hp-finite element method for coupled linear systems of mixed-order differential equations with boundary layer behaviour.

D. Farmer[1]   D. J. Ivers[2]

Abstract

Spectral methods for the horizontal dependence in linearised magnetohydrodynamics in spherical geometries and plane layers give rise to mixed-order (second and fourth) systems of equations in the radial or vertical dependence. The spherical case was covered by Ivers & Phillips [*Geophys J. Intl* 175, 2008]. Previous work of Farmer & Ivers [*ANZIAM J.* 52, 2011] indicated how to extend results appearing by Schwab [1998] to multiple boundary layers retaining robust exponential convergence (in the small parameter appearing in the coefficient of the highest derivative terms). This article investigates the recovery of the interior solution for different model problems. The main result is that the robust convergence results reported by Farmer & Ivers [2011]

 Record comments on this article via
http://journal.austms.org.au/ojs/index.php/ANZIAMJ/comment/add/10173/0

do carry through for coupled mixed-order problems. It is also shown that for a coupled mixed-order system there is little difference between using a C^0 expansion for the second-order variables and C^1 expansion for the fourth-order, versus the algorithmically simpler choice of a C^1 expansion for both. Calibration of the method to determine the boundary layer widths and the degree of approximation for the interior when the exact solution is unknown, is discussed. The results of this article can be applied to solving the linearised magnetohydrodynamics equations in spherical geometries and plane layers.

Contents

1 Introduction

The effective solution of coupled linear systems of mixed-order ordinary differential equations, which may include boundary layers, is of interest when studying the linearised magnetohydrodynamic (MHD) equations in plane layers and spherical geometries [1, 4, 9, 10, 12, 5, 6]. Farmer & Ivers [2] summarised the spherical case. This article investigates different boundary layer behaviour to determine how best to apply the *hp*-finite element method (FEM) for use in MHD applications. The focus of the solution methods is on accurate recovery of the interior solution for smallest degrees of freedom, studied using three model problems.

Problem I is a second-order boundary layer problem: find $y(x) \in C^2[-1, 1]$, such that

$$- \left(d^2 D^2 - 1 \right) y = f, \quad |x| < 1, \tag{1}$$
$$y(\pm 1) = 0, \tag{2}$$

for operator $D = d/dx$ and the prescribed function $f(x)$. The parameter $0 < d \ll 1$ is small and positive.

Problem II is a fourth-order two boundary layer problem: find $y(x) \in C^4[-1, 1]$, such that

$$\left(a^2 d^2 D^2 - 1 \right) \left(d^2 D^2 - 1 \right) y = f, \quad |x| < 1, \tag{3}$$
$$y(\pm 1) = y^{(1)}(\pm 1) = 0. \tag{4}$$

where $0 < a < 1$ is the constant relative thickness of the inner boundary layer (thickness ad) to the outer boundary layer (thickness d) and the function $f(x)$ is prescribed. The notation $y^{(n)}(x)$ denotes the nth derivative of y with respect to its variable x.

Problem III is the mixed fourth-order and second-order system: find $y(x) \in C^4[-1,1]$ and $z(x) \in C^2[-1,1]$, such that

$$\lambda_1 y^{(4)} - \lambda_2 y^{(2)} + c_1 y + c_2 z = f_1, \tag{5}$$

$$-\lambda_3 z^{(2)} + c_3 z - \lambda_4 y^{(2)} + c_4 y = f_2, \quad |x| < 1; \tag{6}$$

$$z(\pm 1) = y(\pm 1) = y^{(1)}(\pm 1) = 0, \tag{7}$$

for constants $0 < \lambda_i \ll 1$ and c_j close to one, and prescribed functions $f_1(x)$ and $f_2(x)$.

Problems I and II have been studied with the constant forcing $f(x) = 1$ by Schwab [11] and Farmer & Ivers [2]. Here they are re-examined, this time with an oscillatory forcing $f(x) = \cos(bx)$ in order to determine the scaling of the boundary elements, focusing on efficient recovery of the interior solution in Subsection 4.1–4.2.

The mixed-order coupled Problem III is new to this article, again studied to determine strategies for the efficient recovery of the interior solution in Subsection 4.3.

How to apply the FEM in practise, when the boundary layer widths and degree of approximation of the interior is unknown, is the topic of Subsection 4.4.

Problems I–III will be used in future research to develop a strategy for solving the linearised MHD equations in plane layers and spherical geometries [1, 2, 4, 9, 10, 12, 5, 6]: it forms a large and sparse coupled linear system of mixed-order differential equations, potentially with boundary layer behaviour. The study of the linearised MHD equations provides valuable insight into the solution of the full nonlinear MHD equations [1], and is useful in time-stepping codes [8].

2   Weak formulation and *hp*-bases

For the purpose of completeness, and to present unpublished results, the weak formulation of Problem III is described in detail, along with the discretisation of the weak formulation in Section 3. For Problem III, multiplying (5) and (6) by test functions σ and η, adding and integrating gives

$$\int_{-1}^{1} \left(\lambda_1 y^{(4)} - \lambda_2 y^{(2)} + c_1 y + c_2 z\right)\sigma + \left(-\lambda_3 z^{(2)} + c_3 z - \lambda_4 y^{(2)} + c_4 y\right)\eta \, dx$$

$$= \int_{-1}^{1} (f_1 \sigma + f_2 \eta) \, dx. \tag{8}$$

Further integration by parts of the derivative terms in (8) can be performed, leading to different weak-forms. The choice of weak-form along with the boundary conditions imposes different restrictions on the underlying function spaces for the test functions (σ and η) and trial functions (y and z). The following integration by parts of (8) imposes minimal regularity on these spaces

$$\lambda_1 y^{(3)} \sigma - \lambda_1 y^{(2)} \sigma^{(1)} - \lambda_3 z^{(1)} \eta - \lambda_4 y^{(1)} \eta \Big]_{-1}^{1} + \int_{-1}^{1} \left(\lambda_1 y^{(2)} \sigma^{(2)}\right.$$

$$+ \lambda_2 y^{(1)} \sigma^{(1)} + c_1 y \sigma + c_2 z \sigma + \lambda_3 z^{(1)} \eta^{(1)} + c_3 z \eta + \lambda_4 y^{(1)} \eta^{(1)} + c_4 y \eta\Big) \, dx$$

$$= \int_{-1}^{1} (f_1 \sigma + f_2 \eta) \, dx. \tag{9}$$

There is a choice to integrate the $y^{(2)}$ terms in (8) or not. If these coefficients were non-constant, then it would be more straightforward to leave the corresponding terms unintegrated. Consider the terms in (9) that arise from integration by parts, since $y^{(3)}$, $y^{(2)}$ and $z^{(1)}$ are unknown on the boundary, the choices of the test spaces $\sigma \in H_0^2 = \{\phi \in H^2 \mid \phi(\pm 1) = \phi'(\pm 1) = 0\}$ and $\eta \in H_0^1 = \{\phi \in H^1 \mid \phi(\pm 1) = 0\}$ are appropriate. The remaining terms in (9) comprise a weak form of Problem III, as in this setting $y \in H_0^2$ and $z \in H_0^1$. The boundary conditions (7) are termed essential: they are imposed directly

on the trial function spaces, as opposed to natural boundary conditions, that arise from the weak form through integration by parts.

The *hp*-FEM discretisation follows by partitioning the domain into a mesh $-1 = x_1 < \cdots < x_{n+1} = 1$ of elements or subintervals $[x_e, x_{e+1}]$ and expanding y, z, σ and η in appropriate polynomial bases.

For global C^0 expansions the following hierarchical or modal basis is used on the standard element $[-1, 1]$

$$\psi_n(\xi) = \begin{cases} (1 - \xi)/2\,, & n = 1\,, \\ (1 + \xi)/2\,, & n = 2\,, \quad |\xi| \leqslant 1\,. \\ \mu_n(1 - \xi^2)P_{n-3}^{1,1}(\xi), & n \geqslant 3\,, \end{cases}$$

Modes ψ_1 and ψ_2 are boundary modes enforcing continuity at the left and right endpoints, respectively, of each element. The remaining modes correspond to global functions that are zero outside the element. These interior modes are scaled so that $\int_{-1}^{1} \left(\psi_n^{(1)}\right)^2 d\xi = 1$ [11, §3.1]. The function $P_n^{\alpha,\beta}(\xi)$ where α, $\beta > -1$ is the Jacobi polynomial of degree n and weight $(1 - \xi)^\alpha(1 + \xi)^\beta$, see Appendix A. The basis $\{\psi_n\}_{n\in\mathbb{N}}$ is hierarchical or modal if $\{\psi_j \mid j \leqslant N_1\} \subset \{\psi_j \mid j \leqslant N_2\}$ for $N_1 < N_2$ (compare with spectral elements [8, §2.3.4.2]). Modal bases are useful in *hp*-FEM codes for *p*-refinement, that is, the increase of the order of approximation on a particular element. The choice of $P_{n-3}^{1,1}(\xi)$ in the interior modes produces sparse matrices $\mathbf{M}\{\psi\} = \left[M\{\psi\}_{i,j}\right]$ where $M\{\psi\}_{i,j} = \int_{-1}^{1} \psi_i \psi_j \, d\xi$, and $\mathbf{L}\{\psi\} = \left[L\{\psi\}_{i,j}\right]$ where $L\{\psi\}_{i,j} = \int_{-1}^{1} \psi_i' \psi_j' \, d\xi$.

As can be seen from (9) in the expansion for z and η the first derivative of ψ_n is required. Although this may be calculated from the $\psi_n(\xi)$ by collocation differentiation, exact expressions are derived. The derivatives of the interior modes are found from substitution into the recurrence relation for the derivatives of Jacobi polynomials (51) in Appendix A,

$$\psi_n'(\xi) = \begin{cases} -1/2\,, & n = 1\,, \\ 1/2\,, & n = 2\,, \quad |\xi| \leqslant 1\,, \\ -\mu_n\left[(n-1)\xi P_{n-3}^{1,1}(\xi) - (n-2)P_{n-4}^{1,1}(\xi)\right], & n \geqslant 3\,, \end{cases}$$

where $P_{-1}^{\alpha,\beta} := 0$, and the symbol $:=$ denotes "is defined by". From the identity (53) and the orthogonality relation (48) it follows that the normalisation factor $\mu_n = (4n - 6)^{1/2}/(4n - 8)$.

Global C^1 expansions are implemented by the basis

$$\chi_n(\xi) = \begin{cases} (1 - \xi)^2(1 + \xi)/4, & n = 1, \\ (1 - \xi)^2(2 + \xi)/4, & n = 2, \\ (1 + \xi)^2(2 - \xi)/4, & n = 3, \quad |\xi| \leqslant 1. \\ -(1 + \xi)^2(1 - \xi)/4, & n = 4, \\ \mu_n(1 - \xi^2)^2 P_{n-5}^{2,2}(\xi), & n \geqslant 5, \end{cases}$$

The basis functions χ_j for $1 \leqslant j \leqslant 4$ are the boundary modes, which enforce continuity of the expansion and its first derivative, respectively, at the left endpoint of each element for modes χ_1, χ_2 and the right endpoint for modes χ_3, χ_4. The choice of $P_{n-5}^{2,2}(\xi)$ in the interior modes is again so that the matrices $\mathbf{M}\{\chi\}$ and $\mathbf{L}\{\chi\}$ and the matrix $\mathbf{H}\{\chi\} = \left[H\{\chi\}_{i,j}\right]$ where $H\{\chi\}_{i,j} = \int_{-1}^1 \chi_i^{(2)} \chi_j^{(2)} \, d\xi$, are sparse. The normalisation factor μ_n is such that $\int_{-1}^1 (\chi_n^{(2)})^2 \, d\xi = 1$ [11, §3.1]. The factor $\mu_n = (4n - 10)^{1/2}/[8(n-4)(n-3)]$ follows by applying the formula (53) twice and using the orthogonality relation (48).

From (9) both first and second derivatives of χ_n are required. Again from (51) the first derivative is

$$\chi_n'(\xi) = \begin{cases} -3(1 - \xi^2)/4, & n = 1, \\ -(1 - \xi)(1 + 3\xi)/4, & n = 2, \\ 3(1 - \xi^2)/4, & n = 3, \\ -(1 + \xi)(1 - 3\xi)/4, & n = 4, \quad |\xi| \leqslant 1, \\ -\mu_n(1 - \xi^2)\left[(n - 1)\xi P_{n-5}^{2,2}(\xi) \right. \\ \qquad \left. - (n - 3)P_{n-6}^{2,2}(\xi)\right], & n \geqslant 5, \end{cases}$$

and the second derivative, using also the Jacobi differential equation (47),

$$
\chi_n^{(2)}(\xi) = \begin{cases}
3\xi/2\,, & n = 1\,, \\
(3\xi - 1)/2\,, & n = 2\,, \\
-3\xi/2\,, & n = 3\,, \\
(3\xi + 1)/2\,, & n = 4\,, \\
\mu_n\Big[(n-1)\big(2-(n-2)(1-\xi^2)\big)P_{n-5}^{2,2}(\xi) \\
\quad - 2(n-3)\xi P_{n-6}^{2,2}(\xi)\Big]\,, & n \geqslant 5\,,
\end{cases}
\quad |\xi| \leqslant 1\,.
$$

3   Implementation

With $\omega := (\sigma, \eta)$ and $u := (y, z)$ write the weak equation (9) for Problem III as

$$
\begin{aligned}
\mathcal{B}(\omega, u) &:= \mathcal{B}^{1,1}(\sigma, y) + \mathcal{B}^{1,2}(\sigma, z) + \mathcal{B}^{2,1}(\eta, y) + \mathcal{B}^{2,2}(\eta, z) \\
&= \mathcal{F}^1(\sigma) + \mathcal{F}^2(\eta) =: \mathcal{F}(\omega),
\end{aligned} \tag{10}
$$

where $\mathcal{B}^{1,1}(\sigma, y) = \int_{-1}^1 \lambda_1 y^{(2)} \sigma^{(2)} + \lambda_2 y' \sigma' + a_1 y \sigma \, dx$, $\mathcal{F}^1(\sigma) = \int_{-1}^1 f_1 \sigma \, dx$, and the remaining bilinear forms and linear functionals are defined in like manner.

For a C^0 expansion of z on the element e, the restriction of z to $[x_e, x_{e+1}]$ is

$$
z|_{[x_e, x_{e+1}]} := \sum_{j=1}^{p_e} z_j^e \psi_j^e\,, \quad \psi_j^e(x) := \psi_j\big((x - m_e)/h_e\big), \quad \text{for } x \in [x_e, x_{e+1}],
$$

and the element midpoint $m_e := (x_e + x_{e+1})/2$, element half-width $h_e := (x_{e+1} - x_e)/2$ and order of approximation p_e. For a C^1 expansion

$$
z|_{[x_e, x_{e+1}]} := \sum_{j=1}^{p_e} z_j^e \chi_j^e\,, \quad \chi_j^e(x) := \nu_j \chi_j\big((x - m_e)/h_e\big), \quad \text{for } x \in [x_e, x_{e+1}],
$$

and $\nu_j = h_e$ if $j = 2, 4$, $\nu_j = 1$ otherwise. The h_e factors arise from the continuity of the derivative at the mesh-points. The functions ψ_j^e and χ_j^e are extended to the whole domain $[-1, 1]$ by zero outside the element $[x_e, x_{e+1}]$.

The C^0 and C^1 element-wise expansions become global by enforcing the continuity conditions across the element boundaries. For a C^0 expansion, modes ψ_2^e and ψ_1^{e+1} are combined so $z_2^e = z_1^{e+1}$. For C^1 continuity, modes χ_3^e and χ_1^{e+1} are combined so $z_3^e = z_1^{e+1}$, and the ν_j factors enable modes χ_4^e and χ_2^{e+1} to be combined into a global C^1 function with $z_4^e = z_2^{e+1}$.

The weak-form (10) does not hold element-wise, there are extra terms from the integration by parts (9), when σ or η is not an interior mode. However, in the assembly of a global matrix system, corresponding element-wise blocks are overlapped and added together to enforce the continuity of the expansion, the extra terms cancelling in the process. Therefore, with a similar expansion for η as for z (C^0 or C^1) and C^1 expansions of order q_e for σ and y, the matrix form for the left and right side blocks of (10) can be written down. Let $\leftrightarrow$ denote corresponding left side and right side blocks in the assembly of the global matrix equation, then the matrix form of (10) on element e is

$$\boldsymbol{\sigma}^\mathsf{T}\boldsymbol{\mathcal{B}}^{1,1}\mathbf{y} + \boldsymbol{\sigma}^\mathsf{T}\boldsymbol{\mathcal{B}}^{1,2}\mathbf{z} + \boldsymbol{\eta}^\mathsf{T}\boldsymbol{\mathcal{B}}^{2,1}\mathbf{y} + \boldsymbol{\eta}^\mathsf{T}\boldsymbol{\mathcal{B}}^{2,2}\mathbf{z} \leftrightarrow \boldsymbol{\sigma}^\mathsf{T}\mathbf{f}^1 + \boldsymbol{\eta}^\mathsf{T}\mathbf{f}^2 , \qquad (11)$$

where $\boldsymbol{\sigma} := \left(\sigma_1^e, \ldots, \sigma_{q_e}^e\right)$, $\boldsymbol{\mathcal{B}}^{1,1} := \left[\mathcal{B}_{j,k}^{1,1}\right]$ where $\mathcal{B}_{j,k}^{1,1} := \mathcal{B}^{1,1}(\chi_j^e, \chi_k^e)$, $\mathbf{f}^1 := \left[f_k^1\right]$ where $f_k^1 := \mathcal{F}_1(\chi_k^e)$ and the remaining terms are defined similarly. For the first element there is no $\sigma_1, \sigma_2, \eta_1, y_1, y_2$ and z_1 whereas the last element has no $\sigma_3, \sigma_4, \eta_2, y_3, y_4$ and z_2 due to the essential boundary conditions and corresponding test and trial spaces.

To assemble the global matrix system, element-wise blocks (11) are reordered so that sub-blocks from corresponding parts of adjacent elements are added together (enforcing the continuity conditions). However, a further step, the static condensation of the interior modes, is performed to reduce the size, and improve conditioning of the resulting linear system. For example, for C^0 expansions of z and η (the changes for a C^1 expansion are minor), in MATLAB notation let $\mathbf{i}_1 := [1, 2, 4, 5, 7 : q_e + 2]$ and $\mathbf{i}_2 := [3, 6, q_e + 3 : q_e +$

p_e] be index vectors, and define vectors $\boldsymbol{\omega}$, $\mathbf{u}$, and $\mathbf{f}$ by

$$\boldsymbol{\omega}(\mathbf{i}_1) := \boldsymbol{\sigma}, \; \boldsymbol{\omega}(\mathbf{i}_2) := \boldsymbol{\eta} ; \quad \mathbf{u}(\mathbf{i}_1) := \mathbf{y}, \; \mathbf{u}(\mathbf{i}_2) := \mathbf{z} ; \quad \mathbf{f}(\mathbf{i}_1) := \mathbf{f}_1, \; \mathbf{f}(\mathbf{i}_2) := \mathbf{f}_2 .$$

Then (11) is written

$$\boldsymbol{\omega}^\mathsf{T} \boldsymbol{\mathcal{B}} \, \mathbf{u} \leftrightarrow \boldsymbol{\omega}^\mathsf{T} \mathbf{f} , \tag{12}$$

where

$$\boldsymbol{\mathcal{B}}(\mathbf{i}_1, \mathbf{i}_1) = \boldsymbol{\mathcal{B}}^{1,1} , \quad \boldsymbol{\mathcal{B}}(\mathbf{i}_1, \mathbf{i}_2) = \boldsymbol{\mathcal{B}}^{1,2} , \quad \boldsymbol{\mathcal{B}}(\mathbf{i}_2, \mathbf{i}_1) = \boldsymbol{\mathcal{B}}^{2,1} , \quad \boldsymbol{\mathcal{B}}(\mathbf{i}_2, \mathbf{i}_2) = \boldsymbol{\mathcal{B}}^{2,2} .$$

Thus the boundary (b) mode terms are placed before the interior (i) mode terms; for example

$$\boldsymbol{\omega} = (\underbrace{\sigma_1, \sigma_2, \eta_1, \sigma_3, \sigma_4, \eta_2}_{\text{b}}, \underbrace{\sigma_5, \ldots, \sigma_{q_e}, \eta_3, \ldots, \eta_{p_e}}_{\text{i}}).$$

Partitioning (12) in this way

$$\boldsymbol{\omega}^\mathsf{T} \begin{bmatrix} \boldsymbol{\mathcal{B}}_{bb} & \boldsymbol{\mathcal{B}}_{bi} \\ \boldsymbol{\mathcal{B}}_{ib} & \boldsymbol{\mathcal{B}}_{ii} \end{bmatrix} \begin{bmatrix} \mathbf{u}_b \\ \mathbf{u}_i \end{bmatrix} \leftrightarrow \boldsymbol{\omega}^\mathsf{T} \begin{bmatrix} \mathbf{f}_b \\ \mathbf{f}_i \end{bmatrix} , \tag{13}$$

then using $\boldsymbol{\mathcal{B}}_{ii}$ to zero out $\boldsymbol{\mathcal{B}}_{bi}$ the reduced system of boundary mode coefficients only is solved. Performing the elimination in (13) gives

$$\begin{bmatrix} \boldsymbol{\mathcal{B}}_{bb} - \boldsymbol{\mathcal{B}}_{bi} \boldsymbol{\mathcal{B}}_{ii}^{-1} \boldsymbol{\mathcal{B}}_{ib} & \mathbf{0} \\ \boldsymbol{\mathcal{B}}_{ib} & \boldsymbol{\mathcal{B}}_{ii} \end{bmatrix} \begin{bmatrix} \mathbf{u}_b \\ \mathbf{u}_i \end{bmatrix} \leftrightarrow \begin{bmatrix} \mathbf{f}_b - \boldsymbol{\mathcal{B}}_{bi} \boldsymbol{\mathcal{B}}_{ii}^{-1} \mathbf{f}_i \\ \mathbf{f}_i \end{bmatrix} ,$$

which decouples into Schur-complement boundary mode blocks. These are then assembled into a condensed matrix system, again overlapping and adding together corresponding parts from consecutive blocks to enforce continuity. The interior mode coefficients are found element-wise (and potentially in parallel) once the boundary mode coefficients are determined from the equality $\boldsymbol{\mathcal{B}}_{ib} \mathbf{u}_b + \boldsymbol{\mathcal{B}}_{ii} \mathbf{u}_i = \mathbf{f}_i$. This equality (rather than correspondence) follows from (9) when σ and η are interior modes.

To see how scaling issues typically arise let $\mathbf{B}\{i\} = \big[B\{i\}_{k,l}\big]$, where

$$B\{i\}_{k,l} := \int_{-1}^{1} \chi_k^{(i)} \chi_l^{(i)} \, d\xi \,.$$

Then on element e

$$\mathcal{B}^{1,1} = h_e \, \mathbf{D} \left(\frac{\lambda_1}{h_e^4} \mathbf{B}\{2\} + \frac{\lambda_2}{h_e^2} \mathbf{B}\{1\} + 2\mathbf{B}\{0\} \right) \mathbf{D} \,, \qquad (14)$$

where $\mathbf{D} = \big[D_{i,j}\big]$ is a diagonal matrix that effects the multiplication of modes 2 and 4 by h_e. So $D_{j,j} = h_e$ for $j = 2, 4$ and $D_{j,j} = 1$ otherwise, with the remaining $D_{i,j} = 0$. The h_e factor that appears at the front of the right side of (14) comes from integrating over $[x_e, x_{e+1}]$. Already apparent from (14) is the dependence on h_e. If h_e varies by orders of magnitude on the mesh, then scaling issues often arise. A method for handling scaling issues resulting from the multiplication of $\mathbf{D}$ was indicated by Farmer and Ivers [2]. The method used in this article is a diagonal scaling. Writing the assembled condensed matrix system as $\mathbf{Ku} = \mathbf{q}$, where $\mathbf{u}$ is a vector of all the boundary mode coefficients, the diagonal matrix $\mathbf{\Lambda}$ defined by $\mathbf{\Lambda}(i, i) = 2^{r_i}$ where r_i is the integer satisfying $\frac{1}{4} \leqslant 2^{2r_i} |\mathbf{K}(i, i)| < 1$, produces the scaled system

$$(\mathbf{\Lambda K \Lambda})(\mathbf{\Lambda}^{-1}\mathbf{u}) = \mathbf{\Lambda q} \,,$$

which is solved in the form

$$\mathbf{u} = \mathbf{\Lambda}(\mathbf{\Lambda K \Lambda})^{-1}(\mathbf{\Lambda q}).$$

The choice of the diagonal factors $\mathbf{\Lambda}(i, i)$ is to avoid roundoff while scaling the diagonal elements of $\mathbf{K}$ to be close to one in magnitude. This scaling seems particularly effective for the hp-FEM, certainly for the problems investigated in this article. The effects of diagonal scaling on the conditioning of the matrices $\mathbf{K}$ for a linear finite element approximation or h-version to an anisotropic diffusion problem has been studied by Kamenski, Huang & Wu [7].

4 Numerical Results

The robust approximation of boundary layer components is achieved in the *hp*-FEM for the model Problems I and II with $f(x) = 1$, by constructing elements proportional to the thickness of the boundary layer [11, 2]. Previous results [2] show that when dealing with fourth-order derivatives, a higher-order (C^1) expansion is preferable to a lower-order expansion (rewriting the fourth-order derivatives as coupled second-order terms).

In Subsection 4.1–4.2 Problems I and II are re-examined, this time to investigate the scaling of the boundary layer elements and the minimum degrees of freedom necessary for accurate recovery of the interior solution. A method for estimating the widths of the boundary layers is presented in Subsection 4.2 based on the approximation of the boundary layer components.

The accuracy of the interior solution when fewer boundary layer elements to boundary layer components are used, is discussed for Problems II and III in Subsection 4.2–4.3. Also how the solution deteriorates depends on the relative thicknesses of the boundary layer components.

Subsection 4.3 also discusses the choice of expansion for the variable z in Problem III, either C^0 or C^1.

The final Subsection 4.4 discusses a numerical method for determining the boundary layer widths and degree of approximation of the interior solution when these cannot be determined by analytical means. This method is then applied to the example presented in Subsection 4.3 without reference to the exact solution.

4.1 The second-order Problem I

The forcing is an oscillatory function $f = \cos(bx)$ where b is a fixed number chosen to produce boundary layer components; that is, b is not an odd

multiple of $\pi/2$. The exact solution is

$$y = \frac{1}{1 + b^2 d^2} \left[\cos(bx) - \cos(b) \frac{\cosh(x/d)}{\cosh(1/d)} \right]. \tag{15}$$

For $0 < d \ll 1$, a more useful expression is found by rewriting (15) as

$$
\begin{aligned}
y &= \frac{1}{1 + b^2 d^2} \left[\cos(bx) - \cos(b) \left(\frac{e^{-(1-x)/d} - e^{-(1+x)/d}}{1 + e^{-2/d}} \right) \right], \\
&= \frac{1}{1 + b^2 d^2} \left[\cos(bx) - \cos(b) \left(e^{-(1+x)/d} + e^{-(1-x)/d} \right) \left(1 - \mathcal{O}\left(e^{-2/d} \right) \right) \right].
\end{aligned}
\tag{16}
$$

The $\mathcal{O}(e^{-2/d})$ term may be neglected for the range of d investigated $(0 < d \leqslant 10^{-2})$ and for double precision, since $e^{-2/d} < 2^{-52}$, the unit roundoff.

Problem I is cast into the following weak-form:

$$\int_{-1}^{1} d^2 y' \phi' + y\phi \, dx = \int_{-1}^{1} f_1 \phi \, dx, \tag{17}$$

for $y, \phi \in H_0^1$, with energy norm defined from the left side of (17)

$$\|y\|_E^2 := \int_{-1}^{1} d^2 y'^2 + y^2 \, dx.$$

The solution of (17) is even, therefore the three-element symmetric mesh/order combination

$$x = (-1, -1 + sd, 1 - sd, 1), \quad p = (n_1, n_0, n_1),$$

could be used, with only the even interior modes included in the computation for the central element. However, for simplicity in calculating the error in the energy-norm, the four-element mesh/order combination

$$x = (-1, -1 + sd, 0, 1 - sd, 1), \quad p = (n_1, n_0, n_0, n_1), \tag{18}$$

was chosen. After checking the symmetry of the computed solution $\tilde{y}$, the element-wise errors in the energy-norm of the first and second elements, defined by

$$\|y - \tilde{y}\|^2_{E,e} := \int_{x_e}^{x_{e+1}} d^2(y^{(1)} - \tilde{y}^{(1)})^2 + (y - \tilde{y})^2 \, dx \,, \tag{19}$$

were plotted versus s to investigate the scaling of the boundary element width.

Some care is needed in calculating (19). Farmer & Ivers [2] used a simplification exploiting the weak form of the solution that is not used here due to its poor accuracy. To briefly explain, the error in the energy norm is

$$\|y - \tilde{y}\|^2_E = \int_{-1}^{1} d^2(y^{(1)} - \tilde{y}^{(1)})^2 + (y - \tilde{y})^2 \, dx \tag{20}$$

$$= \int_{-1}^{1} d^2(y^{(1)})^2 + y^2 \, dx + \int_{-1}^{1} d^2(\tilde{y}^{(1)})^2 + \tilde{y}^2 \, dx$$

$$- 2 \int_{-1}^{1} d^2 y^{(1)} \tilde{y}^{(1)} + y\tilde{y} \, dx \,. \tag{21}$$

The first two integrals in (21) may be simplified using the weak-form (17), and the last with an integration by parts and the original differential equation (1) to give

$$\|y - \tilde{y}\|^2_E = \int_{-1}^{1} fy \, dx - \int_{-1}^{1} f\tilde{y} \, dx \,. \tag{22}$$

The problem with calculating each integral separately in (22) then subtracting, is that the best accuracy that may be achieved in fixed-precision arithmetic is the order of the unit roundoff. Whereas the original expression in (20) may be far smaller, approaching the order of the square of the unit roundoff. There is no discrepancy mathematically here, except that the calculation is in fixed precision, and care has to be exercised so that catastrophic cancellation does not occur.

Two other sources of inaccuracy arise in computing the errors in (19). The first and most straightforward one is when $x \approx -1$; it is better to calculate

$1 + x$ directly in (16) to avoid catastrophic cancellation. For example in the first element, given the vector of Gauss–Lobatto quadrature nodes $\mathbf{N}$, calculate $\mathbf{x}_1 = h_1(\mathbf{N} + 1)$ and then $\mathbf{x} = \mathbf{x}_1 - 1$ rather than the reverse. In the second element $\mathbf{x}_1 = 2h_1 + h_2(\mathbf{N} + 1)$. There is potentially a loss in significant figures for the nodes close to -1 as the maximal quadrature order (assumed fixed) becomes large. One way to avoid this loss is to calculate the nodes in higher precision, for example, quadruple precision in FORTRAN: the nodes are calculated in double precision from the eigenvalues of a symmetric tridiagonal matrix [3], then used as starting values in a quadruple precision Newton–Raphson method. The second source of inaccuracy comes from the cancellation of terms in the first boundary layer element, again as x approaches the boundary at -1. From (16) writing $x_1 = x + 1$ for $x < 0$,

$$
\begin{aligned}
y &= \frac{1}{1 + b^2 d^2} \left[\cos(b(x_1 - 1)) - \cos(b)\left(e^{-x_1} + \mathcal{O}\left(e^{-2/d}\right)\right) \right], \\
&= \frac{1}{1 + b^2 d^2} \left[\cos(b)\left(\cos(bx_1) - e^{-x_1}\right) + \sin(bx_1)\sin(b) - \mathcal{O}\left(e^{-2/d}\right) \right].
\end{aligned}
\tag{23}
$$

In (23) the first terms in the series expansions of $\cos(bx_1)$ and e^{-x_1} cancel as x_1 approaches zero. If this is not accounted for, then catastrophic cancellation again occurs. Thus, rewriting (23)

$$
\begin{aligned}
y = \frac{1}{1 + b^2 d^2} \Big\{ &\cos(b)\left[(\cos(bx_1) - 1) + e^{-x_1}\left(e^{x_1} - 1\right)\right] \\
&+ \sin(bx_1)\sin(b) - \mathcal{O}\left(e^{-2/d}\right) \Big\}.
\end{aligned}
\tag{24}
$$

The rewrite in (24) is trivial mathematically; however, it does show how to accurately compute y when x_1 is close to zero. By computing the series $(\cos(bx_1) - 1)$ and $(e^{x_1} - 1)$, the terms that cancel are removed from the calculation. For $(\cos(bx_1) - 1)$, grouping pairs of terms produces a series of purely negative terms, whereas $(e^{x_1} - 1)$ contains purely positive terms. Both series $(\cos(bx_1) - 1)$ and $(e^{x_1} - 1)$ for $0 < x_1 < 1$, are recursively calculated, adding terms from the smallest in magnitude to the largest in

Figure 1: Problem I: Convergence results for $d = 10^{-6}$, $n_1 = 19$, $n_0 = 12$.
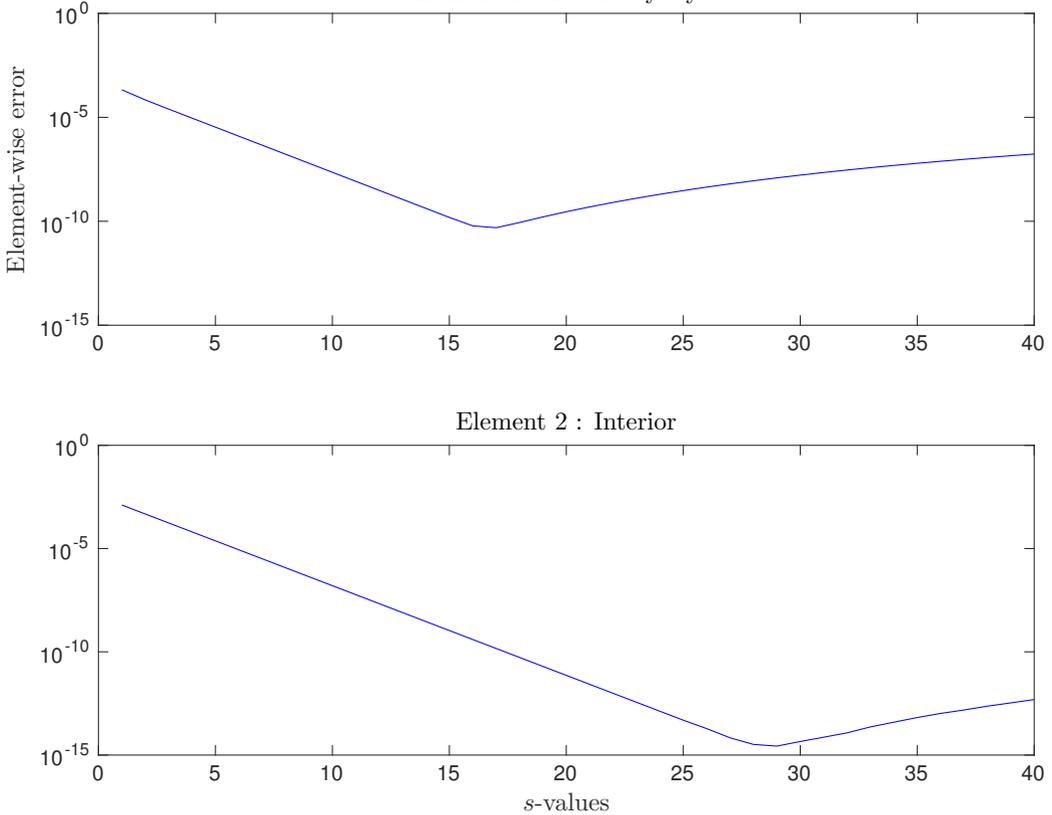
magnitude, using a Taylor series estimate for the tail as a stopping criterion. This approach is used for the computation of the exact solution throughout the three problems investigated, to avoid catastrophic cancellation in the boundary layer.

On applying the hp-FEM varying d, the results indicate that once convergence is reached in the interior, increasing the parameter n_1 decreased the sensitivity of the method with respect to the parameter s. Throughout this article results are taken as converged when the error in the relevant norm of the left-most

Figure 2: Problem I: Increase in convergence interval in s; $d = 10^{-6}$, $n_1 = 25$, $n_2 = 12$.
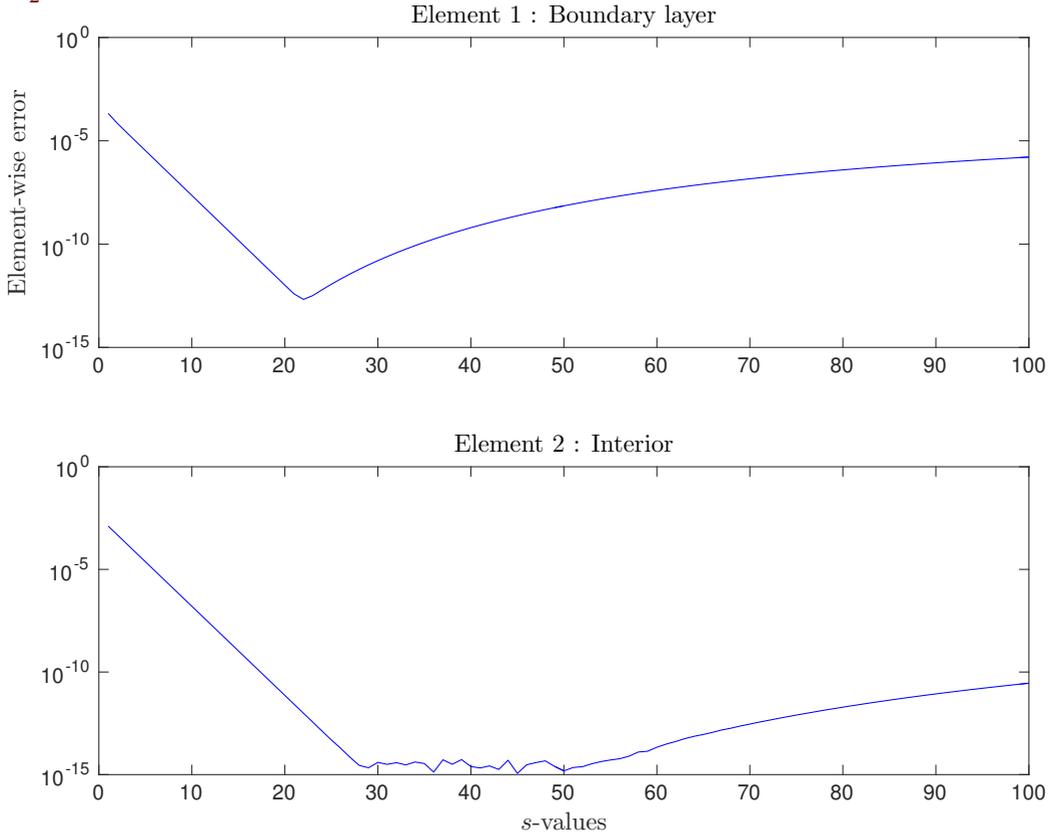
interior element falls below 9×10^{-14}. Taking $b = \pi/3$ and $d = 10^{-6}$ in Problem I, with the four-element mesh/order combination (18), convergence was achieved with $n_1 = 19$, $n_0 = 12$, and $s = 29$ as shown in Figure 1. The sensitivity with respect to s is immediately apparent. Also, the boundary layer components do not have to be resolved to full accuracy to produce full accuracy in the interior. On increasing n_1 to 25, there is a larger range of s for which convergence is achieved, Figure 2. Increasing the parameter n_1 further, eventually increases the convergence interval with respect to s in the

Table 1: Convergence values and intervals

d	$n_1^{\min} / s^{\min}$	$n_1^{\max} / I_s$
10^{-9}	$13 / 25$	$27 / [25, 100]$
10^{-8}	$14 / 25$	$29 / [25, 100]$
10^{-7}	$16 / 25$	$31 / [25, 100]$
10^{-6}	$17 / 25$	$33 / [25, 100]$
10^{-5}	$18 / 25$	$35 / [25, 100]$
10^{-4}	$19 / 25$	$37 / [25, 100]$
10^{-3}	$20 / 26$	$39 / [26, 100]$
10^{-2}	$21 / 26$	$41 / [26, 99]$

first element as well. However, it is the interior solution that is of primary concern. Experimenting for a range of d with b fixed at $\pi/3$ (larger b required choosing larger n_0), produced the results in Table 1 of the parameters n_1 and s for convergence of the interior. The second column of Table 1 lists the minimum values of $n_1 = n_1^{\min}$ with its corresponding value of $s = s^{\min}$ for which convergence is achieved in the left-most interior element. The third column lists the minimum value $n_1 = n_1^{\max}$ for which convergence is attained at the corresponding range of $s \in I_s$. The value is 99 for $d = 10^{-2}$ as s must be less than 100 to produce a valid mesh. The total degrees of freedom is given by $N = 2(n_1 + n_0) - 5$. Table 1 shows that the n_1 values decrease as d is *decreased*: less modes are required for a thinner boundary layer. This counter-intuitive result does seem to be a feature of the method and is consistent with bounds on the energy norm error derived by Schwab [11].

4.2   The fourth-order Problem II

For the fourth-order Problem II the exact solution is

$$y = \frac{1}{(1 + b^2 d^2)(1 + a^2 b^2 d^2)} \left[\cos(bx) + A_1 \frac{\cosh(x/d)}{\cosh(1/d)} + A_2 \frac{\cosh(x/(ad))}{\cosh(1/(ad))} \right],$$

where the A_i are found from the boundary conditions (4) as

$$A_1 = -\left[\frac{abd\sin(b) + \cos(b)\tanh(1/(ad))}{\tanh(1/(ad)) - a\tanh(1/d)}\right], \tag{25}$$

$$A_2 = a\left[\frac{bd\sin(b) + \cos(b)\tanh(1/d)}{\tanh(1/(ad)) - a\tanh(1/d)}\right]. \tag{26}$$

The solution is even in x and takes the form for $x < 0$ and $0 < d \ll 1$

$$y = \left[(1 + b^2d^2)(1 + a^2b^2d^2)\right]^{-1}\left[\cos(bx)\right.$$
$$\left. + (1-a)^{-1}\left(\tilde{A}_1 e^{-(1+x)/d} + \tilde{A}_2 e^{-(1+x)/(ad)} + \mathcal{O}\left(e^{-2/d}\right)\right)\right],$$

where the $\tilde{A}_i$ follows from A_i ($i = 1, 2$) by replacing the tanh terms with 1 in (25)–(26). Again the $\mathcal{O}(e^{-2/d})$ term is discarded for $0 < d \leqslant 10^{-2}$.

The natural weak-form for Problem II, chosen for having a corresponding energy norm is

$$\int_{-1}^{1} a^2 d^4 y^{(2)}\phi^{(2)} + (1+a^2)d^2 y^{(1)}\phi^{(1)} + y\phi\, dx = \int_{-1}^{1} f_2\phi\, dx, \tag{27}$$

where y and ϕ are in H_0^2. The corresponding energy norm to (27) is

$$\|y\|_E^2 := \int_{-1}^{1} a^2 d^4 (y^{(2)})^2 + (1+a^2)d^2 {y^{(1)}}^2 + y^2\, dx. \tag{28}$$

The mesh/order combinations that were tested are the four-element version

$$\mathbf{x} = (-1, -1 + s_1 d, 0, 1 - s_1 d, 1), \quad \mathbf{p} = (n_1, n_0, n_0, n_1), \tag{29}$$

and the six-element version

$$\mathbf{x} = (-1, -1 + s_2 ad, -1 + s_1 d, 0, 1 - s_1 d, 1 - s_2 ad, 1),$$
$$\mathbf{p} = (n_2, n_1, n_0, n_0, n_1, n_2). \tag{30}$$

The scalings s_1 and s_2 of the boundary layer elements are again investigated for the accurate recovery of the interior solution. This is measured in the element-wise energy norm

$$\|y\|_{E,e}^2 := \int_{x_e}^{x_{e+1}} a^2 d^4 (y^{(2)})^2 + (1 + a^2) d^2 {y^{(1)}}^2 + y^2 \, dx. \qquad (31)$$

For the four-element version (28), $\|y - \tilde{y}\|_{E,2}$ is measured, and for the six-element version (29) $\|y - \tilde{y}\|_{E,3}$ measured, from (31).

Experimenting with the different mesh/order combinations and varying the parameters involved shows that convergence of the interior is still achieved with the four-element version (29). As the relative boundary layer thickness a decreases, n_1 must be increased for convergence. Also the maximum accuracy that is achieved for the interior drops once $a < 10^{-2}$. Illustrated in Figure 3 are convergence results for the four-element mesh with $a = 10^{-6}$ and $b = \pi/3$ (again increasing b required increasing the number of modes in the interior, n_0). The maximum accuracy is around 10^{-9} for the interior.

For the six-element mesh, full accuracy of the interior is reached for a range of the parameter a (including 10^{-6}), see Figure 4. In both Figure 3 and 4 the minimum value for the interior is marked with a circle.

Increasing n_1 for the four-element, and n_1 and n_2 for the six-element versions, again increases the convergence interval with respect to s_1 and s_2. Table 2 shows convergence results for the four-element and six-element versions with $a = 0.1$. Numerical tests show full convergence can be achieved for this value of a with the four-element version. For each d value in Table 2 the top result is for the four-element version and the lower values for the six-element. For example, when $d = 10^{-6}$ the minimum n_1 for convergence of the interior using the four-element version is $n_1 = n_1^{min} = 48$ with corresponding value $s_1 = s_1^{min} = 26$. When $n_1 = n_1^{max} = 88$ the convergence interval is for $s_1 \in I_{s_1} = [26, 100]$. The lower values for $d = 10^{-6}$ show that for the six-element version, the minimum values for convergence of the interior is given by $n_1 = n_1^{min} = 20$ and $n_2 = n_2^{min} = 10$ with corresponding values $s_1 = s_1^{min} = 28$

Figure 3: Problem II: Four element; $a = 10^{-6}$, $d = 10^{-6}$, $n_1 = 40$, $n_2 = 12$.

Element 1 : Boundary layer

Element 2 : Interior

$(16,\ 1.7024\text{e-}09)$

s_1-values

Element-wise error
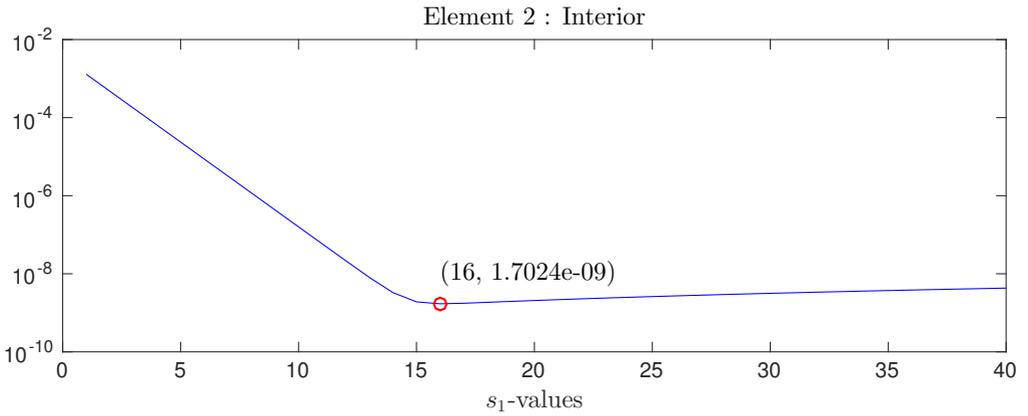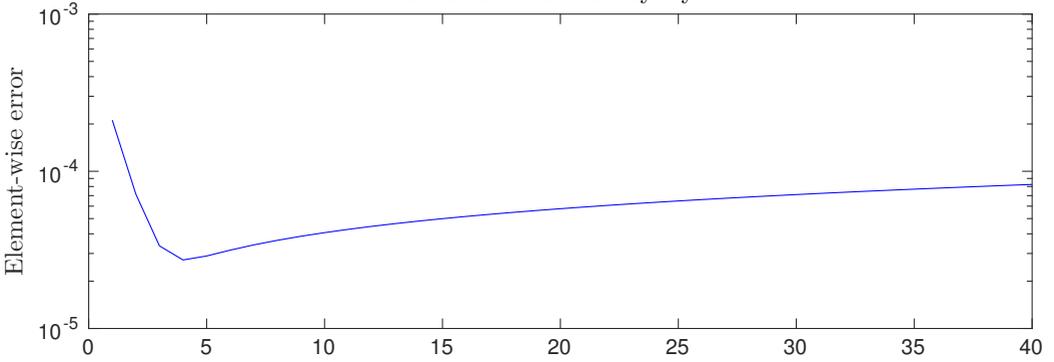
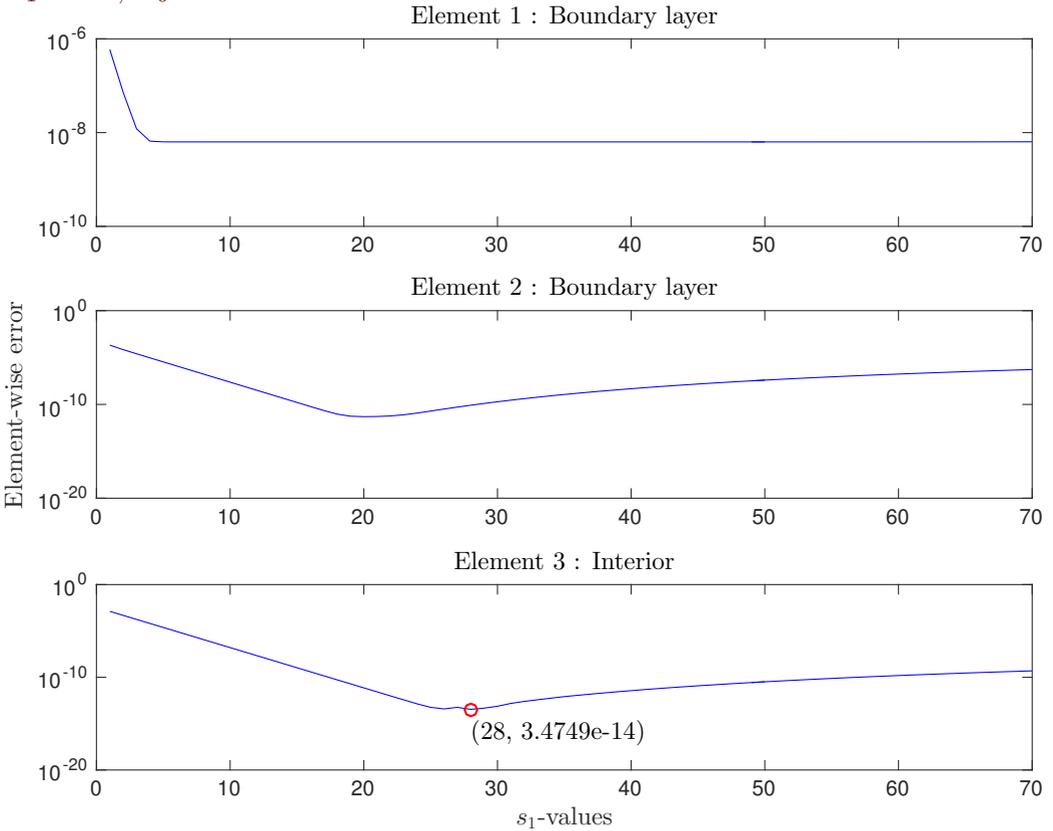Figure 4: Problem II: Six element; $s_2 = 15$, $a = 10^{-6}$, $d = 10^{-6}$, $n_2 = 10$, $n_1 = 23$, $n_0 = 12$.

Table 2: Four and six element convergence values and intervals, $a = 0.1$.

d	n_1^{min} / s_1^{min}	n_2^{min} / s_2^{min}	n_1^{max} / I_{s_1}	n_2^{max} / I_{s_2}
10^{-8}	40 / 26		70 / [26, 100]	
	17 / 28	8 / 7	31 / [28, 100]	21 / [7, 100]
10^{-7}	44 / 27		80 / [27, 100]	
	18 / 28	8 / 8	35 / [28, 100]	23 / [8, 100]
10^{-6}	48 / 26		88 / [26, 100]	
	20 / 28	10 / 13	39 / [28, 100]	25 / [13, 100]
10^{-5}	52 / 26		96 / [26, 100]	
	22 / 28	12 / 13	43 / [28, 100]	27 / [13, 100]
10^{-4}	56 / 26		102 / [26, 100]	
	22 / 28	12 / 16	47 / [28, 100]	29 / [16, 100]
10^{-3}	60 / 27		108 / [27, 100]	
	25 / 28	15 / 17	51 / [28, 100]	31 / [17, 100]
10^{-2}	64 / 27		114 / [27, 100]	
	26 / 28	16 / 18	55 / [28, 99]	33 / [18, 100]

and $s_2 = s_2^{min} = 13$. When $n_1 = n_1^{max} = 39$ and $n_2 = n_2^{max} = 25$, convergence holds for $s_1 \in I_{s_1} = [28, 100]$ and $s_2 \in I_{s_2} = [13, 100]$.

Simple estimates of s_1, s_2, n_1, and n_2, are found by considering the leading order approximations of the boundary layer components. To describe the method, consider the approximation of $f = e^{-x/d}$ on the interval $[sd, 1]$. Given s and the order of approximation N, the least-squares minimiser $\mathcal{P}_N(f)$ may be found. Further, the smallest value of s may be found for which the inequality

$$\int_{sd}^{1} (\mathcal{P}_N(f) - f)^2 \, dx \leqslant \varepsilon^2, \qquad (32)$$

holds for some prescribed ε. Initially N will be the fixed-order approximation of the interior (or left-most interior element for the model problems in this

article). Since f is the slowest decaying boundary layer component, if (32) is satisfied, then faster decaying components will either be negligible or satisfy a similar inequality (replacing f with them). Next find the smallest N' for which

$$\int_0^{sd} (\mathcal{P}_{N'}(f) - f)^2 \, dx \leqslant \varepsilon'^2, \tag{33}$$

where ε' is given, and then solve for the smallest s' for which

$$\int_{s'd'}^{sd} (\mathcal{P}_{N'}(f') - f')^2 \, dx \leqslant \varepsilon'^2, \tag{34}$$

where $f' = e^{-x/d'}$ is the second-most slowest decaying boundary layer component. Again, faster decaying components are negligible or satisfy a similar inequality to (34), whereas the slower decaying component satisfies (33) and so satisfies (34) as well (with f' replaced by f). Continue in this manner until all boundary layer components are resolved. For the fourth-order problem considered in this section the final step is to find the smallest N'' for a given ε'', at which

$$\int_0^{s'd'} (\mathcal{P}_{N''}(f') - f')^2 \, dx \leqslant \varepsilon''^2. \tag{35}$$

Slower decaying components will not vary much on this interval so (35) (with f' replaced by f) will be satisfied for them. This results in the estimates $s_1 = s$, $s_2 = s'$, $n_2 = N''$ and $n_1 = N'$.

To use this method for Problem III in the next Subsection 4.3, further steps are required. Firstly find the smallest s'' for which

$$\int_{s''d''}^{s'd'} (\mathcal{P}_{N''}(f'') - f'')^2 \, dx \leqslant \varepsilon''^2, \tag{36}$$

then find N''' for which, for given ε''',

$$\int_0^{s''d''} (\mathcal{P}_{N'''}(f'') - f'')^2 \, dx \leqslant \varepsilon'''^2. \tag{37}$$

Table 3: Estimated n_i and s_j for $a = 0.1$.

d	n_2	n_1	s_2	s_1
10^{-8}	15	25	14.7	29.3
10^{-7}	16	26	15.8	29.3
10^{-6}	18	27	16.9	29.6
10^{-5}	19	28	17.7	29.6
10^{-4}	20	29	18.6	29.7
10^{-3}	21	30	19.6	30.8
10^{-2}	22	31	20.6	30.0

Values for the ε, ε', ε'' and ε''' occurring on the right-hand sides of (32)–(37) determine how accurately the boundary layer components are approximated. For example in (32), if the interior solution is $\mathcal{O}(1)$ then choosing $\varepsilon^2 = \mathbf{u}^2$ where $\mathbf{u}$ is small, say 10^{-15}, gives an estimate for s_1. In double precision arithmetic it is not possible to choose the remaining εs this small, due to poor convergence of the least-squares minimiser at the endpoints of the approximation interval. However, choosing $\varepsilon' = 10^{-12}$ and $\varepsilon'' = 10^{-10}$ suffice to produce estimates for s_2, n_1 and n_2, in the six-element version (30) of Problem II. Numerical results indicate that the L^2 errors in elements 1 and 2 for the converged results in Table 2 are larger than these values for ε'' and ε'. The estimated values for this choice of ε' and ε'' are in Table 3: s_1 and s_2 are rounded to three significant figures. The value for s_1 is slightly higher than the result from numerical tests. Closer values are found if the method is applied to the interval $[sd, A]$ where $A < 1$ and $e^{-A/d}$ can be safely taken as zero. In double precision arithmetic certainly $e^{-A/d} = 2^{-1074}$. In practice, without prior knowledge of the εs this method is difficult to apply. For the calibration Subsection 4.4 the s values are taken so that $e^{-s} \leqslant 2^{-52}$, the unit roundoff in double precision.

4.3   The mixed-order coupled Problem III

The particular form for Problem III,

$$a_1^2 d^4 y^{(4)} - (1 + a_1^2) d^2 y^{(2)} + 2y + z = \cos(bx), \tag{38}$$

$$- a_1^2 a_2^2 d^2 z^{(2)} + z - a_1^2 a_2^2 d^2 y^{(2)} + y = x \sin(bx), \tag{39}$$

$$z(\pm 1) = y(\pm 1) = y'(\pm 1) = 0, \tag{40}$$

is studied for $0 < d \ll 1$. These choices of the constants λ_i and c_j simplify the resulting equation in y once z is eliminated. The parameters a_1 and a_2 are chosen to produce distinct boundary layer components: $0 < a_1, a_2, a_1 a_2 \neq 1$; and $a_1 d, a_1 a_2 d \ll 1$.

Eliminating z using (38), the differential equation (39) in y becomes

$$(a_1^2 a_2^2 d^2 D^2 - 1)(a_1^2 d^2 D^2 - 1)(d^2 D^2 - 1)y = x \sin(bx) - (1 + a_1^2 a_2^2 d^2 b^2) \cos(bx) \tag{41}$$

which is solved exactly, the boundary layer components found from (40). The solution for y and z is given by equations (54) and (55), respectively, in Appendix B. The weak-form of (38)–(40) is given from (9) as

$$\int_{-1}^{1} a_1^2 d^4 y^{(2)} \sigma^{(2)} + (1 + a_1^2) d^2 y' \sigma' + 2y\sigma + z\sigma + a_1^2 a_2^2 d^2 z' \eta' + z\eta$$

$$+ a_1^2 a_2^2 d^2 y' \eta' + y\eta \, dx = \int_{-1}^{1} \cos(bx)\sigma + x \sin(bx)\eta \, dx. \tag{42}$$

As there are now three boundary layer components, the method can be tested on three mesh/order combinations. Let $\mathbf{v}$ be the vector $(1, a_1, a_1 a_2)$ sorted in descending order. Then the three mesh/order combinations studied are: the four-element version

$$\mathbf{x} = (-1, -1 + s_1 v_1 d, \ 0 \ , 1 - s_1 v_1 d, \ 1 \ ), \quad \mathbf{p} = (n_1, n_0, n_0, n_1);$$

the six-element version

$$\mathbf{x} = (-1, -1 + s_2 v_2 d, -1 + s_1 v_1 d, \ 0 \ , 1 - s_1 v_1 d, 1 - s_2 v_2 d, \ 1 \ ),$$

$$\mathbf{p} = (n_2, n_1, n_0, n_0, n_1, n_2);$$

and the eight element version

$$\mathbf{x}(1{:}5) = (-1, -1 + s_3 v_3 d, -1 + s_2 v_2 d, -1 + s_1 v_1 d, \ 0 \ ),$$
$$\mathbf{p} = (n_3, n_2, n_1, n_0, n_0, n_1, n_2, n_3). \tag{43}$$

In (43) the remainder of $\mathbf{x}$ is to be inferred by symmetry. There is no corresponding energy norm so errors must be calculated in another norm. The L^2 norm was used (on each element) to measure the errors in the calculated y, $y^{(1)}$ and z. With three boundary layer components there are more parameters to investigate. Also it is important not to place too much emphasis on behaviour that may be special to particular choices of the parameters. There is also a choice as to whether to use a C^0 expansion for z and η and a C^1 expansion for y and σ (referred to as C01), or C^1 expansions throughout (C11). Experimenting with a range of a_1, a_2 and d shows little difference between these two choices. The C01 has slightly better convergence properties, whereas the C11 is slightly more accurate, when converged. Figure 5–7 compare the convergence for the (left-most) interior element for the four, six and eight element versions with $a_1 = 10^{-2}$, $a_2 = 10^4$, and $d = 10^{-6}$ (these are typical convergence plots observed over a range of a_1, a_2 and d). The value $b = \pi/3$ and values for s_1, s_2 and s_3 were found by the method in Subsection 4.2. This method is applied with increasing values for the εs, to reflect the observed numerical behaviour: that the boundary layers require less accuracy for elements further from the interior, to produce accurate results in the interior. In particular the values $\varepsilon = 10^{-15}$, $\varepsilon' = 10^{-12}$, $\varepsilon'' = 10^{-10}$, and $\varepsilon''' = 10^{-8}$ were used. The interior was resolved with fixed-order (twelve) elements. In all plots the order of approximation for y and z is the same in each boundary element. That is, plots are taken over a range of n, where $n_3 = n_2 = n_1 = n$.

For the four element version we anticipate from the results in the previous Subsection 4.2, that full convergence can be reached when the relative thickness of the smallest to largest boundary layer is large enough, in Subsection 4.2 the value was around 10^{-2}. However, this is expected to deteriorate until full convergence is no longer possible once these thicknesses decrease. For

Figure 5: Problem III: Four element; $\mathfrak{a}_1 = 10^{-2}$, $\mathfrak{a}_2 = 10^4$, $\mathfrak{d} = 10^{-6}$, $\mathfrak{s}_1 = 29.7$.
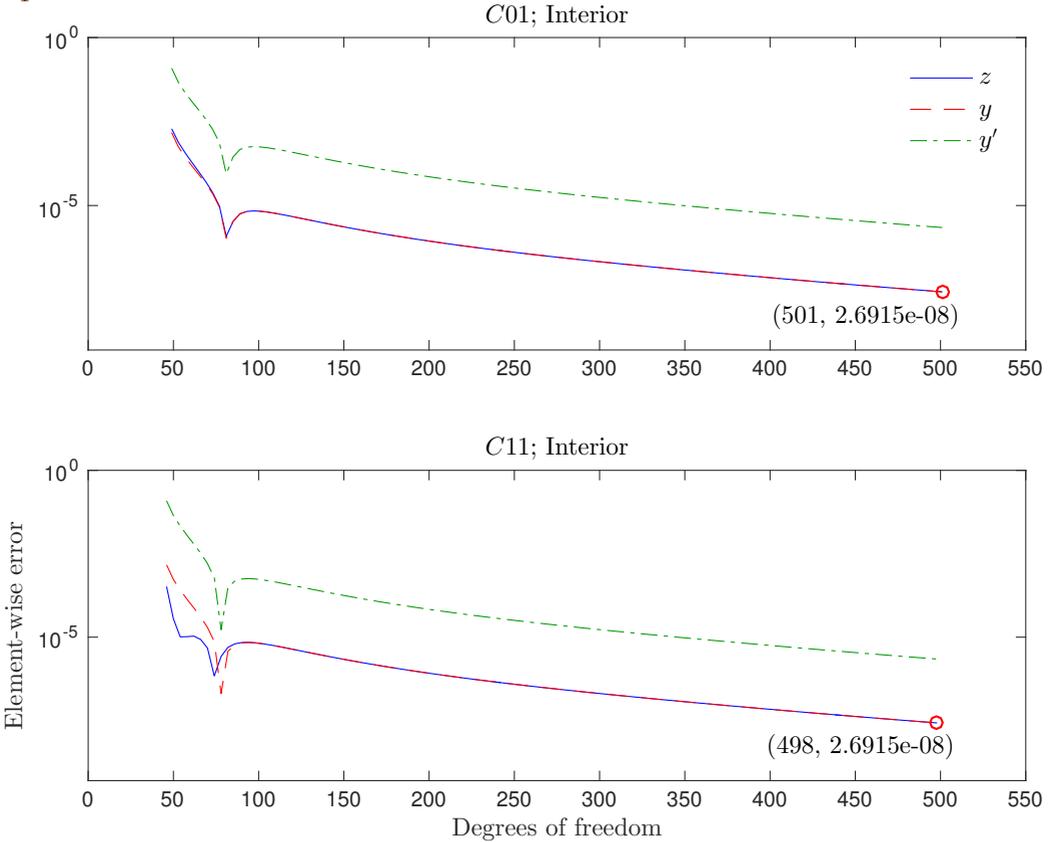
the example presented, the thickness of the layers is 10^{-4}, 10^{-6} and 10^{-7}. Figure 5 shows the convergence rate is much slower for the four element version compared with the six element, Figure 6, and eight element, Figure 7. Also the accuracy achieved for about 500 degrees of freedom is much worse, as indicated by the circle on each plot. However, numerical tests do confirm that if the relative thickness of the smallest to largest boundary layer component is greater than 10^{-2}, then full convergence is achieved for the four element version (although for much higher degrees of freedom than the six and eight element versions).

For the six element version, Figure 6, convergence is faster than the four-element. The observation made for the four element version, about full convergence being achieved when the relative thicknesses of the inner boundary layers to the outer ones is large enough, extends to the six element version. For the six element version the convergence results deteriorate in the maximum accuracy attained, once the relative boundary layer thickness of the second slowest decaying component to the fastest decaying component drops below 10^{-2}.

For the eight element version, Figure 7, full convergence is achieved well within the degrees of freedom range. The strategy that was employed for the order of approximation, taking the same order for each of the boundary elements, is not the most efficient one. The results from Subsection 4.2 suggest that convergence is achieved with fewer degrees of freedom by decreasing the number of modes as the elements get further from the interior. This is indeed the case. Taking $n_3 = 10$, $n_2 = 17$, and $n_1 = 27$ produces full convergence for the C01 version for only 237 degrees of freedom (compared to 333). These values for the n_j were the corresponding values returned from the method in Subsection 4.2 for the values $\varepsilon = 10^{-15}$, $\varepsilon' = 10^{-12}$, $\varepsilon'' = 10^{-10}$ and $\varepsilon''' = 10^{-8}$.

Figure 8 plots the convergence results for the six element version with $a_1 = 10^{-2}$, $a_2 = 10^{-1}$, and $d = 10^{-6}$, and shows y converging faster than z for both the C01 and C11 versions. However, this is behaviour that is special to

Figure 6: Problem III: Six element; $a_1 = 10^{-2}$, $a_2 = 10^4$, $d = 10^{-6}$, $s_1 = 29.7$, $s_2 = 20.2$.

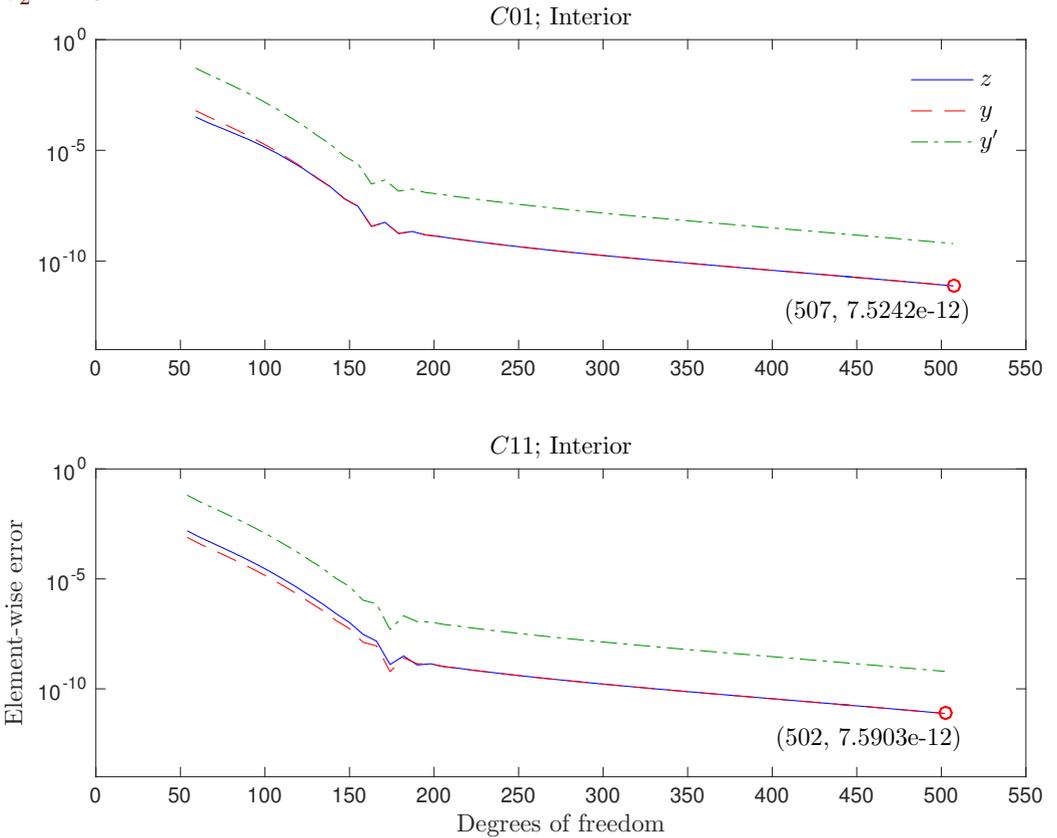Figure 7: Problem III: Eight element; $\mathfrak{a}_1 = 10^{-2}$, $\mathfrak{a}_2 = 10^4$, $\mathfrak{d} = 10^{-6}$, $\mathfrak{s}_1 = 29.7$, $\mathfrak{s}_2 = 20.2$, $\mathfrak{s}_3 = 13.3$.
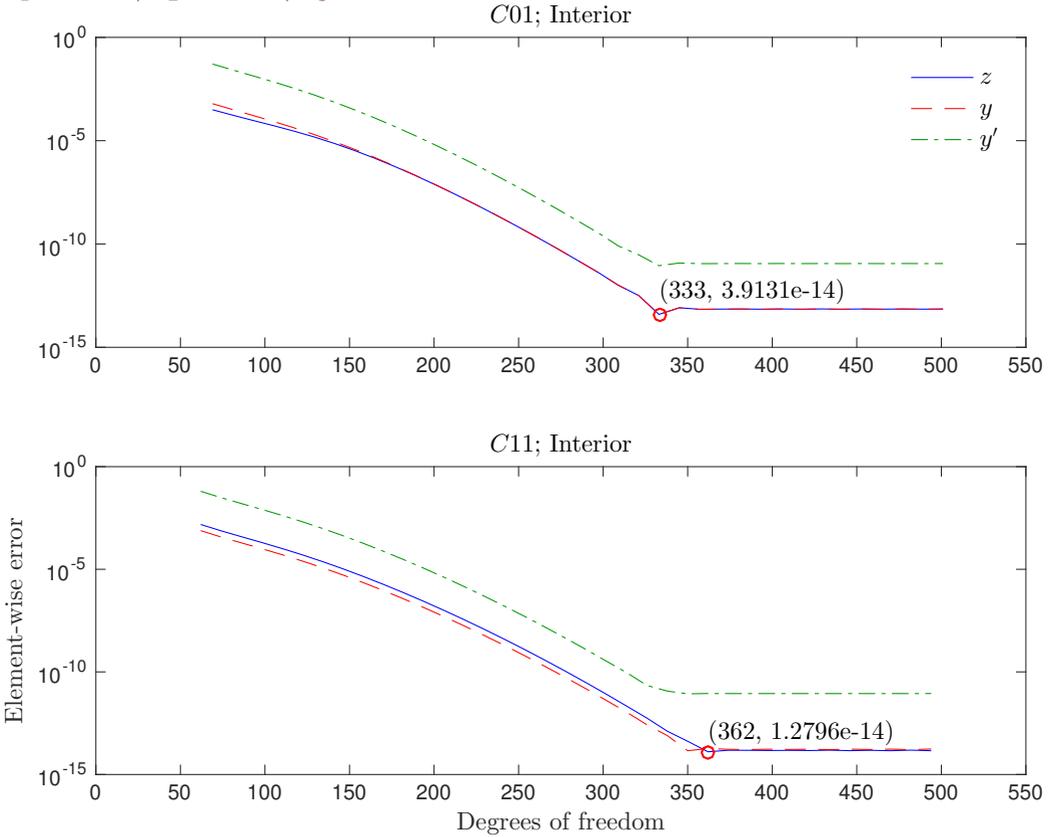
Figure 8: Problem III: Six element; $\mathfrak{a}_1 = 10^{-2}$, $\mathfrak{a}_2 = 10^{-1}$, $\mathfrak{d} = 10^{-6}$, $\mathfrak{s}_1 = 29.6$, $\mathfrak{s}_2 = 18$, $\mathfrak{s}_3 = 10.8$.
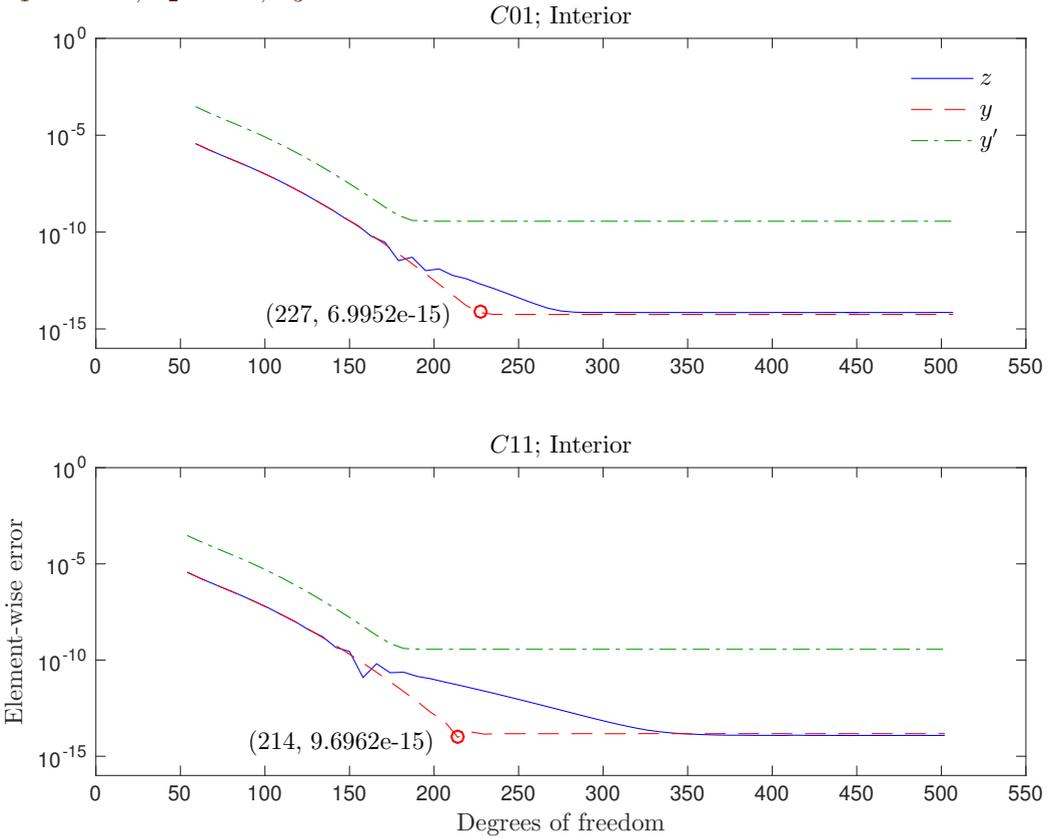
the choice of parameters a_1, a_2 and d. For this choice of parameters a_1, a_2 and d the exact solution for y is not strongly dependent on the resolution of the fastest decaying component. To three significant figures $A_3 = 8.74 \times 10^{-9}$ compared to $A_1 = 3.69 \times 10^{-1}$ and $A_2 = -3.69 \times 10^{-3}$. Therefore it is important not to place too much emphasis on such special behaviour. The poor result for $y^{(1)}$ is due to poor convergence at the endpoint $-1 + s_1 d$. Increasing s_1 improves this result. The method is most importantly robust for a range of the parameters a_1, a_2 and d once there are enough elements and modes to approximate the boundary layer components and interior solution. To apply the method these boundary element widths have to be known, or found numerically. Also the number of modes required for the interior solution has to be determined. This is the topic of the next section.

4.4   Calibration of the method in applications

The previous Subsection 4.1–4.3 are concerned with testing the *hp*-FEM for boundary layer problems where the exact solution was known. This section is concerned with how to implement the method when the exact solution is not known, as is generally the case. If the widths of the boundary layers are known, or estimated somehow analytically, then the method may be applied directly. There may need to be fine-tuning to determine the scaling of the boundary layer widths and orders of approximation of the boundary layer elements and interior element. If the problem is sufficiently complicated, then the widths may need to be determined numerically. For this latter case, consider the second-order Problem I.

Convergence results for the one-element *p*-version with $d = 10^{-1}$, 10^{-2} and 10^{-3} are plotted in Figure 9. The main feature of this plot being the different convergence rates. Firstly, for $d = 10^{-1}$ the rate of convergence is in the super-exponential rate: convergence is reached with relatively few degrees of freedom compared with the other two plots. For $d = 10^{-2}$ convergence is still in the exponential range, but is only reached for significantly higher

Figure 9: Problem I: One-element, p-version; $\mathbf{d} = 10^{-1}$, 10^{-2} and 10^{-3}.
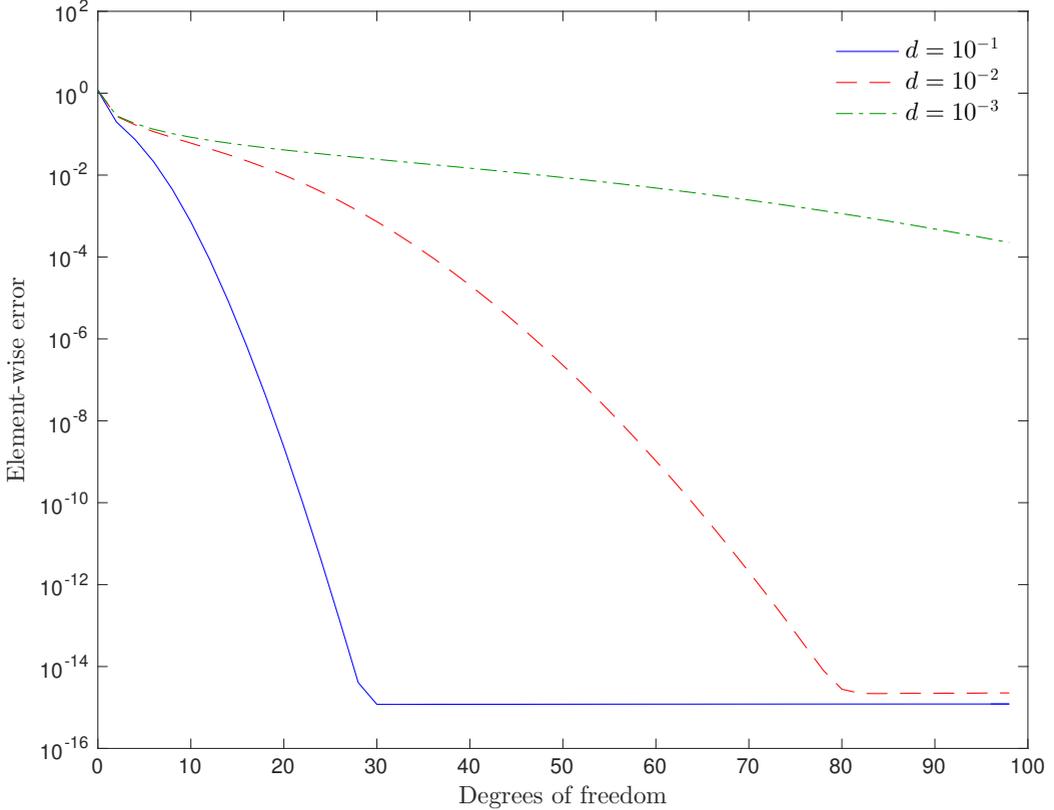
degrees of freedom, roughly 8/3 times the degrees of freedom for $\mathbf{d} = 10^{-1}$. Finally, when $\mathbf{d} = 10^{-3}$ convergence deteriorates to the point where spectral convergence is no longer in the practical range of computation. This convergence behaviour is in complete contrast to the observed hp-FEM behaviour, where convergence is achieved for lower degrees of freedom as $\mathbf{d}$ decreases. This behaviour suggests the following numerical approach in determining the boundary layer widths.

   1. Find the fastest decaying boundary layer component:

- Find J_0 such that
  - $d = 10^{J_0+1}$ is in the exponential range of the p-version,
  - $d = 10^{J_0}$ is not,
- $w_1 = 10^{-(J_0+3)}d$ is the smallest boundary layer width.

2. Find slower decaying boundary layer component:

- Set $d = 10^{J_1}$ for $J_1 = J_0, J_0 - 1, \ldots$, run *hp*-version with mesh/order combination $\{-1, -1+s_1 w_1 d, 1-s_1 w_1 d, 1\}$, (n_1, n_0, n_1), for n_1 constant and n_0 variable;

- If non-exponential convergence, then add $w_2 = 10^{-(J_1+3)}d$ as boundary layer width;

- Repeat process with $d = 10^{J_2}$ for $J_2 = J_1, J_1 - 1, \ldots$, until all boundary layer components found or minimum value for d reached.

3. Read off interior mode scaling from graph once all boundary layer components are found.

The main assumption in the above method is that the convergence of the method in the central element follows a similar behaviour to the p-version, when adding boundary layer elements. The fixed values for the order of approximation of the boundary elements have to be large enough to resolve their corresponding boundary layer components for the corresponding value of d, which as noted at the end of Subsection 4.1 is larger for larger d.

To illustrate the use of the method consider the example presented in Subsection 4.3: solving (42) with $a_1 = 10^{-2}$, $a_2 = 10^4$, and $d = 10^{-6}$. For the one-element p-version, setting $J = -2, -1, 0, \ldots$, and $d = 10^J$ there is exponential convergence for $J = 0$ and non-exponential convergence for $J = -1$ and so $J_0 = -1$, see Figure 10. Consecutive errors in the central element are $\|z_N - z_{N-1}\|_2$ where N is the degrees of freedom and z_N the computed solution for z (strictly it is the difference $\|z_{2m} - z_{2m-2}\|_2$ since the solution is an even function). Thus $w_1 = 10^{-(J_0+3)}d = 10^{-2}d$ is the width of the fastest

Figure 10: Problem III: One-element, p-version: $\mathbf{d} = 10^{-2}$, 10^{-1} and 10^0.
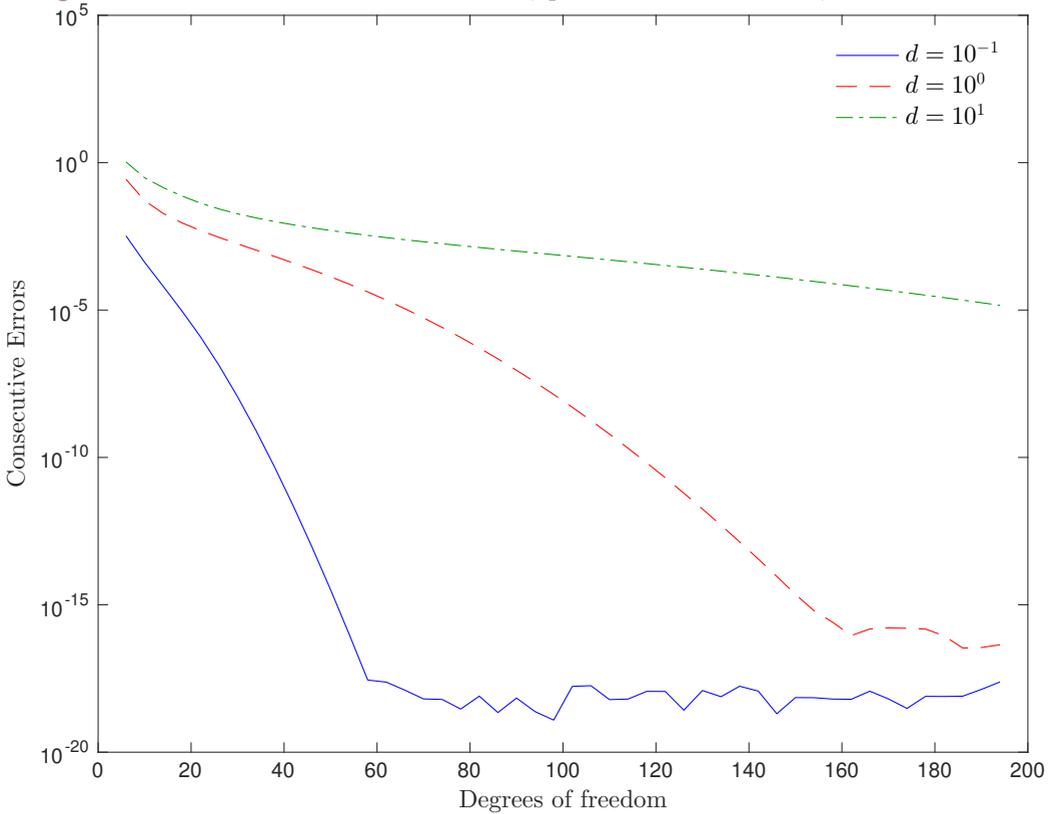
decaying component. If there was not exponential convergence for $J = -2$, then J would be decreased until within the spectral range. Finding this first value without any prior knowledge of the boundary layers would be difficult. However, in practise there are ways of estimating boundary layers, giving a guide where to start.

Now for the remaining boundary layers. With $J_1 = J_0, J_0 - 1, \ldots$, and $\mathbf{d} = 10^{J_1}$ ($J_0 = -1$) the hp-version with three-element mesh/order combination

$$\mathbf{x} = (-1, -1 + s_1 w_1, 1 - s_1 w_1, 1), \quad \mathbf{p} = (n_1, n_0, n_1), \qquad (44)$$

Figure 11: Problem III: Three-element hp-version (44): $d = 10^{J_1}$, $w_1 = 10^{-2}d$, $n_1 = 30$, $s_1 = 36$.
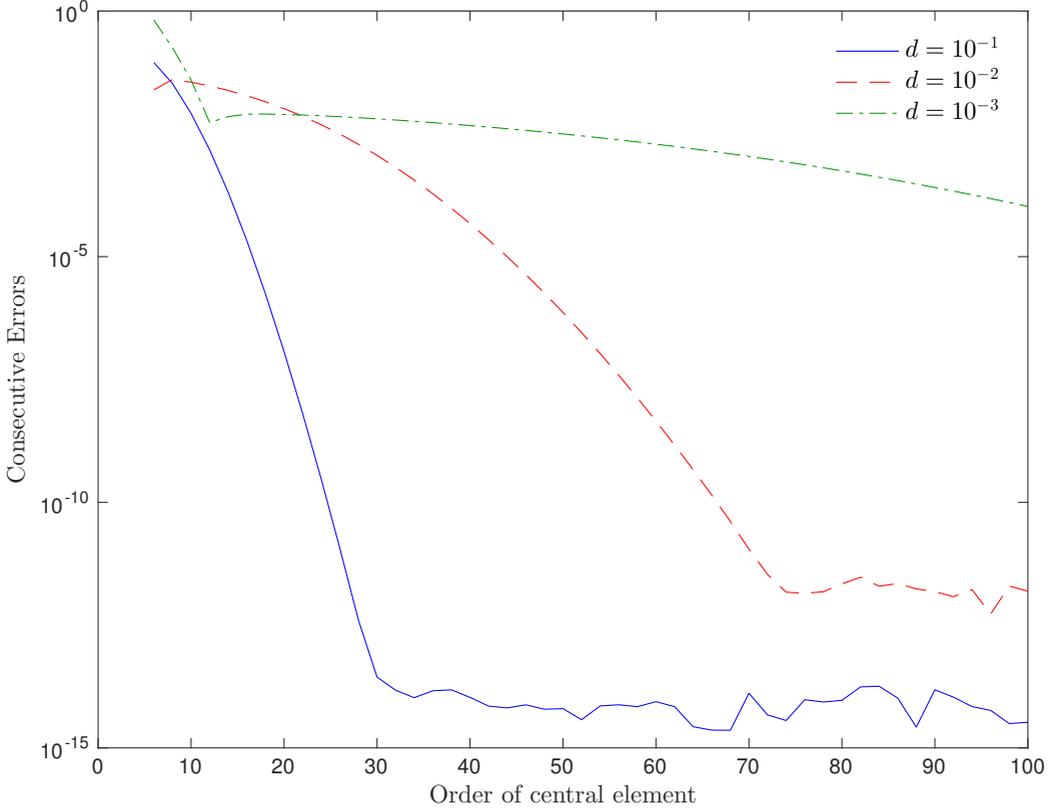

for n_1 fixed and n_0 variable is used to determine where the convergence behaviour changes from exponential to non-exponential. From Figure 11 this occurs when $J_1 = J_0 - 2 = -3$. The next boundary layer width is therefore $w_2 = 10^{-(J_1+3)}d = d$.

The final component is found by applying the hp-version with the five-element mesh/order combination

$$\mathbf{x} = (-1, -1 + s_1 w_1, -1 + s_2 w_2, 1 - s_2 w_2, 1), \quad \mathbf{p} = (n_2, n_1, n_0, n_1, n_2), \quad (45)$$

Figure 12: Problem III: Five-element *hp*-version (45): $d = 10^{J_2}$, $w_1 = 10^{-2}d$, $w_2 = d$, $n_2 = n_1 = 30$, $s_2 = s_1 = 36$.
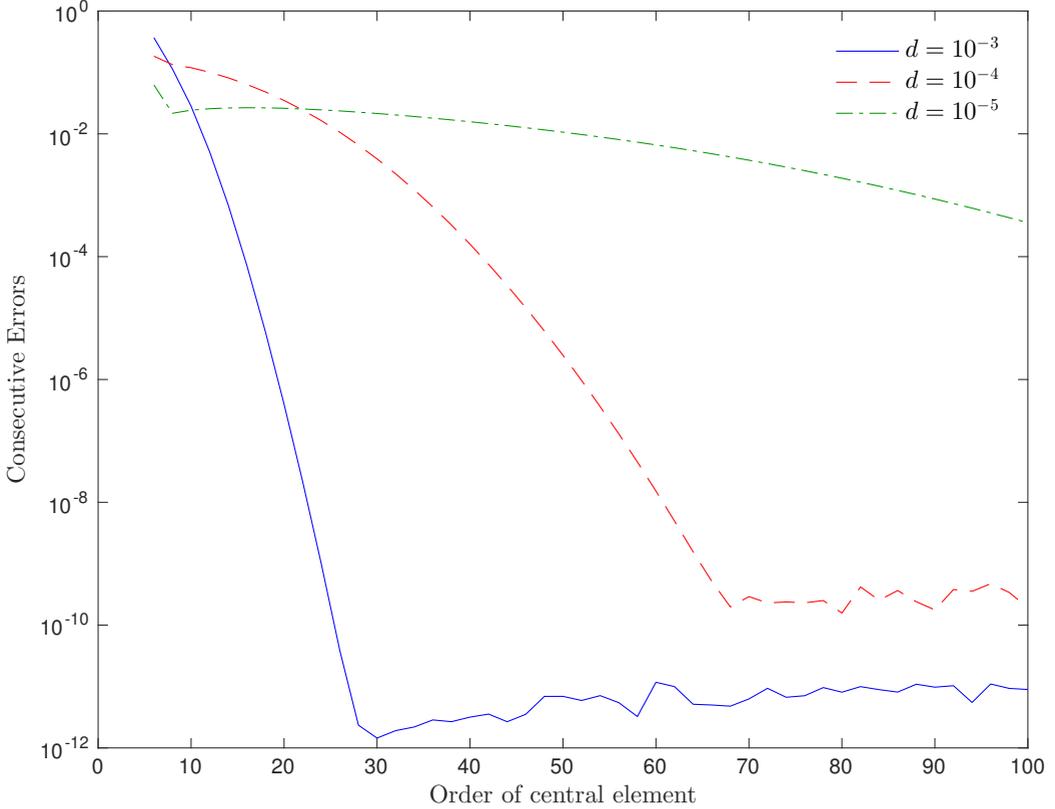
at $d = 10^{J_2}$, for $J_2 = J_1$, $J_1 - 1$, ..., n_2 and n_1 fixed and n_0 variable. The convergence jumps from exponential to non-exponential when $J_2 = J_1 - 2 = -5$ from Figure 12. The last boundary layer width is $w_3 = 10^{-(J_2+3)}d = 10^2 d$.

Subsequent plots with the seven-element mesh/order combination

$$\mathbf{x} = (-1, -1 + s_1 w_1, -1 + s_2 w_2, -1 + s_3 w_3, 1 - s_3 w_3, 1 - s_2 w_2, 1)$$
$$\mathbf{p} = (n_3, n_2, n_1, n_0, n_1, n_2, n_3)$$
(46)

where $d = 10^{J_3}$, for $J_3 = J_2$, $J_2 - 1$, $J_2 - 2$, that is, $d = 10^{-5}$, 10^{-6} and 10^{-7}

Figure 13: Problem III: Seven-element *hp*-version (46): $\mathsf{d} = 10^{\mathsf{J}_3}$, $\mathsf{w}_1 = 10^{-2}\mathsf{d}$, $\mathsf{w}_2 = \mathsf{d}$, $\mathsf{w}_3 = 10^2\mathsf{d}$, $\mathsf{n}_3 = \mathsf{n}_2 = \mathsf{n}_1 = 30$, $\mathsf{s}_3 = \mathsf{s}_2 = \mathsf{s}_1 = 36$.
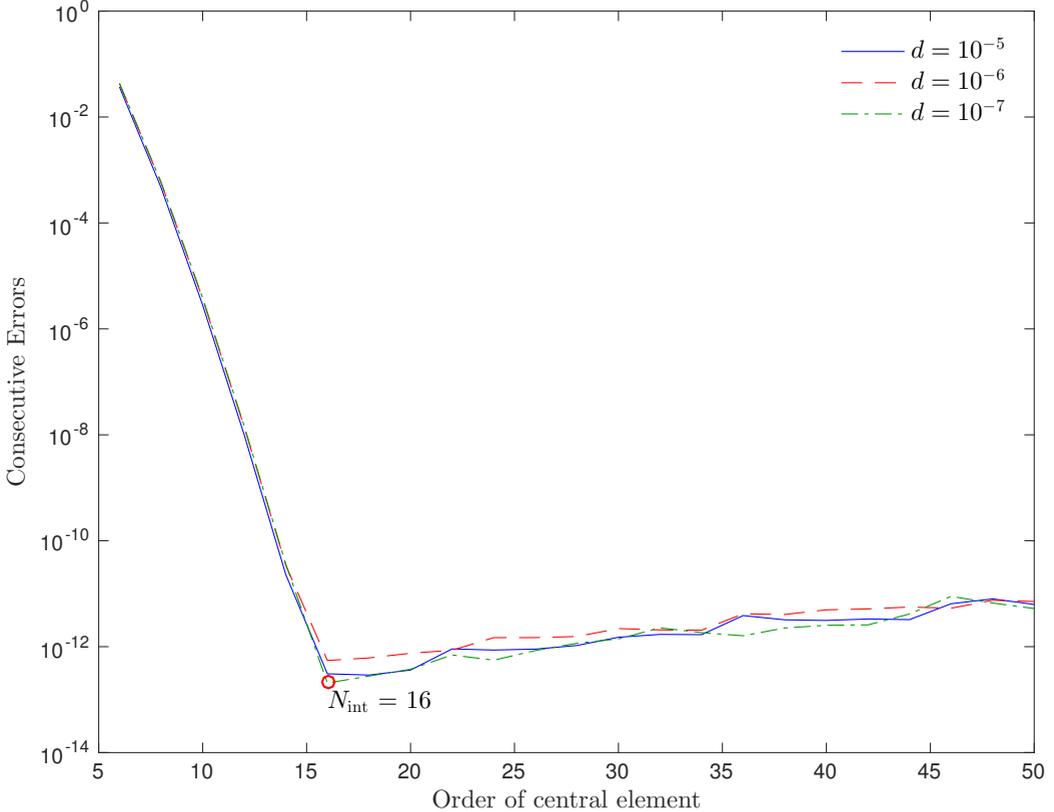
and n_3, n_2, n_1 fixed, n_0 variable, display exponential convergence, Figure 13. All components have therefore been resolved in the range for d. The interior mode order may also be read off this graph as $\mathsf{N}_{\text{int}} = 16$. The final step in calibration would be to find optimal values for s_j and n_j at the d that corresponds to its physical value. This is achievable by applying the method in Subsection 4.2, and testing the accuracy of the interior solution with a converged solution. As a rule of thumb, choose the εs larger in boundary elements further from the interior. This will be studied in future work.

5   Conclusions

The robustness of the *hp*-FEM presented in this article for boundary layer behaviour makes it useful for MHD applications, which typically have magnetic, viscous and thermal diffusive boundary layers. The calibration method in Subsection 4.4 shows it is possible to apply the *hp*-FEM without prior knowledge of all the boundary layer widths or the degree of approximation of the interior solution.

Future work will study the application of the *hp*-FEM to the linearised MHD equations in plane layers and sphere and spherical shell geometries. The aim is a practical approach to applying these methods. An approach where the boundary element widths and scaling factors s_j, and the order of the boundary elements and interior elements n_i, may be found in an efficient manner.

A   Jacobi polynomials

The function $P_n^{\alpha,\beta}(\xi)$, where $\alpha, \beta > -1$, is the Jacobi polynomial of degree n [8, Appendix A]. They are solutions of the singular Sturm–Liouville equation

$$(1 - \xi^2)y^{(2)} + \big[\beta - \alpha - (\alpha + \beta + 2)\xi\big]y^{(1)} = -n(n + \alpha + \beta + 1)y, \quad (47)$$

and satisfy the orthogonality relation

$$\int_{-1}^{1} (1-\xi)^{\alpha}(1+\xi)^{\beta} P_n^{\alpha,\beta} P_m^{\alpha,\beta}\, dx = \frac{\delta_{m,n} 2^{\alpha+\beta+1}}{2n + \alpha + \beta + 1} \frac{\Gamma(n + \alpha + 1)\Gamma(n + \beta + 1)}{n!\,\Gamma(n + \alpha + \beta + 1)}. \tag{48}$$

All orthogonal polynomials satisfy a three-term recurrence relation from which they may be computed. For the Jacobi polynomials the recurrence relation is

$$
\begin{aligned}
P_0^{\alpha,\beta} &= 1\,, \\
P_1^{\alpha,\beta} &= \tfrac{1}{2}(\alpha - \beta) + \tfrac{1}{2}(\alpha + \beta + 2)\xi\,, \\
a_1^n P_{n+1}^{\alpha,\beta} &= (a_2^n + a_3^n \xi)P_n^{\alpha,\beta} - a_4^n P_{n-1}^{\alpha,\beta}\,, \quad n \geqslant 1\,,
\end{aligned}
\tag{49}
$$

where the recursion coefficients a_j^n are

$$
\begin{aligned}
a_1^n &= (2n + 2)(n + \alpha + \beta + 1)(2n + \alpha + \beta), \\
a_2^n &= (2n + \alpha + \beta + 1)(\alpha^2 - \beta^2), \\
a_3^n &= (2n + \alpha + \beta)(2n + \alpha + \beta + 1)(2n + \alpha + \beta + 2), \\
a_4^n &= 2(n + \alpha)(n + \beta)(2n + \alpha + \beta + 2).
\end{aligned}
\tag{50}
$$

The derivatives of the Jacobi polynomials satisfy the recurrence relation

$$
b_1^n \frac{d}{d\xi} P_n^{\alpha,\beta} = b_2^n P_n^{\alpha,\beta} + b_3^n P_{n-1}^{\alpha,\beta}\,, \quad n \geqslant 1\,,
\tag{51}
$$

where

$$
\begin{aligned}
b_1^n &= (2n + \alpha + \beta)\left(1 - \xi^2\right), \\
b_2^n &= n(\alpha - \beta) - n(2n + \alpha + \beta)\xi\,, \\
b_3^n &= 2(n + \alpha)(n + \beta).
\end{aligned}
\tag{52}
$$

The Jacobi polynomials also satisfy the identity

$$
(1 - \xi)^{\alpha+1}(1 + \xi)^{\beta+1} P_{m-1}^{\alpha+1,\beta+1}(\xi) = -2m \int_{-1}^{\xi} (1 - y)^\alpha (1 + y)^\beta P_m^{\alpha,\beta}(y)\, dy\,,
\tag{53}
$$

which is used to calculate the normalisation factor for the interior modes.

B   Solution of differential equation (41)

The solution for y from (41) is

$$y = C_1 x \sin(bx) + C_2 \cos(bx) + A_1 \frac{\cosh(x/d)}{\cosh(1/d)}$$
$$+ A_2 \frac{\cosh(x/(a_1 d))}{\cosh(1/(a_1 d))} + A_3 \frac{\cosh(x/(a_1 a_2 d))}{\cosh(1/(a_1 a_2 d))} , \tag{54}$$

where

$$C_1 = -\left[(1 + b^2 d^2)(1 + a_1^2 b^2 d^2)(1 + a_1^2 a_2^2 b^2 d^2)\right]^{-1},$$

$$C_2 = C_1^2 \Big\{ (1 + b^2 d^2)(1 + a_1^2 b^2 d^2)(1 + a_1^2 a_2^2 b^2 d^2)^2$$
$$- 2bd^2 \left[1 + a_1^2(1 + a_2^2)(1 + 2b^2 d^2) + a_1^4 a_2^2 b^2 d^2(2 + 3b^2 d^2)\right] \Big\},$$

and the coefficients

$$A_1 = A_{1,1}\left[A_{1,2}\tanh(1/(a_1 d)) + A_{1,3}\tanh(1/(a_1 a_2 d))\right],$$
$$A_{1,1} = \left[(-1 + a_2^2)(-1 + a_1^2 a_2^2)(a_1\tanh(1/d) - \tanh(1/(a_1 d)))\right]^{-1},$$
$$A_{1,2} = -(\gamma_1 \cos(b) + \gamma_2 \sin(b)),$$
$$A_{1,3} = a_1^2 a_2^3(\gamma_3 \cos(b) + \gamma_4 \sin(b)),$$

$$A_2 = A_{2,1}\left[A_{2,2}\tanh(1/d) + A_{2,3}\tanh(1/(a_1 a_2 d))\right],$$
$$A_{2,1} = a_1 A_{1,1},$$
$$A_{2,2} = \gamma_1 \cos(b) + \gamma_2 \sin(b),$$
$$A_{2,3} = -a_1 a_2^3(\gamma_3 \cos(b) + \gamma_4 \sin(b)),$$

$$A_3 = A_{3,1}\left[A_{3,2}\cos(b) + A_{3,3}\sin(b)\right],$$
$$A_{3,1} = -a_1^2 a_2^4\left[(-1 + a_2^2)(-1 + a_1^2 a_2^2)\right]^{-1},$$
$$A_{3,2} = \gamma_3,$$
$$A_{3,3} = \gamma_4,$$

where

$$\gamma_1 = -C_2 + a_2^2 \left[C_2 + a_1^2 \left(C_2 - a_2^2 \gamma_5 \right) \right],$$
$$\gamma_2 = C_1 \left(1 + a_1^2 a_2^2 b^2 d^2 \right) \left[-1 + a_2^2 (1 + a_1^2 + a_1^2 b^2 d^2) \right],$$
$$\gamma_3 = C_2 - \gamma_5,$$
$$\gamma_4 = C_1 (1 + b^2 d^2)(1 + a_1^2 b^2 d^2),$$
$$\gamma_5 = 1 + (1 + a_1^2) b (2C_1 - bC_2) d^2 + a_1^2 b^3 (4C_1 - bC_2) d^4.$$

The corresponding solution z may be found from (38) as

$$\begin{aligned}
z = {} & D_1 x \sin(bx) + D_2 \cos(bx) + B_1 \frac{\cosh(x/d)}{\cosh(1/d)} \\
& + B_2 \frac{\cosh(x/(a_1 d))}{\cosh(1/(a_1 d))} + B_3 \frac{\cosh(x/(a_1 a_2 d))}{\cosh(1/(a_1 a_2 d))},
\end{aligned} \qquad (55)$$

where

$$D_1 = -C_1 \left[2 + (1 + a_1^2) b^2 d^2 + a_1^2 b^4 d^4 \right],$$
$$D_2 = 1 - 2C_2 + (1 + a_1^2) b (2C_1 - bC_2) d^2 + a_1^2 b^3 (4C_1 - bC_2) d^4,$$

and

$$B_1 = -A_1, \quad B_2 = -A_2, \quad B_3 = -A_3 \left[2 - a_2^{-2} + (1 - a_2^2) a_1^{-2} a_2^{-4} \right].$$

These constants, and computations involving exact solutions in Subsection 4.1–4.3, were calculated by summing from smallest to largest terms in magnitude as a strategy for improving the accuracy.

References

[1] S. Chandrasekhar. *Hydrodynamic and Hydromagnetic Stability*. Dover Publications, Newburyport, 2013. MR:0128226. Zbl:0142.44103. E100, E101

[2] D. Farmer and D. J. Ivers. Aspects of a hybrid hp-finite element/spectral method for the linearised magnetohydrodynamic equations in spherical geometries. In W. McLean and A. J. Roberts editors, *Proceedings of the 15th Biennial Computational Techniques and Applications Conference, CTAC-2010*, volume 52 of *ANZIAM J.*, pages C379–C394, July 2011. http://journal.austms.org.au/ojs/index.php/ANZIAMJ/article/view/3950 [July 20, 2011] E100, E101, E108, E109, E111

[3] G. H. Golub and J. H. Welsch. Calculation of Gauss Quadrature Rules. *Mathematics of Computation*, 23:221–230, 1969. doi:10.2307/2004418 E112

[4] R. Hide and P. H. Roberts. Hydromagnetic flow due to an oscillating plane. *Reviews of Modern Physics*, 32(4):799–806, 1960. doi:10.1103/revmodphys.32.799 E100, E101

[5] D. J. Ivers and C. G. Phillips. A vector spherical harmonic spectral code for linearised magnetohydrodynamics. In K. Burrage and Roger B. Sidje editors, *Proceedings of the 10th Biennial Computational Techniques and Applications Conference, CTAC-2001*, volume 44 of *ANZIAM J.*, pages C423–C442, April 2003. http://journal.austms.org.au/ojs/index.php/ANZIAMJ/article/view/689 [April 1, 2003] E100, E101

[6] D. J. Ivers and C. G. Phillips. Scalar and vector spherical harmonic spectral equations of rotating magnetohydrodynamics. *Geophysical Journal International*, 175:955–974, 2008. doi:10.1111/j.1365-246x.2008.03944.x E100, E101

[7] L. Kamenski, W. Huang, and H. Xu. Conditioning of finite element equations with arbitrary anisotropic meshes. *Mathematics of computation*, 83(289):2187–2211, 2014. doi:10.1090/s0025-5718-2014-02822-6 E108

[8] G. E. Karniadakis and S. Sherwin. *Spectral hp Element Methods for Computational Fluid Dynamics, Second Edition*. Oxford University

Press, Oxford, 2005. doi:10.1093/acprof:oso/9780198528692.001.0001 E101, E103, E137

[9] D. E. Loper. General solution for the linearized Ekman–Hartmann layer on a spherical boundary. *The Physics of Fluids*, 13(12):2995–2998, 1970. doi:10.1063/1.1692891 E100, E101

[10] D. E. Loper. Steady hydromagnetic boundary layer near a rotating, electrically conducting plate. *The Physics of Fluids*, 13(12):2999–3002, 1970. doi:10.1063/1.1692892 E100, E101

[11] C. Schwab. *p- and hp- Finite Element Methods, Theory and Applications in Solid and Fluid Mechanics*. Oxford University Press, Oxford, 1998. MR:1695813. Zbl:0910.73003 E101, E103, E104, E109, E115

[12] A. M. Soward and R. Hollerbach. Non-axisymmetric magnetohydrodynamic shear layers in a rotating spherical shell. *Journal of Fluid Mechanics*, 408:239–274, 2000. doi:10.1017/s0022112099007776 E100, E101

Author addresses

1. **D. Farmer**, School of Mathematics and Statistics, University of Sydney, Sydney, New South Wales 2006, Australia.
   mailto:davidf@maths.usyd.edu.au
   orcid:0000-0003-3884-0446

2. **D. J. Ivers**, School of Mathematics and Statistics, University of Sydney, Sydney, New South Wales 2006, Australia.
   mailto:david.ivers@sydney.edu.au
   orcid:0000-0001-6794-5824