Journal of Computational Physics 268 (2014) 377–398

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

On *hp*-convergence of prolate spheroidal wave functions and a new well-conditioned prolate-collocation scheme

Li-Lian Wang [a,1], Jing Zhang [b,c,*,2], Zhimin Zhang [d,3]

[a] Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 637371, Singapore
[b] School of Mathematics and Statistics, Huazhong Normal University, Wuhan 430079, China
[c] Beijing Computational Science Research Center, China
[d] Wayne State University, Detroit, MI 48202, United States

ARTICLE INFO

Article history:
Received 11 October 2013
Received in revised form 27 January 2014
Accepted 6 March 2014
Available online 18 March 2014

Keywords:
Prolate spheroidal wave functions
Collocation method
Pseudospectral differentiation matrix
Condition number
hp-convergence
Eigenvalues

ABSTRACT

The first purpose of this paper is to provide further illustrations, from both theoretical and numerical perspectives, for the nonconvergence of *h*-refinement in *hp*-approximation by the prolate spheroidal wave functions (PSWFs), a surprising convergence property that was first discovered by Boyd et al. (2013) [3]. The second purpose is to offer a new basis that leads to prolate-collocation systems with condition numbers independent of (c, N), the intrinsic bandwidth parameter and the number of collocation points. We highlight that the collocation scheme together with a very practical rule for pairing up (c, N) significantly outperforms the Legendre polynomial-based method (and likewise other Jacobi polynomial-based methods) in approximating highly oscillatory bandlimited functions.

1. Introduction

The prolate spheroidal wave functions of order zero (PSWFs) provide an optimal tool for approximating bandlimited functions (whose Fourier transforms are compactly supported), and appear superior to polynomials in approximating nearly bandlimited functions (cf. [37]). PSWFs also offer an alternative to Chebyshev and Legendre polynomials for pseudospectral/collocation and spectral-element algorithms, which enjoy a "plug-and-play" function by simply swapping the cardinal basis, collocation points and differentiation matrices (cf. [4,7,39,3]). With an appropriate choice of the underlying tunable bandwidth parameter c, PSWFs exhibit some advantages: (i) Spectral accuracy can be achieved on quasi-uniform computational grids; (ii) Spatial resolution can be enhanced by a factor of $\pi/2$; and (iii) The resulted method relaxes the Courant–Friedrichs–Lewy (CFL) condition related to an explicit time-stepping scheme. Boyd et al. [3, Table 1] provided an up-to-date review of recent developments since the series of seminal works by Slepian et al. [31,21,29].

While PSWFs enjoy some unique properties (e.g., being bandlimited and orthogonal over both a finite and an infinite interval), they are anyhow a non-polynomial basis, and therefore might lose certain capabilities of polynomials, when they

* Corresponding author at: School of Mathematics and Statistics, Huazhong Normal University, Wuhan 430079, China.
E-mail addresses: lilian@ntu.edu.sg (L.-L. Wang), zjccnu@gmail.com (J. Zhang), ag7761@wayne.edu (Z. Zhang).

[1] The research of this author is partially supported by Singapore MOE AcRF Tier 1 Grant (RG 15/12), MOE AcRF Tier 2 Grant (MOE 2013-T2-1-095 ARC 44/13), and A*STAR-SERC-PSF Grant (122-PSF-007).
[2] The work of this author is supported by the National Natural Science Foundation of China (11201166).
[3] This author is supported in part by the US National Science Foundation under grant DMS-1115530.

http://dx.doi.org/10.1016/j.jcp.2014.03.005
0021-9991/

are used for solving partial differential equations (PDEs). This can be best testified to by the nonconvergence of h-refinement in prolate-element methods, which was discovered by Boyd et al. [3] through simply examining hp-prolate approximation of the trivial function $u(x) = 1$. Indeed, PSWFs lack some properties of polynomials which are important for efficient spectral algorithms. Therefore, a naive extension of existing algorithms to this setting might be unsatisfactory or fail to work sometimes, so the related numerical issues are worthy of investigation.

The purpose of this paper is to have new insights into spectral algorithms using PSWFs as basis functions. The main contributions reside in the following aspects.

- We establish an hp-error bound for a PSWF-orthogonal projection. As a by-product, this provides further illustrations, from an approximation theory viewpoint, for the nonconvergence of h-refinement in hp-approximation. We also present more numerical results to demonstrate this surprising convergence behavior observed by Boyd et al. [3].
- We offer a new PSWF basis of dual nature.
  Firstly, it produces a matrix that nearly inverts the second-order prolate pseudospectral differentiation matrix, in the sense that their product is approximately an identity matrix for large N (see (5.9)). Consequently, it can be used as a preconditioner for the usual prolate-collocation scheme, leading to well-conditioned collocation linear systems. We remark that the idea along this line is mimic to the integration preconditioning (see e.g., [17,11,33]), and also related to the polynomial-based spectral algorithms in e.g., [9,13,38,15,14]. However, the PSWFs lack some properties of polynomials, so the procedure here is different from that for the polynomials.
  Secondly, under the new basis, the matrix of the highest derivative in the collocation linear system is an identity matrix, and the resulted linear system is well-conditioned. In contrast with the preconditioning technique, this does not involve the differentiation matrices, so the round-off errors in forming differentiation matrices can be alleviated.
  It is noteworthy that the non-availability of a quadrature rule exact for products of PSWFs, makes the PSWF-Galerkin method less attractive. We believe that the proposed well-conditioned collocation approach might be the best choice.
- We propose a practical approximation to Kong–Rokhlin's rule (see [19]) for pairing up (c, N), and demonstrate that the collocation scheme using this rule significantly outperforms the Legendre polynomial-based method when the underlying solution is bandlimited. For example, the portion of discrete eigenvalues of the prolate differentiation matrix that approximates the eigenvalues of the continuous operator to 12-digit accuracy is about 87% against 25% for the Legendre case (see Section 3.2). Similar advantages are also observed in solving Helmholtz equations with high wave numbers in heterogeneous media (see Section 5.3).

This paper is organized as follows. In Section 2, we review some basic properties of PSWFs, and the related quadrature rules, cardinal bases and differentiation matrices. In Section 3, we introduce Kong–Rokhlin's rule for pairing up (c, N), and study the discrete eigenvalues of the second-order prolate differentiation matrix. In Section 4, we establish the hp-error bound for a PSWF-projection and explain the nonconvergence of h-refinement in prolate-element methods. In Section 5, we introduce a new PSWF-basis which leads to well-conditioned collocation schemes. We also propose a collocation-based prolate-element method for solving Helmholtz equations with high wave numbers in heterogeneous media. The last section is for extensions and concluding remarks.

2. PSWFs and prolate pseudospectral differentiation

In this section, we review some relevant properties of the PSWFs, and introduce the quadrature rules, cardinal basis and prolate differentiation matrices.

2.1. Prolate spheroidal wave functions

The PSWFs arise from two contexts: (i) in solving the Helmholtz equation in prolate spheroidal coordinates by separation of variables (see e.g., [1]), and (ii) in studying time-frequency concentration problem (see [31]). As highlighted in [31], "*PSWFs form a complete set of bandlimited functions which possesses the curious property of being orthogonal over a given finite interval as well as over* $(-\infty, \infty)$."

Firstly, PSWFs, denoted by $\psi_n(x; c)$, are eigenfunctions of the singular Sturm–Liouville problem:

$$\mathcal{D}_x^c[\psi_n] := -\partial_x\big((1-x^2)\partial_x\psi_n(x;c)\big) + c^2x^2\psi_n(x;c) = \chi_n(c)\psi_n(x;c), \tag{2.1}$$

for $x \in I := (-1, 1)$, and $c \geqslant 0$. Here, $\{\chi_n(c)\}_{n=0}^{\infty}$, are the corresponding eigenvalues, and the positive constant c is dubbed as the "*bandwidth parameter*" (see Remark 2.3). PSWFs are complete and orthogonal in $L^2(I)$ (the space of square integrable functions). Hereafter, we adopt the conventional normalization:

$$\int_{-1}^{1} \psi_n(x;c)\psi_m(x;c)\,dx = \delta_{mn} := \begin{cases} 1, & m = n, \\ 0, & m \neq n. \end{cases} \tag{2.2}$$

The eigenvalues $\{\chi_n(c)\}_{n=0}^{\infty}$ (arranged in ascending order), have the property (cf. [37]):

$$\chi_n(0) < \chi_n(c) < \chi_n(0) + c^2, \quad n \geqslant 0,\ c > 0. \tag{2.3}$$

Asymptotically, for fixed c and large n, we have (cf. [26, (64)]):

$$\chi_n(c) = n(n+1) + \frac{c^2}{2} + \frac{c^2(4+c^2)}{32n^2}\left(1 - \frac{1}{n} + O\left(n^{-2}\right)\right). \tag{2.4}$$

Remark 2.1. Note that when $c = 0$, (2.1) reduces to the Sturm–Liouville equation of the Legendre polynomials. Denote the Legendre polynomials by $P_n(x)$, and assume that they are orthonormal. Then we have $\psi_n(x;0) = P_n(x)$ and $\chi_n(0) = n(n+1)$.

Secondly, Slepian et al. (cf. [31,30]) discovered that PSWFs coincidentally arise from the study of time-frequency concentration problems. Define the integral operator related to the finite Fourier transform:

$$\mathcal{F}_c[\phi](x) := \int_{-1}^{1} e^{\mathrm{i}cxt}\phi(t)\,dt, \quad \forall c > 0, \tag{2.5}$$

where $\mathrm{i} = \sqrt{-1}$ is the complex unit. Remarkably, the differential and integral operators are commutable: $\mathcal{D}_x^c \circ \mathcal{F}_c = \mathcal{F}_c \circ \mathcal{D}_x^c$. This implies that PSWFs are also eigenfunctions of $\mathcal{F}_c$, namely,

$$\mathrm{i}^n \lambda_n(c)\psi_n(x;c) = \int_{-1}^{1} e^{\mathrm{i}cx\tau}\psi_n(\tau;c)\,d\tau, \quad x \in I,\ c > 0. \tag{2.6}$$

The corresponding eigenvalues $\{\lambda_n(c)\}$ (modulo the factor i^n) are all real, positive, simple and ordered as

$$\lambda_0(c) > \lambda_1(c) > \cdots > \lambda_n(c) > \cdots > 0, \quad c > 0. \tag{2.7}$$

We have the following uniform upper bound (cf. [32, (2.14)]):

$$\lambda_n(c) < \frac{\sqrt{\pi}c^n(n!)^2}{(2n)!\Gamma(n+3/2)}, \quad n \geqslant 1,\ c > 0, \tag{2.8}$$

where $\Gamma(\cdot)$ is the Gamma function.

Remark 2.2. As demonstrated in [32], the upper bound in (2.8) provides a fairly accurate approximation to $\lambda_n(c)$ for a wide range of c, n of interest.

Remark 2.3. Recall that a function $f(x)$ defined in $(-\infty, \infty)$, is said to be bandlimited, if its Fourier transform $F(\omega)$, defined by

$$F(\omega) = \int_{-\infty}^{\infty} f(x)e^{\mathrm{i}\omega x}\,dx, \tag{2.9}$$

has a finite support (cf. [31]), that is, $F(\omega)$ vanishes when $|\omega| > \sigma > 0$. Then $f(x)$ can be recovered by the inverse Fourier transform

$$f(x) = \frac{1}{2\pi}\int_{-\sigma}^{\sigma} F(\omega)e^{-\mathrm{i}\omega x}\,d\omega. \tag{2.10}$$

One verifies from (2.6) and the parity: $\psi_n(-x;c) = (-1)^n\psi_n(x;c)$ (see [31]) that

$$\psi_n(x;c) = \frac{\mathrm{i}^n}{c\lambda_n(c)}\int_{-c}^{c}\psi_n\left(\frac{\omega}{c};c\right)e^{-\mathrm{i}\omega x}\,d\omega. \tag{2.11}$$

Hence, the PSWF ψ_n is bandlimited to $[-c, c]$, and c is therefore called the bandwidth parameter. However, its counterpart $P_n(x)$ is not bandlimited. Indeed, we have the following formula (see [12, p. 213]):

$$\int_{-1}^{1} P_n(\omega) e^{-\mathrm{i}\omega x}\, d\omega = (-\mathrm{i})^n (2n+1)\sqrt{\frac{\pi}{2}}\frac{J_{n+1/2}(x)}{\sqrt{x}}, \tag{2.12}$$

where $J_{n+1/2}$ is the Bessel function (cf. [1]). This implies $J_{n+1/2}(x)/\sqrt{x}$ is bandlimited, as its Fourier transform is $P_n(\omega)\chi_I(\omega)$ (up to a constant multiple), where χ_I is the indicate function of $(-1,1)$. Since a function and its Fourier transform cannot both have finite support, $P_n(x)$ is not bandlimited.

The PSWFs provide an optimal tool for approximating bandlimited functions (see e.g., [31,30,37,19]). On the other hand, being the eigenfunctions of a singular Sturm–Liouville problem (cf. (2.1)), the PSWFs offer a spectral basis on quasi-uniform grids with spectral accuracy (see e.g., [4,7,20,32,39,34,3]). However, the PSWFs are non-polynomials, so they lack some important properties that make the naive extension of polynomial algorithms to PSWFs unsatisfactory or fail to work sometimes. For example, Boyd et al. [3] demonstrated the nonconvergence of h-refinement in prolate elements, which turns out to be in distinctive contrast with Legendre polynomials. In addition, we observe that for any

$$\psi_m, \psi_n \in V_N^c := \operatorname{span}\{\psi_n\colon\ 0 \leqslant n \leqslant N\}, \tag{2.13}$$

we have

$$\partial_x \psi_n \notin V_{N-1}^c; \qquad \int \psi_n\, dx \notin V_{N+1}^c; \qquad \psi_n \cdot \psi_m \notin V_{2N}^c, \quad c > 0. \tag{2.14}$$

These will bring about some numerical issues to be addressed later.

Remark 2.4. In what follows, we might drop c and simply denote the PSWFs by $\psi_n(x)$, and likewise for the eigenvalues, whenever no confusion might cause.

2.2. Quadrature rules and grid points

The conventional choice of grid points for pseudospectral and spectral-element methods, is the Gauss–Lobatto points. The quadrature rule using such a set of points has the highest degree of precision (DOP) for polynomials. For example, let $\{\xi_j, \rho_j\}_{j=0}^N$ (with $\xi_0 = -1$ and $\xi_N = 1$) be the Legendre–Gauss–Lobatto (LGL) points (i.e., zeros of $(1-x^2)P_N'(x)$) and quadrature weights. Then we have

$$\int_{-1}^{1} P_n(x)\, dx = \sum_{j=0}^{N} P_n(\xi_j)\rho_j, \quad 0 \leqslant n \leqslant 2N-1. \tag{2.15}$$

It is also exact for all $P_n \cdot P_m \in \mathbb{P}_{2N-1}$ (the set of all algebraic polynomials of degree at most $2N-1$), namely,

$$\int_{-1}^{1} P_n(x) P_m(x)\, dx = \sum_{j=0}^{N} P_n(\xi_j) P_m(\xi_j)\rho_j, \quad 0 \leqslant n+m \leqslant 2N-1.$$

Such an exactness is essential for the superior accuracy of spectral/spectral-element methods based on the Galerkin formulation.

The choice of computational grids for the PSWFs is subtle, largely due to (2.14). The pursuit of the highest DOP leads to the generalized Gaussian quadrature (see e.g., [8,37,4,24]). In particular, the generalized prolate-Gauss–Lobatto (GPGL) quadrature in [4] is based on the fixed points: $x_0 = -1$, $x_N = 1$, and the interior quadrature points $\{x_j\}_{j=1}^{N-1}$ and weights $\{\omega_j\}_{j=0}^N$ being determined by

$$\int_{-1}^{1} \psi_n(x)\, dx = \psi_n(-1)\omega_0 + \sum_{j=1}^{N-1} \psi_n(x_j)\omega_j + \psi_n(1)\omega_N, \quad 0 \leqslant n \leqslant 2N-1. \tag{2.16}$$

Another choice is the prolate-Lobatto (PL) points (see [20,5] and [37,23] for prolate-Gaussian case), which are zeros of $(1-x^2)\partial_x \psi_N(x)$ (still denoted by $\{x_j\}_{j=0}^N$). The corresponding quadrature weights $\{\omega_j\}_{j=0}^N$ are determined by

$$\int_{-1}^{1} \psi_n(x)\, dx = \sum_{j=0}^{N} \psi_n(x_j)\omega_j, \quad 0 \leqslant n \leqslant N, \tag{2.17}$$

which is exact for $\{\psi_n\}_{n=0}^N$.

Remark 2.5. It is noteworthy that in the Legendre case (i.e., $c = 0$), the quadrature rules (2.16) and (2.17) are identical.

Remark 2.6. In view of (2.14), the GPGL quadrature (2.16) is not exact for $\psi_n \cdot \psi_m$ with $0 \leqslant m + n \leqslant 2N - 1$. This makes the spectral-Galerkin method using PSWFs less attractive. On the other hand, when it comes to prolate pseudospectral/collocation approaches, we find there is actually very subtle difference between two sets of points (also see [7]). Moreover, much more effort is needed to compute the GPGL points, so in what follows, we just use the PL points.

2.3. Prolate differentiation matrices

With the grid points at our disposal, we now introduce the cardinal (or nodal) basis. Here, we have two different routes to define the prolate cardinal basis once again due to (2.14).

Let $\{x_j\}_{j=0}^N$ be the PL points. The first approach searches for the cardinal basis $h_k(x) := h_k(x; c) \in V_N^c$ such that

$$h_k(x_j) = \delta_{jk}, \quad 0 \leqslant k, j \leqslant N. \tag{2.18}$$

To compute the basis functions, we write

$$h_k(x) = \sum_{n=0}^{N} t_{nk} \psi_n(x), \tag{2.19}$$

and determine the coefficients $\{t_{nk}\}$ from (2.18). More precisely, introducing the $(N+1)^2$ matrices:

$$\boldsymbol{\Psi}_{jk} = \psi_k(x_j), \qquad \boldsymbol{\Psi}_{jk}^{(m)} = \psi_k^{(m)}(x_j), \qquad \boldsymbol{T}_{nk} = t_{nk}, \qquad \boldsymbol{D}_{jk}^{(m)} = h_k^{(m)}(x_j), \tag{2.20}$$

we have $\boldsymbol{\Psi}\boldsymbol{T} = \boldsymbol{I}_{N+1}$, so $\boldsymbol{T} = \boldsymbol{\Psi}^{-1}$. Thus, the mth-order differentiation matrix is computed by

$$\boldsymbol{D}^{(m)} = \boldsymbol{\Psi}^{(m)} \boldsymbol{\Psi}^{-1}, \quad m \geqslant 1. \tag{2.21}$$

The second approach is to define

$$l_k(x) = \frac{s(x)}{s'(x_k)(x - x_k)}, \quad 0 \leqslant k \leqslant N \quad \text{with } s(x) = \left(1 - x^2\right) \partial_x \psi_N(x). \tag{2.22}$$

Then one verifies readily that

$$l_k(x_j) = \delta_{jk}, \quad 0 \leqslant k, j \leqslant N. \tag{2.23}$$

Different from the previous case, the so-defined $\{l_k\}_{k=0}^N \nsubseteq V_N^c$ for $c > 0$. The differentiation matrix $\widehat{\boldsymbol{D}}^{(m)}$ with the entries $\widehat{\boldsymbol{D}}_{jk}^{(m)} = l_k^{(m)}(x_j)$ for $0 \leqslant k, j \leqslant N$ can be computed by directly differentiating the cardinal basis in (2.22). We provide in Appendix A the explicit formulas for computing the entries of $\widehat{\boldsymbol{D}}^{(1)}$ and $\widehat{\boldsymbol{D}}^{(2)}$, which only involve the function values $\{\psi_N(x_j)\}_{j=0}^N$.

3. Study of eigenvalues of the prolate differentiation matrix

The appreciation of eigenvalue distribution of spectral differentiation matrices is important in many applications of spectral methods (see e.g., [35,36]). For example, for the second-order differentiation matrix, we are interested in the answer to the question: *to what extent can the discrete eigenvalues approximate those of the continuous operator accurately?*

With this in mind, we first introduce the Kong–Rokhlin's rule in [19] for pairing up (c, N) that guarantees high accuracy in numerical integration and differentiation of bandlimited functions, but it requires computing the eigenvalue λ_N of the integral equation (2.6). In this section, we first propose a practical mean for its implementation. We demonstrate that with the choice of (c, N) by this rule, the portion of discrete eigenvalues of the prolate differentiation matrix that approximates the eigenvalues of the continuous operator to 12-digit accuracy is about 87% against 25% for the Legendre case. This implies that the polynomial interpolation cannot resolve the continuous spectrum, while the PSWF interpolation has significant higher resolution.

3.1. Kong–Rokhlin's rule

An important issue related to the PSWFs is the choice of bandlimit parameter c. As commented by [4], the so-called "*transition bandwidth*":

$$c_*(N) = \frac{\pi}{2}\left(N + \frac{1}{2}\right), \tag{3.1}$$

turned out to be very crucial for asymptotic study of PSWFs and all aspects of their applications. In fact, when c is close to $c_*(N)$, $\psi_N(x;c)$ behaves like the trigonometric function $\cos([\pi/2]N(1-x))$, so it's nearly uniformly oscillatory. However, when $c > c_*(N)$, $\psi_N(x;c)$ transits to the region of the scaled Hermite function, so it vanishes near the endpoints $x = \pm 1$. In other words, the PSWFs with $c > c_*(N)$ lose the capability of approximating general functions in $(-1,1)$. Therefore, the feasible bandwidth parameter c should fall into $[0, c_*(N))$. However, this range appears rather loose, as many numerical evidences showed the significant degradation of accuracy when c is close to $c_*(N)$.

A conservative bound was provided in [34] (which improved that in [7]):

$$0 < q_N := \frac{c}{\sqrt{\chi_N}} < \frac{1}{\sqrt[6]{2}} \approx 0.8909. \tag{3.2}$$

Note that $q_N \approx 1$, if $c = c_*(N)$. In practice, a quite safe choice is $c = N/2$ (see e.g., [7,32]).

From a different perspective, Kong and Rokhlin [19] proposed a useful rule for pairing up (c, N). The starting point is a prolate quadrature rule, say (2.17). We know from [37] that it has the accuracy for the complex exponential $e^{\mathrm{i}cax}$:

$$\left| \int_{-1}^{1} e^{\mathrm{i}cax}\, dx - \sum_{j=0}^{N} e^{\mathrm{i}cax_j}\omega_j \right| = O(\lambda_N). \tag{3.3}$$

Furthermore, for a bandlimited function of bandwidth c, defined by

$$f(x) = \int_{-1}^{1} \phi(t) e^{\mathrm{i}cxt}\, dt, \quad \text{for some } \phi \in L^2(-1,1),$$

we have (see [37, Remark 5.1])

$$\left| \int_{-1}^{1} f(x)\, dx - \sum_{j=0}^{N} f(x_j)\omega_j \right| \leqslant \varepsilon \|\phi\|, \tag{3.4}$$

where ε is the maximum error of integration of a single complex exponential as in (3.3). In view of this, Kong and Rokhlin [19] suggested the rule: given c and an error tolerance ε, choose the smallest $N_* = N_*(c,\varepsilon)$ such that

$$\lambda_{N_*}(c) \leqslant \varepsilon \leqslant \lambda_{N_*-1}(c). \tag{3.5}$$

In what follows, we introduce a very practical mean to implement this rule approximately, which does not require computing the eigenvalues $\{\lambda_N\}$. We first use the property $n! = \Gamma(n+1)$ and the formula (see [1, (6.1.38)]):

$$\Gamma(x+1) = \sqrt{2\pi}\, x^{x+\frac{1}{2}} \exp\left(-x + \frac{\theta}{12x}\right), \quad x > 0,\ \theta \in (0,1), \tag{3.6}$$

to derive a more explicit upper bound of λ_N in (2.8):

$$\frac{\sqrt{\pi} c^N (N!)^2}{(2N)!\Gamma(N+3/2)} \leqslant \sqrt{\frac{\pi e}{2}} \left(\frac{ec}{4}\right)^N \left(N+\frac{1}{2}\right)^{-(N+1/2)} e^{1/(6N)} := \nu_N(c). \tag{3.7}$$

We now replace λ_N in (3.5) by ν_N. Accordingly, given a tolerance $\varepsilon > 0$, we look for N_* satisfying the equation: $\nu_{N_*}(c) = \varepsilon$. Taking the common log on both sides of this equation, we are able to convert the problem of finding N_* to locate the root of the algebraic equation: $F_\varepsilon(x;c) = 0$ where

$$F_\varepsilon(x;c) := x \log\frac{ec}{4} - \left(x+\frac{1}{2}\right)\log\left(x+\frac{1}{2}\right) + \frac{1}{6x} + \log\frac{1}{\varepsilon} + \frac{1}{2}\log\frac{\pi e}{2}, \quad x \geqslant 1. \tag{3.8}$$

One verifies that $F'_\varepsilon(x;c) < 0$ for slightly large x, and $F''_\varepsilon(x;c) < 0$ (see Fig. 3.1 (left)). Moreover, $F_\varepsilon(1;c) > 0$ and $F_\varepsilon(\infty;c) < 0$, so $F_\varepsilon(x;c) = 0$ has a unique root x_*. Then we set $N_* = [x_*]$ (i.e., the integer part of x_*).

Remark 3.1. Note that $\nu_N(c)$ provides a fairly accurate approximation to $\lambda_N(c)$ (cf. [32]) and λ_{N_*} decays exponentially with respect to N_*, so we have $\lambda_{N_*} \approx \varepsilon \approx \lambda_{N_*-1}$. In fact, Boyd et al. [3] listed some other approximations to $\lambda_N(c)$, which might be employed here as well.

We compare in Table 3.1 the approximate approach with the "exact" approach in [19], and very similar performance is observed. In Fig. 3.1 (middle), we depict the ratio $\lambda_{N_*}/\varepsilon$ for various c, which are all constant and become smaller for as c increases. We also see from Fig. 3.1 (right) that N_* roughly grows linearly with respect to c.

Table 3.1
A comparison of the pairs (c, N_*) obtained by the approximate approach and (c, N) obtained by the Kong–Rokhlin's rule [19], where $\varepsilon = 10^{-14}$.

c	N_*	λ_{N_*}	N_* [19]	λ_{N_*} [19]	c	N_*	λ_{N_*}	N_* [19]	λ_{N_*} [19]
10	24	1.77e−14	26	8.54e−16	100	94	2.79e−15	96	8.25e−16
20	34	5.96e−15	36	8.54e−16	200	163	8.00e−16	164	7.49e−16
40	50	8.79e−15	52	1.78e−15	400	299	5.20e−16	294	2.69e−15
80	79	1.10e−14	82	7.57e−16	800	571	1.57e−16	554	7.73e−16

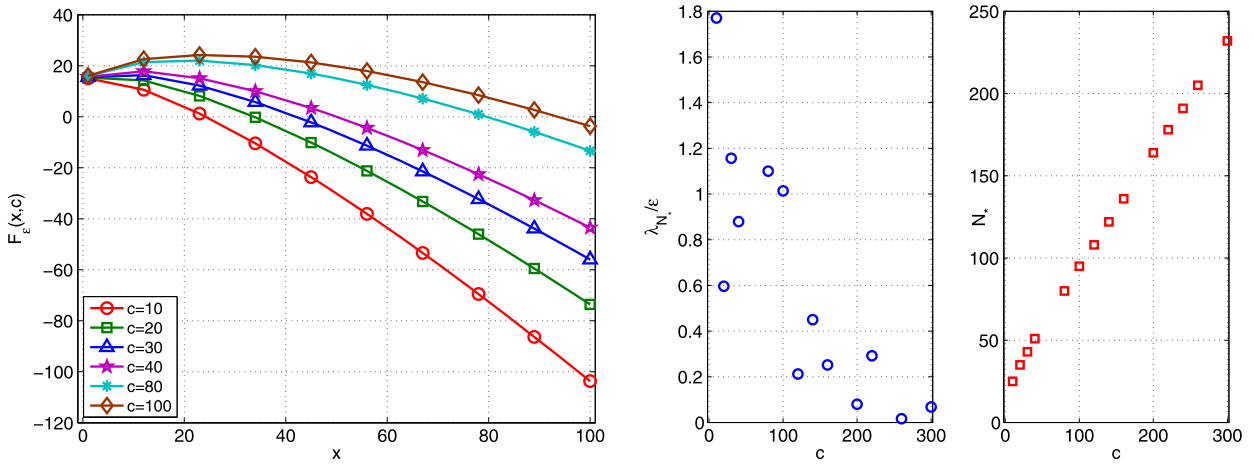

Fig. 3.1. Graphs of $F_\varepsilon(x; c)$ for some c given in the legend (left). The ratio $\lambda_{N_*}/\varepsilon$ (middle) and the pair (c, N_*) (right) for several samples of c. In all cases, $\varepsilon = 10^{-14}$.

3.2. Eigenvalues of the second-order prolate differentiation matrix

Consider the model eigen-problem:

$$\text{Find } (\mu, u) \text{ such that } u''(x) = \mu u(x), \quad x \in (-1, 1); \qquad u(\pm 1) = 0, \tag{3.9}$$

which admits the eigen-pairs (μ_k, u_k):

$$\mu_k = -\frac{k^2\pi^2}{4}, \quad u_k(x) = \sin\frac{k\pi(x+1)}{2}, \quad k \geqslant 1. \tag{3.10}$$

The corresponding discrete eigen-problems are

$$\text{Find } (\tilde{\mu}, \tilde{\boldsymbol{u}}) \text{ such that } \boldsymbol{D}_{\text{in}}^{(2)}\tilde{\boldsymbol{u}} = \tilde{\mu}\tilde{\boldsymbol{u}}; \quad \text{or} \quad \text{Find } (\hat{\mu}, \hat{\boldsymbol{u}}) \text{ such that } \widehat{\boldsymbol{D}}_{\text{in}}^{(2)}\hat{\boldsymbol{u}} = \hat{\mu}\hat{\boldsymbol{u}}. \tag{3.11}$$

Here, $\boldsymbol{D}_{\text{in}}^{(2)}$ and $\widehat{\boldsymbol{D}}_{\text{in}}^{(2)}$ are obtained by deleting the first and last rows and columns of $\boldsymbol{D}^{(2)}$ and $\widehat{\boldsymbol{D}}^{(2)}$, respectively. Recall that $\boldsymbol{D}^{(2)}$ and $\widehat{\boldsymbol{D}}^{(2)}$ are the differentiation matrices associated with two different ways of forming nodal basis functions, see (2.19) and (2.22), respectively.

We examine the relative errors:

$$\tilde{e}_j := \frac{|\tilde{\mu}_j - \mu_j|}{|\mu_j|}, \quad \hat{e}_j := \frac{|\hat{\mu}_j - \mu_j|}{|\mu_j|}, \quad 1 \leqslant j \leqslant N-1.$$

In the computation, (c, N) is paired up by the approximate Kong–Rokhlin's rule (i.e., through finding the root, see (3.8)) with $\varepsilon = 10^{-14}$. We plot in Fig. 3.2 the relative errors between the discrete and continuous eigenvalues of the prolate differentiation matrices with $c = 120\pi$ and $N = 284$, compared with those of the Legendre differentiation matrix at the Legendre–Gauss–Lobatto (LGL) points. Among 283 eigenvalues of $\boldsymbol{D}_{\text{in}}^{(2)}$, 245 (approximately 87%) are accurate to at least 12 digits with respect to the exact eigenvalues, while only 72 (approximately 25%) of the Legendre case are of this accuracy. A very similar number of accurate eigenvalues is also obtained from $\widehat{\boldsymbol{D}}_{\text{in}}^{(2)}$.

Remark 3.2. Some remarks are in order.

- As shown in [35] for the Legendre case, a portion $2/\pi$ of the eigenvalues approximate the eigenvalues of the continuous operator with one or two digit accuracy (about 180 among 283). The errors in the remaining ones are large, which

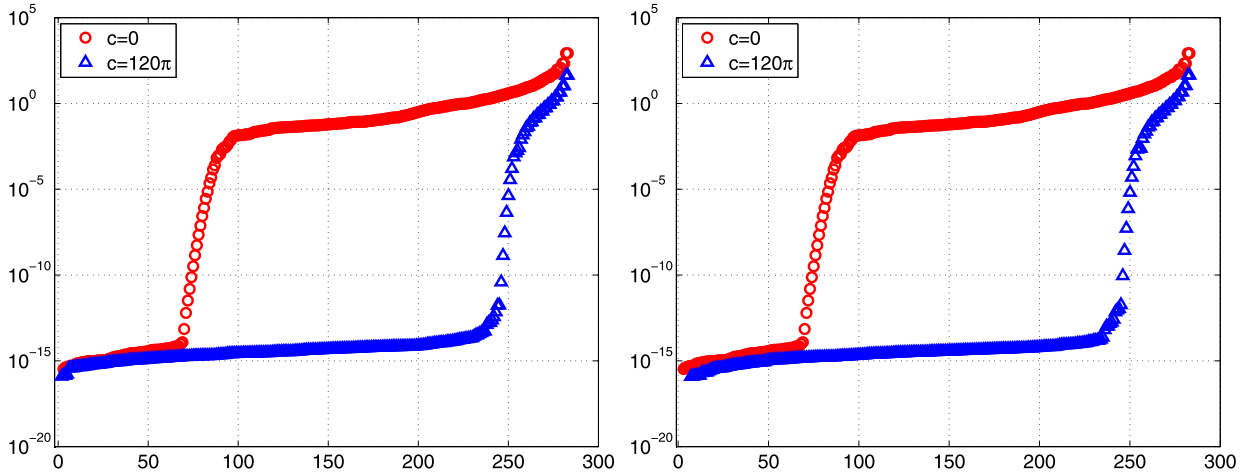

Fig. 3.2. The relative errors $\{\tilde{e}_j\}_{j=1}^{N-1}$ (left) and $\{\hat{e}_j\}_{j=1}^{N-1}$ (right), obtained by $c = 120\pi$, $\varepsilon = 10^{-14}$ and $N = 284$. The prolate differentiation matrices $\boldsymbol{D}_{\text{in}}^{(2)}$ (left, marked by "△") and $\widehat{\boldsymbol{D}}_{\text{in}}^{(2)}$ (right, marked by "△"), against the Legendre case (marked by "∘").

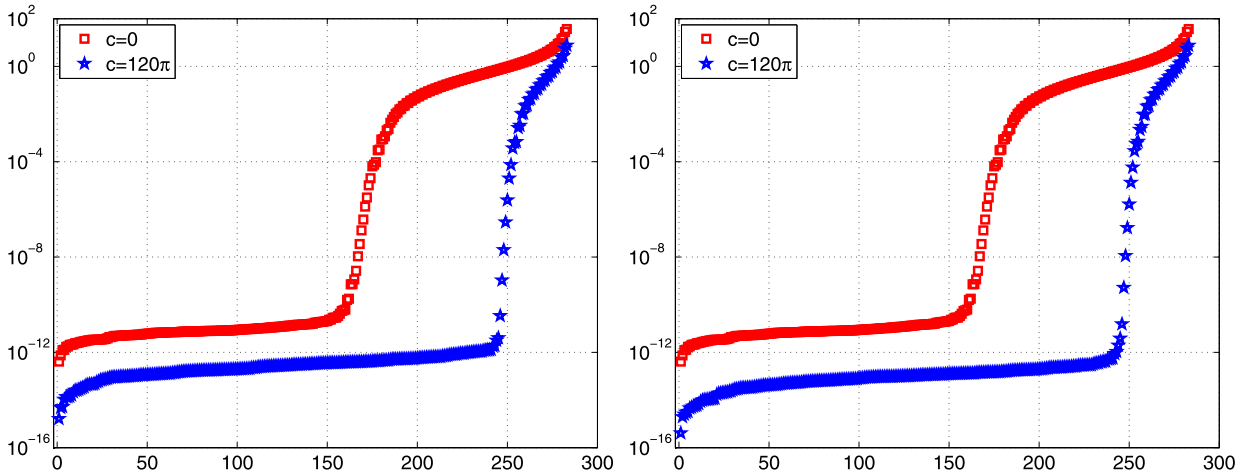

Fig. 3.3. The relative errors $\{\tilde{e}_j\}_{j=1}^{N-1}$ (left) and $\{\hat{e}_j\}_{j=1}^{N-1}$ (right) for (3.12) with $c = 120\pi$, $\varepsilon = 10^{-14}$ and $N = 284$. The prolate differentiation matrices $\boldsymbol{D}_{\text{in}}^{(2)}$ (left, marked by "★") and $\widehat{\boldsymbol{D}}_{\text{in}}^{(2)}$ (right, marked by "★"), against the Legendre case (marked by "□").

cannot be resolved by polynomial interpolation even on spectral grids. However, the prolate interpolation significantly improves the portion to around 95%. We point out that the portion is about 1/3 for piecewise linear finite elements (refer to [41] for a review).

- We remark that the usual prolate differentiation scheme under the approximate Kong–Rokhlin's rule behaves very similarly to the new differentiation scheme proposed by Kong and Rokhlin [19], which however is based on a Gram–Schmidt orthogonalization of certain modal basis.

We next consider the eigen-problem involving the Bessel operator:

$$u''(r) + \frac{1}{r}u'(r) - \frac{1}{r^2}u(r) = \mu u(r), \quad r \in (0,1); \qquad u(0) = u(1) = 0. \tag{3.12}$$

The exact eigenvalues are $\mu_k = -\tau_k^2$, $k \geqslant 1$, where each τ_k is a root of the Bessel function $J_1(\cdot)$. We adopt the same computational setting as for Fig. 3.2, and the relative errors are depicted in Fig. 3.3. Among 283 (discrete) eigenvalues, 245 are accurate to at least 12 digits with respect to the exact eigenvalues. In comparison, there are only 111 eigenvalues produced by Legendre collocation method that are within the same accurate level.

We demonstrate in Fig. 3.4 the growth of the magnitude of the largest and smallest eigenvalues of $\boldsymbol{D}_{\text{in}}^{(2)}$ and $\widehat{\boldsymbol{D}}_{\text{in}}^{(2)}$, compared with the Legendre case, where (c, N) is once again paired up by the approximate Kong–Rokhlin's rule. We observe a much slower growth of the largest eigenvalue, so the condition number of the differentiation matrix is much smaller.

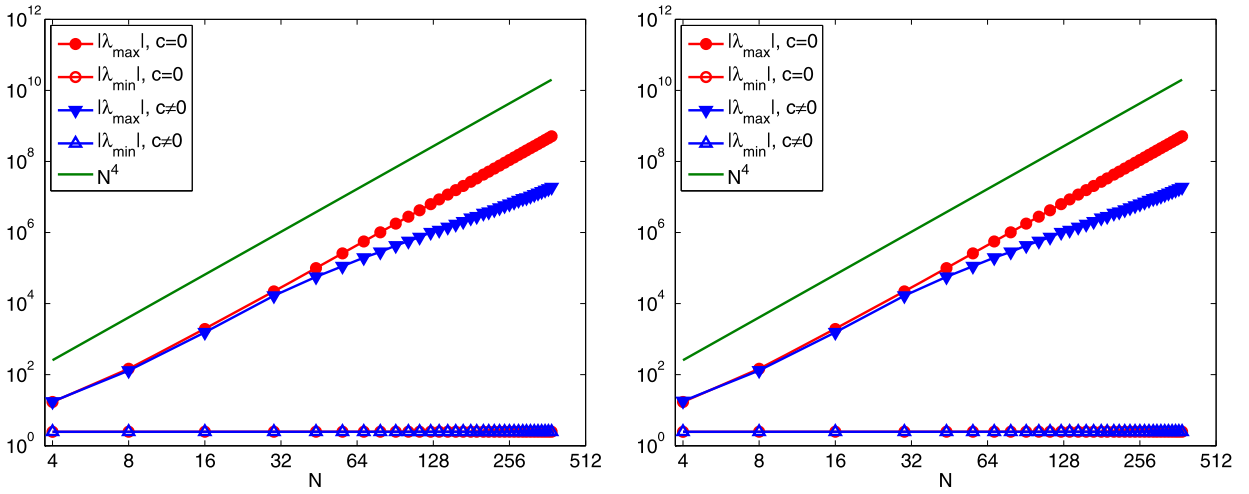

Fig. 3.4. Growth of the magnitude of the largest and smallest eigenvalues of $\boldsymbol{D}_{\text{in}}^{(2)}$ (left) and $\widehat{\boldsymbol{D}}_{\text{in}}^{(2)}$ (right) at the PL points ($c \neq 0$) against the Legendre case at the LGL points ($c = 0$).

4. On nonconvergence of *h*-refinement in prolate elements

As already mentioned, one purpose of this paper is to provide further illustrations, from both theoretical and numerical point of views, for the nonconvergence of h-refinement in prolate-element methods, a surprising property revealed by Boyd et al. [3].

We first introduce the notation and setting for hp-approximation by PSWFs. For simplicity, we partition the interval $\Omega := (a, b)$ uniformly into M non-overlapping subintervals, that is,

$$\bar{\Omega} = \bigcup_{i=1}^{M} \bar{I}_i, \quad I_i := (a_{i-1}, a_i), \quad a_i = a + ih, \quad h = \frac{b-a}{M}, \quad 1 \leqslant i \leqslant M. \tag{4.1}$$

Note that the mapping between I_i and the reference interval $I_{\text{ref}} := (-1, 1)$ is given by

$$x = \frac{h}{2} y + \frac{a_{i-1} + a_i}{2} = \frac{hy + 2a + (2i-1)h}{2}, \quad x \in I_i, \ y \in I_{\text{ref}}. \tag{4.2}$$

For any $u(x)$ defined in Ω, we denote

$$u|_{x \in I_i} = u^{I_i}(x) = \hat{u}^{I_i}(y), \quad x = \frac{hy + 2a + (2i-1)h}{2} \in I_i, \ y \in I_{\text{ref}}. \tag{4.3}$$

Hereafter, for a generic interval Λ, let $H^s(\Lambda)$ with $s \geqslant 0$ be the usual Sobolev space with the norm $\|\cdot\|_{H^s(\Lambda)}$ as in Admas [2]. In particular, for $s = 0$, we denote the L^2-norm by $\|\cdot\|_{L^2(\Lambda)}$.

Let $\hat{\pi}_N^c$ be the $L^2(I_{\text{ref}})$-orthogonal projector upon $V_N^c = \text{span}\{\psi_n : \ 0 \leqslant n \leqslant N\}$, given by

$$\big(\hat{\pi}_N^c \hat{u}\big)(y) = \sum_{n=0}^{N} \hat{u}_n(c) \psi_n(y; c) \quad \text{with } \hat{u}_n(c) = \int_{I_{\text{ref}}} \hat{u}(y) \psi_n(y; c)\, dy. \tag{4.4}$$

Define the approximation space

$$X_{h,N}^c = \big\{ v \in H^1(\Omega) : \ v|_{I_i}(x) = \hat{v}^{I_i}(y) \in V_N^c, \ 1 \leqslant i \leqslant M \big\}. \tag{4.5}$$

Let $\boldsymbol{\pi}_{h,N}^c : H^1(\Omega) \to X_{h,N}^c$ be a mapping, assembled by

$$\big(\boldsymbol{\pi}_{h,N}^c u\big)\big|_{I_i}(x) = \big(\hat{\pi}_N^c \hat{u}^{I_i}\big)(y), \quad 1 \leqslant i \leqslant M, \tag{4.6}$$

where by definition, we have

$$\big(\boldsymbol{\pi}_{h,N}^c u\big)\big|_{I_i}(x) = \sum_{n=0}^{N} \hat{u}_n^{I_i}(c)\, \psi_n(y; c) \quad \text{with } \hat{u}_n^{I_i}(c) = \int_{I_{\text{ref}}} \hat{u}^{I_i}(y) \psi_n(y; c)\, dy. \tag{4.7}$$

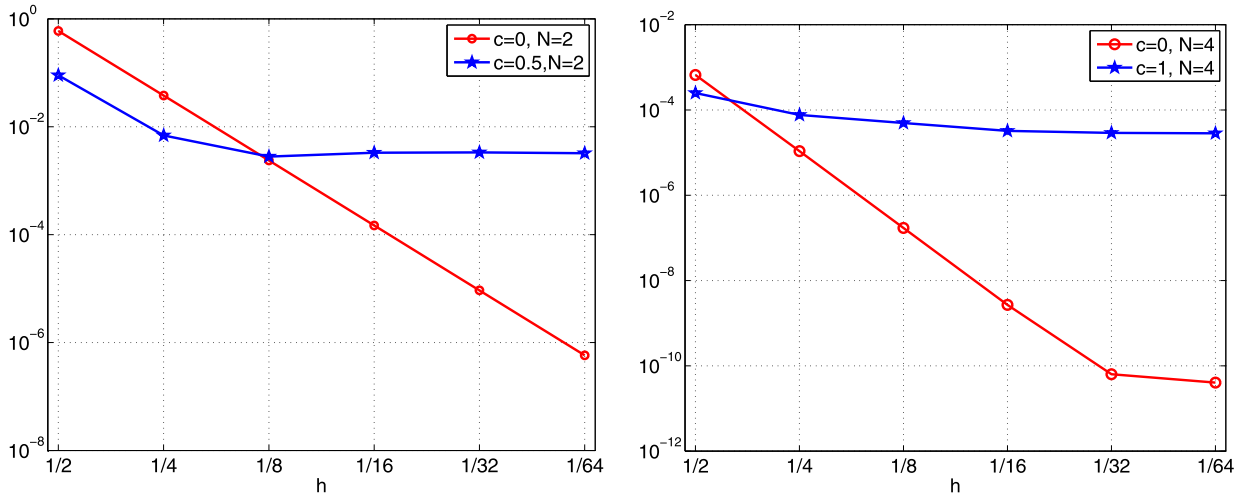

Fig. 4.1. Illustration of nonconvergence of h-refinement in prolate elements. Maximum point-wise errors against h with $N = 2$, $c = 0, 0.5$ (left), and $N = 4$, $c = 0, 1$ (right).

To describe the errors, we introduce the broken Sobolev space:

$$\widetilde{H}^{\sigma}(a,b) = \left\{u\colon\, u^{I_i} \in H^{\sigma}(I_i),\ 1 \leqslant i \leqslant M\right\}, \quad \sigma \geqslant 1, \tag{4.8}$$

equipped with the norm and semi-norm

$$\|u\|_{\widetilde{H}^{\sigma}(a,b)} = \left(\sum_{i=1}^{M} \|u^{I_i}\|^2_{H^{\sigma}(I_i)}\right)^{\frac{1}{2}}, \qquad |u|_{\widetilde{H}^{\sigma}(a,b)} = \left(\sum_{i=1}^{M} \|\partial_x^{\sigma} u^{I_i}\|^2_{L^2(I_i)}\right)^{\frac{1}{2}}.$$

The hp-approximability of $\boldsymbol{\pi}^c_{h,N}u$ to u is stated in the following theorem.

Theorem 4.1. *Let $\boldsymbol{\pi}^c_{h,N}$ be the projector defined in (4.6). For any constant $q_* < 1$, if*

$$\frac{c}{\sqrt{\chi N}} \leqslant \frac{q_*}{\sqrt[6]{2}} \approx 0.8909 q_*, \tag{4.9}$$

then for any $u \in \widetilde{H}^{\sigma}(a,b)$ with $\sigma \geqslant 1$, we have

$$\left\|\boldsymbol{\pi}^c_{h,N}u - u\right\|_{L^2(a,b)} \leqslant D\left\{\sqrt{N}\left(\frac{h}{N}\right)^{\sigma} |u|_{\widetilde{H}^{\sigma}(a,b)} + \frac{1}{\sqrt{\delta \ln(1/q_*)}}(q_*)^{\delta N}\|u\|_{L^2(a,b)}\right\}, \tag{4.10}$$

where D and δ are positive constants independent of u, N and c.

To be not distracted from the main result, we postpone its proof to Appendix B.

Remark 4.1. Some remarks are in orders.

- Observe from (4.10) that the second term in the upper bound is independent of h. This implies that for fixed N (e.g., $N = 2, 3$), the refinement of h does not lead to any convergence in h, as the second term dominates. For the example, $u(x) = 1$, considered in [3], the first term of the upper bound vanishes, so (4.10) indicates no gain in order of h, but the error decays exponentially as N grows.
- This should be in contrast with the Legendre approximation (see e.g., [6,18]), for which we have

$$\left\|\boldsymbol{\pi}^0_{h,N}u - u\right\|_{L^2(a,b)} \leqslant D\left(\frac{h}{N}\right)^{\sigma} |u|_{\widetilde{H}^{\sigma}(a,b)}. \tag{4.11}$$

- For fixed c, the estimate in (4.10) appears sub-optimal due to the factor $\sqrt{N}$. It can be improved to the optimal order by applying [32, Theorem 3.3] to (B.1), but the norm in the upper bound will depend on c implicitly.

Table 4.1
Performance of the prolate-element method with $c = N/4$ and the Legendre spectral-element method.

h	$N(c \neq 0)$					
	2	3	4	6	8	16
1/2	8.98E − 02	4.76E − 03	1.98E − 04	1.97E − 06	4.91E − 08	1.03E − 13
1/4	6.90E − 03	4.32E − 04	7.27E − 05	1.84E − 06	4.77E − 08	7.60E − 12
1/8	2.80E − 03	3.52E − 04	4.47E − 05	1.12E − 06	2.94E − 08	1.27E − 12
1/16	3.30E − 03	3.93E − 04	3.21E − 05	8.58E − 07	2.31E − 08	3.16E − 12
h	$N(c \neq 0)$					
	2	3	4	6	8	16
1/2	5.97E − 01	7.17E − 03	6.60E − 04	1.35E − 06	3.35E − 09	5.91E − 12
1/4	3.79E − 02	3.00E − 04	1.08E − 05	5.89E − 09	7.99E − 12	6.26E − 12
1/8	2.37E − 03	1.06E − 05	1.71E − 07	8.98E − 11	7.29E − 12	1.52E − 11
1/16	1.48E − 04	3.45E − 07	2.68E − 09	4.24E − 11	2.22E − 11	3.26E − 11

We next provide some numerical illustrations for the nonconvergence in h-refinement. Consider the prolate-element method for the equation:

$$-\big(1+x^2\big)u''(x) - (2x + \sin x)u'(x) + u(x) = f(x), \quad x \in (0,1),$$
$$u(0) = 0, \qquad u(1) = u_1, \tag{4.12}$$

where u_1 and $f(x)$ are computed from the exact solution: $u(x) = (x+1)^\alpha \sin(\pi x/2)$ with $\alpha = 13/3$. The prolate-element scheme is based on swapping the points, cardinal basis and differentiation matrices in the standard Legendre spectral-element method (see e.g., [25,5]).

In Fig. 4.1, we plot the maximum point-wise errors against h with fixed $N = 2, 4$ for the prolate and Legendre spectral-element methods. It clearly shows that the prolate elements do not have h-refinement convergence, as opposite to its counterpart.

We tabulate in Table 4.1 the maximum point-wise errors of two methods with various h, N. For fixed N, nonconvergence is observed by refining h for the prolate-element method. Benefited from h-convergence, the Legendre approach appears more accurate for small h and fixed N. However, from the viewpoint of p-version (e.g., for $h = 1/2$), the prolate-element method slightly outperforms the Legendre method.

5. A well-conditioned prolate-collocation method

In this section, we propose a well-conditioned prolate-collocation method for second-order boundary value problems (BVPs). The essential piece of the puzzle is to construct a new basis with dual nature. Firstly, this basis generates a matrix, denoted by $\boldsymbol{B}_{\rm in}$, such that the eigenvalues of $\boldsymbol{B}_{\rm in}\boldsymbol{D}_{\rm in}^{(2)}$ and $\boldsymbol{B}_{\rm in}\widehat{\boldsymbol{D}}_{\rm in}^{(2)}$ are nearly concentrated around one. In other words, the matrix $\boldsymbol{B}_{\rm in}$ is approximately the "inverse" of the second-order differentiation matrix. Therefore, the matrix $\boldsymbol{B}_{\rm in}$ is a nearly optimal preconditioner, leading to a well-conditioned prolate-collocation linear system. Secondly, using the new basis, the matrix of the highest derivative in the linear system of the usual collocation scheme is identity and the condition number of the whole linear system is independent of N and c. In this section, we focus on second-order BVPs, but the idea can be extended to higher-order BVPs (see Section 6).

5.1. New basis

Let $\{\beta_k(x) := \beta_k(x;c)\}_{k=0}^N$ be a set of functions in an $(N+1)$-dimensional space to be specified shortly, which satisfies the conditions:

$$\begin{aligned}
&\beta_0(-1) = 1, \qquad \beta_0''(x_j) = 0, \quad 1 \leqslant j \leqslant N-1, \qquad \beta_0(1) = 0;\\
&\beta_k(-1) = 0, \qquad \beta_k''(x_j) = \delta_{jk}, \quad \beta_k(1) = 0, \quad 1 \leqslant j,k \leqslant N-1;\\
&\beta_N(-1) = 0, \qquad \beta_N''(x_j) = 0, \quad 1 \leqslant j \leqslant N-1, \qquad \beta_N(1) = 1,
\end{aligned} \tag{5.1}$$

where $\{x_j\}$ are the PL points.

If we look for $\{\beta_k\}_{k=0}^N \subseteq V_N^c = \text{span}\{\psi_n\colon\ 0 \leqslant n \leqslant N\}$, then (5.1) is associated with a generalized Birkhoff interpolation problem: Given $u \in C^2(-1,1)$, find $p \in V_N^c$ such that

$$p(-1) = u(-1); \qquad p''(x_j) = u''(x_j), \quad 1 \leqslant j \leqslant N-1; \qquad p(1) = u(1). \tag{5.2}$$

We can express the interpolant as

$$p(x) = u(-1)\beta_0(x) + \sum_{k=1}^{N-1} u''(x_k)\beta_k(x) + u(1)\beta_N(x). \tag{5.3}$$

The basis $\{\beta_k\}$ for (5.2) can be computed by writing $\beta_k(x) = \sum_{k=0}^{N} \alpha_{nk}\psi_n(x)$, and solving the coefficients by the interpolation conditions. However, this process requires the inversion of a matrix as ill-conditioned as $\boldsymbol{\Psi}^{(2)}$ and $\boldsymbol{D}^{(2)}$, which is apparently unstable even for slightly large N. However, this approach works for the Legendre and Chebyshev cases (see [33]), thanks to some properties of orthogonal polynomials (but unfortunately, not available for the PSWFs).

Remark 5.1. The Birkhoff interpolation is typically considered in the polynomial setting (see [22,10,40]). Unlike the Lagrange and Hermite interpolation, it does not interpolate the function and its derivative values consecutively at every point. For example, in (5.2), the data $u(x_j)$ and $u'(x_j)$ are not interpolated at the interior point x_j.

In what follows, we search for the basis $\{\beta_k\}$ and the interpolant p in a different finite dimensional space other than V_N^c. More precisely, set

$$\beta_0(x) = \frac{1-x}{2}, \qquad \beta_N(x) = \frac{1+x}{2}, \tag{5.4}$$

and we look for

$$\beta_k \in W_N^{c,0} := \operatorname{span}\big\{\phi_n\colon\ \phi_n''(x) = \psi_n(x) \text{ with } \phi_n(\pm 1) = 0,\ 0 \leqslant n \leqslant N-2\big\}, \tag{5.5}$$

for $1 \leqslant k \leqslant N-1$, which therefore satisfy $\beta_k(\pm 1) = 0$ in (5.1). Solving the ordinary differential equation in (5.5) directly leads to

$$\phi_n(x) = x\int_{-1}^{x} \psi_n(t)\,dt - \int_{-1}^{x} t\psi_n(t)\,dt + \frac{1+x}{2}\int_{-1}^{1} (t-1)\psi_n(t)\,dt. \tag{5.6}$$

Then we compute $\{\beta_k\}_{k=1}^{N-1}$, by writing

$$\beta_k(x) = \sum_{n=0}^{N-2} \alpha_{nk}\phi_n(x), \quad \text{so} \quad \beta_k''(x) = \sum_{n=0}^{N-2} \alpha_{nk}\psi_n(x). \tag{5.7}$$

Thus we can find the coefficients $\{\alpha_{nk}\}$ by $\beta_k''(x_j) = \delta_{jk}$ with $1 \leqslant k, j \leqslant N-1$, that is,

$$\boldsymbol{A} = \bar{\boldsymbol{\Psi}}^{-1} \quad \text{where } \boldsymbol{A}_{nk} = \alpha_{nk},\ \bar{\boldsymbol{\Psi}}_{jn} = \psi_n(x_j), \tag{5.8}$$

for $1 \leqslant j, k \leqslant N-1$ and $0 \leqslant n \leqslant N-2$.

Remark 5.2. Like the cardinal basis in (2.19), this process only involves inverting a matrix of PSWF values, rather than their derivative values. Hence, the operations are very stable even for very large N.

Remark 5.3. It is important to point out that the use of integrated orthogonal polynomials (e.g., Chebyshev polynomials) as basis functions for devising well-conditioned spectral algorithms can be traced back to Clenshaw [9] in late 1950s. Alternatively, one could cast a differential equation into an integral equation, and then directly employs orthogonal polynomials to expand the unknown (see e.g., [13,38,15,14]). The idea of constructing the basis in (5.5) is actually inspired by these polynomial-based algorithms, and also by the very similar attempts along the line of collocation/pseudospectral approaches [17,11,33]. Unlike polynomial cases, we see from (5.6) that we do not use the iterated integration of the PSWFs, but invert the highest differential operator with homogeneous Dirichlet data to derive the basis.

Define the matrix $\boldsymbol{B}$ with entries $\boldsymbol{B}_{jk} = \beta_k(x_j)$ for $0 \leqslant k, j \leqslant N$, and let $\boldsymbol{B}_{\text{in}}$ be the $(N-1)^2$ matrix obtained by deleting the first and last rows and columns from $\boldsymbol{B}$. Observe from (5.5)–(5.6) that $\boldsymbol{B}_{\text{in}}$ is generated from integration of PSWFs, which turns out to be an "inverse process" of spectral differentiation in the sense that

$$\boldsymbol{D}_{\text{in}}^{(2)}\boldsymbol{B}_{\text{in}} \approx \boldsymbol{I}_{N-1}, \qquad \boldsymbol{B}_{\text{in}}\boldsymbol{D}_{\text{in}}^{(2)} \approx \boldsymbol{I}_{N-1}, \tag{5.9}$$

where $\boldsymbol{I}_{N-1}$ is an $(N-1)^2$ identity matrix.

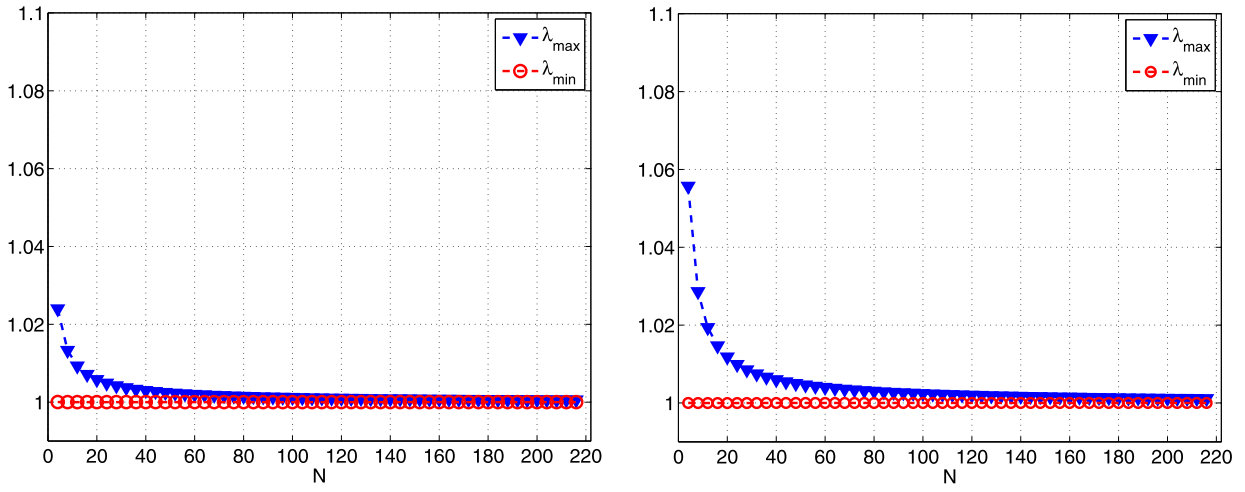

Fig. 5.1. Distribution of the largest and smallest eigenvalues of $\boldsymbol{B}_{\rm in}\boldsymbol{D}_{\rm in}^{(2)}$ (left) and $\boldsymbol{B}_{\rm in}\widehat{\boldsymbol{D}}_{\rm in}^{(2)}$ (right) for various $N \in [4, 218]$ and $c = N/2$.

Remark 5.4. We give an explanation to (5.9). Let $\{h_k\}_{k=0}^N$ be the cardinal basis defined in (2.18), and let $\mathcal{I}_N^c : C[-1,1] \to V_N^c$ be the interpolation operator defined by

$$\big(\mathcal{I}_N^c u\big)(x) = \sum_{p=0}^{N} u(x_p) h_p(x), \quad \forall u \in C[-1,1].$$

It is noteworthy that the estimate of the L^2-orthogonal projection error is based on a sophisticated analysis of the expansion coefficient (cf. (B.3) in Appendix B, and [7,34]). However, the bound of the interpolation error $\mathcal{I}_N^c u - u$ remains open, when $c > 0$. In fact, from the numerical evidences in e.g., [4,7,20], the nodal basis enjoys the approximability of the PSWFs $\{\psi_n\}$. Formally, we write

$$\beta_k''(x) \approx \big(\mathcal{I}_N^c \beta_k\big)''(x) = \sum_{p=1}^{N-1} \beta_k(x_p) h_p''(x), \quad 1 \leqslant k \leqslant N-1. \tag{5.10}$$

Roughly speaking, for large N and c satisfying (3.2), $\max_{|x|\leqslant 1} |(\mathcal{I}_N^c \beta_k)'' - \beta_k''|$ should decay exponentially with respect to N for all k. Indeed, by (5.6)–(5.7), $\{\beta_k\}$ are analytic functions, as they are linear combinations of PSWFs (which are analytic). Since $\beta_k(x_j) = \delta_{jk}$ (cf. (5.1)), letting $x = x_j$ in (5.10) leads to the first property in (5.9).

Similarly, by (5.3),

$$h_j(x) \approx \sum_{k=1}^{N-1} h_j''(x_k)\beta_k(x), \quad 1 \leqslant k \leqslant N-1.$$

We infer the second property in (5.9) from the approximability of the basis $\{\beta_k\}$.

Remark 5.5. The above argument also applies to the cardinal basis $\{l_j\}$ defined in (2.22), so one can replace $\boldsymbol{D}_{\rm in}^{(2)}$ in (5.9) by $\widehat{\boldsymbol{D}}_{\rm in}^{(2)}$.

We depict in Fig. 5.1 the distribution of the largest and smallest eigenvalues of $\boldsymbol{B}_{\rm in}\boldsymbol{D}_{\rm in}^{(2)}$ and $\boldsymbol{B}_{\rm in}\widehat{\boldsymbol{D}}_{\rm in}^{(2)}$ at the PL points. We see that all their eigenvalues for various N with $c = N/2$ are confined in $[\lambda_{\min}, \lambda_{\max}]$, which are concentrated around one for slightly large N. This agrees with (5.9).

5.2. Well-conditioned prolate-collocation methods

Consider the second-order variable coefficient problem:

$$u''(x) + p(x)u'(x) + q(x)u(x) = f(x), \quad x \in I = (-1,1); \qquad u(\pm 1) = u_{\pm}, \tag{5.11}$$

where p, q and f are continuous functions. Let $\{x_j\}_{j=0}^N$ be the PL points as before. Then the usual collocation scheme is: Find $u_N \in V_N^c$ such that

$$u_N''(x_j) + p(x_j)u_N'(x_j) + q(x_j)u_N(x_j) = f(x_j), \quad 1 \leqslant j \leqslant N-1; \qquad u_N(\pm 1) = u_{\pm}. \tag{5.12}$$

Under the cardinal basis $\{h_k\}$ defined in (2.18)–(2.19), the prolate-collocation system reads

$$\big(\boldsymbol{D}_{\rm in}^{(2)} + \boldsymbol{\Lambda}_p \boldsymbol{D}_{\rm in}^{(1)} + \boldsymbol{\Lambda}_q\big)\boldsymbol{u} = \boldsymbol{g}, \tag{5.13}$$

where $\boldsymbol{\Lambda}_p$ is a diagonal matrix with entries $\{p(x_j)\}_{j=1}^{N-1}$ (and likewise for $\boldsymbol{\Lambda}_q$), the unknown vector $\boldsymbol{u} = (u_N(x_1), \ldots, u_N(x_{N-1}))^t$, and $\boldsymbol{g}$ is the vector with elements

$$\boldsymbol{g}_j = f(x_j) - u_-\big(h_0''(x_j) + p(x_j)h_0'(x_j)\big) - u_+\big(h_N''(x_j) + p(x_j)h_N'(x_j)\big), \quad 1 \leqslant j \leqslant N-1.$$

It is known that the system (5.13) is ill-conditioned.

Thanks to (5.9), we precondition the system (5.13):

$$\boldsymbol{B}_{\rm in}\big(\boldsymbol{D}_{\rm in}^{(2)} + \boldsymbol{\Lambda}_p \boldsymbol{D}_{\rm in}^{(1)} + \boldsymbol{\Lambda}_q\big)\boldsymbol{u} = \boldsymbol{B}_{\rm in}\boldsymbol{g}, \tag{5.14}$$

which is well-conditioned (see Table 5.1).

On the other hand, one can directly use $\{\beta_k\}$ as a basis. Different from (5.12), the collocation scheme becomes: Find $v_N \in W_N^c := \text{span}\{\beta_k\colon\ 0 \leqslant k \leqslant N\}$ such that

$$v_N''(x_j) + p(x_j)v_N'(x_j) + q(x_j)v_N(x_j) = f(x_j), \quad 1 \leqslant j \leqslant N-1; \qquad v_N(\pm 1) = u_\pm. \tag{5.15}$$

By writing

$$v_N(x) = u_-\beta_0(x) + \sum_{k=1}^{N-1} w_k\beta_k(x) + u_+\beta_N(x), \tag{5.16}$$

the collocation system becomes

$$\big(\boldsymbol{I}_{N-1} + \boldsymbol{\Lambda}_p \boldsymbol{B}_{\rm in}^{(1)} + \boldsymbol{\Lambda}_q \boldsymbol{B}_{\rm in}\big)\boldsymbol{w} = \boldsymbol{h}, \tag{5.17}$$

where $\boldsymbol{w}$ is the vector of unknowns, and the vector $\boldsymbol{h}$ has the components

$$\boldsymbol{h}_j = f(x_j) - \big(p(x_j) + x_j q(x_j)\big)\frac{u_+ - u_-}{2} - q(x_j)\frac{u_+ + u_-}{2}, \quad 1 \leqslant j \leqslant N-1.$$

Finally, we recover $\boldsymbol{v} = (v_N(x_1), \ldots, v_N(x_{N-1}))^t$—the approximation of the solution, from (5.16):

$$\boldsymbol{v} = \boldsymbol{B}_{\rm in}\boldsymbol{w} + u_-\boldsymbol{b}_0 + u_+\boldsymbol{b}_N, \tag{5.18}$$

where $\boldsymbol{b}_0 = (\beta_0(x_1), \ldots, \beta_0(x_{N-1}))^t$ and $\boldsymbol{b}_N = (\beta_N(x_1), \ldots, \beta_N(x_{N-1}))^t$.

Remark 5.6. Compared with (5.14), the system (5.17) does not involve differentiation matrices. However, the unknowns are not physical values, so an additional step (5.18) is needed to recover the physical values.

Remark 5.7. Similar to the spectral-Galerkin method in [27], an essential idea is to construct an appropriate basis so that the matrix of the highest derivative becomes diagonal or identity. We refer to [28, p. 160] for the proof of the well-conditioning of such spectral-Galerkin schemes. However, a rigorous justification in this context appears challenging. Here, we just provide some intuitions for (5.11) with $p = 0$ and $q = q_0$ (a constant). Let $\lambda_{\min}$ and $\lambda_{\max}$ be the minimum and maximum eigenvalues of $\boldsymbol{D}_{\rm in}^{(2)}$. By (5.9), the eigenvalues of $\boldsymbol{B}_{\rm in}$ in magnitude are roughly confined in $[|\lambda_{\max}|^{-1}, |\lambda_{\min}|^{-1}]$. As a result, the eigenvalues of $\boldsymbol{I}_{N-1} + q_0\boldsymbol{B}_{\rm in}$ in magnitude approximately fall into the range $[1 + q_0|\lambda_{\max}|^{-1}, 1 + q_0|\lambda_{\min}|^{-1}]$. Note that for large N, $|\lambda_{\min}|$ behaves like a constant, while $|\lambda_{\max}|$ grows like $O(N^4)$ (see Fig. 3.4). This implies $\boldsymbol{I}_{N-1} + q_0\boldsymbol{B}_{\rm in}$ is well-conditioned.

We now provide some numerical examples, and compare the condition numbers between (5.13), (5.14) and (5.17). Consider

$$u''(x) - xu'(x) - u(x) = f(x) = \begin{cases} 0, & -1 < x < 0, \\ -3x^2/2, & 0 \leqslant x < 1, \end{cases} \tag{5.19}$$

with the exact solution

$$u(x) = \begin{cases} \exp(\frac{x^2}{2} + 1) + \exp(\frac{x^2}{2}), & -1 \leqslant x < 0, \\ \exp(\frac{x^2}{2} + 1) + \frac{x^2}{2} + 1, & 0 \leqslant x \leqslant 1. \end{cases} \tag{5.20}$$

Note that $f \in C^1(\bar{I})$ and $u \in C^3(\bar{I})$. The linear systems (5.13), (5.14) and (5.17) are neither sparse nor symmetric, so we solve them by the iterative approach—biconjugated gradient stabilized method. In Table 5.1, we tabulate the condition

Table 5.1
Performance of PCOL, P-PCOL and N-COL methods.

N	PCOL			P-PCOL			N-PCOL		
	Cond.	Errors	Steps	Cond.	Errors	Steps	Cond.	Errors	Steps
4	6.64E+00	1.40E−02	3	1.24	1.40E−02	3	1.25	7.71E−03	3
6	1.93E+00	1.30E−04	7	1.31	1.30E−04	6	1.49	1.04E−04	6
8	4.58E+01	1.29E−04	8	1.32	1.29E−04	6	1.59	1.03E−04	6
10	9.68E+01	4.19E−05	11	1.33	4.19E−05	6	1.66	3.98E−05	7
12	1.85E+02	2.01E−05	14	1.33	2.01E−05	6	1.70	1.99E−05	7
14	3.24E+02	1.12E−05	14	1.33	1.12E−05	6	1.72	1.12E−05	7
16	5.32E+02	6.78E−06	23	1.33	6.78E−06	6	1.74	6.78E−06	7
32	7.61E+03	4.80E−07	69	1.33	4.80E−07	6	1.82	4.80E−07	7
64	1.16E+05	3.20E−08	271	1.33	3.20E−08	6	1.86	3.20E−08	7
128	1.82E+06	2.14E−09	1037	1.33	2.07E−09	6	1.38	2.07E−09	7
256	2.88E+07	3.29E−08	6038	1.33	1.32E−10	6	1.88	1.32E−10	7
512	4.60E+08	8.65E−04	65 791	1.33	1.21E−11	6	1.89	8.35E−12	7

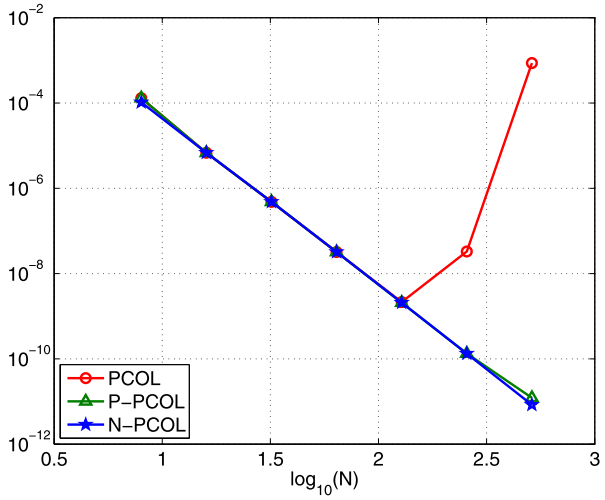

Fig. 5.2. Maximum point-wise errors for PCOL, P-PCOL and N-PCOL methods. The slope of two lines is approximately −3.95.

numbers, iteration steps, and maximum point-wise errors between the numerical and exact solutions obtained from the prolate-collocation scheme (5.13) (PCOL), the preconditioned scheme (5.14) (P-PCOL), and the new collocation scheme (5.17) (N-PCOL), respectively. Here, we choose $c = N/2$. In Fig. 5.2, we plot the maximum point-wise errors for three schemes.

We see that the last two schemes are well-conditioned and the iterative solver converges in a few steps, so they significantly outperform the usual prolate-collocation method using the cardinal basis (2.18)–(2.19). Note that the exact solution $u \in H^{4-\epsilon}(-1,1)$ for some $\epsilon > 0$, so the slope of the line is approximately -3.95 as expected.

5.3. A collocation-based p-version prolate-element method

As already discussed, the prolate-element method does not possess h-refinement convergence, and the Galerkin method is less attractive due to the lack of accurate quadrature rules for products of PSWFs. We therefore propose a p-version prolate-element method using the collocation formulation and the new basis $\{\beta_k\}$. Here, we particularly apply it to problems with discontinuous variable coefficients, e.g., the Helmholtz equation in heterogeneous media with high wave numbers.

Consider the model problem:

$$\begin{aligned} &L[u](x) := -\big(p(x)u'(x)\big)' + q(x)u(x) = f(x), \quad x \in \Omega = (a,b);\\ &u(a) = u_a, \qquad u(b) = u_b. \end{aligned} \tag{5.21}$$

We adopt the same setting as in (4.1)–(4.3). The interval Ω is uniformly partitioned into M non-overlapping subintervals $\{I_i = (a_{i-1}, a_i)\}_{i=1}^{M}$. Recall that the mapping between I_i and the reference interval $I_{\text{ref}} = (-1,1)$ is given by

$$x = \frac{h}{2}y + \frac{a_{i-1}+a_i}{2} = \frac{hy + 2a + (2i-1)h}{2}, \quad x \in I_i,\ y \in I_{\text{ref}}. \tag{5.22}$$

As before, let $W_N^c = \text{span}\{\beta_k\colon\ 0 \leqslant k \leqslant N\}$. Without loss of generality, assume that the same number of points will be used for each subinterval. Introduce the approximation space

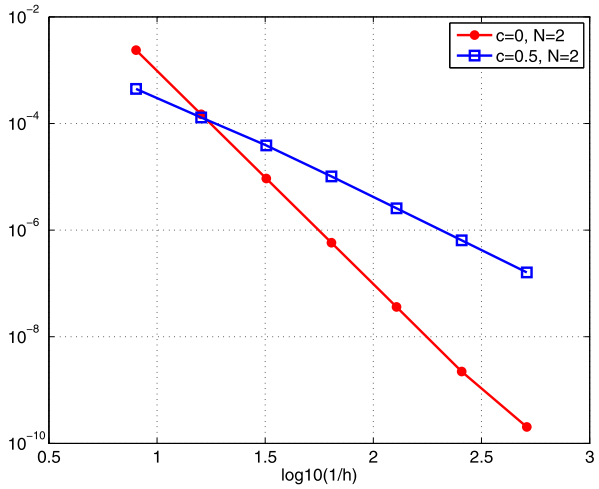

Fig. 5.3. Maximum point-wise errors of the new prolate- and Legendre spectral-element methods for the example (4.12). The slope of line with $c = 0.5$ is approximately -1.99.

$$Y^c_{h,N} := \{u \in H^1(\Omega):\ u(x)|_{x\in I_i} = u^{I_i}(x) = \hat{u}^{I_i}(y)|_{y\in I_{\rm ref}} \in W^c_N,\ 0 \leqslant i \leqslant M\}. \tag{5.23}$$

Define

$$\phi^{I_i}_k(x) = \begin{cases} \beta_k(y), & x = (hy + 2a + (2i-1)h)/2 \in I_i, \\ 0, & \text{otherwise}, \end{cases} \tag{5.24}$$

and at the adjoined points a_i, $1 \leqslant i \leqslant M-1$,

$$\varphi^{a_i}(x) = \begin{cases} (1+y)/2, & x = (hy + 2a + (2i-1)h)/2 \in I_i, \\ (1-y)/2, & x = (hy + 2a + (2i+1)h)/2 \in I_{i+1}, \\ 0, & \text{otherwise}. \end{cases} \tag{5.25}$$

Then we have

$$Y^c_{h,N} := \text{span}\{\{\phi^{I_1}_k\}^{N-1}_{k=0}, \{\phi^{I_2}_k\}^{N-1}_{k=1}, \ldots, \{\phi^{I_{M-1}}_k\}^{N-1}_{k=1}, \{\phi^{I_M}_k\}^{N}_{k=1};\ \{\varphi^{a_i}\}^{M-1}_{i=1}\}, \tag{5.26}$$

and the dimension of $Y^c_{h,N}$ is $MN+1$.

Let $\{y_j\}$ be the PL points in the reference interval $I_{\rm ref}$. Then the grids on each I_i are given by

$$x^{I_i}_j = \frac{hy_j + 2a + (2i-1)h}{2}, \quad 0 \leqslant j \leqslant N,\ 1 \leqslant i \leqslant M. \tag{5.27}$$

The prolate-element method for (5.21) is: Find $v \in Y^c_{h,N}$ such that $v(a) = u_a$, $v(b) = u_b$, and

$$L[v]\big(x^{I_i}_j\big) = f\big(x^{I_i}_j\big), \quad 1 \leqslant j \leqslant N-1,\ 1 \leqslant i \leqslant M, \tag{5.28}$$

and at the joint points a_i,

$$\int_a^b \{p(x)v'(x)\big(\varphi^{a_i}(x)\big)' + q(x)v(x)\varphi^{a_i}(x)\}\,dx = \int_a^b f(x)\varphi^{a_i}(x)\,dx, \quad 1 \leqslant i \leqslant M-1. \tag{5.29}$$

We see that the scheme is collocated at the interior points in each subinterval, and at the joint points, it is built upon the Galerkin-formulation for ease of imposing the continuity across elements. As shown in Section 5.2, the interior solvers (5.28) are well-conditioned, and the differentiation matrices are not involved.

Remark 5.8. As illustrated in Section 4, the prolate-element method using the nodal basis $\{h_j\}$ is not convergent when h is refined with fixed N. However, it is seen from (5.4) that the hat functions are included in the new basis $\{\beta_k\}$, but the basis cannot represent quadratic polynomials exactly. We therefore expect that $\{\beta_k\}$ should possess a second-order convergence in h. With this in mind, we test the method on the same example (4.12) with the same setting as for Fig. 4.1 (left). We plot in Fig. 5.3 the maximum point-wise errors of both the new prolate-element and Legendre spectral-element methods. Indeed, we find the rate of convergence is about second-order, while the Legendre method converges at a rate as predicted in (4.11).

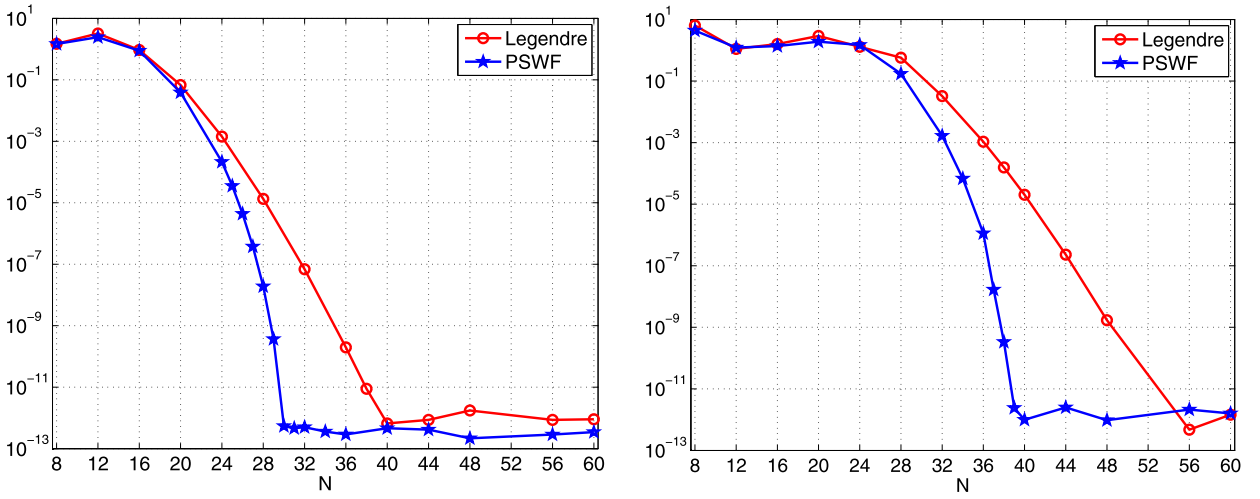

Fig. 5.4. Maximum point-wise errors of Legendre spectral-element and new prolate-element methods for the Helmholtz equation with exact solution (5.31). Left: $k = 60$ and right: $k = 100$.

We next provide some numerical results to show the performance of the new scheme. We focus on the Helmholtz equation in a heterogeneous medium:

$$\begin{aligned} &\big(c^2(x)u'(x)\big)' + k^2 n^2(x)u(x) = 0, \quad x \in \Omega = (a,b);\\ &u(a) = u_a, \qquad \big(cu' - \mathrm{i}knu\big)(b) = 0,\\ &u,\ c^2u' \text{ are continuous on } \Omega, \end{aligned} \tag{5.30}$$

where the wave number $k > 0$, and $c(x)$, $n(x)$ are piecewise smooth such that

$$0 < c_0 \leqslant c(x) \leqslant c_1, \qquad 0 < n_0 \leqslant n(x) \leqslant n_1.$$

Note that $c(x)$ and $n(x)$ represent the local speed of sound and the index of refraction in a heterogeneous medium, respectively.

In the first example, we choose $\Omega = (0,1)$, $n(x) = 1$ and $c(x)$ to be piecewise constant:

$$c(x) = \begin{cases} 2, & 0 < x < 1/2,\\ 1, & 1/2 < x < 1. \end{cases}$$

Then the problem (5.30) admits the exact solution (cf. [16]):

$$u(x) = \begin{cases} (3\exp(\frac{\mathrm{i}k(1+2x)}{4}) + \exp(\frac{\mathrm{i}k(3-2x)}{4}))/4, & 0 < x < 1/2,\\ \exp(\mathrm{i}kx), & 1/2 < x < 1. \end{cases} \tag{5.31}$$

In this case, we partition $\Omega = (0,1)$ into two subintervals $I_1 = (0, 1/2)$ and $I_2 = (1/2, 1)$.

In Fig. 5.4, we plot the maximum point-wise errors for the usual Legendre spectral-element method and the new p-version prolate-element method, where (c, N) is paired up by the approximate Kong–Rokhlin's rule with $\varepsilon = 10^{-14}$ and samples of c in $[2, 52]$. From Fig. 5.4, a much rapid convergence rate of the new approach is observed for high wave numbers.

In the second example, we take $\Omega = (0,1)$, $f(x) = 1$ and consider the problem (5.30) with piecewise smooth coefficients (cf. [16]):

$$c(x) = \begin{cases} 1 + x^2, & 0 < x < 0.25,\\ 1 - x^2, & 0.25 < x < 0.5,\\ 1, & 0.5 < x < 1, \end{cases} \qquad n(x) = \begin{cases} 1.75 + x, & 0 < x < 0.25,\\ 1.25 - x, & 0.25 < x < 0.5,\\ 2, & 0.5 < x < 1. \end{cases}$$

Naturally, we partition Ω into four subintervals of equal length. In this case, we do not have explicit exact solution, so we generate a reference "exact" solution using very fine grids by the new prolate-element method $(c, N) = (177, 144)$ (paired up by the approximate Kong–Rokhlin's rule again). In Fig. 5.5, we plot the real and image parts of the "exact" solution (where $k = 160$) against the numerical solution obtained by very coarse grids with $(c, N) = (36, 48)$, which approximates the highly oscillatory solution with an accuracy about 10^{-6}.

In Fig. 5.6, we make a comparison of convergence similar to that in Fig. 5.4. Here, we sample $c \in [4, 52]$. One again, we observe significantly faster convergence rate for the new approach under the approximate Kong–Rokhlin's rule (with $\varepsilon = 10^{-14}$) of selecting (c, N).

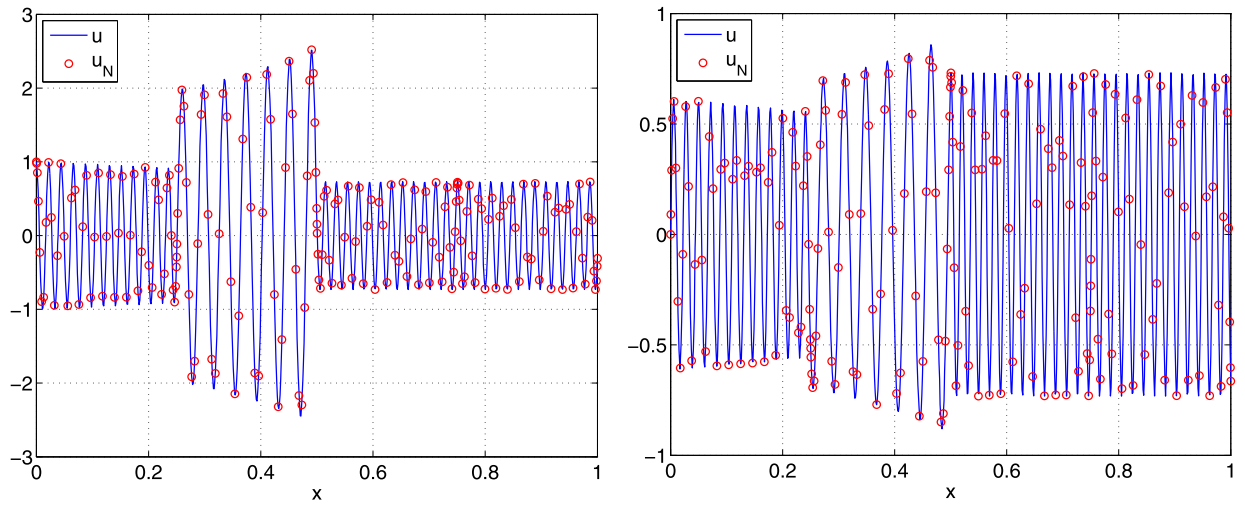

Fig. 5.5. Real part (left) and imaginary part (right) of the reference "exact" solution u computed by $(c, N) = (177, 144)$ and $k = 160$, against the numerical solution u_N of the prolate-element method with $(c, N) = (36, 48)$. The maximum point-wise error is 1.19×10^{-6}.

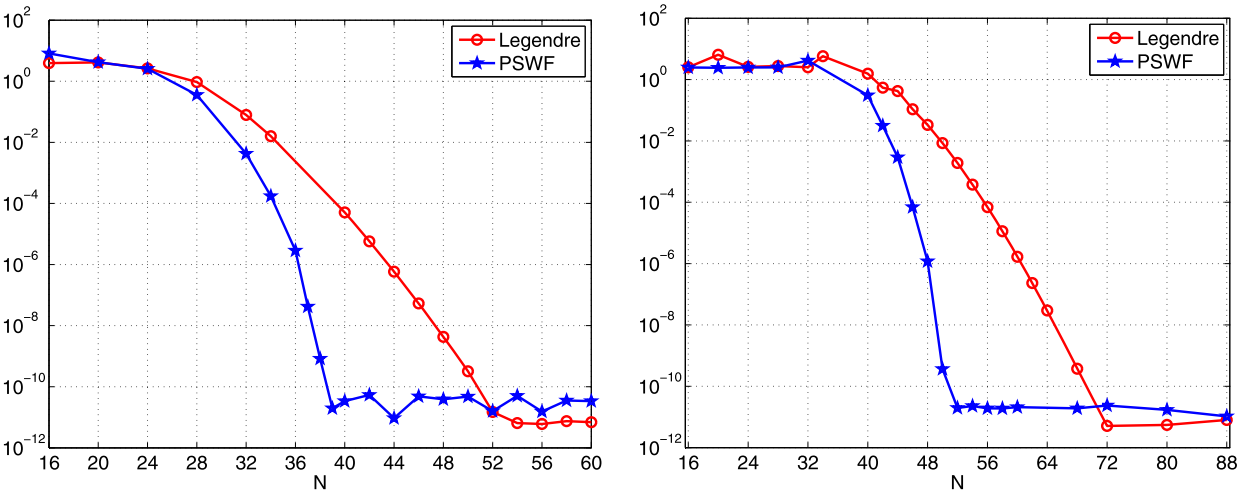

Fig. 5.6. Maximum point-wise errors of Legendre spectral-element and new prolate-element methods. Left: $k = 100$ and right: $k = 160$.

6. Concluding remarks and discussions

In this paper, we provided further illustrations for nonconvergence of h-refinement in prolate elements, which was first claimed by Boyd et al. [3]. We also proposed well-conditioned collocation and collocation-based p-version prolate-element methods using a new PSWF-basis. We demonstrated that the new approach with Kong–Rokhlin's rule of selecting (c, N) significantly outperformed the Legendre polynomial-based method in particular when the underlying solution is bandlimited. Advantages of our proposals were confirmed in solving the Helmholtz equations with high wave numbers in heterogeneous media.

The discussions in Section 5 were centered around the second-order BVPs. We conclude this paper with some remarks on how to devise the new basis for higher-order BVPs.

To fix the idea, we consider a fourth-order problem with given boundary data $\{u(\pm 1), u'(\pm 1)\}$. Like (5.4), we set

$$B_-(x) = \frac{(2+x)(1-x)^2}{4}, \qquad B_+(x) = B_-(-x),$$
$$\widehat{B}_-(x) = \frac{(1+x)(1-x)^2}{4}, \qquad \widehat{B}_+(x) = -\widehat{B}_-(-x), \tag{6.1}$$

which are the Hermite interpolation basis polynomials associated with two interpolating points $x = \pm 1$. In other words, letting

$$H(x) = u(-1)B_-(x) + u(1)B_+(x) + u'(-1)\widehat{B}_-(x) + u'(1)\widehat{B}_+(x),$$

we have $H(\pm1) = u(\pm1)$ and $H'(\pm1) = u'(\pm1)$. As a direct extension of (5.5), we look for $\{B_k\}_{k=1}^{N-1}$ such that

$$B_k \in Y_N^{c,0} := \operatorname{span}\big\{\varphi_n\colon\ \varphi_n^{(4)}(x) = \psi_n(x) \text{ with } \varphi_n(\pm1) = \varphi_n'(\pm1) = 0,\ 0 \leqslant n \leqslant N-2\big\}, \tag{6.2}$$

and

$$B_k(x_j) = \delta_{kj}, \quad 1 \leqslant j \leqslant N-1, \tag{6.3}$$

where $\{x_j\}$ are the PL points as before. To compute the new basis, we solve the fourth-order equation in (6.2), while this appears slightly more involved than the second-order case. Let us introduce the iterated integral operators:

$$\partial_x^{-1}u(x) = \int_{-1}^{x} u(y)\,dy, \qquad \partial_x^{-m}u(x) = \partial_x^{-1}\big(\partial_x^{1-m}u(x)\big), \quad m \geqslant 1.$$

Define $\Psi_n(x) = \partial_x^{-4}\psi_n(x)$, which can be computed straightforwardly from the integrated Legendre polynomials and then modifying the codes in Boyd [5]. Noting that $\Psi_n(-1) = \Psi_n'(-1) = 0$, we have the basis for $Y_N^{c,0}$:

$$\varphi_n(x) = \Psi_n(x) - \Psi_n(1)B_+(x) - \Psi_n'(1)\widehat{B}_+(x), \quad 0 \leqslant n \leqslant N-2.$$

Then we compute $\{B_k\}_{k=1}^{N-1}$ from $\{\varphi_n\}_{n=0}^{N-2}$ stably in a fashion very similar to (5.7)–(5.8).

With the new basis $\{B_\pm, \widehat{B}_\pm\} \cup \{B_k\}_{k=1}^{N-1}$ at our disposal, we are able to design well-conditioned collocation schemes for fourth-order BVPs as in Section 5.

Acknowledgements

The first author would like to thank the hospitality of Beijing Computational Science Research Centre during the visit in June 2013. The authors are grateful to the anonymous referees for many valuable comments.

Appendix A. Formulas for differentiation matrices

To this end, we derive the explicit formulas involving only function values $\{\psi_N(x_j)\}_{j=0}^{N}$ for computing the entries of the first-order and second-order differentiation matrices generated from the cardinal basis (2.22).

A direct derivation from (2.22) leads to

$$l_k'(x_j) = \begin{cases} \frac{1}{x_j - x_k}\frac{s'(x_j)}{s'(x_k)}, & \text{if } j \neq k, \\ \frac{s''(x_k)}{2s'(x_k)}, & \text{if } j = k, \end{cases} \tag{A.1}$$

where $s(x) = (1-x^2)\psi_N'(x)$. By (2.1),

$$s'(x) = \big(c^2x^2 - \chi_N\big)\psi_N(x), \qquad s''(x) = 2c^2x\,\psi_N(x) + \big(c^2x^2 - \chi_N\big)\psi_N'(x). \tag{A.2}$$

As $\{x_k\}_{k=1}^{N-1}$ are zeros of $\psi_N'(x)$, we have

$$s''(x_k) = 2c^2x_k\,\psi_N(x_k), \quad 1 \leqslant k \leqslant N-1. \tag{A.3}$$

Again by (2.1),

$$\psi_N'(-1) = -\frac{1}{2}\big(\chi_N - c^2\big)\psi_N(-1), \qquad \psi_N'(1) = \frac{1}{2}\big(\chi_N - c^2\big)\psi_N(1), \tag{A.4}$$

which, together with (A.2), implies

$$s''(-1) = \big(-2c^2 + \big(c^2 - \chi_N\big)^2/2\big)\psi_N(-1), \qquad s''(1) = \big(2c^2 - \big(c^2 - \chi_N\big)^2/2\big)\psi_N(1). \tag{A.5}$$

Then, (A.1) can be computed by

$$l_k'(x_j) = \begin{cases} -\frac{q^2}{q^2-1} + \frac{\chi_N}{4}(q^2-1), & \text{if } j = k = 0, \\ \frac{1}{x_j - x_k}\frac{q^2x_j^2 - 1}{q^2x_k^2 - 1}\frac{\psi_N(x_j)}{\psi_N(x_k)}, & \text{if } j \neq k,\ 0 \leqslant j,k \leqslant N, \\ \frac{q^2x_k}{q^2x_k^2 - 1}, & \text{if } 1 \leqslant j = k \leqslant N-1, \\ \frac{q^2}{q^2-1} - \frac{\chi_N}{4}(q^2-1), & \text{if } j = k = N, \end{cases} \tag{A.6}$$

where $q = c/\sqrt{\chi_N}$.

We now compute the entries of the second-order differentiation matrix. A direct differentiation of $s(x) = s'(x_k)(x - x_k)l_k(x)$ (cf. (2.22)) yields

$$s''(x) = s'(x_k)(x - x_k)l_k''(x) + 2s'(x_k)l_k'(x). \tag{A.7}$$

Therefore, for $j \neq k$,

$$l_k''(x_j) = \frac{1}{x_j - x_k}\left\{\frac{s''(x_j)}{s'(x_k)} - 2l_k'(x_j)\right\}, \tag{A.8}$$

so the off-diagonal entries of $\widehat{\boldsymbol{D}}^{(2)}$ can be computed from (A.2)–(A.6).

It remains to compute diagonal entries of $\widehat{\boldsymbol{D}}^{(2)}$. Differentiating (A.7) and letting $x = x_k$, gives

$$l_k''(x_k) = \frac{s'''(x_k)}{3s'(x_k)}, \quad 0 \leqslant k \leqslant N.$$

By (A.2),

$$s'''(x) = \big(c^2x^2 - \chi_N\big)\psi_N''(x) + 4c^2x\psi_N'(x) + 2c^2\psi_N(x). \tag{A.9}$$

For $1 \leqslant k \leqslant N-1$, we find from (2.1) and the fact $\psi_N'(x_k) = 0$ that

$$\psi_N''(x_k) = \frac{c^2x_k^2 - \chi_N}{1 - x_k^2}\psi_N(x_k), \quad \text{so} \quad s'''(x_k) = \left\{2c^2 + \frac{(c^2x_k^2 - \chi_N)^2}{1 - x_k^2}\right\}\psi_N(x_k),$$

which, together with (A.2), gives

$$l_k''(x_k) = \frac{s'''(x_k)}{3s'(x_k)} = \frac{2}{3}\frac{q^2}{q^2x_k^2 - 1} + \frac{\chi_N}{3}\frac{q^2x_k^2 - 1}{1 - x_k^2}, \quad 1 \leqslant k \leqslant N-1. \tag{A.10}$$

It is seen from (A.9) that the remaining two entries $l_0''(-1)$ and $l_N''(1)$ involve $\psi_N''(\pm1)$, which can also be represented by $\psi_N(\pm1)$. Indeed, differentiating (2.1) and letting $x = \pm1$, leads to

$$4\psi_N''(\pm1) = \pm\big(\chi_N - 2 - c^2\big)\psi_N'(\pm1) - 2c^2\psi_N(\pm1),$$

so by (A.4), $\psi_N''(\pm1)$ is a multiple of $\psi_N(\pm1)$. Finally, we get

$$l_0''(-1) = l_N''(1) = \frac{2q^2}{3(q^2-1)} + \frac{1}{24}\big(c^2 - \chi_N + 1\big)^2 - \frac{5}{6}c^2 - \frac{1}{24}, \tag{A.11}$$

where $q = c/\sqrt{\chi_N}$ as before.

Appendix B. Proof of Theorem 4.1

We derive from the definition (4.6) that

$$\big\|\boldsymbol{\pi}_{h,N}^c u - u\big\|_{L^2(a,b)}^2 = \sum_{i=1}^{M}\big\|\big(\boldsymbol{\pi}_{h,N}^c u\big)\big|_{I_i} - u^{I_i}\big\|_{L^2(I_i)}^2 = \frac{h}{2}\sum_{i=1}^{M}\big\|\hat{\pi}_N^c\hat{u}^{I_i} - \hat{u}^{I_i}\big\|_{L^2(I_{\rm ref})}^2. \tag{B.1}$$

Thus, it suffices to estimate $L^2(I_{\rm ref})$-orthogonal projection error in the reference interval $I_{\rm ref} = (-1,1)$. To do this, we recall the estimate in [34, Theorem 2.1]: if $c/\sqrt{\chi_n} \leqslant q_*/\sqrt[6]{2}$, then for any

$$\hat{u} \in B^\sigma(I_{\rm ref}) := \big\{\hat{u}\colon\ \big(1-y^2\big)^{k/2}\partial_y^k\hat{u}(y) \in L^2(I_{\rm ref}),\ 0 \leqslant k \leqslant \sigma\big\}, \quad \sigma \geqslant 0, \tag{B.2}$$

we have the estimate for the PSWF expansion coefficient in (4.4):

$$\big|\hat{u}_n(c)\big| \leqslant D\big(n^{-\sigma}\big\|\big(1-y^2\big)^{\sigma/2}\partial_y^\sigma\hat{u}\big\|_{L^2(I_{\rm ref})} + (q_*)^{\delta n}\|\hat{u}\|_{L^2(I_{\rm ref})}\big), \quad n \gg 1, \tag{B.3}$$

where D and δ are generic positive constants independent of $\hat{u}$, n and c. Then we have the following L^2-error estimate for the orthogonal projection defined in (4.4):

$$\big\|\hat{\pi}_N^c\hat{u} - \hat{u}\big\|_{L^2(I_{\rm ref})} \leqslant D\left(N^{1/2-\sigma}\big\|\big(1-y^2\big)^{\sigma/2}\partial_y^\sigma\hat{u}\big\|_{L^2(I_{\rm ref})} + \frac{1}{\sqrt{\delta\ln(1/q_*)}}(q_*)^{\delta N}\|\hat{u}\|_{L^2(I_{\rm ref})}\right). \tag{B.4}$$

for integer $\sigma \geqslant 1$. Indeed, by the orthogonality (2.2) and the bound (B.3),

$$\left\|\hat{\pi}_N^c \hat{u} - \hat{u}\right\|^2_{L^2(I_{\rm ref})} = \sum_{n=N+1}^{\infty} \left|\hat{u}_n(c)\right|^2 \leqslant D\left\{\left(\sum_{n=N+1}^{\infty} n^{-2\sigma}\right)\left\|(1-y^2)^{\sigma/2}\partial_y^\sigma \hat{u}\right\|^2_{L^2(I_{\rm ref})} + \left(\sum_{n=N+1}^{\infty} (q_*)^{2\delta n}\right)\|\hat{u}\|^2_{L^2(I_{\rm ref})}\right\}.$$

Since

$$\sum_{n=N+1}^{\infty} n^{-2\sigma} \leqslant \int_N^{\infty} \frac{1}{x^{2\sigma}}\,dx = \frac{1}{2\sigma-1}N^{1-2\sigma}, \quad \text{if } \sigma > \frac{1}{2},$$

and

$$\sum_{n=N+1}^{\infty} (q_*)^{2\delta n} \leqslant \int_N^{\infty} \left(q_*^2\right)^{\delta x}\,dx \leqslant \frac{1}{2\delta \ln(1/q_*)}(q_*)^{2\delta N},$$

we obtain (B.4).

One verifies readily from (4.3) that for $x \in I_i$ and $y \in I_{\rm ref}$,

$$\partial_y^\sigma \hat{u}^{I_i}(y) = \frac{h^\sigma}{2^\sigma}\partial_x^\sigma u^{I_i}(x), \qquad (1-y^2)^\sigma = 2^{2\sigma}\left(\frac{a_i - x}{h}\right)^\sigma \left(\frac{x - a_{i-1}}{h}\right)^\sigma \leqslant 2^{2\sigma}.$$

Then applying (B.4) to (B.1) leads to the desired result.

References

[1] M. Abramowitz, I. Stegun, Handbook of Mathematical Functions, Dover, New York, 1964.
[2] R.A. Adams, Sobolev Spaces, Academic Press, New York, 1975.
[3] J.P. Boyd, G. Gassner, B.A. Sadiq, The nonconvergence of h-refinement in prolate elements, J. Sci. Comput. 57 (2) (2013) 372–389.
[4] J.P. Boyd, Prolate spheroidal wavefunctions as an alternative to Chebyshev and Legendre polynomials for spectral element and pseudospectral algorithms, J. Comput. Phys. 199 (2) (2004) 688–716.
[5] J.P. Boyd, Algorithm 840: computation of grid points, quadrature weights and derivatives for spectral element methods using prolate spheroidal wave functions—prolate elements, ACM Trans. Math. Softw. 31 (1) (2005) 149–165.
[6] C. Canuto, M.Y. Hussaini, A. Quarteroni, T.A. Zang, Spectral Methods: Fundamentals in Single Domains, Springer-Verlag, Berlin, 2006.
[7] Q.Y. Chen, D. Gottlieb, J.S. Hesthaven, Spectral methods based on prolate spheroidal wave functions for hyperbolic PDEs, SIAM J. Numer. Anal. 43 (5) (2005) 1912–1933.
[8] H. Cheng, V. Rokhlin, N. Yarvin, Nonlinear optimization, quadrature, and interpolation, SIAM J. Optim. 9 (4) (1999) 901–923.
[9] C.W. Clenshaw, The numerical solution of linear differential equations in Chebyshev series, Math. Proc. Camb. Philos. Soc. 53 (1957) 134–149.
[10] F.A. Costabile, E. Longo, A Birkhoff interpolation problem and application, Calcolo 47 (1) (2010) 49–63.
[11] M.E. Elbarbary, Integration preconditioning matrix for ultraspherical pseudospectral operators, SIAM J. Sci. Comput. 28 (3) (2006) 1186–1201.
[12] A. Erdélyi, W. Magnus, F. Oberhettinger, F.G. Tricomi, Higher Transcendental Functions, McGraw-Hill, New York, 1953.
[13] L. Fox, I.B. Parker, Chebyshev Polynomials in Numerical Analysis, Oxford University Press, 1968.
[14] D. Gottlieb, S.A. Orszag, Numerical Analysis of Spectral Methods, SIAM, Philadelphia, PA, 1977.
[15] L. Greengard, Spectral integration and two-point boundary value problems, SIAM J. Numer. Anal. 28 (4) (1991) 1071–1080.
[16] H.D. Han, Z.Y. Huang, A tailored finite point method for the Helmholtz equation with high wave numbers in heterogeneous medium, J. Comput. Math. 26 (5) (2008) 728–739.
[17] J. Hesthaven, Integration preconditioning of pseudospectral operators. I. Basic linear operators, SIAM J. Numer. Anal. 35 (4) (1998) 1571–1593.
[18] Y.Y. Ji, H. Wu, H.P. Ma, B.Y. Guo, Multidomain pseudospectral methods for nonlinear convection–diffusion equations, Appl. Math. Mech. 32 (10) (2011) 1255–1268.
[19] W.Y. Kong, V. Rokhlin, A new class of highly accurate differentiation schemes based on the prolate spheroidal wave functions, Appl. Comput. Harmon. Anal. 33 (2) (2012) 226–260.
[20] N. Kovvali, W. Lin, Z. Zhao, L. Couchman, L. Carin, Rapid prolate pseudospectral differentiation and interpolation with the fast multipole method, SIAM J. Sci. Comput. 28 (2) (2006) 485–497.
[21] H.J. Landau, H.O. Pollak, Prolate spheroidal wave functions, Fourier analysis and uncertainty. III, Bell Syst. Tech. J. 41 (4) (1962) 1295–1336.
[22] G.G. Lorentz, K. Jetter, S.D. Riemenschneider, Birkhoff Interpolation, Cambridge University Press, 1984.
[23] A. Osipov, V. Rokhlin, On the evaluation of prolate spheroidal wave functions and associated quadrature rules, Appl. Comput. Harmon. Anal. 36 (1) (2014) 108–142.
[24] A. Osipov, V. Rokhlin, H. Xiao, Prolate Spheroidal Wave Functions of Order Zero, Springer Ser. Appl. Math. Sci., vol. 187, Springer, 2013.
[25] C. Pozrikidis, Introduction to Finite and Spectral Element Methods Using MATLAB, Chapman and Hall/CRC, 2005.
[26] V. Rokhlin, H. Xiao, Approximate formulae for certain prolate spheroidal wave functions valid for large values of both order and band-limit, Appl. Comput. Harmon. Anal. 22 (1) (2007) 105–123.
[27] J. Shen, Efficient spectral-Galerkin method I. Direct solvers for second- and fourth-order equations by using Legendre polynomials, SIAM J. Sci. Comput. 15 (1994) 1489–1505.
[28] J. Shen, T. Tang, L.L. Wang, Spectral Methods: Algorithms, Analysis and Applications, Springer Ser. Comput. Math., vol. 41, Springer-Verlag, Berlin, Heidelberg, 2011.
[29] D. Slepian, Prolate spheroidal wave functions, Fourier analysis and uncertainity, IV: extensions to many dimensions generalized prolate spheroidal functions, Bell Syst. Tech. J. 43 (1964) 3009–3057.
[30] D. Slepian, Some comments on Fourier analysis, uncertainty and modeling, SIAM Rev. 25 (3) (1983) 379–393.
[31] D. Slepian, H.O. Pollak, Prolate spheroidal wave functions, Fourier analysis and uncertainty. I, Bell Syst. Tech. J. 40 (1961) 43–63.
[32] L.L. Wang, Analysis of spectral approximations using prolate spheroidal wave functions, Math. Comput. 79 (270) (2010) 807–827.

[33] L.L. Wang, M. Samson, X.D. Zhao, A well-conditioned collocation method using pseudospectral integration matrix, SIAM J. Sci. Comput. (2014), forthcoming.
[34] L.L. Wang, J. Zhang, An improved estimate of PSWF approximation and approximation by Mathieu functions, J. Math. Anal. Appl. 379 (1) (2011) 35–47.
[35] J.A.C. Weideman, L.N. Trefethen, The eigenvalues of second-order spectral differentiation matrices, SIAM J. Numer. Anal. 25 (6) (1988) 1279–1298.
[36] B.D. Welfert, On the eigenvalues of second-order pseudospectral differentiation operators, Comput. Methods Appl. Mech. Eng. 116 (1) (1994) 281–292.
[37] H. Xiao, V. Rokhlin, N. Yarvin, Prolate spheroidal wavefunctions, quadrature and interpolation, Inverse Probl. 17 (4) (2001) 805–838.
[38] A. Zebib, A Chebyshev method for the solution of boundary value problems, J. Comput. Phys. 53 (3) (1984) 443–455.
[39] J. Zhang, L.L. Wang, Z.J. Rong, A prolate-element method for nonlinear PDEs on the sphere, J. Sci. Comput. 47 (1) (2011) 73–92.
[40] Z. Zhang, Superconvergence points of polynomial spectral interpolation, SIAM J. Numer. Anal. 50 (5) (2012) 2966–2985.
[41] Z. Zhang, How many numerical eigenvalues can we trust? arXiv:1312.6773, 2013.