J Sci Comput (2017) 70:451–477
DOI 10.1007/s10915-016-0253-2

Optimal Spectral Schemes Based on Generalized Prolate Spheroidal Wave Functions of Order −1

Jing Zhang[1] · **Li-Lian Wang**[2] · **Huiyuan Li**[3] · **Zhimin Zhang**[4,5]

Received: 20 March 2016 / Revised: 6 July 2016 / Accepted: 19 July 2016 / Published online: 25 July 2016
© Springer Science+Business Media New York 2016

Abstract We introduce a family of generalized prolate spheroidal wave functions (PSWFs) of order −1, and develop new spectral schemes for second-order boundary value problems. Our technique differs from the differentiation approach based on PSWFs of order zero in Kong and Rokhlin (Appl Comput Harmon Anal 33(2):226–260, 2012); in particular, our orthogonal basis can naturally include homogeneous boundary conditions without the re-orthogonalization of Kong and Rokhlin (2012). More notably, it leads to diagonal systems or direct "explicit" solutions to 1D Helmholtz problems in various situations. Using a rule optimally pairing the bandwidth parameter and the number of basis functions as in Kong and Rokhlin (2012), we demonstrate that the new method significantly outperforms the Legendre spectral method in approximating highly oscillatory solutions. We also conduct a rigorous

Jing Zhang: The work of this author is partially supported by the National Natural Science Foundation of China (11201166), and the Fundamental Research Funds for the Central Universities (CCNU15A02033).
Li-Lian Wang: The research of this author is partially supported by Singapore MOE AcRF Tier 1 Grants (RG 15/12 and 27/15), and Singapore MOE AcRF Tier 2 Grant (MOE 2013-T2-1-095, ARC 44/13).
Huiyuan Li: The research of this author is partially supported by the National Natural Science Foundation of China (91130014, 11471312 and 91430216).
Zhimin Zhang: The research of this author is supported in part by the National Natural Science Foundation of China (11471031 and 91430216), and the U.S. National Science Foundation (DMS-1419040). The first author would like to thank the supports from both Beijing Computational Sciences and Research Center, Beijing, China and Division of Mathematical Sciences of Nanyang Technological University, Singapore.

✉ Li-Lian Wang
LiLian@ntu.edu.sg

1 School of Mathematics and Statistics and Hubei Key Laboratory of Mathematical Sciences, Central China Normal University, Wuhan 430079, China

2 Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore

3 State Key Laboratory of Computer Science/Laboratory of Parallel Computing, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China

4 Beijing Computational Sciences and Research Center, Beijing 100193, China

5 Department of Mathematics, Wayne State University, Detroit, MI 48202, USA

error analysis of this new scheme. The idea and analysis can be extended to generalized PSWFs of negative integer order for higher-order boundary value and eigenvalue problems.

Keywords Generalized prolate spheroidal wave functions of order -1 · Helmholtz equations · Optimal spectral schemes · Error analysis

Mathematics Subject Classification 65N35 · 65E05 · 65M70 · 41A05 · 41A10 · 41A25

1 Introduction

There have been limited studies and applications of the prolate spheroidal wave functions of order zero (PSWFs) during the first three decades after the seminal works of Slepian et al. (see, e.g., [17,19] in early 1960s). The renewed interest in PSWFs is evident in the monographs by Osipov, Rokhlin and Xiao [14], and Hogan and Lakey [10]. The former provided an up-to-date exposition of the analytic and asymptotic properties, and numerics of PSWFs (but touched only briefly on the wide-ranging applications of PSWFs). The latter elaborated on the applications of PSWFs in sampling and signal processing. We also refer to [7] for a review of many recent publications on PSWFs of order zero.

On the one hand, the PSWFs are eigenfunctions of an integral operator related to the finite Fourier transform and time–frequency concentration problem (cf. [18]). They naturally form an orthogonal basis of the Paley–Wiener space of c-bandlimited functions, and offer an optimal tool for approximating bandlimited functions.

On the other hand, the PSWFs are eigenfunctions of a singular Sturm–Liouville problem, so they are born with the orthogonality and completeness properties in L^2-space as do their counterparts, Legendre polynomials. Indeed, spectral approximations using PSWFs have exhibited some advantages over the Legendre polynomial-based methods: (i) they enable fewer number of points per wavelength to resolve waves; and (ii) they use quasi-uniformly distributed collocation points allowing for larger time steps in explicit time-marching schemes. However, the PSWFs are non-polynomials, which lack some attractive properties of orthogonal polynomials, e.g., three-term recurrence formulas, simple derivative relations, etc. Consequently, one has a trade-off between efficiency and advantage at times. Moreover, the PSWFs might lose certain abilities of polynomials. For instance, Boyd et al. [7] discovered that prolate-element methods are nonconvergent when h-refinement is used in hp-approximation.

We highlight some important attempts to improve the efficiency and performance of PSWF-based algorithms. Kong and Rokhlin [13] proposed a class of accurate prolate spectral-differentiation schemes with the following features:

(i) It is based on the PSWF expansion of a function whose expansion coefficients are determined by a least-square procedure using an oversampling of the nodal values.
(ii) For a given bandwidth c, the approximation with accuracy ε is attained by a minimal number of terms in the PSWF expansion.
(iii) In order to incorporate the homogenous boundary conditions, e.g., $u(\pm 1) = 0$, a linear function: $\mu_j(x) = \psi_j(-1) + (\psi_j(1) - \psi_j(-1))(1+x)/2$, is subtracted from PSWF $\psi_j(x)$, and then a Gram–Schmidt orthogonalization is used to reconstruct an orthogonal basis so that the differentiation in (i) can be implemented.
(iv) The differentiation matrix has much smaller spectral radius. For example, the second-order one has a reduction of spectral radius from $O(N^4)$ to $O(N^2)$.

Following the spirit of [22] (for polynomials), Wang et al. [24] constructed a new PSWF-type basis from a PSWF-Birkhoff interpolation that led to well-conditioned prolate-collocation schemes.

Our approach is different from the above works. We introduce in this paper a family of generalized PSWFs of order -1 (denoted by $\{\psi_n^{(-1)}(x;c)\}_{n=0}^{\infty}$), and construct new spectral-Galerkin schemes for boundary value problems (BVPs) and eigenvalue problems. We outline below the main contributions.

- The generalized PSWFs of order -1 are defined as the eigenfunctions of a Sturm–Liouville problem. They are mutually orthogonal in $L^2_{\omega_{-1}}(-1,1)$ with $\omega_{-1}=(1-x^2)^{-1}$. They naturally build in the boundary conditions: $u(\pm 1)=0$. The matrix of a spectral-Galerkin scheme for the Helmholtz operator d^2/dx^2+c^2 becomes diagonal under the new basis. Thus, it offers an optimal spectral-algorithm for 1D Helmholtz problems with arbitrary high wavenumber c. Indeed, we demonstrate that using the Kong–Rokhlin rule for pairing (c,n) in [13], we can achieve an accuracy and resolution superior to the Legendre–Galerkin method.
- We provide a rigorous error analysis of the new basis in error bounds with an explicit dependence on the bandwidth parameter, thereby justifying that under the rule similar to [13], spectral accuracy can be attained.
- When the bandwidth parameter $c=0$, our approach reduces to the optimal spectral algorithm in [9]. In fact, we can define generalized PSWFs of negative integer order and extend our ideas to solve higher-order BVPs and eigenvalue problems.

The paper is organized as follows. In Sect. 2, we introduce the generalized PSWFs of order -1 and study their properties. In Sect. 3, we construct the spectral schemes for 1D Helmholtz equations and eigenvalue problems, and provide ample numerical results to demonstrate the significant gain in accuracy when one shifts from Legendre approximation to the new method. We conduct error analysis of approximation by the new basis in Sect. 4. We conclude the paper with some remarks and possible extensions.

2 Generalized PSWFs of Order -1 and Their Properties

In this section, we introduce the generalized PSWFs of order -1, and highlight some properties. We also present a rule for optimally truncating the generalized PSWF expansion of a function with bandwidth c and for a given error tolerance ε.

2.1 Generalized PSWFs of Order -1

Define the second-order differential operator:

$$\mathscr{D}_c^{(-1)} := -(1-x^2)\frac{d^2}{dx^2}+c^2x^2 = -(1-x^2)\left(\frac{d^2}{dx^2}+c^2\right)+c^2, \tag{2.1}$$

for $x\in I:=(-1,1)$, and real $c\geq 0$. Consider the eigen-value problem: find $\{\chi,u\}$ such that

$$\mathscr{D}_c^{(-1)}[u](x)=\chi\, u(x),\quad x\in I,\quad u(\pm 1)=0. \tag{2.2}$$

Note that $\mathscr{D}_c^{(-1)}$ is a positive, self-adjoint operator, as for any u, v in the domain of $\mathscr{D}_c^{(-1)}$, we have

$$\big(\mathscr{D}_c^{(-1)}u, v\big)_{\omega_{-1}} = \big(u, \mathscr{D}_c^{(-1)}v\big)_{\omega_{-1}}; \quad \big(\mathscr{D}_c^{(-1)}u, u\big)_{\omega_{-1}} = \|u'\|^2 + c^2\|xu\|^2_{\omega_{-1}} > 0, \quad \forall\, u \neq 0, \tag{2.3}$$

where $(\cdot,\cdot)_{\omega_{-1}}$ and $\|\cdot\|_{\omega_{-1}}$ are the inner product and norm of $L^2_{\omega_{-1}}(I)$ with the weight function $\omega_{-1}(x) = (1-x^2)^{-1}$, respectively. According to the general theory of Sturm–Liouville problems (cf. [2,8]), the eigen-problem (2.2) admits a countable set of eigen-pairs: $\Big\{\chi_n^{(-1)}(c),\, \psi_n^{(-1)}(x;c)\Big\}_{n=0}^{\infty}$, such that

$$\mathscr{D}_c^{(-1)}\Big[\psi_n^{(-1)}\Big](x) = \chi_n^{(-1)}(c)\,\psi_n^{(-1)}(x;c), \quad x \in I; \quad c \geq 0. \tag{2.4}$$

Moreover, we have the following properties common to a Sturm–Liouville problem.

- The eigenfunctions $\Big\{\psi_n^{(-1)}(x;c)\Big\}_{n=0}^{\infty}$ are sufficiently smooth, and form a complete orthonormal system of $L^2_{\omega_{-1}}(I)$, namely,

$$\int_{-1}^{1} \psi_m^{(-1)}(x;c)\psi_n^{(-1)}(x;c)(1-x^2)^{-1}\,dx = \delta_{mn}. \tag{2.5}$$

- The eigenfunctions $\Big\{\psi_n^{(-1)}(x;c)\Big\}_{n=0}^{\infty}$ with even n are even functions of x, and those with odd n are odd.
- The eigenvlaues $\Big\{\chi_n^{(-1)}(c)\Big\}_{n=0}^{\infty}$ are all real, positive, simple and arranged in ascending order. Moreover, $\lim_{n\to\infty}\chi_n^{(-1)}(c) = \infty$.

We call $\psi_n^{(-1)}(x;c)$ *the generalized PSWF of order* -1 *(and of degree* n*)*, where the parameter c is called the (generalized) bandwidth parameter.

Remark 2.1 Wang and Zhang [23] introduced the generalized PSWFs of order $\alpha > -1$. Very recently, Karoui and Souabni [12] extended the generalisation in [23] to a more general setting and related these functions to a generalized energy concentration problem. More precisely, the generalized PSWFs of order $\alpha > -1$ and of degree n in [23], are eigenfunctions of the singular Sturm–Liouville problem:

$$\mathscr{D}_c^{(\alpha)}[\psi_n^{(\alpha)}](x) = \Big\{-(1-x^2)\partial_x^2 + 2(\alpha+1)x\partial_x + c^2x^2\Big\}\,\psi_n^{(\alpha)}(x) = \chi_n^{(\alpha)}(c)\,\psi_n^{(\alpha)}(x), \tag{2.6}$$

for $\alpha > -1$, $c > 0$ and $x \in I$. The generalized PSWFs form a complete orthonormal system in $L^2_{\omega_\alpha}(I)$ (with the weight function $\omega_\alpha(x) = (1-x^2)^\alpha$):

$$\int_{-1}^{1} \psi_m^{(\alpha)}(x)\psi_n^{(\alpha)}(x)\omega_\alpha(x)dx = \delta_{mn}, \quad \alpha > -1. \tag{2.7}$$

Note that for $c = 0$, $\psi_n^{(\alpha)}(x;0) = P_n^{(\alpha)}(x)$, the normalized ultraspherical polynomial of degree n (cf. [20]), and $\chi_n^{(\alpha)}(0) = n(n+2\alpha+1)$.

The generalized PSWFs of order $\alpha > -1$ are eigenfunctions of the integral operator:

$$\mathscr{F}_c^{(\alpha)}[\psi_n^{(\alpha)}](x) = \int_{-1}^{1} e^{\mathrm{i}ctx}\psi_n^{(\alpha)}(t)\omega_\alpha(t)dt = \mathrm{i}^n\lambda_n^{(\alpha)}(c)\psi_n^{(\alpha)}(x), \quad x \in I, \quad c > 0, \tag{2.8}$$

where the eigenvalues $\{\lambda_n^{(\alpha)} := \lambda_n^{(\alpha)}(c)\}$ (modulo the factor i^n) are all real, positive, simple and in descending order (cf. [23]).

We refer to [12,23] for many more properties. In particular, the PSWFs of order zero (i.e., $\alpha = 0$) are well documented in literature, see [17–19,26] and the monographs [10,14]. □

The following proposition shows the intimate relation between generalized PSWFs of order -1 and of order 1.

Proposition 2.1 *Let* $\left\{\psi_n^{(1)}(x;c)\right\}$ *be the generalized PSWFs of order 1, and* $\{\chi_n^{(1)}(c)\}$ *be the corresponding eigenvalues. Then we have*

$$\psi_n^{(-1)}(x;c) = (1-x^2)\psi_n^{(1)}(x;c); \quad \chi_n^{(-1)}(c) = \chi_n^{(1)}(c) + 2, \tag{2.9}$$

for all $x \in I$, $n \geq 0$ *and* $c > 0$.

Proof One verifies readily from (2.1) and (2.6) with $\alpha = 1$ that

$$\begin{aligned}\mathscr{D}_c^{(-1)}[(1-x^2)v](x) &= (1-x^2)\left\{-(1-x^2)v''(x) + 4xv(x) + 2v(x) + c^2x^2v(x)\right\}\\ &= (1-x^2)\mathscr{D}_c^{(1)}[v](x) + 2(1-x^2)v(x).\end{aligned}$$

Thus, letting $v(x) = \psi_n^{(1)}(x)$ in the above, we derive from (2.6) with $\alpha = 1$ that

$$\begin{aligned}\mathscr{D}_c^{(-1)}\left[(1-x^2)\psi_n^{(1)}\right](x) &= (1-x^2)\mathscr{D}_c^{(1)}\left[\psi_n^{(1)}\right](x) + 2(1-x^2)\psi_n^{(1)}(x)\\ &= \left(\chi_n^{(1)} + 2\right)(1-x^2)\psi_n^{(1)}(x).\end{aligned}$$

Comparing with (2.4), we infer that

$$\chi_n^{(-1)}(c) = \chi_n^{(1)}(c) + 2, \quad \psi_n^{(-1)}(x;c) = C_n(1-x^2)\psi_n^{(1)}(x;c),$$

where C_n is any nonzero constant. By (2.5) and (2.7) (with $\alpha = 1$), we have $C_n^2 = 1$. Here, we take $C_n = 1$ by noting that the definition of these PSWFs can differ from a sign. □

Thanks to (2.9), the properties of $\psi_n^{(1)}(x;c)$ can be transplanted to $\psi_n^{(-1)}(x;c)$. Indeed, we derive from (2.8) that the generalized PSWFs of order -1 are eigenfunctions of the following operator.

Proposition 2.2 *We have*

$$\mathscr{F}_c^{(-1)}\left[\psi_n^{(-1)}\right](x) := (1-x^2)\int_{-1}^{1} e^{\mathrm{i}cxt}\psi_n^{(-1)}(t)\,dt = \mathrm{i}^n\lambda_n^{(1)}\,\psi_n^{(-1)}(x), \tag{2.10}$$

for all $x \in I$, $n \geq 0$ *and* $c > 0$.

When $c = 0$, the generalized PSWF of order -1 reduces to the integrated Legendre polynomial.

Proposition 2.3 *For* $c = 0$, *we have*

$$\psi_n^{(-1)}(x;0) = \frac{1}{\sqrt{(n+1)(n+2)}}\int_{-1}^{x}\bar{P}_{n+1}(t)\,dt, \quad n \geq 0, \tag{2.11}$$

where $\bar{P}_{n+1}$ *is the Legendre polynomial of degree* $n+1$.

Proof Denote $\phi_n(x) = \int_{-1}^{x} \bar{P}_{n+1}(t)\,dt$. By [15, (3.171)], we have

$$\begin{aligned} -(1-x^2)\phi_n''(x) &= -(1-x^2)\bar{P}_{n+1}'(x) = (n+1)(n+2)\int_{-1}^{x} \bar{P}_{n+1}(x)dx \\ &= (n+1)(n+2)\phi_n(x). \end{aligned} \tag{2.12}$$

Let $\mathscr{D}_0^{(-1)}$ be the operator defined in (2.1) with $c = 0$. Then by (2.12),

$$\mathscr{D}_0^{(-1)}[\phi_n] = (n+1)(n+2)\phi_n. \tag{2.13}$$

Noting that $\chi_n^{(-1)}(0) = (n+1)(n+2)$ (cf. (2.9)), we infer from (2.4) with $c = 0$ and (2.12) that $\psi_n^{(-1)}(x;0) = C_n\phi_n(x)$, where the constant C_n is determined by the normalization (2.5). Then we can work out this constant by using the properties of normalized Legendre polynomials and derive (2.11). □

2.2 Evaluation of Generalized PSWFs and Their Eigenvalues

Thanks to (2.9), we can compute $\left\{\psi_n^{(-1)}(x;c),\, \chi_n^{(-1)}(c)\right\}$ from $\left\{\psi_n^{(1)}(x;c),\, \chi_n^{(1)}(c)\right\}$ via the Bouwkamp algorithm (cf. [5,23]). Here, we sketch this algorithm and the related formulas will be useful later on.

Let $P_k^{(1)}(x)$ be the normalized ultraspherical polynomial with $\alpha = 1$ defined by the three-term recurrence relation (cf. [20]):

$$x\,P_k^{(1)}(x) = a_k P_{k+1}^{(1)}(x) + b_k P_{k-1}^{(1)}(x), \quad k \ge 1; \quad P_0^{(1)}(x) = \frac{\sqrt{3}}{2}, \quad P_1^{(1)}(x) = \frac{\sqrt{15}}{2}x, \tag{2.14}$$

where

$$a_k = \sqrt{\frac{(k+1)(k+3)}{(2k+3)(2k+5)}}, \quad b_k = \sqrt{\frac{k(k+2)}{(2k+1)(2k+3)}}. \tag{2.15}$$

We expand $\psi_n^{(1)}(x)$ as

$$\psi_n^{(1)}(x) = \sum_{k=0}^{\infty} \beta_k^n P_k^{(1)}(x) \quad \text{with} \quad \beta_k^n = \int_{-1}^{1} \psi_n^{(1)}(x) P_k^{(1)}(x)(1-x^2)\,dx. \tag{2.16}$$

Observe from the parity that $\beta_k^n = 0$, if $n+k$ is odd. From (2.14) and the property (cf. [20]):

$$-(1-x^2)^{-1}\frac{d}{dx}\left((1-x^2)^2\frac{d}{dx}P_k^{(1)}(x)\right) = k(k+3)P_k^{(1)}(x), \tag{2.17}$$

we infer that (2.6) with $\alpha = 1$ is equivalent to the matrix eigenvalue problem:

$$\boldsymbol{A}\boldsymbol{\beta}_n = \chi_n^{(1)}\,\boldsymbol{\beta}_n\,, \quad \forall\, n \ge 0, \tag{2.18}$$

where $\boldsymbol{\beta}_n = (\beta_0^n, \beta_1^n, \ldots)^t$, and $\boldsymbol{A}$ is an infinite symmetric peta-diagonal matrix with non-zeros entries given by

$$\begin{aligned} \tilde{a}_{k,k} &= k(k+3) + \frac{2k(k+3)+1}{(2k+1)(2k+5)} \times c^2, \\ \tilde{a}_{k,k+2} &= \tilde{a}_{k+2,k} = \sqrt{\frac{(k+1)(k+2)(k+3)(k+4)}{(2k+3)(2k+5)^2(2k+7)}} \times c^2. \end{aligned} \tag{2.19}$$

The infinite system (2.18) can be decomposed into two symmetric tridiagonal systems:

$$\boldsymbol{A}^e \boldsymbol{\beta}_n^e = \chi_n^e \boldsymbol{\beta}_n^e, \quad n = 2l; \quad \boldsymbol{A}^o \boldsymbol{\beta}_n^o = \chi_n^o \boldsymbol{\beta}_n^o, \quad n = 2l+1, \tag{2.20}$$

where $\boldsymbol{A}^e$ (resp. $\boldsymbol{A}^o$) consists of even-numbered (resp. odd-numbered) rows and columns of $\boldsymbol{A}$, and $\boldsymbol{\beta}_n^e = (\beta_0^n, \beta_2^n, \ldots)^t$ (resp. $\boldsymbol{\beta}_n^o = (\beta_1^n, \beta_3^n, \ldots)^t$).

In the computation, we have to reduce the infinite eigen-system (2.18) with a suitable cut-off number $M > N$. More precisely, we take the first $(N+1)$ eigen-pairs from $(M+1)$ pairs, and set

$$\psi_n^{(1)}(x) = \sum_{k=0}^{M} \beta_k^n \, P_k^{(1)}(x), \quad 0 \le n \le N, \quad \text{i.e.,} \quad \boldsymbol{\psi}^{(1)}(x) = \boldsymbol{B}\boldsymbol{P}^{(1)}(x), \tag{2.21}$$

where $\boldsymbol{B}$ is an $(N+1)$-by-$(M+1)$ matrix, and

$$\boldsymbol{B} = \left(\beta_k^n\right)_{0\le n\le N}^{0\le k\le M}, \quad \boldsymbol{\psi}^{(1)} = \left(\psi_0^{(1)}, \ldots, \psi_N^{(1)}\right)^t, \quad \boldsymbol{P}^{(1)} = \left(P_0^{(1)}, \ldots, P_M^{(1)}\right)^t. \tag{2.22}$$

Remark 2.2 Boyd [6] suggested a conservative cut-off number: $M = 2N + 30$, which guaranteed a machine zero accuracy for computing the PSWFs of order zero for all $0 < c \le c_*(N) = \pi\,(N + 1/2)\,/2$. This "transition bandwidth" $c_*(N)$ also plays an important role in understanding the decay rate of $\lambda_n(c) := \lambda_n^{(0)}(c)$. Note that by [12, Theorem 2], this rule still works for $\alpha = 1$, as $\lambda_n^{(1)}(c) \le \lambda_n^{(0)}(c) = \lambda_n(c)$. □

The following formula (cf. [23]) provides a stable way to compute $\{\lambda_n^{(1)}(c)\}$:

$$\lambda_n^{(1)}(c) = \begin{cases} \dfrac{2\,\beta_0^n}{\mathrm{i}^n\,\sqrt{3}\,\psi_n^{(1)}(0;\,c)}, & \text{if } n \text{ is even,} \\ \dfrac{2c\,\beta_1^n}{\mathrm{i}^{n-1}\,\sqrt{15}\,\partial_x\psi_n^{(1)}(0;\,c)}, & \text{if } n \text{ is odd,} \end{cases} \tag{2.23}$$

where β_0^n and β_1^n are given in (2.16). □

2.3 Optimal N for Given Bandwidth Parameter c

An important issue related to the application of PSWFs of order zero is the choice of bandwidth parameter c and the number of basis functions N. In most cases, they are independently chosen and sometimes on a trial-and-error basis. Kong and Rokhlin [13] proposed a useful rule for pairing up c and N from a quadrature rule involving PSWFs of zero. More precisely, for given c, one can control the accuracy of prolate-quadrature rule within a prescribed error tolerance ε by choosing the smallest number of points $N_* := N_*(c, \varepsilon)$ such that

$$\lambda_{N_*}(c) \le \varepsilon \le \lambda_{N_*-1}(c). \tag{2.24}$$

It is noteworthy that Wang et al. [24] introduced a practical mean to implement this rule without computing the eigenvalues to find N_* for $c < c_*(N)$.

In this context, we examine this issue from the perspective of best generalized PSWF approximation to c-bandlimited functions of the type:

$$u(x) = \mathscr{F}_c^{(\alpha)}[\phi](x) \quad \text{for} \quad \phi \in L^2_{\omega_\alpha}(I), \quad \alpha = \pm 1, \tag{2.25}$$

where the integral $\mathscr{F}_c^{(\alpha)}$ is defined in (2.8) and (2.10).

We write

$$u(x) = \sum_{n=0}^{\infty} \hat{u}_n^{(\alpha)} \psi_n^{(\alpha)}(x;c) \quad \text{with} \quad \hat{u}_n^{(\alpha)} := \hat{u}_n^{(\alpha)}(c) = \int_{-1}^{1} u(x)\psi_n^{(\alpha)}(x;c)\omega_\alpha(x)dx, \quad (2.26)$$

and likewise for $\phi(x)$ with the expansion coefficients $\{\hat{\phi}_n^{(\alpha)}\}$.

Theorem 2.1 *Let $u(x)$ be defined in (2.25) with bandwidth $c > 0$, and denote $(\pi_{N,c}^{(\alpha)} u)(x) = \sum_{n=0}^{N-1} \hat{u}_n^{(\alpha)} \psi_n^{(\alpha)}(x;c)$ with $\alpha = \pm 1$. Then we have*

$$\left\| u - \pi_{N,c}^{(\alpha)} u \right\|_{\omega_\alpha} \le \lambda_N^{(1)}(c) \|\phi\|_{\omega_\alpha}, \quad \alpha = \pm 1. \quad (2.27)$$

Proof We just provide the proof for $\alpha = -1$ below, as the case with $\alpha = 1$ can be shown in the same fashion. One verifies readily from (2.10) and (2.25) that

$$\begin{aligned}
\hat{u}_n^{(-1)} &= \int_{-1}^{1} \mathscr{F}_c^{(-1)}[\phi](x)\psi_n^{(-1)}(x;c)\omega_{-1}(x)\,dx = \int_{-1}^{1} \left[\int_{-1}^{1} e^{\mathrm{i}cxt}\phi(t)\,dt \right] \psi_n^{(-1)}(x;c)\,dx \\
&= \int_{-1}^{1} \left[\int_{-1}^{1} e^{\mathrm{i}cxt}\psi_n^{(-1)}(x;c)dx \right] \phi(t)\,dt \overset{(2.10)}{=} \mathrm{i}^n \lambda_n^{(1)}(c) \int_{-1}^{1} \phi(t)\psi_n^{(-1)}(t;c)\omega_{-1}(t)\,dt \\
&= \mathrm{i}^n \lambda_n^{(1)}(c)\hat{\phi}_n^{(-1)}.
\end{aligned}$$

Hence, by the orthogonality (2.5) and the decay of $\lambda_n^{(1)}(c)$,

$$\left\| u - \pi_{N,c}^{(-1)} u \right\|_{\omega_{-1}}^2 = \sum_{n=N}^{\infty} |\hat{u}_n^{(-1)}|^2 = \sum_{n=N}^{\infty} |\lambda_n^{(1)}(c)|^2 |\hat{\phi}_n^{(-1)}|^2 \le \left(\lambda_N^{(1)}(c)\right)^2 \|\phi\|_{\omega_{-1}}^2. \quad (2.28)$$

This ends the proof. □

For given $c > 0$, and a prescribed error tolerance ε, we follow the spirit of [13] and look for the optimal N_*, which is the smallest integer such that

$$\lambda_{N_*}^{(1)}(c) \le \varepsilon \le \lambda_{N_*-1}^{(1)}(c). \quad (2.29)$$

This ensures

$$\|u - \pi_{N_*,c}^{(\alpha)} u\|_{\omega_\alpha} = O(\varepsilon). \quad (2.30)$$

Remark 2.3 Note that $\{\lambda_n^{(1)}(c)\}$ is exponentially small for large n (cf. [23]):

$$\lambda_n^{(1)}(c) \approx \nu_n^{(1)}(c) := \sqrt{\frac{2\pi}{c}} \frac{e}{4} \left(\frac{2n+4}{2n+3}\right)^{\frac{3}{2}} \left(\frac{ce}{4n+6}\right)^{n+\frac{1}{2}} \exp\left(\frac{2n+3}{12(n+1)(n+2)}\right), \quad (2.31)$$

where $\nu_n^{(1)}(c)$ provides a good approximation to $\lambda_n^{(1)}(c)$ when $c < (4n+6)/e$. With this, we can extend the practice rule in [24] to find N_* in (2.29). □

3 New Spectral Schemes for Helmholtz Equations and Eigenvalue Problems

Equipped with the generalized PSWFs of order -1, we develop in this section optimal spectral-Galerkin schemes for one-dimensional Helmholtz equations in various situations, and elaborate on some remarkable advantages over the polynomial counterparts.

3.1 An Illustrative Model Problem and the Scheme

To fix the idea, we consider

$$\mathcal{H}_c[u](x) := -u''(x) - c^2 u(x) = f(x), \quad x \in I; \quad u(\pm 1) = 0, \tag{3.1}$$

where $c > 0$ is the wavenumber, and f is independent of c.

As usual, a weak form of (3.1) is to find $u \in H_0^1(I)$ such that

$$\mathcal{A}_c(u, v) := (u', v') - c^2(u, v) = (f, v), \quad \forall\, v \in H_0^1(I). \tag{3.2}$$

Introduce the approximation space

$$\mathcal{V}_N^0 := \text{span}\left\{\psi_n^{(-1)} : 0 \le n \le N-2\right\} \subseteq H_0^1(I). \tag{3.3}$$

The generalized PSWF-spectral scheme is to find $u_N \in \mathcal{V}_N^0$ such that

$$\mathcal{A}_c(u_N, v_N) = \left(\mathbb{I}_L^{(1)} f, v_N\right), \quad \forall\, v_N \in \mathcal{V}_N^0, \tag{3.4}$$

where $\mathbb{I}_L^{(1)} f$ is the polynomial interpolation of $f(x)$ on $(L+1)$ Gegenbauer–Gauss points $\{\xi_j\}_{j=0}^L$ associated with the weight function $\omega_1(x) = 1 - x^2$ (cf. [15, Ch. 3]).

We have the following markedly properties.

Theorem 3.1 *Let $\{\psi_n^{(\alpha)}(x; c)\}$ be the generalized PSWFs of order $\alpha = \pm 1$. Then there holds the orthogonality*

$$\mathcal{A}_c\left(\psi_n^{(-1)}, \psi_m^{(-1)}\right) = \sigma_n\, \delta_{mn}, \tag{3.5}$$

where

$$\sigma_n := \chi_n^{(-1)} - c^2 = \chi_n^{(1)} + 2 - c^2 \neq 0, \quad \forall\, c > 0, \quad \forall\, n \ge 0. \tag{3.6}$$

Define the stiffness matrix $\boldsymbol{S}$ and mass matrix $\boldsymbol{M}$ associated with $\mathcal{V}_N^0$ and with the entries

$$\boldsymbol{S}_{lj} = \left(\partial_x \psi_j^{(-1)}, \partial_x \psi_l^{(-1)}\right), \quad \boldsymbol{M}_{lj} = \left(\psi_j^{(-1)}, \psi_l^{(-1)}\right), \quad 0 \le l, j \le N-2. \tag{3.7}$$

Then we have

$$\boldsymbol{S} - c^2 \boldsymbol{M} = \boldsymbol{\Sigma}, \quad \boldsymbol{M} = \boldsymbol{I}_{N-1} - \boldsymbol{B}\boldsymbol{T}\boldsymbol{T}^t\boldsymbol{B}^t, \tag{3.8}$$

where $\boldsymbol{\Sigma} = \text{diag}(\sigma_0, \ldots, \sigma_{N-2})$, the matrix $\boldsymbol{B}$ is defined in (2.22), and $\boldsymbol{T}$ is a tridiagonal matrix of order $M+1$ with zero main diagonal, and upper and lower diagonals $(a_0, \ldots, a_{M-1})$ and $(b_1, \ldots, b_M)$ in (2.14)–(2.15), respectively.

Proof It follows from (2.1), (2.4) and (2.9) straightforwardly that for any $c > 0$,

$$\mathcal{H}_c\left[\psi_n^{(-1)}\right](x) = \sigma_n\, \psi_n^{(1)}(x). \tag{3.9}$$

Since $\psi_n^{(-1)}(\pm 1) = 0$, we derive from integration by parts, (2.7), (2.9) and (3.9) that

$$\begin{aligned} \mathcal{A}_c\left(\psi_n^{(-1)}, \psi_m^{(-1)}\right) &= \left(\mathcal{H}_c\left[\psi_n^{(-1)}\right], \psi_m^{(-1)}\right) = \sigma_n \left(\psi_n^{(1)}, \psi_m^{(-1)}\right) \\ &= \sigma_n \left(\psi_n^{(1)}, \psi_m^{(1)}\right)_{\omega_1} = \sigma_n\, \delta_{mn}. \end{aligned} \tag{3.10}$$

We next show that $\sigma_n \neq 0$. We argue by contradiction. If $\sigma_n = 0$, we find from (3.9) that $\psi_n^{(-1)}(x) = A\sin(cx) + B\cos(cx)$ for some constants A, B. From $\psi_n^{(-1)}(\pm 1) = 0$, we

conclude that $A = B = 0$ and $\psi_n^{(-1)}(x) \equiv 0$, which contradicts to the fact that $\psi_n^{(-1)}(x)$ is an eigenfunction (cf. (2.4)). Therefore, (3.6) holds.

The first identity in (3.8) follows from (3.5) directly. By (2.9) and the orthogonality of generalized PSWFs $\{\psi_n^{(1)}\}$, we have

$$\begin{aligned} \boldsymbol{M}_{lj} &= \int_{-1}^{1} \psi_j^{(-1)}(x)\psi_l^{(-1)}(x)dx = \int_{-1}^{1} \psi_j^{(1)}(x)\psi_l^{(1)}(x)(1-x^2)^2 dx \\ &= \int_{-1}^{1} \psi_j^{(1)}(x)\psi_l^{(1)}(x)(1-x^2)dx - \int_{-1}^{1} x^2\psi_j^{(1)}(x)\psi_l^{(1)}(x)(1-x^2)dx \\ &= \delta_{lj} - \int_{-1}^{1} \left\{x\psi_j^{(1)}(x)\right\}\left\{x\psi_l^{(1)}(x)\right\}(1-x^2)dx. \end{aligned} \tag{3.11}$$

Using (2.14) and (2.21), we can compute $x\psi_n^{(1)}(x)$ via

$$x\boldsymbol{\psi}^{(1)}(x) = \boldsymbol{B}\boldsymbol{T}\boldsymbol{P}^{(1)}(x), \tag{3.12}$$

where $\boldsymbol{T}$ is a tridiagonal matrix of order $M+1$ associated with (2.14)–(2.15).

In view of (3.12), we can derive the second identity of (3.8) from (3.11) and the orthogonality of $\{P_n^{(1)}\}$ immediately. □

Thanks to Theorem 3.1, we obtain the "explicit" form of the numerical solution in (3.4).

Proposition 3.1 *Let u_N be the solution of* (3.4). *Then we have*

$$u_N(x) = \sum_{n=0}^{N-2} \frac{\breve{f}_n}{\sigma_n}\psi_n^{(-1)}(x), \quad x \in I, \qquad \sigma_n = \chi_n^{(1)}(c) + 2 - c^2, \tag{3.13}$$

with

$$\breve{f}_n = \sum_{l=0}^{L} \tilde{f}_l\, \beta_l^n, \quad \tilde{f}_l = \sum_{j=0}^{L} f(\xi_j) P_l^{(1)}(\xi_j)\omega_j, \tag{3.14}$$

where $\{\xi_j, \omega_j\}_{j=0}^{L}$ are the Gegenbauer–Gauss quadrature points and weights with respect to the weight function $\omega_1 = 1 - x^2$, and $\{\beta_l^n\}$ are the same as in (2.16).

Proof We have the expansion (cf. [15, Ch. 3]):

$$(\mathbb{I}_L^{(1)} f)(x) = \sum_{l=0}^{L} \tilde{f}_l P_l^{(1)}(x), \quad \tilde{f}_l = \sum_{j=0}^{L} f(\xi_j) P_l^{(1)}(\xi_j)\omega_j, \tag{3.15}$$

Writing the numerical solution as $u_N(x) = \sum_{n=0}^{N-2} \tilde{u}_n \psi_n^{(-1)}(x)$, and substituting it into (3.4), we derive from the orthogonality (3.5) and (2.16) immediately that

$$\begin{aligned} \tilde{u}_n &= \frac{1}{\sigma_n}(\mathbb{I}_L^{(1)} f, \psi_n^{(-1)}) = \frac{1}{\sigma_n}(\mathbb{I}_L^{(1)} f, \psi_n^{(1)})_{\omega_1} = \frac{1}{\sigma_n}\sum_{l=0}^{L} \tilde{f}_l\, (P_l^{(1)}, \psi_n^{(1)})_{\omega_1} \\ &= \frac{1}{\sigma_n}\sum_{l=0}^{L} \tilde{f}_l\, \beta_l^n := \frac{\breve{f}_n}{\sigma_n}, \quad 0 \le n \le N-2. \end{aligned} \tag{3.16}$$

This ends the derivation. □

3.2 Treatment of Nonhomogeneous Boundary Conditions

In order to approximate general functions in $H^1(I)$, we introduce two "bubble" functions (or boundary modes). It is evident that by (2.1),

$$\mathscr{D}_c^{(-1)}[\sin(cx)] = c^2\sin(cx), \quad \mathscr{D}_c^{(-1)}[\cos(cx)] = c^2\cos(cx). \tag{3.17}$$

Thus, $\sin(cx)$ and $\cos(cx)$ are eigenfunctions of $\mathscr{D}_c^{(-1)}$, which do not vanish at $x = \pm 1$, so they can supplement the generalized PSWFs $\{\psi_n^{(-1)}(x)\}$ to approximate general functions. For this purpose, we define

$$\psi_-^{(-1)}(x) := \frac{1}{2}\left(\frac{\cos(cx)}{\cos c} - \frac{\sin(cx)}{\sin c}\right), \quad \psi_+^{(-1)}(x) := \frac{1}{2}\left(\frac{\cos(cx)}{\cos c} + \frac{\sin(cx)}{\sin c}\right). \tag{3.18}$$

Then we have

$$\psi_-^{(-1)}(-1) = 1, \quad \psi_-^{(-1)}(1) = 0, \quad \psi_+^{(-1)}(-1) = 0, \quad \psi_+^{(-1)}(1) = 1, \tag{3.19}$$

and by (3.9) and (3.17),

$$\mathcal{H}_c\left[\psi_\pm^{(-1)}\right](x) = 0, \quad \mathscr{D}_c^{(-1)}\left[\psi_\pm^{(-1)}\right](x) = c^2\,\psi_\pm^{(-1)}(x). \tag{3.20}$$

Moreover, one verifies readily that

$$\begin{aligned} &\mathcal{A}_c\left(\psi_\pm^{(-1)}, \psi_\pm^{(-1)}\right) = c\cot(2c), \quad \mathcal{A}_c\left(\psi_\mp^{(-1)}, \psi_\pm^{(-1)}\right) = -c\csc(2c), \\ &\mathcal{A}_c\left(\psi_\pm^{(-1)}, \psi_n^{(-1)}\right) = 0, \quad \forall\, n \geq 0, \quad c > 0. \end{aligned} \tag{3.21}$$

We next consider the Helmholtz equation as a scattering problem with the exact Dirichlet-to-Neumann (DtN) boundary condition at $x = 1$ (cf. [4]):

$$\begin{cases} \mathcal{H}_c[u](x) = -u''(x) - c^2u(x) = f(x), & x \in I; \\ u(-1) = 0, \quad u'(1) - \mathrm{i}c\,u(1) = h. \end{cases} \tag{3.22}$$

Note that if $u(-1) = u_- \neq 0$, we can subtract $u_-\psi_-^{(-1)}$ from u which only affects the value of h. Define ${}_0\mathcal{V}_N := \{\psi_+^{(-1)}\} \cup \mathcal{V}_N^0$. Let u^R and u^I be the real and imaginary parts of u, respectively, and likewise for f^R, f^I, h^R, h^I etc.. The generalized PSWF-Galerkin scheme for (3.22) is to find $u_N = u_N^R + \mathrm{i}u_N^I$ with $u_N^R, u_N^I \in {}_0\mathcal{V}_N$ such that

$$\begin{cases} \mathcal{A}_c(u_N^R, v) + c\,u_N^I(1)v(1) = \left(\mathbb{I}_L^{(1)}f^R, v\right) + h^Rv(1), & \forall\, v \in {}_0\mathcal{V}_N, \\ \mathcal{A}_c(u_N^I, w) - c\,u_N^R(1)w(1) = \left(\mathbb{I}_L^{(1)}f^I, w\right) + h^Iw(1), & \forall\, w \in {}_0\mathcal{V}_N. \end{cases} \tag{3.23}$$

For notational convenience, we denote

$$\hat{f}_n^Z = \left(\mathbb{I}_L^{(1)}f^Z, \psi_n^{(-1)}\right), \quad \hat{f}_+^Z = \left(\mathbb{I}_L^{(1)}f^Z, \psi_+^{(-1)}\right), \quad Z = R, I. \tag{3.24}$$

Proposition 3.2 *The solution of* (3.23) *can be explicitly expressed as*

$$\left\{u_N^R(x), u_N^I(x)\right\} = \left\{u_N^R(1), u_N^I(1)\right\}\psi_+^{(-1)}(x) + \sum_{n=0}^{N-2}\left\{\frac{\hat{f}_n^R}{\sigma_n}, \frac{\hat{f}_n^I}{\sigma_n}\right\}\psi_n^{(-1)}(x), \tag{3.25}$$

where

$$\begin{aligned} u_N^R(1) &= \frac{\sin^2(2c)}{c}\left\{(\hat{f}_+^R + h^R)\cot(2c) - (\hat{f}_+^I + h^I)\right\}, \\ u_N^I(1) &= \frac{\sin^2(2c)}{c}\left\{(\hat{f}_+^I + h^I)\cot(2c) + (\hat{f}_+^R + h^R)\right\}. \end{aligned} \tag{3.26}$$

Proof Since $u_N^R, u_N^I \in {}_0\mathcal{V}_N$, we have

$$u_N^Z(x) = u_N^Z(1)\psi_+^{(-1)}(x) + \sum_{n=0}^{N-2} \tilde{u}_n^Z\, \psi_n^{(-1)}(x), \quad Z = R, I. \tag{3.27}$$

We next determine the expansion coefficients. Thanks to the fact $\psi_n^{(-1)}(\pm 1) = 0$, (3.5) and (3.21), we insert (3.27) into (3.23), and take $v, w = \psi_k^{(-1)}$, that yields

$$\tilde{u}_n^Z = \frac{\hat{f}_n^Z}{\sigma_n}, \quad Z = R, I, \quad 0 \le n \le N-2.$$

Taking $v, w = \psi_+^{(-1)}$ in (3.23) and using (3.21), we obtain the linear system:

$$\begin{cases} c\cot(2c)\, u_N^R(1) + c\, u_N^I(1) = \hat{f}_+^R + h^R, \\ -c\, u_N^R(1) + c\cot(2c)\, u_N^I(1) = \hat{f}_+^I + h^I, \end{cases} \tag{3.28}$$

whose solution is given by (3.26). Then we have the solution in (3.25). □

Remark 3.1 Note that $\hat{f}_n^R$ and $\hat{f}_n^I$ in (3.25) can be evaluated as in (3.14). This can avoid the use of numerical quadrature rules related to PSWFs, whose nodes and weights are complicated to evaluate. On the other hand, we see from (3.18) that $\hat{f}_+^R$, $\hat{f}_+^I$ involve highly oscillatory integrands when $c \gg 1$. In fact, they can be computed exactly by using an explicit formula (cf. [3]):

$$\int_{-1}^{1} P_n(x) e^{\mathrm{i}xy}\, dx = \mathrm{i}^n (2n+1)\sqrt{\frac{\pi}{2}}\frac{J_{n+1/2}(y)}{\sqrt{y}}, \quad y > 0, \tag{3.29}$$

where $J_{n+1/2}$ is the Bessel function, and P_n is the normalized Legendre polynomial of degree n. Recall the formula (cf. [15]):

$$P_l^{(1)}(x) = d_l P_{l+1}'(x), \quad d_l = \sqrt{\frac{2}{(l+1)(l+2)}}.$$

By (3.15),

$$\hat{f}_+^Z = \left(\mathbb{I}_L^{(1)} f^Z, \psi_+^{(-1)}\right) = \sum_{l=0}^{L} d_l \tilde{f}_l^Z \left(P_{l+1}', \psi_+^{(-1)}\right), \quad Z = R, I,$$

where by (3.18)–(3.19) and integration by parts, we derive from (3.29) that

$$\begin{aligned} \left(P_{l+1}', \psi_+^{(-1)}\right) &= 1 - \left(P_{l+1}, (\psi_+^{(-1)})'\right) = 1 - \frac{c}{2\sin c}(P_{l+1}, \cos cx) + \frac{c}{2\cos c}(P_{l+1}, \sin cx) \\ &= 1 + \left(l + \frac{3}{2}\right)\sqrt{\frac{\pi c}{2}} J_{l+3/2}(c) \times \begin{cases} \frac{(-1)^{l/2}}{\cos c}, & \text{if } l \text{ is even}, \\ \frac{(-1)^{(l-1)/2}}{\sin c}, & \text{if } l \text{ is odd}. \end{cases} \end{aligned}$$

With this, we can compute the highly oscillatory integrals accurately. □

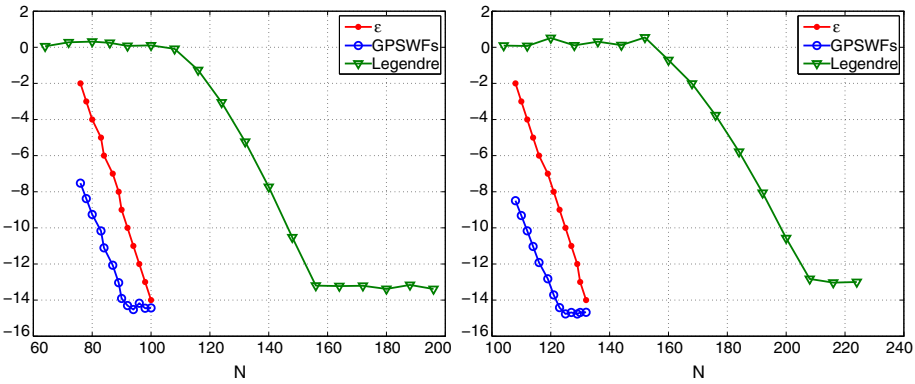

Fig. 1 Test for (3.30). *Left* $c = 35\pi$, *right* $c = 50\pi$

Remark 3.2 We point out that the Legendre–Galerkin method using the basis form by the integrated Legendre polynomials in (2.11), leads to sparse linear system, whose condition number behaves like $O(c^2)$, to solve (cf. [16]). However, the use of generalized PSWFs leads to a direct solution. Moreover, the new approximation enjoys spectral accuracy (see Sect. 4) and has a much faster convergence rate than the Legendre approximation (see Fig. 1).

We also remark that the use of the differentiation scheme in [13] results in dense matrix systems, and also additional efforts are needed to precompute the basis functions to incorporate boundary conditions. □

3.3 Numerical Results

In what follows, we present some numerical results obtained by the "explicit" formulas in Proposition 3.1 and Proposition 3.2. We also consider an example of the Helmholtz problem with variable coefficients. Here, we put the emphasis on the comparison with the spectral scheme using integrated Legendre polynomials.

We first consider (3.1) with $f(x) = \sin x$, which has the exact solution:

$$u(x) = \cos(cx) + \frac{\sin(cx) - c\sin(x)}{c^3 - c}, \quad c > 1. \tag{3.30}$$

Note that it does not meet the homogeneous Dirichlet boundary conditions, so we subtract $u_*(x) = u(-1)\psi_-^{(-1)}(x) + u(1)\psi_+^{(-1)}(x)$, from the solution and then derive from Proposition 3.1 the numerical solution:

$$u_N(x) = u_*(x) + \sum_{n=0}^{N-2} \frac{\breve{f}_n}{\sigma_n} \psi_n^{(-1)}(x). \tag{3.31}$$

In Fig. 1, we plot the logarithm of the maximum point-wise errors for the usual Legendre–Galerkin method, and the new generalized PSWF-Galerkin method for $c = 35\pi$ (left) and $c = 50\pi$ (right). Here, N is paired up with c by the rule in Sect. 2.3 (cf. (2.29)) with different $\varepsilon \in [10^{-14}, 10^{-2}]$, and the number of integrated Legendre basis functions is $N - 1$.

Some observations from Fig. 1 are in order.

(i) The new generalized PSWFs offer a much more accurate approximation than the Legendre polynomials. As expected, the error curve of Legendre approximation plunges into

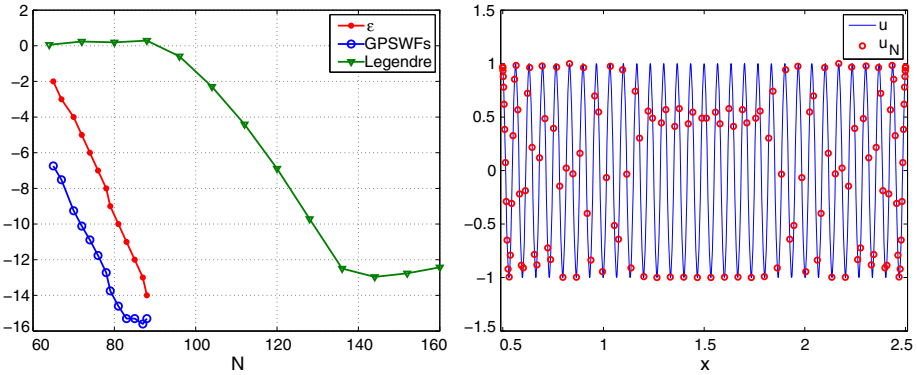

Fig. 2 Test for (3.32) with $c = 30\pi$. *Left* Logarithm of the maximum point-wise errors of Legendre and generalized PSWF-Galerkin methods. *Right* The real part of the numerical solution with $N = 70$ by generalized PSWF approximation against the exact solution

the exponential decaying region roughly when $N > c$. However, the generalized PSWFs decays exponentially even for much smaller N.

(ii) The errors of the generalized PSWF approximation is controlled by the error tolerance ε. We shall make a rigorous justification of this in the forthcoming section.

We next test an example related to (3.22):

$$-u'' - c^2 u = e^{-x^2/100} e^{25x\mathrm{i}}, \quad x \in (a, b); \quad u(a) = e^{-\mathrm{i}c}, \quad u'(b) - \mathrm{i}c\, u(b) = 0, \tag{3.32}$$

whose exact solution can be evaluated by the solution formula for a second-order ordinary differential equation. Then we can directly compute the numerical solution by using (3.25)–(3.26).

In Fig. 2 (left), we make a comparison of convergence behaviour as in Fig. 1, where $a = 0.5, b = 2.5, c = 30\pi$, and the optimal N corresponding to various $\varepsilon \in [10^{-14}, 10^{-2}]$ is identified by the rule in (2.29) as before. In Fig. 2 (right), we plot the real part of the numerical solution by generalized PSWFs with $N = 70$ (note: the maximum point-wise error is 4.91×10^{-10}) against the exact solution. Once again, we observe that one needs significantly smaller N to achieve a similar accuracy, and the error of the new scheme is actually controlled by ε. Thus, the generalized PSWFs enjoy a much higher resolution for highly oscillatory waves.

Finally, we consider the Bessel-type equation:

$$\begin{aligned} &-r^2 \frac{d^2 v}{dr^2} - r\frac{dv}{dr} + (n^2 - k^2 r^2) v = f(r), \quad r \in (a, b), \\ &v(a) = v_a, \quad v'(b) - \mathrm{i}k\, v(b) = h, \end{aligned} \tag{3.33}$$

where $a > 0$ and $n = 0, \pm 1, \ldots$. It arises from acoustic scattering problems with a cylindrical scatterer (cf. [16,21]). We make a change of variable to remove the first-order derivative (cf. [21]), that is,

$$r = a + \frac{x+1}{2}(b-a), \quad u(x) = \sqrt{r}\, v(r), \quad r \in (a, b), \quad x \in (-1, 1). \tag{3.34}$$

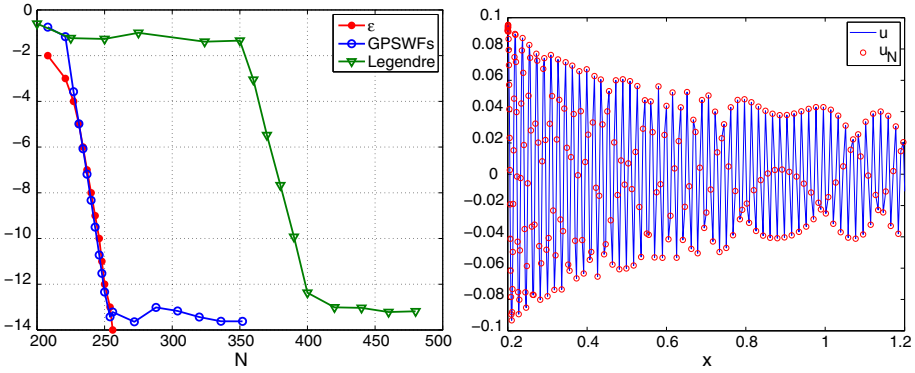

Fig. 3 Test for (3.33) with $c = 111\pi$. *Left* Logarithm of the maximum point-wise error of Legendre spectral and generalized PSWFs-Galerkin methods. *Right* The numerical solution u_N (with $N = 248$) of the generalized PSWF-Galerkin method against the exact solution in [0.2, 1.2]. Note that the maximum point-wise error is 2.99×10^{-12}

Then, we can convert (3.33) into

$$\begin{aligned} &-u''(x) - c^2 u(x) + s(x)u(x) = \tilde{f}, \quad x \in (-1, 1), \\ &u(-1) = u_-, \quad u'(1) - \eta\, u(1) = \tilde{h}, \end{aligned} \tag{3.35}$$

where $\tilde{f}(x) = f(r)$, and

$$c = \frac{b-a}{2}k, \quad s = \left(\frac{b-a}{2}\right)^2 \frac{4n^2-1}{4r^2}, \quad u_- = \sqrt{a}\, v_a, \quad \tilde{h} = \frac{b-a}{2}\sqrt{b}\, h,$$
$$\eta = \left(\mathrm{i}k + \frac{1}{2b}\right)\frac{b-a}{2}.$$

It is clear that we can subtract $u_- \psi_-^{(-1)}$ from the solution u and then the boundary condition at $x = -1$ becomes homogeneous. Assuming $u_- = 0$, we can obtain a Galerkin scheme similar to (3.23) but with the extra term from $s(x)u$, which can be accurately evaluated by a Jacobi–Gauss quadrature rule with the weight function $(1 - x^2)$. Thanks to Theorem 2.2, the matrix in the Galerkin scheme of the "leading" part: $-v'' - c^2 v$ is diagonal like the previous two cases. Thus, robust iterative solvers can be applied.

In the computation, we take the exact solution: $v(r) = J_n^{(1)}(cr)$ in (3.33), and set $n = 1, a = 0, b = 2.2$ and $c = 111\pi$. We report the results in Fig. 3 with a setting very similar to that in Fig. 2. Note that in the right figure, we only depict the solutions over the interval [0.2, 1.2] for better zooming in the oscillatory solutions. Indeed, we observe the same convergence behaviour as the previous two cases, even for variable coefficient problems and high wavenumbers.

3.4 Approximability to Spectrum of Laplacian

To have some insights into the approximability of the new basis (in comparison with polynomials again), we next study the generalized PSWF approximation of the Laplacian eigenvalue problems as in [13,25]. Consider

$$-\Delta u = \mu\, u \quad \text{in} \quad \Omega = (-1, 1)^d, \quad d = 1, 2; \quad u|_{\partial\Omega} = 0, \tag{3.36}$$

which has the eigen-pairs (μ_k, u_k) or (μ_{ij}, u_{ij}), respectively:

(i) for $d = 1$,

$$\mu_k = \frac{k^2\pi^2}{4}, \quad u_k(x) = \sin\frac{k\pi(x+1)}{2}, \quad k \geq 1; \tag{3.37}$$

(ii) for $d = 2$,

$$\mu_{ij} = \frac{(i^2+j^2)\pi^2}{4}, \quad u_{ij}(x,y) = \sin\frac{i\pi(x+1)}{2}\sin\frac{j\pi(y+1)}{2}, \quad i,j \geq 1. \tag{3.38}$$

The corresponding discrete eigen-problems are

(i) for $d = 1$,

$$\text{Find } (\tilde{\mu}, \hat{\boldsymbol{u}}) \text{ such that } \boldsymbol{S}\hat{\boldsymbol{u}} = \tilde{\mu}\boldsymbol{M}\hat{\boldsymbol{u}} \text{ or } \boldsymbol{\Sigma}\hat{\boldsymbol{u}} = (\tilde{\mu} - c^2)\boldsymbol{M}\hat{\boldsymbol{u}}, \tag{3.39}$$

where $\hat{\boldsymbol{u}} = (\hat{u}_0, \ldots, \hat{u}_{N-2})^t$ and the (discrete) eigenfunctions are computed by

$$u_N(x) = \sum_{n=0}^{N-2} \hat{u}_n \psi_n^{(-1)}(x). \tag{3.40}$$

(ii) for $d = 2$,

$$\begin{aligned} &\text{Find } (\tilde{\mu}, \widehat{\boldsymbol{U}}) \text{ such that } \boldsymbol{S}\widehat{\boldsymbol{U}}\boldsymbol{M} + \boldsymbol{M}\widehat{\boldsymbol{U}}\boldsymbol{S} = \tilde{\mu}\boldsymbol{M}\widehat{\boldsymbol{U}}\boldsymbol{M}, \\ &\qquad \text{or } \boldsymbol{\Sigma}\widehat{\boldsymbol{U}}\boldsymbol{M} + \boldsymbol{M}\widehat{\boldsymbol{U}}\boldsymbol{\Sigma} = (\tilde{\mu} - 2c^2)\boldsymbol{M}\widehat{\boldsymbol{U}}\boldsymbol{M}, \end{aligned} \tag{3.41}$$

where $\widehat{\boldsymbol{U}} = (\hat{u}_{ij})_{0\leq i,j\leq N-2}$ and the (discrete) eigenfunctions are computed by

$$u_N(x,y) = \sum_{i,j=0}^{N-2} \hat{u}_{ij} \psi_i^{(-1)}(x)\psi_j^{(-1)}(y). \tag{3.42}$$

We compare the new scheme with the Legendre–Galerkin method and examine the relative errors:

$$e_k = \frac{|\tilde{\mu}_k - \mu_k|}{|\mu_k|}, \quad 1 \leq k \leq (N-1)^d, \quad d = 1, 2,$$

where in the two-dimensional case, we arrange the eigenvalues in ascending order.

In Fig. 4 (left), we depict the relative errors between the discrete and continuous eigenvalues obtained by two methods for $d = 1$, where $c = 600\pi$ and the corresponding $N = 1307$ (obtained by the rule with $\varepsilon = 10^{-14}$ in Sect. 2.3). According to [25,27], there are about $2/\pi$ portion of "trusted" eigenvalues for the polynomial spectral method in 1D, where "trusted" means at least $O(N^{-1})$ accuracy. In a striking contrast, the generalized PSWF approximation leads to a portion of about 94 % "trusted" eigenvalues. In Table 1, we tabulate the percentages of "trusted" discrete eigenvalues obtained by two methods for many more N. Observe again that the generalized PSWF method leads to a significant higher portion of "trusted" eigenvalues. Based on the argument in [25], the generalized PSWFs have a better resolution of waves with fewer number of points per wavelength than polynomials.

In Fig. 4 (right), we plot $\{e_k\}$ against k for $d = 2$, where $(c, N) = (2980, 2048)$ with a total of about 10^6 discrete eigenvalues. As with [27], about 40.62 % discrete eigenvalues are "trustable" for the Legendre approximation. The portion increases to 89.99 % for the generalized PSWF approximation.

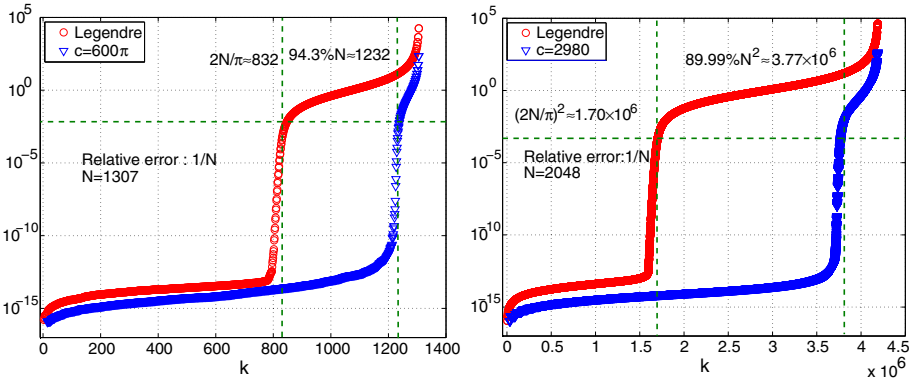

Fig. 4 Relative errors $\{e_k\}$ for $k = 1, 2, \ldots, (N-1)^d$ between the discrete and exact eigenvalues. *Left* $d = 1$ and $(c, N) = (600\pi, 1307)$. *Right* $d = 2$ and $(c, N) = (2980, 2048)$. Here, $\varepsilon = 10^{-14}$ in the rule of pairing up (c, N)

Table 1 Comparison of percentage of "trusted" eigenvalues

N	Generalized PSWFs		Legendre
	c	Percentage	Percentage ($\approx 2/\pi$)
67	20π	$52/67 \approx 77.6\,\%$	$42/67 \approx 62.7\,\%$
101	35π	$82/101 \approx 81.1\,\%$	$62/98 \approx 63.3\,\%$
454	200π	$417/454 \approx 91.9\,\%$	$289/454 \approx 63.6\,\%$
668	300π	$621/668 \approx 93.0\,\%$	$425/668 \approx 63.6\,\%$
882	400π	$825/882 \approx 93.5\,\%$	$561/882 \approx 63.6\,\%$
987	450π	$926/987 \approx 93.8\,\%$	$628/987 \approx 63.6\,\%$
1093	500π	$1028/1093 \approx 94.0\,\%$	$696/1093 \approx 63.6\,\%$
1307	600π	$1232/1307 \approx 94.3\,\%$	$832/1307 \approx 63.6\,\%$

Remark 3.3 The generalized PSWF-Galerkin approximation enjoys a performance very similar to the PSWF-based differentiation scheme in Kong and Rokhlin [13] for the one-dimensional eigenvalue problem. However, it is noteworthy that the modal PSWF basis in [13] was constructed by using more points than the number of modes, and a Gram–Schmidt orthogonalization was implemented to incorporate homogeneous boundary conditions. □

4 Error and Convergence Analysis

In this section, we conduct error estimates of approximation by generalized PSWFs which can provide theoretical justification of convergence behaviours observed in the previous section.

The following bound of $\{\beta_k^n\}$ (defined in (2.16)) plays an important role in the analysis. It is noteworthy that the argument of the analysis follows that of [12, Thm 1], but we correct the bound in [12, (52)] from $(2/q_n)^k$ to $(2/\sqrt{q_n})^k$, and improve the constant in the upper bound for $\alpha = 1$. In view of this, we therefore provide its proof in Appendix 1.

Lemma 4.1 *Denote*

$$q_n := q(n,c) = c^2/\chi_n^{(1)}(c), \quad n \geq 0, \quad c > 0. \tag{4.1}$$

If $q_n \leq 1$, then for all positive integer k such that $k(k+3) \leq \chi_n^{(1)}(c)$ and $n+k$ is even, we have

$$|\beta_0^n| \leq \sqrt{\frac{3}{2}}\,\lambda_n^{(1)}(c); \quad |\beta_k^n| \leq \Upsilon_k^{5/2,1}\,\frac{2}{\sqrt{\pi}}\left(\frac{2}{\sqrt{q_n}}\right)^k \lambda_n^{(1)}(c), \quad k \geq 1, \tag{4.2}$$

where $\Upsilon_k^{5/2,1}$ is defined in (4.4). Note that if $n+k$ is odd, then $\beta_k^n = 0$.

Remark 4.1 In the proofs, we need to use the property of the Gamma function ([1,28]): for any constants a, b, we have that for $n \geq 1$, $n+a > 1$ and $n+b > 1$,

$$\frac{\Gamma(n+a)}{\Gamma(n+b)} \leq \Upsilon_n^{a,b} n^{a-b}, \tag{4.3}$$

where

$$\Upsilon_n^{a,b} = \exp\left(\frac{a-b}{2(n+b-1)} + \frac{1}{12(n+a-1)} + \frac{(a-b)^2}{n}\right). \tag{4.4}$$

4.1 Main Result

We state the main result as follows.

Theorem 4.2 *Let u and u_N be the solutions of (3.1) and (3.4), respectively. For $c > 0$, let $q_{N-1} := c^2/\chi_{N-1}^{(1)}(c) \leq 1$, and $M(\leq N-1)$ be the largest integer such that*

$$M(M+3) + c^2 \leq \chi_{N-1}^{(1)}(c). \tag{4.5}$$

If $(1-x^2)^{l/2} f^{(l)} \in L^2_{\omega_1}(I)$ for $1 \leq l \leq r \in \mathbb{N}$, then we have that for $1 \leq r \leq M$,

$$\begin{aligned}\|u-u_N\|_{\omega_{-1}} \leq & C\left\{L^{-r}\left(\sum_{n=0}^{N-2}\frac{1}{|\sigma_n|^2}\right)^{1/2} + M^{-r}\left(\sum_{n=N-1}^{\infty}\frac{1}{|\sigma_n|^2}\right)^{1/2}\right\}\left\|(1-x^2)^{r/2} f^{(r)}\right\|_{\omega_1} \\ & + C\left\{\sum_{n=N-1}^{\infty}\left(\frac{\lambda_n^{(1)}(c)}{\sigma_n}\right)^2\left(\frac{2}{\sqrt{q_n}}\right)^{2M}\right\}^{1/2}\|f\|_{\omega_1},\end{aligned} \tag{4.6}$$

where $L+1$ is the number of quadrature nodes used in the interpolation of f (cf. (3.15)), $\sigma_n = \chi_n^{(-1)}(c) - c^2\,(\neq 0)$ is as defined in (3.9), and C is a generic positive constant independent of L, N, M, c, f and u.

Proof Like (3.13), we can write the solution u of (3.1) as

$$u(x) = \sum_{n=0}^{\infty}\frac{\hat{f}_n}{\sigma_n}\psi_n^{(-1)}(x), \quad \text{where } \hat{f}_n := (f, \psi_n^{(-1)}) = (f, \psi_n^{(1)})_{\omega_1}. \tag{4.7}$$

Then by (3.13) and the orthogonality (2.5), we have

$$\begin{aligned}\|u-u_N\|^2_{\omega_{-1}} &= \left\|\sum_{n=0}^{N-2}\frac{\hat{f}_n-\breve{f}_n}{\sigma_n}\psi_n^{(-1)}(x)+\sum_{n=N-1}^{\infty}\frac{\hat{f}_n}{\sigma_n}\psi_n^{(-1)}(x)\right\|^2_{\omega_{-1}}\\ &= \sum_{n=0}^{N-2}\left|\frac{\hat{f}_n-\breve{f}_n}{\sigma_n}\right|^2+\sum_{n=N-1}^{\infty}\left|\frac{\hat{f}_n}{\sigma_n}\right|^2 := \mathcal{I}_1+\mathcal{I}_2,\end{aligned} \tag{4.8}$$

where $\breve{f}_n := (\mathbb{I}_L^{(1)}f,\psi_n^{(1)})_{\omega_1}$ (cf. (3.16)). Next, we estimate the two terms $\mathcal{I}_1$ and $\mathcal{I}_2$ in (4.8), separately.

Firstly, using the Cauchy–Schwarz inequality, (2.7) and the fundamental approximation result on Jacobi–Gauss interpolation (cf. [15, Thm. 3.41]), we derive that for $1\le r\le N$,

$$\begin{aligned}\mathcal{I}_1 &= \sum_{n=0}^{N-2}\left|\frac{\hat{f}_n-\breve{f}_n}{\sigma_n}\right|^2=\sum_{n=0}^{N-2}\frac{\left|(f-\mathbb{I}_L^{(1)}f,\psi_n^{(1)})_{\omega_1}\right|^2}{|\sigma_n|^2}\le\sum_{n=0}^{N-2}\frac{\|f-\mathbb{I}_L^{(1)}f\|^2_{\omega_1}\|\psi_n^{(1)}\|^2_{\omega_1}}{|\sigma_n|^2}\\ &=\left(\sum_{n=0}^{N-2}\frac{1}{|\sigma_n|^2}\right)\|f-\mathbb{I}_L^{(1)}f\|^2_{\omega_1}\le CL^{-2r}\left(\sum_{n=0}^{N-2}\frac{1}{|\sigma_n|^2}\right)\|(1-x^2)^{r/2}f^{(r)}\|^2_{\omega_1}.\end{aligned} \tag{4.9}$$

Now, we estimate $|\hat{f}_n|$ (with $n\ge N-1$) involved in $\mathcal{I}_2$ in (4.8). Given M satisfying (4.5), let f_M be the truncated Gegenbauer series:

$$f_M(x) := (\pi_M^{(1)}f)(x)=\sum_{k=0}^{M}\hat{g}_kP_k^{(1)}(x),\quad\text{where}\quad\hat{g}_k=\int_{-1}^{1}f(x)P_k^{(1)}(x)\omega_1(x)\,dx. \tag{4.10}$$

Then we have

$$|\hat{f}_n|\le\left|\int_{-1}^{1}(f-f_M)\psi_n^{(1)}\omega_1dx\right|+\left|\int_{-1}^{1}f_M\psi_n^{(1)}\omega_1dx\right|:=\mathbb{J}_{n,M}^{(1)}+\mathbb{J}_{n,M}^{(2)},\quad n\ge N-1. \tag{4.11}$$

We obtain from the Cauchy–Schwartz inequality and Jacobi polynomial approximation result (cf. [15, Thm. 3.35]),

$$\mathbb{J}_{n,M}^{(1)}=\left|\int_{-1}^{1}(f-f_M)\psi_n^{(1)}\omega_1\,dx\right|\le\|f-\pi_M^{(1)}f\|_{\omega_1}\|\psi_n^{(1)}\|_{\omega_1}\le CM^{-r}\left\|(1-x^2)^{r/2}f^{(r)}\right\|_{\omega_1}. \tag{4.12}$$

Thus, it remains to estimate $\mathbb{J}_{n,M}^{(2)}$ in (4.11). By (4.10), (2.7) and (2.16),

$$\begin{aligned}\mathbb{J}_{n,M}^{(2)} &= \left|\int_{-1}^{1}f_M\psi_n^{(1)}\omega_1dx\right|=\left|\sum_{k=0}^{M}\hat{g}_k\int_{-1}^{1}P_k^{(1)}\psi_n^{(1)}\omega_1dx\right|\\ &\le\left(\sum_{k=0}^{M}(\hat{g}_k)^2\right)^{\frac12}\left(\sum_{k=0}^{M}\left(\int_{-1}^{1}P_k^{(1)}\psi_n^{(1)}\omega_1dx\right)^2\right)^{\frac12}\le\|f\|_{\omega_1}\left(\sum_{k=0}^{M}(\beta_k^n)^2\right)^{\frac12}.\end{aligned}$$

For fixed $c>0$, we know that $\chi_n^{(1)}(c)$ increases with respect to n, so we have from (4.5) that

$$M(M+3)\le\chi_{N-1}^{(1)}-c^2\le\chi_n^{(1)}(c)-c^2,\quad\forall n\ge N-1.$$

We first assume that both n and M are even. Then by (2.16), $\{\beta_{2k}^n\}_{k=0}^{M/2}$ (note that $M \le N-1 \le n$) are non-zero, so we have from Lemma 4.1 that for all $n \ge N-1$,

$$\sum_{k=0}^{M}(\beta_k^n)^2 = \sum_{k=0}^{M/2}(\beta_{2k}^n)^2 \le C\left\{\lambda_n^{(1)}(c)\right\}^2 \sum_{k=0}^{M/2}\left(\frac{16}{q_n^2}\right)^k \le C\left\{\lambda_n^{(1)}(c)\right\}^2 \frac{1-(16/q_n^2)^{M/2+1}}{1-(16/q_n^2)}$$
$$= C\left\{\lambda_n^{(1)}(c)\right\}^2 \left(\frac{4}{q_n}\right)^M \frac{(16/q_n^2)-(q_n^2/16)^{M/2}}{(16/q_n^2)-1} \le C\left\{\lambda_n^{(1)}(c)\right\}^2 \left(\frac{2}{\sqrt{q_n}}\right)^{2M}, \tag{4.13}$$

where we have used the condition $q_n \le 1$ to derive the last inequality. Similarly, we can show (4.13) is valid when both n and M are odd.

With the above bounds for (4.11), we can estimate $\mathcal{I}_2$ in (4.8) as follows:

$$\begin{aligned}\mathcal{I}_2 &= \sum_{n=N-1}^{\infty}\left|\frac{\hat{f}_n}{\sigma_n}\right|^2 \le 2\sum_{n=N-1}^{\infty}\left|\frac{\mathbb{J}_{n,M}^{(1)}}{\sigma_n}\right|^2 + 2\sum_{n=N-1}^{\infty}\left|\frac{\mathbb{J}_{n,M}^{(2)}}{\sigma_n}\right|^2 \\ &\le CM^{-2r}\left(\sum_{n=N-1}^{\infty}\frac{1}{\sigma_n^2}\right)\left\|(1-x^2)^{r/2}f^{(r)}\right\|_{\omega_1}^2 \\ &\quad + C\left\{\sum_{n=N-1}^{\infty}\left(\frac{\lambda_n^{(1)}(c)}{\sigma_n}\right)^2\left(\frac{2}{\sqrt{q_n}}\right)^{2M}\right\}\|f\|_{\omega_1}^2.\end{aligned} \tag{4.14}$$

Then the estimate (4.6) follows from (4.8), (4.9) and (4.14). □

4.2 Asymptotic Estimates

Observe from (4.6) that the upper bound depends on M, q_n, $\lambda_n^{(1)}$ etc.. In order to have more insights into the estimates, it is of practical interest to consider

$$c = \kappa n \quad \text{with} \quad 0 < \kappa < \frac{4}{e}\,(\approx 1.4715) < \frac{\pi}{2}, \tag{4.15}$$

within the "transition bandwidth" (cf. Remark 2.2), and $\lambda_n^{(1)}(c)$ begins to plunge into the range of exponential decay (cf. (2.31)). It is noteworthy that the constant $\frac{4}{e}$ is in fact the optimal constant ensuring the super-exponential decay rate of the $\lambda_n^{(1)}(c)$ (cf. [11, Corollary 5]).

Note that for fixed c and large n, we have the asymptotic estimate (cf. [23, (3.59)]):

$$\chi_n^{(1)}(c) = \widetilde{\chi}_n^{(1)}(c) + O\left(\frac{c^2}{n^3}\right), \quad \widetilde{\chi}_n^{(1)}(c) := n(n+3) + \frac{c^2}{2} + \frac{c^2(c^2+28)}{32n^2}. \tag{4.16}$$

In fact, under (4.15), $\widetilde{\chi}_n^{(1)}(c)$ offers a quite satisfactory approximation to $\chi_n^{(1)}(c)$ for large n. As some numerical illustrations, we tabulate in Table 2 the order in c:

$$\chi_n^{(1)}(c) = \widetilde{\chi}_n^{(1)}(c) + O(c^\tau), \tag{4.17}$$

and the quantity q_n in (4.1) for several n and various κ, satisfying (4.15). Observe that $\tau < 2$ and $q_n < 1$ in all cases.

For $0 < \kappa < \frac{4}{e}$, we obtain from (4.17) that

$$\chi_n^{(1)}(c) = W(\kappa)n^2 + O(n^\tau), \quad 1 \le \tau < 2; \quad W(t) = 1 + \frac{t^2}{2} + \frac{t^4}{32}, \tag{4.18}$$

Table 2 The order τ with different n and κ

n	κ	q_n	τ	n	κ	q_n	τ
64	1	0.63	0.38	1024	1	0.65	0.91
64	1.1	0.71	0.39	1024	1.1	0.73	1.03
64	1.2	0.79	0.31	1024	1.2	0.84	1.12
64	1.4	0.91	0.59	1024	1.4	0.93	1.29
512	1	0.65	0.76	2048	1	0.65	1.03
512	1.1	0.73	0.90	2048	1.1	0.73	1.12
512	1.2	0.80	1.01	2048	1.2	0.80	1.20
512	1.4	0.93	1.21	2048	1.4	0.92	1.36

for large n, so we have

$$\sigma_n = \chi_n^{(1)}(c) + 2 - c^2 = S(\kappa)n^2 + O(n^\tau); \quad S(t) := 1 - \frac{t^2}{2} + \frac{t^4}{32}. \tag{4.19}$$

Note that $S(t)$ is monotonically decreasing for $0 < t = \kappa < 4/e$, so we have

$$0.065 \approx S(4/e) < S(\kappa) < S(0) = 1, \quad \text{for } 0 < \kappa < \frac{4}{e}. \tag{4.20}$$

Moreover, by (2.31) and (4.15), we have that for $n \geq N - 1$,

$$\lambda_n^{(1)}(c) \approx \frac{e}{4}\sqrt{\frac{2\pi}{c}}\left(\frac{ce}{4n+6}\right)^{n+\frac{1}{2}} \approx \frac{e}{4}\sqrt{\frac{2\pi}{\kappa n}}\left(\frac{\kappa e}{4}\right)^{n+\frac{1}{2}} = \left(\frac{e}{4}\right)^{3/2}\sqrt{\frac{2\pi}{n}}\left(\frac{\kappa e}{4}\right)^n. \tag{4.21}$$

We are now ready to estimate the last term in (4.6). Letting $M \leq \delta(N-1)$ for some $\delta < 1$, we have that for all $n \geq N - 1$,

$$\begin{aligned}\left(\frac{\lambda_n^{(1)}(c)}{\sigma_n}\right)^2\left(\frac{2}{\sqrt{q_n}}\right)^{2M} &\leq \left(\frac{e}{4}\right)^3\frac{2\pi}{S^2(\kappa)n^5}\left(\frac{\kappa^2e^2}{16}\right)^n\left(\frac{4W(\kappa)}{\kappa^2}\right)^M \\ &\leq \left(\frac{e}{4}\right)^3\frac{2\pi}{S^2(\kappa)n^5}\left\{\frac{\kappa^2e^2}{16}\left(\frac{4W(\kappa)}{\kappa}\right)^\delta\right\}^n = \left(\frac{e}{4}\right)^3\frac{2\pi}{S^2(\kappa)n^5}\{G(\kappa,\delta)\}^n,\end{aligned} \tag{4.22}$$

where we denoted

$$G(\kappa,\delta) := \frac{\kappa^2e^2}{16}\left(\frac{4W(\kappa)}{\kappa^2}\right)^\delta = e^2 4^{\delta-2}\kappa^{2(1-\delta)}W^\delta(\kappa), \tag{4.23}$$

for $\kappa \in (0, 4/e)$ and $\delta \in (0, 1)$.

In Fig. 5, we depict $G(\kappa, \delta)$ for various κ and $\delta \in (0, 1)$, which is an increasing function of δ. We also observe that for fixed δ, $G(\kappa, \delta)$ is ascending with respect to κ. In Table 3, we list the values of δ_* solved from $G(\kappa, \delta_*) = 1$ for various $\kappa < 4/e (\approx 1.4715)$. Note that for $\kappa = 4/e$, $G(\kappa, \delta_*) = 1$ implies $\delta_* = 0$. Thus, for $0 < \kappa < 4/e$, we choose $\delta < \delta_* < 1$ (so $M < \delta(N-1)$), and by (4.23), there exists $0 < p < 1$, such that $G(k, \delta) \leq p^2$. Then by (4.22),

$$\left(\frac{\lambda_n^{(1)}(c)}{\sigma_n}\right)^2\left(\frac{2}{\sqrt{q_n}}\right)^{2M} \leq \frac{C}{n^5}p^{2n}, \quad n \geq N-1. \tag{4.24}$$

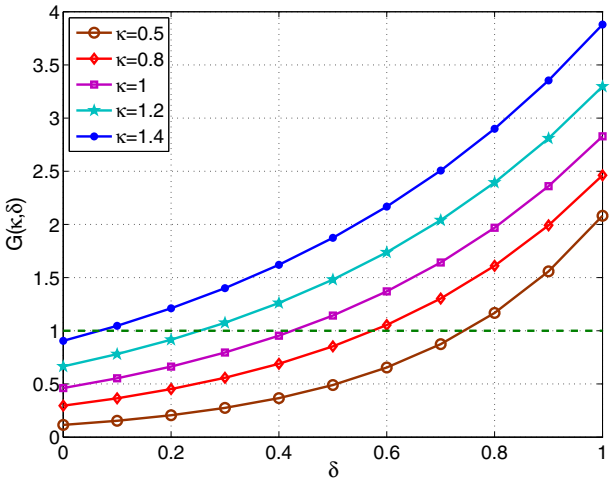

Fig. 5 Graphs of $G(\kappa, \delta)$ for various κ

Table 3 Various (κ, δ_*) such that $G(\kappa, \delta_*) = 1$

$G(\kappa, \delta_*)$	κ	δ_*
1	0.5	0.75
1	0.7	0.64
1	0.8	0.58
1	0.9	0.50
1	1	0.43
1	1.1	0.34
1	1.2	0.25
1	1.3	0.16
1	1.4	0.07

Using (4.19), we can estimate the bound of the following factor in (4.6) as

$$\sum_{n=N-1}^{\infty} \frac{1}{|\sigma_n|^2} \le C \sum_{n=N-1}^{\infty} \frac{1}{n^4} \le CN^{-3}. \tag{4.25}$$

In summary, with the above analysis, we can obtain from Theorem 4.2 the following more explicit estimate.

Corollary 4.3 *Let $c = \kappa N$ with $0 < \kappa < 4/e$. Then there exists a constant $0 < p < 1$, such that*

$$\|u - u_N\|_{\omega_{-1}} \le C \left\{ L^{-r} \left(\sum_{n=0}^{N-2} \frac{1}{|\sigma_n|^2} \right)^{1/2} + N^{-3/2-r} \right\} \times \left\| (1-x^2)^{r/2} f^{(r)} \right\|_{\omega_1} + CN^{-2}\, p^N \|f\|_{\omega_1},$$

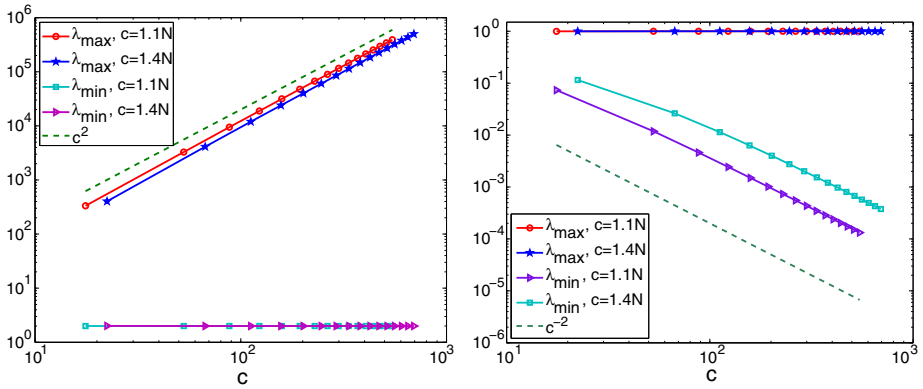

Fig. 6 The maximum and minimum eigenvalues of the stiffness matrix $\boldsymbol{S}$ and mass matrix $\boldsymbol{M}$ with $c = \kappa N$ for various κ

where $\sigma_n = \chi_n^{(-1)}(c) - c^2 (\neq 0)$ *is as defined in* (3.9)*, and* C *is a generic positive constant independent of* N, L, c, f *and* u.

4.3 Concluding Remarks

To conclude the paper, we reiterate some important contributions of the paper, and touch on the extension to generalized PSWFs of negative integer order.

It is seen that the use of generalized PSWFs of order -1 leads to optimal and spectrally accurate schemes for 1D Helmholtz problems with arbitrary high wave-numbers. Compared with the spectral differentiation approach based on PSWFs of order zero in [13], our approach can naturally build in boundary conditions, leads to a diagonal linear system for the Helmholtz operator, and enjoys much higher percentage of "trusted" discrete eigenvalues than the polynomial approach (so the approximability of the new basis as good as that in [13]). The stiffness and mass matrices have the attractive relation (3.8), so the algorithm can be simplified with a much better conditioning. In fact, the conditioning of both the stiffness matrix and mass matrix is $O(c^2)$. As an illustration, we plot in Fig. 6 (left) the smallest and largest eigenvalues of stiffness matrix $\boldsymbol{S}$ in (3.8) for various $c = \kappa N$. Observe that for each κ, the largest eigenvalue grows like c^2, while the smallest one remains a constant. In view of the relation (3.8), the largest eigenvalue of the mass matrix $\boldsymbol{M}$ should remain a constant, while the smallest eigenvalue decays at a rate of c^{-2}. Indeed, we can observe this from Fig. 6(right). As a result, the condition numbers of the stiffness matrix $\boldsymbol{S}$ and mass matrix $\boldsymbol{M}$ behave like $O(c^2)$ for $c = \kappa N$.

In fact, we can define the generalized PSWFs of negative integer order $-k$ with $k \geq 1$, denoted by $\{\psi_n^{(-k)}(x; c)\}$, as the eigenfunctions of the operator in (2.6) (but with $\alpha = -k$):

$$\mathscr{D}_c^{(-k)}[u](x) = \chi\, u(x), \quad x \in (-1, 1); \quad u^{(l)}(\pm 1) = 0, \quad l = 0, 1, \ldots, k-1.$$

Then we can employ the basis for the $2k$-th order BVPs. In particular, when $c = 0$, it reduces to the optimal spectral algorithms based on generalized Jacobi polynomials in [9]. Moreover, we can extend the idea to introduce (anisotropic) generalized PSWFs $\psi_n^{(-k,-l)}(x; c)$ from the Jacobi polynomials with negative integer parameters for odd-order BVPs as in [9]. We leave these extensions to the interested readers.

Appendix 1: Proof of Lemma 4.1

We first recall the following bound in [12, Appendix].

Proposition 4.1 *Let q_n be defined as in (4.1). If $q_n \le 1$, then for all positive integer k such that $k(k+3) \le \chi_n^{(1)}(c)$ and $n+k$ is even, we have*

$$|\partial_x^k \psi_n^{(1)}(0)| \le \sqrt{2}\left(\chi_n^{(1)}(c)\right)^{k/2}. \tag{4.26}$$

Note that if $n+k$ is odd, then $\partial_x^k \psi_n^{(1)}(0) = 0$.

Proposition 4.2 *Let $\{\beta_k^n\}$ be defined in (2.16). For given $c > 0$ and $n \in \mathbb{N}$, let*

$$k(k+3) + c^2 \le \chi_n^{(1)}(c). \tag{4.27}$$

We have

(i) *If both n and k are even, then $\beta_0^n, \beta_2^n, \beta_4^n, \ldots$ have the same sign, and $|\beta_k^n| \le |\beta_{k+2}^n|$ for all $k = 0, 2, 4, \ldots$, satisfying (4.27);*
(ii) *If both n and k are odd, then $\beta_1^n, \beta_3^n, \beta_5^n, \ldots$ have the same sign, and $|\beta_k^n| \le |\beta_{k+2}^n|$ for all $k = 1, 3, 5, \ldots$, satisfying (4.27).*

Note that if $n+k$ is odd, then $\beta_k^n = 0$.

Proof From (2.16) and the parity of $\psi_n^{(1)}(x)$ and $P_k^{(1)}(x)$, we have $\beta_k^n = 0$, if $n+k$ is odd.

We first justify the statement (i). By (2.18),

$$\beta_{k+2}^n = \frac{1}{F(k+2)c^2}\left\{\left(\chi_n^{(1)}(c) - k(k+3) - G(k)c^2\right)\beta_k^n - F(k)c^2\beta_{k-2}^n\right\}, \quad k \ge 0, \tag{4.28}$$

with $\beta_{-2}^n = \beta_{-1}^n = 0$, where

$$F(k) = \sqrt{\frac{(k-1)k(k+1)(k+2)}{(2k-1)(2k+1)^2(2k+3)}}, \quad G(k) = \frac{2k(k+3)+1}{(2k+1)(2k+5)}. \tag{4.29}$$

One verifies that for $k \ge 2$,

$$\frac{1}{5} < \frac{(k-1)(k+2)}{(2k+1)^2} < F(k) = \sqrt{\frac{k(k+1)}{(2k+1)^2}}\sqrt{\frac{(k-1)(k+2)}{(2k-1)(2k+3)}} < \sqrt{\frac{1}{4}}\sqrt{\frac{1}{4}} = \frac{1}{4}, \tag{4.30}$$

and for $k \ge 1$,

$$\frac{2}{5} \le \frac{k}{2k+1} < \frac{2k(k+3)}{(2k+1)(2k+5)} < G(k) < \frac{(2k+1)(k+\frac{5}{2})}{(2k+1)(2k+5)} = \frac{1}{2}. \tag{4.31}$$

We proceed with the proof by induction. For $k = 0$, we find from (4.28) that

$$\beta_2^n = \frac{5}{2c^2}\sqrt{\frac{7}{2}}\left(\chi_n^{(1)} - \frac{c^2}{5}\right)\beta_0^n. \tag{4.32}$$

It is evident that with a looser condition than (4.27), the factor in front of β_0^n is positive, so β_2^n has the same sign as β_0^n. Moreover, by (4.27) with $k = 0$,

$$|\beta_2^n| = \frac{5}{2c^2}\sqrt{\frac{7}{2}}\left(\chi_n^{(1)} - \frac{c^2}{5}\right)|\beta_0^n| \ge 2\sqrt{\frac{7}{2}}|\beta_0^n| \ge |\beta_0^n|, \tag{4.33}$$

We next assume that $\beta_{k-2}^n \beta_k^n > 0$ and $|\beta_{k-2}^n| \le |\beta_k^n|$ for all $k \ge 2$. Then we derive from (4.28), (4.30)–(4.31) and (4.27) that

$$\begin{aligned}\beta_k^n \beta_{k+2}^n &= \frac{1}{F(k+2)c^2}\left\{\left(\chi_n^{(1)}(c) - k(k+3) - G(k)c^2\right)(\beta_k^n)^2 - F(k)c^2\beta_{k-2}^n\beta_k^n\right\}\\ &\ge \frac{1}{F(k+2)c^2}\left\{\chi_n^{(1)}(c) - k(k+3) - G(k)c^2 - F(k)c^2\right\}(\beta_k^n)^2\\ &> \frac{1}{F(k+2)c^2}\left\{\chi_n^{(1)}(c) - k(k+3) - \frac{3}{4}c^2\right\}(\beta_k^n)^2 > 0,\end{aligned} \tag{4.34}$$

which implies $\beta_{k+2}^n \beta_k^n > 0$. Now, we show that $|\beta_k^n| \le |\beta_{k+2}^n|$. We show this by contradiction. Assuming that $|\beta_k^n| > |\beta_{k+2}^n|$, we find from (4.34) and (4.30) that

$$\begin{aligned}0 &> \frac{1}{F(k+2)c^2}\left\{\chi_n^{(1)}(c) - k(k+3) - \frac{3}{4}c^2\right\}|\beta_k^n| - |\beta_{k+2}^n|\\ &> \frac{1}{F(k+2)c^2}\left\{\chi_n^{(1)} - k(k+3) - \frac{3c^2}{4} - F(k+2)c^2\right\}|\beta_k^n|\\ &> \frac{1}{F(k+2)c^2}\left\{\chi_n^{(1)} - k(k+3) - c^2\right\}|\beta_k^n|,\end{aligned} \tag{4.35}$$

which contradicts to (4.27). Thus, we have $|\beta_k^n| \le |\beta_{k+2}^n|$ for even n and k.

The statement (ii) can be justified similarly. In fact, the derivations in (4.34)–(4.35) also hold for odd n, k, so it suffices to verify the initial of the induction. Like (4.32)–(4.33), we can show that

$$\beta_3^n = \frac{7}{2c^2}\sqrt{\frac{3}{2}}\left(\chi_n^{(1)} - 4 - \frac{3}{7}c^2\right)\beta_1^n,$$

and

$$|\beta_3^n| = \frac{7}{2c^2}\sqrt{\frac{3}{2}}\left(\chi_n^{(1)} - 4 - \frac{3}{7}c^2\right)|\beta_1^n| \ge \sqrt{6}|\beta_1^n| \ge |\beta_1^n|,$$

which implies β_3^n has the same sign as β_1^n, and $|\beta_1^n| \le |\beta_3^n|$. Then, we use (4.34)–(4.35) to complete the proof of Proposition A.2. □

Proof of Lemma 4.1 With the above propositions, we are now ready to prove Lemma 4.1. □

The bound for β_0^n follows directly from (2.23) and (4.26) with $k = 0$.

We carry out the proof by estimating the moment: $\int_{-1}^1 t^m \psi_n^{(1)}(t)\omega_1(t)dt$. Note that

$$t^j = \sum_{k=0}^{j} \hat{p}_{jk} P_k^{(1)}(t), \quad \text{where} \quad \hat{p}_{jk} = \int_{-1}^{1} t^j P_k^{(1)}(t)(1-t^2)dt, \tag{4.36}$$

where we can find the formula of $\hat{p}_{jk}$ from [12, (16)-(17)], and have

$$\hat{p}_{jk} = 0, \quad \text{if } k+j \text{ is odd}; \quad \hat{p}_{jk} > 0, \quad \text{if } k+j \text{ is even}; \quad \hat{p}_{jj} = \frac{\sqrt{2\pi(2j+3)j!(j+2)!}}{2^{j+2}\Gamma(j+5/2)}. \tag{4.37}$$

Thus, we obtain from (2.16) that

$$\int_{-1}^{1} t^j \psi_n^{(1)}(t)\omega_1(t)dt = \sum_{k=0}^{j} \hat{p}_{jk}\beta_k^n = \begin{cases} 0, & \text{if } n+j \text{ is odd,} \\ \sum_{k=0}^{j/2} \hat{p}_{j,2k}\,\beta_{2k}^n, & \text{if } n,\, j \text{ are even,} \\ \sum_{k=0}^{(j-1)/2} \hat{p}_{j,2k+1}\beta_{2k+1}^n, & \text{if } n,\, j \text{ are odd,} \end{cases} \tag{4.38}$$

where we used the property: $\beta_k^n = 0$, if $k+n$ is odd.

On the other hand, taking the jth derivative at $x = 0$ on both sides of (2.8) with $\alpha = 1$, yields

$$\int_{-1}^{1} t^j\, \psi_n^{(1)}(t)(1-t^2)\, dt = (-1)^j\, \mathrm{i}^{n+j} c^{-j}\, \lambda_n^{(1)}(c)\, \partial_x^j \psi_n^{(1)}(0), \tag{4.39}$$

which vanishes, if $j+n$ is odd.

Thus, we have from Lemma A.2 (i), and (4.37)–(4.39) that for even n, j,

$$\left| \sum_{k=0}^{j/2} \hat{p}_{j,2k}\, \beta_{2k}^n \right| = \sum_{k=0}^{j/2} \hat{p}_{j,2k}\, \left|\beta_{2k}^n\right| = c^{-j} \lambda_n^{(1)}(c)\, \left|\partial_x^j \psi_n^{(1)}(0)\right|, \tag{4.40}$$

and further by (4.26),

$$\hat{p}_{jj}|\beta_j^n| \le \sum_{k=0}^{j/2} \hat{p}_{j,2k}\, \left|\beta_{2k}^n\right| = c^{-j}\lambda_n^{(1)}(c)|\partial_x^j \psi_n^{(1)}(0)| \le \sqrt{2}\left(\frac{1}{\sqrt{q_n}}\right)^j \lambda_n^{(1)}(c). \tag{4.41}$$

Similarly, for odd n, j,

$$\hat{p}_{jj}|\beta_j^n| \le \sum_{k=0}^{(j-1)/2} \hat{p}_{j,2k+1}\, \left|\beta_{2k+1}^n\right| = c^{-j}\lambda_n^{(1)}(c)|\partial_x^j \psi_n^{(1)}(0)| \le \sqrt{2}\left(\frac{1}{\sqrt{q_n}}\right)^j \lambda_n^{(1)}(c). \tag{4.42}$$

Thus, by (4.37), and (4.3)–(4.4),

$$\hat{p}_{jj}^{-1} = \frac{2^{j+2}\sqrt{j}}{\sqrt{2\pi(2j+3)}} \sqrt{\Upsilon_j^{5/2,1}\Upsilon_j^{5/2,3}} \le \Upsilon_j^{5/2,1}\, \frac{2^{j+1}}{\sqrt{\pi}}. \tag{4.43}$$

Then we obtain the bound (4.2).

References

1. Abramowitz, M., Stegun, I.A.: Handbook of Mathematical Functions. Dover, New York (1972)
2. Al-Gwaiz, M.A.: Sturm–Liouville Theory and Its Applications. Springer, Berlin (2008)
3. Arfken, G.B., Weber, H.J.: Mathematical Methods for Physicists. Harcourt/Academic Press, San Diego (2001)
4. Babuška, I., Sauter, S.A.: Is the pollution effect of the FEM avoidable for the Helmholtze equation considering high wave number? SIAM Rev. **34**(6), 2392–2423 (1997)
5. Bouwkamp, C.J.: On the theory of spheroidal wave functions of order zero. Ned. Akad. Wetensch. Proc. **53**, 931–944 (1950)
6. Boyd, J.P.: Large mode number eigenvalues of the prolate spheroidal differential equation. Appl. Math. Comput. **145**(2), 881–886 (2003)
7. Boyd, J.P., Gassner, G., Sadiq, B.A.: The nonconvergence of h-refinement in prolate elements. J. Sci. Comput. **57**(2), 372–389 (2013)

8. Coddington, E.A., Levinson, N.: Theory of Ordinary Differential Equations. McGraw-Hill, New York (1955)
9. Guo, B.Y., Shen, J., Wang, L.L.: Optimal spectral-Galerkin methods using generalized Jacobi polynomials. J. Sci. Comput. **27**(1), 305–322 (2006)
10. Hogan, J.A., Lakey, J.D.: Duration and Bandwidth Limiting: Prolate Functions, Sampling, and Applications. Birkhäuser, New York (2012)
11. Bonami, A., Karoui, A.: Spectral decay of time and frequency limiting operator. Appl. Comput. Harmon. Anal. (2015). doi:10.1016/j.acha.2015.05.003.2015
12. Karoui, A., Souabni, A.: Generalized prolate spheroidal wave functions: spectral analysis and approximation of almost band-limited functions. J. Fourier Anal. Appl. **22**(2), 383–412 (2016)
13. Kong, W.Y., Rokhlin, V.: A new class of highly accurate differentiation schemes based on the prolate spheroidal wave functions. Appl. Comput. Harmon. Anal. **33**(2), 226–260 (2012)
14. Osipov, A., Rokhlin, V., Xiao, H.: Prolate Spheroidal Wave Functions of Order Zero. Springer, Berlin (2013)
15. Shen, J., Tang, T., Wang, L.L.: Spectral Methods: Algorithms, Analysis and Applications. Springer, Berlin (2011)
16. Shen, J., Wang, L.L.: Spectral approximation of the Helmholtz equation with high wave numbers. SIAM J. Numer. Anal. **43**(2), 623–644 (2005)
17. Slepian, D.: Prolate spheroidal wave functions, Fourier analysis and uncertainity. IV. Bell Syst. Tech. J. **43**(6), 3009–3057 (1964)
18. Slepian, D.: Some comments on Fourier analysis, uncertainty and modeling. SIAM Rev. **25**(3), 379–393 (1983)
19. Slepian, D., Pollak, H.O.: Prolate spheroidal wave functions, Fourier analysis and uncertainty. I. Bell Syst. Tech. J. **40**, 43–63 (1961)
20. Szegö, G.: Orthogonal Polynomials, 4th edn. AMS Colloquium Publications, Providence (1975)
21. Wang, K., Wong, Y.S., Deng, J.: Efficient and accurate numerical solutions for Helmholtz equation in polar and spherical coordinates. Commun. Comput. Phys. **17**(3), 779–807 (2015)
22. Wang, L.L., Samson, M.D., Zhao, X.D.: A well-conditioned collocation method using a pseudospectral integration matrix. SIAM J. Sci. Comput. **36**(3), 907–929 (2014)
23. Wang, L.L., Zhang, J.: A new generalization of the PSWFs with applications to spectral approximations on quasi-uniform grids. Appl. Comput. Harmon. Anal. **29**(3), 525–545 (2010)
24. Wang, L.L., Zhang, J., Zhang, Z.: On hp-convergence of prolate spheroidal wave functions and a new well-conditioned prolate-collocation scheme. J. Comput. Phys. **268**(2), 377–398 (2014)
25. Weideman, J.A.C., Trefethen, L.N.: The eigenvalues of second-order spectral differentiation matrices. SIAM J. Numer. Anal. **25**(6), 1279–1298 (1988)
26. Xiao, H., Rokhlin, V., Yarvin, N.: Prolate spheroidal wavefunctions, quadrature and interpolation. Inverse Probl. **17**(4), 805–838 (2001)
27. Zhang, Z.: How many numerical eigenvalues can we trust? J. Sci. Comput. **65**(2), 1–12 (2015)
28. Zhao, X.D., Wang, L.L., Xie, Z.Q.: Sharp error bounds for Jacobi expansions and Gegenbauer–Gauss quadrature of analytic functions. SIAM J. Numer. Anal. **51**(3), 1443–1469 (2013)