting of the Helmholtz equation, and an explicit form the discrete dispersion relation is obtained for elements of arbitrary order. It is shown that the numerical dispersion displays three different types of behaviour depending on the size of the order of the method relative to the mesh-size and the wave number. Quantitative estimates are obtained for the behaviour and rates of decay of the dispersion error in the differing regimes. All estimates are fully explicit and are shown to be sharp. Limits are obtained for where transitions between the different regimes occurs, and used to provide guidelines for the selection of the mesh-size and the polynomial order in terms of the wave number so that the dispersion error is controlled.

Key words Discrete dispersion relation. High wave number. *hp*-finite element method.

AMS subject classifications Primary 65N50. 65N15, 65N30, 35A40, 35J05.

1 Introduction

Wave propagation phenomena arising in practical applications typically require large wave number (or frequency) ω. Accurate numerical simulation of such applications is thwarted by a number of issues, perhaps the most acute of which is numerical dispersion. This refers to the effect whereby the numerical scheme fails to propagate waves at the correct speed resulting a phase-lead or lag in the numerical approximation. Numerical dispersion is often responsible not only for poor resolution but approximations that are even qualitatively incorrect.

Finite elements are often the method of choice for engineer's interested in problems of continuum mechanics posed over complicated domains. It is therefore not surprising that finite elements are frequently used in numerical

*Mathematics Department, Strathclyde University, 26 Richmond Street, Glasgow G1 1XH, Scotland. M.Ainsworth@strath.ac.uk. The work of this author was supported in part by the Engineering and Physical Sciences Research Council of Great Britain under grant GR/M59426. This work was completed while the author was visiting the Newton Institute for Mathematical Sciences, Cambridge, UK.

wave propagation [3]. The importance of numerical dispersion is widely recognised, and is often used in assessing the quality of a numerical scheme and as a basis for ranking different finite element methods. For instance, Harari and co-workers [16–19] consider the use of stabilised and Galerkin least squares finite element formulations for combatting the problem of dispersion in the solution of acoustic scattering problems, typically in the context of low order finite elements. Pinsky and co-workers [1, 25] also study the discrete dispersive properties of various lower order finite element methods (such as the 8-node trilinear element, 20-node serendipity, 27-node tri-quadratic element) for the approximation of the scalar wave equation in three dimensions. More recently, Christon [6] considered the dispersive behaviour of a variety of finite element schemes for the second-order wave equation and performed a computational study of the discrete phase and group velocities.

Advantages of using higher order elements have also been widely recognised. For example, Harari and Avraham [18] compare the efficiency of first and second order elements for the solution of acoustic scattering problems, and view their work as justifying the extension of the ideas to higher order (p-version) finite elements. Thompson and Pinsky [29] study the dispersive and attenuation properties of finite elements up to fifth order for the one dimensional scalar Helmholtz equation, and on the basis of numerical evidence, conjecture that elements of degree p provide a $2p$-th order accurate approximation of the dispersion relation in the limit ωh tends to zero.

Applications are not confined to applications in acoustic scattering, for instance, Dyson [12] proposes the use of high (up to fifteenth) order schemes for propagation of waves for Euler equations. Monk and Cohen [7–9, 24] have considered the dispersive behaviour of lower order finite element methods for Maxwell's equations. The use of high order finite element and spectral element schemes for the approximation of Maxwell's equations has attracted much interest [10, 20].

The first systematic study of the properties of finite element methods for high wave number applications was carried out in a series of papers by Babuška and Ihlenburg. In [4, 21], the convergence properties in H^1-norm of first order finite elements for the one dimensional model Helmholtz problem are studied working under the assumption that $\omega h < 1$, and it is show that the pollution error is of the same order as the dispersion error. These ideas are extended to higher dimensions by Deraemaeker et. al. [11] who undertake a numerical study of the dispersive behaviour for various finite element formulations in higher dimensions that allow one to include topological effects of the meshes, while Gerdes and Ihlenburg [14] study convergence of an h-version Galerkin finite element method for a 3D-problem of rigid scattering with mesh refinement in the radial direction, and show that the error bound contains pollution effects similar to those observed in the one dimensional analysis. A detailed study of the dispersion and approximation behaviour of hp-finite elements for the Helmholtz equation in one dimension is undertaken by Ihlenburg and Babuška [22].

Despite extensive investigations, several important issues concerning the dispersive properties of standard finite element schemes remain unresolved, particularly in the context of high order elements. The aim of the present work

is to give a sharp analysis of the dispersive properties of high order finite element schemes in the setting of the Helmholtz equation, to identify thresholds (relating the order p of the method and the mesh-size h to the wave number ω) where the dispersion error begins to decay, and to obtain sharp quantitative estimates on the rates of decay of the error in the differing regimes. A clear understanding of the dispersive properties of a scheme is not only of academic interest. Accurate quantitative information on the dispersion effects can serve as a practical guideline for the construction of a mesh and a polynomial order that will lead to a reasonable first approximation.

The analysis hinges on knowledge of an explicit form for the discrete dispersion relation valid for elements of arbitrary order. This is obviously a valuable tool for the study of numerical dispersion, and, to the best of our knowledge, has not been obtained before. We derive a neat closed form expression for the discrete dispersion relation for elements of arbitrary order in terms of Padé approximants.

We study the behaviour of the error in the discrete dispersion relation in two important limits: (i) the small wave number limit where $\omega h \ll 1$ limit, where we provide a confirmation of the conjecture of Thompson and Pinsky [29]; and (ii), in the important practical case of high wave number, where $\omega h \gg 1$. The analysis provides a practical guideline for choosing the order p of the elements and the mesh-size h in order that the dispersion error is properly controlled:

$$p + \frac{1}{2} > \frac{\omega h}{2} + C(\omega h)^{1/3}. \tag{1}$$

In fact, when p is increased in this regime, the error decays at a super-exponential rate, and an explicit expression for the error is given. In the limit where p is much larger than $\omega h/2$, the expression decays as $(\omega he/2(2p+1))^{2p+1}$, which is compatible with the upper bounds derived by Ihlenburg and Babuška [22]. More importantly, criterion (1) is shown to be sharp in the sense that if p does not satisfy (1), i.e. if

$$p + \frac{1}{2} < \frac{\omega h}{2} - o(\omega h)^{1/3},$$

then the dispersion error will not decay, and may even increase significantly as the order p is increased. These results show that there is essentially no pre-asymptotic error reduction which one might have expected based on the analysis for the positive definite case [28]. Strictly speaking, the error does begin to decay when p enters the transition zone

$$p + \frac{1}{2} \in \left(\frac{\omega h}{2} - o(\omega h)^{1/3}, \frac{\omega h}{2} + o(\omega h)^{1/3} \right).$$

However, it is shown that the decay in this phase is only algebraic: $\mathcal{O}(p^{-1/3})$, and the comparatively narrow transition zone means that the pre-asymptotic decay is too short-lived to be of any real practical significance.

The results obtained here improve on the upper bounds given in [22]. In particular, all estimates are given explicitly and do not involve generic constants. This enables us to show that these estimates are the best ones possible.

Our analysis is restricted to schemes of uniform order on tensor product meshes. Nevertheless, this type of scheme is used locally in regions remote from the scatterer where the main issue is to control numerical dispersion. Equally well, our analysis only deals with the issue of numerical dispersion. The actual accuracy of the approximation is a separate issue which is considered by Ihlenburg and Babuška [22], where it is shown that the error in the H^1-norm will be properly controlled only if $\omega^2 h$ is sufficiently small.

The remainder of this article is organised as follows. Firstly, we review the standard framework leading to the derivation of the discrete dispersion relation in the setting of the wave equation in one dimension, and describe the relevance of this to the multi-dimensional case where tensor product meshes are used. The main results are outlined in the following section. The remaining sections deal with the technical details and proofs of the results.

2 The Discrete Dispersion Relation

2.1 The Setting

It is well-known that the general solution of the homogeneous wave equation in one space dimension

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 0,$$

can be expressed as a superposition of plane waves in the form

$$u(x,t) = \int_{\mathbb{R}} \left[a(k) e^{i(kx+\omega t)} + b(k) e^{i(kx-\omega t)} \right] \mathrm{d}k$$

for suitable functions a and b, where ω and k are related by the *dispersion relation*

$$\omega^2 = k^2.$$

Suppose that a uniform grid of size $h > 0$ is placed on the real line with nodes located at $h\mathbb{Z}$, and let V_h denote the set of continuous piecewise linear functions relative to the grid. By analogy with the continuous problem, we may seek solutions of the form

$$u_h(x,t) = e^{i\omega t} U_h(x) \tag{2}$$

so that $U_h \in V_h$ must satisfy

$$B_\omega(U_h, v_h) = 0 \text{ for all } v_h \in V_h \tag{3}$$

where

$$B_\omega(U_h, v_h) = (U_h', v_h') - \omega^2 (U_h, v_h)$$

and $(\cdot,\cdot)$ denotes the L_2-inner product on $\mathbb{R}$.

The invariance of the grid under translation by h prompts use to seek Bloch wave [26] solutions of the homogeneous eq. (3) in the form

$$U_h(x) = \alpha \sum_{m\in\mathbb{Z}} e^{imkh} \theta_m(x) \tag{4}$$

where α and k are constants to be determined. Here, θ_m are the usual piecewise linear hat functions defined by

$$\theta_m(nh) = \delta_{mn}, \quad m, n \in \mathbb{Z}. \tag{5}$$

The translation invariance of the grid means that

$$\theta_m(x + nh) = \theta_{m-n}(x), \quad x \in \mathbb{R}, \quad n \in \mathbb{Z},$$

which in turn implies that U_h has the characteristic property of a Bloch wave: for each $n \in \mathbb{Z}$,

$$U_h(x + nh) = e^{iknh} U_h(x), \quad x \in \mathbb{R}. \tag{6}$$

This means that (3) is equivalent to the condition

$$B_\omega(U_h, \theta_0) = 0,$$

or, inserting the expression (4) for U_h,

$$\alpha \sum_{m\in\mathbb{Z}} e^{imkh} B_\omega(\theta_m, \theta_0) = 0.$$

Hence, a non-trivial Bloch wave exists provided that

$$\sum_{m\in\mathbb{Z}} e^{imkh} B_\omega(\theta_m, \theta_0) = 0.$$

This expression may be simplified further using properties of the hat functions giving

$$e^{-ikh} B_\omega(\theta_{-1}, \theta_0) + B_\omega(\theta_0, \theta_0) + e^{ikh} B_\omega(\theta_1, \theta_0) = 0,$$

and, again by exploiting translation invariance of the grid, we obtain the *discrete dispersion relation*

$$2\cos(kh) B_\omega(\theta_1, \theta_0) + B_\omega(\theta_0, \theta_0) = 0. \tag{7}$$

We refer to eq. (7) as the dispersion relation for the following reason. For the first order standard Galerkin scheme considered above, we find that

$$B_\omega(\theta_0, \theta_0) = \frac{2}{h}\left(1 - \frac{1}{3}\omega^2 h^2\right); \quad B_\omega(\theta_1, \theta_0) = -\frac{1}{h}\left(1 + \frac{1}{6}\omega^2 h^2\right)$$

and the discrete dispersion relation simplifies to give the well-known result

$$kh = \cos^{-1}\left(\frac{6 - 2\omega^2 h^2}{6 + \omega^2 h^2}\right) = \omega h - \frac{1}{24}(\omega h)^3 + \ldots.$$

2.2 Extension to High Order Schemes

The discrete dispersion relation for higher order schemes may be obtained by modifying the previous arguments. Let V_{hp} denote the set of continuous piecewise polynomials of degree p on the grid $h\mathbb{Z}$, and let $V_{hp}^\flat$ denote the subspace

$$V_{hp}^\flat = \{v_{hp} \in V_{hp} : v_{hp}(mh) = 0, m \in \mathbb{Z}\}.$$

As before, we seek a Bloch wave solution $U_{hp} \in V_{hp}$ satisfying

$$B_\omega(U_{hp}, v_{hp}) = 0 \text{ for all } v_{hp} \in V_{hp}.$$

Accordingly, we write U_{hp} in the form

$$U_{hp}(x) = \sum_{m\in\mathbb{Z}} e^{ikmh}\left[\alpha\theta_m^{(p)}(x) + \beta\psi^{(p)}(x - mh)\right] \tag{8}$$

where α and β are constants, and $\psi^{(p)} \in V_{hp}^\flat$ is supported on $(0, h)$. By analogy with (5), the function $\theta_m^{(p)} \in V_{hp}$ has nodal values given by

$$\theta_m^{(p)}(nh) = \delta_{mn}, \quad m, n \in \mathbb{Z}, \tag{9}$$

but is instead extended to the element interiors as a polynomial of degree p by requiring that

$$B_\omega(\theta_m^{(p)}, v_{hp}) = 0 \text{ for all } v_{hp} \in V_{hp}^\flat. \tag{10}$$

If ω does not correspond to a discrete eigenvalue, then $\theta_m^{(p)}$ is uniquely defined by this condition. It is not difficult to show that the function U_{hp} satisfies the Bloch wave condition: for all $n \in \mathbb{Z}$,

$$U_{hp}(x + nh) = e^{iknh}U_{hp}(x), \quad x \in \mathbb{R}.$$

Consequently, exploiting translation invariance of the grid, it suffices to require

$$\left.\begin{aligned} B_\omega(U_{hp}, \theta_0^{(p)}) &= 0 \\ B_\omega(U_{hp}, \psi^{(p)}) &= 0. \end{aligned}\right\}$$

By inserting the expression (8) for U_{hp} into the latter equation, using (10) and the fact that ω does not correspond to a discrete eigenvalue, we conclude that $\psi^{(p)}$ must vanish identically. Consequently, the expression (8) collapses to the form considered in the case of first order schemes,

$$U_{hp}(x) = \alpha \sum_{m\in\mathbb{Z}} e^{ikmh}\theta_m^{(p)}(x), \tag{11}$$

and by analogy with (7), the higher order discrete dispersion relation assumes the form

$$2\cos(kh)B_\omega(\theta_1^{(p)}, \theta_0^{(p)}) + B_\omega(\theta_0^{(p)}, \theta_0^{(p)}) = 0. \tag{12}$$

The same expression was obtained by Ihlenburg and Babuška [4, 22]. A detailed study of the discrete dispersion relation for higher order elements is postponed until the next section.

2.3 Relevance to Multi-Dimensional Problems

Information on the discrete dispersion relation for the scalar Helmholtz equation in one dimension may be used to derive explicit forms for the discrete dispersion relation for finite element approximation of problems in higher dimensions on tensor product meshes. Here, we describe the case of the wave equation in detail. An extension of the argument to the approximation of Maxwell equations using Nédélec elements will be found in [2].

Consider the wave equation in d-dimensions,

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} - \Delta u = 0$$

and assume that a tensor product grid $h\mathbb{Z}^d$ is introduced on $\mathbb{R}^d$. Let V_{hp} denote the space of piecewise polynomials of total degree p in each variable on the grid, then seeking a discrete solution of the form (2) leads to the problem of determining $U_{hp} \in V_{hp}^{(d)}$ such that

$$\sum_{r=1}^{d} \left(\frac{\partial U_{hp}}{\partial x_r}, \frac{\partial v_{hp}}{\partial x_r} \right) - \kappa^2 (U_{hp}, v_{hp}) = 0, \text{ for all } v_{hp} \in V_{hp}^{(d)}.$$

The tensor product structure prompts us to seek a solution of the form

$$U_{hp}(x_1, \ldots, x_d) = \alpha \prod_{r=1}^{d} X_p(k_r; x_r)$$

where α is a constant, and

$$X_p(k; s) = \sum_{m \in \mathbb{Z}} e^{imkh} \phi_m^{(p)}(s) \tag{13}$$

with $\phi_m^{(p)}$ defined as above. In particular, we recall that $X_p(k_r) \in V_{hp}$ satisfies

$$(X_p'(k_r), v') - \kappa_r^2 (X_p(k_r), v) = 0 \text{ for all } v \in V_{hp} \tag{14}$$

where κ_r is related to k_r by the discrete dispersion relation (12).

Choosing the test function v_{hp} to be a product of one dimensional functions $\prod v_r$, leads to the following necessary condition for the existence of non-trivial solutions

$$\sum_{q=1}^{d} (\Phi_p'(k_q), v_q') \prod_{r \neq q} (X_p(k_r), v_r) - \kappa^2 \prod_{r=1}^{d} (X_p(k_r), v_r) = 0,$$

Therefore, in view of (14), we obtain

$$\left(\sum_{r=1}^{d} \kappa_r^2 - \kappa^2 \right) \prod_{q=1}^{d} (X_p(k_q), v_q) = 0,$$

and it follows that the discrete dispersion relation for the multi-dimensional scheme is given by

$$\sum_{r=1}^{d} \kappa_r^2 = \kappa^2 \tag{15}$$

where κ_r is related to k_r by the discrete dispersion relation (12). An alternative proof of this result for Gauss point mass lumped finite element schemes will be found the book [9, p. 228]. The above argument extends immediately to these schemes.

3 Higher Order Discrete Dispersion Relation

An explicit expression for the higher order dispersion relation could, at least in principle, be derived by proceeding directly as in the first order case. Unfortunately, such a direct computation rapidly becomes intractable with increasing order, as pointed out in [9], and the general result seems to be unavailable. In Sect. 4, we prove that the discrete dispersion relation is given explicitly in terms of Padé approximants [5]:

Theorem 1 *Suppose $p \in \mathbb{N}$, and define set $N_e = \lfloor p/2 \rfloor$ and $N_o = \lfloor (p+1)/2 \rfloor$. Let $[2N_e + 2/2N_e]_{\kappa \tan \kappa}$ and $[2N_o/2N_o - 2]_{\kappa \cot \kappa}$ denote the Padé approximants to $\kappa \tan \kappa$ and $\kappa \cot \kappa$ respectively. Then, the discrete dispersion relation is given by*

$$\cos(kh) = R_p(h\omega) \tag{16}$$

where R_p is the rational function

$$R_p(2\kappa) = \frac{[2N_o/2N_o - 2]_{\kappa \cot \kappa} - [2N_e + 2/2N_e]_{\kappa \tan \kappa}}{[2N_o/2N_o - 2]_{\kappa \cot \kappa} + [2N_e + 2/2N_e]_{\kappa \tan \kappa}}. \tag{17}$$

Despite the apparent simplicity, this result has been hitherto unknown in the literature, although special cases for lower order elements can be found in many sources. Table 1 shows the discrete dispersion relation and the leading term in the error in the dispersion relation for methods of order $p = 1$ to $p = 4$. For example, in the case of first order approximation $p = 1$,

$$\cos(kh) = R_1(\omega h) = \frac{6 - 2\omega^2 h^2}{6 + \omega^2 h^2},$$

which agrees with the result obtained for first order approximation in the previous section.

3.1 Accuracy at Small Wave Numbers

The following general result for the leading term for the error in the dispersion relation, valid for wave numbers satisfying $\omega h \ll 1$, is proved in Sect. 4:

Theorem 2 *Let $p \in \mathbb{N}$. Then, the error in the discrete dispersion relation is given by*

$$\cos kh - \cos \omega h = \frac{1}{2} \left[\frac{p!}{(2p)!} \right]^2 \frac{(\omega h)^{2p+2}}{2p+1} + \mathcal{O}(\omega h)^{2p+4}, \tag{18}$$

Order p	$R_p(\Omega)$	$\cos^{-1} R_p(\Omega) - \Omega$
1	$\dfrac{-2\Omega^2+6}{\Omega^2+6}$	$-\dfrac{\Omega^3}{24}$
2	$\dfrac{3\Omega^4-104\Omega^2+240}{\Omega^4+16\Omega^2+240}$	$-\dfrac{\Omega^5}{1440}$
3	$\dfrac{-4\Omega^6+540\Omega^4-11520\Omega^2+25200}{\Omega^6+30\Omega^4+1080\Omega^2+25200}$	$-\dfrac{\Omega^6}{201600}$
4	$\dfrac{5\Omega^8-1800\Omega^6+134064\Omega^4-2378880\Omega^2+5080320}{\Omega^8+48\Omega^6+3024\Omega^4+161280\Omega^2+5080320}$	$-\dfrac{\Omega^7}{50803200}$

Table 1: Discrete dispersion relation $\cos(kh) = R_p(\omega h)$ for order p approximation given in Theorem 1. The leading term in the series expansion for the error when $\Omega \ll 1$ (see Theorem 2) is also indicated.

or, if kh is sufficiently small,

$$kh - \omega h = -\frac{1}{2}\left[\frac{p!}{(2p)!}\right]^2 \frac{(\omega h)^{2p+1}}{2p+1} + \mathcal{O}(\omega h)^{2p+3}. \tag{19}$$

The result implies that

$$\frac{k}{\omega} - 1 = \mathcal{O}\left(\omega h\right)^{2p} \tag{20}$$

meaning that the dispersion relation for a p-th order scheme is accurate to order $2p$. This is consistent with the conjecture made by Thompson and Pinsky [29, eq. (41)] on the basis of numerical evidence in the particular cases of elements of order $p = 1$ to 5.

3.2 Accuracy at Large Wave Numbers

While error estimates for small values of ωh are not without interest, the most interesting case in practice is the high wave number limit, where the practical limitations on the size of mesh means that product ωh is large even though h is small. The next result describes the behaviour of the error as the order of approximation p is increased, so that both p and ωh are large:

Theorem 3 *Suppose that $\omega h \gg 1$. Then, the error $\mathcal{E}^p = \cos kh - \cos \omega h$ in the discrete dispersion relation passes through three distinct phases as the order $p \in \mathbb{N}$ is increased:*

1. *Oscillatory Phase: For $2p+1 < \omega h - o(\omega h)^{1/3}$, $\mathcal{E}^p$ oscillates, but does not decay, as p is increased;*
2. *Transition Zone: For $\omega h - o(\omega h)^{1/3} < 2p+1 < \omega h + o(\omega h)^{1/3}$, the error $\mathcal{E}^p$ decays algebraically at a rate $\mathcal{O}(p^{-1/3})$;*
3. *Super-Exponential Decay: For $2p+1 > \omega h + o(\omega h)^{1/3}$, $\mathcal{E}^p$ decreases at a super-exponential rate:*

$$\mathcal{E}^p \approx \frac{\sin(\omega h)}{2} f\left(\sqrt{1-(\omega h/(2p+1))^2}\right)^{p+1/2}, \tag{21}$$

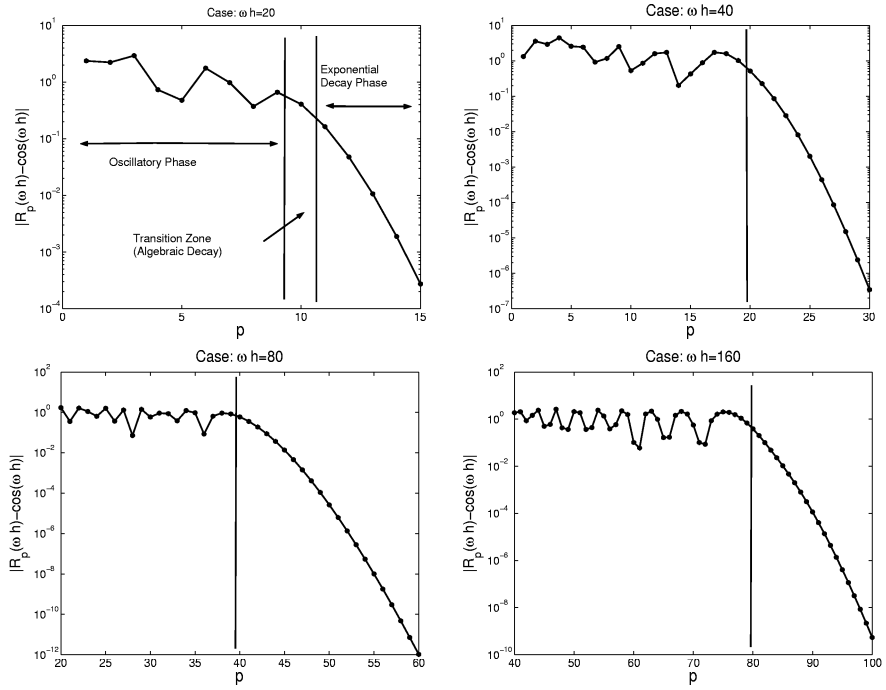

Figure 1: Behaviour of error in discrete dispersion relation for high wave numbers $\omega h \gg 1$ as the order p is increased. The transition region between the oscillatory phase and the super-exponential decay of the error is indicated in each case (cf. Theorem 3).

where $f : w \to (1-w)/(1+w)\exp(2w)$, *so that in the case where* $2p+1 > \omega he/2$,

$$\mathcal{E}^p \approx \frac{\sin(\omega h)}{2}\left[\frac{\omega he}{2(2p+1)}\right]^{2p+1}. \tag{22}$$

Observe that the term appearing in parentheses in eq. (21) has magnitude less than unity, which follows from the fact that the function $f : w \to (1-w)/(1+w)\exp(2w)$ is non-negative and monotonic decreasing on $[0,1]$ and the observation that $f(0) = 1$.

Figure 1 shows the behaviour of the actual error as the order p is increased for a range of values of ωh. The oscillatory region and the transition to the super-exponential decay in the error described in Theorem 3 can be clearly discerned.

Theorem 3 provides clear guidelines for the construction of meshes and choice of order for numerical resolution of waves using finite elements. In order to ensure that the dispersion error is properly controlled, it is desirable that we work in the super-exponential regime. For this reason, it is recommended to choose the order p and the mesh-size h so that

$$2p + 1 > \omega h + C(\omega h)^{1/3}. \tag{23}$$

The numerical results shown in Fig. 1 support this criterion. It is interesting that the same type of criterion is found in the implementation of the Fast Multipole Method for scattering problems [23, Eq. (3.38)].

4 Proofs of the Results

4.1 Basic polynomials

Let $\widehat{B}$ denote the bilinear form

$$\widehat{B}(u,v) = \int_{-1}^{1} \left(u'v' - \kappa^2 uv\right)\, \mathrm{d}s,$$

where $\kappa > 0$ is a constant. We introduce the following basic polynomials, Φ_o^p and Φ_e^p, of degree at most $p \in \mathbb{N}$ satisfying:

$$\Phi_e^p(\pm 1) = 1: \quad \widehat{B}(\Phi_e^p, v) = 0 \quad \forall v \in \mathbb{P}_p \cap H_0^1(-1,1) \tag{24}$$

and

$$\Phi_o^p(\pm 1) = \pm 1: \quad \widehat{B}(\Phi_o^p, v) = 0 \quad \forall v \in \mathbb{P}_p \cap H_0^1(-1,1). \tag{25}$$

Throughout, it will be assumed that κ does not coincide with an eigenvalue for this problem so that the polynomials are defined uniquely by these conditions.

It is easy to see that Φ_o^p and Φ_e^p are odd and even functions, respectively. The first result gives explicit closed forms for the expressions $\widehat{B}(\Phi_o^p, \Phi_o^p)$ and $\widehat{B}(\Phi_e^p, \Phi_e^p)$ which will be needed later.

Theorem 4 *Let $p \in \mathbb{N}$ satisfy $p \geq 2$. Then:*

1. *if $\kappa \neq (m + 1/2)\pi$ for all $m \in \mathbb{Z}$,*

$$\widehat{B}(\Phi_e, \Phi_e) = 2\kappa \frac{J_{2N+3/2}(\kappa)\cos\kappa + Y_{2N+3/2}(\kappa)\sin\kappa}{J_{2N+3/2}(\kappa)\sin\kappa - Y_{2N+3/2}(\kappa)\cos\kappa} \tag{26}$$

   where $N = \lfloor p/2 \rfloor$, and $\lfloor\cdot\rfloor$ denotes the integer part;

2. *if $\kappa \neq m\pi$ for all $m \in \mathbb{Z}$,*

$$\widehat{B}(\Phi_o^p, \Phi_o^p) = -2\kappa \frac{J_{2N+1/2}(\kappa)\sin\kappa - Y_{2N+1/2}(\kappa)\cos\kappa}{J_{2N+1/2}(\kappa)\cos\kappa + Y_{2N+1/2}(\kappa)\sin\kappa} \tag{27}$$

   where $N = \lfloor (p+1)/2 \rfloor$.

Proof. Symmetry arguments reveal that the function Φ_e^p is an even order polynomial of degree $2N$, where $N = \lfloor p/2 \rfloor$. For the remainder of the proof superscripts will be omitted since no confusion is likely to arise. Using definition (24) and integrating by parts shows that

$$\int_{-1}^{1} \left(\Phi_e'' + \kappa^2 \Phi_e\right) v \, \mathrm{d}x = 0 \quad \forall v \in \mathbb{P}_{2N} \cap H_0^1(-1,1)$$

where the superscript p has been omitted. The term in parentheses is a polynomial of degree $2N$, which may be written in the form

$$\Phi_e'' + \kappa^2 \Phi_e = \sum_{k=1}^{2N+1} \mu_k L_k'(x)$$

for suitable scalars μ_k, where L_k is the Legendre polynomial [15] of degree k. Inserting $v = (1-x^2)L_k'(x)$, with $j = 1, 2, \ldots, 2N-1$, and recalling the orthogonality property [15]

$$\int_{-1}^{1} (1-x^2) L_j'(x)\, L_k'(x)\, \mathrm{d}x = 0 \text{ for } j \neq k,$$

leads to the conclusion $\mu_j = 0$ for $j = 1, 2, \ldots, 2N-1$. Furthermore, the fact that L_{2N}' is an odd function and a parity argument shows $\mu_{2N} = 0$. Hence,

$$\Phi_e'' + \kappa^2 \Phi_e = \mu_{2N+1} L_{2N+1}'. \tag{28}$$

It is not difficult to verify that the function w_{2N} defined by

$$w_{2N}(x) = \sum_{j=0}^{N} \left(-\frac{1}{\kappa^2} \right)^{j+1} L_{2N+1}^{(2j+1)}(x) \tag{29}$$

is a polynomial of degree $2N$ satisfying

$$w_{2N}'' + \kappa^2 w_{2N} = -L_{2N+1}'.$$

Consequently, Φ_e may be written in the form

$$\Phi_e(x) = \frac{w_{2N}(x)}{w_{2N}(1)}$$

provided that $w_{2N}(1)$ is non-zero, and moreover, inserting this form into eq. (28) reveals that

$$\mu_{2N+1} = -1/w_{2N}(1).$$

With the aid of these results, we obtain

$$\begin{aligned}
\widehat{B}(\Phi_e, \Phi_e) &= \left[\Phi_e' \Phi_e\right]_{-1}^{1} - \int_{-1}^{1} \Phi_e \left(\Phi_e'' + \kappa^2 \Phi_e \right) \mathrm{d}x \\
&= 2\Phi_e'(1) - \mu_{2N+1} \int_{-1}^{1} \Phi_e L_{2N+1}'\, \mathrm{d}x \\
&= 2\Phi_e'(1) - \mu_{2N+1} \left[\Phi_e L_{2N+1}\right]_{-1}^{1} \\
&= 2\left(\Phi_e'(1) - \mu_{2N+1} \right) \\
&= \frac{2}{w_{2N}(1)} \left(1 + w_{2N}'(1) \right),
\end{aligned}$$

where standard properties of Legendre polynomials have been used such as the fact that L_{2N+1} is orthogonal to any polynomial of lower degree with respect

to the $L_2(-1,1)$ inner product, and $L_{2N+1}(\pm 1) = \pm 1$. The values of w and its derivative at $x = 1$ are given by the following formula which is obtained using eq. $(8.910)_2$ in [15]:

$$L_n^{(d)}(\pm 1) = \begin{cases} \dfrac{(n+d)!}{d!(n-d)!}\dfrac{(\pm 1)^{d+n}}{2^d}, & \text{for } d = 0, \ldots, n \\ 0, & \text{otherwise.} \end{cases} \tag{30}$$

giving, after some manipulation,

$$1 + w'_{2N}(1) = a_{2N+1} \tag{31}$$

and

$$w_{2N}(1) = -b_{2N+1}/\kappa. \tag{32}$$

Here, a_n and b_n are the expressions defined, for non-negative integers n, by

$$a_n = \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{(-1)^k}{(2k)!}\frac{(n+2k)!}{(n-2k)!}\frac{1}{(2\kappa)^{2k}} \tag{33}$$

and

$$b_n = \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} \frac{(-1)^k}{(2k+1)!}\frac{(n+2k+1)!}{(n-2k-1)!}\frac{1}{(2\kappa)^{2k+1}}. \tag{34}$$

These series appear in formulae (8.461) and (8.465) of [15], and satisfy the identity

$$\begin{bmatrix} \sin(\kappa - \pi n/2) & \cos(\kappa - \pi n/2) \\ \cos(\kappa - \pi n/2) & -\sin(\kappa - \pi n/2) \end{bmatrix} \begin{bmatrix} a_n \\ b_n \end{bmatrix} = \sqrt{\frac{\pi\kappa}{2}} \begin{bmatrix} J_{n+1/2}(\kappa) \\ (-1)^{n-1} Y_{n+1/2}(\kappa) \end{bmatrix} \tag{35}$$

where $J_{n+1/2}$ and $Y_{n+1/2}$ are Bessel functions of the first and second kind respectively. Equation (35) may be rearranged to obtain expressions for the series a_n and b_n, leading to the conclusion,

$$\widehat{B}(\Phi_e, \Phi_e) = -2\kappa\frac{a_{2N+1}}{b_{2N+1}} = 2\kappa\frac{J_{2N+3/2}(\kappa)\cos\kappa + Y_{2N+3/2}(\kappa)\sin\kappa}{J_{2N+3/2}(\kappa)\sin\kappa - Y_{2N+3/2}(\kappa)\cos\kappa} \tag{36}$$

which completes the proof in the even case.

The proof of the odd order case follows similar lines, leading to the following analogue of eq. (36),

$$\widehat{B}(\Phi_o, \Phi_o) = -2\kappa\frac{a_{2N}}{b_{2N}} \tag{37}$$

where $N = \lfloor (p+1)/2 \rfloor$. Inserting expressions for the series a_{2N} and b_{2N}, and simplifying leads to the result claimed. ∎

Equations (26) and (27) provide compact representations for the terms $\widehat{B}(\Phi_e^p, \Phi_e^p)$ and $\widehat{B}(\Phi_o^p, \Phi_o^p)$, but hide the fact that they are actually rational functions of κ. Interestingly enough, the expressions are actually certain types of Padé approximant [5]:

Theorem 5 *Let $p \in \mathbb{N}$ satisfy $p \geq 2$. Then:*

1. *$\widehat{B}(\Phi_e^p, \Phi_e^p)$ is the $[2N+2/2N]$-Padé approximant of $-2\kappa \tan\kappa$, where $N = \lfloor p/2 \rfloor$. Furthermore, if $\kappa \neq (m+1/2)\pi$, $m \in \mathbb{Z}$, then*

$$\begin{aligned} \mathcal{E}_e^p(\kappa) &= \widehat{B}(\Phi_e^p, \Phi_e^p) + 2\kappa \tan\kappa \\ &= \frac{1}{2}\left[\frac{(2N+1)!}{(4N+2)!}\right]^2 \frac{(2\kappa)^{4N+4}}{4N+3} + \mathcal{O}(\kappa^{4N+6}). \end{aligned} \tag{38}$$

2. *$\widehat{B}(\Phi_o^p, \Phi_o^p)$ is the $[2N/2N-2]$-Padé approximant of $2\kappa \cot\kappa$, where $N = \lfloor (p+1)/2 \rfloor$. If $\kappa \neq m\pi$, $m \in \mathbb{Z}$, then*

$$\begin{aligned} \mathcal{E}_o^p(\kappa) &= \widehat{B}(\Phi_o^p, \Phi_o^p) - 2\kappa \cot\kappa \\ &= 2\left[\frac{(2N)!}{(4N)!}\right]^2 \frac{(2\kappa)^{4N}}{4N+1} + \mathcal{O}(\kappa^{4N+2}). \end{aligned} \tag{39}$$

Proof. Firstly, eq. (36) shows that $\widehat{B}(\Phi_e^p, \Phi_e^p)$ is given by $-2\kappa a_{2N+1}/b_{2N+1}$ where a_{2N+1} and b_{2N+1} are defined in (33)-(34). It is not difficult to see that $\kappa^{2N+1}a_{2N+1}$ and $\kappa^{2N+1}b_{2N+1}$ are polynomials in κ of degree $2N+1$ and $2N$ respectively. Hence, $\widehat{B}(\Phi_e^p, \Phi_e^p)$ is a rational function of degree $[2N+2/2N]$. Straightforward manipulation beginning with the expression (26) gives

$$\widehat{B}(\Phi_e^p, \Phi_e^p) + 2\kappa \tan\kappa = -\frac{2\kappa}{\cos^2\kappa} Q_{2N+3/2}(\kappa) \left(1 - Q_{2N+3/2}(\kappa) \tan\kappa\right)^{-1} \tag{40}$$

where

$$Q_{2N+3/2}(\kappa) = \frac{J_{2N+3/2}(\kappa)}{Y_{2N+3/2}(\kappa)}.$$

The behaviour of $Q_{2N+3/2}(\kappa)$ is studied in the Appendix, where the following estimate is proved in Lemma 6,

$$Q_{2N+3/2}(\kappa) = -\frac{1}{2}\left[\frac{(2N+1)!}{(4N+2)!}\right]^2 \frac{(2\kappa)^{4N+3}}{4N+3} + \ldots.$$

With the aid of this estimate, we obtain that

$$\widehat{B}(\Phi_e^p, \Phi_e^p) + 2\kappa \tan\kappa = \frac{1}{2}\left[\frac{(2N+1)!}{(4N+2)!}\right]^2 \frac{(2\kappa)^{4N+4}}{4N+3} + \ldots$$

as claimed. Summarising, we have shown that $\widehat{B}(\Phi_e^p, \Phi_e^p)$ is rational function of degree $[2N+2/2N]$ which approximates $-2\kappa \tan\kappa$ to order $4N+4$. Consequently, $\widehat{B}(\Phi_e^p, \Phi_e^p)$ is the $[2N+2/2N]$-Padé approximant of $-2\kappa \tan\kappa$.

The assertions concerning $\widehat{B}(\Phi_o^p, \Phi_o^p)$ are proved in a similar fashion. In particular, using eq. (37) it is easy to see that $\widehat{B}(\Phi_o^p, \Phi_o^p)$ is a rational function of type $[2N/2N-2]$. With the aid of eq. (27), we derive

$$\widehat{B}(\Phi_o^p, \Phi_o^p) - 2\kappa \cot\kappa = -\frac{2\kappa}{\sin^2\kappa} Q_{2N+1/2}(\kappa) \left(1 + Q_{2N+1/2}(\kappa) \cot\kappa\right)^{-1}, \tag{41}$$

where

$$Q_{2N+1/2}(\kappa) = \frac{J_{2N+1/2}(\kappa)}{Y_{2N+1/2}(\kappa)},$$

and then, applying Lemma 6, we deduce that

$$\widehat{B}(\Phi_o^p, \Phi_o^p) - 2\kappa\cot\kappa = 2\left[\frac{(2N)!}{(4N)!}\right]^2 \frac{(2\kappa)^{4N}}{4N+1} + \dots$$

as claimed. It follows that $\widehat{B}(\Phi_o^p, \Phi_o^p)$ is the $[2N/2N-2]$ Padé approximant of $2\kappa\cot\kappa$. ∎

4.2 Proof of Theorem 1

We are now in a position to present the proof of Theorem 1:

Proof. Fix $\kappa = \omega h/2$. Firstly, we claim that for $x \in (0,h)$, the function $\theta_m^{(p)}$ defined in eqs (9)-(10) may be expressed in terms of the basic polynomials as follows:

$$\theta_0^{(p)}(x) = \frac{1}{2}\left[\Phi_e^p(s) - \Phi_o^p(s)\right]$$

and

$$\theta_1^{(p)}(x) = \frac{1}{2}\left[\Phi_e^p(s) + \Phi_o^p(s)\right]$$

where $s = 2x/h - 1 \in (-1,1)$. It is easy to verify that the expression for $\theta_0^{(p)}$ takes the correct values at the endpoints $x = 0$ and $x = h$. Moreover, since Φ_e^p is a polynomial of degree p (in both x and s), it suffices to show that the orthogonality condition (10) is satisfied. Let $v_{hp} \in V_{hp}^\flat$ be supported on $(0,h)$ and define $V \in \mathbb{P} \cap H_0^1(-1,1)$ by $V(s) = v_{hp}(x)$, $x \in (0,h)$. A simple change of variable reveals that

$$B_\omega(\theta_0^{(p)}, v_{hp}) = h^{-1}\widehat{B}(\Phi_e^p - \Phi_o^p, V)$$

and conditions (24)-(25) show that this vanishes. Similar arguments may be applied in the case of $\theta_1^{(p)}$.

It is clear from symmetry considerations that $\theta_0^{(p)}$ is an even function. This fact, combined with a simple change of variable shows that, since $\omega h = 2\kappa$,

$$B_\omega(\theta_0^{(p)}, \theta_0^{(p)}) = h^{-1}\widehat{B}(\Phi_e^p - \Phi_o^p, \Phi_e^p - \Phi_o^p)$$

then, exploiting the parities of Φ_e^p and Φ_o^p, we obtain

$$B_\omega(\theta_0^{(p)}, \theta_0^{(p)}) = h^{-1}\left[\widehat{B}(\Phi_e^p, \Phi_e^p) + \widehat{B}(\Phi_o^p, \Phi_o^p)\right].$$

Similar arguments reveal that

$$B_\omega(\theta_0^{(p)}, \theta_1^{(p)}) = (2h)^{-1}\left[\widehat{B}(\Phi_e^p, \Phi_e^p) - \widehat{B}(\Phi_o^p, \Phi_o^p)\right].$$

Substituting these results into (12) gives

$$\cos(kh) = \frac{\widehat{B}(\Phi_o^p, \Phi_o^p) + \widehat{B}(\Phi_e^p, \Phi_e^p)}{\widehat{B}(\Phi_o^p, \Phi_o^p) - \widehat{B}(\Phi_e^p, \Phi_e^p)}. \tag{42}$$

Theorem 5 identifies the terms appearing in this quotient as Padé approximants, and substituting for these expressions leads to the result claimed in Theorem 1. ■

4.3 Error for Small ωh

The general result for the leading term given in Theorem 2 for the error in the dispersion relation for small wave number is proved using Theorem 5 as follows:

Proof. Fix $\kappa = \omega h/2 \ll 1$, and let $\mathcal{E}_e^p(\kappa)$ and $\mathcal{E}_o^p(\kappa)$ be defined as in Theorem 5. By writing $\widehat{B}(\Phi_e^p, \Phi_e^p)$ and $\widehat{B}(\Phi_o^p, \Phi_o^p)$ in terms of $\mathcal{E}_e^p$ and $\mathcal{E}_o^p$ respectively, substituting into eq. (42), followed by a lengthy but otherwise straightforward computation, one arrives at the following expression for the error in the discrete dispersion relation:

$$\cos kh - \cos \omega h = \frac{\sin \omega h}{\omega h} \left\{ \mathcal{E}_o^p \sin^2(\omega h/2) + \mathcal{E}_e^p \cos^2(\omega h/2) \right\} \left\{ 1 + \frac{\sin \omega h}{2\omega h} (\mathcal{E}_o^p - \mathcal{E}_e^p) \right\}^{-1}. \tag{43}$$

(Here, the argument κ of $\mathcal{E}_e^p$ and $\mathcal{E}_o^p$ has been suppressed.) In particular, for small argument, Theorem 5 implies that

$$\mathcal{E}_o^p = 2 \left[\frac{(2N_o)!}{(4N_o)!} \right]^2 \frac{(\omega h)^{4N_o}}{4N_o + 1} + \dots$$

where $N_o = \lfloor (p+1)/2 \rfloor$, and

$$\mathcal{E}_e^p = \frac{1}{2} \left[\frac{(2N_e + 1)!}{(4N_e + 2)!} \right]^2 \frac{(\omega h)^{4N_e + 4}}{4N_e + 3} + \dots$$

where $N_e = \lfloor p/2 \rfloor$. It then follows that

$$\cos kh - \cos \omega h = \left(\frac{\omega h}{2} \right)^2 \mathcal{E}_o^p + \mathcal{E}_e^p + \dots$$

where

$$\left(\frac{\omega h}{2} \right)^2 \mathcal{E}_o^p = \frac{1}{2} \left[\frac{(2N_o)!}{(4N_o)!} \right]^2 \frac{(\omega h)^{4N_o + 2}}{4N_o + 1} + \dots$$

and $\mathcal{E}_e^p$ is given above. There are two cases, depending on the parity of the polynomial order p:

- If p is even, then $2N_e = 2N_o = p$ and the term involving $\mathcal{E}_o^p$ dominates the error, giving

$$\cos kh - \cos \omega h = \frac{1}{2} \left[\frac{p!}{(2p)!} \right]^2 \frac{(\omega h)^{2p+2}}{2p + 1} + \dots$$

- If p is odd, then $2N_e = p-1$ and $2N_o = p+1$ and the term involving $\mathcal{E}_e^p$ now dominates the error, giving

$$\cos kh - \cos \omega h = \frac{1}{2}\left[\frac{p!}{(2p)!}\right]^2 \frac{(\omega h)^{2p+2}}{2p+1} + \dots$$

This concludes the proof of (18). Estimate (19) then follows immediately from (18) using the approximation

$$\cos kh - \cos \omega h = -(kh - \omega h)\sin \omega h + \dots$$

valid for small $kh - \omega h$. ■

In examining this proof, we observe that the leading term in the remainder is the same, regardless of the parity of the polynomial order p. This effect occurs despite terms of different parities alternately dominating the error.

4.4 Error for Large ωh

Now fix $\kappa = \omega h/2 \gg 1$. Equation (40) implies that, with $\kappa = \omega h/2$,

$$\mathcal{E}_e^p(\kappa)\frac{\cos^2(\omega h/2)}{\omega h} = -Q_{2N_e+3/2}(\kappa)\left\{1 - Q_{2N_e+3/2}(\kappa)\tan\kappa\right\}^{-1} \tag{44}$$

while eq. (41) implies that

$$\mathcal{E}_o^p(\kappa)\frac{\sin^2(\omega h/2)}{\omega h} = -Q_{2N_o+1/2}(\kappa)\left\{1 + Q_{2N_o+1/2}(\kappa)\cot\kappa\right\}^{-1}. \tag{45}$$

Theorem 3 is proved using the foregoing results along with estimates for the behaviour of the quotient Q_m studied in Theorem 7 of the Appendix:

Proof. Firstly, consider the pre-asymptotic regime where $2p+1 < \omega h - o(\omega h)^{1/3}$. For p in this range, neither $2N_e$ nor $2N_o$ exceed $\kappa - o(\kappa^{1/3})$, where $\kappa = \omega h/2$. Therefore, we are in the situation covered by the first part of Theorem 7, where both $Q_{2N_o+1/2}(\kappa)$ and $Q_{2N_e+3/2}(\kappa)$ oscillate, but do not decay. Consequently, with the aid of the identities (44)-(45) and the expression for the error given in eq. (43), we are led to the conclusion that $\mathcal{E}^p$ oscillates, but does not decay, as p is increased in this range.

For p in the transition region where $\omega h - o(\omega h)^{1/3} < 2p+1 < \omega h + o(\omega h)^{1/3}$, it follows that both $2N_e$ and $2N_o$ lie in the transition region $[\kappa - o(\kappa^{1/3}), \kappa + o(\kappa^{1/3})]$ dealt with in the second part of Theorem 7. Here, the term appearing in the denominator of eq. (43) is $\mathcal{O}(1)$ for $\omega h \gg 1$. The identities (44)-(45) show that the error in the dispersion relation is dictated by the behaviour of the sum of $Q_{2N_e+3/2}(\kappa)$ and $Q_{2N_o+1/2}(\kappa)$. Applying Theorem 7, we conclude that the error decays algebraically at a rate $\mathcal{O}(p^{-1/3})$.

The proof in the case where $2p+1 > \omega h + o(\omega h)^{1/3}$ follows the same lines as the argument used in the transition region, and will not be elaborated upon further. ■

A Behaviour of $Q_m(\kappa)$

The quotient Q_m defined by

$$Q_m(\kappa) = \frac{J_m(\kappa)}{Y_m(\kappa)}, \quad m = \text{integer} + \frac{1}{2}, \tag{46}$$

appears in the expression for the error in the Padé approximants considered in Sect. 4. The following estimate, valid for small values of κ, was used in the proof of Theorem 5:

Lemma 6 *Let $m = \text{integer} + 1/2$ and let Q_m be defined as above. Then, for $\kappa \ll 1$,*

$$Q_m(\kappa) = -\frac{1}{2}\left[\frac{(m-\frac{1}{2})!}{(2m-1)!}\right]^2 \frac{(2\kappa)^{2m}}{2m} + \dots. \tag{47}$$

Proof. Write $m = n + 1/2$ where $n \in \mathbb{Z}$. For small κ, identity (8.440) of [15] gives

$$J_{n+1/2}(\kappa) = \frac{1}{\Gamma(3/2+n)}\left(\frac{\kappa}{2}\right)^{n+1/2} + \dots$$

while combining identities $(8.465)_1$ and (8.440) of [15] gives

$$Y_{n+1/2}(\kappa) = (-1)^{n-1}J_{-n-1/2}(\kappa) = \frac{(-1)^{n-1}}{\Gamma(1/2-n)}\left(\frac{\kappa}{2}\right)^{-n-1/2} + \dots$$

where Γ denotes the gamma function. Therefore, using formulae (8.339) of [15] gives, after some simplification,

$$Q_m(\kappa) = -\frac{1}{2}\left[\frac{n!}{(2n)!}\right]^2 \frac{(2\kappa)^{2n+1}}{2n+1} + \dots,$$

and rewriting in terms of m gives the result claimed. ∎

Lemma 6 shows that $Q_m(\kappa)$ decays algebraically as κ becomes small. However, it will be useful to consider the ratio in the regime $\kappa \gg 1$, with particular attention to the behaviour as the order m of the Bessel functions becomes large. Figure 2 shows the behaviour of $Q_m(\kappa)$ when $\kappa = 20$ as the order m is increased. It is found that there are three distinct phases depending on the size of the order m. Initially, $Q_m(\kappa)$ oscillates around unity. As the order m passes through κ, there is a relatively short-lived transition zone where $Q_m(\kappa)$ begins to decay at an algebraic rate as m is increased. Finally, as m is increased further, $Q_m(\kappa)$ decays at an exponential rate.

Our objective in the remainder of this section will be to show that the behaviour observed in this particular case is typical. The following result provides sharp estimates for the values at which the different phases occur and quantifies the rates of decay:

Theorem 7 *Let Q_m be defined as above, and $m = \text{integer} + 1/2$. Then, as m is increased, $Q_m(\kappa)$ passes through three phases:*

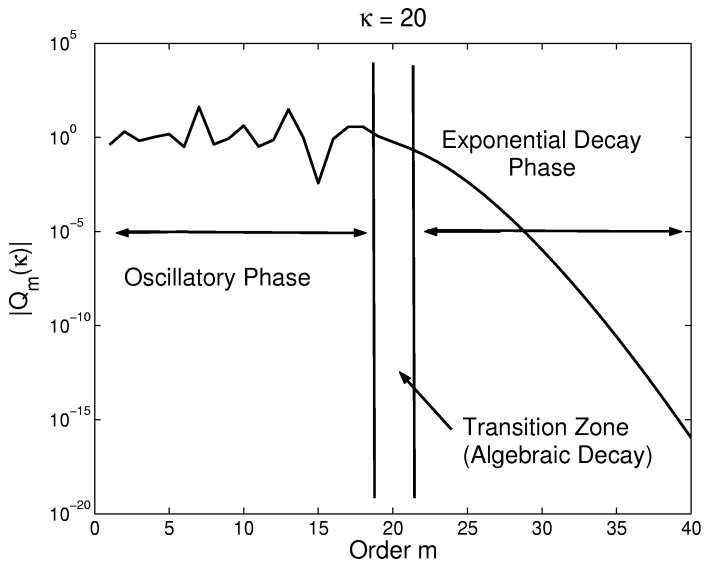

Figure 2: Graph showing the three phases in the behaviour of $|Q_m(\kappa)|$ for $\kappa = 20$ as the order m is increased.

1. *for* $m < \kappa - o(\kappa^{1/3})$, $Q_m(\kappa)$ *oscillates around unity, but does not decay as* m *is increased;*

2. *for* $\kappa - o(\kappa^{1/3}) < m < \kappa + o(\kappa^{1/3})$, *then* $Q_m(\kappa)$ *decays algebraically at a rate* $\mathcal{O}(m^{-1/3})$. *More precisely,*

$$Q_m(\kappa) \approx -\frac{1}{\sqrt{3}} - \frac{1}{\pi}\Gamma\left(\frac{2}{3}\right)^2 (\kappa - m)\left(\frac{6}{m}\right)^{1/3} + \dots \tag{48}$$

3. *for* $m > \kappa + o(\kappa^{1/3})$, $Q_m(\kappa)$ *decays at a super-exponential rate:*

$$Q_m(\kappa) \approx -\frac{1}{2}\left[\frac{1-\sqrt{1-\kappa^2/m^2}}{1+\sqrt{1-\kappa^2/m^2}} e^{2\sqrt{1-\kappa^2/m^2}}\right]^m, \tag{49}$$

so that, for $m \gg \kappa$,

$$Q_m(\kappa) \approx -\frac{1}{2}\left[\frac{\kappa e}{2m}\right]^{2m}. \tag{50}$$

The proof of this result is divided into four distinct cases covered in the following sections.

A.1 Pre-asymptotic Regime: $m < \kappa$

We start by discussing the behaviour of $Q_m(\kappa)$ in the pre-asymptotic regime where the value of the argument κ exceeds the order m of the Bessel functions. Langer's formulas [13, Sect. 7.13.4] provide uniform asymptotic expansions for

Bessel functions of large order and large argument, and give

$$Q_m(\kappa) = \frac{J_{1/3}(z)\cos(\pi/6) - Y_{1/3}(z)\sin(\pi/6) + \mathcal{O}(m^{-4/3})}{J_{1/3}(z)\sin(\pi/6) + Y_{1/3}(z)\cos(\pi/6) + \mathcal{O}(m^{-4/3})} \tag{51}$$

where

$$z = m(w - \tan^{-1} w) \text{ and } w = \sqrt{\kappa^2/m^2 - 1}.$$

Bounds on the accuracy of the approximation obtained when the higher order terms are dropped in eq. (51) could be obtained using the uniform asymptotic expansions with remainder given in Olver [27]. However, we shall content ourselves with making the approximation

$$Q_m(\kappa) \approx \frac{J_{1/3}(z)\cos(\pi/6) - Y_{1/3}(z)\sin(\pi/6)}{J_{1/3}(z)\sin(\pi/6) + Y_{1/3}(z)\cos(\pi/6)}. \tag{52}$$

A.1.1 Oscillatory Phase: $m < \kappa - o(\kappa^{1/3})$

In the pre-asymptotic range where m is small relative to κ, the ratio $Q_m(\kappa)$ tends to oscillate and has magnitude of order unity. While it is difficult to make quantitative statements concerning the erratic behaviour observed in Fig. 2, it is possible to give a qualitative explanation. When m is small relative to κ, the argument z of the Bessel functions appearing in eq. (52) will be large and positive. Asymptotic expansions for Bessel functions of large argument are given in $(8.440)_1$ of [15]:

$$J_\nu(z) \sim \left(\frac{\pi z}{2}\right)^{-1/2} \cos\left(z - \frac{1}{2}\nu\pi - \frac{\pi}{4}\right)$$

and in $(8.440)_2$ of [15]:

$$Y_\nu(z) \sim \left(\frac{\pi z}{2}\right)^{-1/2} \sin\left(z - \frac{1}{2}\nu\pi - \frac{\pi}{4}\right).$$

Together with (52), these expressions show that $Q_m(\kappa)$ will tend to oscillate without a decay in the magnitude as m is increased. Indeed, these expressions into the right hand side of (52) and simplifying gives

$$\cot\left(z - \frac{\pi}{4}\right). \tag{53}$$

Of course, we would not expect this expression to necessarily agree closely with $Q_m(\kappa)$. Nevertheless, this expression actually provides a surprisingly good representation of the qualitative behaviour in the pre-asymptotic regime even in the case of relatively modest values of κ, as shown in Fig. 3.

A.1.2 Transition Zone: $\kappa - o(\kappa^{1/3}) < m < \kappa$

We consider the behaviour in the zone where m approaches κ from below. For m in this range,

$$1 < \frac{\kappa}{m} < 1 + o(m^{-1/3})$$

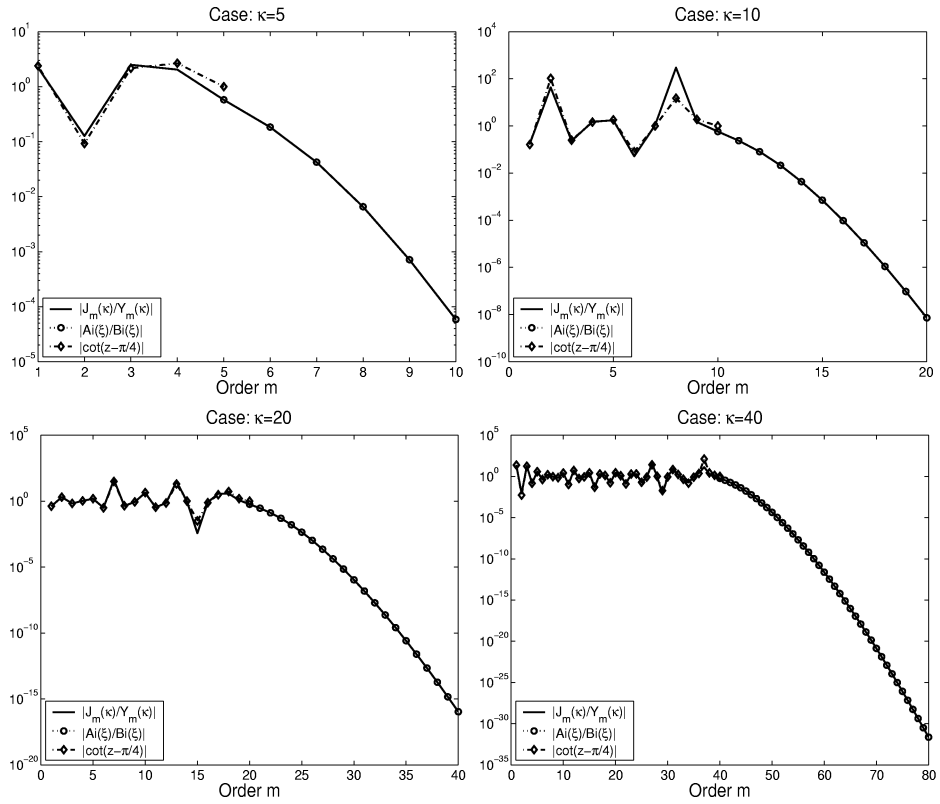

Figure 3: Graphs of $|Q_m(\kappa)|$ in (46) for $m = 1, \ldots, 2\kappa$, $|\cot(z - \pi/4)|$ in (53) for $m = 1, \ldots, \kappa$ and $|\operatorname{Ai}(\xi)/\operatorname{Bi}(\xi)|$ in (56) for $m = \kappa + 1, \ldots, 2\kappa$. Values of $\kappa = 5$, 10, 20 and 40 are shown. Observe the oscillatory behaviour of $|Q_m|$ and the good qualitative agreement provided by the cotangent in the pre-asymptotic regime $m < \kappa$. Furthermore, note the quantitative agreement between $|Q_m|$ and $|\operatorname{Ai}(\xi)/\operatorname{Bi}(\xi)|$ in the asymptotic regime $m > \kappa$.

so that,

$$w \approx \left[\frac{\kappa - m}{(m/2)}\right]^{1/2} = o(1),$$

and therefore,

$$z \approx \frac{1}{3} m w^3 \approx \frac{2}{3}\left[\frac{\kappa - m}{(m/2)^{1/3}}\right]^{3/2} = o(1).$$

Using the series representations for Bessel functions [15, eq. (8.440)], we obtain

$$Q_m(\kappa) \approx -\frac{1}{\sqrt{3}} - \frac{3}{\pi}\Gamma\left(\frac{2}{3}\right)^2 \left(\frac{z}{2}\right)^{2/3} + \ldots$$

and, by substituting for z and simplifying, we arrive at the conclusion

$$Q_m(\kappa) \approx -\frac{1}{\sqrt{3}} - \frac{1}{\pi}\Gamma\left(\frac{2}{3}\right)^2 (\kappa - m)\left(\frac{6}{m}\right)^{1/3} + \ldots \tag{54}$$

valid for $\kappa - o(\kappa^{1/3}) < m < \kappa$. As a matter of fact, this result could also have been obtained formally using Nicholson's formulas [13, Sect. 7.13.3].

A.2 Asymptotic Regime: $m > \kappa$

We now study the behaviour in the regime where the order of the Bessel functions exceeds the argument. Langer's formulas [13, Sect. 7.13.4] imply that

$$Q_m(\kappa) = -\frac{\pi^{-1}K_{1/3}(z) + \mathcal{O}(m^{-4/3})}{I_{1/3}(z) + I_{1/3}(z) + \mathcal{O}(m^{-4/3})} \tag{55}$$

where, in this case,

$$z = m\left(\tanh^{-1} w - w\right) \text{ and } w = \sqrt{1 - \kappa^2/m^2}.$$

As before, it is possible to use the results of Olver [27] to obtain bounds on the accuracy of the approximation obtained when the higher order terms are dropped in eq. (55), although we will not pursue this further here. Writing $z = \frac{2}{3}\xi^{3/2}$ and using formulas (11.1.04) and (11.1.12) from [27] gives

$$Q_m(\kappa) \approx -\frac{\pi^{-1}K_{1/3}(z)}{I_{1/3}(z) + I_{1/3}(z)} = -\frac{\mathrm{Ai}(\xi)}{\mathrm{Bi}(\xi)} \tag{56}$$

where Ai and Bi denote Airy functions of the first and second kinds respectively [15]. The accuracy of this approximation is indicated in Fig. 3.

The behaviour of the ratio of Airy functions for positive ξ may be deduced from the results quoted in Olver [27, pp. 392-393]. However, the following simple approximations will suffice for present purposes. For small argument $\xi \leq 2$,

$$\frac{\mathrm{Ai}(\xi)}{\mathrm{Bi}(\xi)} \approx \frac{1}{\sqrt{3}} \frac{1 - 3^{5/6}\Gamma\left(\frac{2}{3}\right)^2 \frac{\xi}{2\pi} + \frac{\xi^3}{12}}{1 + 3^{5/6}\Gamma\left(\frac{2}{3}\right)^2 \frac{\xi}{2\pi} + \frac{\xi^3}{12}} \tag{57}$$

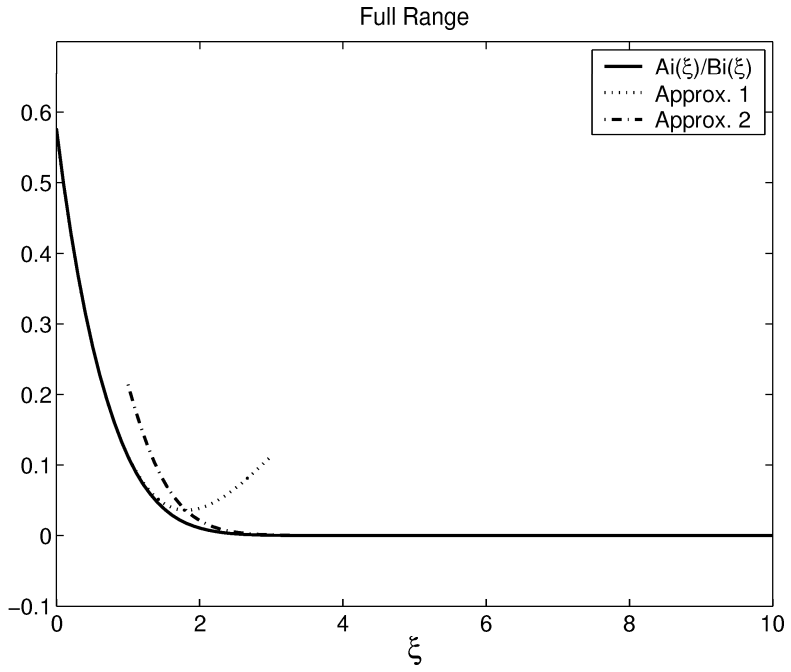

Figure 4: Graph showing the ratio $\mathrm{Ai}(\xi)/\mathrm{Bi}(\xi)$ compared with the approximations for small and large arguments given in eqs (57) and (58) respectively.

while for larger arguments where $\xi > 2$,

$$\frac{\mathrm{Ai}(\xi)}{\mathrm{Bi}(\xi)} \approx \frac{e^{-2z}}{2}\frac{1-\frac{15}{216z}}{1+\frac{15}{216z}}, \quad z = \frac{2}{3}\xi^{3/2}. \tag{58}$$

Here Γ denotes the gamma function [15]. Together, these approximations provide an accurate picture of the behaviour of the ratio throughout the full range of argument, as indicated in Fig. 4. In particular, it is observed that the ratio initially decays linearly,

$$\frac{\mathrm{Ai}(\xi)}{\mathrm{Bi}(\xi)} \approx \frac{1}{\sqrt{3}} - \frac{3^{1/3}}{\pi}\Gamma\left(\frac{2}{3}\right)^2 \xi, \tag{59}$$

before undergoing a rapid transition to an exponential rate of decay given by

$$\frac{\mathrm{Ai}(\xi)}{\mathrm{Bi}(\xi)} \approx \frac{e^{-2z}}{2}. \tag{60}$$

A.2.1 Transition Zone: $\kappa < m < \kappa + o(\kappa^{1/3})$

As the order m passes through κ, we have

$$1 - o(m^{-1/3}) < \frac{\kappa}{m} < 1$$

and using similar arguments to those used before, we obtain

$$z \simeq \frac{1}{3}mw^3 \simeq \frac{2}{3}\left[\frac{m-\kappa}{(m/2)^{1/3}}\right]^{3/2}$$

or, equally well,

$$\xi \simeq \left(\frac{2}{m}\right)^{1/3}(m-\kappa).$$

Since $m-\kappa = o(\kappa^{1/3}) = o(m^{1/3})$, it follows that $\xi = o(1)$ and, using (59), we obtain

$$Q_m(\kappa) \approx -\frac{\mathrm{Ai}(\xi)}{\mathrm{Bi}(\xi)} \simeq -\frac{1}{\sqrt{3}} + \frac{1}{\pi}\Gamma\left(\frac{2}{3}\right)^2 (m-\kappa)\left(\frac{6}{m}\right)^{1/3} + \ldots. \tag{61}$$

valid for $\kappa < m < \kappa + o(\kappa^{1/3})$. This form agrees with the result (54) obtained when m lies the transition region to the left of κ.

A.2.2 Exponential Decay Phase: $m > \kappa + o(\kappa^{1/3})$

If m exceeds $\kappa + o(\kappa^{1/3})$, then w is no longer small, and in turn, z and ξ will be large. By substituting for z in terms of w in the expression (60) and applying elementary manipulations, we arrive at

$$\frac{\mathrm{Ai}(\xi)}{\mathrm{Bi}(\xi)} \approx \frac{1}{2}\left[\frac{1-w}{1+w}e^{2w}\right]^m,$$

or, substituting for w,

$$\frac{\mathrm{Ai}(\xi)}{\mathrm{Bi}(\xi)} \approx \frac{1}{2}\left[\frac{1-\sqrt{1-\kappa^2/m^2}}{1+\sqrt{1-\kappa^2/m^2}}e^{2\sqrt{1-\kappa^2/m^2}}\right]^m.$$

The function $f : w \to (1-w)/(1+w)\exp(2w)$ is monotonic decreasing on $[0,1]$ from 0 to 1. Therefore, the term in parentheses is less than unity and we have an exponential rate of decay when m is close to κ. This rate of decay increases as m becomes even larger relative to κ, and in the limiting case, we find that

$$f\left(\sqrt{1-\kappa^2/m^2}\right) \simeq \left[\frac{\kappa e}{2m}\right]^2.$$

Therefore, when $m > \kappa e/2$, we obtain a super-exponential rate of decay

$$\frac{\mathrm{Ai}(\xi)}{\mathrm{Bi}(\xi)} \approx \frac{1}{2}\left[\frac{\kappa e}{2m}\right]^{2m}.$$

References

[1] N.N. Abboud and P.M. Pinsky. Finite element dispersion analysis for the 3-dimensional 2nd- order scalar wave-equation. *Internat. J. Numer. Methods Engrg.*, 35(6):1183–1218, October 1992.

[2] M. Ainsworth. Dispersive properties of high order Nédélec/edge element approximation of the time-harmonic Maxwell equations. *Phil. Trans. Roy. Soc. Series A*, Submitted for publication.

[3] J. Astley, K. Gerdes, D. Givoli, and I. Harari. Special issue on Finite elements for wave problems–Preface. *J. Comput. Acoust.*, 8(1):vii–ix, March 2000.

[4] I. Babuška and F. Ihlenburg. Dispersion analysis and error estimation of Galerkin finite element methods for the Helmholtz equation. *Internat. J. Numer. Methods Engrg.*, 38:3745–3774, 1995.

[5] G.A. Baker and P. Graves-Morris. *Padé Approximants.* Encyclopedia of Mathematics and Its Applications. Cambridge University Press, 2nd edition, 1995.

[6] M.A. Christon. The influence of the mass matrix on the dispersive nature of the semi-discrete, second-order wave equation. *Comput. Methods Appl. Mech. Engrg.*, 173:146–166, 1999.

[7] G. Cohen and P. Monk. Gauss point mass lumping schemes for Maxwell's equations. *Numer. Meth. PDE.*, 14:63–88, 1998.

[8] G. Cohen and P. Monk. Mur-Nédélec finite element schemes for Maxwell's equations. *Comput. Methods Appl. Mech. Engrg.*, 169:197–217, 1999.

[9] G.C. Cohen. *Higher-Order Numerical Methods for Transient Wave Equations.* Springer Series in Scientific Computation. Springer-Verlag, Berlin, Heidelberg, New York, 2002.

[10] L. Demkowicz and L. Vardapetyan. Modeling of electromagnetic absorption/scattering problems using hp-adaptive finite elements. *Comput. Methods Appl. Mech. Engrg.*, 152(1-2):103–124, 1998.

[11] A. Deraemaeker, I. Babuška, and P. Bouillard. Dispersion and pollution of the FEM solution for the Helmholtz equation in one, two and three dimensions. *Internat. J. Numer. Methods Engrg.*, 46(4):471–499, October 1999.

[12] R.W. Dyson. Technique for very high order nonlinear simulation and validation. *J. Comput. Acoust.*, 10(2):211–229, June 2002.

[13] A. Erdélyi, W. Magnus, F. Oberhettinger, and F.G. Tricomi. *Higher transcendental functions.* Bateman Manuscript Project. McGraw-Hill, New York, London, 1953–55.

[14] K. Gerdes and F. Ihlenburg. On the pollution effect in finite element solutions of the 3D Helmholtz equation. *Comput. Methods Appl. Mech. Engrg.*, 170(1-2):155–172, February 1999.

[15] I.S. Gradshteyn and I.M. Ryzhik. *Table of Integrals, Series and Products.* Academic Press Ltd, A. Jeffrey (ed.) United Kingdom 5th edition, 1994.

[16] I. Harari. Reducing spurious dispersion, anisotropy and reflection in finite element analysis of time-harmonic acoustics. *Comput. Methods Appl. Mech. Engrg.*, 140(1-2):39–58, January 1997.

[17] I. Harari. Finite element dispersion of cylindrical and spherical acoustic waves. *Comput. Methods Appl. Mech. Engrg.*, 190(20-21):2533–2542, 2001.

[18] I. Harari and D. Avraham. High-order finite element methods for acoustic problems. *J. Comput. Acoust.*, 5(1):33–51, March 1997.

[19] I. Harari and T.J.R. Hughes. Finite-element methods for the Helmholtz equation in an exterior domain-model problems. *Comput. Methods Appl. Mech. Engrg.*, 87(1):59–96, May 1991.

[20] J.S. Hesthaven and T. Warburton. Nodal high-order methods on unstructured grids. I. Time-domain solution of Maxwell's equations. *J. Comput. Phys.*, 181(1):186–221, September 2002.

[21] F. Ihlenburg and I. Babuška. Finite element solution of the Helmholtz equation with high wave-number. Part 1. The h-version of the finite element method. *Comput. Math. Appl.*, 30(9):9–37, November 1995.

[22] F. Ihlenburg and I. Babuška. Finite element solution of the Helmholtz equation with high wave number. Part 2. The h-p version of the finite element method. *SIAM J. Numer. Anal.*, 34(1):315–358, 1997.

[23] S. Koc, Jiming Song, and W.C. Chew. Error analysis for the numerical evaluation of the diagonal forms of the scalar spherical addition theorem. *SIAM J. Numer. Anal.*, 36(3):906–921, April 1999.

[24] P.B. Monk and A.K. Parrott. Dispersion analysis of finite element methods for Maxwell's equations. *SIAM J. Sci. Comput.*, 15(4):916–937, 1994.

[25] A.A. Oberai and P.M. Pinsky. A numerical comparison of finite element methods for the Helmholtz equation. *J. Comput. Acoust.*, 8(1):211–221, March 2000.

[26] F. Odeh and J.B. Keller. Partial differential equations with periodic coeffcients and Bloch waves in crystals. *J. Math. Phys.*, 5(11):1499–1504, November 1964.

[27] F.W.J. Olver. *Asymptotics and Special Functions.* Computer Science and Applied Mathematics. Academic Press, 1974.

[28] C. Schwab and M. Suri. The p and hp versions of the finite element method for problems with boundary layers. *Math. Comp.*, 65(216):1403–1429, October 1996.

[29] L.L. Thompson and P.M. Pinsky. Complex wavenumber Fourier analysis of the p-version finite element method. *Comput. Mech.*, 13:255–275, 1994.