Numerical Algorithms manuscript No.
(will be inserted by the editor)

A numerical study of the Legendre-Galerkin method for the evaluation of the prolate spheroidal wave functions

Sebastian Schmutzhard · Tomasz Hrycak · Hans G. Feichtinger

Abstract We conduct a numerical study of the Legendre-Galerkin method for the evaluation of the prolate spheroidal wave functions (PSWFs) viewed as the eigenfunctions of the prolate differential operator with boundary conditions of continuity. Our experiments indicate that the minimal dimension N of the Legendre-Galerkin matrix for the evaluation of the nth prolate with precision ε is $\mathcal{O}\left(n+\sqrt{nc}\right)$, as $n, c \to \infty$, where $c > 0$ is the bandwidth parameter. The behavior of N, when either c or n is held constant, is also examined. As a consequence, we obtain an upper bound on the complexity of the evaluation of the prolates. We also study the condition number of the approximate Legendre coefficients, computed as an eigenvector of the Legendre-Galerkin matrix. We observe experimentally that for fixed precision ε, an error estimate based on this condition number is $\mathcal{O}\left(n+c\right)$ as $n, c \to \infty$. We conclude that the Legendre-Galerkin method is accurate for fairly large values of n and c.

Keywords prolate spheroidal wave functions · Legendre-Galerkin method · discretization · condition number · accuracy · complexity

Mathematics Subject Classification (2000) 33E10 · 65L15 · 65L60

1 Introduction

1.1 Motivation

The prolate spheroidal wave functions (PSWFs) have been long studied e.g. in relation to the Helmholtz equation in space, or as solutions to the energy

S. Schmutzhard · H.G. Feichtinger
Department of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Wien, Austria, e-mail: sebastian.schmutzhard@univie.ac.at, hans.feichtinger@univie.ac.at

T. Hrycak
Acoustics Research Institute, Wohllebengasse 12-14, 1040 Wien, Austria

concentration problem for bandlimited functions. More recently, they have been used as spectral elements for the discretization for time-dependent partial differential equations [8], and in modeling of wireless channels [14].

In this paper, we are concerned with the evaluation of the prolate spheroidal wave functions of zero order (PSWFs) on the interval $[-1, 1]$. The PSWFs with the bandwidth parameter $c > 0$ are the eigenfunctions of the prolate differential operator L_c defined by the formula

$$L_c f(x) = -\left(\left(1-x^2\right) f'(x)\right)' + c^2 x^2 f(x), \tag{1}$$

that are, moreover, continuous at the endpoints $x = \pm 1$. The most common way to discretize the eigenproblem

$$L_c f = \chi f \tag{2}$$

with boundary conditions of continuity is to use the Galerkin method with the Legendre polynomials [5, 8–10, 13]. This approach amounts to computing approximate Legendre coefficients of the PSWFs, followed by the evaluation of the resulting Legendre series. Bouwkamp [5] obtains the Legendre coefficients as eigenvectors of a real tridiagonal, but non-symmetric matrix. Hodge [13] uses a real symmetric tridiagonal matrix instead, which is diagonally conjugated to Bouwkamp's.

One of the main advantages of this method is its spectral convergence. Since every prolate is the restriction of an entire function, the Legendre coefficients of the prolate decay super-exponentially (i.e. faster than every exponential function), see [7, Theorem 7.3].

Even though there are several publications on the subject, some important aspects of the Legendre-Galerkin method are not well understood.

First, it has not been satisfactorily explained, how the dimension of the discretized eigenproblem depends on the parameter c, the mode number n of the prolate, and the required precision ε.

Second, it is not clear what precision can be obtained, especially when the mode number n or the bandwidth parameter c is large. Conditioning of the eigenvectors depends on the separation of the eigenvalues of the Galerkin matrix, and little is known on this subject.

Both questions are difficult to analyze with current methods. Therefore, we have decided to conduct numerical experiments, and then formulate conjectures, which may lead to rigorous proofs.

We note that there exists an analytically equivalent definition of the PSWFs as the eigenfunctions of the prolate integral operator P_c defined on $L^2(-1, 1)$ as follows

$$P_c f(x) = \frac{1}{\pi} \int_{-1}^{1} \frac{\sin\ c(x-y)}{x-y} f(y) dy. \tag{3}$$

This description is commonly used to study analytic properties of the PSWFs [15, 16, 23, 24]. However, this definition is unsuitable for the numerical evaluation for the following reasons:

1. The eigenvalues of the integral equation decay super-exponentially fast, [26, Corollary 2]. Therefore, the eigenvectors corresponding to smaller eigenvalues are inaccurate.
2. For large parameters c, the initial eigenvalues are close to each other [11, Theorem 1] and, consequently, the initial eigenfunctions are ill-conditioned.

Therefore, we consider the Legendre-Galerkin method to be the best choice for the numerical evaluation of the prolates. We note that an old approach presented e.g. in [5] used Newton's method to improve accuracy of the eigenvalues given their initial approximations. The current, robust approach uses a modern eigensolver for real symmetric tridiagonal matrices.

1.2 Previous work

Even though the Legendre-Galerkin algorithm is the method of choice for the numerical evaluation of the prolates, little is known about its overall computational complexity or accuracy, which depend on the mode number n and the bandwidth c.

The question of the dimension of the Galerkin matrix has been studied by J. P. Boyd in a special case when $c \leqslant c_*(n)$, where $c_*(n) = \frac{\pi}{2}(n + \frac{1}{2})$ is the *transition bandwidth*. Boyd has verified experimentally that the dimension $N_{trunc} = 2n + 30$ [8, equation (30)] is sufficient to guarantee double precision. We study this case in Sec. 3.4, and obtain results consistent with Boyd's. Boyd explains in [8, Section 2.3] that the assumption $c \leqslant c_*(n)$ prevents the first n prolates from having uniformly small values at the endpoints $x = \pm 1$, and is necessary when the prolates are used as a basis for a numerically stable expansion.

Some works like [17,18] deal exclusively with the *continuous* eigenproblem, e.g. the eigenvalues of the prolate differential operator with certain boundary conditions. By contrast, in this paper we study a *discrete* eigenproblem, e.g. the eigenvalues of finite Galerkin matrices. Similarly, [20] also deals mostly with the continuous eigenproblem, with some exceptions, e.g. formula (10.3) on p. 354. This formula sets the dimension of Galerkin matrix to $N = 1.1c + n + 1000$, but without any justification. In [19, Remark 9], the dimension is assumed to be $\mathcal{O}(n + c)$, again without any justification.

In [27], Xiao et al. use a truncation of the Galerkin matrices that guarantees only that the omitted Legendre coefficients are smaller in magnitude than a fixed $\varepsilon > 0$. This does not guarantee that the prolate functions are computed with this precision. In fact, Xiao et al. make no statement about accuracy of the prolates computed with their approach.

By contrast, we study the smallest dimension $N = N(n, c, \varepsilon)$ of the Legendre-Galerkin matrix required to compute the nth prolate with precision ε. We determine accuracy of our approach by comparing the nth eigenvector of the Galerkin matrix with the exact Legendre coefficients of the nth prolate, see (33).

1.3 Contributions

We conduct a numerical study of the Legendre-Galerkin method applied to the eigenproblem (2) with boundary conditions of continuity addressing the following aspects. Throughout this paper, we use the big O notation to describe the limiting behavior as $n, c \to \infty$ and $\varepsilon \to 0^+$.

Our main contributions are as follows.

1. *Discretization.* We observe experimentally that for a fixed $\varepsilon > 0$, the minimal dimension N of the Legendre-Galerkin matrix required to achieve precision ε satisfies

$$N = \mathcal{O}\left(n + \sqrt{nc}\right), \tag{4}$$

   where c is the bandwidth parameter, and n is the mode number of the prolate, see Section 3.1.
2. *Discretization.* In Section 3.4, we observe that if $c = c_*(n)$, then

$$N = \mathcal{O}\left(n + \sqrt{n}\,|\log\varepsilon|\right). \tag{5}$$

3. *Conditioning.* We study the condition number of the approximate Legendre coefficients viewed as an eigenvector of the Legendre-Galerkin matrix. In Section 3.5, we observe experimentally that for fixed precision ε, an error estimate based on this condition number is $\mathcal{O}(n + c)$.

Two additional bounds on N are presented in Section 3.2 and Section 3.3. In Section 3.7, we obtain from (4) the following upper bound on the operation count L required for the evaluation of the nth prolate at M points in fixed precision

$$L = \mathcal{O}\left((n + \sqrt{nc})(\log^2 n + \log^2 c) + M\right), \tag{6}$$

which is the first estimate for the overall complexity of the Legendre-Galerkin algorithm in terms of the parameters n and c.

Using computations in extended precision, we confirm experimentally that the error of the PSWFs evaluated by the Legendre-Galerkin method grows slowly with n and c, and is within an order of magnitude from our error estimates based on the condition number of the eigenvectors of the corresponding Galerkin matrix, see Section 3.5.

2 Preliminaries

In this section, we review basic properties of the Legendre polynomials and the PSWFs, and outline the Galerkin method. We also present some error estimates for the symmetric eigenvalue problem.

2.1 Legendre polynomials

The Legendre polynomials $P_0, P_1, \ldots$, are defined by the recurrence relation ([1], *22.7.10*)

$$P_0(x) = 1, \tag{7}$$

$$P_1(x) = x, \tag{8}$$

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x) \quad \text{for } n \geqslant 1. \tag{9}$$

It follows from the definition, that P_n is an even function when n is even, and an odd function when n is odd. The Legendre polynomial P_n is a solution of the differential equation ([1], *22.6.13*)

$$-\left(\left(1-x^2\right)f'\right)' = n(n+1)f. \tag{10}$$

The Legendre polynomials are orthogonal on the interval $[-1,1]$, but not orthonormal. It is convenient to use the normalized Legendre polynomials ([1], *22.2.10*)

$$\widetilde{P}_n := \sqrt{n+\tfrac{1}{2}}\,P_n \quad \text{for } n \geqslant 0. \tag{11}$$

The normalized Legendre polynomials $\left\{\widetilde{P}_n\right\}_{n\geqslant 0}$ form an orthonormal basis of the Lebesgue space $L^2(-1,1)$, ([4], Theorem 4.8.5).

2.2 Prolate spheroidal wave functions

For a fixed $c > 0$, we are concerned with real-valued solutions $w = w(x)$ of the prolate differential equation

$$L_c w - \chi w = -\left((1-x^2)w'\right)' + \left(c^2x^2 - \chi\right)w = 0 \tag{12}$$

We only consider solutions that are continuous at $x = \pm 1$. It is well-known ([10], Section 3.1) that there exists an increasing sequence of distinct real numbers,

$$0 < \chi_0(c) < \chi_1(c) < \chi_2(c) < \ldots, \tag{13}$$

such that for $\chi = \chi_n(c)$ equation (12) has, up to a multiplicative constant, a unique solution $\psi_n(x;c)$ that is continuous at $x = \pm 1$. The numbers $\chi_0, \chi_1, \ldots$ are thus the eigenvalues of the boundary value problem given by (12) with the boundary condition of continuity. The corresponding eigenfunctions $\psi_n(x) = \psi_n(x;c)$ are called the *prolate spheroidal wave functions* (PSWFs) of mode number n and order 0. Following [7], we call the positive number c the bandwidth parameter of the PSWFs $\psi_n(x;c)$.

It follows from the symmetry of the prolate differential equation that ψ_n is an even function when n is even, and an odd function when n is odd.

2.3 Legendre-Galerkin method

The Legendre-Galerkin method (see [6, Section 3.1]) uses the basis of the normalized Legendre polynomials $\widetilde{P}_0, \widetilde{P}_1, \dots$ to discretize the eigenvalue problem for the differential operator L_c with the boundary conditions of continuity. The Galerkin matrix A for this problem is given by

$$A_{i,j} = \left\langle L_c \widetilde{P}_j, \widetilde{P}_i \right\rangle \quad i,j = 0,1,\dots. \tag{14}$$

Combining (9), (10) and (11) (see [8, Section 2]), one can show that the only nonzero elements of the matrix A are given by

$$A_{i,i} = i(i+1) + c^2 \frac{2i(i+1)-1}{(2i-1)(2i+3)} \qquad i = 0,1,\dots, \tag{15}$$

$$A_{i,i-2} = A_{i-2,i} = c^2 \frac{i(i-1)}{(2i-1)\sqrt{(2i-3)(2i+1)}} \qquad i = 2,3,\dots. \tag{16}$$

Thus eigenproblem (2) is reduced to a matrix eigenvalue problem of the form

$$A\mathbf{a} = \chi \mathbf{a}, \tag{17}$$

where $\mathbf{a} = \{a_j(n,c)\}_{j \geqslant 0}$ are the coefficients of the PSWF $\psi_n(x;c)$ with the mode number n and bandwidth parameter c with respect to the normalized Legendre polynomials

$$\psi_n(x;c) = \sum_{j \geqslant 0} a_j(n,c)\, \widetilde{P}_j(x). \tag{18}$$

We notice that nonzero entries of A appear only on the main diagonal, the second subdiagonal, and the second superdiagonal. Therefore (17) decouples into two tridiagonal eigenproblems, see also [8, Section 2]. If n is even, the first N even Legendre coefficients of ψ_n are approximated by the eigenvector corresponding to the $\left(\frac{n}{2}+1\right)$st smallest eigenvalue of the $N \times N$ matrix

$$A_N^e := \{A_{2i,2j}\}_{i,j=0,\dots,N-1}. \tag{19}$$

If n is odd, the first N odd Legendre coefficients of ψ_n are approximated by the eigenvector corresponding to the $\left(\frac{n+1}{2}\right)$th eigenvalue of the $N \times N$ matrix

$$A_N^o := \{A_{2i+1,2j+1}\}_{i,j=0,\dots,N-1}. \tag{20}$$

In other words, approximate Legendre coefficients $\tilde{\mathbf{a}} = \tilde{\mathbf{a}}(n,c,N)$ of ψ_n are constructed from the corresponding eigenvectors $b = (b_1, b_2, \dots, b_N)^T$ of the matrices (19) and (20), respectively, interleaved with zeros and zero-padded to sequences in $\ell^2(\mathbb{N}_0)$, i.e.

$$\tilde{\mathbf{a}} = \begin{cases} (b_1, 0, b_2, 0, \dots, b_N, 0, 0, 0, 0, \dots)^T & \text{for } n \text{ even}, \\ (0, b_1, 0, b_2, \dots, 0, b_N, 0, 0, 0, \dots)^T & \text{for } n \text{ odd}. \end{cases} \tag{21}$$

Both matrices A_N^e and A_N^o are real and symmetric. They are also positive definite if $c > 0$. This follows from (1) and (14) by integration by parts.

	operation count
all eigenvalues	$\mathcal{O}(N \log N)$
nth eigenvector	$\mathcal{O}(N)$
evaluation of the Legendre series	$\mathcal{O}\left(N \log^2 N + M\right)$

Table 1 Operation count of the evaluation of ψ_n at M nodes.

2.4 Complexity of the Legendre-Galerkin algorithm

In this section, we derive the operation count of the Legendre-Galerkin algorithm for the numerical evaluation of the PSWF ψ_n described in Section 2.3.

If n is even we compute all eigenvalues of the matrix A_N^e, see (19), if n is odd we compute all eigenvalues of the matrix A_N^o, see (20). Since both matrices are symmetric and tridiagonal, the divide-and-conquer algorithm [12] finds the eigenvalues in $\mathcal{O}(N \log N)$ flops. We note that this operation count is sufficient to compute the eigenvalues with the absolute precision roughly equal to the machine epsilon times the norm of the matrix. However, if needed, the eigenvalues and eigenvectors of tridiagonal symmetric positive definite matrices A_N^e and A_N^o can be computed to high *relative* precision using more sophisticated methods [2].

Then, for even n, we find an eigenvector corresponding to the $\left(\frac{n}{2}+1\right)$st smallest eigenvalue of A^e. For odd n, we find an eigenvector corresponding to the $\left(\frac{n+1}{2}\right)$th smallest eigenvalue of A^o. The convergence rate and the complexity of this step depends on the distribution of the eigenvalues. In our case, the eigenvalues of both matrices appear experimentally to be separated by a fixed constant, see Section 3.6. Consequently, inverse iterations with shifts converge to precision ε in $\mathcal{O}(|\log \varepsilon|)$ iterations, which results in $\mathcal{O}(N)$ operations per eigenvector.

As described in Section 2.3, the entries of the eigenvector approximate the first N Legendre coefficients in (18). At M arguments, the N-th partial sum of the Legendre series can be evaluated in $\mathcal{O}(MN)$ flops. However, if the number of arguments M is large, the evaluation of the Legendre series can be accelerated by the fast polynomial transform [21] to $\mathcal{O}\left(N \log^2 N + M\right)$ flops. Thus in fixed precision, the resulting total operation count L satisfies

$$L = \mathcal{O}\left(N \log^2 N + M\right). \tag{22}$$

The operation count of the evaluation of ψ_n at M nodes is summarized in Table 1.

2.5 Perturbation of Eigenvectors

Let A be an $N \times N$ matrix with the eigenvalues $\lambda_1, \ldots, \lambda_N$. For $z \in \mathbb{C}$, we denote the distance from z to all eigenvalues of A different from a fixed eigen-

value λ by

$$\mathrm{gap}_A(z,\lambda) := \min_{\lambda_k \neq \lambda} |z - \lambda_k| \,. \tag{23}$$

Let A be a real symmetric matrix with a simple eigenvalue λ, and let u be a corresponding normalized eigenvector. The condition number of the eigenvector u measures the rate of change of the eigenvector under infinitesimally small perturbations of the matrix. Following [22, Section 3.3.2], we define the condition number of u with respect to the matrix operator norm as

$$\mathrm{cond}(u) := \max_{\|E\|=1} \lim_{t\to 0} \frac{1}{t} \|\mathbf{u}(A+tE) - u\| \,, \tag{24}$$

where the maximum is taken over all real $N \times N$ matrices E, and $\mathbf{u}(A+tE)$ is an eigenvector of $A+tE$ chosen so that $\mathbf{u}(A+tE)$ is a continuous function of t defined in a neighborhood of $t=0$ with $\mathbf{u}(0) = u$. The following expression for the absolute condition number is proved in [22, (3.45)]

$$\mathrm{cond}(u) = \frac{1}{\mathrm{gap}_A(\lambda,\lambda)}. \tag{25}$$

The condition number is used to estimate forward errors ε_f occurring in the computation of u as follows,

$$\varepsilon_f \approx \mathrm{cond}(u)\, \varepsilon_b. \tag{26}$$

The backward error ε_b is typically equal to the matrix norm times the precision of the computations ε,

$$\varepsilon_b \approx \varepsilon \, \|A\| \,. \tag{27}$$

Other possible estimates for the backward error include a slowly growing function of the matrix dimensions. Typically, the backward error grows linearly, quadratically, or cubically with the dimension of the matrix, see [25, p. 106]. Combining (25), (26) and (27), we arrive at the following error estimate of the error for u

$$\varepsilon_f \approx \varepsilon \, \frac{\|A\|}{\mathrm{gap}_A(\lambda,\lambda)}. \tag{28}$$

If A is Hermitian with non-negative eigenvalues, then

$$\varepsilon_f \approx \varepsilon \, \frac{\lambda_{\max}(A)}{\mathrm{gap}_A(\lambda,\lambda)}, \tag{29}$$

where $\lambda_{\max}(A)$ is the largest eigenvalue of A.

3 Numerical experiments

The results presented below are selected from a larger set of numerical simulations covering wide ranges of parameters. However, our conclusions are consistent with experiments that we do not include for the sake of brevity. Thus our conclusions are based on a consistent behavior of the experimental curves for all parameters, and the asymptotic regime has been reached in our simulations.

We are interested in how the minimal dimension $N = N(n, c, \varepsilon)$ of the Legendre-Galerkin matrix required to achieve precision ε depends on the prolate mode number n and the bandwidth parameter c. In particular, we investigate the limiting behavior as $n, c \to \infty$ and $\varepsilon \to 0^+$.

We study properties of the matrices A_N^e and A_N^o defined by (19) and (20), respectively. For the sake of brevity, we include the results for A_N^e only. Thus Figures 1(a), 3(a), 7(a), 8(a) and 10(a) show the curves restricted to the even values of n.

Our experiments have been done in MATLAB R2012a. Some computations in Section 3.5 have been done in extended precision using the Multiple Precision Toolbox for MATLAB created by Ben Barrowes [3].

3.1 Dimension of the Legendre-Galerkin matrix for fixed precision

We assume that the prolates are normalized in $L^2(-1, 1)$, so that their normalized Legendre coefficients $\mathbf{a} = \{a_j(n, c)\}_{j \geqslant 0}$ in (18) are unit vectors in $\ell^2(\mathbb{N}_0)$. With our normalization, the absoute and relative errors coincide

$$\frac{\|\mathbf{a} - \tilde{\mathbf{a}}\|}{\|\mathbf{a}\|} = \|\mathbf{a} - \tilde{\mathbf{a}}\|. \tag{30}$$

We also normalize approximate Legendre coefficients, so that $\|\tilde{\mathbf{a}}\| = 1$.

We use Parseval's identity to express the L^2-norm of the kth remainder of the series (18) in terms of the coefficients $a_j(n, c)$, $k = 0, 1, \ldots$,

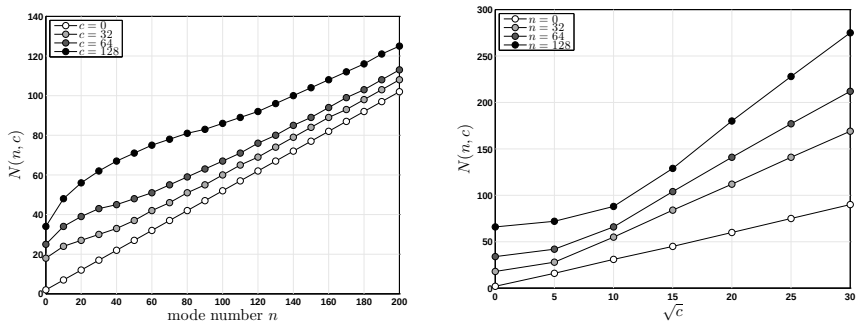
$$r_k(n, c) := \left\| \sum_{j \geqslant k+1} a_j(n, c)\, \widetilde{P}_j \right\|_2 = \left(\sum_{j \geqslant k+1} |a_j(n, c)|^2 \right)^{\frac{1}{2}}. \tag{31}$$

For every $\varepsilon > 0$, we denote by $d(n, c, \varepsilon)$ the smallest index k such that $r_{2k-1}(n, c) \leqslant \varepsilon$,

$$d(n, c, \varepsilon) = \min \{k : r_{2k-1}(n, c) \leqslant \varepsilon\}. \tag{32}$$

Thus d measures the rate of decay of the Legendre coefficients of ψ_n.

As explained in Section 2.3, approximate coefficients $\tilde{\mathbf{a}}(n, c, M)$ are computed as the entries of an eigenvector corresponding to the $\left(\frac{n}{2} + 1\right)$st or $\left(\frac{n+1}{2}\right)$th smallest eigenvalue of A_M^e or A_M^o, respectively, interleaved with zeros and



(a) (b)

Fig. 1 The minimal dimension $N(n, c, \varepsilon)$ of the Legendre-Galerkin matrix required to compute the nth prolate with precision $\varepsilon = 10^{-8}$.

zero-padded in the remaining entries. We denote by $N(n, c, \varepsilon)$ the smallest dimension of the Legendre-Galerkin matrix required to compute the nth prolate with precision ε,

$$N(n, c, \varepsilon) = \min \{M : \|\mathbf{a}(n, c) - \tilde{\mathbf{a}}(n, c, M)\| \leqslant \varepsilon\}, \tag{33}$$

We sometimes write $d(n, c)$, $N(n, c)$ when ε is fixed.

According to (21), the only possibly nonzero entries of $\tilde{\mathbf{a}}(n, c, M) = \{\tilde{a}_j\}_{j\geqslant 0}$ are $\tilde{a}_0, \tilde{a}_2, \ldots, \tilde{a}_{2M-2}$ if n is even, or $\tilde{a}_1, \tilde{a}_3, \ldots, \tilde{a}_{2M-1}$ if n is odd. Thus $\tilde{a}_j = 0$ for $j > 2M - 1$, and $r_{2M-1} \leqslant \|\mathbf{a} - \tilde{\mathbf{a}}\|$. If $M = N(n, c, \varepsilon)$, then $r_{2M-1} \leqslant \|\mathbf{a} - \tilde{\mathbf{a}}\| \leqslant \varepsilon$. Consequently,

$$d(n, c, \varepsilon) \leqslant N(n, c, \varepsilon). \tag{34}$$

In our experiments, we have observed that the two quantities are very close to each other. In the studied examples, the minimal dimension $N(n, c, \varepsilon)$ of the Legendre-Galerkin matrix exceeds $d(n, c, \varepsilon)$ by at most 1. We conjecture that

$$N(n, c, \varepsilon) \leqslant d(n, c, \varepsilon) + \mathcal{O}(|\log \varepsilon|). \tag{35}$$

Figure 1 shows the quantity $N(n, c, \varepsilon)$ for $\varepsilon = 10^{-8}$, the corresponding plot for $d(n, c)$ is visually identical, and is not included. Figure 1(a) shows $N(n, c, \varepsilon)$ as a function of n, while Figure 1(b) shows as a function of $\sqrt{c}$. We notice that all curves in Figure 1(b) asymptote to linear functions of $\sqrt{c}$.

On the basis of our experiments, we formulate the following estimate, which is one of our main contributions (4).

Observation 1 For a fixed $\varepsilon > 0$,

$$N(n, c, \varepsilon) = \mathcal{O}\left(n + \sqrt{nc}\right). \tag{36}$$

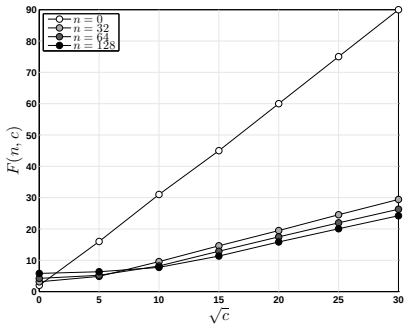


Fig. 2 The quotient $F(n,c) = \frac{N(n,c,\varepsilon)}{\sqrt{n+1}}$ for $\varepsilon = 10^{-8}$.

To support our claim, we introduce the function

$$F(n,c,\varepsilon) = \frac{N(n,c,\varepsilon)}{\sqrt{n+1}}. \tag{37}$$

For $c = 0$, the Legendre-Galerkin matrix is a diagonal, see (16), and all Galerkin eigenvectors are the exact Legendre coefficients of the respective eigenfunctions. Thus the minimal dimension to compute the nth eigenvector after the decoupling is equal to $\left\lceil \frac{n+1}{2} \right\rceil$, i.e.

$$N(n,0) = \left\lceil \frac{n+1}{2} \right\rceil. \tag{38}$$

Consequently,

$$F(n,0) = \frac{N(n,0)}{\sqrt{n+1}} = \mathcal{O}\left(\sqrt{n}\right). \tag{39}$$

Figure 2 shows $F(n,c)$ as a function of $\sqrt{c}$ for $n = 0, 32, 64, 128$. We observe that

$$F(0,c) = \mathcal{O}\left(\sqrt{c}\right). \tag{40}$$

Moreover, for a fixed argument $\sqrt{c}$, the slopes of the curves decrease when n grows, which implies that

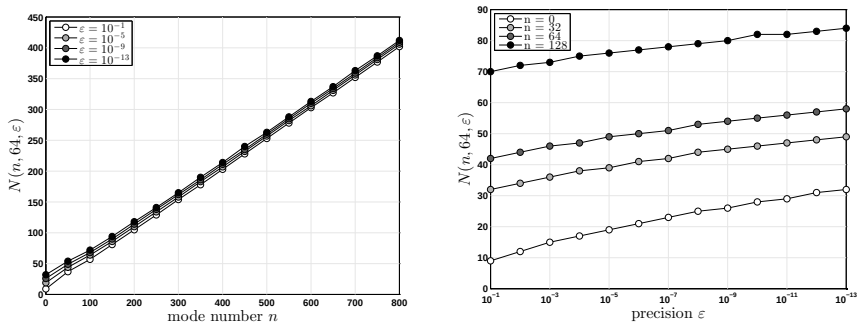
$$F(n,c) - F(n,0) \leqslant F(0,c) - F(0,0). \tag{41}$$

Combining (39), (40) and (41), we deduce that

$$F(n,c) = \mathcal{O}\left(\sqrt{n} + \sqrt{c}\right). \tag{42}$$

Now (36) follows from (37) and (42).

We note that this discussion reflects a general pattern discovered experimentally, but leaves out discretization effects. Due to a discrete nature of the quantities N and F, we expect that some exceptions to the statement that the



(a) (b)

Fig. 3 The minimal dimension $N(n, c, \varepsilon)$ of the Legendre-Galerkin matrix for $c = 64$.

slopes in Figure 2 decrease with n, especially for small values of c. A rigorous proof of (36) should take that into account.

The dimension of the Galerkin matrix determines the operation count of the Galerkin method. Substituting (36) into (22), we obtain the following upper bound on the operation count L required for the evaluation of the nth prolate at M points in fixed precision

$$L = \mathcal{O}\left((n + \sqrt{nc})(\log^2 n + \log^2 c) + M\right). \tag{43}$$

3.2 Dimension of the Legendre-Galerkin matrix for a fixed bandwidth

Figure 3 shows the quantity $N(n, c, \varepsilon)$ for $c = 64$. Figure 3(a) shows $N(n, c, \varepsilon)$ as a function of n, while Figure 3(b) shows $N(n, c, \varepsilon)$ as a function of ε. We notice that all curves in Figure 3(a) asymptote to linear functions of n. On the basis of our experiments, we formulate the following estimate.

Observation 2 For a fixed $c > 0$,

$$N(n, c, \varepsilon) = \mathcal{O}\left(n + |\log \varepsilon|\right). \tag{44}$$

Our justification of (44) is similar to the one used for (36). We observe in Figure 3(b) that the slopes of $N(n, c, \varepsilon)$ viewed as a function of ε decrease when n grows. Consequently,

$$N(n, c, \varepsilon) - N(n, c, 0.1) \leqslant N(0, c, \varepsilon) - N(0, c, 0.1). \tag{45}$$

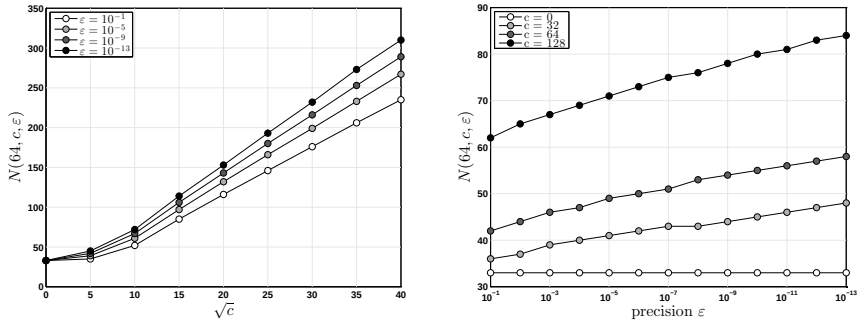
We also observe in Figure 3(b) that

$$N(0, c, \varepsilon) = \mathcal{O}\left(|\log \varepsilon|\right). \tag{46}$$

We observe in Figure 3(a) that

$$N(n, c, 0.1) = \mathcal{O}\left(n\right). \tag{47}$$

Combining (45), (46), (47), we obtain (44).



(a) (b)

Fig. 4 The minimal dimension $N(n, c, \varepsilon)$ of the Legendre-Galerkin matrix for $n = 64$.

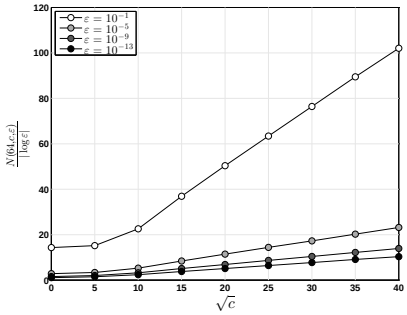


Fig. 5 The quantity $N(n, c, \varepsilon)/|\log \varepsilon|$ for $n = 64$.

3.3 Dimension of the Legendre-Galerkin matrix for a fixed mode number

Figure 4 shows the quantity $N(n, c, \varepsilon)$ for $n = 64$. Figure 4(a) shows $N(n, c, \varepsilon)$ as a function of $\sqrt{c}$, while Figure 4(b) shows $N(n, c, \varepsilon)$ as a function of ε. We notice that all curves in Figure 4(a) asymptote to linear functions of $\sqrt{c}$. On the basis of our experiments, we formulate the following estimate.

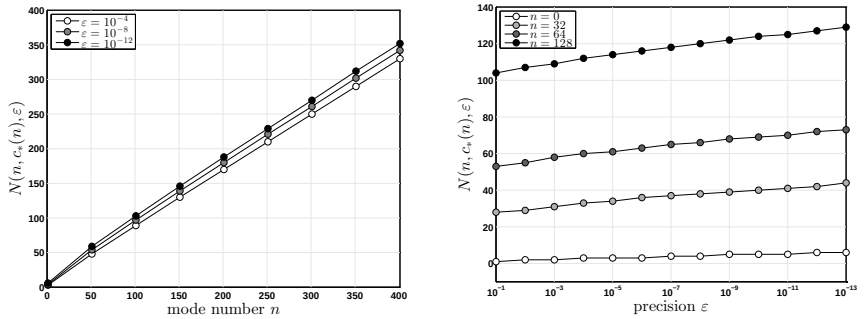
Observation 3 For a fixed n,

$$N(n, c, \varepsilon) = \mathcal{O}\left(\sqrt{c}\,|\log \varepsilon|\right). \tag{48}$$

We observe in Figure 5 that

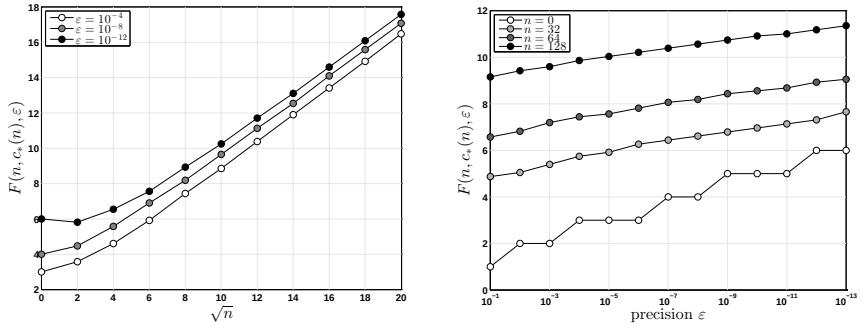
$$\frac{N(n, c, \varepsilon)}{|\log \varepsilon|} = \mathcal{O}\left(\sqrt{c}\right), \tag{49}$$

which implies (48).



(a) (b)

Fig. 6 The minimal dimension $N(n, c, \varepsilon)$ of the Legendre-Galerkin matrix for $c = c_*(n)$.



(a) (b)

Fig. 7 The quotient $F(n, c_*(n), \varepsilon) = \frac{N(n, c_*(n), \varepsilon)}{\sqrt{n+1}}$.

3.4 Dimension of the Legendre-Galerkin matrix for $\mathbf{c} = \mathbf{c}_*(\mathbf{n})$

The behavior of the prolates as finite elements depends crucially on the so called transition bandwidth defined as

$$c_*(n) := \frac{\pi}{2}\left(n + \tfrac{1}{2}\right), \tag{50}$$

where n is the largest mode number. The assumption $c \leqslant c_*(n)$, see [7, Theorem 5.3], is necessary for the numerical stability of the prolate expansion.

We study the minimal dimension of the Galerkin matrix $N(n, c_*(n), \varepsilon)$ needed to compute the nth PSWF with the bandwidth parameter $c = c_*(n)$ with precision ε. Figure 6 shows the quantity $N(n, c_*(n), \varepsilon)$. Figure 6(a) shows $N(n, c_*(n), \varepsilon)$ as a function of n, while Figure 6(b) shows $N(n, c_*(n), \varepsilon)$ as a function of ε. On the basis of our experiments, we formulate the following estimate.

Observation 4

$$N(n, c_*(n), \varepsilon) = \mathcal{O}\left(n + \sqrt{n}\,|\log \varepsilon|\right). \tag{51}$$

Our justification of (51) is analogous to the one used for (36). Figure 7(a) shows the quotients

$$F(n, c_*(n), \varepsilon) = \frac{N(n, c_*(n), \varepsilon)}{\sqrt{n+1}} \tag{52}$$

as functions of $\sqrt{n}$ for several values of ε. We observe that for a fixed argument $\sqrt{n}$, the slopes of the curves decrease when $|\log \varepsilon|$ increases. From Figure 7(b), we infer that

$$F(0, c_*(0), \varepsilon) = \mathcal{O}(|\log \varepsilon|), \tag{53}$$

and thus (51) holds.

A brief comment about (53) is in order. We have observed experimentally (35), in particular

$$F(0, c_*(0), \varepsilon) = N(0, c_*(0), \varepsilon) \leqslant d(0, c_*(0), \varepsilon) + \mathcal{O}(|\log \varepsilon|). \tag{54}$$

Therefore (53) follows from

$$d(0, c_*(0), \varepsilon) = \mathcal{O}(|\log \varepsilon|). \tag{55}$$

In turn, (55) follows from exponential decay of the Legendre coefficients of the prolate $\psi_0(x, c_*(0))$, see [7, Theorem 7.3].

We note that (51) is consistent with the dimension $N_{trunc} = 2n + 30$ of the Galerkin matrix (before decoupling) used in [8, equation (30)] in double precision.

3.5 Conditioning of the Legendre coefficients in fixed precision

In this section, we study the accuracy of the approximations of the Legendre coefficients in (18) by the eigenvectors of the Legendre-Galerkin matrix. For n even, the even Legendre coefficients of ψ_n are approximated by an eigenvector corresponding to the $\left(\frac{n}{2}+1\right)$st smallest eigenvalue of the matrix A_N^e defined in (19). For an odd n, the odd Legendre coefficients of ψ_n are approximated by an eigenvector corresponding to the $\left(\frac{n+1}{2}\right)$th smallest eigenvalue of the matrix A_N^o defined in (20).

Backward errors given by (27) result from rounding. They have irregular behavior, so both backward and forward errors are difficult to investigate experimentally. Therefore, we use estimate (29) for the forward error, and study experimentally the forward errors together with their analytic estimates.

According to (29), the forward error ε_f of the normalized eigenvector u_n can be estimated as follows

$$\varepsilon_f \approx \varepsilon_0 \, \frac{\lambda_{\max}(B)}{\mathrm{gap}_B(\chi_n, \chi_n)}, \tag{56}$$

where ε_0 is the machine precision, $\lambda_{\max}(B)$ is the maximal eigenvalue of B, χ_n is the eigenvalue of u_n, and $B = A_N^e$ or $B = A_N^o$ depending on whether n

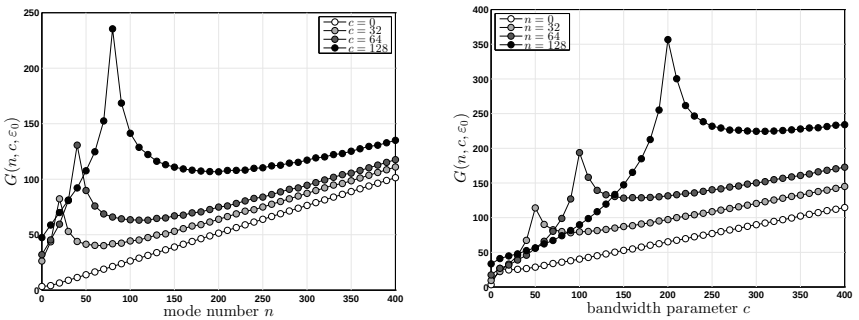


Fig. 8 The function $G(n, c, \varepsilon)$ for $\varepsilon = 2^{-52}$.

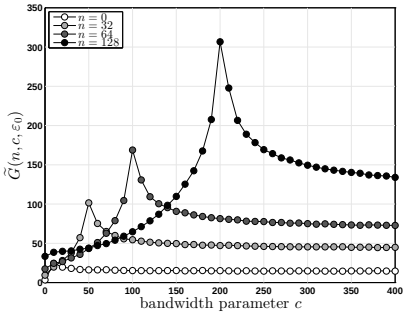


Fig. 9 The quantity $\tilde{G}(n, c, \varepsilon) = G(n, c, \varepsilon) - \frac{c}{4}$ for $\varepsilon = 2^{-52}$.

is even or odd. The entries and the dimension N of the matrix B depend on the parameters n and c, and the precision ε chosen in the algorithm. For the sake of readability, we do not show this dependence.

Let G denote the function

$$G(n, c, \varepsilon) = \frac{\lambda_{\max}(B)}{\mathrm{gap}_B(\chi_n, \chi_n)}, \tag{57}$$

so that the approximate error ε_f of the normalized eigenvector u_n satisfies

$$\varepsilon_f \approx \varepsilon_0 G(n, c, \varepsilon_0). \tag{58}$$

Figure 8 shows the quantity $G(n, c, \varepsilon)$ for $\varepsilon = 2^{-52}$. As a function of c, $G(n, c, \varepsilon)$ has a local maximum located close to the transition bandwidth $c = c_*(n)$. Based on our experiments, we make the following claim.

Observation 5 For a fixed $\varepsilon > 0$,

$$G(n, c, \varepsilon) = \mathcal{O}\left(n + c\right). \tag{59}$$

To justify this claim, we consider the function $\tilde{G}(n, c, \varepsilon) := G(n, c, \varepsilon) - \frac{c}{4}$, see Figure 9. As a function of c, $\tilde{G}(n, c, \varepsilon)$ is unimodal, and has a maximum

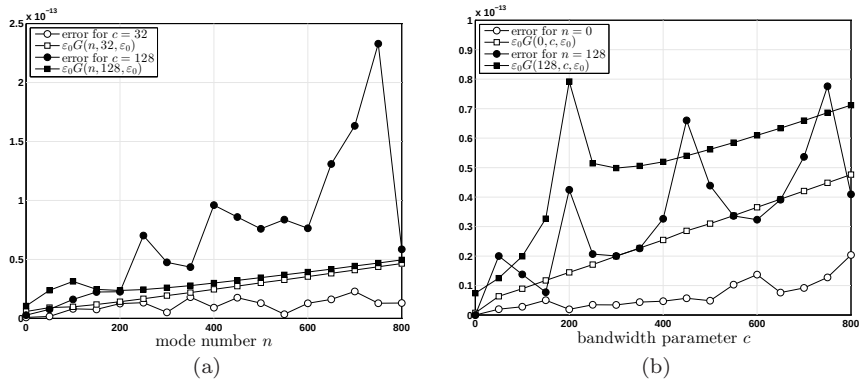


Fig. 10 The error and the estimated error.

value, which grows at most linearly with n. Thus $\tilde{G}(n, c, \varepsilon) = \mathcal{O}(n)$, and (59) follows. It is noteworthy that the growth in c is described by the function $\frac{c}{4}$ independently of the precision ε.

Figure 10 shows the errors of the normalized Legendre coefficients of prolates. We also show estimated errors obtained via (56) with $\varepsilon_0 = 2^{-52}$.

In these experiments, we use the Galerkin matrix of dimension $N(n, c, 2^{-52})$, so that the truncation error is equal to the machine epsilon in double precision. We notice that our error estimates based on the condition number of the eigenvectors of the corresponding Galerkin matrix are within an order of magnitude from the actual errors. We note that our error estimates are approximations of the errors, but they are not necessarily upper bounds. Thus we do not expect the computed errors to be majorized by the estimated ones.

We have also conducted experiments on conditioning of the eigenvectors of the Galerkin matrices using the perturbation bound for scaled diagonally dominant matrices given in [2, Theorem 6]. However, we have not observed any improvement over (59).

3.6 Distribution of the eigenvalues

Figure 11 shows the minimal gaps between consecutive eigenvalues of the Galerkin matrices. For the matrices A_N^e, the gaps between the consecutive eigenvalues equal at least 6, for the matrices A_N^o, the gaps are at least 10. The separation of the eigenvalues of the Galerkin matrices affects both complexity and accuracy of the Legendre-Galerkin method, see Section 2.4 and Section 3.5, respectively.

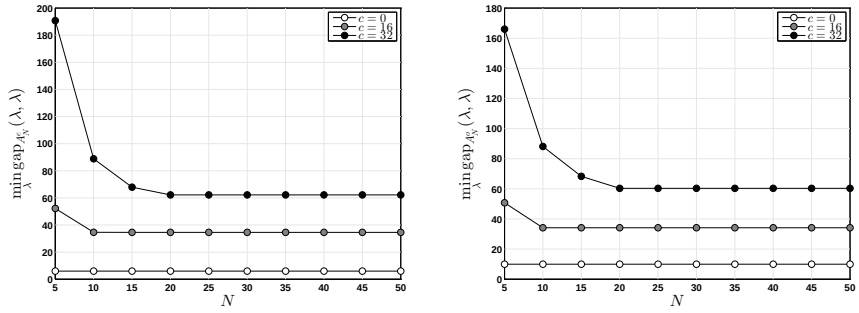


Fig. 11 The minimal gaps between consecutive eigenvalues of the Galerkin matrices.

3.7 Summary of the experiments

We study the minimal dimension $N = N(n, c, \varepsilon)$ of the Legendre-Galerkin matrix required to compute the nth prolate of bandwidth c with precision ε. The operation count required for the evaluation of the nth prolate at M points in fixed precision ε is denoted by L. The function G defined by (57) is used to estimate the error of an appropriate eigenvector of the Galerkin matrix.

1. For a fixed $\varepsilon > 0$
$$N(n, c, \varepsilon) = \mathcal{O}\left(n + \sqrt{nc}\right). \tag{60}$$
2. For a fixed $\varepsilon > 0$
$$L(n, c, \varepsilon, M) = \mathcal{O}\left((n + \sqrt{nc})(\log^2 n + \log^2 c) + M\right). \tag{61}$$
3. For a fixed $c > 0$
$$N(n, c, \varepsilon) = \mathcal{O}\left(n + |\log \varepsilon|\right). \tag{62}$$
4. For a fixed n
$$N(n, c, \varepsilon) = \mathcal{O}\left(\sqrt{c}\,|\log \varepsilon|\right). \tag{63}$$
5. If $c = c_*(n)$
$$N(n, c_*(n), \varepsilon) = \mathcal{O}\left(n + \sqrt{n}\,|\log \varepsilon|\right). \tag{64}$$
6. For a fixed $\varepsilon > 0$
$$G(n, c, \varepsilon) = \mathcal{O}\left(n + c\right). \tag{65}$$
7. The gaps between the consecutive eigenvalues of the Galerkin matrices are uniformly bounded from below.

2. On p. 18, we explain that our conclusions are based on a consistent behavior of the experimental curves for all parameters, and the asymptotic regime has been reached in our simulations.

4 Conclusions

Observations 1, 2 and 3 combined lead to a conjectural upper bound on the minimal dimension of the Legendre-Galerkin matrix required to achieve precision ε

$$N(n, c, \varepsilon) = \mathcal{O}\left(n + \sqrt{nc} + \sqrt{c}\,|\log \varepsilon|\right). \tag{66}$$

However, this estimate is fairly complex, and our numerical evidence for this conjecture is inconclusive.

It is also interesting to notice that this dimension $N(n, c, \varepsilon)$ is very close to the minimal length $d(n, c, \varepsilon)$ of the Legendre series needed to approximate ψ_n with precision ε. This means that the Legendre-Galerkin method efficiently computes very accurate approximations to the Legendre coefficients. The behavior of the difference $N(n, c, \varepsilon) - d(n, c, \varepsilon)$ requires further studies.

In Subsection 3.5, we observe that the error estimate based on the condition number of the eigenvector containing the Legendre coefficients of the nth prolate is $\mathcal{O}(n + c)$ as $n, c \to \infty$. Using computations in extended precision, we observe that the errors of the computed PSWFs grow in a similar way. This loss of precision is gradual and fairly moderate, so the Legendre-Galerkin method is accurate for reasonably large values of n and c.

In Subsection 3.6, we observe that the consecutive eigenvalues of the Galerkin matrices are separated from each other by a fixed constant. We are not aware of a rigorous proof of this fact.

While our analysis of the experiments is not itself rigorous, it reveals the underlying structure of the studied quantities, and may help formulate rigorous proofs.

Acknowledgments

The first author is supported by the Austrian Science Fund (FWF) NFN S106 grant Signal and Information Processing in Science and Engineering (SISE). The second author is supported by the FWF grant I-1018-N25.

The authors are grateful for the comments of the editors and the reviewers.

References

1. M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1965.
2. J. Barlow and J. Demmel. Computing accurate eigensystems of scaled diagonally dominant matrices. *SIAM J. Numer. Anal.*, 27(3):762–791, 1990.
3. B. Barrowes. Multiple Precision Toolbox for MATLAB, Dec 2004. MATLAB Central File Exchange. Retrieved Jan 28, 2013 from http://www.mathworks.com/matlabcentral/fileexchange/6446.
4. R. Beals and R. Wong. *Special Functions - A Graduate Text*. Cambridge University Press, 2010.
5. C. J. Bouwkamp. On spheroidal wave functions of order zero. *J. Math. Phys. Mass. Inst. Tech.*, 26:79–92, 1947.

6. J. P. Boyd. *Chebyshev and Fourier Spectral Methods*. Dover Publications, Inc., 2nd (revised) edition, 2001.
7. 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):688–716, 2004.
8. 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 Transactions on Mathematical Software*, 31(1):149–165, March 2005.
9. P. E. Falloon, P. C. Abbott, and J. B. Wang. Theory and computation of spheroidal wavefunctions. *J. Phys. A*, 36(20):5477–5495, 2003.
10. C. Flammer. *Spheroidal wave functions*. Stanford University Press, Stanford, California, 1957.
11. W. H. J. Fuchs. On the eigenvalues of an integral equation arising in the theory of band-limited signals. *J. Math. Anal. Appl.*, 9:317–330, 1964.
12. M. Gu and S. C. Eisenstat. A divide-and-conquer algorithm for the symmetric tridiagonal eigenproblem. *SIAM J. Matrix Anal. Appl.*, 16(1):172–191, 1995.
13. D. B. Hodge. Eigenvalues and eigenfunctions of the spheroidal wave equation. *J. Mathematical Phys.*, 11:2308–2312, 1970.
14. T. Hrycak, S. Das, and G. Matz. Inverse methods for reconstruction of channel taps in OFDM systems. *IEEE Transactions on Signal Processing*, 60(5):2666–2671, 2012.
15. H. J. Landau and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty. II. *Bell System Tech. J.*, 40:65–84, 1961.
16. H. J. Landau and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty. III. The dimension of the space of essentially time- and band-limited signals. *Bell System Tech. J.*, 41:1295–1336, 1962.
17. A. Osipov. Certain inequalities involving prolate spheroidal wave functions and associated quantities. *Applied and Computational Harmonic Analysis*, 35(3):359 – 393, 2013.
18. A. Osipov. Certain upper bounds on the eigenvalues associated with prolate spheroidal wave functions. *Applied and Computational Harmonic Analysis*, 35(2):309 – 340, 2013.
19. A. Osipov and V. Rokhlin. On the evaluation of prolate spheroidal wave functions and associated quadrature rules. *Applied and Computational Harmonic Analysis*, 36(1):108 – 142, 2014.
20. A. Osipov, V. Rokhlin, and H. Xiao. *Prolate spheroidal wave functions of order zero. Mathematical tools for bandlimited approximation.* Berlin: Springer, 2013.
21. D. Potts. Fast algorithms for discrete polynomial transforms on arbitrary grids. *Linear Algebra Appl.*, 366:353–370, 2003. Special issue on structured matrices: analysis, algorithms and applications (Cortona, 2000).
22. Y. Saad. *Numerical Methods for Large Eigenvalue Problems, Revised Edition*. SIAM, 2011.
23. D. Slepian. Prolate spheroidal wave functions, Fourier analysis and uncertainity. IV. Extensions to many dimensions; generalized prolate spheroidal functions. *Bell System Tech. J.*, 43:3009–3057, 1964.
24. D. Slepian and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty. I. *Bell System Tech. J.*, 40:43–63, 1961.
25. L. N. Trefethen and D. Bau. *Numerical Linear Algebra*. SIAM: Society for Industrial and Applied Mathematics, June 1997.
26. H. Widom. Asymptotic behavior of the eigenvalues of certain integral equations. II. *Arch. Rational Mech. Anal.*, 17:215–229, 1964.
27. H Xiao, V Rokhlin, and N Yarvin. Prolate spheroidal wavefunctions, quadrature and interpolation. *Inverse Problems*, 17(4):805–838, 2001.