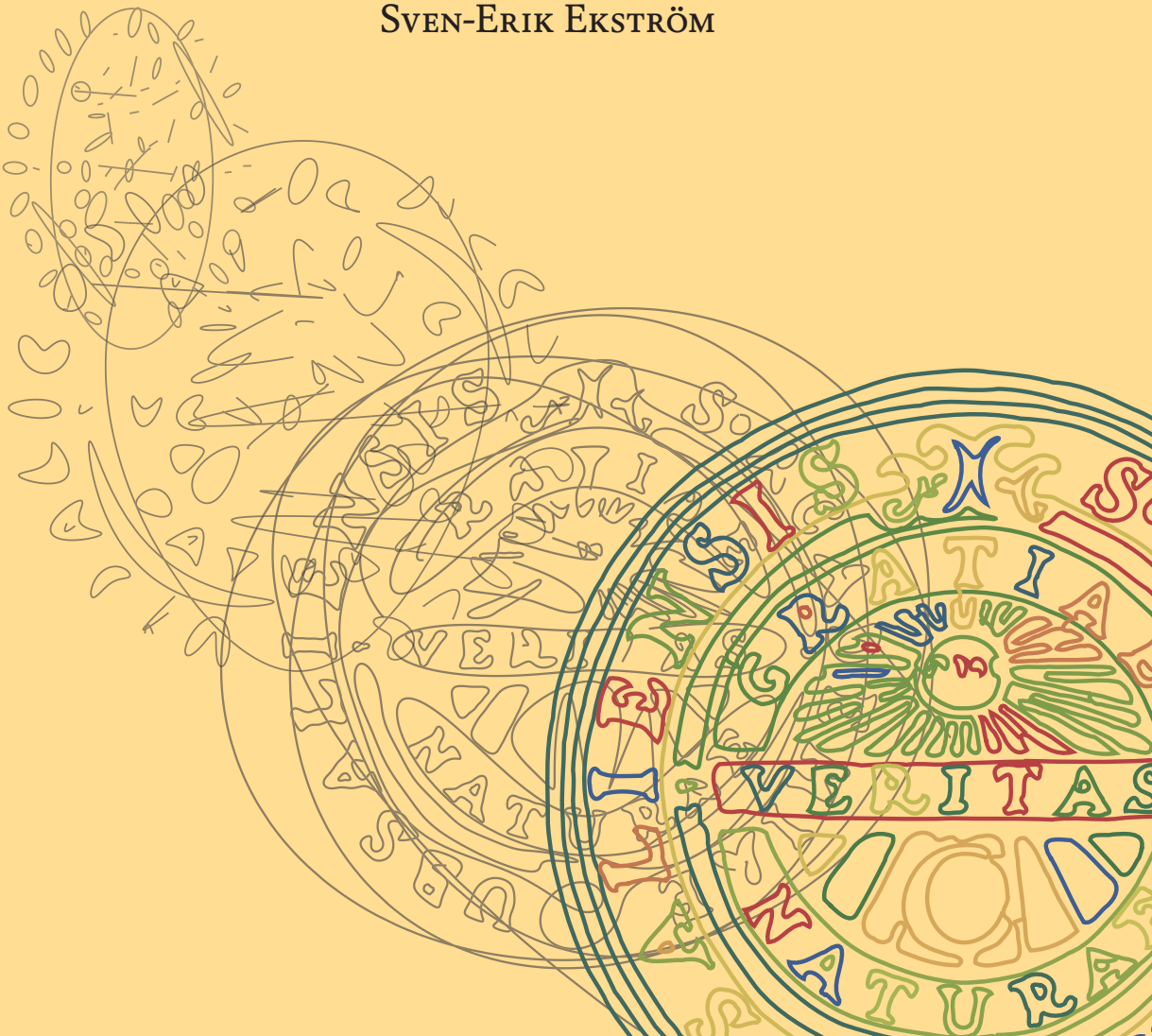
UPPSALA
UNIVERSITET

Matrix-Less Methods for Computing Eigenvalues of Large Structured Matrices

Sven-Erik Ekström



UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1652

Matrix-Less Methods for Computing Eigenvalues of Large Structured Matrices

SVEN-ERIK EKSTRÖM

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2018

ISSN 1651-6214
ISBN 978-91-513-0288-1
urn:nbn:se:uu:diva-346735

Dissertation presented at Uppsala University to be publicly examined in 2446 ITC, Lägerhyddsvägen 2, hus 2, Uppsala, Friday, 18 May 2018 at 10:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Lothar Reichel (Department of Mathematical Sciences, Kent State University).

Abstract

Ekström, S.-E. 2018. Matrix-Less Methods for Computing Eigenvalues of Large Structured Matrices. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 1652. 81 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-0288-1.

When modeling natural phenomena with linear partial differential equations, the discretized system of equations is in general represented by a matrix. To solve or analyze these systems, we are often interested in the spectral behavior of these matrices. Whenever the matrices of interest are Toeplitz, or Toeplitz-like, we can use the theory of Generalized Locally Toeplitz (GLT) sequences to study the spectrum (eigenvalues). A central concept in the theory of GLT sequences is the so-called symbol, that is, a function associated with a sequence of matrices of increasing size. When sampling the symbol and when the related matrix sequence is Hermitian (or quasi-Hermitian), we obtain an approximation of the spectrum of a matrix of a fixed size and we can therefore see its general behavior. However, the so-computed approximations of the eigenvalues are often affected by errors having magnitude of the reciprocal of the matrix size.

In this thesis we develop novel methods, which we call "matrix-less" since they neither store the matrices of interest nor depend on matrix-vector products, to estimate these errors. Moreover, we exploit the structures of the considered matrices to efficiently and accurately compute the spectrum.

We begin by considering the errors of the approximate eigenvalues computed by sampling the symbol on a uniform grid, and we conjecture the existence of an asymptotic expansion for these errors. We devise an algorithm to approximate the expansion by using a small number of moderately sized matrices, and we show through numerical experiments the effectiveness of the algorithm. We also show that the same algorithm works for preconditioned matrices, a result which is important in practical applications. Then, we explain how to use the approximated expansion on the whole spectrum for large matrices, whereas in earlier works its applicability was restricted only to certain matrix sizes and to a subset of the spectrum. Next, we demonstrate how to use the so-developed techniques to investigate, solve, and improve the accuracy in the eigenvalue computations for various differential problems discretized by the isogeometric analysis (IgA) method. Lastly, we discuss a class of non-monotone symbols for which we construct the sampling grid yielding exact eigenvalues and eigenvectors.

To summarize, we show, both theoretically and numerically, the applicability of the presented matrix-less methods for a wide variety of problems.

Keywords: Toeplitz matrices, eigenvalues, eigenvalue asymptotics, polynomial interpolation, extrapolation, generating function and spectral symbol

Sven-Erik Ekström, Department of Information Technology, Division of Scientific Computing, Box 337, Uppsala University, SE-751 05 Uppsala, Sweden.

ISSN 1651-6214
ISBN 978-91-513-0288-1
urn:nbn:se:uu:diva-346735 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-346735)

Dedicated to Rita & Simone

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I S.-E. Ekström, C. Garoni, and S. Serra-Capizzano,
Are the Eigenvalues of Banded Symmetric Toeplitz Matrices Known in Almost Closed Form?,
Experimental Mathematics (2017) (in press)
https://doi.org/10.1080/10586458.2017.1320241

II F. Ahmad, E. S. Al-Aidarous, D. A. Alrehaili, S.-E. Ekström, I. Furci, and S. Serra-Capizzano,
Are the Eigenvalues of Preconditioned Banded Symmetric Toeplitz Matrices Known in Almost Closed Form?,
Numerical Algorithms (2017) (in press)
https://doi.org/10.1007/s11075-017-0404-z

III S.-E. Ekström and C. Garoni,
A Matrix-Less and Parallel Interpolation–Extrapolation Algorithm for Computing the Eigenvalues of Preconditioned Banded Symmetric Toeplitz Matrices,
Numerical Algorithms (2018) (in press)
https://doi.org/10.1007/s11075-018-0508-0

IV S.-E. Ekström, I. Furci, C. Garoni, and S. Serra-Capizzano,
Are the Eigenvalues of the B-spline IgA Approximation of $-\Delta u = \lambda u$ Known in Almost Closed Form?,
Tech. Rep. 2017-016, Department of Information Technology, Uppsala University, Aug. 2017 (submitted)
http://www.it.uu.se/research/publications/reports/2017-016

V S.-E. Ekström and S. Serra-Capizzano,
Eigenvalues and Eigenvectors of Banded Toeplitz Matrices and the Related Symbols,
Numerical Linear Algebra with Applications (2018) (in press)
https://doi.org/10.1002/nla.2137

Related Works

The following papers, although not included, are related to the contents of the present thesis. They are numbered according to the list of references.

[31] S.-E. Ekström, I. Furci, and S. Serra-Capizzano,
Exact Formulae and Matrix-Less Eigensolvers for Block Banded Symmetric Toeplitz Matrices,
Tech. Rep. 2018-005, Department of Information Technology, Uppsala University, Mar. 2018 (submitted)
http://www.it.uu.se/research/publications/reports/2018-005

[32] S.-E. Ekström, C. Garoni, T. J. Hughes, A. Reali, S. Serra-Capizzano, and H. Speleers,
Finite Element and Isogeometric B-spline Discretizations of Eigenvalue Problems: Symbol-Based Analysis, (in preparation)

Contents

1. Introduction

> Indeed, in order to understand the great mathematical events, the comprehensive theories, long schooling and persistent application would be required. But this is also true with music. On going to a concert for the first time one is not able to appreciate fully Bach's "The Art of Fugue," nor can one immediately visualize the structure of a symphony. But besides the great works of music there are the smaller pieces which have something of great sublimity and whose spirit reveals itself to everyone.
>
> Otto Toeplitz
> The Enjoyment of Mathematics

1.1 Background and Motivation

When modeling many natural phenomena, a partial differential equation (PDE) is often the resulting mathematical representation. Whenever the considered model is linear, the PDE takes the form

$$\mathscr{A}u = b,$$

where $\mathscr{A}$ is the linear differential operator, u is the unknown, and b is the source term. Since for many problems the analytical solution u is not available, we have to resort to one of its numerical approximations. For this purpose, we select an existing (or design an ad hoc) numerical method to discretize the continuous problem. The computation of the numerical solution is then reduced to solving a linear system of the form

$$A_n \mathbf{u}_n = \mathbf{b}_n,$$

where the matrix $A_n \in \mathbb{C}^{n\times n}$ is the discrete counterpart of $\mathscr{A}$, $\mathbf{u}_n \in \mathbb{C}^n$ is the numerical solution vector, and $\mathbf{b}_n \in \mathbb{C}^n$ is the source term vector. Under the assumption of convergence of the chosen numerical method, we get closer and closer to the analytical solution in a certain metric, as we increase n. For many types of discretizations, the matrices A_n possess the so-called Toeplitz or Toeplitz-like structure. We stress that, when discussing Toeplitz-like structures we are not only referring to small perturbations of Toeplitz matrices, but also the more general class of generalized locally Toeplitz structures [40, 41, 67, 68, 73]. In this thesis, we focus on these types of matrices.

We develop new theoretical and algorithmic tools for analyzing the spectrum of A_n. This is motivated by the fact that

- it is inherently difficult to solve the linear system $A_n \mathbf{u}_n = \mathbf{b}_n$, due to the conditioning of A_n,
- the convergence rate of mainstream iterative solvers, such as multigrid methods and preconditioned Krylov, when applied to the matrix A_n,
- the design of appropriate multigrid smoothers, prolongation and restriction operators, as well as Krylov preconditioners for the matrix A_n,

are all topics that are strongly related to the eigenvalues (and singular values) of A_n. From the theoretical viewpoint, we conjecture the existence of an asymptotic expansion for the eigenvalues of A_n, and we validate the conjecture through several numerical experiments. From the algorithmic viewpoint, based on the conjectured expansion, we design fast and accurate matrix-less methods for computing the whole spectrum of A_n. We use the new term "matrix-less", instead of the classical "matrix-free", in order to stress that our methods neither need to compute any matrix-vector products, nor need to store the entries of A_n. Indeed, the proposed algorithms only make use of a few matrices A_k for a limited number of different sizes k, much smaller than n.

1.2 Toeplitz Matrices

The first matrices of interest in this thesis, Toeplitz matrices, are named after the German mathematician Otto Toeplitz (1881–1940) [19, 76]. A Toeplitz matrix is a square matrix of size $n \times n$ with constant diagonals, that is, a matrix

$$[a_{i-j}]_{i,j=1}^{n} = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \dots & \dots & a_{1-n} \\ a_1 & \ddots & \ddots & \ddots & & \vdots \\ a_2 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & a_{-2} \\ \vdots & & \ddots & \ddots & \ddots & a_{-1} \\ a_{n-1} & \dots & \dots & a_2 & a_1 & a_0 \end{bmatrix}.$$

Given a function $f \in L^1(-\pi, \pi)$ we can associate with it a sequence of Toeplitz matrices $\{T_n(f)\}_n$ with $T_n(f) = [\hat{f}_{i-j}]_{i,j=1}^{n}$. The constants $\hat{f}_\omega$ are the Fourier coefficients of f, that is,

$$\hat{f}_\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-\mathbf{i}\omega\theta} \mathrm{d}\theta, \quad \mathbf{i}^2 = -1, \quad \omega \in \mathbb{Z}. \tag{1.2.1}$$

The function f is referred to as the generating function of the matrix sequence $\{T_n(f)\}_n$, but it is also called the symbol of $\{T_n(f)\}_n$ because of the Weyl distribution results [3, 57, 70] that we address in Section 1.3, and a more general version presented in [74, 75, 77]. The symbol f generates the finite-dimensional

matrices $T_n(f)$, but also $T(f) = [\hat{f}_{i-j}]_{i,j=1}^{\infty} = [\hat{f}_{i-j}]_{i,j=-\infty}^{\infty}$, which is an infinite (or bi-infinite) dimensional matrix. However, the study of infinite Toeplitz matrices, also known as Toeplitz operators, is out of the scope of this thesis; details on the topic can be found in [17, 19, 20]. From the definition of the Fourier coefficients, we observe that if f is real-valued then $T_n(f)$ is Hermitian ($A_n = A_n^*$) for all n; if f is real-valued and even ($f(\theta) = f(-\theta) \in \mathbb{R}$), almost everywhere, then $T_n(f)$ is real and symmetric for all n. For a given banded Toeplitz matrix $T_n(f)$ with bandwidth $2p+1$, and with a generic central row

$$(T_n(f))_i = \left[0, \ldots, 0, \hat{f}_p, \ldots, \hat{f}_1, \hat{f}_0, \hat{f}_{-1}, \ldots, \hat{f}_{-p}, 0, \ldots, 0\right],$$

the corresponding symbol is,

$$f(\theta) = \sum_{\omega=-p}^{p} \hat{f}_\omega e^{\mathbf{i}\omega\theta}. \tag{1.2.2}$$

In the case of a symmetric banded Toeplitz matrix we have $\hat{f}_\omega = \hat{f}_{-\omega}$ for $\omega = 1, \ldots, p$, and in view of the relation $\hat{f}_\omega(e^{\mathbf{i}\omega\theta} + e^{-\mathbf{i}\omega\theta}) = 2\hat{f}_\omega \cos(\omega\theta)$, (1.2.2) becomes,

$$f(\theta) = \hat{f}_0 + 2\sum_{\omega=1}^{p} \hat{f}_\omega \cos(\omega\theta). \tag{1.2.3}$$

When $\hat{f}_\omega \in \mathbb{R}$ for all $\omega = 0, \ldots p$, we call a symbol of the form (1.2.3) a real cosine trigonometric polynomial (RCTP). Toeplitz matrices generated by RCTPs (in some cases with the addition of a low-rank, with respect to n, correction matrix) represent the starting point of the results in Papers I–IV. Furthermore, in Paper V we deal with a special class of Toeplitz matrices, that we call "symmetrically sparse tridiagonal", generated by symbols of the form (1.2.2), where the only non-zero elements are $\hat{f}_{-p}$, $\hat{f}_0$, and $\hat{f}_p$, for some p.

Example 1.2.1. To illustrate the construction given above, we consider a simple but practical and important example of an RCTP symbol,

$$f(\theta) = 2 - 2\cos(\theta).$$

The Fourier coefficients of $f(\theta)$, by (1.2.1), are $\hat{f}_0 = 2$, $\hat{f}_1 = \hat{f}_{-1} = -1$, and $\hat{f}_\omega = 0$, for all $\omega \neq \{-1, 0, 1\}$. Hence, the bandwidth is $2p + 1 = 3$. Since $f(\theta)$ is of the form of (1.2.3), and $p = 1$, it is an RCTP symbol. The Toeplitz matrix of order $n = 5$, generated by f, is given by,

$$T_5(f) = \begin{bmatrix} \hat{f}_0 & \hat{f}_{-1} & 0 & 0 & 0 \\ \hat{f}_1 & \hat{f}_0 & \hat{f}_{-1} & 0 & 0 \\ 0 & \hat{f}_1 & \hat{f}_0 & \hat{f}_{-1} & 0 \\ 0 & 0 & \hat{f}_1 & \hat{f}_0 & \hat{f}_{-1} \\ 0 & 0 & 0 & \hat{f}_1 & \hat{f}_0 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix}.$$

The matrix $T_n(f)$ is the matrix resulting from discretizing the Laplacian with second order finite differences, that is,

$$-\Delta u = -\frac{\partial^2 u}{\partial x^2} \approx \frac{1}{h^2} T_n(f)\mathbf{u}_n = \frac{1}{h^2}\begin{bmatrix} 2u_1 - u_2 \\ -u_1 + 2u_2 - u_3 \\ \vdots \\ -u_{j-1} + 2u_j - u_{j+1} \\ \vdots \\ -u_{n-2} + 2u_{n-1} - u_n \\ -u_{n-1} + 2u_n \end{bmatrix},$$

where $h = 1/(n+1)$, $\mathbf{u}_n = [u_1, \ldots, u_n]^{\mathrm{T}}$, and we are assuming Dirichlet boundary conditions; for more details, see [40, Section 10.5.1].

If the matrix at hand is banded and Toeplitz, for example from a discretized PDE, and the matrix is assumed to be generated by a symbol, then, we can construct the symbol. Consider, the matrix

$$A_5 = \begin{bmatrix} -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} \hat{f}_0 & \hat{f}_{-1} & \hat{f}_{-2} & 0 & 0 \\ 0 & \hat{f}_0 & \hat{f}_{-1} & \hat{f}_{-2} & 0 \\ 0 & 0 & \hat{f}_0 & \hat{f}_{-1} & \hat{f}_{-2} \\ 0 & 0 & 0 & \hat{f}_0 & \hat{f}_{-1} \\ 0 & 0 & 0 & 0 & \hat{f}_0 \end{bmatrix}.$$

A symbol that generates this non-Hermitian matrix $A_5 = T_5(f)$, is constructed by using (1.2.2), that is,

$$f(\theta) = \sum_{\omega=-p}^{p} \hat{f}_\omega e^{\mathbf{i}\omega\theta} = -1 + 2e^{-\mathbf{i}\theta} - e^{-\mathbf{i}2\theta},$$

where $p = 2$, $\hat{f}_1 = \hat{f}_2 = 0$, $\hat{f}_0 = \hat{f}_{-2} = -1$, and $\hat{f}_{-1} = 2$. This is a symbol for the second order forward finite difference discretization of the Laplacian in one dimension, that is,

$$-\Delta u = -\frac{\partial^2 u}{\partial x^2} \approx \frac{1}{h^2} T_n(f)\mathbf{u}_n = \frac{1}{h^2}\begin{bmatrix} -u_1 + 2u_2 - u_3 \\ -u_2 + 2u_3 - u_4 \\ \vdots \\ -u_j + 2u_{j+1} - u_{j+2} \\ \vdots \\ -u_{n-1} + 2u_n \\ -u_n \end{bmatrix},$$

where again $h = 1/(n+1)$, $\mathbf{u}_n = [u_1, \ldots, u_n]^{\mathrm{T}}$, and we are assuming Dirichlet boundary conditions.

1.3 Spectrum of Toeplitz Matrices

As mentioned in Section 1.2, there is a close relation between the symbol and the spectrum of the generated Toeplitz matrices. We present here a brief timeline of the important historical results that led to the theory on which the current thesis is based.

Szegő's limit theorem [70], which dates back to 1920, states that for any real-valued function $f \in L^\infty(-\pi,\pi)$, the eigenvalues of $T_n(f)$, denoted by $\lambda_j\left(T_n(f)\right)$, $j=1,\ldots,n$, satisfy

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}F(\lambda_j(T_n(f)))=\frac{1}{2\pi}\int_{-\pi}^{\pi}F(f(\theta))\mathrm{d}\theta, \tag{1.3.1}$$

for every continuous function $F:\mathbb{R}\to\mathbb{C}$ with bounded support. The so-called Avram–Parter theorem [3, 57], which dates back to the 1980s, extends Szegő's limit theorem to complex-valued functions $f \in L^\infty(-\pi,\pi)$ and the singular values, denoted by $\sigma_j(T_n(f))$, $j=1,\ldots,n$. It states that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}F(\sigma_j(T_n(f)))=\frac{1}{2\pi}\int_{-\pi}^{\pi}F\left(|f(\theta)|\right)\mathrm{d}\theta, \tag{1.3.2}$$

for every continuous and compactly supported function $F:\mathbb{R}\to\mathbb{C}$. In the case where f is real-valued (and hence the matrices $T_n(f)$ are Hermitian), it can be shown that the Avram–Parter theorem is equivalent to Szegő's limit theorem. At the end of the 1990s, Tyrtyshnikov and Zamarashkin [77] proved, using matrix-theory arguments, that (1.3.1) holds for all real-valued $f \in L^1(-\pi,\pi)$ and (1.3.2) holds for all $f \in L^1(-\pi,\pi)$. In 1998, Tilli [74] generalized the theory to the multivariate and block cases, also allowing rectangular matrices. The case of preconditioned matrices was treated in [22, 63], also with reference to block structures [64, 65].

The need to consider matrices arising from the discretization of variable-coefficient PDEs, defined over generic Peano–Jordan measurable domains [47], and by means of either uniform or non-uniform meshes, led to the introduction of the theory of Generalized Locally Toeplitz sequences (GLT sequences) by Tilli and Serra-Capizzano [67, 68, 73]. The theory of GLT sequences, which extends the theory of Toeplitz matrices to include Toeplitz-like matrices in a broad sense, is the mathematical foundation on which this thesis is based. Here we only recall the essentials of the theory that are needed to understand the results presented herein. For the interested reader we refer to the more detailed introductions [16, 39] or to the comprehensive reviews [40, 41, 43].

Figure 1.3.1 presents a conceptual visualization of the results by Tilli [75]; how the eigenvalues of a matrix generated by a symbol f relate to the symbol. In the left panel we have a symbol $f:[-\pi,\pi]\to\mathbb{C}$. In the middle panel we see that the components A and B disconnects the complex plane. We have U_0, that is a unique unbounded connected component of $\mathbb{C}\backslash\mathscr{ER}(f)$, where $\mathscr{ER}(f)$ is

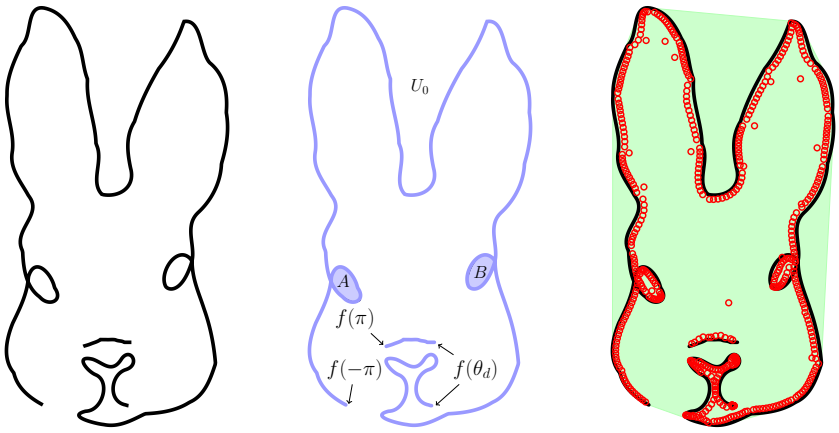


Figure 1.3.1. Visualization of Tilli's results [75] on the spectrum of the Toeplitz matrices generated by $f : [-\pi, \pi] \to \mathbb{C}$. Left: A discontinuous symbol $f \in L^\infty(-\pi, \pi)$. Middle: Symbol f is discontinuous at θ_d. The set $Z = \mathbb{C} \backslash U_0$, indicated in blue, is a cluster for the spectrum of $T_n(f)$. Right: The eigenvalues of $T_n(f)$, for $n = 500$. Each eigenvalue is represented by a red circle. The convex hull of $\mathscr{ER}(f)$ is indicated in green.

the essential range of f. Tilli proved that the set $Z = \mathbb{C} \backslash U_0 = \mathscr{ER}(f) \cup A \cup B$, indicated in blue, is a cluster for the spectrum of $T_n(f)$. This means that for every $\varepsilon > 0$ the number of eigenvalues of $T_n(f)$ which do not belong to the ε-neighborhood of Z is $o(n)$ as $n \to \infty$. In the right panel of Figure 1.3.1 is shown, as red circles, the eigenvalues of the matrix $T_n(f)$ generated by f for $n = 500$. As can be seen, a small number of the eigenvalues lie outside of Z, but still inside the convex hull of the essential range of f, $\mathscr{ER}(f)$, indicated in green, as stated by Brown and Halmos [19, Theorem 1.18].

If a symbol f is real-valued (and not constant), then any eigenvalue of $T_n(f)$ belongs to the open set (m_f, M_f) with m_f and M_f being the essential infimum and essential supremum of f, respectively. If $M_f > 0$ and f is nonnegative almost everywhere, then $T_n(f)$ is Hermitian positive definite.

1.4 Generalized Locally Toeplitz Sequences

Unlike the notion of Toeplitz matrices, the notion of GLT sequences does not apply to a single matrix. It is only an asymptotic concept that can be given for a sequence of matrices of increasing dimension. The original construction [67, 68, 73] is rather complex; it involves special diagonal matrix-sequences associated with given weight functions, Toeplitz sequences with their generating functions (or symbols), Kronecker products, and a notion of approximation theory for matrix-sequences [66], which induces a convergence in metric spaces

(a.c.s. convergence – convergence in measure [4, 5, 40]; approximating classes of sequences (a.c.s.) are defined in end of this section). Given all these ingredients, the definition is a construction made by several steps, including also topological closures with respect to the a.c.s. convergence.

In this thesis we do not report the original construction, owing to its intrinsic difficulty. Instead, we report in Table 1.4.1 a set of axioms [16, 39, 40] which identifies the GLT class and which can be viewed as an alternative, simpler definition of GLT sequences. Indeed, the axioms can be used in a constructive way, starting from the basic elements of the GLT class reported in axiom GLT 3 and using axioms GLT 4–GLT 8 as a way for constructing and defining the whole GLT class. To fully understand the GLT axioms of Table 1.4.1, we now introduce a few fundamental concepts. In what follows, for any $p \in [1, \infty]$ we denote by $\|\cdot\|_p$ the Schatten p-norm of matrices, which is defined as the p-norm of the vector of singular values [11]; in particular, $\|\cdot\|_\infty = \|\cdot\|$ is the classical spectral norm. Moreover, for any matrix A we denote by $A^\dagger$ the Moore–Penrose pseudoinverse of A [45].

Matrix-sequences. A matrix-sequence is a sequence of the form $\{A_n\}_n$, where $A_n \in \mathbb{C}^{n\times n}$ is a matrix of order n. The matrix-sequence $\{A_n\}_n$ is called Hermitian if each A_n is Hermitian.

Singular value and eigenvalue distribution of a matrix-sequence. We use the notation $\{A_n\}_n \sim_\sigma f$ to indicate that the sequence $\{A_n\}_n$ has a singular value distribution described by $f : D \subset \mathbb{R}^k \to \mathbb{C}$, a measurable function defined on a set D with $0 < \mu_k(D) < \infty$. This means that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} F(\sigma_j(A_n)) = \frac{1}{\mu_k(D)} \int_D F\left(|f(y_1, \ldots, y_k)|\right) \mathrm{d}y_1 \ldots \mathrm{d}y_k,$$

for all $F \in C_0(\mathbb{C})$. Here μ_k is the Lesbegue measure in $\mathbb{R}^k$ and $C_0(\mathbb{C})$ is the space of continuous complex-valued functions with bounded support.

Analogously, we use the notation $\{A_n\}_n \sim_\lambda f$ to indicate that the sequence $\{A_n\}_n$ has an eigenvalue distribution described by f, that is,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} F(\lambda_j(A_n)) = \frac{1}{\mu_k(D)} \int_D F\left(f(y_1, \ldots, y_k)\right) \mathrm{d}y_1 \ldots \mathrm{d}y_k,$$

for all $F \in C_0(\mathbb{C})$.

Zero-distributed sequences. A matrix-sequence $\{Z_n\}_n$ is referred to as a zero-distributed sequence when $\{Z_n\} \sim_\sigma 0$. If $\|Z_n\| \to 0$ then $\{Z_n\}_n \sim_\sigma 0$. For further properties of zero-distributed sequences, see [40, Section 3.4]. We mention that zero-distributed sequences appear in the approximation of integral operators or, more generally, in the approximation of compact operators; see for example [1, 34, 44] and [40, Section 10.4].

Diagonal sampling matrices. If $n \in \mathbb{N}$ and $a : [0,1] \to \mathbb{C}$, the diagonal sampling matrix $D_n(a)$ associated with a is the $n \times n$ diagonal matrix given by

$$(D_n(a))_{i,i} = a\,(i/n)\,, \qquad i = 1, \ldots, n.$$

Approximating classes of sequences. A fundamental concept on which the theory of GLT sequences is based. Let $\{A_n\}_n$ be a matrix-sequence, and let $\{\{B_{n,m}\}_n\}_m$ be a sequence of matrix-sequences. Then $\{\{B_{n,m}\}_n\}_m$ is an a.c.s. for $\{A_n\}_n$ if, for all sufficiently large m, the sequence $\{B_{n,m}\}_n$ approximates $\{A_n\}_n$ in the sense that A_n eventually is equal to $B_{n,m}$ plus a small rank matrix plus a small norm matrix. The notion of a.c.s. is a notion of convergence in the space of matrix-sequences and for this reason we use the notation $\{B_{n,m}\}_n \xrightarrow{\text{a.c.s.}} \{A_n\}_n$ to indicate that $\{\{B_{n,m}\}_n\}_m$ is an a.c.s. for $\{A_n\}_n$. For deeper definitions and discussions on a.c.s., see [4, 5] and [40, Chapter 5].

GLT sequences. A GLT sequence $\{A_n\}_n$ is a special matrix-sequence that is equipped with a measurable function $f : [0,1] \times [0,\pi] \to \mathbb{C}$, called the symbol (or kernel). The notation $\{A_n\}_n \sim_{\text{GLT}} f$ means that $\{A_n\}_n$ is a GLT sequence with symbol f.

In axioms GLT 1–GLT 8 of Table 1.4.1, we summarize the main properties of the GLT sequences (for further details, see [16, 40]).

1.5 Applying GLT Theory to Approximate the Spectrum

We start by considering the case of Toeplitz matrix-sequences generated by even trigonometric polynomials $f : [0,\pi] \to \mathbb{C}$. Afterwards, we discuss other cases where the theory of GLT sequences is applied in more generality.

If $\{T_n(f)\}_n \sim_\lambda f$, then the eigenvalues $\lambda_j(T_n(f))$ of a matrix $T_n(f)$ can be approximated by sampling the symbol $f(\theta)$, using a grid $\theta_{j,n}$, $j = 1, \ldots, n$, for large enough n. In particular, we have

$$\lambda_j(T_n(f)) = f(\theta_{j,n}) + E_{j,n}, \tag{1.5.1}$$

where $E_{j,n}$ are the errors of the approximations, typically of order $\mathcal{O}(h)$, where h depends on n. If we let $\theta_{j,n}$ be an equispaced grid in $[0,\pi]$, then $E_{j,n}$ tends to zero as $n \to \infty$, again of order $\mathcal{O}(h)$.

In Table 1.5.1 we list seven examples of uniform grids, with varying n. The general notation for a grid, where the type is defined by context, is $\theta_{j,n}$, where n is the number of grid points, and j are the indices $j \in \{1, \ldots, n\}$. We can also define a set of indices $j_{\mathcal{S}} \subseteq \{1, \ldots, n\}$, and then $\theta_{j_{\mathcal{S}},n}$ means $\theta_{j,n}$ for indices $j \in j_{\mathcal{S}}$. The grid fineness parameter h for the respective grids is also presented in Table 1.5.1. The names of the different grids are chosen in view of their relations with the τ-algebras [15, see (19), (22), and (23)]. Note that the τ-algebras are closed under inversion in the sense that if an invertible

Table 1.4.1. GLT axioms.

GLT 1. If $\{A_n\}_n \sim_{\text{GLT}} f$ then $\{A_n\}_n \sim_\sigma f$. If, moreover, the matrices A_n are Hermitian, then $\{A_n\}_n \sim_\lambda f$.

GLT 2. If $\{A_n\}_n \sim_{\text{GLT}} f$ and $\{A_n\}_n$ is quasi-Hermitian, that is, $A_n = X_n + Y_n$, where
- $\|X_n\|, \|Y_n\| \leq C$ for some constant C independent of n,
- every X_n is Hermitian,
- $\lim_{n\to\infty} \frac{\|Y_n\|_1}{n} = 0$,

then $\{A_n\}_n \sim_\lambda f$.

GLT 3. We have
- $\{T_n(f)\}_n \sim_{\text{GLT}} f(x,\theta) = f(\theta)$ if $f \in L^1(-\pi,\pi)$,
- $\{D_n(a)\}_n \sim_{\text{GLT}} f(x,\theta) = a(x)$ if $a : [0,1] \to \mathbb{C}$ is continuous almost everywhere,
- $\{Z_n\}_n \sim_{\text{GLT}} f(x,\theta) = 0$ if and only if $\{Z_n\}_n \sim_\sigma 0$.

GLT 4. If $\{A_n\}_n \sim_{\text{GLT}} f$ then $\{A_n^*\}_n \sim_{\text{GLT}} \bar{f}$, where A_n^* is the conjugate transpose of A_n.

GLT 5. If $\{A_n\}_n \sim_{\text{GLT}} f$ and $\{B_n\}_n \sim_{\text{GLT}} g$, where A_n and B_n have the same size, then
- $\{\alpha A_n + \beta B_n\}_n \sim_{\text{GLT}} \alpha f + \beta g$, for all $\alpha, \beta \in \mathbb{C}$,
- $\{A_n B_n\}_n \sim_{\text{GLT}} fg$.

GLT 6. If $\{A_n\}_n \sim_{\text{GLT}} f$ and $f \neq 0$ a.e. then $\{A_n^\dagger\}_n \sim_{\text{GLT}} f^{-1}$.

GLT 7. If $\{A_n\}_n \sim_{\text{GLT}} f$ and each A_n is Hermitian, then $\{F(A_n)\}_n \sim_{\text{GLT}} F(f)$ for every continuous function $F : \mathbb{C} \to \mathbb{C}$.

GLT 8. $\{A_n\}_n \sim_{\text{GLT}} f$ if and only if there exist GLT sequences $\{B_{n,m}\}_n \sim_{\text{GLT}} f_m$ such that $\{B_{n,m}\}_n \xrightarrow{\text{a.c.s.}} \{A_n\}_n$ and $f_m \to f$ in measure over $[0,1] \times [-\pi,\pi]$.

matrix A_n belongs to a τ-algebra, then the inverse A_n^{-1} belongs to the same τ-algebra; this is a consequence of the fact that all τ-algebras are closed under linear combinations and matrix-matrix multiplication. Unless otherwise stated, the τ_n-grid is the default choice in this thesis.

Hence, it has been known for a long time that for some special matrices, sampling the symbol with the appropriate grid gives the exact eigenvalues. Now consider a tridiagonal Toeplitz matrix, that is, $\hat{f}_0, \hat{f}_1, \hat{f}_{-1} \in \mathbb{C}$ are the only non-zero components in the generic central row of the matrix. This matrix belongs to the τ-algebra [15], and the symbol is

$$f(\theta) = \hat{f}_0 + \hat{f}_1 e^{\mathbf{i}\theta} + \hat{f}_{-1} e^{-\mathbf{i}\theta}.$$

Table 1.5.1. Seven examples of uniform grids. Typically the τ_n-grid is the default choice, unless other grids provide more accurate, or even exact, eigenvalues when sampling the symbol.

Name	Grid	j	h	Description
τ_n	$j\pi/(n+1)$	$1,\ldots,n$	$1/(n+1)$	$\tau_n(0,0)$
τ_{n-1}	$j\pi/n$	$1,\ldots,n-1$	$1/n$	$\tau_{n-1}(0,0)$
τ_{n-2}	$j\pi/(n-1)$	$1,\ldots,n-2$	$1/(n-1)$	$\tau_{n-2}(0,0)$
τ_{n-1}^{0}	$(j-1)\pi/n$	$1,\ldots,n$	$1/n$	$\tau_n(1,1) = 0 \cup \tau_{n-1}(0,0)$
τ_{n-1}^{π}	$j\pi/n$	$1,\ldots,n$	$1/n$	$\tau_n(-1,-1) = \tau_{n-1}(0,0) \cup \pi$
$\tau_{n-2}^{0,\pi}$	$(j-1)\pi/(n-1)$	$1,\ldots,n$	$1/(n-1)$	$0 \cup \tau_{n-2}(0,0) \cup \pi$
$\tau_{n-1}^{0,\pi}$	$(j-1)\pi/n$	$1,\ldots,n+1$	$1/n$	$0 \cup \tau_{n-1}(0,0) \cup \pi$

The eigenvalues of the generated matrix $T_n(f)$ are exactly given by sampling a "symmetrized" symbol g, for which $T_n(f) \sim T_n(g)$, that is,

$$\lambda_j(T_n(f)) = g(\theta_{j,n}) = \hat{f}_0 + 2\sqrt{\hat{f}_1\hat{f}_{-1}}\cos(\theta_{j,n}), \tag{1.5.2}$$

where $\theta_{j,n}$ is the τ_n-grid in Table 1.5.1; see [18, 35, 55]. The eigenvector that corresponds to the eigenvalue $\lambda_j(T_n(f))$ is

$$\mathbf{x}_{j,n} = \left[x_1^{(j,n)}, \ldots, x_k^{(j,n)}, \ldots, x_n^{(j,n)}\right]^{\mathrm{T}}, \tag{1.5.3}$$

where

$$x_k^{(j,n)} = \left(\sqrt{\hat{f}_1/\hat{f}_{-1}}\right)^k \sin\left(k\theta_{j,n}\right). \tag{1.5.4}$$

Remark 1.5.1. It is important to note that $\sqrt{\hat{f}_1\hat{f}_{-1}}$ and $\sqrt{\hat{f}_1/\hat{f}_{-1}}$ in (1.5.2) and (1.5.4) should be interpreted as follows: We have in exponential form $\hat{f}_1 = \left|\hat{f}_1\right| e^{\mathbf{i}\phi_1}$ and $\hat{f}_{-1} = \left|\hat{f}_{-1}\right| e^{\mathbf{i}\phi_{-1}}$, for $\phi_1, \phi_{-1} \in [0, 2\pi]$, and we set

$$\sqrt{\hat{f}_1\hat{f}_{-1}} = \sqrt{\hat{f}_1}\sqrt{\hat{f}_{-1}} = \sqrt{\left|\hat{f}_1\right|\left|\hat{f}_{-1}\right|}e^{\mathbf{i}(\phi_1+\phi_{-1})/2},$$
$$\sqrt{\hat{f}_1/\hat{f}_{-1}} = \sqrt{\hat{f}_1}/\sqrt{\hat{f}_{-1}} = \sqrt{\left|\hat{f}_1\right|/\left|\hat{f}_{-1}\right|}e^{\mathbf{i}(\phi_1-\phi_{-1})/2}.$$

For instance, for the case $\hat{f}_1 = \hat{f}_{-1} = -1 = e^{\mathbf{i}\pi}$, we have

$$\sqrt{(-1)(-1)} = \sqrt{-1}\sqrt{-1} = \mathbf{i}\cdot\mathbf{i} = -1,$$
$$\sqrt{(-1)/(-1)} = \sqrt{-1}/\sqrt{-1} = \mathbf{i}/\mathbf{i} = 1.$$

We now present a few examples to illustrate the type of problems that this thesis aims to tackle.

Example 1.5.1. We return to Example 1.2.1 and show that through equations (1.5.2), (1.5.3), and (1.5.4) we can exactly compute the eigenvalues and eigenvectors of $T_n(f)$. Recall that in Example 1.2.1 we have $n = 5$ and $f(\theta) = 2 - 2\cos(\theta)$. From (1.5.2) we infer that $g(\theta) = f(\theta)$, and the τ_n-grid is $\theta_{j,5} = j\pi/6$, for $j = 1, \ldots, 5$. Table 1.5.2 exhibits the eigenvalues and the corresponding eigenvectors for $T_5(f)$; and, indeed, the equation $T_5(f)\mathbf{x}_{j,5} = \lambda_j(T_5(f))\mathbf{x}_{j,5}$ is true for all $j = 1, \ldots, 5$.

Table 1.5.2. Example 1.5.1: The eigenvalues, and corresponding eigenvectors, for the generated Toeplitz matrix $T_5(f)$, where the symbol is $f(\theta) = 2 - 2\cos(\theta)$.

j	$\lambda_j(T_5(f))$	$\mathbf{x}_{j,5}$
1	$2 - 2\cos(\pi h)$	$[\sin(\pi h), \sin(2\pi h), \sin(3\pi h), \sin(4\pi h), \sin(5\pi h)]^{\mathrm{T}}$
2	$2 - 2\cos(2\pi h)$	$[\sin(2\pi h), \sin(4\pi h), \sin(6\pi h), \sin(8\pi h), \sin(10\pi h)]^{\mathrm{T}}$
3	$2 - 2\cos(3\pi h)$	$[\sin(3\pi h), \sin(6\pi h), \sin(9\pi h), \sin(12\pi h), \sin(15\pi h)]^{\mathrm{T}}$
4	$2 - 2\cos(4\pi h)$	$[\sin(4\pi h), \sin(8\pi h), \sin(12\pi h), \sin(16\pi h), \sin(20\pi h)]^{\mathrm{T}}$
5	$2 - 2\cos(5\pi h)$	$[\sin(5\pi h), \sin(10\pi h), \sin(15\pi h), \sin(20\pi h), \sin(25\pi h)]^{\mathrm{T}}$

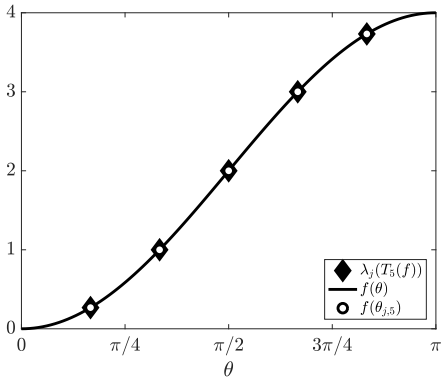


Figure 1.5.1. Example 1.5.1: The symbol $f(\theta) = 2 - 2\cos(\theta)$ shown for $\theta \in [0, \pi]$ (black line). The eigenvalues $\lambda_j(T_5(f))$ (black diamonds), and the samplings of $f(\theta)$ with the τ_5-grid (white circles) overlap exactly.

From axiom GLT 1, and the fact that all matrices $T_n(f)$ are Hermitian, we deduce that this is an example of a matrix-sequence satisfying both $\{T_n(f)\}_n \sim_{\text{GLT}} f$ and $\{T_n(f)\}_n \sim_{\sigma,\lambda} f$. Additionally, since we know that the matrix $T_n(f)$ belongs to the τ-algebra, described in [15] and originally studied in [12], we do know the grid that gives the exact eigenvalues when sampling the symbol.

We now construct an example of a matrix $\tilde{T}_n(f)$, belonging to the $\tau_n(1,1)$-algebra, for which the grid $\tau_n(1,1) = \tau^0_{n-1}$ gives the exact eigenvalues, when sampling f. Define $\tilde{T}_n(f) = T_n(f) + R_n$, where $T_n(f)$ is the Toeplitz matrix generated by f, and R_n is a low rank correction. R_n is a matrix with all zeros, except for two elements, namely, $(R_n)_{1,1} = \hat{f}_1 = -1$ and $(R_n)_{n,n} = \hat{f}_1 = -1$.

By the same construction, but with $\tilde{T}_n(f)$ belonging to the $\tau_n(-1,-1)$-algebra, we have that the grid $\tau_n(-1,-1) = \tau^{\pi}_{n-1}$ gives the exact eigenvalues when sampling f. Here we have instead $\tilde{T}_n(f) = T_n(f) - R_n$, with R_n defined as before.

Example 1.5.2. In contrast to Examples 1.2.1 and 1.5.1, we now focus on a case where a tridiagonal matrix $T_n(f)$ is not Hermitian. We construct an example where

$$\{T_n(f)\}_n \sim_{\text{GLT}} f, \qquad \{T_n(f)\}_n \sim_{\sigma} f, \qquad \{T_n(f)\}_n \not\sim_{\lambda} f.$$

In other words, the sequence $\{T_n(f)\}_n$ is a GLT sequence and f describes its singular value distribution. However, as $n \to \infty$ the errors $E_{j,n} = \lambda_j(T_n(f)) - f(\theta_{j,n})$ does not tend to zero, since f does not describe the eigenvalue distribution of $\{T_n(f)\}_n$; see Figure 1.5.2. Consider the symbol

$$f(\theta) = 2 - e^{\mathbf{i}\theta} + 2\mathbf{i}e^{-\mathbf{i}\theta},$$

and the generated Toeplitz matrix of order $n = 5$,

$$T_5(f) = \begin{bmatrix} 2 & 2\mathbf{i} & 0 & 0 & 0 \\ -1 & 2 & 2\mathbf{i} & 0 & 0 \\ 0 & -1 & 2 & 2\mathbf{i} & 0 \\ 0 & 0 & -1 & 2 & 2\mathbf{i} \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix}.$$

From (1.5.2) we have that a symbol $g(\theta)$ is such that $T_n(f) \sim T_n(g)$; that is, the nth Toeplitz matrix generated by f shares the same eigenvalues as the nth Toeplitz matrix generated by g. In this example we have

$$g(\theta) = 2 + 2\sqrt{2}e^{\mathbf{i}3\pi/4}\cos(\theta),$$

and the generated Toeplitz matrix of order $n = 5$,

$$T_5(g) = \begin{bmatrix} 2 & \sqrt{2}e^{\mathbf{i}3\pi/4} & 0 & 0 & 0 \\ \sqrt{2}e^{\mathbf{i}3\pi/4} & 2 & \sqrt{2}e^{\mathbf{i}3\pi/4} & 0 & 0 \\ 0 & \sqrt{2}e^{\mathbf{i}3\pi/4} & 2 & \sqrt{2}e^{\mathbf{i}3\pi/4} & 0 \\ 0 & 0 & \sqrt{2}e^{\mathbf{i}3\pi/4} & 2 & \sqrt{2}e^{\mathbf{i}3\pi/4} \\ 0 & 0 & 0 & \sqrt{2}e^{\mathbf{i}3\pi/4} & 2 \end{bmatrix}.$$

Note that $T_n(g)$ is a symmetrization of $T_n(f)$, in the sense that there is an invertible diagonal matrix D_n such that $T_n(g) = D_n T_n(f) D_n^{-1}$, see [58]. With $\hat{f}_1\hat{f}_{-1} \neq 0$ and $n = 5$, a possible D_5 is given by

$$D_5 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 & 0 \\ 0 & 0 & \gamma^2 & 0 & 0 \\ 0 & 0 & 0 & \gamma^3 & 0 \\ 0 & 0 & 0 & 0 & \gamma^4 \end{bmatrix},$$

where $\gamma = \sqrt{\hat{f}_{-1}/\hat{f}_1} = \sqrt{2\mathbf{i}}/\sqrt{-1} = 1-\mathbf{i}$. To conclude, we have the relations $\{T_n(f)\}_n \sim_{\text{GLT},\sigma} f$, $\{T_n(f)\}_n \not\sim_\lambda f$, $\{T_n(f)\}_n \sim_\lambda g$, and $\{T_n(g)\}_n \sim_{\text{GLT},\sigma,\lambda} g$.

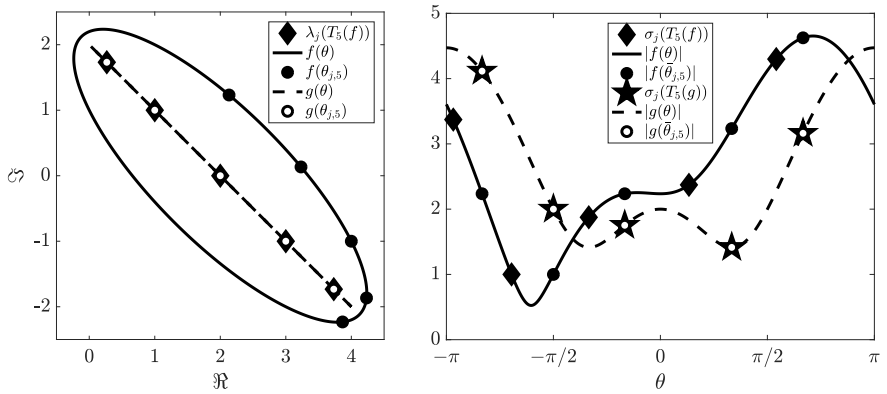


Figure 1.5.2. Example 1.5.2: Left: Eigenvalue distribution. Symbols f (black line) and g (dashed line) drawn on the complex plane. Sampling g with τ_5-grid gives eigenvalues of $T_5(f) \sim T_5(g)$. Right: Singular value distribution. Absolute values of the symbols $|f(\theta)|$ (black line) and $|g(\theta)|$ (dashed line) for $\theta \in [-\pi, \pi]$.

In the left panel of Figure 1.5.2, we illustrate the eigenvalue distribution of the generated matrix $T_n(f)$ (black diamonds). We display the two symbols $f : [-\pi, \pi] \to \mathbb{C}$ (black line) and $g : [-\pi, \pi] \to \mathbb{C}$ (dashed line). For any n, the eigenvalues of $T_n(f)$ and $T_n(g)$ lie on the dashed line. Sampling f with the τ_5-grid (black circles) gives bad approximations of the eigenvalues. Sampling g with the τ_5-grid (white circles) gives the exact eigenvalues, both for $T_5(f)$ and $T_5(g)$ which are similar matrices. In the right panel of Figure 1.5.2, we show the singular value distribution of $T_n(f)$ (black diamonds) and $T_n(g)$ (black stars). Absolute values of the symbols $|f(\theta)|$ (black line) and $|g(\theta)|$ (dashed line) are shown for $\theta \in [-\pi, \pi]$. Since $|f|$ and $|g|$ are non-monotone on $\theta \in [-\pi, \pi]$, the correct sampling grid is not uniquely defined. Three grid-points of the standard τ_5-grid have been shifted to $[-\pi, 0]$, to define a new sampling grid $\bar{\theta}_{j,5}$. The samplings $|f(\bar{\theta}_{j,5})|$ (black circles) do not give an exact approximation of $\sigma_j(T_5(f))$. The matrix $T_5(g)$ is normal ($A^*A = AA^*$), and thus $\sigma_j(T_5(g))$ are equal to the moduli of the eigenvalues, $|\lambda_j(T_5(g))|$. Both the grid $\bar{\theta}_{j,n}$ and the τ_5-grid gives the exact singular values of $T_5(g)$ when sampling $|g(\theta)|$ (white circles is $|g(\bar{\theta}_{j,n})|$).

Example 1.5.3. In this example we extend Example 1.2.1, in the sense that we allow variable coefficients, that is, the matrix we consider does not have constant diagonals. Consider the differential equation,

$$\begin{cases} -(a(x)u'(x))' = b(x), & x \in (0,1), \\ u(x) = 0, & x \in \{0,1\}, \end{cases} \tag{1.5.5}$$

for some $a(x)$ and $b(x)$. Note that if we have $a(x) = 1$, then we get $-u''(x)$ on the left hand side and we are then in the case of Example 1.2.1. When discretizing (1.5.5) we have

$$-(a(x)u'(x))' = -a'(x)u'(x) - a(x)u''(x) \approx A_n\mathbf{u}_n = \mathbf{b}_n,$$

where, by using second order finite differences,

$$A_n = \begin{bmatrix} a_{1/2}+a_{3/2} & -a_{3/2} & & & & \\ -a_{3/2} & a_{3/2}+a_{5/2} & -a_{5/2} & & & \\ & -a_{5/2} & \ddots & \ddots & & \\ & & \ddots & \ddots & -a_{n-1/2} & \\ & & & -a_{n-1/2} & a_{n-1/2}+a_{n+1/2} \end{bmatrix},$$

with a generic central row given by

$$(A_n)_i = \left[0, \ldots, 0, -a_{i-1/2}, a_{i-1/2} + a_{i+1/2}, -a_{i+1/2}, 0, \ldots, 0\right],$$

where $a_i = a(x_i)$, $x_i = ih$, $i = 0, \ldots, n+1$, $h = 1/(n+1)$, $b_i = b(x_i)$, $\mathbf{u}_n = [u_1, \ldots, u_n]^{\mathrm{T}}$, and $\mathbf{b}_n = [b_1, \ldots, b_n]^{\mathrm{T}}$.

Now we show that $\{A_n\}_n$ is a particular non-trivial example of a GLT sequence. Let $D_n(a)$ be the diagonal matrix generated by $a(x)$, as defined in axiom GLT 3, and let $T_n(f)$ be the Toeplitz matrix generated by $f(\theta) = 2 - 2\cos(\theta)$. We have

$$D_n(a) = \begin{bmatrix} a_1 & & & \\ & a_2 & & \\ & & \ddots & \\ & & & a_n \end{bmatrix}, \quad T_n(f) = \begin{bmatrix} 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{bmatrix},$$

and thus

$$D_n(a)T_n(f) = \begin{bmatrix} 2a_1 & -a_1 & & & \\ -a_2 & 2a_2 & -a_2 & & \\ & -a_3 & \ddots & \ddots & \\ & & \ddots & \ddots & -a_{n-1} \\ & & & -a_n & 2a_n \end{bmatrix}.$$

We observe that $A_n = D_n(a)T_n(2 - 2\cos(\theta)) + E_n$ where,

$$(E_n)_{i,j} = \begin{cases} a_{(2i-1)/2} - 2a_{2i/2} + a_{(2i+1)/2}, & i = j, i = 1, \ldots, n, \\ a_{2i/2} - a_{(2i-1)/2}, & i = j+1, i = 2, \ldots, n, \\ a_{2i/2} - a_{(2i+1)/2}, & i = j-1, i = 1, \ldots, n-1, \\ 0, & \text{otherwise.} \end{cases}$$

Under the assumption that $a(x)$ is continuous almost everywhere, the norm of the matrix E_n is such that $\|E_n\| \to 0$ when $n \to \infty$; see Figure 1.5.3. As a consequence $\{E_n\}_n \sim_\sigma 0$, that is, $\{E_n\}_n$ is zero-distributed, and hence $\{E_n\}_n \sim_{\text{GLT}} 0$ by axiom GLT 3. Since $\{D_n(a)\}_n \sim_{\text{GLT}} a(x)$, $\{T_n(f)\}_n \sim_{\text{GLT}} f(\theta)$, $\{E_n\}_n \sim_{\text{GLT}} 0$ by axioms GLT 2 and GLT 5, we conclude that $\{A_n\}_n \sim_{\text{GLT}} a(x)(2-2\cos(\theta))$. Therefore, by axiom GLT 1 and by the fact that A_n is real symmetric, we infer that $\{A_n\}_n \sim_{\sigma,\lambda} a(x)(2-2\cos(\theta))$.
As a practical example, consider (1.5.5) with a discontinuous $a(x)$ defined by

$$a(x) = \begin{cases} 2 + x/2, & x \in [0, 1/3), \\ e^{\pi x/2}, & x \in [1/3, 2/3), \\ 2 + \cos(3x), & x \in [2/3, 1]. \end{cases} \tag{1.5.6}$$

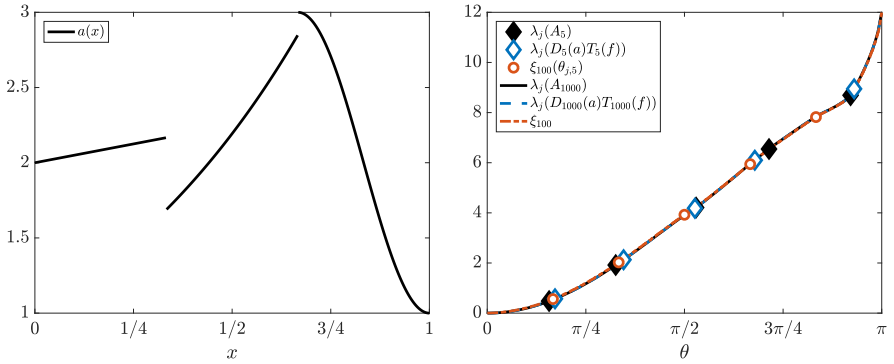


Figure 1.5.3. Example 1.5.3. Left: Discontinuous $a(x)$ in (1.5.5), defined in (1.5.6). Right: The spectrum of A_n and $D_n(a)T_n(f)$ for $n = \{5, 1000\}$. Also shown the samplings ξ_r, for $r = 100$, and the interpolated samplings $\xi_r(\theta_{j,5})$.

The left panel of Figure 1.5.3 shows the graph of the function $a(x)$. In the right panel of Figure 1.5.3, we show the eigenvalues of the matrices A_5 (black diamonds) and $D_5(a)T_5(f)$ (white diamonds with blue edge). Also shown is an approximation of the eigenvalues (white circles with red edges), by an interpolated sampling of a vector ξ_{100} of length $100^2 = 10000$, using the τ_5-grid. The vector ξ_r is obtained in the following way: for a chosen r, define the grid

$$\mathcal{G}_r = \{(i/n, j/(n+1)); i, j = 1, \ldots, r\},$$

and then compute all the samples of $a(x)f(\theta)$ at the points $(x, \theta) \in \mathcal{G}_r$. These r^2 samples are stored, after being sorted in non-decreasing order, in a vector ξ_r. We assume that the vector ξ_r contains samples of some function $\xi(\theta)$ and in the limit, as $r \to \infty$, ξ_r describes the spectrum of $D_n(a)T_n(f)$. However, we do not know $\xi(\theta)$ analytically, so we cannot sample $\xi(\theta)$ with a grid $\theta_{j,n}$, but we

can interpolate the data in the vector ξ_r to the grid points $\theta_{j,n}$. In the right panel of Figure 1.5.3, it is clear that for a large $n = 1000$, the eigenvalues of the two matrices A_n and $D_n(a)T_n(f)$ overlap well with ξ_{100}, over the whole spectrum.
Various problems, with variable coefficients and discretized with isogeometric analysis (IgA), are studied in [32]. For further discussion on variable coefficients and quite general choices of differential operators and approximation methods, we refer to [40, 41, 43], and the references therein.

Example 1.5.4. The symbol can also be matrix-valued, that is, each sampling returns a matrix. In what follows, matrix-valued symbols are denoted in bold, for example $\mathbf{f}^{(s)}(\theta)$, and $s \times s$ is the size of the matrix returned for each sampling. The size of a generated matrix $T_n(\mathbf{f}^{(s)})$ is $N \times N$, where $N = ns$. The case $s = 1$, is the scalar case discussed in previous examples. This type of generated matrices are called block Toeplitz matrices, that is, we generate a Toeplitz matrix, where the constant diagonal elements, instead of scalars, are matrices (or blocks) of order s.

We now show how to view a single Toeplitz matrix, namely, $T_6(f)$ with symbol $f(\theta) = 2 - 2\cos(\theta)$, in several different ways depending on the block structure we choose. Consequently, we have different generating functions and symbols depending on our choice. We have

$$T_6(f) = \begin{bmatrix} 2 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & -1 & 2 \end{bmatrix} = T_6(\mathbf{f}^{(1)}),$$

where we have chosen the scalar case $s = 1$, and thus,

$$\mathbf{f}^{(1)}(\theta) = \underbrace{\begin{bmatrix} 2 \end{bmatrix}}_{\hat{\mathbf{f}}_0^{(1)}} + \underbrace{\begin{bmatrix} -1 \end{bmatrix}}_{\hat{\mathbf{f}}_1^{(1)}} e^{\mathbf{i}\theta} + \underbrace{\begin{bmatrix} -1 \end{bmatrix}}_{\hat{\mathbf{f}}_{-1}^{(1)}} e^{-\mathbf{i}\theta} = \begin{bmatrix} 2 \end{bmatrix} + 2 \begin{bmatrix} -1 \end{bmatrix} \cos(\theta).$$

We now view the same matrix as a block Toeplitz matrix, with blocks of order $s = 2$,

$$T_3(\mathbf{f}^{(2)}) = \begin{bmatrix} 2 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 6 & 0 & 0 & -1 & 2 & -1 \\ 6 & 0 & 0 & 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{f}}_0^{(2)} & \hat{\mathbf{f}}_{-1}^{(2)} & \mathbf{0} \\ \hat{\mathbf{f}}_1^{(2)} & \hat{\mathbf{f}}_0^{(2)} & \hat{\mathbf{f}}_{-1}^{(2)} \\ \mathbf{0} & \hat{\mathbf{f}}_1^{(2)} & \hat{\mathbf{f}}_0^{(2)} \end{bmatrix},$$

where the matrix-valued symbol is

$$\mathbf{f}^{(2)}(\theta) = \underbrace{\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}}_{\hat{\mathbf{f}}_0^{(2)}} + \underbrace{\begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix}}_{\hat{\mathbf{f}}_1^{(2)}} e^{\mathbf{i}\theta} + \underbrace{\begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}}_{\hat{\mathbf{f}}_{-1}^{(2)}} e^{-\mathbf{i}\theta}.$$

Note that $\hat{\mathbf{f}}_1^{(2)} = (\hat{\mathbf{f}}_{-1}^{(2)})^{\mathrm{T}}$. Thus, $T_6(f) = T_3(\mathbf{f}^{(2)})$. In general, $T_N(f) = T_n(\mathbf{f}^{(s)})$, where $N = ns$ is the size of the matrix, n is the number of blocks, each of order s. So, to sample the matrix-valued symbol to get the eigenvalues of $T_3(\mathbf{f}^{(2)})$, we here would use the grid

$$\theta_{j,n} = 2\theta_{j,2n} = 2\frac{j\pi}{2n+1} = \frac{j\pi}{n+1/2}, \quad j = 1, \ldots, n,$$

to get the exact eigenvalues when sampling $\mathbf{f}^{(2)}(\theta)$. Each sampling yields a local 2×2 matrix, for which we need to solve the eigenvalue problem, and get two eigenvalues.

Now choose yet another block structure of the matrix Toeplitz $T_6(f)$, consisting of blocks of order $s = 3$, with a new symbol $\mathbf{f}^{(3)}$,

$$T_2(\mathbf{f}^{(3)}) = \begin{bmatrix} 2 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{f}}_0^{(3)} & \hat{\mathbf{f}}_{-1}^{(3)} \\ \hat{\mathbf{f}}_1^{(3)} & \hat{\mathbf{f}}_0^{(3)} \end{bmatrix},$$

that is, the symbol for this chosen block size of 3×3 is

$$\mathbf{f}^{(3)}(\theta) = \underbrace{\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}}_{\hat{\mathbf{f}}_0^{(3)}} + \underbrace{\begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\hat{\mathbf{f}}_1^{(3)}} e^{\mathbf{i}\theta} + \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}}_{\hat{\mathbf{f}}_{-1}^{(3)}} e^{-\mathbf{i}\theta}.$$

The grid that gives the exact eigenvalues, for a general symbol $\mathbf{f}^{(s)}$ with blocks of order s, for the given tridiagonal structure is

$$\theta_{j,n} = s\theta_{j,N} = \frac{j\pi}{n+1/s}, \quad j = 1, \ldots, n.$$

In Figure 1.5.4 we show the eigenfunctions for the three different symbols $f = \mathbf{f}^{(1)}$, $\mathbf{f}^{(2)}$, and $\mathbf{f}^{(3)}$, previously defined in the current example. Each eigenfunction is denoted as $\mathbf{f}_q^{(s)}$, where s is the order of the blocks of the symbol, and $q = 1, \ldots, s$ are the numbers of the eigenfunctions. As expected, the three symbols describe the same spectrum as f, but split into different eigenfunctions for $s > 1$.

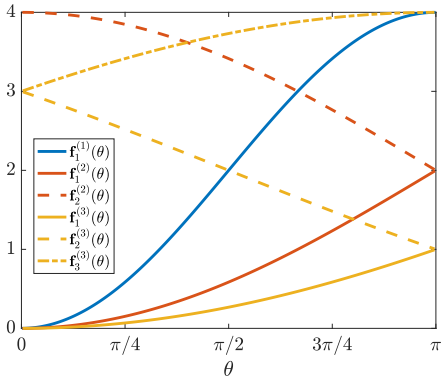


Figure 1.5.4. Example 1.5.4. The eigenfunctions $\mathbf{f}_q^{(s)}$, $q = 1, \ldots, s$, of the three symbols $\mathbf{f}^{(1)}$, $\mathbf{f}^{(2)}$, and $\mathbf{f}^{(3)}$ describe the same spectrum.

We can also get exact information on the spectrum for a Toeplitz-like block matrix that is not a full block matrix in the sense that its size is not $ns \times ns$. Consider the matrix-valued symbol $\mathbf{f}^{(2)}(\theta)$, defined earlier in this example, and generate the matrix $T_n(\mathbf{f}^{(2)})$. Then remove the last row and column, so as to obtain a matrix A_{N-1}. To get the exact eigenvalues, do the following procedure: (a) Sample $\mathbf{f}^{(2)}(\theta)$ with the τ_{n-1}^{π}-grid, defined in Table 1.5.1. This results in $2n$ samplings, but the matrix A_{N-1} is of order $N-1 = 2n-1$. (b) Discard one of the two (equal) samplings from $\mathbf{f}^{(2)}(\theta_{n,n})$. Hence, we have used the different grids τ_{n-1} and τ_{n-1}^{π} on the two eigenfunctions $\mathbf{f}_1^{(2)}$ and $\mathbf{f}_2^{(2)}$.
Note also that in general, we do not know what grid we should use to get exact eigenvalues, and then we often use the standard τ_n-grid, unless we can find an "optimal" grid, better suited for the problem under consideration.

We now consider a more applicable example of a matrix-valued symbol. In [31] we extend the results of this thesis to matrix-valued symbols, but we also present matrices, where we get the exact eigenvalues by sampling the symbol with a special grid. Take the stiffness matrix constructed by discretizing a second order elliptic differential problem by the $\mathbb{Q}_{\boldsymbol{p}}$ Lagrangian finite element method; for details see [42]. For $n = 3$ and $p = 3$ we get the following normalized stiffness matrix

$$K_3^{(3)} = \frac{1}{40}\left[\begin{array}{rrr|rrr|rr} 432 & -297 & 54 & 0 & 0 & 0 & 0 & 0 \\ -297 & 432 & -189 & 0 & 0 & 0 & 0 & 0 \\ 54 & -189 & 296 & -189 & 54 & -13 & 0 & 0 \\ \hline 0 & 0 & -189 & 432 & -297 & 54 & 0 & 0 \\ 0 & 0 & 54 & -297 & 432 & -189 & 0 & 0 \\ 0 & 0 & -13 & 54 & -189 & 296 & -189 & 54 \\ \hline 0 & 0 & 0 & 0 & 0 & -189 & 432 & -297 \\ 0 & 0 & 0 & 0 & 0 & 54 & -297 & 432 \end{array}\right].$$

The natural block structure for the matrix $K_3^{(3)}$ is shown in different colors, and the associated symbol is

$$\mathbf{f}^{(3)}(\theta) = \begin{bmatrix} 432 & -297 & 54 \\ -297 & 432 & -189 \\ 54 & -189 & 296 \end{bmatrix} + \begin{bmatrix} 0 & 0 & -189 \\ 0 & 0 & 54 \\ 0 & 0 & -13 \end{bmatrix} e^{\mathbf{i}\theta} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -189 & 54 & -13 \end{bmatrix} e^{-\mathbf{i}\theta}.$$

As is seen, the matrix $K_3^{(3)}$ is not a full block matrix; the last row and column has been removed due to boundary conditions. Using a special grid we can obtain the exact eigenvalues of the matrix $K_3^{(3)}$ when sampling $\mathbf{f}^{(3)}(\theta)$; for details see [31].

For further discussion on matrix-valued symbols we refer to [30, 43] and the references therein.

Example 1.5.5. Symbols that have multiple arguments, such as $f(\theta_1, \theta_2, \ldots, \theta_k)$, are called multivariate. Multivariate symbols generate Toeplitz matrices whose elements themselves are Toeplitz matrices. These structures are referred to as multilevel Toeplitz matrices [77]. Consider the bi-variate symbol

$$\begin{aligned} f(\theta_1, \theta_2) &= 6 - 6\cos(\theta_1) - 4\cos(\theta_2) - 4\cos(\theta_1)\cos(\theta_2) \\ &= 2(3 - 2\cos(\theta_2)) - 2\cos(\theta_1)(3 + 2\cos(\theta_2)), \end{aligned}$$

where, by convention, θ_1 is the "outer" and θ_2 is the "inner" variable. The generated matrix from this symbol has a tensor structure. We denote by $\mathbb{I}_n^{(\omega)}$ the matrix of size n with ones on the diagonal ω and zeros elsewhere, that is, the identity matrix of order n is $\mathbb{I}_n^{(0)}$. We can construct the matrix $T_{n_1,n_2}(f)$ as follows, where the Kronecker product is denoted by $\otimes$

$$\begin{aligned} A_{n_2} &= T_{n_2}(3 - 2\cos(\theta_2)) = 3\mathbb{I}_{n_2}^{(0)} - \left(\mathbb{I}_{n_2}^{(1)} + \mathbb{I}_{n_2}^{(-1)}\right), \\ B_{n_2} &= T_{n_2}(3 + 2\cos(\theta_2)) = 3\mathbb{I}_{n_2}^{(0)} + \left(\mathbb{I}_{n_2}^{(1)} + \mathbb{I}_{n_2}^{(-1)}\right), \\ \mathbf{g}(\theta_1) &= 2A_{n_2} - 2B_{n_2}\cos(\theta_1), \\ T_{n_1,n_2}(f) &= T_{n_1}(\mathbf{g}) = 2\mathbb{I}_{n_1}^{(0)} \otimes A_{n_2} - \left(\mathbb{I}_{n_1}^{(1)} + \mathbb{I}_{n_1}^{(-1)}\right) \otimes B_{n_2}, \end{aligned}$$

and $T_{n_1,n_2}(f)$ is then a matrix of size $N \times N$ where $N = n_1 n_2$. The sampling of the symbol $f(\theta_1, \theta_2)$ is done as follows. The sampling grids in each dimension are the standard τ_{n_1} and τ_{n_2}, that is,

$$\theta^{(1)}_{j,n_1} = \frac{j\pi}{n_1+1}, \quad j = 1, \ldots, n_1,$$
$$\theta^{(2)}_{j,n_2} = \frac{j\pi}{n_2+1}, \quad j = 1, \ldots, n_2,$$

and the approximated eigenvalues are

$$\tilde{\lambda}_{i,j}(T_{n_1,n_2}(f)) = f(\theta^{(1)}_{i,n_1}, \theta^{(2)}_{j,n_2}).$$

In this example we have $\lambda_{i,j}(T_{n_1,n_2}(f)) = \tilde{\lambda}_{i,j}(T_{n_1,n_2}(f))$, since the symbol gives rise to tridiagonal structures in both dimensions for which the τ_n-grid returns the exact eigenvalues. The matrix $T_{n_1,n_2}(f)$ is shown below for $n_1 = 3$ and $n_2 = 4$. Its "blocks" are matrices of order $n_2 = 4$ and are shown in yellow $(2A_{n_2})$ and red (B_{n_2}).

$$T_{3,4}(f) = \left[\begin{array}{cccc|cccc|cccc}
6 & -2 & 0 & 0 & -3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
-2 & 6 & -2 & 0 & -1 & -3 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & -2 & 6 & -2 & 0 & -1 & -3 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & -2 & 6 & 0 & 0 & -1 & -3 & 0 & 0 & 0 & 0 \\
\hline
3 & -1 & 0 & 0 & 6 & -2 & 0 & 0 & -3 & -1 & 0 & 0 \\
-1 & -3 & -1 & 0 & -2 & 6 & -2 & 0 & -1 & -3 & -1 & 0 \\
0 & -1 & -3 & -1 & 0 & -2 & 6 & -2 & 0 & -1 & -3 & -1 \\
0 & 0 & -1 & -3 & 0 & 0 & -2 & 6 & 0 & 0 & -1 & -3 \\
\hline
0 & 0 & 0 & 0 & -3 & -1 & 0 & 0 & 6 & -2 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & -3 & -1 & 0 & -2 & 6 & -2 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & -3 & -1 & 0 & -2 & 6 & -2 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & -3 & 0 & 0 & -2 & 6
\end{array}\right]$$

We have the "sequence" of matrices $\{T_{n_1,n_2}(f)\}_{n_1,n_2}$ in which n_1 and n_2 can grow independently. Often, in the discretization of PDEs we have $n_i = \nu c_i$, $i = 1, 2$, where ν is related to the fineness, or discretization, parameter and grows to infinity, while the quantities c_i, $i = 1, 2$, are fixed proportionality constants. In Figure 1.5.5 we show the spectrum of $T_{n_1,n_2}(f)$, for $n_1 = 10$ and $n_2 = 50$, with black dots. The symbol $f(\theta_1, \theta_2)$ is shown in color.
For more details on multilevel Toeplitz matrices and their applications we refer the reader to [8, 30, 41, 42, 51], and the references therein.

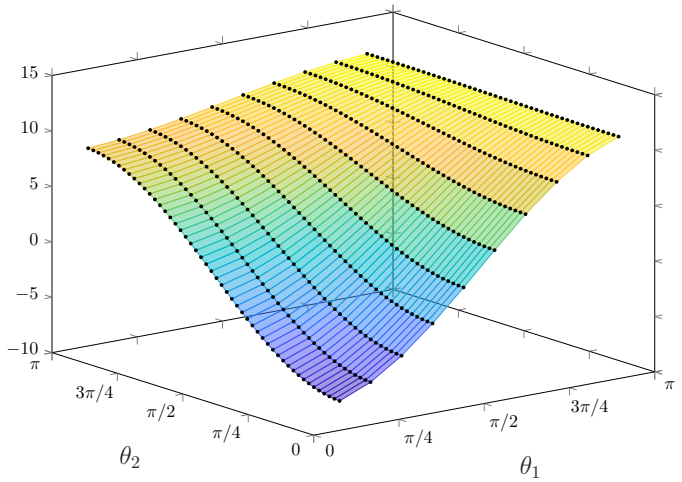


Figure 1.5.5. Example 1.5.5. The spectrum of the matrix $T_{10,50}(f)$ (black dots), with the symbol $f(\theta_1, \theta_2) = 2(3 - 2\cos(\theta_2)) - 2\cos(\theta_1)(3 + 2\cos(\theta_2))$ (colored surface).

Example 1.5.6. In this example we discuss the case of non-uniform discretizations in the context of finite differences; for the finite element setting, see [7]. Consider the problem

$$-(a(x)u'(x))' = b(x), \quad x \in (0,1),$$

with Dirichlet boundary conditions. Suppose we discretize this one-dimensional problem with the second order central finite difference method based on a non-uniform grid, such that, $0 = x_0 < x_1 < \ldots < x_{n+1} = 1$, as described in [40, Section 10.5.4]. Suppose further that the non-uniform grid is obtained as the mapping of the uniform grid $0 = \hat{x}_0 < \hat{x}_1 < \ldots < \hat{x}_{n+1} = 1$ through a fixed (increasing and bijective) function $G : [0,1] \to [0,1]$, that is, $x_j = G(\hat{x}_j)$ for all $j = 0, \ldots, n+1$. We assume that $G \in C^1([0,1])$ and that there exists at most finitely many points $\hat{x}$ such that $G'(\hat{x}) = 0$. Then, the resulting discretization matrix $A_{G,n}$ is approximately equal to $(n+1)D_n\,(a(G(\hat{x}))/G'(\hat{x}))\,T_n(2-2\cos(\theta))$ in the sense that

$$\frac{1}{n+1}A_{G,n} = D_n\left(\frac{a(G(\hat{x}))}{G'(\hat{x})}\right)T_n(2-2\cos(\theta)) + E_n,$$

where $\{E_n\}_n$ is a zero-distributed sequence. Thus, taking also into account that $A_{G,n}$ is symmetric, we conclude by axioms GLT 1, GLT 3, and GLT 5 that

$$\left\{\frac{1}{n+1}A_{G,n}\right\}_n \sim_{\text{GLT},\sigma,\lambda} \frac{a(G(\hat{x}))}{G'(\hat{x})}(2-2\cos(\theta)).$$

For more details, we refer to [40, Section 10.5.4].

For example, consider the map $G(\hat{x}) = 1 - \cos(\pi\hat{x})$. Note that the non-uniform grid generated by G is (almost) the grid consisting of the Chebyshev nodes in $[0,1]$. Since G' vanishes only at the points $\hat{x} = 0$ and $\hat{x} = 1$, we have

$$\left\{\frac{1}{n+1}A_{G,n}\right\}_n \sim_{\text{GLT},\sigma,\lambda} \frac{a(1-\cos(\pi\hat{x}))}{\pi\sin(\pi\hat{x})}(2-2\cos(\theta)).$$

To finalize this section, we here collect a few important references, with further details on the application of the theory of GLT sequences to different types of discretizations of differential equations:

- finite difference [40, Section 10.5], [41, Section 8.5], and [16, 67, 68],
- finite volume [9],
- discontinuous Galerkin [8, 30],
- finite element [40, Section 10.6] and [7, 16, 29, 42, 68],
- isogeometric analysis [40, Section 10.7], [41, Section 8.5], and [26, 32, 37, 38, 60],
- fractional differential equations [27, 52].

Integral equations are treated in [40, Section 10.4], and [1, 61]. Regarding preconditioning in Krylov methods, GLT based multigrid, and combinations of these methods in the spirit of multi-iterative solvers [62], see [23, 24, 25, 28, 40] and references therein.

1.6 The Approximation Errors

In this section we discuss the approximation errors

$$E_{j,n} = \lambda_j(A_n) - f(\theta_{j,n}),$$

that arise when $\{A_n\}_n \sim_{\text{GLT},\sigma,\lambda} f$, and $\theta_{j,n}$ is a grid that does not return the exact eigenvalues, when used to sample the symbol f.

Example 1.6.1. From Example 1.2.1, we know that the symbol for the second order finite difference discretization of the Laplacian is $f(\theta) = 2 - 2\cos(\theta)$. Assume we apply the Laplacian operator twice, commonly known as the bi-Laplacian or the biharmoic operator,

$$\nabla^4 u = \Delta^2 u,$$

and suppose we want to study the spectral properties of this operator. From axioms GLT 3 and GLT 5, we have $\{B_n\}_n \sim_{\text{GLT}} f^2$, where

$$\begin{aligned} B_n &= (T_n(f))^2, \\ g(\theta) &= f(\theta)^2 = (2 - 2\cos(\theta))^2 = 6 - 8\cos(\theta) + 2\cos(2\theta). \end{aligned}$$

The matrix B_n is the discretized bi-Laplacian, in the sense that we twice apply the Laplace operator, discretized by second order finite differences. Since $\{B_n\}_n \sim_{\text{GLT}} g$, and B_n is Hermitian, we have from axiom GLT 1 that $\{B_n\}_n \sim_{\sigma,\lambda} g$. We have,

$$\begin{aligned} B_n = T_5(f)T_5(f) &= \begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 5 & -4 & 1 & 0 & 0 \\ -4 & 6 & -4 & 1 & 0 \\ 1 & -4 & 6 & -4 & 1 \\ 0 & 1 & -4 & 6 & -4 \\ 0 & 0 & 1 & -4 & 5 \end{bmatrix}. \end{aligned} \tag{1.6.1}$$

Indeed, sampling $g(\theta)$ with the τ_n-grid returns the exact eigenvalues of B_n, that is $\lambda_j(B_n) = g(\theta_{j,n})$. Note the two corner elements, indicated in red, which deviate from the pure Toeplitz structure.

We focus now on $g(\theta)$ and $T_n(g)$. The generated matrix $T_n(g)$ is the discretized bi-Laplacian in the classical second order finite difference sense. We have the Fourier coefficients $\hat{g}_0 = 6$, $\hat{g}_1 = \hat{g}_{-1} = -4$, and $\hat{g}_2 = \hat{g}_{-2} = 1$, and thus

$$T_5(g) = \begin{bmatrix} \hat{g}_0 & \hat{g}_{-1} & \hat{g}_{-2} & 0 & 0 \\ \hat{g}_1 & \hat{g}_0 & \hat{g}_{-1} & \hat{g}_{-2} & 0 \\ \hat{g}_2 & \hat{g}_1 & \hat{g}_0 & \hat{g}_{-1} & \hat{g}_{-2} \\ 0 & \hat{g}_2 & \hat{g}_1 & \hat{g}_0 & \hat{g}_{-1} \\ 0 & 0 & \hat{g}_2 & \hat{g}_1 & \hat{g}_0 \end{bmatrix} = \begin{bmatrix} 6 & -4 & 1 & 0 & 0 \\ -4 & 6 & -4 & 1 & 0 \\ 1 & -4 & 6 & -4 & 1 \\ 0 & 1 & -4 & 6 & -4 \\ 0 & 0 & 1 & -4 & 6 \end{bmatrix}. \quad (1.6.2)$$

We note the difference between B_n of (1.6.1) and $T_n(g)$ of (1.6.2); the elements $(B_n)_{1,1}$ and $(B_n)_{n,n}$ (indicated in red) are not six, but five, a consequence of the multiplication of the two matrices $T_n(f)$ and $T_n(f)$ to attain B_n. The matrix B_n is equal to $T_n(g)$ except for a low-rank correction matrix R_n (in this case of rank 2),

$$R_n = B_n - T_n(g) = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

The low-rank correction R_n is zero-distributed according to the definition in Section 1.4. This means that as $n \to \infty$ the eigenvalues of B_n and $T_n(g)$ are the same. However, we work with finite-dimensional matrices $T_n(g)$ and we do not know an explicit grid $\theta_{j,n}$, which yields the exact eigenvalues of $\lambda_j(T_n(g))$ when sampling $g(\theta)$. Thus,

$$\lambda_j(T_n(g)) = g(\theta_{j,n}) + E_{j,n}, \quad j = 1, \ldots, n,$$

where $E_{j,n}$ are the errors of the approximations.
We display the graph of the errors $E_{j,n}$ in the left panel of Figure 1.6.1, when using the τ_n-grid for three different $n \in \{100, 200, 400\}$. The errors $E_{j,n}$ are of order $\mathcal{O}(h)$, where $h = 1/(n+1)$ for each n. Hence, it decreases linearly in n as n increases. It is also clear that the shape of the "error curve", given by plotting $E_{j,n}$ for all $j = 1, \ldots, n$, is retained as n increases. In the right panel of Figure 1.6.1, we show $E_{j,n}/h$ for each n, and the three curves overlap perfectly.

The overlap of the curves $E_{j,n}/h$ in the right panel of Figure 1.6.1 is due to the fact that for many types of symbols we have an asymptotic expansion of the errors $E_{j,n}$ of the form

$$\begin{aligned} E_{j,n} &= \lambda_j(T_n(f)) - f(\theta_{j,n}) \\ &= \sum_{k=0}^{\alpha} c_k(\theta_{j,n}) h^k + E_{j,n,\alpha}, \end{aligned}$$

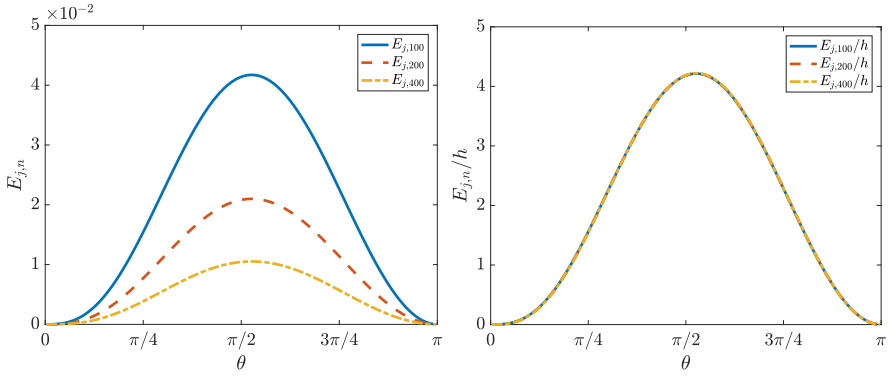


Figure 1.6.1. Example 1.6.1: The errors $E_{j,n}$, when approximating $\lambda_j(T_n(g))$ with samplings of $g(\theta) = 6 - 8\cos(\theta) + 2\cos(2\theta)$ for different n. Left: The approximation errors, $E_{j,n} = \mathcal{O}(h)$, decrease as n increases. Right: The scaled errors $E_{j,n}/h$, with $h = 1/(n+1)$, overlap for the differrent n.

for some constant α, a set of functions $c_k(\theta)$, and remaining errors $E_{j,k,\alpha}$. Note that $c_0(\theta) = 0$, and the summation is for $k = 0, \ldots, \alpha$, so if we choose $\alpha = 0$ then $E_{j,n} = E_{j,n,0}$. Hence, the right panel of Figure 1.6.1 shows

$$E_{j,n}/h = c_1(\theta_{j,n}) + E_{j,n,1}/h,$$

for three different n. Here $E_{j,n,1}/h$ is small, and $E_{j,n}/h$ is approximately equal to the samplings $c_1(\theta_{j,n})$, and since the function $c_1(\theta)$ is the same for all n, the three curves overlap.

The major contribution of this thesis is two-fold. We describe methods that efficiently estimate the functions $c_k(\theta)$, and then we use these approximations to accurately reduce the error of eigenvalue approximations for large structured matrices. Furthermore, using precomputed estimations $\tilde{c}_k(\theta)$ leads to highly efficient matrix-less methods. As shown in the following section, our approach is novel and uses a framework different from well-developed iterative eigenvalue solvers, such as the Lanczos and Arnoldi methods. We also comment on the generality of the approach and its applicability to analyze spectral properties of matrices obtained from differential operators.

2. Main Results and Contributions

> I want to climb this little mountain on my own,
> before I see the maps of the great explorers.
>
> ANONYMOUS

The main results and contributions of this thesis are summarized in this section. Each subsection is dedicated to respective Paper I–V.

2.1 Asymptotic Expansion of the Approximation Errors

In Paper I we investigate the errors of the approximated eigenvalues of banded symmetric Toeplitz (BST) matrices, given by sampling the generating function f. The generating function f is also the symbol by Szegő's limit theorem (1.3.1), since f is bounded and real-valued.

Introduction

We first recall a few fundamental facts from Section 1. Consider a real cosine trigonometric polynomial (RCTP)

$$f(\theta) = \hat{f}_0 + 2\sum_{\omega=1}^{p} \hat{f}_\omega \cos(\omega\theta), \quad \hat{f}_0, \ldots, \hat{f}_p \in \mathbb{R},$$

whose (possible) nonzero Fourier coefficients are $\hat{f}_\omega$, $\omega = 0, \ldots, p$. This RCTP is the generating function and the symbol of the sequence of BST matrices

$$T_n(f) = \begin{bmatrix} \hat{f}_0 & \ldots & \hat{f}_p & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \hat{f}_p & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \hat{f}_p \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & \hat{f}_p & \ldots & \hat{f}_0 \end{bmatrix}, \quad n = 1, 2, \ldots$$

If we sample $f(\theta)$ on a uniform grid in $[0, \pi]$, for example $\theta_{j,n} = j\pi h$ with $j = 1, \ldots, n$ and $h = 1/(n+1)$, we obtain an approximation of the eigenvalues of the matrix $T_n(f)$, that is,

$$\lambda_j(T_n(f)) = f(\theta_{j,n}) + E_{j,n}, \quad j = 1, \ldots, n,$$

where $E_{j,n}$ are the errors associated with the chosen grid points $\theta_{j,n}$. These errors $E_{j,n}$ are typically of order $\mathcal{O}(h)$, a fact that can be proved by using matrix-theory arguments such as the embedding in the τ_n-algebra; see [12]. However, when certain conditions are met, we show that one can approximate the errors in an efficient way with the asymptotic expansion

$$E_{j,n} = \sum_{k=0}^{\alpha} c_k(\theta_{j,n}) h^k + E_{j,n,\alpha},$$

where $c_k(\theta)$ are functions associated with the chosen grid type, $c_0(\theta) = 0$, and $E_{j,n} = E_{j,n,0}$. We thus have an expression for the eigenvalues of the matrix $T_n(f)$, namely

$$\begin{aligned} \lambda_j(T_n(f)) &= f(\theta_{j,n}) + E_{j,n} \\ &= f(\theta_{j,n}) + \sum_{k=0}^{\alpha} c_k(\theta_{j,n}) h^k + E_{j,n,\alpha}. \end{aligned} \tag{2.1.1}$$

The errors $E_{j,n,\alpha}$ are of order $O(h^{\alpha+1})$, or more precisely, as shown in Paper I, we have $|E_{j,n,\alpha}| \leq C_\alpha h^{\alpha+1}$, where C_α is a constant depending on α and f.

The Algorithm to Approximate $c_k(\theta)$, $k = 1, \ldots, \alpha$

In Paper I we introduce Algorithm 1, defined below, to find an approximation $\tilde{c}_k(\theta)$ of the functions $c_k(\theta)$, for $k = 1, \ldots, \alpha$. Assume we have a monotone non-decreasing RCTP $f(\theta)$, and we choose an integer α, and an integer n_1, typically small. In Paper I we only studied the τ_{n_1}-grid, that is $\theta_{j_1,n_1} = j_1 \pi h_1$ for the set of indices $j_1 = \{1, \ldots, n_1\}$ and $h_1 = 1/(n_1 + 1)$, however, Algorithm 1 can be modified to work for other types of grids. From (2.1.1) and the fact that $c_0(\theta) = 0$ and $\alpha > 0$, we have

$$E_{j_1,n_1} = \lambda_{j_1}(T_{n_1}(f)) - f(\theta_{j_1,n_1}) = \sum_{k=1}^{\alpha} c_k(\theta_{j_1,n_1}) h_1^k + E_{j_1,n_1,\alpha},$$

where it is assumed that the eigenvalues of each Toeplitz matrix generated by $f(\theta)$ are sorted in non-decreasing order. We now define $\alpha - 1$ additional matrix sizes by the rule $n_k = 2^{k-1}(n_1 + 1) - 1$, which is then true for $k = 1, \ldots, \alpha$. We also define $h_k = 1/(n_k + 1)$ and the sets of indices $j_k = 2^{k-1} j_1$, for all $k = 1, \ldots, \alpha$, because then, $\theta_{j_k,n_k} = j_k \pi h_k$ is the same for all k. Thus, we have for all $k = 1, \ldots, \alpha$,

$$E_{j_k,n_k} = \lambda_{j_k}(T_{n_k}(f)) - f(\theta_{j_1,n_1}) = \sum_{k=1}^{\alpha} c_k(\theta_{j_1,n_1}) h_k^k + E_{j_k,n_k,\alpha}.$$

Now, assuming n_1 is small enough, we compute all the errors E_{j_k,n_k} using any standard solver. This is done, for all k, by generating the matrix $T_{n_k}(f)$ and using an eigenvalue solver to find all eigenvalues $\lambda_j(T_{n_k}(f))$, sorted in non-decreasing order. Then, we sample the symbol $f(\theta)$ on the grid points θ_{j,n_k}. If

the samples $f(\theta_{j,n_k})$ are not sorted in non-decreasing order, that is, if $f(\theta)$ is not non-decreasing and monotone, we introduce a permutation σ on $\{1,\ldots,n_k\}$ that sorts the samples in non-decreasing order. We then consider the inverse permutation $\rho = \sigma^{-1}$, that sorts the eigenvalues $\lambda_{\rho(j)}(T_{n_k}(f))$ according to the samples $f(\theta_{j,n_k})$. Then, $E_{j_k,n_k} = \lambda_{\rho(j)}(T_{n_k}(f)) - f(\theta_{j,n_k})$. However, the assumption was that $f(\theta)$ is non-decreasing and monotone, so for clarity we omit ρ henceforth.

Because the error terms $E_{j_k,n_k,\alpha}$ are of order $\mathcal{O}(h_k^{\alpha+1})$, we decide to remove them from the system since they are small enough for our given purposes; by the choice of α and n_1. We then introduce $\tilde{c}_k(\theta)$ instead of $c_k(\theta)$, since we no longer solve the exact system, after ignoring all the error terms $E_{j_k,n_k,\alpha}$. Hence, we have

$$\begin{aligned}
E_{j_1,n_1} &= \tilde{c}_1(\theta_{j_1,n_1})h_1 + \tilde{c}_2(\theta_{j_1,n_1})h_1^2 + \ldots + \tilde{c}_\alpha(\theta_{j_1,n_1})h_1^\alpha,\\
&\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots\\
E_{j_k,n_k} &= \tilde{c}_1(\theta_{j_1,n_1})h_k + \tilde{c}_2(\theta_{j_1,n_1})h_k^2 + \ldots + \tilde{c}_\alpha(\theta_{j_1,n_1})h_k^\alpha,\\
&\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots\\
E_{j_\alpha,n_\alpha} &= \tilde{c}_1(\theta_{j_1,n_1})h_\alpha + \tilde{c}_2(\theta_{j_1,n_1})h_\alpha^2 + \ldots + \tilde{c}_\alpha(\theta_{j_1,n_1})h_\alpha^\alpha,
\end{aligned}$$

which is written in compact form as

$$\mathbf{E} = \mathbf{V}\mathbf{C}, \tag{2.1.2}$$

where $\mathbf{E} \in \mathbb{R}^{\alpha\times n_1}, \mathbf{V} \in \mathbb{R}^{\alpha\times\alpha}, \mathbf{C} \in \mathbb{R}^{\alpha\times n_1}$ are given by

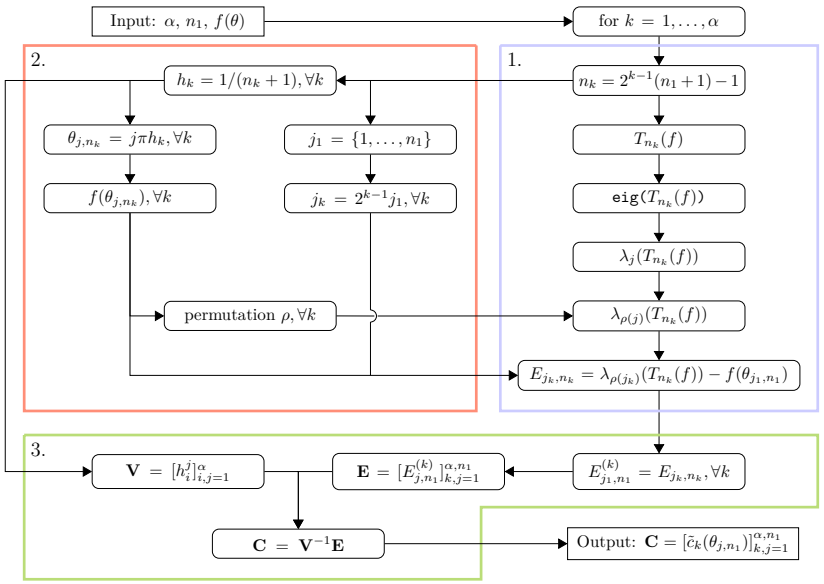
$$\mathbf{E} = \begin{bmatrix} E_{j_1,n_1} \\ E_{j_2,n_2} \\ \vdots \\ E_{j_\alpha,n_\alpha} \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} h_1 & h_1^2 & \ldots & h_1^\alpha \\ h_2 & h_2^2 & \ldots & h_2^\alpha \\ \vdots & \vdots & \ddots & \vdots \\ h_\alpha & h_\alpha^2 & \ldots & h_\alpha^\alpha \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \tilde{c}_1(\theta_{j_1,n_1}) \\ \tilde{c}_2(\theta_{j_1,n_1}) \\ \vdots \\ \tilde{c}_\alpha(\theta_{j_1,n_1}) \end{bmatrix}.$$

By solving the system (2.1.2) we obtain our desired approximation of $c(\theta_{j_1,n_1})$,

$$\mathbf{C} = \mathbf{V}^{-1}\mathbf{E}.$$

The Vandermonde matrix $\mathbf{V}$ is typically ill-conditioned and the system (2.1.2) is commonly solved using LAPACK [2] or similar packages, and not by inversion of the matrix $\mathbf{V}$. In Algorithm 1, we present the process for computing the approximations of the $c_k(\theta)$ functions in a flowchart, for a specified input of α, n_1, and f.

Algorithm 1 Approximate $c_k(\theta_{j,n_1})$ for specified α, n_1, and $f(\theta)$.



Step	Description
	Input to algorithm: $\alpha \in \mathbb{N}, n_1 \in \mathbb{N}$, and $f(\theta)$ monotone RCTP.
1.	Loop over $k = 1, \ldots, \alpha$.
1.1	Define $n_k = 2^{k-1}(n_1 + 1) - 1$.
1.2	Generate matrix $T_{n_1}(f)$.
1.3	Compute eigenvalues of $T_{n_1}(f)$.
1.4	Sort eigenvalues in non-decreasing order, $\lambda_j(T_{n_1}(f))$.
1.5	Input: ρ (2.1.3). Reorder with permutation ρ, $\lambda_{\rho(j)}(T_{n_1}(f))$.
1.6	Input: $f(\theta_{j_1,n_1})$ (2.1.2), j_k (2.2.1). Compute errors E_{j_k,n_k}.
2.	Input: n_k (1.1).
2.1	Define $h_k = 1/(n_k + 1)$.
2.1.1	Define $\theta_{j,n_k} = j\pi h_k$.
2.1.2	Compute $f(\theta_{j,n_k})$ for $j = 1, \ldots, n$.
2.1.3	Define permutation ρ.
2.2	Define $j_1 = \{1, \ldots, n_1\}$.
2.2.1	Define $j_k = 2^{k-1} j_1$.
3.	Input: E_{j_k,n_k} (1.6), h_k (2.1).
3.1	Define $E_{j_1,n_1}^{(k)} = E_{j_k,n_k}$.
3.2	Define error matrix $\mathbf{E} = [E_{j,n_1}^{(k)}]_{k,j=1}^{\alpha,n_1}$.
3.3	Define Vandermonde matrix $\mathbf{V} = [h_i^j]_{i,j=1}^{\alpha}$.
3.4	Compute $\mathbf{C}$ by solving $\mathbf{VC} = \mathbf{E}$.
	Output of algorithm: $\mathbf{C} = [\tilde{c}_k(\theta_j, n_1)]_{k,j=1}^{\alpha,n_1}$.

Example 2.1.1. We consider the same RCTP symbol as in Example 1.6.1, namely, $f(\theta) = (2 - 2\cos(\theta))^2 = 6 - 8\cos(\theta) + 2\cos(2\theta)$, and use Algorithm 1 to compute an approximation to the expansion (2.1.1), that is, $\tilde{c}_k(\theta_{j_1,n_1})$. First we focus on a simple case, that is, $\alpha = 3$ and $n_1 = 5$.

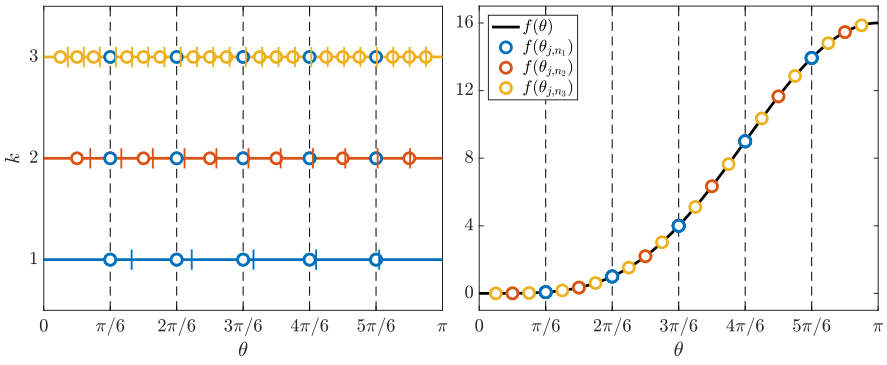


Figure 2.1.1. Example 2.1.1: τ_{n_k}-grids and symbol $f(\theta) = 6 - 8\cos(\theta) + 2\cos(2\theta)$. Left: Three grids, θ_{j,n_k}, where $n_1 = 5$, $n_2 = 11$, and $n_3 = 23$. Blue circles are grid points θ_{j_k,n_k}. Colored vertical lines are grid points that give exact eigenvalues, when sampling $f(\theta)$. Right: Symbol $f(\theta)$, and the three grids, θ_{j,n_k}.

In the left panel of Figure 2.1.1, we show the three grids θ_{j,n_k} that correspond to the sizes $n_k = 2^{k-1}(n_1 + 1) - 1$, $k = 1, 2, 3$, that is, $n_1 = 5$ (blue), $n_2 = 11$ (red), and $n_3 = 23$ (yellow). The blue circles are the three sets of indices j_1, j_2, and j_3, such that θ_{j_k,n_k} is the same for all k. In addition, with vertical colored lines we show the grid points for which the exact eigenvalues $\lambda_j(T_{n_k}(f))$ are obtained, when sampling the symbol $f(\theta)$. In the right panel of Figure 2.1.1, we plot the symbol (black line), and the three sampling grids, with the same coloring as in the left panel.

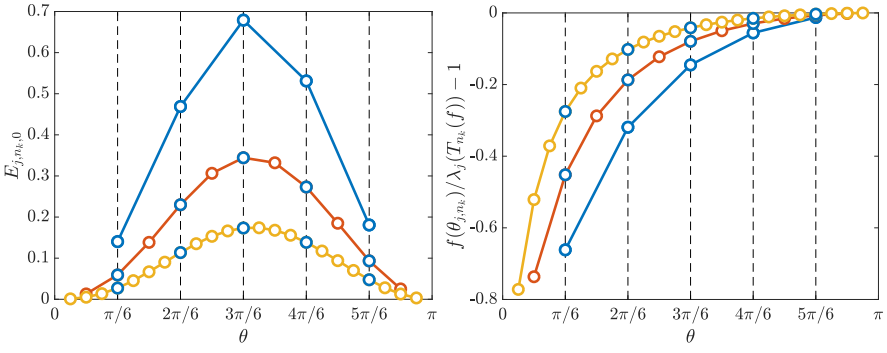


Figure 2.1.2. Example 2.1.1: Errors when approximating eigenvalues by sampling the symbol, with grids of sizes $n_1 = 5$ (blue), $n_2 = 11$ (red), and $n_3 = 23$ (yellow). Left: Absolute errors, $E_{j,n_k,0} = \lambda_j(T_{n_k}(f)) - f(\theta_{j,n_k})$. Right: Relative errors, $f(\theta_{j,n_k})/\lambda_j(T_{n_k}(f)) - 1$.

In Figure 2.1.2 we show the errors that arise when approximating the eigenvalues by sampling the symbol with the grids θ_{j,n_k}, for $k = 1, 2, 3$. In the left panel we visualize the absolute errors $E_{j,n_k,0} = \lambda_j(T_{n_k}(f)) - f(\theta_{j,n_k})$. Note that the errors $E_{j_k,n_k,0}$ (blue circles) are the errors used in Algorithm 1. In the right panel we present the relative errors, that is, $f(\theta_{j,n_k})/\lambda_j(T_{n_k}(f)) - 1$.

Remark 2.1.1. Since the symbol $f(\theta)$ does not comply with the simple-loop conditions described, for example, in [6, 13, 14] — which require that $f'(\theta) \neq 0$ for $\theta \in (0, \pi)$ and $f''(\theta) \neq 0$ for $\theta \in \{0, \pi\}$ — because $f''(\theta) = 8\cos(\theta) - 8\cos(2\theta) = 0$ for $\theta = 0$, the expansion (2.1.1) is point-wise not true for all eigenvalues, which are close to zero [6]. We now demonstrate why this in practice, when using standard double precision computations, is not a problem when using Algorithm 1.

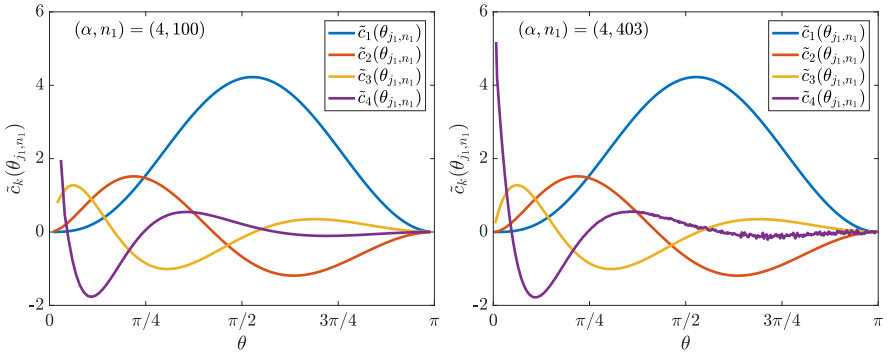


Figure 2.1.3. Example 2.1.1: The approximations $\tilde{c}_k(\theta_{j,n_1}), k = 1, \ldots, \alpha$, for $\alpha = 4$, and the symbol $f(\theta) = 6 - 8\cos(\theta) + 2\cos(2\theta)$. Point-wise erratic approximations are removed. Left: Smooth behavior for all four $\tilde{c}_k$ when $n_1 = 100$. Right: For $n_1 = 403$, we have oscillatory behavior in $\tilde{c}_4$, due to the double precision computations.

Now we use Algorithm 1 to compute approximations of $c_k(\theta)$ for $\alpha = 4$, and two different n_1; $n_1 = 100$ and $n_1 = 2^{3-1}(100+1)-1 = 403$. In Figure 2.1.3 we show the resulting approximations, $\tilde{c}_k$. The points $\tilde{c}_3(\theta_{1,n_1})$, $\tilde{c}_4(\theta_{1,n_1})$, and $\tilde{c}_4(\theta_{2,n_1})$ have been removed from both panels, because of their erratic behavior, as mentioned in Remark 2.1.1. In the left panel of Figure 2.1.3, we show the results with $n_1 = 100$, and in the right panel $n_1 = 403$. As is shown, the functions $\tilde{c}_3$ and $\tilde{c}_4$ behave well in the domain and the erratic behavior is only point-wise, close to zero for the three removed approximations. Moreover, $\tilde{c}_4$ in the right panel shows an erratic behavior in the right part of the domain $[0, \pi]$, but this is due to using double precision arithmetic for the computations. Increasing n_1 would make $\tilde{c}_4$ oscillatory, unless we increase the precision of the computations.

Another approach to remedy the issue of point-wise errors in the expansion, is to simply increase α in the computations. The resulting approximated $\tilde{c}_k$ for larger ks are oscillatory for standard double precision arithmetics, so they are in themselves not useful. However, as is demonstrated in Figure 2.1.4, the point-wise error for $\tilde{c}_3(\theta_{1,n_1})$ is effectively suppressed. In the left panel, we see the behavior of $\tilde{c}_3$ in a neighborhood of zero, when computed for four different n_1. We choose $n_1^{(q)} = 2^{q-1}(n_1^{(1)} + 1) - 1$ where $n_1^{(1)} = 100$ and $q = 1, \ldots, 4$, since then the resulting approximations $\tilde{c}_k$ overlap. Clearly the first eigenvalue for each $n_1^{(q)}$ behaves erratically, in the sense that the expansion point-wise does not work as expected; this behavior is described in detail in [6]. In the right panel, we see the same approximations $\tilde{c}_k$ as in the left panel, however, we have used $\alpha = 6$ to compute them. The erratic behavior of the approximation of $\tilde{c}_3$ for the first grid point has been suppressed enough to be considered negligibly small.

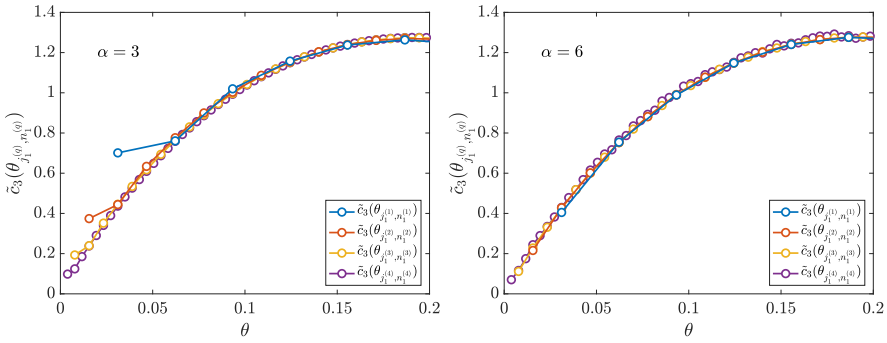


Figure 2.1.4. Example 2.1.1: Erratic behavior for the approximation $\tilde{c}_3(\theta_{1,n_1^{(q)}})$, for $q = 1, \ldots, 4$. Left: Computations with $\alpha = 3$. Erratic $\tilde{c}_3(\theta_{1,n_1^{(q)}})$, for all q. Right: Computations with $\alpha = 6$ suppresses the erratic $\tilde{c}_3(\theta_{1,n_1^{(q)}})$, for all q.

In Figure 2.1.5 we show the results of Algorithm 1 with $\alpha = 3$ and $n_1 = 100$. As usual, we define $n_k = 2^{k-1}(n_1 + 1) - 1$, $h_k = 1/(n_k + 1)$, $j_1 = \{1, \ldots, n_1\}$, and $j_k = 2^{k-1} j_1$. In the top left panel, the errors $E_{j_k,n_k,0} = \lambda_{j_k}(T_n(f)) - f(\theta_{j_k,n_k})$, for $k = 1, 2, 3$, are shown. In the middle left panel, we present the computed errors $\tilde{E}_{j_k,n_k,1} = E_{j_k,n_k,0} - \tilde{c}_1(\theta_{j_k,n_k})h_k$. In the bottom left panel, we finally show $\tilde{E}_{j_k,n_k,2} = \tilde{E}_{j_k,n_k,1} - \tilde{c}_2(\theta_{j_k,n_k})h_k^2$. The errors $\tilde{E}_{j_k,n_k,3} = \tilde{E}_{j_k,n_k,2} - \tilde{c}_3(\theta_{j_k,n_k})h_k^3$ are not shown since they are zero, to machine precision. In the right panels of Figure 2.1.5, we show $E_{j_k,n_k,0}/h_k$, $\tilde{E}_{j_k,n_k,1}/h_k^2$, and $\tilde{E}_{j_k,n_k,2}/h_k^3$. These curves are in essence the computed $\tilde{c}_k$ for $k = 1, 2, 3$; compare with the $\tilde{c}_k$ in Figure 2.1.3.

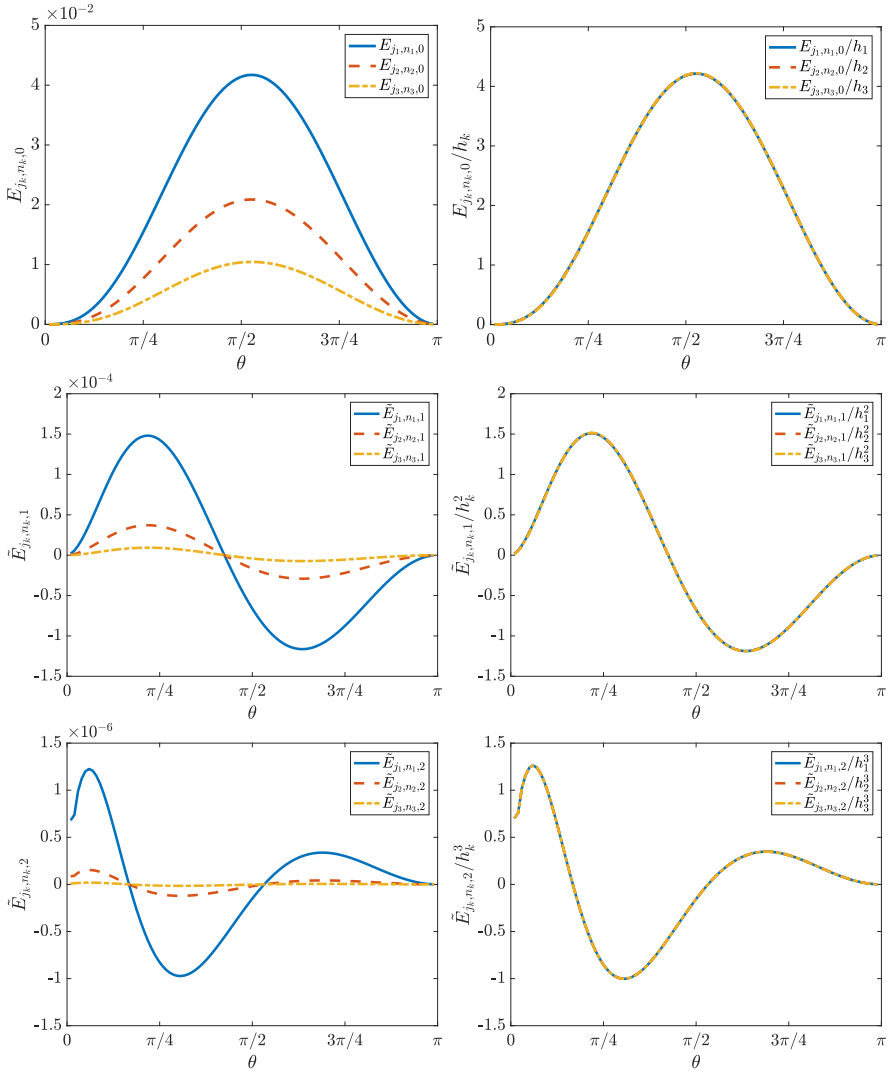


Figure 2.1.5. Example 2.1.1: The symbol $f(\theta) = 6 - 8\cos(\theta) + 2\cos(2\theta)$, $\alpha = 3$, and $n_1 = 100$. Left: $E_{j_k,n_k,0}$, $\tilde{E}_{j_k,n_k,1}$, and $\tilde{E}_{j_k,n_k,2}$ for $k = 1, 2, 3$. Right: $E_{j_k,n_k,0}/h_k$, $\tilde{E}_{j_k,n_k,1}/h_k^2$, and $\tilde{E}_{j_k,n_k,2}/h_k^3$ for $k = 1, 2, 3$. The errors $\tilde{E}_{j_k,n_k,3}$ are zero for all k.

We now consider how many of the functions c_k that we need to approximate, in order to obtain "good" results when we use them to approximate the eigenvalues of a large matrix. Assume we are interested in the eigenvalues, which are of order $\mathcal{O}(1)$, of a large matrix, say in the order of $n = \mathcal{O}(10^6)$, and as seen in Figure 2.1.3, the c_k are of order $\mathcal{O}(1)$. Then, when using the approximated $\tilde{c}_k$

in (2.1.1), and $h = \mathcal{O}(10^{-6})$, we have $\tilde{c}_3 h^3 = \mathcal{O}(10^{-18})$. For standard double precision arithmetic computations, this is beyond machine precision, which is in the order of $\mathcal{O}(10^{-16})$. Hence, it is usually sufficient to approximate the first two or three c_k, and we should do it as accurately as possible.

Using Approximations $\tilde{c}_k(\theta_{j_1,n_1})$ on a Large Matrix

When we have the approximations $\tilde{c}_k(\theta_{j_1,n_1})$ from Algorithm 1, we use them to approximate the errors for a subset of the eigenvalues of a much larger matrix. We are restricted to the matrices of size $n_m = 2^{m-1}(n_1 - 1) + 1$, for some integer $m > \alpha$. The eigenvalues for which we estimate the approximation, are the ones corresponding to indices $j_m = 2^{m-1} j_1$. With $h_m = 1/(n_m + 1)$, we have

$$\lambda_{j_m}(T_{n_m}(f)) \approx \tilde{\lambda}_{j_m}(T_{n_m}(f)) = f(\theta_{j_1,n_1}) + \sum_{k=1}^{\alpha} \tilde{c}_k(\theta_{j_1,n_1}) h_m^k. \qquad (2.1.3)$$

Example 2.1.2. In this example, we use approximations $\tilde{c}_k(\theta_{j_1,n_1})$, from Example 2.1.1, to approximate the eigenvalues with indices j_m of matrices $T_{n_m}(f)$ by (2.1.3).

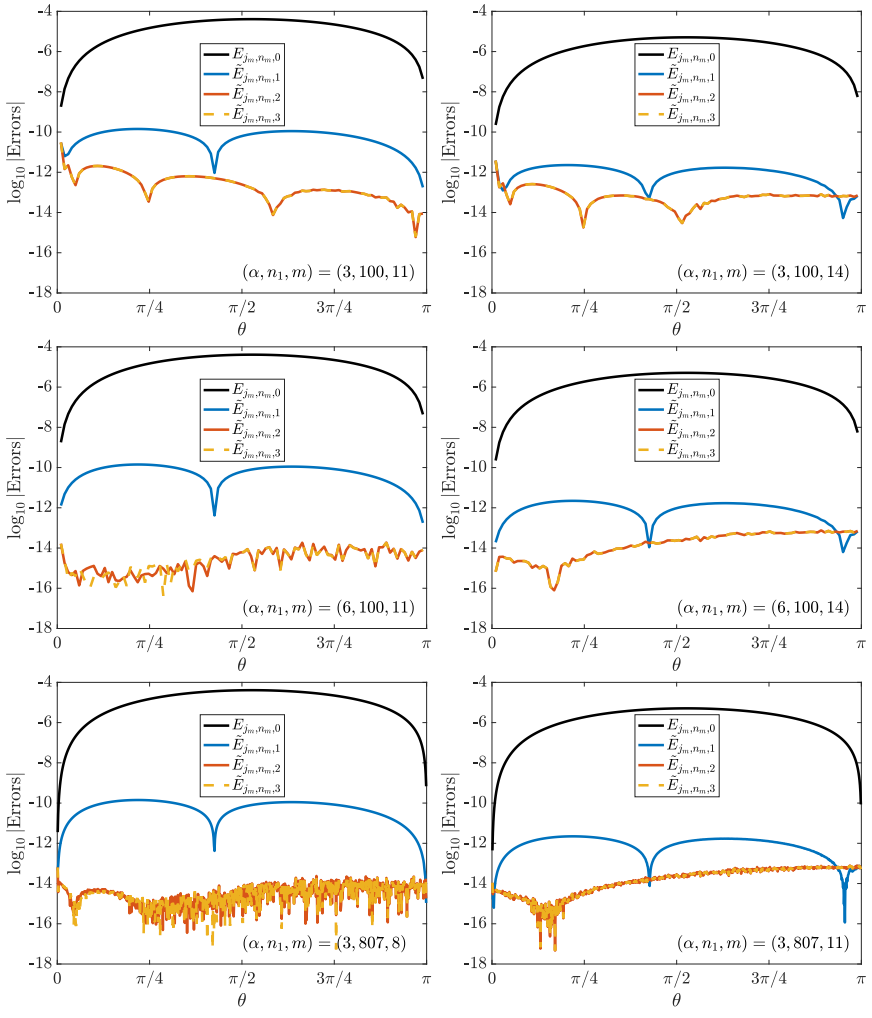


Figure 2.1.6. Example 2.1.2: Matrix $T_n(f)$. Reducing the errors for indices j_m, for matrix orders $n_m = 104323$ (left) and $n_m = 827391$ (right). Different (α, n_1, m). Errors $E_{j_m,n_m,0}$ (black) and reduced errors $\tilde{E}_{j_m,n_m,1}$ (blue), $\tilde{E}_{j_m,n_m,2}$ (red), and $\tilde{E}_{j_m,n_m,3}$ (yellow). Top: α and n_1 are too small to obtain a reduction to $\mathcal{O}(10^{-14})$. Middle: α increased; a reduction to $\mathcal{O}(10^{-14})$ is attained. Bottom: n_1 increased instead of α; a reduction to $\mathcal{O}(10^{-14})$ is attained for more eigenvalues, than in middle panels, and at a lower cost.

In Figure 2.1.6 we show the errors for different sets of parameters (α, n_1, m). The left panels are associated with a large matrix of order $n_m = 104323$, whereas in the right panels the order is $n_m = 827391$. We use only $\tilde{c}_1$, $\tilde{c}_2$ and $\tilde{c}_3$ to approximate the errors $E_{j_m,n_m,0}$ (black). Since, as explained earlier, the term $\tilde{c}_3 h_m^3$ is too small to affect the approximation in any significant way, we only include it for demonstrating this point. In the two top panels, we compute the approximations $\tilde{c}_k$ using $\alpha = 3$ and $n_1 = 100$. The resulting errors $\tilde{E}_{j_m,n_m,3}$

(yellow) is relatively large, especially close to zero. In the two middle panels, we instead compute the approximations $\tilde{c}_k$ using $\alpha = 6$ and $n_1 = 100$. The result is clearly better than in the top panels, also in the region close to zero. In the two bottom panels, we compute the approximations $\tilde{c}_k$ using $\alpha = 3$ and $n_1 = 807$. The order of the errors is the same as for the computations in the middle panels, however, we now have 807 approximations, compared to 100, and we have computed them with less work; computing the eigenvalues of $T_{n_k}(f)$ for $n_k = 2^{k-1}(100+1) - 1$ and $k = 1, \ldots, 6$ in the middle panels, and only $k = 4, 5, 6$ in the bottom panels. This confirms the general rule that it is better to increase n_1 than to increase α to attain a better approximation of the c_k functions. It is also advantageous to increase α beyond the number of desired c_k approximations.

Example 2.1.3. As is seen in Figure 2.1.3, in Example 2.1.1, we quickly reach the limit of machine precision even for moderate n_1, when using standard 64 bit double precision computations. In this example we construct an artificial case, namely we use the symbol from Example 1.5.2, that is, $f(\theta) = 2 - 2\cos(\theta)$. For this symbol we can, as stated in (1.5.2), attain the exact eigenvalues of the generated matrix $T_n(f)$, by sampling with the τ_n-grid. We now sample the symbol with the τ^{π}_{n-1}-grid instead, and approximate the expansion with Algorithm 1. The matrix sizes, used in Algorithm 1, are then defined as $n_k = 2^{k-1}n_1$, indices $j_k = 2^{k-1}j_1$ where $j_1 = \{1, \ldots, n_1\}$, $h_k = 1/n_k$, and $\theta_{j_k,n_k} = j_k\pi/n_k$.

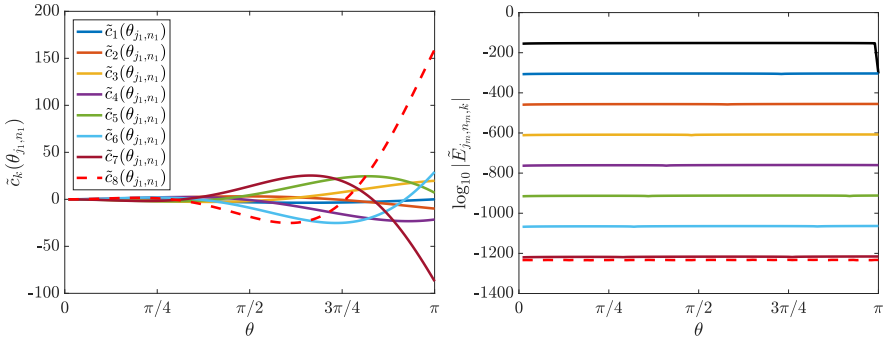


Figure 2.1.7. Example 2.1.3: High precision computations, 4096 bits, $n_1 = 100$. Left: The eight first approximations $\tilde{c}_k(\theta_{j_1,n_1})$. Right: The errors of eigenvalue approximations for a matrix of order $n_m = n_{500} = 2^{500-1} \cdot 100$, when using the computed $\tilde{c}_k(\theta_{j_1,n_1})$ in (2.1.3). Black line is $E_{j_m,\underline{n}_m,0} = \tilde{E}_{j_m,n_{\underline{m}},0} = \lambda_{j_m}(T_{n_m}(f)) - f(\theta_{j_1,n_1})$, and each consecutive colored line is $\tilde{E}_{j_m,n_m,k} = E_{j_m,n_m,k-1} - \tilde{c}_k(\theta_{j_1,n_1})h_m^k$, for $k = 1, \ldots, 8$. Accuracy to machine precision of order $\mathcal{O}(10^{-1233})$ is attained.

In the left panel of Figure 2.1.7 the first eight $\tilde{c}_k$ are presented. These computations are performed with $\alpha = 100$ and $n_1 = 100$, and with 4096 bit precision (machine precision is $1.915 \cdot 10^{-1233}$) in JULIA [10]. On double precision it is not possible to compute more than five $\tilde{c}_k$, with $n_1 = 100$, before the accuracy is disturbed by machine limitation. The right panel of Figure 2.1.7 shows the errors when applying the first eight of the 100 computed $\tilde{c}_k$ using (2.1.3), to the errors of a large matrix [1] of order $n_{500} = 2^{500-1} \cdot 100$, that is, $h_m = \mathcal{O}(10^{-153})$. As expected this is enough to attain close to machine precision accuracy of order $\mathcal{O}(10^{-1233})$, since $\tilde{E}_{j_m,n_m,8} \approx E_{j_m,n_m,8} \leq C_8 h_m^9 = \mathcal{O}(10^{-1377})$.

Asymptotic Expansion of Errors for Non-Monotone Symbols

In Paper I, we also discuss the case of symbols that are non-monotone in parts of the domain, and the possibility of using the expansion of the errors for them. We show through numerical experiments that it is possible to use the expansion in subintervals of the domain where the symbol is monotone. In Paper V, we compute the spectrum for a class of non-monotone symbols, where a technique of multiple sub-grids can be used to obtain a solution also for non-monotone symbols.

Example 2.1.4. Consider the same partially non-monotone symbol as in Example 6 of Paper I, namely $f(\theta) = 2 - \cos(\theta) - \cos(3\theta)$. The left panel of Figure 2.1.8 shows the graph of the symbol f. The non-monotone region, corresponding to the interval $[\hat{\theta}, \pi - \hat{\theta}]$, is depicted in red. In the right panel, we present the computed $\tilde{c}_k(\theta_{j_1,n_1})$, for $\alpha = 5$ and $n_1 = 100$. Compare with Figure 7 of Paper I, where we only compute $\tilde{c}_1$. For an $\alpha > 5$, the computed $\tilde{c}_k$ for $k > 5$, if only double precision is used, have oscillatory behavior due to the low precision arithmetics.

2.2 Extending the Expansion to the Preconditioned Case

In Paper II we extend the results from Paper I to a more general case, which we refer to as the preconditioned case. More precisely, in Paper II we are interested in the eigenvalues of a "preconditioned" matrix of the form $T_n^{-1}(g)T_n(f)$, where f and g are real-valued cosine trigonometric polynomials (RCTPs). This is the definition of the term preconditioned matrix in this thesis, unless otherwise stated. Recall from axiom GLT 6 that if $\{A_n\}_n \sim_{\text{GLT}} f$ and $f \neq 0$ a.e., then $\{A_n^\dagger\}_n \sim_{\text{GLT}} f^{-1}$. In particular, assuming that $g \neq 0$ a.e., we have $\{T_n^{-1}(g)T_n(f)\}_n \sim_{\text{GLT}} g^{-1}f$. We show that the same type of asymptotic ex-

[1] $n_{500} = 2^{500-1} \cdot 100 = 16366953039480709350065948484137995761083210230215323947416456840480668982023372774416350461629520785754433420637800355046086282729426965266642637946880$0.

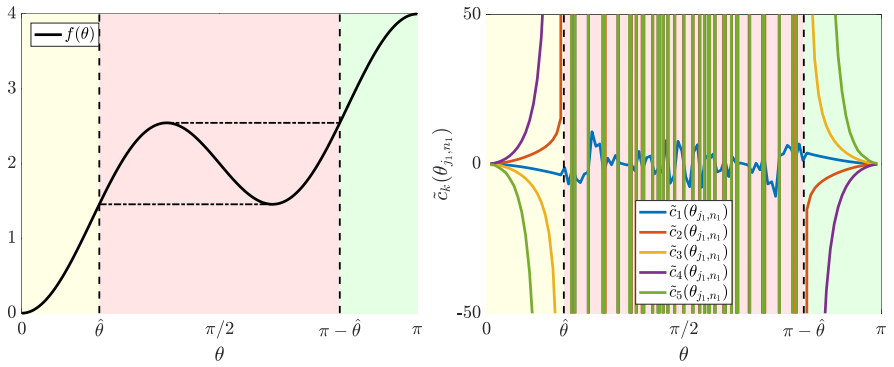


Figure 2.1.8. Example 2.1.4: Symbol $f(\theta) = 2 - \cos(\theta) - \cos(3\theta)$. Left: The symbol f, with monotone regions (yellow and green) and non-monotone region (red). Right: The five computed $\tilde{c}_k(\theta_{j_1,n_1})$, being smooth (yellow and green) and oscillatory (red).

pansion of the errors as (2.1.1) holds in the preconditioned case, namely

$$\begin{aligned}\lambda_j(T_n^{-1}(g)T_n(f)) &= \frac{f(\theta_{j,n})}{g(\theta_{j,n})} + E_{j,n} \\ &= \frac{f(\theta_{j,n})}{g(\theta_{j,n})} + \sum_{k=0}^{\alpha} c_k(\theta_{j,n})h^k + E_{j,n,\alpha}. \end{aligned} \tag{2.2.1}$$

It is important to note that if we define $r(\theta) = f(\theta)/g(\theta)$, then we are here approximating the eigenvalues of the matrix $T_n^{-1}(g)T_n(f)$ and not the Toeplitz matrix $T_n(r)$, generated by r.

The motivation to study the spectrum of these types of matrices arises in several different settings. When using an iterative method to numerically solve the system $A_n\mathbf{u}_n = \mathbf{b}_n$, where A_n is symmetric positive definite, one measure of the difficulty is the so-called spectral condition number,

$$\varkappa(A_n) = \frac{\lambda_{\max}(A_n)}{\lambda_{\min}(A_n)}.$$

If $\varkappa(A_n)$ is large, that is, the quotient of the largest and smallest eigenvalue is large, it is in general difficult to solve the system by iterative methods. One can then introduce a preconditioner, that is, a matrix P^{-1} such that

$$P^{-1}A_n\mathbf{u}_n = P^{-1}\mathbf{b}_n,$$

is easier to solve, because P^{-1} is chosen such that $\varkappa(P^{-1}A_n) = \mathcal{O}(1)$. Ideally $P = A_n$, since then the system is solved. Examples of applications of the GLT theory to the design of preconditioners can be found in [23, 24, 25, 28, 40, 62] and the references therein.

In Paper IV, described in Section 2.4, we use the results of Paper II to solve the eigenvalue problem $-\Delta u = \lambda u$, which in discretized form, by isogeometric analysis (IgA), reads as the following generalized eigenvalue problem

$$K_n^{[p]}\mathbf{u}_n = \lambda M_n^{[p]}\mathbf{u}_n \qquad \Rightarrow \qquad \underbrace{\left(M_n^{[p]}\right)^{-1} K_n^{[p]}}_{L_n^{[p]}}\mathbf{u}_n = \lambda \mathbf{u}_n$$

The discretized system is solved by approximating the eigenvalues of $L_n^{[p]}$, and this can be done since we have the symbols of $nM_n^{[p]}$ and $n^{-1}K_n^{[p]}$, and thus we accurately find the eigenvalues of $n^{-2}L_n^{[p]}$ by using (2.2.1).

Example 2.2.1. We now extend Example 2.1.1 further, by preconditioning the matrix $T_5(f)$ by a matrix $T_5^{-1}(g)$, which is the inverse of the matrix $T_5(g)$ generated by a symbol g. We consider the following specific choices of symbols

$$\begin{aligned} f(\theta) &= (2-2\cos(\theta))^2 = 6 - 8\cos(\theta) + 2\cos(2\theta), \\ g(\theta) &= 3 + 2\cos(\theta), \\ r(\theta) &= \frac{f(\theta)}{g(\theta)} = \frac{(2-2\cos(\theta))^2}{3+2\cos(\theta)}. \end{aligned}$$

The matrix of interest is thus $T_n^{-1}(g)T_n(f)$, with the symbol $r = g^{-1}f$.

Remark 2.2.1. The symbol g in this example is arbitrary, in the sense that it only serves to demonstrate the applicabilty of Algorithm 1 to approximate the spectrum of matrices of the form $T_n^{-1}(g)T_n(f)$, when $r = g^{-1}f$ is monotone. The design and construction of preconditioners to improve iterative solution methods is out of the scope of this thesis.

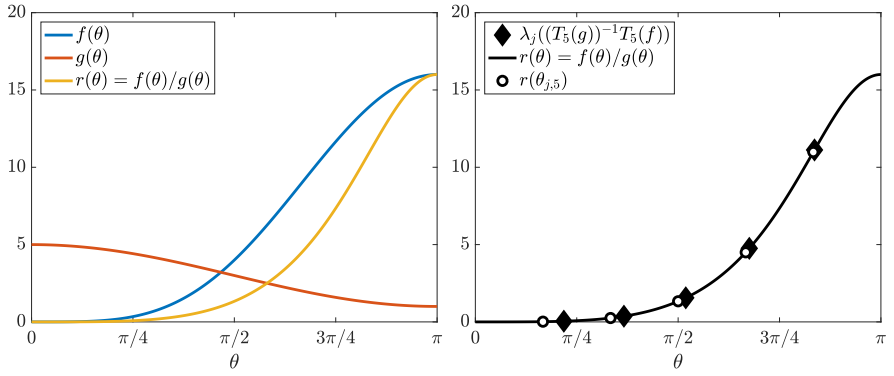


Figure 2.2.1. Example 2.2.1: Preconditioned matrix $T_n^{-1}(g)T_n(f)$. Left: Symbols f, g, and $r = g^{-1}f$. Right: Eigenvalues $\lambda_j(T_5^{-1}(g)T_5(f))$, symbol $r(\theta) = f(\theta)/g(\theta)$, and samplings $r(\theta_{j,5})$ with the τ_5-grid.

In the left panel of Figure 2.2.1, we show the graphs of the three symbols f, g, and $r = g^{-1}f$. In the right panel of Figure 2.2.1, we see the five eigenvalues $\lambda_j(T_5^{-1}(g)T_5(f))$, the symbol r, and samplings of $r(\theta_{j,5})$, where $\theta_{j,5}$ is the τ_5-grid.

From axiom GLT 1, we have $\{T_n(f)\}_n \sim_{\text{GLT},\sigma,\lambda} f$, $\{T_n(g)\}_n \sim_{\text{GLT},\sigma,\lambda} g$, $\{T_n(g^{-1})\}_n \sim_{\text{GLT},\sigma,\lambda} g^{-1}$, $\{T_n(g^{-1}f)\}_n \sim_{\text{GLT},\sigma,\lambda} g^{-1}f$, and from axiom GLT 6, we have $\{T_n^{-1}(g)\}_n \sim_{\text{GLT},\sigma,\lambda} g^{-1}$. Finally, from axiom GLT 5, GLT 7 and with the same symmetrization argument as in [40, solution of Exercise 8.4, pp. 291–292] we obtain $\{T_n^{-1}(g)T_n(f)\}_n \sim_{\text{GLT},\sigma,\lambda} g^{-1}f$. When $n = 5$ we have

$$T_5(f) = \begin{bmatrix} 6 & -4 & 1 & 0 & 0 \\ -4 & 6 & -4 & 1 & 0 \\ 1 & -4 & 6 & -4 & 1 \\ 0 & 1 & -4 & 6 & -4 \\ 0 & 0 & 1 & -4 & 6 \end{bmatrix}, \quad T_5(g) = \begin{bmatrix} 3 & 1 & 0 & 0 & 0 \\ 1 & 3 & 1 & 0 & 0 \\ 0 & 1 & 3 & 1 & 0 \\ 0 & 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 1 & 3 \end{bmatrix},$$

and the inverse of $T_5(g)$ is

$$\begin{aligned} T_5^{-1}(g) &= \frac{1}{144}\begin{bmatrix} 55 & -21 & 8 & -3 & 1 \\ -21 & 63 & -24 & 9 & -3 \\ 8 & -24 & 64 & -24 & 8 \\ -3 & 9 & -24 & 63 & -21 \\ 1 & -3 & 8 & -21 & 55 \end{bmatrix} \\ &\approx \begin{bmatrix} 0.3819 & -0.1458 & 0.0556 & -0.0208 & 0.0069 \\ -0.1458 & 0.4375 & -0.1667 & 0.0625 & -0.0208 \\ 0.0556 & -0.1667 & 0.4444 & -0.1667 & 0.0556 \\ -0.0208 & 0.0625 & -0.1667 & 0.4375 & -0.1458 \\ 0.0069 & -0.0208 & 0.0556 & -0.1458 & 0.3819 \end{bmatrix}. \end{aligned}$$

By computing the Fourier coefficients for g^{-1} through (1.2.1), we have

$$T_5(g^{-1}) \approx \begin{bmatrix} 0.4472 & -0.1708 & 0.0652 & -0.0249 & 0.0095 \\ -0.1708 & 0.4472 & -0.1708 & 0.0652 & -0.0249 \\ 0.0652 & -0.1708 & 0.4472 & -0.1708 & 0.0652 \\ -0.0249 & 0.0652 & -0.1708 & 0.4472 & -0.1708 \\ 0.0095 & -0.0249 & 0.0652 & -0.1708 & 0.4472 \end{bmatrix}.$$

The resulting preconditioned matrix is thus

$$T_5^{-1}(g)T_5(f) = \frac{1}{144}\begin{bmatrix} 422 & -381 & 200 & -75 & 26 \\ -402 & 567 & -456 & 225 & -78 \\ 208 & -456 & 592 & -456 & 208 \\ -78 & 225 & -456 & 567 & -402 \\ 26 & -75 & 200 & -381 & 422 \end{bmatrix}$$

$$\approx \begin{bmatrix} 2.9306 & -2.6458 & 1.3889 & -0.5208 & 0.1806 \\ -2.7917 & 3.9375 & -3.1667 & 1.5625 & -0.5417 \\ 1.4444 & -3.1667 & 4.1111 & -3.1667 & 1.4444 \\ -0.5417 & 1.5625 & -3.1667 & 3.9375 & -2.7917 \\ 0.1806 & -0.5208 & 1.3889 & -2.6458 & 2.9306 \end{bmatrix}.$$

The corresponding generated matrix by the same symbol $g^{-1}f$ is

$$T_5(g^{-1}f) \approx \begin{bmatrix} 4.1803 & -3.2705 & 1.6312 & -0.6231 & 0.2380 \\ -3.2705 & 4.1803 & -3.2705 & 1.6312 & -0.6231 \\ 1.6312 & -3.2705 & 4.1803 & -3.2705 & 1.6312 \\ -0.6231 & 1.6312 & -3.2705 & 4.1803 & -3.2705 \\ 0.2380 & -0.6231 & 1.6312 & -3.2705 & 4.1803 \end{bmatrix}.$$

Remark 2.2.2. For comparison reasons, to see how well our matrix-less method works, we need to compute the eigenvalues of the matrix $T_n^{-1}(g)T_n(f)$ as a reference solution. A standard procedure is to solve the generalized eigenvalue problem, $T_n(f)u = \lambda T_n(g)u$ where $\lambda_j(T_n^{-1}(g)T_n(f))$ are the solutions, with for example LAPACK. As noted, the matrix $T_n^{-1}(g)T_n(f)$ is full, but since we know that both $T_n(f)$ and $T_n(g)$ are banded, we can use Crawford's algorithm [21] and split Cholesky [48, 79], to transform the problem into a standard eigenvalue problem $A_n u = \lambda u$, where A_n has the same bandwidth as the largest bandwidth of $T_n(f)$ and $T_n(g)$. This standard eigenvalue problem is then solved efficiently. To compute the reference solutions in this thesis we use Julia with BandedMatrices.jl [56], which is a wrapper to LAPACK.

We now use Algorithm 1 to approximate the functions $c_k(\theta)$ in (2.2.1), but for the preconditioned matrix. Note again that now $r = g^{-1}f$ is no longer the generating function of the matrix in question, so in Step 1.2 we generate $T_n(f)$ and $T_n(g)$ instead of $T_n(r)$. In Step 1.3 we solve the generalized eigenvalue problem by `eig`$(T_n(f), T_n(g))$ instead of the more numerically sensitive and costly `eig`$(T_n^{-1}(g)T_n(f))$.

We have the same behavior of the errors in the left and right panels of Figure 2.2.2 as in the equivalent Figure 2.1.3, but with the new symbol $r = g^{-1}f$.

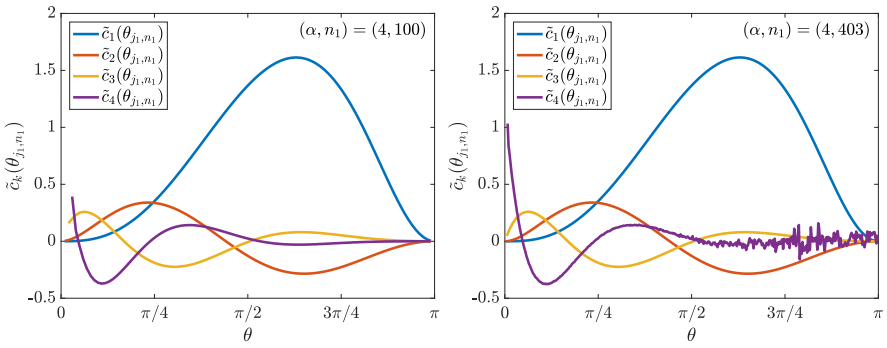


Figure 2.2.2. Example 2.2.1: Preconditioned matrix $T_n^{-1}(g)T_n(f)$. The approximations $\tilde{c}_k(\theta_{j,n_1})$, for $\alpha = 4$, and the symbol $r(\theta) = (2 - 2\cos(\theta))^2/(3 + 2\cos(\theta))$. Point-wise erratic approximations are removed. Left: Smooth behavior for all four $\tilde{c}_k$ when $n_1 = 100$. Right: For $n_1 = 403$, compare oscillatory behavior in Figure 2.1.3.

In Figure 2.2.3 we present the same computations as in Figure 2.1.6, but with the preconditioned matrix $T_{n_m}^{-1}(g)T_{n_m}(f)$. We thus have,

$$\tilde{\lambda}_{j_m}(T_{n_m}^{-1}(g)T_{n_m}(f)) = f(\theta_{j_1,n_1})/g(\theta_{j_1,n_1}) + \sum_{k=1}^{\alpha} \tilde{c}_k(\theta_{j_1,n_1})h_m^k, \qquad (2.2.2)$$

where $h_m = 1/(n_m + 1)$.

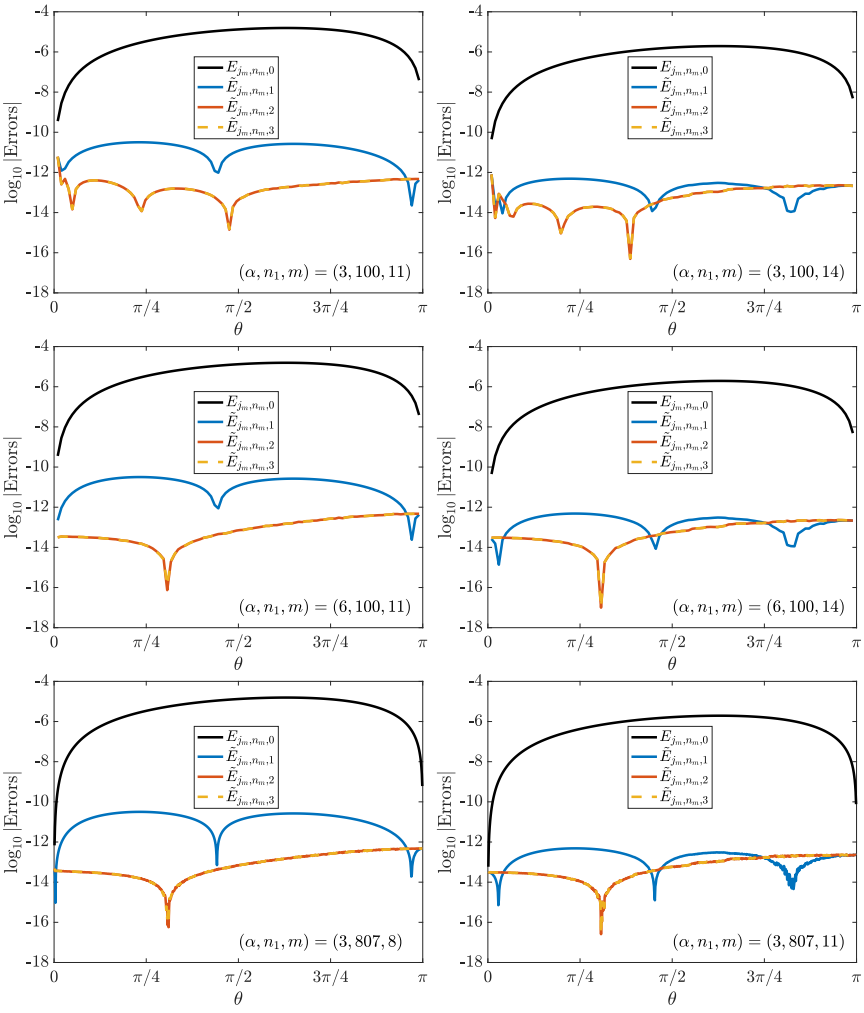
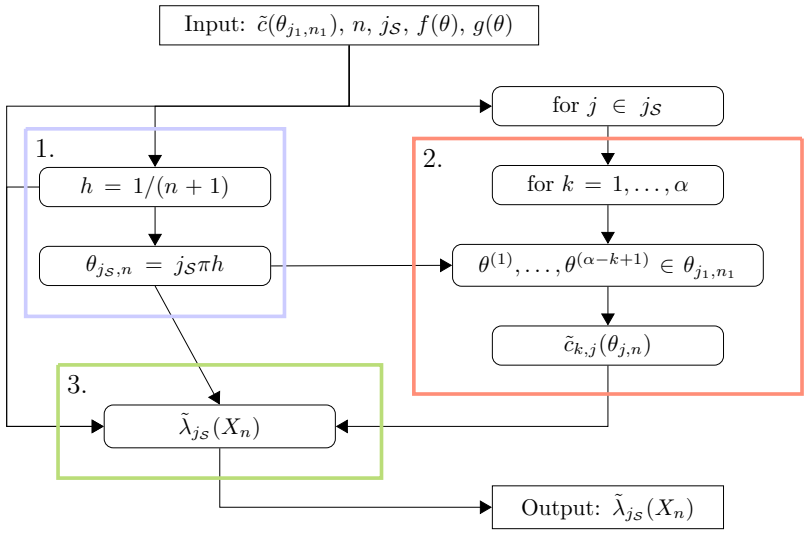


Figure 2.2.3. Example 2.2.1: Preconditioned matrix $T_n^{-1}(g)T_n(f)$. Reducing the errors for indices j_m, for matrix orders $n_m = 104323$ (left) and $n_m = 827391$ (right). Different (α, n_1, m). Errors $E_{j_m,n_m,0}$ (black) and reduced errors $\tilde{E}_{j_m,n_m,1}$ (blue), $\tilde{E}_{j_m,n_m,2}$ (red), and $\tilde{E}_{j_m,n_m,3}$ (yellow). Top: a reduction to $\mathcal{O}(10^{-13})$ is attained except close to zero. Middle: α increased; a reduction to $\mathcal{O}(10^{-13})$ is attained, also close to zero. Bottom: n_1 increased instead of α; a reduction to $\mathcal{O}(10^{-13})$ is attained for more eigenvalues, than in the middle panels, and at a lower cost.

2.3 Extending the Expansion to the Whole Spectrum

In Paper III we show how to utilize the computed $\tilde{c}_k(\theta_{j_1,n_1})$, from Algorithm 1, to accurately approximate a subset, or all, of the eigenvalues for a matrix of the same type and of arbitrary order n. The procedure is explained as a flowchart in Algorithm 2, and relies on efficient interpolation–extrapolation of the data $\tilde{c}_k(\theta_{j_1,n_1})$ to $\tilde{c}_{k,j_\mathcal{S}}(\theta_{j_\mathcal{S},n})$, where $j_\mathcal{S} \subseteq \{1,\dots,n\}$. As in Paper II, we allow the matrix to be of the form $X_n = T_n^{-1}(g)T_n(f)$, with symbol $r = g^{-1}f$.

Algorithm 2 Approximate the eigenvalues $\lambda_j(X_n) = \lambda_j(T_n^{-1}(g)T_n(f))$, for $j \in j_\mathcal{S} \subseteq \{1,\dots,n\}$ by interpolation–extrapolation of $\tilde{c}_k(\theta_{j_1,n_1})$ to $\tilde{c}_k(\theta_{j_\mathcal{S},n})$.



Step	Description
	Input to algorithm: $\tilde{c}_k(\theta_{j_1,n_1})$ from Algorithm 1, $n \in \mathbb{N}$, $j_\mathcal{S}$, $f(\theta)$, $g(\theta)$.
1.	Input: n and $j_\mathcal{S}$.
1.1	Define $h = 1/(n+1)$.
1.2	Define $\theta_{j_\mathcal{S},n} = j_\mathcal{S}\pi h$.
2.	Loop over indices $j \in j_\mathcal{S}$.
2.1	Loop over $k = 1,\dots,\alpha$.
2.1.1	Input: $\theta_{j_\mathcal{S},n}$ (1.2). Determine the $\alpha - k + 1$ points $\theta^{(1)},\dots,\theta^{(\alpha-k+1)} \in \theta_{j_1,n_1}$, which are closest to $\theta_{j,n}$.
2.1.2	Compute $\tilde{c}_{k,j}(\theta_{j,n})$, where $\tilde{c}_{k,j}(\theta)$ is interpolation polynomial of $\left(\theta^{(1)},\tilde{c}_k(\theta^{(1)})\right),\dots,\left(\theta^{(\alpha-k+1)},\tilde{c}_k(\theta^{(\alpha-k+1)})\right)$.
3	Input: $f(\theta)$, $g(\theta)$, h (1.1), $\theta_{j_\mathcal{S},n}$ (1.2), and $\tilde{c}_{k,j_\mathcal{S}}(\theta_{j_\mathcal{S},n})$ (2.1.2).
3.1	Compute $\tilde{\lambda}_{j_\mathcal{S}}(X_n) = f(\theta_{j_\mathcal{S},n})/g(\theta_{j_\mathcal{S},n}) + \sum_{k=1}^{\alpha} \tilde{c}_{k,j_\mathcal{S}}(\theta_{j_\mathcal{S},n})h^k$.
	Output of algorithm: $\tilde{\lambda}_{j_\mathcal{S}}(X_n)$.

Algorithm 2 outputs the approximation of the eigenvalues $\lambda_{j_S}(X_n)$ by a formula like (2.2.2), namely, in Step 3.1

$$\tilde{\lambda}_{j_S}(X_n) = f(\theta_{j_S,n})/g(\theta_{j_S,n}) + \sum_{k=0}^{\alpha} \tilde{c}_{k,j_S}(\theta_{j_S,n})h^k, \tag{2.3.1}$$

where $h = 1/(n+1)$. When r is non-monotone, Algorithm 2 should be slightly modified. Then we define a set j_a of admissible indices j for θ_{j,n_1} in Step 2.1.1 in Algorithm 2. The indices j in j_a are those for which the corresponding points θ_{j,n_1} lie in monotone regions of r. For example, in Figure 2.1.8 of Example 2.1.4 we have the red region where r is non-monotone, we define j_a such that $\theta_{j,n_1} \in [0, \hat{\theta}]$ or $\theta_{j,n_1} \in [\pi - \hat{\theta}, \pi]$, for all $j \in j_a$.

Example 2.3.1. We here further extend Examples 2.1.1 and Example 2.2.1, by interpolating–extrapolating the approximations $\tilde{c}_k$ to the whole spectrum, and use them to reduce the errors for all approximated eigenvalues of different large matrices. Hence, we use (2.3.1) to compute $\tilde{\lambda}_{j_S}(X_n)$ for the two examples.

In Figure 2.3.1 we show the symbol $f(\theta) = 6 - 8\cos(\theta) + 2\cos(\theta)$ from Example 2.1.1. By Algorithm 1 we compute $\tilde{c}_k$ by $\alpha = 3$ and $n_1 = 807$. We then use Algorithm 2 for different large n. In the left panel we have $n = 10^5$ and in the right panel $n = 10^6$. The black line shows the errors $E_{j,n,0} = \lambda_j(T_n(f)) - \tilde{\lambda}_j(T_n(f))$, where $\tilde{\lambda}_j(T_n(f)) = f(\theta_{j,n})$, given by (2.3.1) with $\alpha = 0$. The yellow line is $\tilde{E}_{j,n,3} = \lambda_j(T_n(f)) - \tilde{\lambda}_j(T_n(f))$, where $\tilde{\lambda}_j(T_n(f))$, given by (2.3.1) with $\alpha = 3$. To avoid clutter in the figures, we only show these two errors.

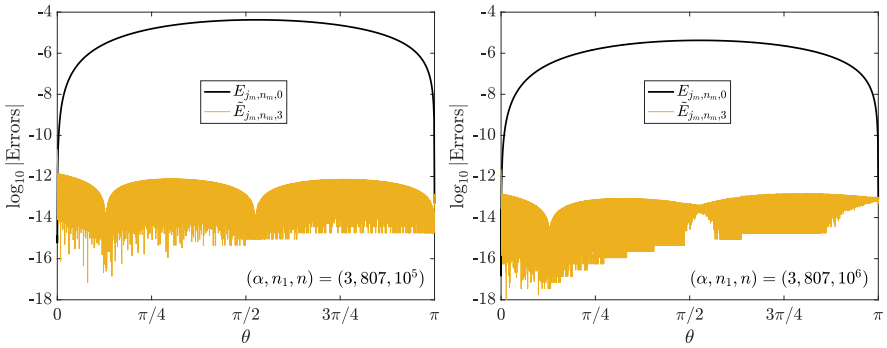


Figure 2.3.1. Example 2.3.1: Reducing errors for the whole spectrum of two large matrices $T_n(f)$. Symbol $f(\theta) = (2 - 2\cos(\theta))^2$. Left: $\tilde{c}_k$ computed with $\alpha = 3$ and $n_1 = 807$. Matrix order $n = 10^5$. Errors of order $\mathcal{O}(10^{-12})$. Right: $\tilde{c}_k$ computed with $\alpha = 3$ and $n_1 = 807$. Matrix order $n = 10^6$. Errors of order $\mathcal{O}(10^{-13})$.

In Figure 2.3.2 we make the same computations as in Figure 2.3.1, but for the symbol $f(\theta) = (6 - 8\cos(\theta) + 2\cos(\theta))/(3 + 2\cos(\theta))$ from Example 2.2.1.

Note that in all computations for Figures 2.1.6, 2.2.3, 2.3.1, and 2.3.2 the point-wise erratic expansion approximations $\tilde{c}_k$, described in Section 2.1, are not removed. The reason is to demonstrate their low impact on the solution.

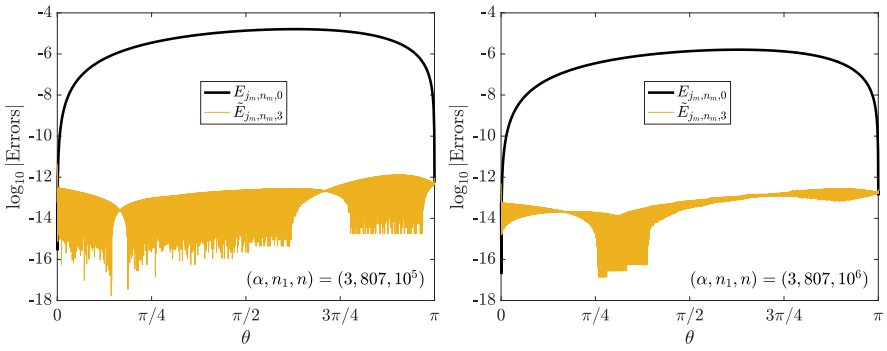


Figure 2.3.2. Example 2.3.1: Reducing errors for the whole spectrum of two large preconditioned matrices $T_n^{-1}(g)T_n(f)$. Symbol $r(\theta) = (2-2\cos(\theta))^2/(3+2\cos(\theta))$. Left: $\tilde{c}_k$ computed with $\alpha = 3$ and $n_1 = 807$. Matrix order $n = 10^5$. Errors of order $\mathcal{O}(10^{-12})$. Right: $\tilde{c}_k$ computed with $\alpha = 3$ and $n_1 = 807$. Matrix order $n = 10^6$. Errors of order $\mathcal{O}(10^{-13})$.

Remark 2.3.1. The timings presented in Paper III show that computing the spectrum using Algorithms 1 and 2 is faster than MATLAB's `eig` command. For the computations in Example 2.3.1 we use JULIA with BANDEDMATRICES.JL to call LAPACK, in order to solve the standard and generalized eigenvalue problems for banded matrices. Proper timing experiments have not been conducted, but for the examples of $n = 10^6$, LAPACK takes in the order of 10 hours of computations, whereas our matrix-less method takes in the order of 10 minutes, where the bottleneck is a naive sorting algorithm in Step 2.1.1 of the Algorithm 2. For computer specifications see the Colophon at the end of this thesis. A more efficient implementation of Algorithms 1 and 2, and taking into account algorithmic optimization and parallelization, should significantly reduce the execution time of the matrix-less method. The accuracy, as seen in Figures 2.3.1 and 2.3.2, is enough for many real world problems. Further implementation and research is warranted, for more complex symbols and discretizations.

2.4 Solving a Model Differential Eigenvalue Problem

In Paper IV we apply the results from Papers I–III to analyze and solve the discrete eigenvalue problem, resulting from the discretization of the Laplacian eigenproblem $-\Delta u = \lambda u$ with isogeometric analysis (IgA). In addition, we prove several spectral and structural properties of the associated discretization matrices.

In this section we focus on the one-dimensional version of the Laplacian eigenvalue problem, namely

$$\begin{cases} -u''(x) = \lambda u(x), & x \in (0,1), \\ u(x) = 0 & x \in \{0,1\}. \end{cases} \tag{2.4.1}$$

We want to find eigenvalues $\lambda \in \mathbb{R}_+$ and eigenfunctions $u \in H_0^1(0,1)$ such that, for all $v \in H_0^1(0,1)$,

$$\int_0^1 v'(x)u'(x)\mathrm{d}x = \lambda \int_0^1 v(x)u(x)\mathrm{d}x.$$

Discretizing this problem with IgA based on B-splines — for a more complete description of the method, see Paper IV and [36] — we have

$$K_n^{[p]}\mathbf{u}_n = \lambda M_n^{[p]}\mathbf{u}_n,$$

where the two matrices $K_n^{[p]}$ and $M_n^{[p]}$ are respectively the stiffness and the mass matrix. Here n is a size parameter and p is the degree of the used B-splines. The size of the matrices is $N \times N$ where $N = n + p - 2$. Multiplying both sides by the inverse of $M_n^{[p]}$, we obtain

$$L_n^{[p]}\mathbf{u}_n = \lambda \mathbf{u}_n,$$

where $L_n^{[p]} = (M_n^{[p]})^{-1}K_n^{[p]}$ is the preconditioned matrix whose eigenvalues converge to the eigenvalues of the original problem (2.4.1) as $n \to \infty$. From [40, Section 10.7] we know that the normalized matrices $n^{-1}K_n^{[p]}$ and $nM_n^{[p]}$ have an approximate Toeplitz structure. Indeed,

$$\begin{aligned} n^{-1}K_n^{[p]} &= T_{n+p-2}(f_p) + R_n^{[p]}, & \operatorname{rank}(R_n^{[p]}) &\le 4(p-1), \\ nM_n^{[p]} &= T_{n+p-2}(g_p) + V_n^{[p]}, & \operatorname{rank}(V_n^{[p]}) &\le 4(p-1), \end{aligned}$$

where f_p and g_p are RCTPs which depend on the spline degree p. In addition,

$$\begin{aligned} \{n^{-1}K_n^{[p]}\}_n &\sim_{\text{GLT},\sigma,\lambda} f_p, \\ \{nM_n^{[p]}\}_n &\sim_{\text{GLT},\sigma,\lambda} g_p, \\ \{n^{-2}L_n^{[p]}\}_n &\sim_{\text{GLT},\sigma,\lambda} g_p^{-1}f_p = e_p. \end{aligned}$$

Assuming the asymptotic expansion for the eigenvalues of the normalized and preconditioned matrix $n^{-2}L_n^{[p]}$, we use Algorithms 1 and 2, with slight modifications, to approximate $c_k(\theta)$ in (2.2.1). For $p > 2$ there are outliers among the eigenvalues of $n^{-2}L_n^{[p]}$, that is, eigenvalues that lie outside the range of e_p. The number of outliers is $n_p^{\text{out}} = p + \operatorname{mod}(p,2) - 2$, and each outlier always has multiplicity two. The outliers are constant to machine precision even for

small n, so they are easily computed, and then used for any large n. We denote the number of eigenvalues that are not outliers by

$$\hat{N}_p = N - n_p^{\text{out}} = n - \text{mod}(p, 2).$$

In Paper IV it is proved that an approximation of all the eigenvalues, that are not outliers, is obtained by sampling the symbol with the following grid,

$$\theta_{j,n,p} = \frac{j\pi}{n}, \qquad j = 1, \dots, \hat{N}_p. \tag{2.4.2}$$

Using the notation of Table 1.5.1, this is a τ_{n-1}-grid for odd p and τ_{n-1}^{π}-grid for even p. For the two cases $p = 1, 2$, the grid (2.4.2) gives the exact eigenvalues when sampling the symbol. To the best of the authors' knowledge, the result for $p = 2$ was not known before Paper IV. Indeed, the matrices for $p = 1$ belong to the $\tau_{n-1}(0, 0)$-algebra, and the matrices for $p = 2$ belong to the $\tau_n(-1, -1)$-algebra [15].

Example 2.4.1. In this example, we investigate some structural properties of the stiffness and mass matrices from the IgA discretization of the Laplacian eigenproblem, when $p = 2$ and $n = 5$ (the following is true for any n). This presents a slightly different viewpoint than in Section 2.3 of Paper IV, and rather reflects a practical example how the connection to the $\tau_n(-1, -1)$-algebra was discovered. The normalized stiffness matrix is

$$\begin{aligned} 5^{-1}K_5^{[2]} &= \frac{1}{6}\begin{bmatrix} 8 & -1 & -1 & 0 & 0 \\ -1 & 6 & -2 & -1 & 0 \\ -1 & -2 & 6 & -2 & -1 \\ 0 & -1 & -2 & 6 & -1 \\ 0 & 0 & -1 & -1 & 8 \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} 3 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 3 \end{bmatrix}}_{F_5} \underbrace{\frac{1}{6}\begin{bmatrix} 3 & 1 & 0 & 0 & 0 \\ 1 & 4 & 1 & 0 & 0 \\ 0 & 1 & 4 & 4 & 0 \\ 0 & 0 & 1 & 4 & 1 \\ 0 & 0 & 0 & 1 & 3 \end{bmatrix}}_{G_5}, \end{aligned}$$

where the red elements indicate deviations from the pure Toeplitz structures. We have $\{n^{-1}K_n^{[2]}\}_n \sim_{\text{GLT}} f_2$, where

$$f_2(\theta) = 1 - \frac{2}{3}\cos(\theta) - \frac{1}{3}\cos(2\theta) = \underbrace{(2 - 2\cos(\theta))}_{f_1}\underbrace{\frac{1}{6}(4 + 2\cos(\theta))}_{g_1}.$$

The Toeplitz-like matrices F_5 and G_5 are

$$F_5 = T_5(f_1) + R_5^f,$$
$$G_5 = T_5(g_1) + R_5^g,$$

where R_5^f and R_5^g are the low-rank corrections that change $T_5(f_1)$ and $T_5(g_1)$, which belong to the $\tau_5(0,0)$-algebra, into the respective matrices F_5 and G_5, which belong to the $\tau_5(-1,-1)$-algebra. The non-zero elements of R_5^f and R_5^g are

$$(R_5^f)_{1,1} = (R_5^f)_{5,5} = -(T_5(f_1))_{1,2} = 1,$$
$$(R_5^g)_{1,1} = (R_5^g)_{5,5} = -(T_5(g_1))_{1,2} = -1/6.$$

The same decomposition is done for the normalized mass matrix,

$$5M_5^{[2]} = \frac{1}{120}\begin{bmatrix} 40 & 25 & 1 & 0 & 0 \\ 25 & 66 & 26 & 1 & 0 \\ 1 & 26 & 66 & 26 & 1 \\ 0 & 1 & 26 & 66 & 25 \\ 0 & 0 & 1 & 25 & 40 \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} \alpha-\gamma & \gamma & 0 & 0 & 0 \\ \gamma & \alpha & \gamma & 0 & 0 \\ 0 & \gamma & \alpha & \gamma & 0 \\ 0 & 0 & \gamma & \alpha & \gamma \\ 0 & 0 & 0 & \gamma & \alpha-\gamma \end{bmatrix}}_{G_5^{(1)}} \underbrace{\begin{bmatrix} \beta-\gamma & \gamma & 0 & 0 & 0 \\ \gamma & \beta & \gamma & 0 & 0 \\ 0 & \gamma & \beta & \gamma & 0 \\ 0 & 0 & \gamma & \beta & \gamma \\ 0 & 0 & 0 & \gamma & \beta-\gamma \end{bmatrix}}_{G_5^{(2)}},$$

where $\alpha = (13 - \sqrt{105})/\sqrt{120}$, $\beta = (13 + \sqrt{105})/\sqrt{120}$, and $\gamma = 1/\sqrt{120}$. We have $\{nM_n^{[2]}\}_n \sim_{\text{GLT}} g_2$, where

$$g_2(\theta) = \frac{1}{120}\left(66 + 52\cos(\theta) + 2\cos(2\theta)\right)$$
$$= \underbrace{\frac{1}{\sqrt{120}}\left(13 - \sqrt{105} + 2\cos(\theta)\right)}_{g_2^{(1)}} \underbrace{\frac{1}{\sqrt{120}}\left(13 + \sqrt{105} + 2\cos(\theta)\right)}_{g_2^{(2)}}.$$

Thus, $5M_5^{[2]}$ also belongs to the $\tau_5(-1,-1)$-algebra. Since the τ-algebras are closed under inversion, and $5M_5^{[2]}$ is symmetric positive definite, $(5M_5^{[2]})^{-1}$ exists and belongs to the $\tau_5(-1,-1)$-algebra. Hence, also $5^{-2}L_5^{[2]} = (5M_5^{[2]})^{-1}5^{-1}K_5^{[2]}$ belongs to the same algebra. This is why we can sample the symbol $e_2(\theta) = f_2(\theta)/g_2(\theta)$ to attain the exact eigenvalues of $L_5^{[2]}$, by the use of the τ_4^{π}-grid.

In Paper IV we also show the successful application of Algorithms 1 and 2 for the IgA matrices $n^{-1}K_n^{[p]}$, $nM_n^{[p]}$, and $n^{-2}L_n^{[p]}$ for $p > 2$, with slight modifications since we do not use the standard τ_n-grid.

A last interesting fact regarding the symbols of the normalized IgA matrices $n^{-1}K_n^{[p]}$ and $nM_n^{[p]}$ is that we do not just have a closed form expression for f_p (the symbol associated with $n^{-1}K_n^{[p]}$), but we also have a general expression for g_p (the symbol associated with $nM_n^{[p]}$). Indeed,

$$\begin{aligned} f_p(\theta) &= (2 - 2\cos(\theta))g_{p-1}(\theta), \\ g_p(\theta) &= \frac{1}{(2p+1)!}\sum_{k=0}^{2p} \left\langle {2p+1 \atop k} \right\rangle \cos((p-k)\theta). \end{aligned}$$

Here,

$$\left\langle {n \atop m} \right\rangle = \sum_{k=0}^{m+1} (-1)^k \binom{n+1}{k} (m+1-k)^n$$

is the formula for the Eulerian numbers [33], and

$$(2p+1)! = \sum_{k=0}^{2p} \left\langle {2p+1 \atop k} \right\rangle.$$

This result regarding the symbol g_p was previously known, but was rediscovered by the author in part by the use of [69]. The relationship between the Eulerian numbers, splines, and IgA is discussed, for example, in [46, 49, 78]. Moreover, the factorization $f_p(\theta) = (2 - 2\cos(\theta))g_{p-1}(\theta)$ was originally proved in [36].

2.5 Some Analytical Results

In Paper V we present formulae for the exact eigenvalues, and the corresponding eigenvectors, for a special class of matrices that we call symmetrically sparse tridiagonal (SST). These matrices have a symbol of form

$$f(\theta) = \hat{f}_0 + \hat{f}_\omega e^{\mathbf{i}\omega\theta} + \hat{f}_{-\omega} e^{-\mathbf{i}\omega\theta},$$

where $\hat{f}_0, \hat{f}_\omega, \hat{f}_{-\omega} \in \mathbb{C}$, $0 < \omega < n$, and $\omega \in \mathbb{Z}_+$. In Paper V, we also put this result into a wider context, that is, $f(\theta)$ is a non-monotone symbol of which we know the exact sampling grid for computing the eigenvalues of $T_n(f)$. The nth Toeplitz matrix generated by f is shown below to the left,

$$T_n(f) = \begin{bmatrix} \hat{f}_0 & 0 & \cdots & 0 & \hat{f}_{-\omega} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \hat{f}_{-\omega} \\ \hat{f}_\omega & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \hat{f}_\omega & 0 & \cdots & 0 & \hat{f}_0 \end{bmatrix} \sim \begin{bmatrix} \hat{g}_0 & 0 & \cdots & 0 & \hat{g}_\omega & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \hat{g}_\omega \\ \hat{g}_\omega & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \hat{g}_\omega & 0 & \cdots & 0 & \hat{g}_0 \end{bmatrix} = T_n(g_\omega).$$

By symmetrizing the matrix $T_n(f)$, we attain a similar matrix, sharing all the eigenvalues of $T_n(f)$. This can be done in the same way as in the tridiagonal case presented in (1.5.2); that is, we define a new symbol $g_\omega(\theta) = \hat{g}_0 + 2\hat{g}_\omega \cos(\omega\theta)$, where $\hat{g}_0 = \hat{f}_0$ and $\hat{g}_\omega = \sqrt{\hat{f}_\omega}\sqrt{\hat{f}_{-\omega}}$. The nth Toeplitz matrix generated by g_ω is shown above, to the right.

In Paper V we construct a grid that, when sampling a third symbol g, gives the exact eigenvalues of $T_n(f) \sim T_n(g_\omega)$. The new modified symbol is defined as $g(\theta) = \hat{g}_0 + 2\hat{g}_\omega \cos(\theta)$, that is, we have removed the constant ω from the symbol $g_\omega(\theta)$. Here follows the construction of this special grid, for a given ω and n. First define

$$\beta = \operatorname{mod}(n, \omega), \quad n_\omega = \frac{n - \beta}{\omega}. \tag{2.5.1}$$

Then define the two sub-grids, τ_{n_ω} and $\tau_{n_{\omega+1}}$,

$$\theta_{j,n_\omega} = \frac{j\pi}{n_\omega + 1}, \quad j = 1, \ldots, n_\omega, \tag{2.5.2}$$

$$\theta_{j,n_\omega+1} = \frac{j\pi}{n_\omega + 2}, \quad j = 1, \ldots, n_\omega + 1. \tag{2.5.3}$$

The matrices of interest have eigenvalues of different multiplicity depending on n and ω in the following way. We have the two sub-grids repeated as follows

$$\tilde{\theta}^{(1)}_{r_1,j,n_\omega(\omega-\beta)} = \theta_{j,n_\omega}, \quad r_1 = 1, \ldots, \omega - \beta, \quad j = 1, \ldots, n_\omega,$$
$$\tilde{\theta}^{(1)}_{r_2,j,(n_\omega+1)\beta} = \theta_{j,n_\omega+1}, \quad r_2 = 1, \ldots, \beta, \quad j = 1, \ldots, n_\omega + 1,$$

that is, the first grid (2.5.2) is repeated $\omega - \beta$ times, which when sampling $g(\theta)$ gives eigenvalues of this multiplicity. The second grid is repeated β times. Now define the following two grids,

$$\tilde{\theta}^{(1)}_{n_\omega(\omega-\beta)} = \left\{ \left\{ \tilde{\theta}^{(1)}_{r_1,j,n_\omega(\omega-\beta)} \right\}_{r_1=1}^{\omega-\beta} \right\}_{j=1}^{n_\omega},$$
$$\tilde{\theta}^{(2)}_{(n_\omega+1)\beta} = \left\{ \left\{ \tilde{\theta}^{(2)}_{r_2,j,(n_\omega+1)\beta} \right\}_{r_2=1}^{\beta} \right\}_{j=1}^{n_\omega+1},$$

of which the union is our final sampling grid for $g(\theta)$,

$$\tilde{\theta}_n = \tilde{\theta}^{(1)}_{n_\omega(\omega-\beta)} \bigcup \tilde{\theta}^{(2)}_{(n_\omega+1)\beta}. \tag{2.5.4}$$

Note again that the grid (2.5.4) typically consists of many grid points with the same value. An addition to Paper V is that, for cases when it is desirable to retain the original symmetrized symbol $g_\omega(\theta)$, instead of $g(\theta)$, and to have a corresponding grid over $\theta_n \in [0, \pi]$, we can use the following (non-unique) grid construction. This new grid has the advantage of not having any grid points of multiplicity greater than one, and is used for example in [31]. We have the same definition of the parameters β and n_ω as in (2.5.1), and the same sub-grids as in (2.5.2) and (2.5.3). We now introduce the notation $n_1 = n_\omega(\omega - \beta)$, and $n_2 = (n_\omega + 1)\beta$, for the total number of grid points corresponding to the two sub-grid types, and $n = n_1 + n_2$. Then,

$$\theta^{(1)}_{n_1} = \left\{ \bigcup_{r_1=1}^{\omega-\beta} (\theta_{n_\omega} + (r_1 - 1)\pi) \right\},$$
$$\theta^{(2)}_{n_2} = \left\{ \bigcup_{r_2=1}^{\beta} (\theta_{n_\omega+1} + (r_2 - 1)\pi + (\omega - \beta)\pi) \right\}.$$

The final grid, now associated with $g_\omega(\theta)$, is defined as

$$\theta_n = \frac{1}{\omega} \left\{ \theta^{(1)}_{n_1}, \theta^{(2)}_{n_2} \right\}. \tag{2.5.5}$$

The corresponding eigenvectors are constructed as follows, where the non-zero components $x^{(j,n)}_k$ of an eigenvector $\mathbf{x}_j$, for the eigenpair $(\lambda_j, \mathbf{x}_j)$, are defined below. For $j = 1, \ldots, n_1$ we have

$$r_1 = \frac{j + \mathrm{mod}(n_1 - j, n_\omega)}{n_\omega},$$
$$x^{(j,n)}_{\omega(k_1-1)+r_1+\beta} = \left(\frac{\sqrt{f_\omega}}{\sqrt{f_{-\omega}}} \right)^{k_1} \sin\left(\omega k_1 \theta_{j,n}\right), \quad k_1 = 1, \ldots, n_\omega, \tag{2.5.6}$$

and for $j = n_1 + 1, \ldots, n$ we have

$$r_2 = \frac{j - n_1 + \mathrm{mod}(n - j, n_\omega + 1)}{n_\omega + 1},$$
$$x^{(j,n)}_{\omega(k_2-1)+r_2} = \left(\frac{\sqrt{f_\omega}}{\sqrt{f_{-\omega}}} \right)^{k_2} \sin\left(\omega k_2 \theta_{j,n}\right), \quad k_2 = 1, \ldots, n_\omega + 1. \tag{2.5.7}$$

Example 2.5.1. We illustrate the use of the grid (2.5.5) with the following symbol

$$g_5(\theta) = 2 - 2\cos(5\theta).$$

Since $\omega = 5$, and by choosing $n = 17$, we obtain $\beta = 2$ and $n_\omega = 3$ from (2.5.1). We thus have three $(\omega - \beta)$ sub-grids derived from θ_{n_ω} of (2.5.2) and two (β) sub-grids derived from $\theta_{n_\omega+1}$ of (2.5.3). The complete grid, defined by (2.5.5) on $[0, \pi]$ has these five sub-grids; in Figure 2.5.1 we see the sub-intervals associated with θ_{n_ω} in yellow, and the sub-intervals associated with $\theta_{n_\omega+1}$ in green.

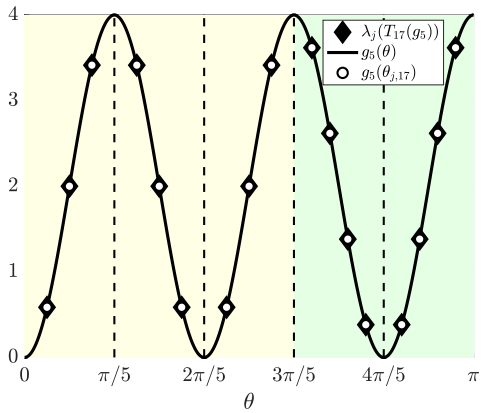


Figure 2.5.1. Example 2.5.1: Symbol $g_5(\theta) = 2 - 2\cos(5\theta)$, for $n = 17$ and $\omega = 5$, and thus, $\beta = 2$ and $n_\omega = 3$. Intervals colored for sub-grids of size n_ω (yellow) and $n_\omega + 1$ (green). Eigenvalues, $\lambda_j(T_{17}(g_5))$ (black diamonds) and samplings of the symbol, $g_5(\theta_{j,5})$ (white circles), with grid $\theta_{j,5}$ from (2.5.5), match exactly.

If the grids (2.5.4) or (2.5.5) were not known, and we use the standard τ_n-grid to estimate the eigenvalues, we have errors, as described in Paper V. Figure 2.5.2 shows the errors $E_{j,n,0}$, for $n = 102$, for two different permutations. The order n is chosen such that we have the same $\beta = \mathrm{mod}(n, \omega) = 2$ as in Figure 2.5.1. The left panel presents the "error modes" described in Section 3.1 of Paper V, that is, we order the samplings $g_\omega(\theta_{j,n})$ in a non-decreasing order, by the permutation σ, and compute the errors $E_{j,n,0}$. We define the $\omega = 5$ error modes by associating the indices $j_q = q, q + \omega, q + 2\omega, \ldots$ for $q = 1, \ldots, 5$ to the respective error mode $E^{\{q\}} = E_{j_q,n,0}$. This ordering of the errors gave a hint of the existence, and consequently the discovery, of the grids (2.5.4) and (2.5.5). As is seen in Figure 2 of Paper V, we observe this oscillatory behavior also for some other non-monotone cases, albeit in some more complex fashion, which warrants further research. The right panel of Figure 2.5.2 shows the errors $E_{j,n,0}$ when the eigenvalues are ordered as the samplings $g_\omega(\theta_{j,n})$, with a permutation $\rho = \sigma^{-1}$. By the coloring of the errors we indicate what error mode the individual error belongs to; i.e. $E^{\{5\}}_{j,20}$, $E^{\{1\}}_{j,21}$, $E^{\{4\}}_{j,20}$, $E^{\{2\}}_{j,21}$, and $E^{\{3\}}_{j,20}$ as shown in the left panel. The background coloring indicates, as in Figure 2.5.1, the number of grid points in the interval, that is, n_ω (yellow) or $n_\omega + 1$ (green).

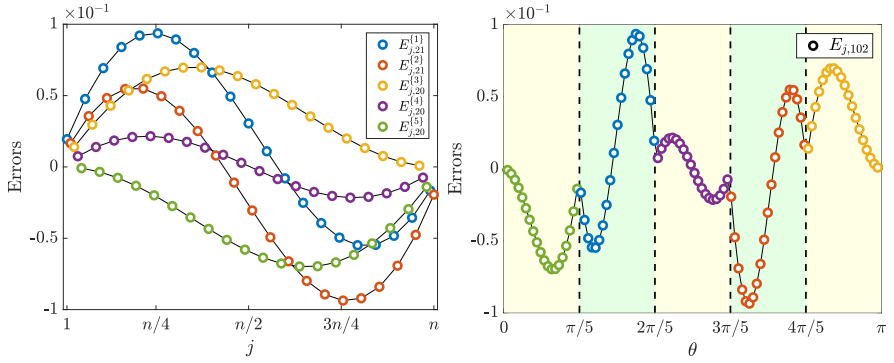


Figure 2.5.2. Example 2.5.1: Different errors depending on ordering, with symbol $g_5(\theta) = 2 - 2\cos(5\theta)$. Left: Permutation of $g_5(\theta_{j,n})$, by σ, to match the ordering of $\lambda_j(T_n(f))$. Errors $E_{j,102}$ are split into five error modes. Right: Permutation of $\lambda_j(T_n(f))$, by $\rho = \sigma^{-1}$, to match the ordering of $g_5(\theta_{j,n})$. The colors of the circles indicate, for each error, the corresponding error mode in the left panel.

An alternative viewpoint for explaining the grid (2.5.4), originally presented in Paper V, is the following: Consider the symbol

$$f(\theta) = \hat{f}_0 + 2\hat{f}_1 \cos(3\theta),$$

that is, $\omega = 3$. Now generate the Toeplitz matrix of order $n = 9$,

$$T_9(f) = \left[\begin{array}{ccc|ccc|ccc} \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 \\ \hline \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 \\ 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 \\ 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 \\ \hline 0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 \end{array}\right].$$

We have $\beta = 0$ and $n_\omega = 3$ from (2.5.1). As indicated by the coloring, we view the matrix $T_9(f)$ as a block matrix where the blocks are of order $s = \omega = 3$. We thus have $T_9(f) = T_3(\mathbf{f}^{(3)})$, where

$$\mathbf{f}^{(3)}(\theta) = \hat{\mathbf{f}}_0^{(3)} + \hat{\mathbf{f}}_1^{(3)} \cos(\theta) = \left[\begin{array}{ccc} \hat{f}_0 & 0 & 0 \\ 0 & \hat{f}_0 & 0 \\ 0 & 0 & \hat{f}_0 \end{array}\right] + 2 \left[\begin{array}{ccc} \hat{f}_1 & 0 & 0 \\ 0 & \hat{f}_1 & 0 \\ 0 & 0 & \hat{f}_1 \end{array}\right] \cos(\theta)$$

$$= (\hat{f}_0 + 2\hat{f}_1 \cos(\theta))\mathbb{I}_3.$$

The eigenvalues are given exactly by sampling the symbol $\mathbf{f}$ with the τ_{n_ω}-grid, given in (2.5.2). By a similarity transformation, we obtain a block diagonal matrix $\tilde{T}_9(f) = P^{-1}T_9(f)P \sim T_9(f)$, by some matrix P, for example

$$P = \begin{bmatrix} 0 & 0 & 9 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 9 \\ 0 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6 & 0 \\ 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 \end{bmatrix}.$$

We then have,

$$\tilde{T}_9(f) = \left[\begin{array}{ccc|ccc|ccc} \hat{f}_0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hat{f}_1 & \hat{f}_0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \hat{f}_1 & \hat{f}_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & \hat{f}_0 & \hat{f}_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \hat{f}_1 & \hat{f}_0 & \hat{f}_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \hat{f}_1 & \hat{f}_0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & \hat{f}_0 & \hat{f}_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \hat{f}_1 & \hat{f}_0 & \hat{f}_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat{f}_1 & \hat{f}_0 \end{array}\right],$$

which can be viewed as a block diagonal matrix of the form

$$\tilde{T}_9(f) = \begin{bmatrix} B^{(1)} & & \\ & B^{(1)} & \\ & & B^{(1)} \end{bmatrix},$$

where the block $B^{(1)}$ is

$$B^{(1)} = \left[\begin{array}{ccc} \hat{f}_0 & \hat{f}_1 & 0 \\ \hat{f}_1 & \hat{f}_0 & \hat{f}_1 \\ 0 & \hat{f}_1 & \hat{f}_0 \end{array}\right].$$

Hence, the multiplicity of eigenvalues is the same as the number of blocks which is three ($\omega - \beta = 3$), and the exact eigenvalues of one block $B^{(1)}$ is given, since it is tridiagonal, by sampling the symbol $g(\theta) = \hat{f}_0 + 2\hat{f}_1 \cos(\theta_{j,3})$, where $\theta_{j,3}$ corresponds to (2.5.2), with $n_\omega = 3$.

We now consider the generated matrix of order $n = 10$, that is, $\beta = 1$ and

$$T_{10}(f) = \left[\begin{array}{ccc|ccc|ccc|c}
\hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 \\
\hline
\hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 \\
0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 \\
0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 \\
\hline
0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 \\
0 & 0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0
\end{array}\right].$$

It is not possible to write it as a block Toeplitz matrix with blocks of order $s = 3$. We "extend" the matrix $T_{10}(f)$ with two rows and two columns of zeros, and call the new matrix $\bar{T}_{10}(f)$. This does not change the spectrum of the matrix, except adding two zeros. Thus,

$$\bar{T}_{10}(f) = \left[\begin{array}{ccc|ccc|ccc|cc}
\hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 \\
\hline
\hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 \\
0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 \\
0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 \\
\hline
0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 \\
0 & 0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right].$$

The matrix $\bar{T}_{10}(f)$ is not a Toeplitz or block Toeplitz matrix, but Toeplitz-like. We now use a new matrix P such that $\tilde{\bar{T}}_{10}(f) = P^{-1}\bar{T}_{10}(f)P \sim \bar{T}_{10}(f)$, so

$$\tilde{\bar{T}}_{10}(f) = \left[\begin{array}{cccc|cccc|cccc}
\hat{f}_0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hat{f}_1 & \hat{f}_0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \hat{f}_1 & \hat{f}_0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \hat{f}_1 & \hat{f}_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & \hat{f}_0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \hat{f}_1 & \hat{f}_0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \hat{f}_1 & \hat{f}_0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat{f}_0 & \hat{f}_1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat{f}_1 & \hat{f}_0 & \hat{f}_1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat{f}_1 & \hat{f}_0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right],$$

which is block diagonal, with the two blocks

$$\bar{B}^{(1)} = \left[\begin{array}{cccc}
\hat{f}_0 & \hat{f}_1 & 0 & 0 \\
\hat{f}_1 & \hat{f}_0 & \hat{f}_1 & 0 \\
0 & \hat{f}_1 & \hat{f}_0 & 0 \\
0 & 0 & 0 & 0
\end{array}\right], \quad B^{(2)} = \left[\begin{array}{cccc}
\hat{f}_0 & \hat{f}_1 & 0 & 0 \\
\hat{f}_1 & \hat{f}_0 & \hat{f}_1 & 0 \\
0 & \hat{f}_1 & \hat{f}_0 & \hat{f}_1 \\
0 & 0 & \hat{f}_1 & \hat{f}_0
\end{array}\right],$$

and

$$\tilde{\bar{T}}_{10}(f) = \begin{bmatrix} B^{(2)} & & \\ & \bar{B}^{(1)} & \\ & & \bar{B}^{(1)} \end{bmatrix}.$$

The block $\bar{B}^{(1)}$ is the "extended" block $B^{(1)}$, with an extra row and column of zeros. The block $\bar{B}^{(1)}$ is repeated twice ($\omega - \beta = 2$) and the block $B^{(2)}$ is repeated once ($\beta = 1$). The eigenvalues of $\bar{B}^{(1)}$ are the same, except for a zero, as for $B^{(1)}$, so we use the $\tau_{n_\omega}-$grid in (2.5.2), where $n_\omega = 3$, to sample $g(\theta)$. The eigenvalues of $B^{(2)}$ are given by sampling $g(\theta)$ with the $\tau_{n_\omega+1}$-grid in (2.5.3).

Finally, we study the generated matrix of order $n = 11$, for which $\beta = 2$,

$$T_{11}(f) = \left[\begin{array}{ccc|ccc|ccc|cc}
\hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 \\
\hline
\hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 \\
0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 \\
0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 \\
\hline
0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 \\
0 & 0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 \\
0 & 0 & 0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0
\end{array}\right].$$

Extending the matrix $T_{11}(f)$ with one row and one column of zeros, in a similar way as for extending $T_{10}(f)$ to $\bar{T}_{10}(f)$, we obtain the matrix

$$\bar{T}_{11}(f) = \left[\begin{array}{ccc|ccc|ccc|ccc}
\hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
\hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 \\
0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 & 0 \\
0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 & 0 \\
0 & 0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & \hat{f}_1 & 0 \\
0 & 0 & 0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat{f}_1 & 0 & 0 & \hat{f}_0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right].$$

Again we follow the same procedure as for $\bar{T}_{10}(f)$, that is, we choose a matrix P such that $\bar{\tilde{T}}_{11}(f) = P^{-1}\bar{T}_{11}(f)P \sim \bar{T}_{11}(f)$, where

$$\bar{\tilde{T}}_{11}(f) = \left[\begin{array}{cccc|cccc|cccc}
\hat{f}_0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hat{f}_1 & \hat{f}_0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \hat{f}_1 & \hat{f}_0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \hat{f}_1 & \hat{f}_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & \hat{f}_0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \hat{f}_1 & \hat{f}_0 & \hat{f}_1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \hat{f}_1 & \hat{f}_0 & \hat{f}_1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \hat{f}_1 & \hat{f}_0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat{f}_0 & \hat{f}_1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat{f}_1 & \hat{f}_0 & \hat{f}_1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \hat{f}_1 & \hat{f}_0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right],$$

that has the block structure

$$\bar{\tilde{T}}_{11}(f) = \begin{bmatrix} B^{(2)} & & \\ & B^{(2)} & \\ & & \bar{B}^{(1)} \end{bmatrix}.$$

The previously defined block $\bar{B}^{(1)}$ is repeated once ($\omega - \beta = 1$), and block $B^{(2)}$ is repeated twice ($\beta = 2$). Again, the grids for the blocks correspond to (2.5.4), and is used to sample $g(\theta)$ for obtaining the exact eigenvalues. For $n = 12$ we are again back to the case $\beta = 0$. This procedure can be done for any combination of ω and n.

3. Conclusions and Future Works

Snerik räknar ägg!

RITA VON HOFSTEN

This thesis presents various aspects of the conjectured asymptotic expansion for the eigenvalues of large Toeplitz-like matrices. This topic was first addressed in Paper I for the case of pure Toeplitz matrices. In Paper II we extended the results from Paper I to the case of preconditioned Toeplitz matrices. In Paper III we developed an efficient "matrix-less" and parallel interpolation–extrapolation algorithm, based on the asymptotic expansion, for computing the whole spectrum of large preconditioned Toeplitz matrices. In Paper IV we applied the results from Papers I–III to the Toeplitz-like matrices arising from the isogeometric analysis (IgA) discretization of second-order differential problems; we also computed the exact eigenvalues of these matrices in some special cases. In Paper V we derived closed formulae for the exact eigenvalues and eigenvectors of so-called "symmetrically sparse tridiagonal" Toeplitz matrices.

We conclude this thesis by listing a few topics of interest for future research, which are closely related to the content of the present thesis.

1. Extend the results of [31] for preconditioned block Toeplitz-like matrices.
2. Explore the existence of expansions for the eigenvalues of Toeplitz-like matrices related to non-monotone spectral symbols; develop appropriate techniques to exploit these expansions for the design of fast interpolation–extrapolation algorithms for computing the spectrum of such matrices. Note that this was implemented for monotone symbols in Papers I–IV.
3. Perform the previous item for multivariate Toeplitz-like matrices.
4. Study so-called "non-symmetrically sparse tridiagonal" Toeplitz matrices, in analogy with what has been done in Paper V for "symmetrically sparse tridiagonal" Toeplitz matrices.

Beside these specific items, there are of course many other open questions regarding the spectrum of Toeplitz and Toeplitz-like matrices. We here mention a few additional topics of interest, related to the contents of this thesis in a broader sense: (a) the application of the results presented herein to improve existing solution methods, for example deflation techniques, multigrid, and iterative solvers such as the Chebyshev iteration method; (b) the study of outlier eigenvalues for Toeplitz-like matrices; (c) a deeper study of "optimal" and exact sampling grids for approximating the spectrum of various Toeplitz and Toeplitz-like matrices; (d) further studies using high precision computations.

4. Acknowledgments

The journey to complete this thesis has been a long and winding road, or a short sprint, depending on one's perspective. It took almost twelve years, or less than two, starting with the work on vertex-centered Discontinuous Galerkin methods (DG), ending up with a thesis on eigenvalues of large matrices.

It is impossible to acknowledge everybody that has had an impact on this thesis in one way or another, so collectively to all of you, thank you!

First of all, I want to express my profound gratitude to my first main advisor, Martin Berggren, for giving me the opportunity to begin my scientific journey which led to this thesis. You taught me not to trust "known truths", and this attitude has been a requisite for many of the discoveries in this thesis. Secondly, I am deeply thankful to my second main advisor, Maya Neytcheva, who explained to me so many "known truths", that I should have known, and thus widened my perspectives. You introduced me to the beautiful theory of Generalized Locally Toeplitz (GLT) sequences, and let me diverge from my planned tasks, which has been truly rewarding.

I am especially grateful to my secondary advisor, Stefano Serra-Capizzano, who is the most important mentor and reason for this thesis. Thank you for providing me with the right questions, and giving me enough foundation and encouragement for asking my own. Finding my "own truths" has been a pure joy under your guidance. Also, thank you for your patience with my engineering mathematics. For tirelessly coaching me in the art of GLT sequences, Carlo Garoni has been a great teacher and secondary advisor. Your writing is a true inspiration! Thank you both for discussions, hospitality, and friendship.

Also, I extend a big thank you to all my other secondary advisors during my work on DG, Jan Nordström, Axel Målqvist, and Per Lötstedt.

Thank you to Isabella Furci for our educating, heated, and joyful discussions, collaboration, and for making me feel at home in Como. Also, thank you to Hendrik Speelers, whose codes facilitated the experiments of this thesis in many ways. I also acknowledge my other co-authors, Fayyaz Ahmad, Eman Salem Al-Aidarous, and Dina Abdullah Alrehaili. I gratefully acknowledge all the anonymous reviewers of the papers included in this thesis.

Olivier Amoignon, Shervin Bagheri, Andreas Hellander, and Eddie Wadbro, who I met in the beginning of my PhD studies, thank you all for the fun times and inspiring me to continue. It is great to see how far you all have come! Ali Dorostkar who I met at the end of my PhD studies, thank you for interesting discussions on GLT and programming, and the extracurricular activities. Thank

you to Jens Berg, Andreas Eklind, Magnus Grandin, Stefan Hellander, Karen Ramirez, and Martin Tillenius for food, drinks, music, discussions and insights. Also, thanks to Davide Bianchi, Ken Mattsson, Murtazo Nazarov, and Saleh Rezaeiravesh for fruitful discussions.

I acknowledge all the personnel at TDB and the IT department at Uppsala University, especially Dick Elfström, Anna-Lena Forsberg, Carina Lindgren, Elisabeth Lindqvist, Marina Nordholm, Tom Smedsaas, and Lina von Sydow for making everything work.

Thank you to my colleagues and friends at Uppsala University, the Graduate School in Mathematics and Computing (FMB), the ADIGMA project, the Swedish Defence Research Agency (FOI), Umeå University, and Università degli Studi dell'Insubria.

I am also grateful to everybody on IRC, and especially the Julia community and Andreas Noack, for discussions and support. A sincere thank you to all the proofreaders of this thesis.

A special thank you goes to Bengt Smedmyr and Göran Laurell, and the rest of the staff at Uppsala University Hospital, for saving my life the first and second time, and thus, in a literal sense made this thesis possible.

Lastly, I want to thank my family. My parents, Birgitta and Curt Ekström, thank you for your encouragement, inspiration, and help both practically and scientifically. Our discussions has led to several experimental discoveries. My brother, Per-Anders Ekström, for stem cells, feedback, and algorithmic advice, and sister-in-law Theresa von Hofsten and nieces Rita and Simone von Hofsten, who all bring such happiness and joy into my life.

5. Svensk sammanfattning

Modellering av fysikaliska fenomen ger ofta i ett lineärt ekvationssystem som kan representeras av en eller flera matriser. Matrisernas egenvärden kan vara av intresse rent fysikaliskt, men även som hjälpmedel för att analysera och lösa ekvationssystemet. Att numeriskt räkna fram egenvärden för en matris är oftast tidsödande och kostsamt, speciellt om matrisen är stor.

Teorin om Generalized Locally Toeplitz-sekvenser (GLT-sekvenser) beskriver egenvärdenas beteende för sekvenser av så kallade Toeplitz och Toeplitz-liknande matriser. I denna avhandling använder vi oss av denna teori för att utveckla nya noggranna och effektiva metoder för att beräkna stora matrisers egenvärden. En Toeplitz-matris är av formen

$$T_n(f) = [\hat{f}_{i-j}]_{i,j=1}^{n} = \begin{bmatrix} \hat{f}_0 & \hat{f}_{-1} & \hat{f}_{-2} & \cdots & \cdots & \hat{f}_{1-n} \\ \hat{f}_1 & \hat{f}_0 & \hat{f}_{-1} & \ddots & & \vdots \\ \hat{f}_2 & \hat{f}_1 & \hat{f}_0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \hat{f}_{-2} \\ \vdots & & \ddots & \ddots & \ddots & \hat{f}_{-1} \\ \hat{f}_{n-1} & \cdots & \cdots & \hat{f}_2 & \hat{f}_1 & \hat{f}_0 \end{bmatrix}, \tag{5.1}$$

det vill säga en kvadratisk matris (av storleken $n \times n$) med konstanta diagonaler. Toeplitz-matrisen $T_n(f)$ genereras av en funktion f, kallad symbolen, när matrisens element, enligt (5.1), är Fourier-koefficienterna för f

$$\hat{f}_\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-\mathbf{i}\omega\theta} \mathrm{d}\theta, \quad \mathbf{i}^2 = -1, \quad \omega \in \mathbb{Z}.$$

Egenvärdena för den genererade Toeplitz-matrisen $T_n(f)$ betecknas $\lambda_j(T_n(f))$. Toeplitz-liknande matriser har samma grundstruktur som (5.1), men kan t.ex. ha ändringar av låg rank, vara av block-typ, ha variabla koefficienter eller vara av andra typer av subklasser. Ett antal varianter av Toeplitz-liknande matristyper behandlas i denna avhandling.

För en Toeplitz eller Toeplitz-liknande matris kan man använda symbolen $f(\theta)$ för att effektivt uppskatta egenvärdena (eller singulärvärdena). För monotona symboler som är reella trigonometriska polynom, bestående av cosinustermer, görs detta genom att man beräknar symbolens funktionsvärden på ett likformigt nät $\theta_{j,n} \in [0, \pi]$, där $j = 1, \ldots, n$. För dessa approximationer är felen $E_{j,n} = \lambda_j(T_n(f)) - f(\theta_{j,n})$ av ordning $\mathcal{O}(h)$, där $h = 1/(n+1)$. Felen

$E_{j,n}$ kan beskrivas av en asymptotisk expansion

$$\begin{aligned}\lambda_j(T_n(f)) &= f(\theta_{j,n}) + E_{j,n} \\ &= f(\theta_{j,n}) + \sum_{k=0}^{\alpha} c_k(\theta_{j,n})h^k + E_{j,n,\alpha},\end{aligned} \tag{5.2}$$

där $f(\theta) \in L^1(-\pi, \pi)$. Felen $E_{j,n,\alpha}$ är av ordning $\mathcal{O}(h^{\alpha+1})$. Genomgående i denna avhandling används varianter av (5.2) för att beräkna egenvärdena för olika typer av strukturerade matriser. De metoder vi utvecklar approximerar funktionerna $c_k(\theta)$, som sedan används för att beräkna egenvärdena för stora matriser, med stor noggrannhet. Vi kallar dessa nya metoder för ”matris-lösa”; då man varken behöver skapa matrisen vars egenvärden man söker, eller utföra några matris-vektor-multiplikationer vilket görs i så kallade ”matris-fria” metoder. Det enda som behövs för att beräkna dessa egenvärden är funktionsevalueringar av symbolen, och de approximerade funktionerna $c_k(\theta)$.

I den första artikeln (Paper I) beskrivs en algoritm som, genom valet av parametrarna α och n_1, kan uppskatta funktionsvärdena $c_k(\theta_{j_1,n_1})$ i (5.2), där $k = 1, \ldots, \alpha$ ($c_0(\theta) = 0$) och $j_1 = \{1, \ldots, n_1\}$. Dessa beräkningar kräver endast lösningen av ett fåtal små egenvärdesproblem. För en stor matris av storleken $n_m \times n_m$ där $n_m = 2^{m-1}(n_1 + 1) - 1$ (antag $m \in \mathbb{Z}_+$ och $m > \alpha$), kan de beräknade $\tilde{c}_k(\theta_{j_1,n_1})$ användas för en bättre approximation av egenvärdena med index $j_m = 2^{m-1}j_1$. Felen $\tilde{E}_{j_1,n_1,\alpha}$ är i storleksordningen $\mathcal{O}(h^{\alpha+1})$. I den andra artikeln (Paper II) beskrivs hur man kan använda expansionen (5.2) när matrisen är förkonditionerad. Förkonditionerad betyder här att matrisen har formen $T_n^{-1}(g)T_n(f)$. Symbolen är $r = g^{-1}f$ och det är egenvärdena för matrisen $T_n^{-1}(g)T_n(f)$ som vi approximerar, och inte för den genererade matrisen $T_n(g^{-1}f)$. I den tredje artikeln (Paper III) introduceras en algoritm, som bygger på interpolation och extrapolation, för att approximera hela spektrumet, för förkonditionerade matriser av godtycklig ordning n. I den fjärde artikeln (Paper IV) används resultaten från de tre första artiklarna (Paper I–III) för att beräkna egenvärdena för olika matriser som fås genom diskretiseringar med isogeometrisk analys (IgA). Även exakta uttryck för egenvärdena ges för vissa av matriserna. De beaktade matriserna är Toeplitz-liknande, då de är Toeplitz-matriser plus lågranks-korrektioner. I den femte och sista artikeln i avhandlingen (Paper V) presenteras analytiska uttryck för både egenvärden och egenvektorer för en speciell klass av matriser – ”symmetriskt glesa tridiagonala” Toeplitz-matriser. Vi diskuterar även expansionen (5.2), då matriserna genereras av icke-monotona symboler.

I avhandlingen presenteras olika aspekter, både teoretiska och numeriska, av expansionen (5.2) för Toeplitz och Toeplitz-liknande matriser. Effektiva matrislösa metoder utvecklas för att beräkna egenvärdena för stora matriser. Det finns stor potential för att kunna nyttja dessa tekniker för att förbättra och analysera existerande metoder, samt för att utveckla nya.

List of Examples

References

[1] A. S. Al-Fhaid, S. Serra-Capizzano, D. Sesana, and M. Zaka Ullah, Singular-value (and eigenvalue) distribution and Krylov preconditioning of sequences of sampling matrices approximating integral operators, Numerical Linear Algebra with Applications, 21 (2014), pp. 722–743. (cited on pp. 17 and 32)

[2] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, LAPACK Users' Guide, Society for Industrial and Applied Mathematics, Philadelphia, PA, third ed., 1999. (cited on p. 39)

[3] F. Avram, On bilinear forms in Gaussian random variables and Toeplitz matrices, Probability Theory and Related Fields, 79 (1988), pp. 37–45. (cited on pp. 12 and 15)

[4] G. Barbarino, Equivalence between GLT sequences and measurable functions, Linear Algebra and its Applications, 529 (2017), pp. 397–412. (cited on pp. 17 and 18)

[5] G. Barbarino and C. Garoni, From convergence in measure to convergence of matrix-sequences through concave functions and singular values, Electronic Journal of Linear Algebra, 32 (2017), pp. 500–513. (cited on pp. 17 and 18)

[6] M. Barrera, A. Böttcher, S. M. Grudsky, and E. A. Maximenko, Eigenvalues of even very nice Toeplitz matrices can be unexpectedly erratic, arXiv preprint arXiv:1710.05243, (2017). (cited on pp. 42 and 43)

[7] B. Beckermann and S. Serra-Capizzano, On the Asymptotic Spectrum of Finite Element Matrix Sequences, SIAM Journal on Numerical Analysis, 45 (2007), pp. 746–769. (cited on p. 32)

[8] P. Benedusi, C. Garoni, R. Krause, X. Li, and S. Serra-Capizzano, Discontinuous Galerkin Discretization of the Heat Equation in Any Dimension: the Spectral Symbol, Tech. Rep. 2018-002, Department of Information Technology, Uppsala University, Jan. 2018. SIAM Journal on Matrix Analysis and Applications (to appear). (cited on pp. 30 and 32)

[9] D. Bertaccini, M. Donatelli, F. Durastante, and S. Serra-Capizzano, Optimizing a multigrid Runge–Kutta smoother for variable-coefficient convection–diffusion equations, Linear Algebra and its Applications, 533 (2017), pp. 507–535. (cited on p. 32)

[10] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, Julia: A Fresh Approach to Numerical Computing, SIAM Review, 59 (2017), pp. 65–98. (cited on pp. 48 and 211)

[11] R. Bhatia, Matrix Analysis, Springer New York, 1997. (cited on p. 17)

[12] D. Bini and M. Capovani, Spectral and computational properties of band symmetric Toeplitz matrices, Linear Algebra and its Applications, 52-53 (1983), pp. 99–126. (cited on pp. 21 and 38)

[13] J. M. Bogoya, A. Böttcher, S. M. Grudsky, and E. A. Maximenko, Eigenvalues of Hermitian Toeplitz matrices with smooth simple-loop symbols, Journal of Mathematical Analysis and Applications, 422 (2015), pp. 1308–1334. (cited on p. 42)

[14] J. M. Bogoya, S. M. Grudsky, and E. A. Maximenko, Eigenvalues of Hermitian Toeplitz Matrices Generated by Simple-loop Symbols with Relaxed Smoothness, in Large Truncated Toeplitz Matrices, Toeplitz Operators, and Related Topics, Springer International Publishing, 2017, pp. 179–212. (cited on p. 42)

[15] E. Bozzo and C. Di Fiore, On the Use of Certain Matrix Algebras Associated with Discrete Trigonometric Transforms in Matrix Displacement Decomposition, SIAM Journal on Matrix Analysis and Applications, 16 (1995), pp. 312–326. (cited on pp. 18, 19, 21, and 59)

[16] A. Böttcher, C. Garoni, and S. Serra-Capizzano, Exploration of Toeplitz-like matrices with unbounded symbols: not a purely academic journey (Исследование обобщенно-тёплицевых матриц с неограниченными символами — это не только академическое занятие), Matematicheskii Sbornik (Математический сборник), 208:11 (2017), pp. 29–55. English translation in Sbornik: Mathematics, 208:11 (2017), pp. 1602–1627. (cited on pp. 15, 17, 18, and 32)

[17] A. Böttcher and S. M. Grudsky, Toeplitz Matrices, Asymptotic Linear Algebra, and Functional Analysis, Birkhäuser Basel, 2000. (cited on p. 13)

[18] A. Böttcher and S. M. Grudsky, Spectral Properties of Banded Toeplitz Matrices, Society for Industrial and Applied Mathematics, jan 2005. (cited on p. 20)

[19] A. Böttcher and B. Silbermann, Introduction to Large Truncated Toeplitz Matrices, Springer New York, 1999. (cited on pp. 12, 13, and 16)

[20] A. Böttcher and B. Silbermann, Analysis of Toeplitz Operators, Springer Berlin Heidelberg, 2006. (cited on p. 13)

[21] C. R. Crawford, Reduction of a band-symmetric generalized eigenvalue problem, Communications of the ACM, 16 (1973), pp. 41–44. (cited on p. 52)

[22] F. Di Benedetto, G. Fiorentino, and S. Serra-Capizzano, C. G. preconditioning for Toeplitz matrices, Computers & Mathematics with Applications, 25 (1993), pp. 35–45. (cited on p. 15)

[23] M. Donatelli, A. Dorostkar, M. Mazza, M. Neytcheva, and S. Serra-Capizzano, Function-based block multigrid strategy for a two-dimensional linear elasticity-type problem, Computers & Mathematics with Applications, 74 (2017), pp. 1015–1028. (cited on pp. 32 and 49)

[24] M. Donatelli, C. Garoni, C. Manni, S. Serra-Capizzano, and H. Speleers, Robust and optimal multi-iterative techniques for IgA Galerkin linear systems, Computer Methods in Applied Mechanics and Engineering, 284 (2015), pp. 230–264. (cited on pp. 32 and 49)

[25] M. Donatelli, C. Garoni, C. Manni, S. Serra-Capizzano, and H. Speleers, Robust and optimal multi-iterative techniques for IgA collocation linear systems, Computer Methods in Applied Mechanics and Engineering, 284 (2015), pp. 1120–1146. (cited on pp. 32 and 49)

[26] M. Donatelli, C. Garoni, C. Manni, S. Serra-Capizzano, and H. Speleers, Spectral analysis and spectral symbol of matrices in isogeometric collocation methods, Mathematics of Computation, 85 (2016), pp. 1639–1680. (cited on p. 32)

[27] M. Donatelli, M. Mazza, and S. Serra-Capizzano, Spectral analysis and structure preserving preconditioners for fractional diffusion equations, Journal of Computational Physics, 307 (2016), pp. 262–279. (cited on p. 32)

[28] A. Dorostkar, Analysis and Implementation of Preconditioners for Prestressed Elasticity Problems: Advances and Enhancements, PhD thesis, Uppsala University, 12 2017. (cited on pp. 32 and 49)

[29] A. Dorostkar, M. Neytcheva, and S. Serra-Capizzano, Spectral analysis of coupled PDEs and of their Schur complements via Generalized Locally Toeplitz sequences in 2D, Computer Methods in Applied Mechanics and Engineering, 309 (2016), pp. 74–105. (cited on p. 32)

[30] M. Dumbser, F. Fambri, I. Furci, M. Mazza, S. Serra-Capizzano, and M. Tavelli, Staggered discontinuous Galerkin methods for the incompressible Navier–Stokes equations: Spectral analysis and computational results, Numerical Linear Algebra with Applications, (2018). (in press). (cited on pp. 29, 30, and 32)

[31] S.-E. Ekström, I. Furci, and S. Serra-Capizzano, Exact Formulae and Matrix-Less Eigensolvers for Block Banded Symmetric Toeplitz Matrices, Tech. Rep. 2018-005, Department of Information Technology, Uppsala University, Mar. 2018. (cited on pp. ix, 28, 29, 63, and 71)

[32] S.-E. Ekström, C. Garoni, T. J. Hughes, A. Reali, S. Serra-Capizzano, and H. Speleers, Finite element and isogeometric B-spline discretizations of eigenvalue problems: symbol-based analysis, (in preparation), (2018). (cited on pp. ix, 26, and 32)

[33] L. Euler, Institutiones calculi differentialis, Teubner, 1755. (cited on p. 61)

[34] D. Fasino and P. Tilli, Spectral clustering properties of block multilevel Hankel matrices, Linear Algebra and its Applications, 306 (2000), pp. 155–163. (cited on p. 17)

[35] F. R. Gantmakher and M. G. Krein, Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems (Осцилляционные матрицы и ядра и малые колебания механических систем), Gostekhizdat, 1950. (cited on p. 20)

[36] C. Garoni, C. Manni, F. Pelosi, S. Serra-Capizzano, and H. Speleers, On the spectrum of stiffness matrices arising from isogeometric analysis, Numerische Mathematik, 127 (2014), pp. 751–799. (cited on pp. 58 and 61)

[37] C. Garoni, C. Manni, S. Serra-Capizzano, D. Sesana, and H. Speleers, Spectral analysis and spectral symbol of matrices in isogeometric Galerkin methods, Mathematics of Computation, 86 (2016), pp. 1343–1373. (cited on p. 32)

[38] C. Garoni, C. Manni, S. Serra-Capizzano, D. Sesana, and H. Speleers, Lusin theorem, GLT sequences and matrix computations: An application to the spectral analysis of PDE discretization matrices, Journal of Mathematical Analysis and Applications, 446 (2017), pp. 365–382. (cited on p. 32)

[39] C. Garoni and S. Serra-Capizzano, The Theory of Generalized Locally Toeplitz Sequences: a Review, an Extension, and a Few Representative Applications, in Large Truncated Toeplitz Matrices, Toeplitz Operators, and Related Topics, Springer International Publishing, 2017, pp. 353–394. (cited on pp. 15 and 17)

[40] C. Garoni and S. Serra-Capizzano, Generalized Locally Toeplitz Sequences: Theory and Applications. Vol. 1, Springer International Publishing, 2017. (cited on pp. 11, 14, 15, 17, 18, 26, 32, 49, 51, and 58)

[41] C. Garoni and S. Serra-Capizzano, Generalized Locally Toeplitz Sequences: Theory and Applications. Vol. 2, Tech. Rep. 2017-002, Department of Information Technology, Uppsala University, Feb. 2017. (cited on pp. 11, 15, 26, 30, and 32)

[42] C. Garoni, S. Serra-Capizzano, and D. Sesana, Spectral Analysis and Spectral Symbol of d-variate $\mathbb{Q}_{\boldsymbol{p}}$ Lagrangian FEM Stiffness Matrices, SIAM Journal on Matrix Analysis and Applications, 36 (2015), pp. 1100–1128. (cited on pp. 28, 30, and 32)
[43] C. Garoni, S. Serra-Capizzano, and D. Sesana, The Theory of Block Generalized Locally Toeplitz Sequences, Tech. Rep. 2018-001, Department of Information Technology, Uppsala University, Jan. 2018. (cited on pp. 15, 26, and 29)
[44] L. Golinskii and S. Serra-Capizzano, The asymptotic properties of the spectrum of nonsymmetrically perturbed Jacobi matrix sequences, Journal of Approximation Theory, 144 (2007), pp. 84–102. (cited on p. 17)
[45] G. H. Golub and C. F. Van Loan, Matrix Computations, Johns Hopkins University Press, 4 ed., 2012. (cited on p. 17)
[46] T.-X. He, Eulerian polynomials and B-splines, Journal of Computational and Applied Mathematics, 236 (2012), pp. 3763–3773. (cited on p. 61)
[47] C. Jordan, Cours d'Analyse de l'École Polytechnique, vol. 1, Gauthier-Villars, Paris, 1909. (cited on p. 15)
[48] L. Kaufman, Banded Eigenvalue Solvers on Vector Machines, ACM Transactions on Mathematical Software, 10 (1984), pp. 73–85. (cited on p. 52)
[49] A. Mantzaflaris and B. Jüttler, Exploring Matrix Generation Strategies in Isogeometric Analysis, in Mathematical Methods for Curves and Surfaces, Springer Berlin Heidelberg, 2014, pp. 364–382. (cited on p. 61)
[50] MATLAB, version 9.3.0.713579 64-bit (maci64) (R2017b), The MathWorks Inc., Natick, Massachusetts, 2017. (cited on p. 211)
[51] M. Mazza, A. Ratnani, and S. Serra-Capizzano, Spectral Analysis and Spectral Symbol for the 2D curl-curl (Stabilized) Operator with Applications to the Related Iterative Solutions, Tech. Rep. 2017-009, Department of Information Technology, Uppsala University, Apr. 2017. Mathematics of Computation (in press). (cited on p. 30)
[52] H. Moghaderi, M. Dehghan, M. Donatelli, and M. Mazza, Spectral analysis and multigrid preconditioners for two-dimensional space-fractional diffusion equations, Journal of Computational Physics, 350 (2017), pp. 992–1011. (cited on p. 32)
[53] A. Noack, Elemental.jl. `https://github.com/JuliaParallel/Elemental.jl`, 2018. (cited on p. 211)
[54] A. Noack, LinearAlgebra.jl. `https://github.com/andreasnoack/LinearAlgebra.jl`, 2018. (cited on p. 211)
[55] S. Noschese, L. Pasquini, and L. Reichel, Tridiagonal Toeplitz matrices: properties and novel applications, Numerical Linear Algebra with Applications, 20 (2012), pp. 302–326. (cited on p. 20)
[56] S. Olver, BandedMatrices.jl. `https://github.com/JuliaMatrices/BandedMatrices.jl`, 2018. (cited on pp. 52 and 211)
[57] S. V. Parter, On the distribution of the singular values of Toeplitz matrices, Linear Algebra and its Applications, 80 (1986), pp. 115–130. (cited on pp. 12 and 15)
[58] S. V. Parter and J. W. T. Youngs, The symmetrization of matrices by diagonal matrices, Journal of Mathematical Analysis and Applications, 4 (1962), pp. 102–110. (cited on p. 22)

[59] J. Poulson, B. Marker, R. A. van de Geijn, J. R. Hammond, and N. A. Romero, Elemental, ACM Transactions on Mathematical Software, 39 (2013), pp. 1–24. (cited on p. 211)
[60] F. Roman, C. Manni, and H. Speleers, Spectral analysis of matrices in Galerkin methods based on generalized B-splines with high smoothness, Numerische Mathematik, 135 (2016), pp. 169–216. (cited on p. 32)
[61] E. Salinelli, S. Serra-Capizzano, and D. Sesana, Eigenvalue-eigenvector structure of Schoenmakers–Coffey matrices via Toeplitz technology and applications, Linear Algebra and its Applications, 491 (2016), pp. 138–160. (cited on p. 32)
[62] S. Serra-Capizzano, Multi-iterative methods, Computers & Mathematics with Applications, 26 (1993), pp. 65–87. (cited on pp. 32 and 49)
[63] S. Serra-Capizzano, An ergodic theorem for classes of preconditioned matrices, Linear Algebra and its Applications, 282 (1998), pp. 161–183. (cited on p. 15)
[64] S. Serra-Capizzano, Asymptotic Results on the Spectra of Block Toeplitz Preconditioned Matrices, SIAM Journal on Matrix Analysis and Applications, 20 (1998), pp. 31–44. (cited on p. 15)
[65] S. Serra-Capizzano, Spectral and Computational Analysis of Block Toeplitz Matrices Having Nonnegative Definite Matrix-Valued Generating Functions, Bit Numerical Mathematics, 39 (1999), pp. 152–175. (cited on p. 15)
[66] S. Serra-Capizzano, Distribution results on the algebra generated by Toeplitz sequences: a finite-dimensional approach, Linear Algebra and its Applications, 328 (2001), pp. 121–130. (cited on p. 16)
[67] S. Serra-Capizzano, Generalized locally Toeplitz sequences: spectral analysis and applications to discretized partial differential equations, Linear Algebra and its Applications, 366 (2003), pp. 371–402. (cited on pp. 11, 15, 16, and 32)
[68] S. Serra-Capizzano, The GLT class as a generalized Fourier analysis and applications, Linear Algebra and its Applications, 419 (2006), pp. 180–233. (cited on pp. 11, 15, 16, and 32)
[69] N. J. A. Sloane, The Encyclopedia of Integer Sequences, 2018. (cited on pp. 61 and 211)
[70] G. Szegő, Beiträge zur Theorie der Toeplitzschen Formen, Mathematische Zeitschrift, 6 (1920), pp. 167–202. (cited on pp. 12 and 15)
[71] T. Tantau, The TikZ and PGF Packages. (cited on p. 211)
[72] The Inkscape Project, Inkscape, Version 0.92. (cited on p. 211)
[73] P. Tilli, Locally Toeplitz sequences: spectral properties and applications, Linear Algebra and its Applications, 278 (1998), pp. 91–120. (cited on pp. 11, 15, and 16)
[74] P. Tilli, A note on the spectral distribution of Toeplitz matrices, Linear and Multilinear Algebra, 45 (1998), pp. 147–159. (cited on pp. 12 and 15)
[75] P. Tilli, Some Results on Complex Toeplitz Eigenvalues, Journal of Mathematical Analysis and Applications, 239 (1999), pp. 390–401. (cited on pp. 12, 15, and 16)
[76] O. Toeplitz, Zur Theorie der quadratischen und bilinearen Formen von unendlichvielen Veränderlichen, Mathematische Annalen, 70 (1911), pp. 351–376. (cited on p. 12)
[77] E. E. Tyrtyshnikov and N. L. Zamarashkin, Spectra of multilevel Toeplitz matrices: Advanced theory via simple matrix relationships, Linear Algebra and its Applications, 270 (1998), pp. 15–27. (cited on pp. 12, 15, and 29)

[78] R.-H. Wang, Y. Xu, and Z.-Q. Xu, Eulerian numbers: A spline perspective, Journal of Mathematical Analysis and Applications, 370 (2010), pp. 486–490. (cited on p. 61)

[79] J. H. Wilkinson, Some recent advances in numerical linear algebra, in The state of the art in numerical analysis: proceedings of the Conference on the State of the Art in Numerical Analysis held at the University of York, April 12th-15th, 1976, D. A. H. Jacobs, ed., Mathematics and Its Application Series, Academic Press, 1977, pp. 3–53. (cited on p. 52)

[80] Wolfram Research, Inc., Mathematica, Version 11.2. Champaign, IL, 2018. (cited on p. 211)

Papers

Contributions by the Author

Here is presented a summary of the contributions by the author of this thesis in each of Papers I–V.

Paper I

The author of this thesis discovered the basic structure of the asymptotic expansion of the eigenvalues numerically. The co-authors refined the idea mathematically, algorithmically, and put it in a relevant scientific context.

Paper II

The author of this thesis cooperated with the other authors in various ways: mainly with algorithm design, code development, and numerical experiments. The co-authors developed the mathematical proofs.

Paper III

The author of this thesis had the general idea of the algorithm and supported it through numerical experiments. The co-author refined the idea mathematically and algorithmically.

Paper IV

The author of this thesis discovered the exact solution for $p = 2$, the "optimal" use of the uniform τ_{n-1} and τ_{n-1}^{π}-grids for $p > 2$, and the multi-dimensional extension of the one-dimensional setting. The co-authors refined the ideas mathematically, algorithmically, and put them in a relevant scientific context.

Paper V

The author of this thesis discovered the grid to attain the exact eigenvalues when sampling the symbol. Also the corresponding eigenvectors, the proof, and numerical experiments have been worked out by the author. The co-author refined the article and put it in a mathematically and scientifically relevant context.

Paper I [redacted]

EXPERIMENTAL MATHEMATICS
https://doi.org/10.1080/10586458.2017.1320241

Taylor & Francis
Taylor & Francis Group

Are the Eigenvalues of Banded Symmetric Toeplitz Matrices Known in Almost Closed Form?

Sven-Erik Ekström [a], Carlo Garoni [b,c], and Stefano Serra-Capizzano [a,c]

[a]Department of Information Technology, Division of Scientific Computing, Uppsala University, ITC, Uppsala, Sweden; [b]University of Italian Switzerland (USI), Institute of Computational Science, Lugano, Switzerland; [c]Department of Science and High Technology, University of Insubria, Como, Italy

ABSTRACT
Bogoya, Böttcher, Grudsky, and Maximenko have recently obtained for the eigenvalues of a Toeplitz matrix, under suitable assumptions on the generating function, the precise asymptotic expansion as the matrix size goes to infinity. In this article we provide numerical evidence that some of these assumptions can be relaxed. Moreover, based on the eigenvalue asymptotics, we devise an extrapolation algorithm for computing the eigenvalues of banded symmetric Toeplitz matrices with a high level of accuracy and a relatively low computational cost.

KEYWORDS
eigenvalue asymptotics; eigenvalues; extrapolation; polynomial interpolation; Toeplitz matrix

2010 AMS SUBJECT CLASSIFICATION
15B05; 65F15; 65D05; 65B05

1. Introduction

A matrix of the form

$$[a_{i-j}]_{i,j=1}^{n} = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & \cdots & a_{-(n-1)} \\ a_1 & \ddots & \ddots & \ddots & & \vdots \\ a_2 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & a_{-2} \\ \vdots & & \ddots & \ddots & \ddots & a_{-1} \\ a_{n-1} & \cdots & \cdots & a_2 & a_1 & a_0 \end{bmatrix},$$

whose entries are constant along each diagonal, is called a Toeplitz matrix. Given a function $f : [-\pi, \pi] \to \mathbb{C}$ belonging to $L^1([-\pi, \pi])$, the nth Toeplitz matrix associated with f is defined as

$$T_n(f) = [\hat{f}_{i-j}]_{i,j=1}^{n},$$

where the numbers $\hat{f}_k$ are the Fourier coefficients of f,

$$\hat{f}_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta)\, \mathrm{e}^{-\mathrm{i}k\theta} \mathrm{d}\theta, \qquad k \in \mathbb{Z}.$$

We refer to $\{T_n(f)\}_n$ as the Toeplitz sequence generated by f, which in turn is called the generating function or the symbol of $\{T_n(f)\}_n$. In the case where f is real, all the matrices $T_n(f)$ are Hermitian and much is known about their spectral properties, from the localization of the eigenvalues to the asymptotic spectral distribution in the Weyl sense; see [Böttcher and Silbermann 99, Garoni and Serra-Capizzano 17] and the references therein.

The present article focuses on the case where f is a real cosine trigonometric polynomial (RCTP), that is, a function of the form

$$f(\theta) = \hat{f}_0 + 2\sum_{k=1}^{m} \hat{f}_k \cos(k\theta), \qquad \hat{f}_0, \hat{f}_1, \ldots, \hat{f}_m \in \mathbb{R}, \quad m \in \mathbb{N}.$$

We say that the RCTP f is monotone if it is either increasing or decreasing over the interval $[0, \pi]$. The nth Toeplitz matrix generated by f is the real symmetric banded matrix given by

$$T_n(f) = \begin{bmatrix} \hat{f}_0 & \hat{f}_1 & \cdots & \hat{f}_m & & & & & & & \\ \hat{f}_1 & \ddots & \ddots & & \ddots & & & & & & \\ \vdots & \ddots & \ddots & \ddots & & \ddots & & & & & \\ \hat{f}_m & & \ddots & \ddots & \ddots & & \ddots & & & & \\ & \ddots & & \ddots & \ddots & \ddots & & \ddots & & & \\ & & \boxed{\hat{f}_m} & \cdots & \hat{f}_1 & \hat{f}_0 & \hat{f}_1 & \cdots & \hat{f}_m & & \\ & & & \ddots & & \ddots & \ddots & \ddots & & \ddots & \\ & & & & \ddots & & \ddots & \ddots & \ddots & & \hat{f}_m \\ & & & & & \ddots & & \ddots & \ddots & \ddots & \vdots \\ & & & & & & \ddots & & \ddots & \ddots & \hat{f}_1 \\ & & & & & & & \hat{f}_m & \cdots & \hat{f}_1 & \hat{f}_0 \end{bmatrix}.$$

In [Bogoya et al. 15a, Bogoya et al. 17, Böttcher et al. 10] it was proved that if the RCTP f is monotone and

CONTACT Sven-Erik Ekström sven-erik.ekstrom@it.uu.se Department of Information Technology, Division of Scientific Computing, Uppsala University, ITC, Lägerhyddsv. 2, Hus 2, P.O. Box 337, SE-751 05 Uppsala, Sweden.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uexm.

satisfies certain additional assumptions, which include the requirements that $f'(\theta) \neq 0$ for $\theta \in (0, \pi)$ and $f''(\theta) \neq 0$ for $\theta \in \{0, \pi\}$, then, for every integer $\alpha \geq 0$, every n and every $j = 1, \ldots, n$, the following asymptotic expansion holds:

$$\lambda_j(T_n(f)) = f(\theta_{j,n}) + \sum_{k=1}^{\alpha} c_k(\theta_{j,n}) h^k + E_{j,n,\alpha}, \quad (1\text{–}1)$$

where:

- The eigenvalues of $T_n(f)$ are arranged in non-decreasing or non-increasing order, depending on whether f is increasing or decreasing.
- $\{c_k\}_{k=1,2,\ldots}$ is a sequence of functions from $[0, \pi]$ to $\mathbb{R}$ which depends only on f.
- $h = \frac{1}{n+1}$ and $\theta_{j,n} = \frac{j\pi}{n+1} = j\pi h$.
- $E_{j,n,\alpha} = O(h^{\alpha+1})$ is the remainder (the error), which satisfies the inequality $|E_{j,n,\alpha}| \leq C_\alpha h^{\alpha+1}$ for some constant C_α depending only on α and f.

The symbols

$$f_q(\theta) = (2 - 2\cos\theta)^q, \qquad q = 1, 2, \ldots \quad (1\text{–}2)$$

arise in the discretization of differential equations and are therefore of particular interest. Unfortunately, for these symbols the requirement that $f''(0) \neq 0$ is not satisfied if $q \geq 2$. The first purpose of this article is to provide numerical evidence that the higher-order approximation (1–1) holds even in this "degenerate case." Actually, based on our numerical experiments, we conjecture that (1–1) holds for all monotone RCTPs f.

In [Bogoya et al. 15a], the authors also briefly mentioned that the asymptotic expansion (1–1) can be used to compute an accurate approximation of $\lambda_j(T_n(f))$ for very large n, provided the values $\lambda_{j_1}(T_{n_1}(f)), \lambda_{j_2}(T_{n_2}(f)), \lambda_{j_3}(T_{n_3}(f))$ are available for moderately sized n_1, n_2, n_3 with $\theta_{j_1,n_1} = \theta_{j_2,n_2} = \theta_{j_3,n_3} = \theta_{j,n}$. The second and main purpose of this article is to carry out this idea and to support it by numerical experiments accompanied by an appropriate error analysis. In particular, we devise an algorithm to compute $\lambda_j(T_n(f))$ with a high level of accuracy and a relatively low computational cost. The algorithm is completely analogous to the extrapolation procedure which is employed in the context of Romberg integration to obtain high precision approximations of an integral from a few coarse trapezoidal approximations [Stoer and Bulirsch 02, Section 3.4]. In this regard, the asymptotic expansion (1–1) plays here the same role as the Euler–Maclaurin summation formula [Stoer and Bulirsch 02, Section 3.3].

In the case where the monotonicity assumption on f is violated, a first-order asymptotic formula for the eigenvalues was established by Bogoya, Böttcher, Grudsky, and Maximenko in [Bogoya et al. 15b]. In particular, following the argument used for the proof of [Bogoya et al. 15b, Theorem 1.6], one can show that for every RCTP f, every n and every $j = 1, \ldots, n$, we have

$$\lambda_{\rho_n(j)}(T_n(f)) = f(\theta_{j,n}) + E_{j,n,0}, \quad (1\text{–}3)$$

where:

- The eigenvalues of $T_n(f)$ are arranged in non-decreasing order, $\lambda_1(T_n(f)) \leq \cdots \leq \lambda_n(T_n(f))$.
- $\rho_n = \sigma_n^{-1}$, where σ_n is a permutation of $\{1, \ldots, n\}$ such that $f(\theta_{\sigma_n(1),n}) \leq \cdots \leq f(\theta_{\sigma_n(n),n})$.
- $h = \frac{1}{n+1}$ and $\theta_{j,n} = \frac{j\pi}{n+1} = j\pi h$.
- $E_{j,n,0} = O(h)$ is the error, which satisfies the inequality $|E_{j,n,0}| \leq C_0 h$ for some constant C_0 depending only on f.

The third and last purpose of this article is to formulate, on the basis of numerical experiments, a conjecture on the higher-order asymptotics of the eigenvalues if the monotonicity assumption on f is not in force. We also illustrate how this conjecture can be used along with our extrapolation algorithm in order to compute some of the eigenvalues of $T_n(f)$ in the case where f is non-monotone.

1.1. Ideas from numerical linear algebra

Before entering into the details of the article, we allow us a digression. Our aim is to highlight that the first-order expansion (1–3) may be proved by purely linear algebra arguments in combination with the results about the so-called quantile function obtained in [Bogoya et al. 15b, Bogoya et al. 16]. Let us outline the scheme of a linear algebra proof of this kind. We will make use of the so-called τ matrices and the related properties [Bini and Capovani 83, Serra-Capizzano 96].

Let $\tau_n(f)$ be the τ matrix of size n generated by f. Then, $\tau_n(f)$ is a real symmetric matrix with the following properties:

- $T_n(f) = \tau_n(f) + R_n^+ + R_n^-$, where R_n^+ is a symmetric nonnegative definite matrix of rank k^+, R_n^- is a symmetric nonpositive definite matrix of rank k^-, and $k^+ + k^- \leq 2(m-1)$, with m being the degree of f.
- The eigenvalues of $\tau_n(f)$ are $f(\theta_{j,n})$, $j = 1, \ldots, n$.

Using a classical interlacing theorem for the eigenvalues (see [Bhatia 97, Exercise III.2.4] or [Garoni and Serra-Capizzano 17, Theorem 2.12]), we obtain

$$f(\theta_{\sigma_n(j-k^-),n}) \leq \lambda_j(T_n(f)) \leq f(\theta_{\sigma_n(j+k^+),n}), \quad j = k^- + 1, \ldots, n - k^+. \quad (1\text{–}4)$$

Moreover, it is known that

$$\lambda_j(T_n(f)) \in [m_f, M_f], \qquad j = 1, \ldots, n, \quad (1\text{–}5)$$

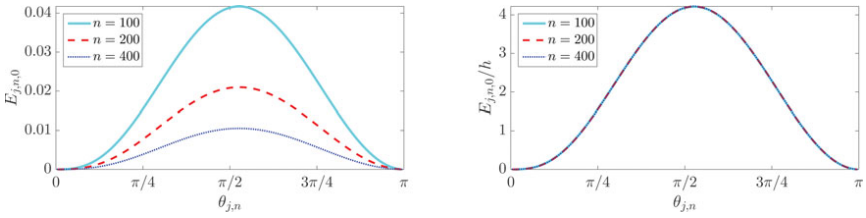


Figure 1. Example 1: Errors $E_{j,n,0}$ and scaled errors $E_{j,n,0}/h$ versus $\theta_{j,n}$ for $j = 1, \dots, n$ and $n = 100, 200, 400$ in the case of the symbol $f(\theta) = (2 - 2\cos\theta)^2$.

where $m_f = \min f$ and $M_f = \max f$; see [Böttcher and Silbermann 99, Garoni and Serra-Capizzano 17]. Considering that f is an RCTP and hence a Lipschitz continuous function, the result (1–3) intuitively follows from (1–4) and (1–5). For a formal derivation, however, it is necessary to resort to the quantile function of f, which is monotone and Lipschitz continuous whenever f is Lipschitz continuous; see [Bogoya et al. 15b, Proposition 2.7].

The relation (1–4) is known in the numerical linear algebra community since more than 30 years and was used in [Serra-Capizzano 96] to study the asymptotics of the extreme eigenvalues of Toeplitz matrices. In particular, if $\alpha \geq 2$ denotes the minimum order of the zeros of $f - \min f$, it was proved in [Serra-Capizzano 96] that the errors $E_{j,n,0}$ corresponding to the smallest eigenvalues of $T_n(f)$ are $O(h^\alpha)$ and not only $O(h)$. More precisely, whenever j is constant with respect to n, we have $|E_{j,n,0}| \leq c_j h^\alpha$ for some constant c_j depending only on f and j.

2. Numerical experiments in support of the asymptotic expansion

We present in this section a few numerical examples, with the purpose of supporting the conjecture that the asymptotic expansion (1–1) is satisfied for all monotone RCTPs f, including those which do not meet the requirements $f'(\theta) \neq 0$ for $\theta \in (0, \pi)$ and $f''(\theta) \neq 0$ for $\theta \in \{0, \pi\}$.

Example 1. Let f be the monotone RCTP defined by (1–2) for $q = 2$,

$$f(\theta) = f_2(\theta) = (2 - 2\cos\theta)^2 = 6 - 8\cos\theta + 2\cos(2\theta).$$

Note that $f''(0) = 0$. The expansion (1–1) with $\alpha = 1$ would say that, for every n and every $j = 1, \dots, n$,

$$\lambda_j(T_n(f)) - f(\theta_{j,n}) = E_{j,n,0} = c_1(\theta_{j,n})h + E_{j,n,1}, \quad (2\text{–}6)$$

where $|E_{j,n,1}| \leq C_1 h^2$ and both the function $c_1 : [0, \pi] \to \mathbb{R}$ and the constant C_1 depend only on f. In particular, the scaled errors $E_{j,n,0}/h$ should be equal to the equispaced samples $c_1(\theta_{j,n})$ (and should therefore reproduce the graph of the function c_1) in the limit where $n \to \infty$. In Figure 1 we plot the errors $E_{j,n,0}$ and the scaled errors $E_{j,n,0}/h$ versus $\theta_{j,n}$ for $j = 1, \dots, n$ and $n = 100, 200, 400$. It is clear that the scaled errors overlap perfectly, thus supporting the conjecture that the expansion (2–6) holds despite the fact that $f''(0) = 0$. In particular, the right pane of Figure 1 displays the graph of c_1 over $[0, \pi]$.

Example 2. Let

$$f(\theta) = 1 + 24\cos\theta - 12\cos(2\theta) + 8\cos(3\theta) - 3\cos(4\theta).$$

The function f is a monotone decreasing RCTP such that $f'(\pi/2) = f''(\pi/2) = f''(0) = 0$. Figure 2 is obtained in the same way as Figure 1. Again, we see that the scaled errors overlap perfectly, thus supporting the conjecture that the expansion (2–6) holds even for this function f, despite the fact that f violates both the conditions $f'(\theta) \neq 0$ for $\theta \in (0, \pi)$ and $f''(\theta) \neq 0$ for $\theta \in \{0, \pi\}$.

Example 3. Let f be the same as in Example 2. The expansion (1–1) with $\alpha = 2$ would say that, for every n and

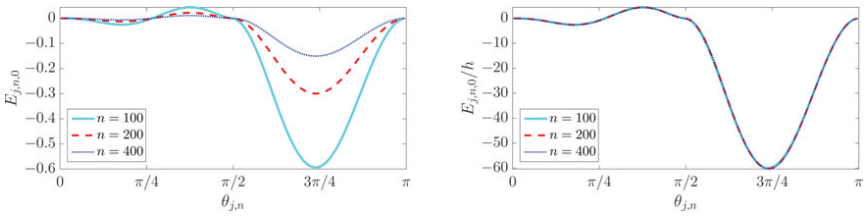


Figure 2. Example 2: Errors $E_{j,n,0}$ and scaled errors $E_{j,n,0}/h$ versus $\theta_{j,n}$ for $j = 1, \dots, n$ and $n = 100, 200, 400$ in the case of the symbol $f(\theta) = 1 + 24\cos\theta - 12\cos(2\theta) + 8\cos(3\theta) - 3\cos(4\theta)$.

every $j = 1, \dots, n$,

$$\lambda_j(T_n(f)) - f(\theta_{j,n}) - c_1(\theta_{j,n})h = E_{j,n,1} = c_2(\theta_{j,n})h^2 + E_{j,n,2}, \tag{2–7}$$

where $|E_{j,n,2}| \le C_2 h^3$ and both the function $c_2 : [0, \pi] \to \mathbb{R}$ and the constant C_2 depend only on f. In particular, the scaled errors $E_{j,n,1}/h^2$ should be equal to the equispaced samples $c_2(\theta_{j,n})$ (and should therefore reproduce the graph of the function c_2) in the limit where $n \to \infty$. Unfortunately, the values $E_{j,n,1}$ are not available, because the function c_1 is unknown. To work around this problem, we fix $n' \gg n$ such that $(n' + 1)$ is a multiple of $(n + 1)$ and we approximate $E_{j,n,1}$ by

$$\tilde{E}_{j,n,1} = \lambda_j(T_n(f)) - f(\theta_{j,n}) - \tilde{c}_1(\theta_{j,n})h,$$

where $\tilde{c}_1$ is the approximation of c_1 obtained from the scaled errors $E_{j',n',0}/h'$ corresponding to the fine parameter n'. In other words, $\tilde{c}_1$ is defined at every point $\theta_{j',n'}$ as

$$\tilde{c}_1(\theta_{j',n'}) = \frac{E_{j',n',0}}{h'} = \frac{\lambda_{j'}(T_{n'}(f)) - f(\theta_{j',n'})}{h'} = c_1(\theta_{j',n'}) + \frac{E_{j',n',1}}{h'}, \quad j' = 1, \dots, n', \quad h' = \frac{1}{n'+1}.$$

Note that $\tilde{c}_1$ is also defined at every point $\theta_{j,n}$, because $(n' + 1)$ is a multiple of $(n + 1)$ and hence every $\theta_{j,n}$ is equal to some $\theta_{j',n'}$ (indeed, $\theta_{j,n} = \theta_{j',n'}$ for $j' = j\frac{n'+1}{n+1}$). When approximating $c_2(\theta_{j,n})$ by $\tilde{E}_{j,n,1}/h^2$ instead of $E_{j,n,1}/h^2$, the error can be estimated as follows:

$$\begin{aligned}
&\left| \frac{\tilde{E}_{j,n,1}}{h^2} - c_2(\theta_{j,n}) \right| \\
&= \left| \frac{E_{j,n,1} + h\big[\tilde{c}_1(\theta_{j,n}) - c_1(\theta_{j,n})\big]}{h^2} - c_2(\theta_{j,n}) \right| \\
&\le \left| \frac{E_{j,n,1}}{h^2} - c_2(\theta_{j,n}) \right| + \frac{1}{h}\left|\tilde{c}_1(\theta_{j,n}) - c_1(\theta_{j,n})\right| \\
&= \left| \frac{E_{j,n,2}}{h^2} \right| + \frac{1}{h}\left|\tilde{c}_1(\theta_{j',n'}) - c_1(\theta_{j',n'})\right| \\
&\qquad \text{(here } j' = j\tfrac{n'+1}{n+1} \text{ so that } \theta_{j',n'} = \theta_{j,n}\text{)} \\
&= \left| \frac{E_{j,n,2}}{h^2} \right| + \frac{1}{h}\left| \frac{E_{j',n',1}}{h'} \right| \\
&\le C_2 h + C_1 \frac{h'}{h}.
\end{aligned}$$

We may then expect that the errors $|\tilde{E}_{j,n,1}/h^2 - c_2(\theta_{j,n})|$ are of the same order as the errors $|E_{j,n,1}/h^2 - c_2(\theta_{j,n})| = |E_{j,n,2}/h^2|$ provided that $h' = O(h^2)$. In Figure 3 we plot the approximated errors $\tilde{E}_{j,n,1}$ and the approximated scaled errors $\tilde{E}_{j,n,1}/h^2$ versus $\theta_{j,n}$ for $j = 1, \dots, n$ and $n = 100, 200, 400$, with $n' = \lceil \frac{n+1}{12} \rceil (n + 1) - 1$. With this choice of n', we ensure that $(n' + 1)$ is a multiple of $(n + 1)$ and $h' \approx 12h^2$ for all n. The figure reveals that the approximated scaled errors converge to a limit function c_2, thus supporting the conjecture that the expansion (2–7) holds despite the fact that f violates both the conditions $f'(\theta) \neq 0$ for $\theta \in (0, \pi)$ and $f''(\theta) \neq 0$ for $\theta \in \{0, \pi\}$.

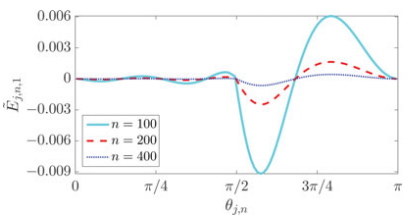
3. Algorithm for computing the eigenvalues with high accuracy

In Section 2 we showed through numerical examples that the asymptotic expansion (1–1) is likely to be satisfied for every monotone RCTP f. We now illustrate how (1–1) can be used to compute an accurate approximation of $\lambda_j(T_n(f))$ for large n.

Let f be a monotone RCTP, fix $n \in \mathbb{N}$ and $j \in \{1, \dots, n\}$. Suppose $\lambda_{j_1}(T_{n_1}(f)), \dots, \lambda_{j_m}(T_{n_m}(f))$ are available for some $(j_1, n_1), \dots, (j_m, n_m)$ such that $j_1 h_1 = \cdots = j_m h_m = jh$, where $h_1 = \frac{1}{n_1+1}, \dots, h_m = \frac{1}{n_m+1}$, $h = \frac{1}{n+1}$. In this situation we have $\theta_{j_1,n_1} = \cdots = \theta_{j_m,n_m} = \theta_{j,n} = \bar{\theta}$ for some $\bar{\theta} \in (0, \pi)$, and the application of (1–1) with $\alpha = m$ yields

$$E_{j_i,n_i,0} = \lambda_{j_i}(T_{n_i}(f)) - f(\bar{\theta}) = \sum_{k=1}^{m} c_k(\bar{\theta}) h_i^k + E_{j_i,n_i,m}, \quad i = 1, \dots, m, \tag{3–8}$$

$$E_{j,n,0} = \lambda_j(T_n(f)) - f(\bar{\theta}) = \sum_{k=1}^{m} c_k(\bar{\theta}) h^k + E_{j,n,m}, \tag{3–9}$$



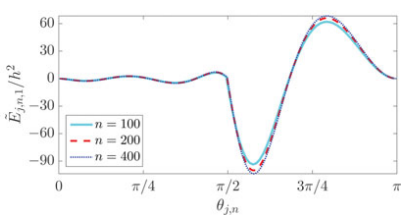


Figure 3. Example 3: Approximated errors $\tilde{E}_{j,n,1}$ and approximated scaled errors $\tilde{E}_{j,n,1}/h^2$ versus $\theta_{j,n}$ for $j = 1, \dots, n$ and $n = 100, 200, 400$ in the case of the symbol $f(\theta) = 1 + 24\cos\theta - 12\cos(2\theta) + 8\cos(3\theta) - 3\cos(4\theta)$.

where

$$|E_{j_i,n_i,m}| \le C_m h_i^{m+1}, \qquad i = 1, \dots, m, \quad (3\text{–}10)$$

$$|E_{j,n,m}| \le C_m h^{m+1}. \quad (3\text{–}11)$$

We are interested in a linear combination of the errors $E_{j_i,n_i,0}$ which "reconstructs" as much as possible the error $E_{j,n,0}$. More precisely, we look for a linear combination

$$\sum_{i=1}^{m} a_i E_{j_i,n_i,0} = \sum_{k=1}^{m} c_k(\bar{\theta}) \sum_{i=1}^{m} a_i h_i^k + \sum_{i=1}^{m} a_i E_{j_i,n_i,m} \quad (3\text{–}12)$$

such that

$$\sum_{i=1}^{m} a_i h_i^k = h^k, \qquad k = 1, \dots, m. \quad (3\text{–}13)$$

If $[\hat{a}_1, \dots, \hat{a}_m]$ is a vector satisfying the conditions (3–13), then

$$\sum_{i=1}^{m} \hat{a}_i E_{j_i,n_i,0} = E_{j,n,0} + \sum_{i=1}^{m} \hat{a}_i E_{j_i,n_i,m} - E_{j,n,m}, \quad (3\text{–}14)$$

and in view of (3–10) and (3–11) the linear combination $\sum_{i=1}^{m} \hat{a}_i E_{j_i,n_i,0}$ is supposed to be an accurate reconstruction of $E_{j,n,0}$. This immediately yields the following high precision approximation for $\lambda_j(T_n(f))$:

$$\lambda_j(T_n(f)) = f(\bar{\theta}) + E_{j,n,0} \approx f(\bar{\theta}) + \sum_{i=1}^{m} \hat{a}_i E_{j_i,n_i,0}. \quad (3\text{–}15)$$

By (3–10), (3–11), and (3–14), an estimate for the error of this approximation is given by

$$\begin{aligned} &\left| \lambda_j(T_n(f)) - f(\bar{\theta}) - \sum_{i=1}^{m} \hat{a}_i E_{j_i,n_i,0} \right| \\ &= \left| E_{j,n,0} - \sum_{i=1}^{m} \hat{a}_i E_{j_i,n_i,0} \right| = \left| \sum_{i=1}^{m} \hat{a}_i E_{j_i,n_i,m} - E_{j,n,m} \right| \\ &\le C_m \left[\sum_{i=1}^{m} |\hat{a}_i| h_i^{m+1} + h^{m+1} \right]. \end{aligned} \quad (3\text{–}16)$$

Theorem 1. *There exists a unique vector $[\hat{a}_1, \dots, \hat{a}_m] \in \mathbb{R}^m$ satisfying the conditions (3–13) and, moreover, the special linear combination $\sum_{i=1}^{m} \hat{a}_i E_{j_i,n_i,0}$ coincides with $hp(h)$, where $p(x)$ is the interpolation polynomial for the data $(h_1, E_{j_1,n_1,0}/h_1), \dots, (h_m, E_{j_m,n_m,0}/h_m)$.*

Proof. Let $V(h_1, \dots, h_m)$ be the Vandermonde matrix corresponding to the nodes $h_1, \dots, h_m$:

$$V(h_1, \dots, h_m) = \begin{bmatrix} 1 & h_1 & \cdots & h_1^{m-1} \\ 1 & h_2 & \cdots & h_2^{m-1} \\ \vdots & \vdots & & \vdots \\ 1 & h_m & \cdots & h_m^{m-1} \end{bmatrix}.$$

We recall two properties of $V(h_1, \dots, h_m)$ that can be found, e.g., in [Bevilacqua et al. 92, Chapter 5] or [Davis 75, Chapter II]. First, since it is implicitly assumed that $n_1, \dots, n_m$ (and hence also $h_1, \dots, h_m$) are all distinct, the matrix $V(h_1, \dots, h_m)$ is invertible. Second, for any $\mathbf{y} = [y_1, \dots, y_m]^T \in \mathbb{R}^m$, the vector $\mathbf{q} = [V(h_1, \dots, h_m)]^{-1}\mathbf{y} = [q_1, \dots, q_m]^T$ is such that $q(x) = q_1 + q_2 x + \cdots + q_m x^{m-1}$ is the interpolation polynomial for the data $(h_1, y_1), \dots, (h_m, y_m)$.

The conditions (3–13) can be rewritten as

$$\begin{bmatrix} h_1 & h_2 & \cdots & h_m \\ h_1^2 & h_2^2 & \cdots & h_m^2 \\ \vdots & \vdots & & \vdots \\ h_1^m & h_2^m & \cdots & h_m^m \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} h \\ h^2 \\ \vdots \\ h^m \end{bmatrix}. \quad (3\text{–}17)$$

If we define

$$D = \begin{bmatrix} h_1 & & & \\ & h_2 & & \\ & & \ddots & \\ & & & h_m \end{bmatrix},$$

then the matrix A of the linear system (3–17) satisfies

$$A = AD^{-1}D = [V(h_1, \dots, h_m)]^T D.$$

It follows that A is invertible and so the linear system (3–17) has a unique solution $[\hat{a}_1, \dots, \hat{a}_m]^T$. Moreover, we have

$$\begin{aligned} A \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_m \end{bmatrix} = \begin{bmatrix} h \\ h^2 \\ \vdots \\ h^m \end{bmatrix} \\ \iff [\hat{a}_1, \hat{a}_2, \dots, \hat{a}_m]A^T = [h, h^2, \dots, h^m] \\ \iff [\hat{a}_1, \hat{a}_2, \dots, \hat{a}_m] = h[1, h, \dots, h^{m-1}]A^{-T}. \end{aligned}$$

If we denote by $p(x) = p_1 + p_2 x + \cdots + p_m x^{m-1}$ the interpolation polynomial for the data $(h_1, E_{j_1,n_1,0}/h_1), \dots, (h_m, E_{j_m,n_m,0}/h_m)$, then

$$\begin{aligned} &\sum_{i=1}^{m} \hat{a}_i E_{j_i,n_i,0} \\ &= [\hat{a}_1, \hat{a}_2, \dots, \hat{a}_m] \begin{bmatrix} E_{j_1,n_1,0} \\ E_{j_2,n_2,0} \\ \vdots \\ E_{j_m,n_m,0} \end{bmatrix} \\ &= h[1, h, \dots, h^{m-1}]A^{-T} \begin{bmatrix} E_{j_1,n_1,0} \\ E_{j_2,n_2,0} \\ \vdots \\ E_{j_m,n_m,0} \end{bmatrix} \end{aligned}$$

$$= h[1, h, \dots, h^{m-1}][V(h_1, \dots, h_m)]^{-1} D^{-1} \begin{bmatrix} E_{j_1,n_1,0} \\ E_{j_2,n_2,0} \\ \vdots \\ E_{j_m,n_m,0} \end{bmatrix}$$

$$= h[1, h, \dots, h^{m-1}][V(h_1, \dots, h_m)]^{-1} \begin{bmatrix} E_{j_1,n_1,0}/h_1 \\ E_{j_2,n_2,0}/h_2 \\ \vdots \\ E_{j_m,n_m,0}/h_m \end{bmatrix}$$

$$= h[1, h, \dots, h^{m-1}] \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_m \end{bmatrix} = h \sum_{i=1}^{m} p_i h^{i-1} = hp(h).$$

□

We remark that n is normally much larger than $n_1, \dots, n_m$. Indeed, the idea behind the algorithm we are describing here is to obtain a high precision approximation of $\lambda_j(T_n(f))$ at the sole price of computing a few eigenvalues $\lambda_{j_1}(T_{n_1}(f)), \dots, \lambda_{j_m}(T_{n_m}(f))$ with $n_1, \dots, n_m \ll n$. Due to the moderate sizes $n_1, \dots, n_m$, the latter eigenvalues can be efficiently computed by a standard eigensolver, and the desired approximation of $\lambda_j(T_n(f))$ is then obtained via equation (3–15) with the $\hat{a}_i$ given by Theorem 1, i.e.,

$$\lambda_j(T_n(f)) = f(\bar{\theta}) + E_{j,n,0} \approx f(\bar{\theta}) + \sum_{i=1}^{m} \hat{a}_i E_{j_i,n_i,0}$$
$$= f(\bar{\theta}) + hp(h). \tag{3–18}$$

An estimate for the error of this approximation is given by (3–16):

$$\left|\lambda_j(T_n(f)) - f(\bar{\theta}) - hp(h)\right| \leq C_m \left[\sum_{i=1}^{m} |\hat{a}_i| h_i^{m+1} + h^{m+1}\right]. \tag{3–19}$$

The procedure of evaluating the interpolation polynomial $p(x)$ at $x = h$ is referred to as extrapolation, because $p(x)$ is evaluated at a point which lies outside the convex hull of the interpolation nodes $h_1, \dots, h_m$. A completely analogous extrapolation procedure is employed in the context of Romberg integration to obtain high precision approximations of an integral from a few coarse trapezoidal approximations; see [Stoer and Bulirsch 02, Section 3.4]. For more details on extrapolation methods, we refer the reader to [Brezinski and Redivo Zaglia 91].

Algorithm 1. With the notation of this article, given f and $m+1$ pairs $(j_1, n_1), \dots, (j_m, n_m), (j, n)$ such that $j_1 h_1 = \cdots = j_m h_m = jh$, we compute a high precision approximation of $\lambda_j(T_n(f))$ as follows:

- Compute the eigenvalues $\lambda_{j_1}(T_{n_1}(f)), \dots, \lambda_{j_m}(T_{n_m}(f))$ using a standard eigensolver.
- Compute the errors $E_{j_i,n_i,0} = \lambda_{j_i}(T_{n_i}(f)) - f(\bar{\theta})$ for $i = 1, \dots, m$, where $\bar{\theta} = \theta_{j,n} = j\pi h$.
- Compute $p(h)$, where $p(x)$ is the interpolation polynomial for the data $(h_i, E_{j_i,n_i,0}/h_i),\ i = 1, \dots, m$.
- Return $f(\bar{\theta}) + hp(h)$.

Example 4. As in Examples 2 and 3, let f be the monotone decreasing RCTP defined by

$$f(\theta) = 1 + 24\cos\theta - 12\cos(2\theta) + 8\cos(3\theta) - 3\cos(4\theta).$$

Suppose we are interested in the jth largest eigenvalue $\lambda_j(T_n(f))$ for $(j, n+1) = (100, 1000)$. Note that n is not dramatically large in this case, so we may compute $\lambda_j(T_n(f))$ by a standard eigensolver, thus obtaining

$$\lambda_j(T_n(f)) = 17.89119035373482\dots \tag{3–20}$$

Let us now compute the approximation of $\lambda_j(T_n(f))$ given by Algorithm 1 with $(j_1, n_1+1) = (4, 40)$, $(j_2, n_2+1) = (5, 50)$, $(j_3, n_3+1) = (10, 100)$. We follow the algorithm step by step.

- Due to the small size of n_1, n_2, n_3, the eigenvalues $\lambda_{j_1}(T_{n_1}(f))$, $\lambda_{j_2}(T_{n_2}(f))$, $\lambda_{j_3}(T_{n_3}(f))$ can be efficiently computed by, say, the MATLAB `eig` function, which yields the values

$$\lambda_{j_1}(T_{n_1}(f)) = 17.86119786677332\dots$$
$$\lambda_{j_2}(T_{n_2}(f)) = 17.86764984932256\dots$$
$$\lambda_{j_3}(T_{n_3}(f)) = 17.88024043750535\dots$$

- In this example we have $\bar{\theta} = \theta_{j,n} = \pi/10$, and the errors $E_{j_1,n_1,0}$, $E_{j_2,n_2,0}$, $E_{j_3,n_3,0}$ are given by

$$E_{j_1,n_1,0} = \lambda_{j_1}(T_{n_1}(f)) - f(\bar{\theta}) = -0.03118562702593\dots$$
$$E_{j_2,n_2,0} = \lambda_{j_2}(T_{n_2}(f)) - f(\bar{\theta}) = -0.02473364447669\dots$$
$$E_{j_3,n_3,0} = \lambda_{j_3}(T_{n_3}(f)) - f(\bar{\theta}) = -0.01214305629390\dots$$

- Let $p(x)$ be the interpolation polynomial for the data $(h_1, E_{j_1,n_1,0}/h_1)$, $(h_2, E_{j_2,n_2,0}/h_2)$, $(h_3, E_{j_3,n_3,0}/h_3)$. The value $p(h)$ can be computed from the Lagrange form of $p(x)$:

$$p(h) = \frac{E_{j_1,n_1,0}}{h_1}\frac{(h-h_2)(h-h_3)}{(h_1-h_2)(h_1-h_3)} + \frac{E_{j_2,n_2,0}}{h_2}\frac{(h-h_1)(h-h_3)}{(h_2-h_1)(h_2-h_3)} + \frac{E_{j_3,n_3,0}}{h_3}\frac{(h-h_1)(h-h_2)}{(h_3-h_1)(h_3-h_2)} = -1.19315109114712\dots$$

Table 1. Example 5: Comparison between $\lambda_j(T_n(f))$ and $f(\bar\theta)+hp(h)$ for several RCTPs f.

f	$\lambda_j(T_n(f))$	$f(\bar\theta)+hp(h)$	Error $\lvert\lambda_j(T_n(f))-f(\bar\theta)-hp(h)\rvert$	Error Estimate $C_3\left[\sum_{i=1}^3\lvert\hat a_i\rvert h_i^4+h^4\right]$
f_2	1.07487275461020	1.07487275470961	$9.94\cdot10^{-11}$	$C_3\cdot 9.47\cdot10^{-10}$
f_3	1.11519899215300	1.11519899090697	$1.25\cdot10^{-9}$	$C_3\cdot 9.47\cdot10^{-10}$
f_4	1.15757333445321	1.15757329396605	$4.05\cdot10^{-8}$	$C_3\cdot 9.47\cdot10^{-10}$

- The approximation of $\lambda_j(T_n(f))$ returned by the algorithm is

$$\lambda_j(T_n(f)) \approx f(\bar\theta)+hp(h) = 17.89119034270811\ldots \quad (3\text{–}21)$$

A direct comparison between (3–20) and (3–21) shows that $|\lambda_j(T_n(f))-f(\bar\theta)-hp(h)|\approx 1.10\cdot10^{-8}$(!). Assuming we have no information about the exact value (3–20), we can estimate the error $|\lambda_j(T_n(f))-f(\bar\theta)-hp(h)|$ via (3–19). The coefficients $\hat a_1,\hat a_2,\hat a_3$ are easily computed by solving the linear system (3–17), which in this case becomes

$$\begin{bmatrix} h_1 & h_2 & h_3\\ h_1^2 & h_2^2 & h_3^2\\ h_1^3 & h_2^3 & h_3^3\end{bmatrix}\begin{bmatrix}\hat a_1\\ \hat a_2\\ \hat a_3\end{bmatrix}=\begin{bmatrix} h\\ h^2\\ h^3\end{bmatrix}$$
$$\Longleftrightarrow \quad \begin{bmatrix}\hat a_1\\ \hat a_2\\ \hat a_3\end{bmatrix}=\begin{bmatrix} 0.0912\\ -0.216\\ 0.304\end{bmatrix}.$$

By (3–19),

$$|\lambda_j(T_n(f))-f(\bar\theta)-hp(h)|\le C_3\cdot 7.33\cdot10^{-8},$$

where C_3 is a constant depending only on f.

Example 5. In this example, for several RCTPs f and for the fixed pair $(j,n)=(1700,5000)$, we compare $\lambda_j(T_n(f))$ to its approximation $f(\bar\theta)+hp(h)$ provided by Algorithm 1 with $(j_1,n_1+1)=(17,50)$, $(j_2,n_2+1)=(34,100)$, $(j_3,n_3+1)=(68,200)$. The results of this comparison are collected in Table 1 for $f=f_q$ and $q=2,3,4$, where f_q is defined in (1–2). Note that the error estimate in the last column seems to be the same in all cases, but it must be recalled that the constant C_3 depends on f.

4. Numerical experiments and a conjecture for the non-monotone case

Consider the non-monotone RCTP $f(\theta)=2+2\cos\theta-2\cos(2\theta)$, whose graph over $[0,\pi]$ is depicted in Figure 4. Note that f restricted to the interval $I=(2\pi/3,\pi]$ is monotone and $f^{-1}(f(I))=I$, where $f(I)=\{f(\theta):\ \theta\in I\}=[-2,2)$ and $f^{-1}(f(I))=\{\theta\in[0,\pi]:\ f(\theta)\in f(I)\}$. Let $\lambda_1(T_n(f)),\ldots,\lambda_n(T_n(f))$ be the eigenvalues of $T_n(f)$ arranged in non-decreasing order, and let σ_n be a permutation of $\{1,\ldots,n\}$ which sorts the samples $f(\theta_{1,n}),\ldots,f(\theta_{n,n})$ in non-decreasing order, i.e., $f(\theta_{\sigma_n(1),n})\le\cdots\le f(\theta_{\sigma_n(n),n})$. Note that the inverse permutation $\rho_n=\sigma_n^{-1}$ is supposed to sort the eigenvalues $\lambda_1(T_n(f)),\ldots,\lambda_n(T_n(f))$ so that they match the samples $f(\theta_{1,n}),\ldots,f(\theta_{n,n})$, i.e., $\lambda_{\rho_n(j)}(T_n(f))$ should be approximately equal to $f(\theta_{j,n})$ for all $j=1,\ldots,n$. In Figure 5 we plot the errors

$$E_{j,n,0}=\lambda_{\rho_n(j)}(T_n(f))-f(\theta_{j,n}) \quad (4\text{–}22)$$

and the scaled errors $E_{j,n,0}/h$ versus $\theta_{j,n}$ for $j=1,\ldots,n$ and $n=100,200,400$. The fundamental observation is that, as long as $\theta_{j,n}\in I$, the errors $E_{j,n,0}$ draw a smooth curve and the scaled errors $E_{j,n,0}/h$ overlap perfectly, just as in the case of monotone RCTPs (see Figures 1 and 2). We may therefore conjecture that the asymptotic expansion (1–1) holds for the eigenvalues of $T_n(f)$ corresponding in (4–22) to the samples $f(\theta_{j,n})$ with $\theta_{j,n}\in I$. These are essentially the eigenvalues belonging to $f(I)=[-2,2)$. The precise statement of our conjecture is reported below along with a further example supporting it.

Conjecture 1. Let f be an RCTP such that f restricted to the interval $I\subseteq[0,\pi]$ is monotone and $f^{-1}(f(I))=I$. Then, for every integer $\alpha\ge0$, every n and every $j=1,\ldots,n$ such that $\theta_{j,n}\in I$, the following asymptotic expansion holds:

$$\lambda_{\rho_n(j)}(T_n(f))=f(\theta_{j,n})+\sum_{k=1}^{\alpha}c_k(\theta_{j,n})h^k+E_{j,n,\alpha}, \quad (4\text{–}23)$$

where:

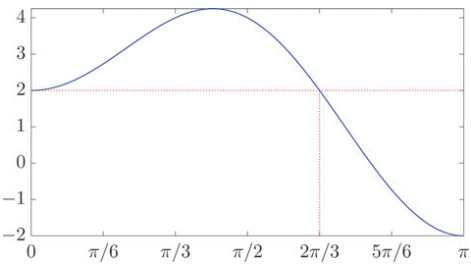


Figure 4. Graph of $f(\theta)=2+2\cos\theta-2\cos(2\theta)$ over $[0,\pi]$.

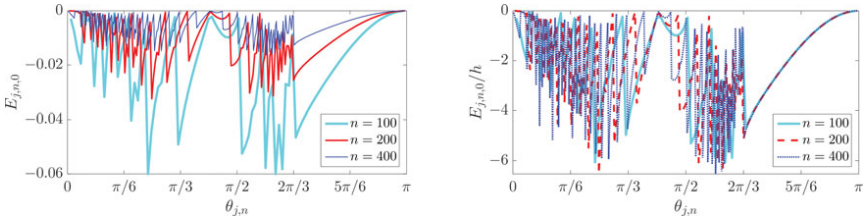


Figure 5. Errors $E_{j,n,0}$ and scaled errors $E_{j,n,0}/h$ versus $\theta_{j,n}$ for $j = 1, \dots, n$ and $n = 100, 200, 400$ in the case of the symbol $f(\theta) = 2 + 2\cos\theta - 2\cos(2\theta)$.

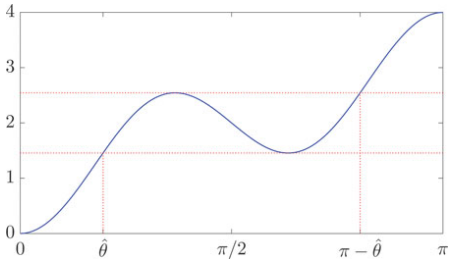


Figure 6. Example 6: Graph of $f(\theta) = 2 - \cos\theta - \cos(3\theta)$ over $[0, \pi]$.

- The eigenvalues of $T_n(f)$ are arranged in non-decreasing order, $\lambda_1(T_n(f)) \le \cdots \le \lambda_n(T_n(f))$.
- $\rho_n = \sigma_n^{-1}$, where σ_n is a permutation of $\{1, \dots, n\}$ such that $f(\theta_{\sigma_n(1),n}) \le \cdots \le f(\theta_{\sigma_n(n),n})$.
- $\{c_k\}_{k=1,2,\dots}$ is a sequence of functions from I to $\mathbb{R}$ which depends only on f.
- $h = \frac{1}{n+1}$ and $\theta_{j,n} = \frac{j\pi}{n+1} = j\pi h$.
- $E_{j,n,\alpha} = O(h^{\alpha+1})$ is the error, which satisfies the inequality $|E_{j,n,\alpha}| \le C_\alpha h^{\alpha+1}$ for some constant C_α depending only on α and f.

For $\alpha = 0$, this conjecture is the same as Bogoya, Böttcher, Grudsky, and Maximenko's result (1–3).

Example 6. Let

$$f(\theta) = 2 - \cos\theta - \cos(3\theta).$$

The graph of f is depicted in Figure 6. The hypotheses of Conjecture 1 are satisfied with either $I = [0, \hat{\theta})$ or $I = (\pi - \hat{\theta}, \pi]$, where $\hat{\theta} = 0.61547970867038\dots$. To fix the ideas, let $I = [0, \hat{\theta})$. Conjecture 1 with $\alpha = 1$ would say that, for every n and every $j = 1, \dots, n$ such that $\theta_{j,n} \in I$,

$$\lambda_{\rho_n(j)}(T_n(f)) - f(\theta_{j,n}) = E_{j,n,0} = c_1(\theta_{j,n}) + E_{j,n,1},$$

where $|E_{j,n,1}| \le C_1 h^2$ and both the function $c_1 : I \to \mathbb{R}$ and the constant C_1 depend only on f. In particular, the scaled errors $E_{j,n,0}/h$ corresponding to the points $\theta_{j,n}$ in I should be equal to the equispaced samples $c_1(\theta_{j,n})$ (and should therefore reproduce the graph of c_1) in the limit where $n \to \infty$. In Figure 7 we plot the errors and the scaled errors versus $\theta_{j,n}$ for $j = 1, \dots, n$ and $n = 100, 200, 400$. Clearly, the scaled errors overlap perfectly over I, thus supporting Conjecture 1. We remark that nothing would have changed in the reasoning if we had chosen $I = (\pi - \hat{\theta}, \pi]$.

Assuming Conjecture 1, we can follow the derivation of Section 3 to work out an algorithm, analogous to Algorithm 1, for computing a high precision approximation of $\lambda_{\rho_n(j)}(T_n(f))$ from $\lambda_{\rho_{n_1}(j_1)}(T_{n_1}(f)), \dots, \lambda_{\rho_{n_m}(j_m)}(T_{n_m}(f))$, provided the corresponding point $\theta_{j_1,n_1} = \cdots = \theta_{j_m,n_m} = \theta_{j,n} = \bar{\theta}$ belongs to an interval $I \subseteq [0, \pi]$ such that $f_{|I}$ is monotone and $f^{-1}(f(I)) = I$. We report here the algorithm for the reader's convenience.

Algorithm 2. With the notation of this article, given f and $m + 1$ pairs $(j_1, n_1), \dots, (j_m, n_m), (j, n)$ such that $j_1 h_1 = \cdots = j_m h_m = jh$, we compute a high precision approximation of $\lambda_{\rho_n(j)}(T_n(f))$ as follows:

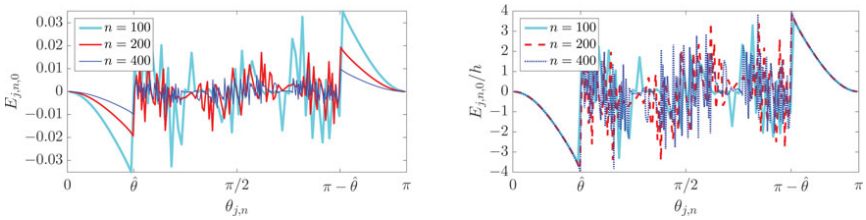


Figure 7. Example 6: Errors $E_{j,n,0}$ and scaled errors $E_{j,n,0}/h$ versus $\theta_{j,n}$ for $j = 1, \dots, n$ and $n = 100, 200, 400$ in the case of the symbol $f(\theta) = 2 - \cos\theta - \cos(3\theta)$.

Table 2. Example 7: Comparison between $\lambda_j(T_n(f))$ and $f(\bar\theta)+hp(h)$ for $m=1,\dots,5$.

m	$\lambda_j(T_n(f))$	$f(\bar\theta)+hp(h)$	Error $\lvert\lambda_j(T_n(f))-f(\bar\theta)-hp(h)\rvert$	Error estimate $C_m\left[\sum_{i=1}^{m}\lvert\hat a_i\rvert h_i^{m+1}+h^{m+1}\right]$
1	0.46103961732270	0.46104722829886	$7.61\cdot10^{-6}$	$C_1\cdot 3.34\cdot10^{-6}$
2	0.46103961732270	0.46103991187671	$2.94\cdot10^{-7}$	$C_2\cdot 2.65\cdot10^{-7}$
3	0.46103961732270	0.46103962607810	$8.76\cdot10^{-9}$	$C_3\cdot 1.08\cdot10^{-8}$
4	0.46103961732270	0.46103961753594	$2.13\cdot10^{-10}$	$C_4\cdot 3.01\cdot10^{-10}$
5	0.46103961732270	0.46103961733097	$8.27\cdot10^{-12}$	$C_5\cdot 6.39\cdot10^{-12}$

- Compute the eigenvalues $\lambda_{\rho_{n_1}(j_1)}(T_{n_1}(f)),\dots,\lambda_{\rho_{n_m}(j_m)}(T_{n_m}(f))$ using a standard eigensolver.
- Compute the errors $E_{j_i,n_i,0}=\lambda_{\rho_{n_i}(j_i)}(T_{n_i}(f))-f(\bar\theta)$ for $i=1,\dots,m$, where $\bar\theta=\theta_{j,n}=j\pi h$.
- Compute $p(h)$, where $p(x)$ is the interpolation polynomial for the data $(h_i, E_{j_i,n_i,0}/h_i)$, $i=1,\dots,m$.
- Return $f(\bar\theta)+hp(h)$.

Example 7. Let f be the same as in Example 6. Suppose we are interested in the jth smallest eigenvalue $\lambda_j(T_n(f))$ for $(j,n+1)=(1000,10000)$. The point $\bar\theta=\theta_{j,n}=\pi/10$ lies in $I=[0,\hat\theta)$, $f_{|I}$ is monotone and $f^{-1}(f(I))=I$ (see Figure 6). Moreover, it is clear that the permutation σ_n which sorts the samples $f(\theta_{1,n}),\dots,f(\theta_{n,n})$ in non-decreasing order is such that $\sigma_n(\ell)=\ell$ for all $\ell=1,2,\dots,\hat\ell$, where $\hat\ell$ is the first index such that $\theta_{\hat\ell+1,n}\geq\hat\theta$. As a consequence, $\rho_n(j)=j$. In Table 2 we compare $\lambda_j(T_n(f))$ to its approximations $f(\bar\theta)+hp(h)$ provided by Algorithm 2 with $m=1,\dots,5$ and $(j_1,n_1+1)=(3,30)$, $(j_2,n_2+1)=(5,50)$, $(j_3,n_3+1)=(7,70)$, $(j_4,n_4+1)=(9,90)$, $(j_5,n_5+1)=(11,110)$. Note that, for the same reasoning as above, $\rho_{n_m}(j_m)=j_m$ for all $m=1,\dots,5$.

5. Conclusions and perspectives

After supporting through numerical experiments the conjecture that the higher-order approximation (1–1) holds for all monotone RCTPs f, we illustrated how (1–1) can be used along with an extrapolation procedure to compute high precision approximations of the eigenvalues of $T_n(f)$ for large n. Moreover, based on numerical experiments, we formulated a conjecture on the eigenvalue asymptotics of $T_n(f)$ in the case where f is non-monotone, and we showed how the conjecture can be used, again in combination with an extrapolation procedure, to compute high precision approximations of some eigenvalues of $T_n(f)$ for large n.

We conclude this work with a list of possible future lines of research.

- Conjecture 1 does not say anything about "fully non-monotone" symbols such as $f(\theta)=2-2\cos(\omega\theta)$, where $\omega\geq2$ is an integer. However, based on numerical experiments, it seems that even in this case a "regular" asymptotics is available for the eigenvalues of $T_n(f)$. For more insights into this topic we refer the reader to papers [Barrera and Grudsky 17] and [Ekström and Serra-Capizzano].
- A noteworthy theoretical objective would be to obtain a precise analytic expression for the error of Algorithm 1, namely $|\lambda_j(T_n(f))-f(\bar\theta)-hp(h)|$. A way to achieve this goal could be to exploit the information about the functions c_k provided in [Bogoya et al. 15a, Bogoya et al. 17, Böttcher et al. 10] and follow the steps in the derivation of the analytic expression for the error of Romberg integration [Bauer 61, Bauer et al. 63].
- With any multi-index $n=(n_1,\dots,n_d)\in\mathbb{N}^d$ and any multivariate matrix-valued function $f:[-\pi,\pi]^d\to\mathbb{C}^{s\times s}$ whose components f_{ij} belong to $L^1([-\pi,\pi]^d)$, we associate the so-called multilevel block Toeplitz matrix $T_n(f)$, which is defined, e.g., in [Tilli 98]. In view of the design of fast extrapolation algorithms for the computation of the eigenvalues, it would be interesting to know whether an asymptotic expansion such as (1–1) or (4–23) holds even for this kind of matrices. Numerical evidence indicates that the answer should be affirmative if

$$f(\theta_1,\dots,\theta_d)=\sum_{i=1}^{d}f_q(\theta_i),\quad q=1,2,\dots \tag{5–24}$$

where f_q is given by (1–2). The d-variate function f is especially interesting as it arises in the discretization of partial differential equations over d-dimensional domains. For this function, however, we do not need any asymptotic expansion to efficiently compute the eigenvalues of $T_n(f)$. Indeed, due to the specific structure of f, it can be shown that

$$T_n(f)=\sum_{i=1}^{d}I_{n_1}\otimes\cdots\otimes I_{n_{i-1}}\otimes T_{n_i}(f_q)\otimes I_{n_{i+1}}\otimes\cdots\otimes I_{n_d},$$

where I_m is the $m\times m$ identity matrix and $\otimes$ denotes the (Kronecker) tensor product of matrices. By the

properties of tensor products, the eigenvalues of $T_n(f)$ are given by

$$\lambda_j(T_n(f)) = \sum_{i=1}^{d} \lambda_{j_i}(T_{n_i}(f_q)), \qquad 1 \le j_1 \le n_1, \ \ldots, \ 1 \le j_d \le n_d,$$

and their computation reduces to the computation of the eigenvalues of the unilevel Toeplitz matrices $T_m(f_q)$, which can be performed through Algorithm 1. For functions f more general than (5–24), the reduction to the unilevel setting is not possible. In this case, an extrapolation algorithm for the computation of the eigenvalues of $T_n(f)$ should directly rely on the asymptotic expansion, and establishing whether the latter exists or not is an interesting subject for future research.

Funding

Sven-Erik Ekström is a PhD student at TDB (Division of Scientific Computing, Uppsala University); his research is cofinanced by the ADIGMA Project, the Graduate School in Mathematics and Computing (FMB), and Uppsala University. Carlo Garoni is a Marie-Curie fellow of the Italian INdAM (Istituto Nazionale di Alta Matematica); his research is cofinanced by INdAM and the European "Marie-Curie Actions" Programme through the Grant PCOFUND-GA-2012-600198. The research of Stefano Serra-Capizzano is partially financed by the INdAM GNCS (Gruppo Nazionale per il Calcolo Scientifico).

ORCID

Sven-Erik Ekström http://orcid.org/0000-0002-7875-7543
Carlo Garoni http://orcid.org/0000-0001-9720-092X
Stefano Serra-Capizzano http://orcid.org/0000-0001-9477-109X

References

[Barrera and Grudsky 17] M. Barrera and S. M. Grudsky. "Asymptotics of Eigenvalues for Pentadiagonal Symmetric Toeplitz Matrices." *Oper. Theory Adv. Appl.* 259 (2017), 51–77.

[Bauer 61] F. L. Bauer. "La méthode d'intégration numérique de Romberg." Colloque sur l'analyse numérique , Librairie Universitaire, Louvain, 22–24 Mars 1961 à Mons, 119–129, 1961.

[Bauer et al. 63] F. L. Bauer, H. Rutishauser, and E. Stiefel. "New Aspects in Numerical Quadrature." *Proc. Symp. Appl. Math.* 15 (1963), 199–218.

[Bevilacqua et al. 92] R. Bevilacqua, D. Bini, M. Capovani, and O. Menchi. *Metodi Numerici*. Bologna, Italy: Zanichelli, 1992.

[Bhatia 97] R. Bhatia. *Matrix Analysis*. New York: Springer, 1997.

[Bini and Capovani 83] D. Bini and M. Capovani. "Spectral and Computational Properties of Band Symmetric Toeplitz Matrices." *Linear Algebra Appl.* 52–53 (1983), 99–126.

[Bogoya et al. 15a] J. M. Bogoya, A. Böttcher, S. M. Grudsky, and E. A. Maximenko. "Eigenvalues of Hermitian Toeplitz Matrices with Smooth Simple-loop Symbols." *J. Math. Anal. Appl.* 422 (2015), 1308–1334.

[Bogoya et al. 15b] J. M. Bogoya, A. Böttcher, S. M. Grudsky, and E. A. Maximenko. "Maximum Norm Versions of the Szegő and Avram–Parter Theorems for Toeplitz Matrices." *J. Approx. Theory* 196 (2015), 79–100.

[Bogoya et al. 16] J. M. Bogoya, A. Böttcher, and E. A. Maximenko. "From Convergence in Distribution to Uniform Convergence." *Bol. Soc. Mat. Mex.* 22 (2016), 695–710.

[Bogoya et al. 17] J. M. Bogoya, S. M. Grudsky, and E. A. Maximenko. "Eigenvalues of Hermitian Toeplitz Matrices Generated by Simple-loop Symbols with Relaxed Smoothness." *Oper. Theory Adv. Appl.* 259 (2017), 179–212.

[Böttcher et al. 10] A. Böttcher, S. M. Grudsky, and E. A. Maximenko. "Inside the Eigenvalues of Certain Hermitian Toeplitz Band Matrices." *J. Comput. Appl. Math.* 233 (2010), 2245–2264.

[Böttcher and Silbermann 99] A. Böttcher and B. Silbermann. *Introduction to Large Truncated Toeplitz Matrices*. New York: Springer, 1999.

[Brezinski and Redivo Zaglia 91] C. Brezinski and M. Redivo Zaglia. *Extrapolation Methods: Theory and Practice*. North-Holland: Elsevier Science Publishers B.V., 1991.

[Davis 75] P. J. Davis. *Interpolation and Approximation*. Mineola, NY: Dover Publications, 1975.

[Ekström and Serra-Capizzano] S.-E. Ekström and S. Serra-Capizzano. "Eigenvalues and Eigenvectors of Banded Toeplitz Matrices and the Related Symbols." In preparation.

[Garoni and Serra-Capizzano 17] C. Garoni and S. Serra-Capizzano. *Generalized Locally Toeplitz Sequences: Theory and Applications (Volume I)*. Cham, Switzerland: Springer International Publishing AG, 2017.

[Serra-Capizzano 96] S. Serra-Capizzano. "On the Extreme Spectral Properties of Toeplitz Matrices Generated by L^1 Functions with Several Minima/Maxima." *BIT* 36 (1996), 135–142.

[Stoer and Bulirsch 02] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*. Third edition, New York: Springer, 2002.

[Tilli 98] P. Tilli. "A Note on the Spectral Distribution of Toeplitz Matrices." *Linear Multilinear Algebra* 45 (1998), 147–159.

Paper II [redacted]

Numer Algor
DOI 10.1007/s11075-017-0404-z

ORIGINAL PAPER

Are the eigenvalues of preconditioned banded symmetric Toeplitz matrices known in almost closed form?

Fayyaz Ahmad[1] · Eman Salem Al-Aidarous[2] · Dina Abdullah Alrehaili[2] · Sven-Erik Ekström[3] · Isabella Furci[1] · Stefano Serra-Capizzano[1,3]

Received: 28 June 2017 / Accepted: 15 August 2017

Abstract Bogoya, Böttcher, Grudsky, and Maximenko have recently obtained the precise asymptotic expansion for the eigenvalues of a sequence of Toeplitz matrices $\{T_n(f)\}$, under suitable assumptions on the associated generating function f. In this paper, we provide numerical evidence that some of these assumptions can be relaxed and extended to the case of a sequence of preconditioned Toeplitz matrices $\{T_n^{-1}(g)T_n(f)\}$, for f trigonometric polynomial, g nonnegative, not identically zero trigonometric polynomial, $r = f/g$, and where the ratio r plays the same role as f

✉ Sven-Erik Ekström
sven-erik.ekstrom@it.uu.se

Fayyaz Ahmad
fahmad@uninsubria.it

Eman Salem Al-Aidarous
ealaidarous@kau.edu.sa

Dina Abdullah Alrehaili
dalrehaili@stu.kau.edu.sa

Isabella Furci
ifurci@uninsubria.it

Stefano Serra-Capizzano
stefano.serrac@uninsubria.it; stefano.serra@it.uu.se

[1] Department of Science and High Technology, University of Insubria, Via Valleggio 11, 22100 Como, Italy

[2] Faculty of Science, Department of Mathematics, King Abdulaziz University, P.O. Box 80203, 21589 Jeddah, Saudi Arabia

[3] Division of Scientific Computing, Department of Information Technology, ITC, Uppsala University, Lägerhyddsv. 2, Hus 2, P.O. Box 337, 751 05 Uppsala, Sweden

Published online: 31 August 2017

in the nonpreconditioned case. Moreover, based on the eigenvalue asymptotics, we devise an extrapolation algorithm for computing the eigenvalues of preconditioned banded symmetric Toeplitz matrices with a high level of accuracy, with a relatively low computational cost, and with potential application to the computation of the spectrum of differential operators.

Keywords (Preconditioned) Toeplitz matrix · Mass and stiffness matrix · Eigenvalues · Eigenvalue asymptotics · Polynomial interpolation · Extrapolation

Mathematics Subject Classifications (2010) 15B05 · 65F15 · 65D05 · 65B05

1 Introduction

A matrix of size n, having a fixed entry along each diagonal, is called Toeplitz and enjoys the expression

$$\left[a_{i-j}\right]_{i,j=1}^{n} = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & \cdots & a_{-(n-1)} \\ a_1 & \ddots & \ddots & \ddots & & \vdots \\ a_2 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & a_{-2} \\ \vdots & & \ddots & \ddots & \ddots & a_{-1} \\ a_{n-1} & \cdots & \cdots & a_2 & a_1 & a_0 \end{bmatrix}.$$

Given a complex-valued Lebesgue integrable function $\phi : [-\pi, \pi] \to \mathbb{C}$, the nth Toeplitz matrix generated by ϕ is defined as

$$T_n(\phi) = \left[\hat{\phi}_{i-j}\right]_{i,j=1}^{n},$$

where the quantities $\hat{\phi}_k$ are the Fourier coefficients of ϕ, which means

$$\hat{\phi}_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi(\theta)\, \mathrm{e}^{-\mathbf{i}k\theta} \mathrm{d}\theta, \qquad k \in \mathbb{Z}.$$

We refer to $\{T_n(\phi)\}_n$ as the Toeplitz sequence generated by ϕ, which in turn is called the generating function of $\{T_n(\phi)\}_n$. In the case where ϕ is real-valued, all the matrices $T_n(\phi)$ are Hermitian and much is known about their spectral properties, from the localization of the eigenvalues to the asymptotic spectral distribution in the Weyl sense: in particular ϕ is the spectral symbol of $\{T_n(\phi)\}_n$, see [7, 14] and the references therein.

More in detail, if ϕ is real-valued and not identically constant, then any eigenvalue of $T_n(\phi)$ belongs to the open set (m_ϕ, M_ϕ), with m_ϕ, M_ϕ being the essential infimum, the essential supremum of ϕ, respectively. The case of a constant ϕ is trivial: in that case if $\phi = m$ almost everywhere then $T_n(\phi) = m\mathbb{I}_n$ with $\mathbb{I}_n$ denoting the identity of size n. Hence if $M_\phi > 0$ and ϕ is nonnegative almost everywhere, then $T_n(\phi)$ is Hermitian positive definite.

In this paper, we focus our attention on the following setting.

- We consider two real-valued cosine trigonometric polynomials (RCTPs) f, g, that is

$$f(\theta) = \hat{f}_0 + 2\sum_{k=1}^{m_1} \hat{f}_k \cos(k\theta), \qquad \hat{f}_0, \hat{f}_1, \ldots, \hat{f}_{m_1} \in \mathbb{R}, \qquad m_1 \in \mathbb{N},$$

$$g(\theta) = \hat{g}_0 + 2\sum_{k=1}^{m_2} \hat{g}_k \cos(k\theta), \qquad \hat{g}_0, \hat{g}_1, \ldots, \hat{g}_{m_2} \in \mathbb{R}, \qquad m_2 \in \mathbb{N},$$

  so that $T_n(f)$, $T_n(g)$ are both real symmetric.
- We assume that $M_g = \max g > 0$ and $m_g = \min g \geq 0$, so that $T_n(g)$ is positive definite.
- We consider $\mathcal{P}_n(f, g) = T_n^{-1}(g)T_n(f)$ the "preconditioned" matrix and we define the new symbol $r = f/g$.

The nth Toeplitz matrix generated by $\phi \in \{f, g\}$ is the real symmetric banded matrix of bandwidth $2m + 1$, $m \in \{m_1, m_2\}$ ($m = m_1$ if $\phi = f$ and $m = m_2$ if $\phi = g$), given by

$$T_n(\phi) = \begin{bmatrix} \hat{\phi}_0 & \hat{\phi}_1 & \cdots & \hat{\phi}_m & & & & & & & \\ \hat{\phi}_1 & \ddots & \ddots & & \ddots & & & & & & \\ \vdots & \ddots & \ddots & \ddots & & \ddots & & & & & \\ \hat{\phi}_m & & \ddots & \ddots & \ddots & & \ddots & & & & \\ & \ddots & & \ddots & \ddots & \ddots & & \ddots & & & \\ & & \hat{\phi}_m & \cdots & \hat{\phi}_1 & \hat{\phi}_0 & \hat{\phi}_1 & \cdots & \hat{\phi}_m & & \\ & & & \ddots & & \ddots & \ddots & \ddots & & \ddots & \\ & & & & \ddots & & \ddots & \ddots & \ddots & & \hat{\phi}_m \\ & & & & & \ddots & & \ddots & \ddots & \ddots & \vdots \\ & & & & & & \ddots & & \ddots & \ddots & \hat{\phi}_1 \\ & & & & & & & \hat{\phi}_m & \cdots & \hat{\phi}_1 & \hat{\phi}_0 \end{bmatrix}.$$

Matrices of the form $\mathcal{P}_n(f, g)$ are important for the fast solution of large Toeplitz linear systems (in connection with the preconditioned conjugate gradient method [9–11, 18] or of more general preconditioned Krylov methods [15, 16]). Furthermore, up to low rank corrections, they appear in the context of the spectral approximation of differential operators in which a low rank correction of $T_n(g)$ is the mass matrix and a low rank correction of $T_n(f)$ is the stiffness matrix.

Their spectral features have been studied in detail. More precisely, under the assumption that $r = m$ identically $\mathcal{P}_n(f, g) = r\mathbb{I}_n$, while if $m_r < M_r$, then any eigenvalue of $\mathcal{P}_n(f, g)$ belongs to the open set (m_r, M_r), see [11], and the whole sequence $\{\mathcal{P}_n(f, g)\}_n$ is spectrally distributed in the Weyl sense as $r = f/g$ (see [19]).

In our context, we say that a function is monotone if it is either increasing or decreasing over the interval $[0, \pi]$.

Under the assumption that $r = f/g$ is monotone, in this paper, we show experimentally that for every integer $\alpha \geq 0$, every n and every $j = 1, \ldots, n$, the following asymptotic expansion holds:

$$\lambda_j(\mathcal{P}_n(f, g)) = r(\theta_{j,n}) + \sum_{k=1}^{\alpha} c_k(\theta_{j,n})h^k + E_{j,n,\alpha}, \tag{1}$$

where:

- the eigenvalues of $\mathcal{P}_n(f, g)$ are arranged in nondecreasing or nonincreasing order, depending on whether r is increasing or decreasing;
- $\{c_k\}_{k=1,2,\ldots}$ is a sequence of functions from $[0, \pi]$ to $\mathbb{R}$ which depends only on r;
- $h = \frac{1}{n+1}$ and $\theta_{j,n} = \frac{j\pi}{n+1} = j\pi h$;
- $E_{j,n,\alpha} = O(h^{\alpha+1})$ is the remainder (the error), which satisfies the inequality $|E_{j,n,\alpha}| \leq C_\alpha h^{\alpha+1}$ for some constant C_α depending only on α and r.

In the pure Toeplitz case, that is for $g = 1$ identically, so that $\mathcal{P}_n(f, g) = T_n(f)$ and $r = f$, the result is proven in [4–6], if the RCTP f is monotone and satisfies certain additional assumptions, which include the requirements that $f'(\theta) \neq 0$ for $\theta \in (0, \pi)$ and $f''(\theta) \neq 0$ for $\theta \in \{0, \pi\}$. The symbols

$$f_q(\theta) = (2 - 2\cos\theta)^q, \qquad q = 1, 2, \ldots, \tag{2}$$

arise in the discretization of differential equations and are therefore of particular interest. Unfortunately, for these symbols, the requirement that $f''(0) \neq 0$ is not satisfied if $q \geq 2$. In [13], several numerical evidences are reported, showing that the higher order approximation (1) holds even in this "degenerate case."

Here, as first purpose, we show numerically the same for the preconditioned matrices $\mathcal{P}_n(f, g)$ and, from a theoretical point of view, the numerical testing is complemented by the proof of the above conjecture in the basic case of $\alpha = 0$.

Furthermore, in [13], the authors employed the asymptotic expansion (1) for computing an accurate approximation of $\lambda_j(T_n(f))$ for very large n, provided that the values $\lambda_{j_1}(T_{n_1}(f)), \ldots, \lambda_{j_s}(T_{n_s}(f))$ are available for moderate sizes $n_1, \ldots, n_s$ with $\theta_{j_1,n_1} = \cdots = \theta_{j_s,n_s} = \theta_{j,n}$, $s \geq 2$. The second and main purpose of this paper is to carry out this idea and to support it by numerical experiments, accompanied by an appropriate error analysis in the more general case of the preconditioned matrices $\mathcal{P}_n(f, g)$. In particular, we devise an algorithm to compute $\lambda_j(\mathcal{P}_n(f, g))$ with a high level of accuracy and a relatively low computational cost. The algorithm is completely analogous to the extrapolation procedure, which is employed in the context of Romberg integration (to obtain high precision approximations of an integral from a few coarse trapezoidal approximations [20, Section 3.4], see also [8] for more advanced algorithms). In this regard, the asymptotic expansion (1) plays here the same role as the Euler–Maclaurin summation formula [20, Section 3.3].

The third and last purpose of this paper is to formulate, on the basis of numerical experiments, a conjecture on the higher-order asymptotic of the eigenvalues if the monotonicity assumption on $r = f/g$ is not in force. We also illustrate how this

conjecture can be used along with our extrapolation algorithm in order to compute some of the eigenvalues of $\mathcal{P}_n(f, g)$ in the case where r is nonmonotone.

2 Error bounds for the coefficients c_k in the asymptotic expansion

We start this section by manipulating the error expression implicitly given in (1), the goal being that of using extrapolation methods [8]. In fact, if we assume that the relations in (1) hold, then we can write

$$E_{j,n,0} = \sum_{k=1}^{\alpha} c_k(\theta_{j,n})\, h^k + E_{j,n,\alpha}\,, \tag{3}$$

where $E_{j,n,0} = \lambda_j(\mathcal{P}_n(f, g)) - r(\theta_{j,n})$.

We now suppose to know the eigenvalues for different (small) n_i namely $\{(n_1, \lambda_{j_1}(\mathcal{P}_{n_1}(f, g))),\ (n_2, \lambda_{j_2}(\mathcal{P}_{n_2}(f, g))),\ \cdots,\ (n_\alpha, \lambda_{j_\alpha}(\mathcal{P}_{n_\alpha}(f, g)))\}$, where $n_1, n_2, \cdots, n_\alpha$ and $j_1, j_2, \cdots, j_\alpha$ are chosen in such a way that $j_1/(n_1+1) = j_2/(n_2+1) = \cdots = j_\alpha/(n_\alpha+1)$.

By defining $h_1 = 1/(n_1+1)$, $h_2 = 1/(n_2+1), \ldots, h_\alpha = 1/(n_\alpha+1)$, for a given set of eigenvalues, (3) can be written as

$$\begin{aligned}
E_{j_1,n_1,0} &= \sum_{k=1}^{\alpha} c_k(\theta_{j_1,n_1})\, h_1^k + E_{j_1,n_1,\alpha},\\
E_{j_2,n_2,0} &= \sum_{k=1}^{\alpha} c_k(\theta_{j_2,n_2})\, h_2^k + E_{j_2,n_2,\alpha},\\
E_{j_3,n_3,0} &= \sum_{k=1}^{\alpha} c_k(\theta_{j_3,n_3})\, h_3^k + E_{j_3,n_3,\alpha},\\
&\vdots\\
E_{j_\alpha,n_\alpha,0} &= \sum_{k=1}^{\alpha} c_k(\theta_{j_\alpha,n_\alpha})\, h_\alpha^k + E_{j_\alpha,n_\alpha,\alpha}.
\end{aligned} \tag{4}$$

Let $c, \tilde{c}$ be the vectors

$$c = [c_1, c_2, \ldots, c_\alpha]^T; \qquad \tilde{c} = [\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_\alpha]^T,$$

and let A be the coefficient matrix of size $\alpha \times \alpha$ with $(A)_{i,j} = h_i^j$. Hence, the set of (4) can be written in matrix form as

$$Ac = b_0 - b_\alpha\,, \tag{5}$$

where $b_0 = [E_{j_1,n_1,0}, E_{j_2,n_2,0}, \ldots, E_{j_\alpha,n_\alpha,0}]^T$ and $b_\alpha = [E_{j_1,n_1,\alpha}, E_{j_2,n_2,\alpha}, \ldots, E_{j_\alpha,n_\alpha,\alpha}]^T$. Furthermore, by neglecting the higher order errors, we may define an approximation $\tilde{c}$ of c according to the expression below

$$A\tilde{c} = b_0\,. \tag{6}$$

By solving the linear system of equations above, the approximation of c is easily obtained since the matrix size is very small. In a subsequent step, we derive upper-bounds for $|\tilde{c} - c|$: in reality, (5) and (6) leads to

$$A(\tilde{c} - c) = b_\alpha. \tag{7}$$

If we define $\Delta c = \tilde{c} - c$ and $\eta_i = \frac{E_{j_i,n_i,\alpha}}{h_i^{\alpha+1}}$ for $i = 1, \ldots, \alpha$, then the system (7) can be written as

$$A\Delta c = \begin{bmatrix} \eta_1 h_1^{\alpha+1} \\ \eta_2 h_2^{\alpha+1} \\ \vdots \\ \eta_\alpha h_\alpha^{\alpha+1} \end{bmatrix}, \tag{8}$$

with $|\eta_i| \leq C_\alpha$ for $i = 1, \ldots, \alpha$, where C_α is a constant. The coefficient matrix can be expressed as

$$A = \begin{bmatrix} h_1 & h_1^2 & \ldots & h_1^\alpha \\ h_2 & h_2^2 & \ldots & h_2^\alpha \\ \vdots & \vdots & & \vdots \\ h_\alpha & h_\alpha^2 & \ldots & h_\alpha^\alpha \end{bmatrix} = \begin{bmatrix} h_1 & & & \\ & h_2 & & \\ & & \ddots & \\ & & & h_\alpha \end{bmatrix} V(h_1, \ldots, h_\alpha),$$

where $V(h_1, \ldots, h_\alpha)$ is the Vandermonde matrix of order α corresponding to $h_1, \ldots, h_\alpha$.
By assuming $W = V^{-1}(h_1, \ldots, h_\alpha)$, we deduce

$$(W)_{i,j} = \begin{cases} (-1)^{\alpha-i} \left(\dfrac{\displaystyle\sum_{\substack{1 \leq k_1 < \ldots < k_{\alpha-i} \leq \alpha \\ k_1, \ldots, k_{\alpha-i} \neq j}} h_{k_1} \cdots h_{k_{\alpha-i}}}{\displaystyle\prod_{\substack{1 \leq k \leq \alpha \\ k \neq j}} (h_j - h_k)} \right) & 1 \leq i < \alpha, \\ \dfrac{1}{\displaystyle\prod_{\substack{1 \leq k \leq \alpha \\ k \neq j}} (h_j - h_k)} & i = \alpha. \end{cases} \tag{9}$$

Therefore for the inversion of the matrix A, we have

$$(A^{-1})_{i,j} = \begin{cases} (-1)^{\alpha-i} \left(\dfrac{\displaystyle\sum_{\substack{1 \leq k_1 < \ldots < k_{\alpha-i} \leq \alpha \\ k_1, \ldots, k_{\alpha-i} \neq j}} h_{k_1} \cdots h_{k_{\alpha-i}}}{h_j \displaystyle\prod_{\substack{1 \leq k \leq \alpha \\ k \neq j}} (h_j - h_k)} \right) & 1 \leq i < \alpha, \\ \dfrac{1}{h_j \displaystyle\prod_{\substack{1 \leq k \leq \alpha \\ k \neq j}} (h_j - h_k)} & i = \alpha, \end{cases} \tag{10}$$

and we can obtain an explicit expression for $(\Delta c)_i$, $i = 1, \ldots, \alpha$, that is

$$(\Delta c)_i = \sum_{j=1}^{\alpha} (A^{-1})_{i,j} \eta_j h_j^{\alpha+1}. \tag{11}$$

Case 1 If $i = \alpha$, then

$$(\Delta c)_\alpha = \sum_{j=1}^{\alpha} \frac{\eta_j h_j^{\alpha+1}}{h_j \prod\limits_{\substack{1 \le k \le \alpha \\ k \ne j}} \left(h_j - h_k\right)}.$$

Whence, from the fact that $|\eta_i| \le C_\alpha$ for $i = 1, \ldots, \alpha$,

$$|(\Delta c)_\alpha| \le \sum_{j=1}^{\alpha} \frac{|\eta_j| h_j^{\alpha+1}}{h_j \prod\limits_{\substack{1 \le k \le \alpha \\ k \ne j}} |h_j - h_k|} \le \sum_{j=1}^{\alpha} \frac{C_\alpha h_j^{\alpha}}{\prod\limits_{\substack{1 \le k \le \alpha \\ k \ne j}} |h_j - h_k|}.$$

With the choice $h_j = \frac{1}{m^{j-1}} h_1$ for $j = 1, \ldots, \alpha$, m positive integer, we have

$$\begin{aligned} |(\Delta c)_\alpha| &\le C_\alpha \sum_{j=1}^{\alpha} \frac{(\frac{h_1}{m^{j-1}})^\alpha}{\prod\limits_{\substack{1 \le k \le \alpha \\ k \ne j}} h_1 \left|\frac{1}{m^{j-1}} - \frac{1}{m^{k-1}}\right|} = C_\alpha h_1^\alpha \sum_{j=1}^{\alpha} \frac{(\frac{1}{m^{j-1}})^\alpha}{h_1^{\alpha-1} \prod\limits_{\substack{1 \le k \le \alpha \\ k \ne j}} \left|\frac{1}{m^{j-1}} - \frac{1}{m^{k-1}}\right|} \\ &= h_1 C_\alpha \sum_{j=1}^{\alpha} \frac{(\frac{1}{m^{j-1}})^\alpha}{\prod\limits_{\substack{1 \le k \le \alpha \\ k \ne j}} \left|\frac{1}{m^{j-1}} - \frac{1}{m^{k-1}}\right|} = O(h_1). \end{aligned}$$

Case 2 If $i = 1, \ldots, \alpha - 1$, then

$$(\Delta c)_i = \sum_{j=1}^{\alpha} (-1)^{\alpha-i} \eta_j h_j^{\alpha+1} \frac{\sum\limits_{\substack{1 \le k_1 < \ldots < k_{\alpha-i} \le \alpha \\ k_1, \ldots, k_{\alpha-i} \ne j}} h_{k_1} \cdots h_{k_{\alpha-i}}}{h_j \prod\limits_{\substack{1 \le k \le \alpha \\ k \ne j}} \left(h_j - h_k\right)},$$

that is different from the case $i = \alpha$ just for the numerator

$$\sum_{\substack{1 \le k_1 < \ldots < k_{\alpha-i} \le \alpha \\ k_1, \ldots, k_{\alpha-i} \ne j}} h_{k_1} \cdots h_{k_{\alpha-i}}.$$

As a consequence,

$$|(\Delta c)_i| \le C_\alpha \sum_{j=1}^{\alpha} h_j^\alpha \frac{\sum\limits_{\substack{1 \le k_1 < \ldots < k_{\alpha-i} \le \alpha \\ k_1, \ldots, k_{\alpha-i} \ne j}} h_{k_1} \cdots h_{k_{\alpha-i}}}{\prod\limits_{\substack{1 \le k \le \alpha \\ k \ne j}} \left|h_j - h_k\right|}.$$

With the choice $h_j = \frac{1}{m^{j-1}} h_1$ for $j = 1, \ldots, \alpha$, we infer

$$
\begin{aligned}
|(\Delta c)_i| &\leq C_\alpha \sum_{j=1}^{\alpha} \left(\frac{h_1}{m^{j-1}}\right)^{\alpha} \frac{\sum_{\substack{1 \leq k_1 < \ldots < k_{\alpha-i} \leq \alpha \\ k_1, \ldots, k_{\alpha-i} \neq j}} h_1^{\alpha-i} \left(\frac{1}{m^{k_1-1}} \frac{1}{m^{k_2-1}} \cdots \frac{1}{m^{k_{\alpha-i}-1}}\right)}{\prod_{\substack{1 \leq k \leq \alpha \\ k \neq j}} h_1 \left|\frac{1}{m^{j-1}} - \frac{1}{m^{k-1}}\right|} \\
&= C_\alpha \sum_{j=1}^{\alpha} \left(\frac{1}{m^{j-1}}\right)^{\alpha} \left(\frac{h_1^{\alpha} h_1^{\alpha-i}}{h_1^{\alpha-1}}\right) \frac{\sum_{\substack{1 \leq k_1 < \ldots < k_{\alpha-i} \leq \alpha \\ k_1, \ldots, k_{\alpha-i} \neq j}} \left(\frac{1}{m^{k_1-1}} \frac{1}{m^{k_2-1}} \cdots \frac{1}{m^{k_{\alpha-i}-1}}\right)}{\prod_{\substack{1 \leq k \leq \alpha \\ k \neq j}} \left|\frac{1}{m^{j-1}} - \frac{1}{m^{k-1}}\right|} \\
&= h_1^{\alpha-i+1} C_\alpha \sum_{j=1}^{\alpha} \left(\frac{1}{m^{j-1}}\right)^{\alpha} \frac{\sum_{\substack{1 \leq k_1 < \ldots < k_{\alpha-i} \leq \alpha \\ k_1, \ldots, k_{\alpha-i} \neq j}} \left(\frac{1}{m^{k_1-1}} \frac{1}{m^{k_2-1}} \cdots \frac{1}{m^{k_{\alpha-i}-1}}\right)}{\prod_{\substack{1 \leq k \leq \alpha \\ k \neq j}} \left|\frac{1}{m^{j-1}} - \frac{1}{m^{k-1}}\right|} = O(h_1^{\alpha-i+1}).
\end{aligned}
$$

As a conclusion, with the choice $h_j = \frac{1}{m^{j-1}} h_1$ for $j = 1, \ldots, \alpha$ and under the assumption that the asymptotic expansion reported in (1) is true, we deduce

$$
|(\Delta c)_i| = O(h_1^{\alpha-i+1}), \tag{12}
$$

for $i = 1, \ldots, \alpha$.

3 Error bounds for numerically approximated eigenvalues

The goal of this short section is to provide error bounds based on the linear system in (6) for the computation of the eigenvalues of $\mathcal{P}_n(f, g)$: of course, these error bounds are based on the conjecture that the relations reported in (1) are true. However, as we can see in Section 4, the numerical tests fully support the existence of the considered asymptotic expansion.

Indeed, as already observed, by solving (6), we can approximate c_k. Once we have the values of c_k, we can approximate the eigenvalues λ_{j_β} of a large dimension matrix of size n_β, here $n_\beta + 1 = m^{\beta-1}(n_1 + 1)$. The asymptotic expansion (3) can be written as

$$
E_{j_\beta, n_\beta, 0} = \bar{h}_\beta^T c + E_{j_\beta, n_\beta, \alpha}\,. \tag{13}
$$

By subtraction $\bar{h}_\beta^T \tilde{c}$ from both sides of the equation above, we find

$$
\begin{aligned}
E_{j_\beta, n_\beta, 0} - \bar{h}_\beta^T \tilde{c} &= \bar{h}_\beta^T (c - \tilde{c}) + E_{j_\beta, n_\beta, \alpha}, \\
\lambda_j(\mathcal{P}_{n_\beta}(f, g)) - r(\theta_{j, n_\beta}) - \bar{h}_\beta^T \tilde{c} &= \bar{h}_\beta^T \Delta c + E_{j_\beta, n_\beta, \alpha}, \\
\left|\lambda_j(\mathcal{P}_{n_\beta}(f, g)) - r(\theta_{j, n_\beta}) - \bar{h}_\beta^T \tilde{c}\right| &\leq \sum_{i=1}^{\alpha} h_\beta^i |(\Delta c)_i| + |E_{j_\beta, n_\beta, \alpha}|, \\
\left|\lambda_j(\mathcal{P}_{n_\beta}(f, g)) - r(\theta_{j, n_\beta}) - \bar{h}_\beta^T \tilde{c}\right| &\leq \sum_{i=1}^{\alpha} h_\beta^i |(\Delta c)_i| + C_\alpha h_\beta^{\alpha+1},
\end{aligned} \tag{14}
$$

where $\bar{h}_\beta = [h_\beta, h_\beta^2, \cdots, h_\beta^\alpha]^T$, $|E_{j_\beta,n_\beta,\alpha}| \leq C_\alpha h_\beta^{\alpha+1}$ for some constant C_α and $|(\Delta c)_i|$ is given in (12).

4 Numerical tests

In this section, we want to present a few numerical experiments to support the asymptotic expansion (1) in the case where one or more properties of the following list are satisfied:

1. $f''(0) \neq 0$ (see Example 1, Example 3, and Example 5),
2. $f''(0) = 0$ (see Example 2 and Example 4),
3. $\min g > 0$ (see Example 1, Example 2, and Example 5),
4. $\min g = 0$ (see Example 3 and Example 4),
5. $r = f/g$ is non monotone (see Example 5).

The approximation of eigenvalues of large matrices in each case is also computed. The expansion (1) for $\alpha = 4$ is

$$\begin{aligned}\lambda_j(\mathcal{P}_n(f,g)) &= r(\theta_{j,n}) + c_1(\theta_{j,n})\,h + c_2(\theta_{j,n})\,h^2 + c_3(\theta_{j,n})\,h^3 + c_4(\theta_{j,n})\,h^4 + E_{j,n,4},\\ E_{j,n,0} &= \lambda_j(\mathcal{P}_n(f,g)) - r(\theta_{j,n}) = c_1(\theta_{j,n})\,h + c_2(\theta_{j,n})\,h^2 + c_3(\theta_{j,n})\,h^3\\ &\quad + c_4(\theta_{j,n})\,h^4 + E_{j,n,4}\,. \end{aligned} \tag{15}$$

In all numerical examples, we choose four matrix-size values, that is n_i for $i \in \{1, 2, 3, 4\}$, in a way that they satisfy $n_i = m^{i-1}(n_1 + 1) - 1$, with m being a positive integer. The expansion (15) for the set of the four dimensions n_i can be written as

$$\begin{aligned} E_{j_1,n_1,0} &= c_1(\theta_{j_1,n_1})\,h_1 + c_2(\theta_{j_1,n_1})\,h_1^2 + c_3(\theta_{j_1,n_1})\,h_1^3 + c_4(\theta_{j_1,n_1})\,h_1^4 + E_{j_1,n_1,4},\\ E_{j_2,n_2,0} &= c_1(\theta_{j_2,n_2})\,h_2 + c_2(\theta_{j_2,n_2})\,h_2^2 + c_3(\theta_{j_2,n_2})\,h_2^3 + c_4(\theta_{j_2,n_2})\,h_2^4 + E_{j_2,n_2,4},\\ E_{j_3,n_3,0} &= c_1(\theta_{j_3,n_3})\,h_3 + c_2(\theta_{j_3,n_3})\,h_3^2 + c_3(\theta_{j_3,n_3})\,h_3^3 + c_4(\theta_{j_3,n_3})\,h_3^4 + E_{j_3,n_3,4},\\ E_{j_4,n_4,0} &= c_1(\theta_{j_4,n_4})\,h_4 + c_2(\theta_{j_4,n_4})\,h_4^2 + c_3(\theta_{j_4,n_4})\,h_4^3 + c_4(\theta_{j_4,n_4})\,h_4^4 + E_{j_4,n_4,4}, \end{aligned} \tag{16}$$

where $h_i = \frac{1}{n_i+1}$ and $j_i = m^{i-1}\,j_1$ for $i \in \{1, 2, 3, 4\}$. Notice that $\theta_{j_i,n_i} = \theta_{j_1,n_1} = \bar{\theta}$ for a fixed $j_1 \in \{1, 2, \cdots, n_1\}$. We are interested in the numerical approximation of $c_i(\bar{\theta})$ for $i \in \{1, 2, 3, 4\}$ and then in the precise numerical approximation of the eigenvalue of $\mathcal{P}_n(f, g)$ for large n. The set of (16) can be written as

$$\begin{aligned} E_{j_1,n_1,0} &= \tilde{c}_1(\bar{\theta})\,h_1 + \tilde{c}_2(\bar{\theta})\,h_1^2 + \tilde{c}_3(\bar{\theta})\,h_1^3 + \tilde{c}_4(\bar{\theta})\,h_1^4,\\ E_{j_2,n_2,0} &= \tilde{c}_1(\bar{\theta})\,h_2 + \tilde{c}_2(\bar{\theta})\,h_2^2 + \tilde{c}_3(\bar{\theta})\,h_2^3 + \tilde{c}_4(\bar{\theta})\,h_2^4,\\ E_{j_3,n_3,0} &= \tilde{c}_1(\bar{\theta})\,h_3 + \tilde{c}_2(\bar{\theta})\,h_3^2 + \tilde{c}_3(\bar{\theta})\,h_3^3 + \tilde{c}_4(\bar{\theta})\,h_3^4,\\ E_{j_4,n_4,0} &= \tilde{c}_1(\bar{\theta})\,h_4 + \tilde{c}_2(\bar{\theta})\,h_4^2 + \tilde{c}_3(\bar{\theta})\,h_4^3 + \tilde{c}_4(\bar{\theta})\,h_4^4. \end{aligned} \tag{17}$$

We solve the system of linear equations above for $j_1 \in \{1, 2, \cdots, n_1\}$ to compute $\tilde{c}_i(\bar{\theta})$. The computed $\tilde{c}_i$ are used to approximate the eigenvalues of large size n_β by exploiting the following relation

$$\tilde{\lambda}_{j_\beta}(\mathcal{P}_{n_\beta}(f,g)) = r(\theta_{j_\beta,n_\beta}) + \bar{h}_\beta^T \tilde{c}\,. \tag{18}$$

Example 1 Let g, f, and r be the functions defined as

$$\begin{aligned} f(\theta) &= 4 - 2\cos(\theta) - 2\cos(2\theta) = (2 - 2\cos(\theta))(3 + 2\cos(\theta)), \\ g(\theta) &= 3 + 2\cos(\theta), \\ r(\theta) &= \frac{f(\theta)}{g(\theta)} = 2 - 2\cos(\theta), \end{aligned}$$

where $\theta \in [0, \pi]$. The graphs of generating functions are shown in left panel of Fig. 1, and the approximations $\tilde{c}_k$, for $k = 1, 2, 3, 4$ are shown in the right panel. Note that $g(\theta) > 0, \forall\, \theta \in [0, \pi]$, $f''(0) \neq 0$, and furthermore r is monotone. We set $n = n_1 \in \{40, 60, 80, 100\}$ and $m = 2$.

Example 2 Let g, f, and r be the functions defined as

$$\begin{aligned} f(\theta) &= 20 - 30\cos(\theta) + 12\cos(2\theta) - 2\cos(3\theta) = (2 - 2\cos(\theta))^3, \\ g(\theta) &= 3 + 2\cos(\theta), \\ r(\theta) &= \frac{f(\theta)}{g(\theta)} = \frac{(2 - 2\cos(\theta))^3}{3 + 2\cos(\theta)}, \end{aligned}$$

where $\theta \in [0, \pi]$. The graphs of generating functions are shown in left panel of Fig. 2, and the approximations $\tilde{c}_k$, for $k = 1, 2, 3, 4$ are shown in the right panel. Remark that $g(\theta) > 0, \forall\, \theta \in [0, \pi]$, $f''(0) = 0$, and furthermore r is monotone. We set $n = n_1 \in \{40, 60, 80, 100\}$ and $m = 2$.

There is an important issue to discuss here. Both the functions f and r attain the minimum at $\theta = 0$ with a very high order. Indeed, we have $f(\theta), r(\theta) \approx \theta^6$, with $\phi_1 \approx \phi_2$ being the symmetric, transitive relation telling that there exist positive constants $c, C > 0$ such that $c\phi_1 \leq \phi_2 \leq C\phi_1$ on the whole definition domain $[0, \pi]$. Therefore for fixed j (independent of n) the jth smallest eigenvalue of $\mathcal{P}_n(f, g)$ is asymptotic to $k_j h^6$, k_j positive constant depending on j but not on n: the reader is refereed to [17] for the preconditioned case with the limitation $j = 1$ and to [1] and references therein for very elegant and precise estimates regarding the pure Toeplitz case.

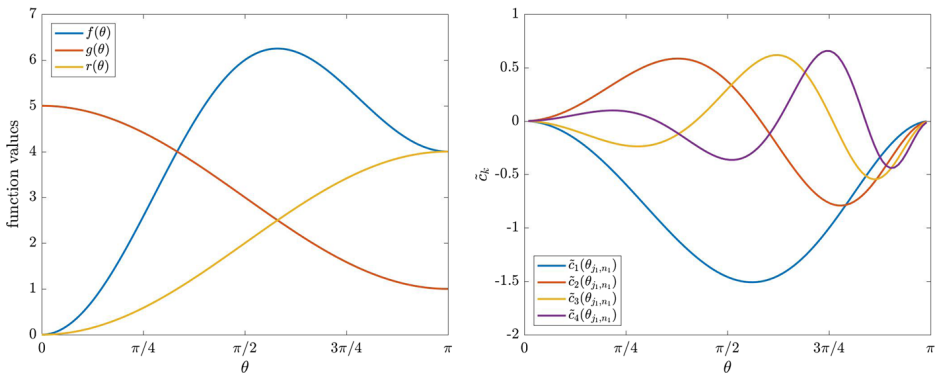


Fig. 1 Example 1: Generating functions (f, g, and r) and $\tilde{c}_k$ for $k = 1, 2, 3, 4$

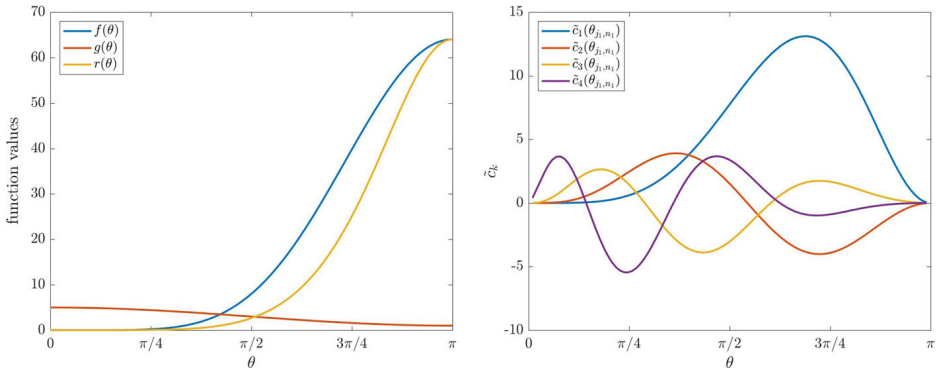


Fig. 2 Example 2: generating functions (f, g, and r) and $\tilde{c}_k$ for $k = 1, 2, 3, 4$

Now if we fix j and we put together $\lambda_j(\mathcal{P}_n(f, g)) \approx h^6$ with relations (3)–(4), then the only possibility for avoiding a contradiction is that the functions $c_1(\theta), c_2(\theta), c_3(\theta), c_4(\theta), c_5(\theta)$ all vanish at $\theta = 0$.

The approximations $\tilde{c}_k$, for $k = 1, 2, 3, 4$ shown in the right panel of Fig. 2 are coherent with the above mathematical conclusion and in fact all these approximations vanish simultaneously at $\theta = 0$ (the fifth is not displayed, but we computed it and it also equals to zero at $\theta = 0$, while, as expected from an extension of the results by [1] to the preconditioned Toeplitz case, the sixth is nonzero at $\theta = 0$).

Since the argument and the conclusions are the very same, we anticipate that the discussion can be repeated verbatim for Example 4, where the functions f and r attain the minimum at $\theta = 0$ with order 10. As a consequence, we expect that the functions $c_1(\theta), \ldots, c_9(\theta)$ all simultaneously vanish at $\theta = 0$, while $c_{10}(0) \neq 0$: this is confirmed for the first four of them as reported in the right panel of Fig. 4.

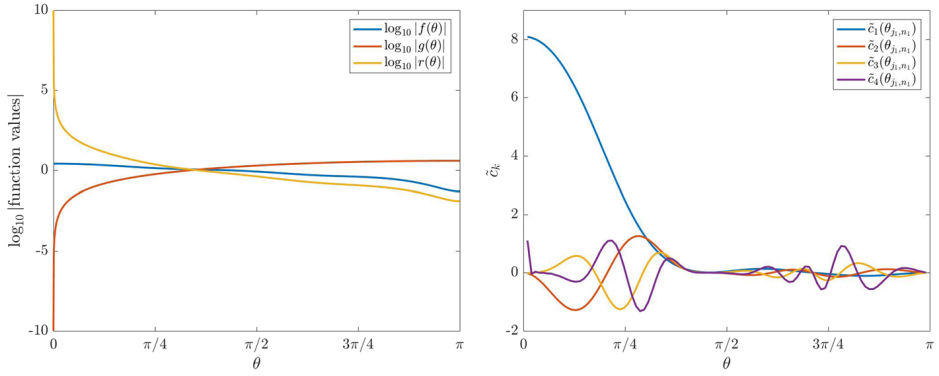


Fig. 3 Example 3: generating functions (f, g, and r) and $\tilde{c}_k$ for $k = 1, 2, 3, 4$

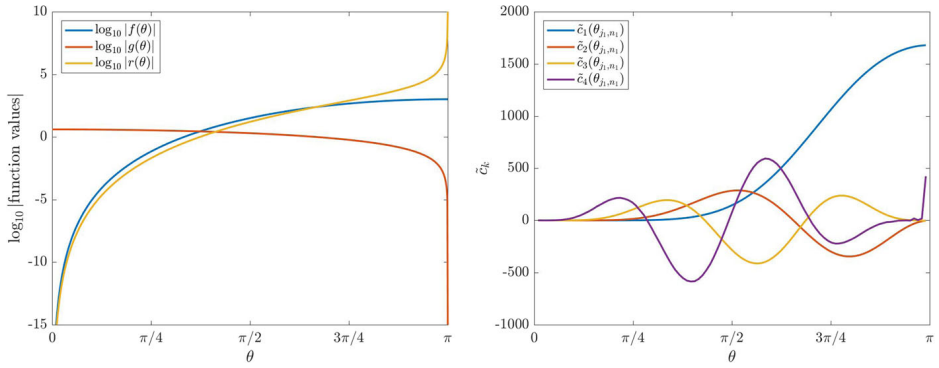


Fig. 4 Example 4: generating functions (f, g, and r) and $\tilde{c}_k$ for $k = 1, 2, 3, 4$

Example 3 Let g, f, and r be the functions defined as

$$
\begin{aligned}
f(\theta) &= 1 + \cos(\theta) + \frac{1}{4}\cos(2\theta) + \frac{1}{5}\cos(3\theta) + \frac{1}{10}\cos(4\theta) + \frac{1}{10}\cos(5\theta)\,,\\
g(\theta) &= 2 - 2\cos(\theta)\,,\\
r(\theta) &= \frac{f(\theta)}{g(\theta)} = \frac{1 + \cos(\theta) + \frac{1}{4}\cos(2\theta) + \frac{1}{5}\cos(3\theta) + \frac{1}{10}\cos(4\theta) + \frac{1}{10}\cos(5\theta)}{2 - 2\cos(\theta)}\,,
\end{aligned}
$$

where $\theta \in [0, \pi]$. The graphs of generating functions are shown in left panel of Fig. 3, and the approximations $\tilde{c}_k$, for $k = 1, 2, 3, 4$ are shown in the right panel. Note that $\min g(\theta) = 0$, $\forall\, \theta \in [0, \pi]$, $f''(0) \neq 0$, and furthermore r is monotone. We set $n = n_1 \in \{40, 60, 80, 100\}$ and $m = 2$.

Example 4 Let g, f, and r be the functions defined as

$$
\begin{aligned}
f(\theta) &= 252 - 420\cos(\theta) + 240\cos(2\theta) - 90\cos(3\theta) + 20\cos(4\theta) - 2\cos(5\theta) = (2 - 2\cos(\theta))^5\,,\\
g(\theta) &= 2 + 2\cos(\theta)\,,\\
r(\theta) &= \frac{f(\theta)}{g(\theta)} = \frac{(2 - 2\cos(\theta))^5}{2 + 2\cos(\theta)}\,,
\end{aligned}
$$

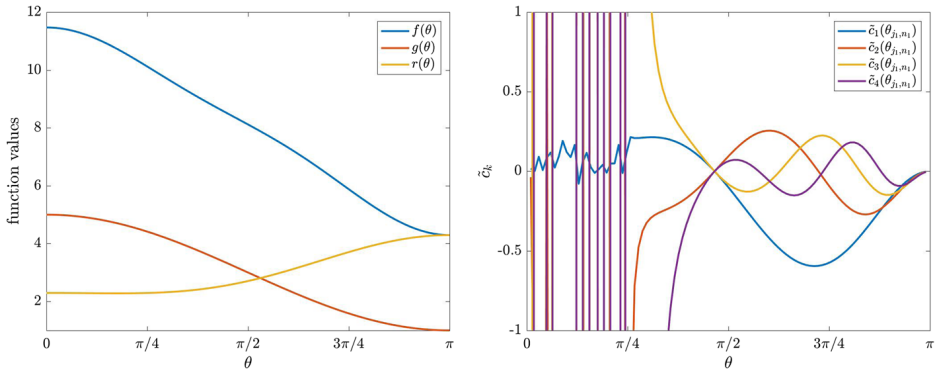


Fig. 5 Example 5: generating functions (f, g, and r) and $\tilde{c}_k$ for $k = 1, 2, 3, 4$

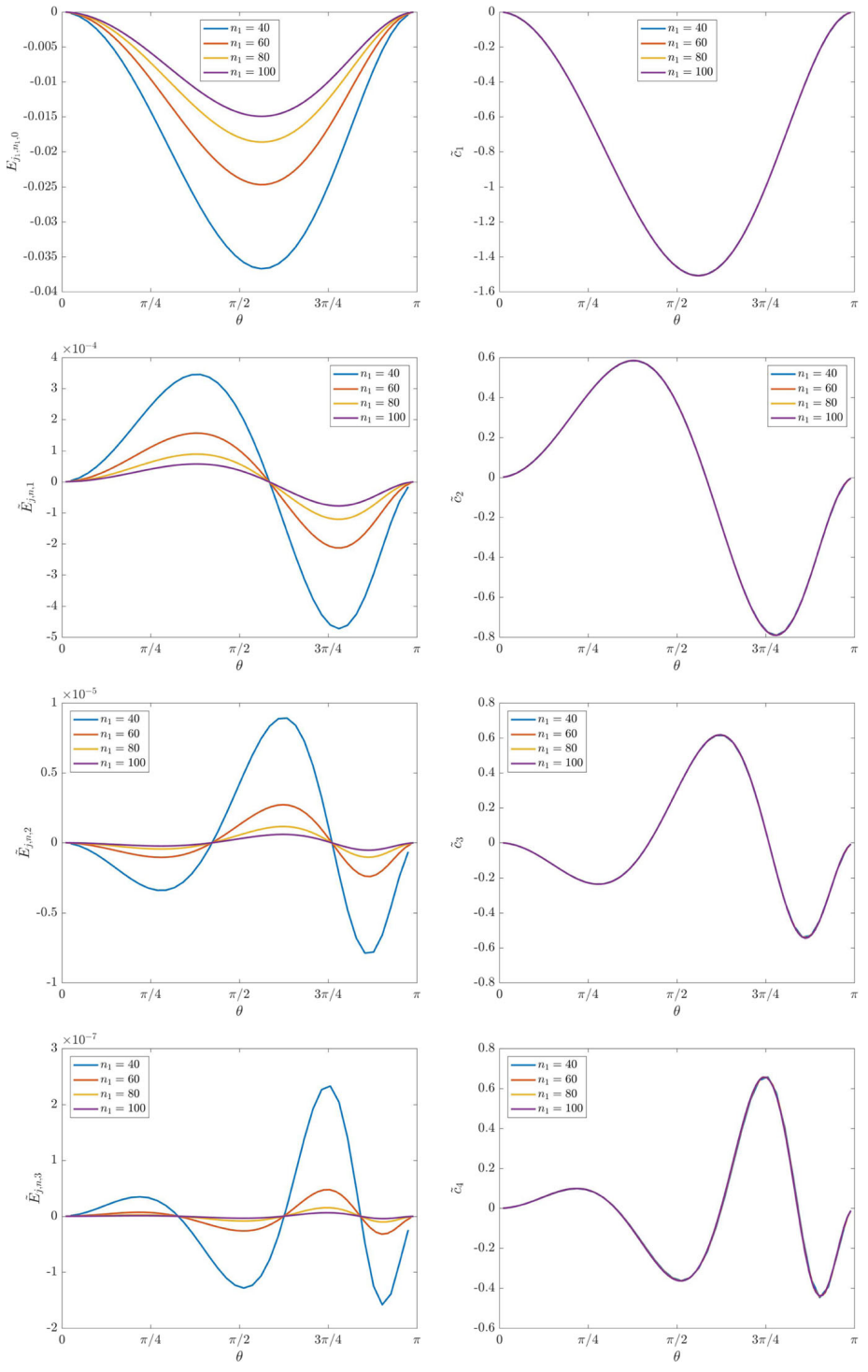


Fig. 6 Example 1: $E_{j,n,0}$, $\tilde{E}_{j,n,k}$ $(k = 1, 2, 3)$, and $\tilde{c}_k$ $(k = 1, 2, 3, 4)$, for $n = n_1 = \{40, 60, 80, 100\}$

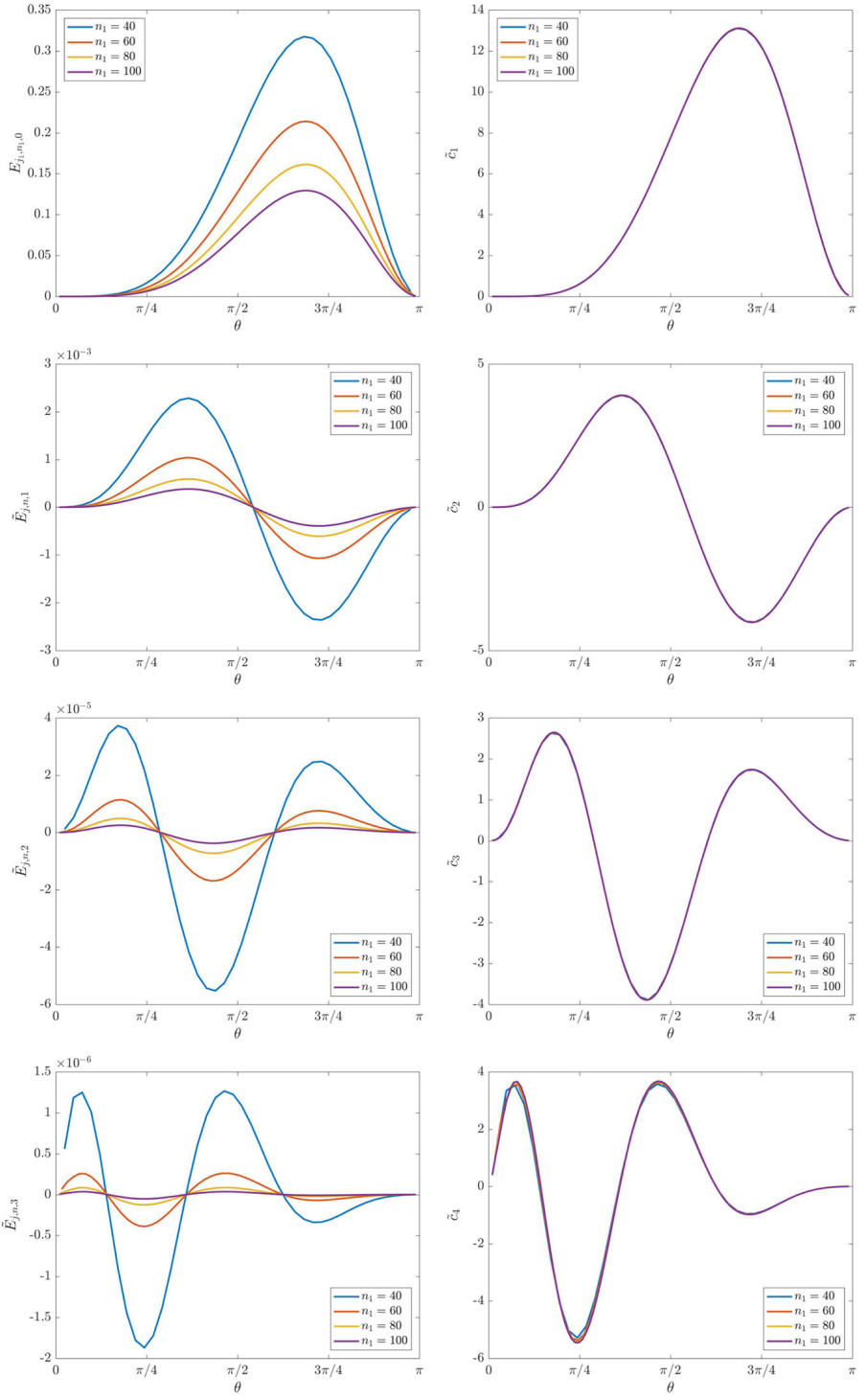


Fig. 7 Example 2: $E_{j,n,0}$, $\tilde{E}_{j,n,k}$ $(k = 1, 2, 3)$, and $\tilde{c}_k$ $(k = 1, 2, 3, 4)$, for $n = n_1 = \{40, 60, 80, 100\}$

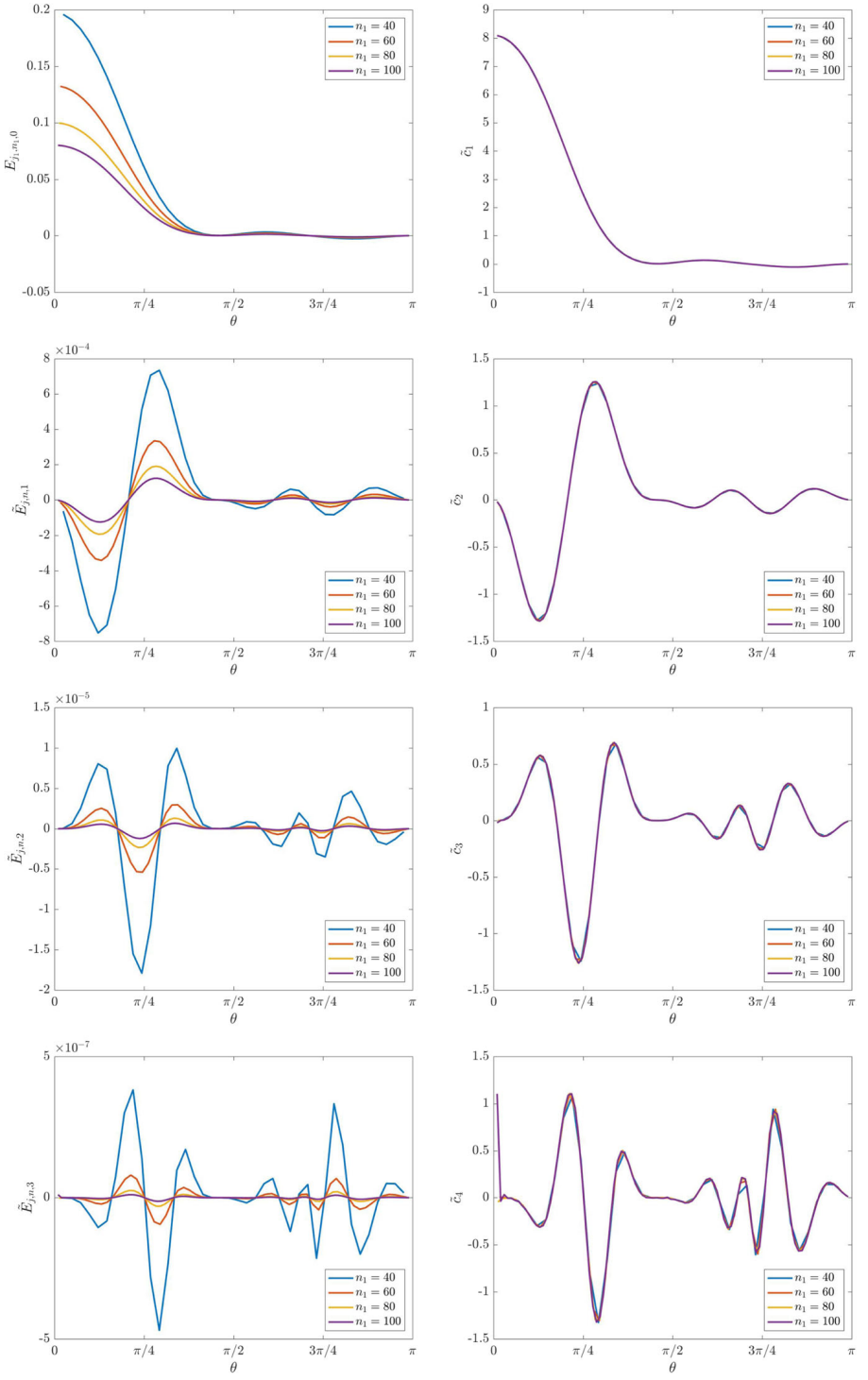


Fig. 8 Example 3: $E_{j,n,0}$, $\tilde{E}_{j,n,k}$ $(k = 1, 2, 3)$, and $\tilde{c}_k$ $(k = 1, 2, 3, 4)$, for $n = n_1 = \{40, 60, 80, 100\}$

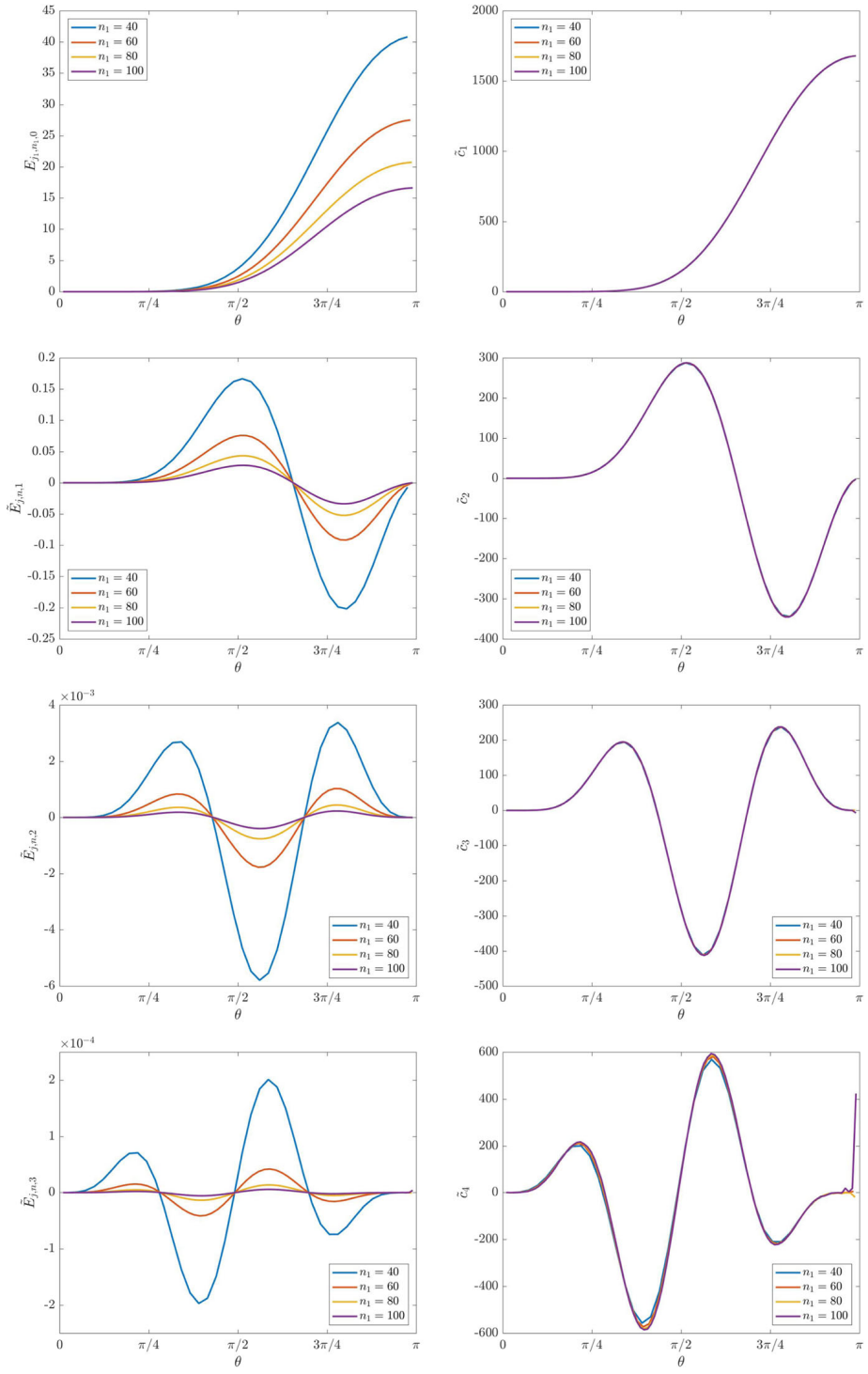


Fig. 9 Example 4: $E_{j,n,0}$, $\tilde{E}_{j,n,k}$ $(k = 1, 2, 3)$, and $\tilde{c}_k$ $(k = 1, 2, 3, 4)$, for $n = n_1 = \{40, 60, 80, 100\}$

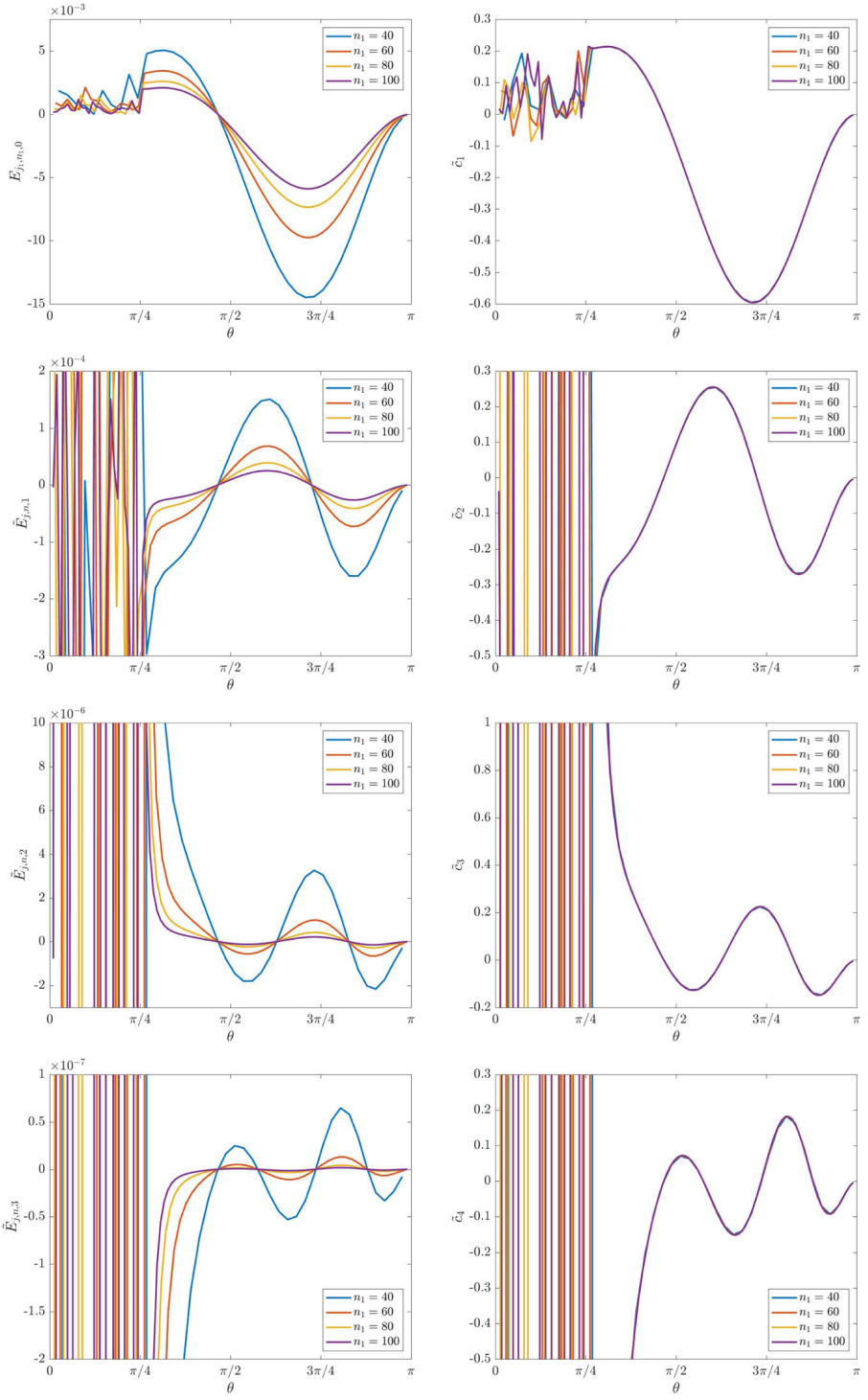


Fig. 10 Example 5: $E_{j,n,0}$, $\tilde{E}_{j,n,k}$ $(k = 1, 2, 3)$, and $\tilde{c}_k$ $(k = 1, 2, 3, 4)$, for $n = n_1 = \{40, 60, 80, 100\}$

where $\theta \in [0, \pi]$. The graphs of generating functions are shown in left panel of Fig. 4, and the approximations $\tilde{c}_k$, for $k = 1, 2, 3, 4$ are shown in the right panel. Remark that $\min g(\theta) = 0, \forall\, \theta \in [0, \pi]$, $f''(0) = 0$, and furthermore r is monotone. We set $n = n_1 \in \{40, 60, 80, 100\}$ and $m = 2$.

Example 5 Let g, f, and r be the functions defined as

$$
\begin{aligned}
f(\theta) &= \frac{136}{17} + \frac{56}{17}\cos(\theta) - \frac{2}{17}\cos(2\theta) + \frac{5}{17}\cos(3\theta) = (3 - \cos(\theta) + \frac{5}{17}\cos(2\theta))(3 + 2\cos(\theta))\,,\\
g(\theta) &= 3 + 2\cos(\theta)\,,\\
r(\theta) &= \frac{f(\theta)}{g(\theta)} = 3 - \cos(\theta) + \frac{5}{17}\cos(2\theta)\,,
\end{aligned}
$$

where $\theta \in [0, \pi]$. The graphs of generating functions are shown in left panel of Fig. 5, and the approximations $\tilde{c}_k$, for $k = 1, 2, 3, 4$ are shown in the right panel. Notice that $\min g(\theta) > 0, \forall\, \theta \in [0, \pi]$, $f''(0) \neq 0$, and furthermore r is non monotone. We set $n = n_1 \in \{40, 60, 80, 100\}$ and $m = 2$.

The numerical tests related to Examples 1 and 2, as in Figs. 6 and 7, show that the error expansion (1) behaves as expected. In Fig. 11, we also see that the approximated $\tilde{c}_k$ can be used for a large n to approximate the error term to (or almost to) machine precision.

In the numerical tests associated with Examples 3 and 4, as in Figs. 8 and 9, we observe again that the error expansion is in accordance with (1). We also note a slight deviation for the largest eigenvalue and this has to be expected since we have $r(\theta_{1,n}) \to \infty$ as $n \to \infty$ for Example 3 (on the other hand for Example 4 we notice $r(\theta_{n,n}) \to \infty$ as $n \to \infty$). However, the approximation of the eigenvalues of $\mathcal{P}_n(f, g)$ is excellent and almost to machine precision as reported in Fig. 12.

In the numerical test related to Example 5, we have a non monotone region for $\theta \in [0, 2\tan^{-1}(\sqrt{3/17})]$ where the proposed expansion does not work. Indeed, additional errors are introduced when compared to $E_{j,n,0}$, since the sampling of $r(\theta_{j_1,n_1})$ leads to a poorer approximation after ordering than the procedure given by sampling $r(\theta_{j,n_7})$ first and then picking samples after ordering. However, the expansion is

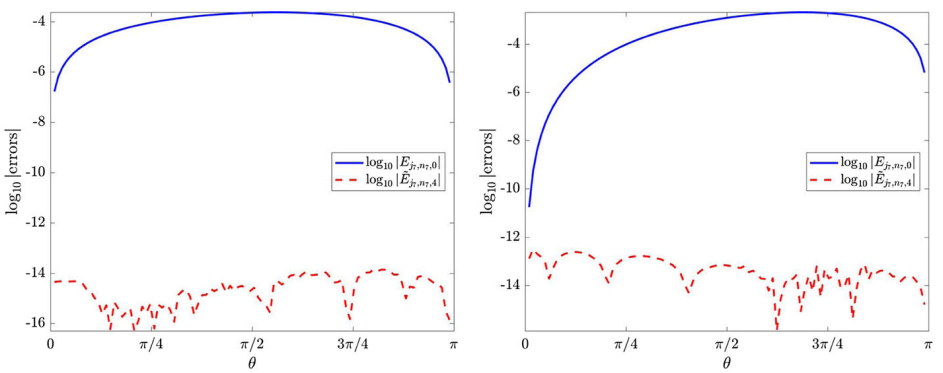


Fig. 11 Example 1 and 2: the errors $\log_{10}|E_{j_7,n_7,0}|$ and $\log_{10}|\tilde{E}_{j_7,n_7,4}|$ for the 100 indices j_7 of $n_7 = 6463$ in (18), corresponding to $n_1 = 100$, and using $\tilde{c}_k$, $k = 1, 2, 3, 4$, computed with $m = 2$

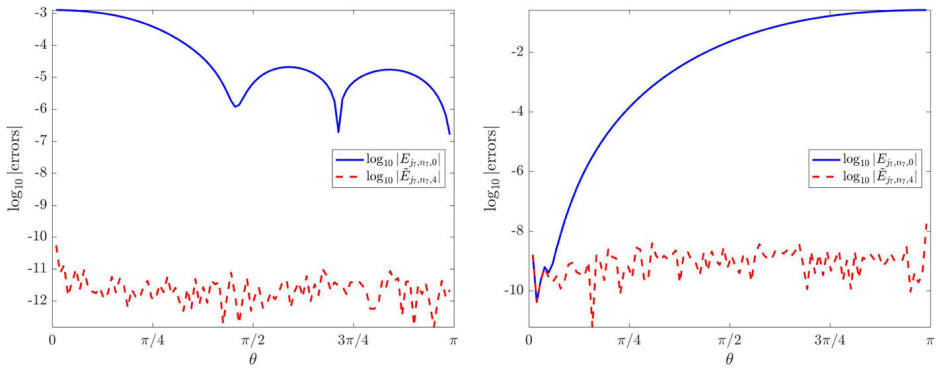


Fig. 12 Example 3 and 4: the errors $\log_{10}|E_{j_7,n_7,0}|$ and $\log_{10}|\tilde{E}_{j_7,n_7,4}|$ for the 100 indices j_7 of $n_7 = 6463$ in (18), corresponding to $n_1 = 100$, and using $\tilde{c}_k$, $k = 1, 2, 3, 4$, computed with $m = 2$

confirmed for the rest of the domain, as seen in Fig. 10. Furthermore, in Fig. 13, the expansion works well again for the monotone part, by allowing an approximation almost to machine precision of the eigenvalues of $\mathcal{P}_n(f, g)$.

However, even if the eigenvalues lying in the non monotone region give raise to an irregular error pattern, it seems that there exists a kind of 'deformed' periodicity in the error, like it is formally proven, without deformations, for the eigenvalues of $T_n(f)$, $f(\theta) = 2 - 2\cos(\omega\theta)$, $\omega \geq 2$ integer, and $g(\theta) = 1$ (see [12]). The latter observation indicates that a more complete study of this 'deformed' periodicity has to be considered in the future.

We finally observe that remarkable numerical results for the eigenvalues of $\mathcal{P}_n(f, g)$, as reported in Figs. 11, 12 and 13, really answer in the positive to the question posed in the title of the paper. In fact, we obtain almost machine precision for

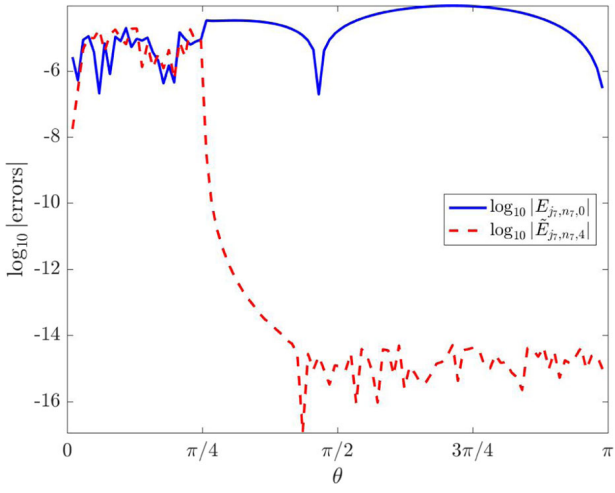


Fig. 13 Example 5: the errors $\log_{10}|E_{j_7,n_7,0}|$ and $\log_{10}|\tilde{E}_{j_7,n_7,4}|$ for the 100 indices j_7 of $n_7 = 6463$ in (18), corresponding to $n_1 = 100$, and using $\tilde{c}_k$, $k = 1, 2, 3, 4$, computed with $m = 2$. Note the non monotone part, $\theta \in [0, 2\tan^{-1}(\sqrt{3/17})]$, where the error is not improved

the computation of the spectrum of $\mathcal{P}_n(f, g)$, for large n and only working with few really small matrices.

5 Conclusions

Bogoya et al. [4–6] have recently obtained the precise asymptotic expansion for the eigenvalues of a sequence of Toeplitz matrices $\{T_n(f)\}$, under suitable assumptions on the associated generating function f. In this paper, we have shown numerical evidence that some of these assumptions can be relaxed and extended to the case of a sequence of preconditioned Toeplitz matrices $\{\mathcal{P}_n(f, g) = T_n^{-1}(g)T_n(f)\}$, for f trigonometric polynomial, g nonnegative, not identically zero trigonometric polynomial, $r = f/g$, and where the ratio r plays the same role as f in the non-preconditioned case. The first-order asymptotic term of the expansion has been also proven using purely linear algebra tools.

Moreover, based on the eigenvalue asymptotics, we devised an extrapolation algorithm for computing the eigenvalues of preconditioned banded symmetric Toeplitz matrices with a high level of accuracy, with a relatively low computational cost, and with potential application to the computation of the spectrum of differential operators. In fact, up to low rank corrections, matrices of the form $\mathcal{P}_n(f, g)$ appear in the context of the spectral approximation of differential operators in which a low rank correction of $T_n(g)$ is the mass matrix and a low rank correction of $T_n(f)$ is the stiffness matrix. We carried out also preliminary numerical tests confirming that the same kind of asymptotic expansion holds, at least in the context of the Isogeometric approximation of second-order differential operators.

Therefore, a plan for the future has to include:

- the theoretical proof of the asymptotic expansion in (1) for $\alpha > 1$;
- the analysis of the non monotone case and its relations with the study in [12] for the special case where $f(\theta) = 2 - 2\cos(\omega\theta)$, $\omega \geq 2$ integer, and $g(\theta) = 1$;
- the extension of the results by [1] to the preconditioned Toeplitz case and the study of its connection with the general expansion in (1);
- the extension of the numerical and theoretical study to a multidimensional, block setting, with special attention to the matrices coming from the approximation of elliptic differential operators.

Acknowledgements The research of Eman Salem Al-Aidarous was funded by King Abdulaziz University during scientific communication year 2017–2018. The research of Sven-Erik Ekström is cofinanced by the Graduate School in Mathematics and Computing (FMB) and Uppsala University. The research of the Isabella Furci and Stefano Serra-Capizzano is cofinanced by INdAM-GNCS (Istituto Nazionale di Alta Matematica - Gruppo Nazionale di Calcolo Scientifico).

Finally, a special thanks to the referee for pertinent comments, which helped us to improve the quality of the paper.

Appendix

Theorem 1 *Let f, g be real-valued cosine trigonometric polynomials (RCTP) on $[0,\pi]$ with $M_g = \max g > 0$ and $m_g = \min g \geq 0$. If $r = \frac{f}{g}$ is monotone on $[0,\pi]$ then $\exists\, C > 0$ such that*

$$\left|\lambda_j(\mathcal{P}_n(f,g)) - r\left(\frac{j\pi}{n+1}\right)\right| \leq Ch \quad \forall\, j, \forall\, n, \tag{19}$$

where

- *$\mathcal{P}_n(f,g)$ is the "preconditioned" matrix $\mathcal{P}_n(f,g) = T_n^{-1}(g)T_n(f)$,*
- *$\lambda_1(\mathcal{P}_n(f,g)), \lambda_2(\mathcal{P}_n(f,g)), \ldots, \lambda_n(\mathcal{P}_n(f,g))$ are the eigenvalues of $\mathcal{P}_n(f,g)$, arranged in nondecreasing or nonincreasing order, depending on whether r is increasing or decreasing,*
- *$h = \frac{1}{n+1}$ and $\theta_{j,n} = \frac{j\pi}{n+1} = j\pi h$.*

Proof For the sake of simplicity, we assume that r is nondecreasing (the other case has a similar proof).

Notice that the conditions on f and g imply that $T_n(g)$ is positive definite and, by setting $\sim$ the symbol representing similarity between matrices, we find $\mathcal{P}_n(f,g) \sim T_n^{-1/2}(g)T_n(f)T_n^{-1/2}(g)$ so we can order the eigenvalues of $\mathcal{P}_n(f,g)$ as follows

$$\lambda_1(\mathcal{P}_n(f,g)) \leq \lambda_2(\mathcal{P}_n(f,g)) \leq \cdots \leq \lambda_n(\mathcal{P}_n(f,g)).$$

We remark that

$$\begin{aligned} T_n(f) &= \tau_n(f) + H_n(f), \\ T_n(g) &= \tau_n(g) + H_n(g), \end{aligned} \tag{20}$$

where, for ψ RCTP of degree m and $Q = \left(\sqrt{\frac{2}{n+1}}\sin\left(\frac{ij\pi}{n+1}\right)\right)_{i,j=1}^{n}$, $\tau_n(\psi)$ is the following τ matrix [3] of size n generated by ψ

$$\tau_n(\psi) = Q \operatorname*{diag}_{1\leq j\leq n}\left(\psi\left(\frac{j\pi}{n+1}\right)\right)Q, \quad Q = Q^T = Q^{-1},$$

and $H_n(\psi)$ is the Hankel matrix

$$H_n(\phi) = \begin{bmatrix} \hat{\psi}_2 & \hat{\psi}_3 & \cdots & \hat{\psi}_m & & & \\ \hat{\psi}_3 & & \iddots & & & & \\ \vdots & \iddots & & & & & \\ \hat{\psi}_m & & & & & & \\ & & & & & \hat{\psi}_m & \\ & & & & \iddots & \vdots & \\ & & & \iddots & & \hat{\psi}_3 & \\ & & \hat{\psi}_m & \cdots & \hat{\psi}_3 & \hat{\psi}_2 & \end{bmatrix}.$$

with $\operatorname{rank}(H_n(\psi)) \le 2(m-1)$.
Hence,

$$\begin{aligned} R_f &:= \operatorname{rank}(H_n(f)) \le 2(\deg(f)-1), \\ R_g &:= \operatorname{rank}(H_n(g)) \le 2(\deg(g)-1), \\ R_{f,g} &:= \max\{R_f, R_g\} \le 2\left(\max\{\deg(f), \deg(g)\}-1\right). \end{aligned} \tag{21}$$

Let P_n^τ be the matrix $\tau_n^{-1}(g)\tau_n(f)$,

$$\begin{aligned} P_n^\tau &= Q\left(\operatorname*{diag}_{1\le j\le n}\left(g\left(\tfrac{j\pi}{n+1}\right)\right)\right)^{-1} Q\, Q \operatorname*{diag}_{1\le j\le n}\left(f\left(\tfrac{j\pi}{n+1}\right)\right) Q \\ &= Q \operatorname*{diag}_{1\le j\le n}\left(\tfrac{f}{g}\left(\tfrac{j\pi}{n+1}\right)\right) Q \\ &= Q \operatorname*{diag}_{1\le j\le n}\left(r\left(\tfrac{j\pi}{n+1}\right)\right) Q. \end{aligned}$$

Hence, for $j = 1, \ldots, n$

$$\lambda_j(P_n^\tau) = r\left(\frac{j\pi}{n+1}\right). \tag{22}$$

By observing that $T_n^{-1}(g)T_n(f)$ is similar to $T_n^{-1/2}(g)T_n(f)T_n^{-1/2}(g)$, using the MinMax spectral characterization for Hermitian matrices [2], fixed $j \in \{R_{f,g}+1, \ldots, n-R_{f,g}\}$ and $T \subset \mathbb{C}^n$, $\dim(T) = n+1-j$, we obtain

$$\begin{aligned} \lambda_j(\mathcal{P}_n(f,g)) &= \lambda_j\left(T_n^{-1}(g)T_n(f)\right) \\ &= \lambda_j\left(T_n^{-1/2}(g)T_n(f)T_n^{-1/2}(g)\right) \\ &= \max_{\dim(T)=n+1-j}\left(\min_{\substack{x\in T,\\ x\neq \underline{0}}}\left(\frac{x^* T_n^{-1/2}(g)T_n(f)T_n^{-1/2}(g)x}{x^*x}\right)\right) \\ &= \max_{\dim(T)=n+1-j}\left(\min_{\substack{x\in T,\ x\neq 0\\ y=T_n^{-1/2}(g)x}}\left(\frac{y^*T_n(f)y}{y^*T_n(g)y}\right)\right) \\ &= \max_{\dim(\hat{T})=n+1-j}\left(\min_{\substack{y\in \hat{T},\\ y\neq \underline{0}}}\left(\frac{y^*T_n(f)y}{y^*T_n(g)y}\right)\right), \end{aligned} \tag{23}$$

because $T_n^{-1/2}(g)$ is a full rank matrix and, if $\dim(T) = n+1-j$, then $\hat{T} := \{y : y = T_n^{-1/2}(g)x,\ x \neq \underline{0}, x \in T\}$ is a new vector space having the same dimension $n+1-j$ as T.

Let F be the subspace of $\mathbb{C}^n$ generated by the union of the columns of matrices $H_n(f)$ and $H_n(g)$. Because of the particular structure of the columns of Hankel matrices $H_n(f)$ and $H_n(g)$, we deduce

$$\dim(F) = \max\{\operatorname{rank}(H_n(g)), \operatorname{rank}(H_n(f))\} = R_{f,g},$$

so that

$$\dim(F^\perp) = n - R_{f,g}.$$

Let us define $W_{f,g} = \hat{T} \cap F^\perp$,

$$\begin{aligned} n+1-j \ge \dim(W_{f,g}) \ge \max\{0, \dim(\hat{T}) + \dim(F^\perp) - n\} &= n+1-j \\ &+n - R_{f,g} - n = n+1-(j+R_{f,g}), \end{aligned}$$

because $n+1-(j+R_{f,g}) \geq 1$ for $j \leq n-R_{f,g}$. The latter implies in particular that $W_{f,g} \neq \emptyset$. Thus, due to the orthogonality, $\forall\, y \neq \underline{0} \in W_{f,g}$, we find

$$H_n(f)y = \underline{0}, \quad H_n(g)y = \underline{0},$$

so that

$$y^* H_n(f)y = 0, \quad y^* H_n(g)y = 0.$$

Hence, from (23)

$$\begin{aligned}
\lambda_j(\mathcal{P}_n(f,g)) &= \max_{\dim(\hat{T})=n+1-j} \left(\min_{\substack{y\in\hat{T},\\ y\neq\underline{0}}} \left(\frac{y^*(\tau_n(f)+H_n(f))y}{y^*(\tau_n(g)+H_n(g))y} \right) \right) \\
&\leq \max_{\dim(\hat{T})=n+1-j} \left(\min_{\substack{y\in W_{f,g}\\ y\neq\underline{0}}} \left(\frac{y^*(\tau_n(f)+H_n(f))y}{y^*(\tau_n(g)+H_n(g))y} \right) \right) \\
&= \max_{\dim(\hat{T})=n+1-j} \left(\min_{\substack{y\in W_{f,g}\\ y\neq\underline{0}}} \left(\frac{y^*\tau_n(f)y}{y^*\tau_n(g)y} \right) \right) \\
&= \max_{\substack{W_{f,g}=\hat{T}\cap F^\perp\\ \dim(\hat{T})=n+1-j}} \left(\min_{\substack{y\in W_{f,g},\\ y\neq\underline{0}}} \left(\frac{y^*\tau_n(f)y}{y^*\tau_n(g)y} \right) \right) \\
&\leq \max_{n+1-j\geq\dim(\hat{W}_{f,g})\geq n+1-(j+R_{f,g})} \left(\min_{\substack{y\in\hat{W}_{f,g},\\ y\neq\underline{0}}} \left(\frac{y^*\tau_n(f)y}{y^*\tau_n(g)y} \right) \right) \\
&= \max_{n+1-j\geq\dim(\hat{W})\geq n+1-(j+R_{f,g})} \left(\min_{\substack{y\in\hat{W}_{f,g},\\ y\neq\underline{0}\\ x=\tau_n^{1/2}(g)y}} \left(\frac{x^*\tau_n^{-1/2}(g)\tau_n(f)\tau_n^{-1/2}(g)x}{x^*x} \right) \right) \\
&= \max\{\lambda_j(P_n^\tau), \lambda_{j+1}(P_n^\tau), \ldots, \lambda_{j+R_{f,g}}(P_n^\tau)\} \\
&= \lambda_{j+R_{f,g}}(P_n^\tau).
\end{aligned} \tag{24}$$

By fixing $j \in \{R_{f,g}+1, \ldots, n-R_{f,g}\}$ and $T \subset \mathbb{C}^n$, $\dim(T) = j$, analogously we obtain

$$\begin{aligned}
\lambda_j(\mathcal{P}_n(f,g)) &= \min_{\dim(T)=j} \left(\max_{\substack{x\in T,\\ x\neq\underline{0}}} \left(\frac{x^* T_n^{-1/2}(g)T_n(f)T_n^{-1/2}(g)x}{x^*x} \right) \right) \\
&= \min_{\dim(T)=j} \left(\max_{\substack{x\in T,\\ x\neq\underline{0}\\ y=T_n^{-1/2}(g)x}} \left(\frac{y^*T_n(f)y}{y^*T_n(g)y} \right) \right) \\
&= \min_{\dim(\hat{T})=j} \left(\max_{\substack{y\in\hat{T},\\ y\neq\underline{0}}} \left(\frac{y^*T_n(f)y}{y^*T_n(g)y} \right) \right) \\
&= \min_{\dim(\hat{T})=j} \left(\max_{\substack{y\in\hat{T},\\ y\neq\underline{0}}} \left(\frac{y^*(\tau_n(f)+H_n(f))y}{y^*(\tau_n(g)+H_n(g))y} \right) \right).
\end{aligned} \tag{25}$$

Let us define $W_{f,g} = \hat{T} \cap F^\perp$,

$j \geq \dim(W_{f,g}) \geq \max\{0, \dim(\hat{T}) + \dim(F^\perp) - n\} = j + n - R_{f,g} - n = j - R_{f,g}$,

because $j - R_{f,g} \geq 1$ for $j \geq R_{f,g}+1$. The latter implies in particular that $W_{f,g} \neq \emptyset$, and hence, due to the orthogonality, $\forall\, y \neq \underline{0} \in W_{f,g}$, we have

$$H_n(f)y = \underline{0}, \quad H_n(g)y = \underline{0},$$

and therefore

$$y^* H_n(f) y = 0, \quad y^* H_n(g) y = 0.$$

Thus, from (25)

$$\begin{aligned}
\lambda_j(\mathcal{P}_n(f,g)) &\geq \min_{\dim(\hat{T})=j} \left(\max_{\substack{y \in W_{f,g}, \\ y \neq \underline{0}}} \left(\frac{y^*(\tau_n(f)+H_n(f))y}{y^*(\tau_n(g)+H_n(g))y} \right) \right) \\
&= \min_{\dim(\hat{T})=j} \left(\max_{\substack{y \in W_{f,g}, \\ y \neq \underline{0}}} \left(\frac{y^*\tau_n(f)y}{y^*\tau_n(g)y} \right) \right) \\
&= \min_{\substack{W_{f,g}=\hat{T}\cap F^\perp \\ \dim(\hat{T})=j}} \left(\max_{\substack{y \in W_{f,g}, \\ x \neq \underline{0}}} \left(\frac{y^*\tau_n(f)y}{y^*\tau_n(g)y} \right) \right) \qquad (26) \\
&\geq \min_{j \geq \dim(\hat{W}_{f,g}) \geq j - R_{f,g}} \left(\max_{\substack{y \in W_{f,g}, \\ y \neq \underline{0}}} \left(\frac{y^*\tau_n(f)y}{y^*\tau_n(g)y} \right) \right) \\
&= \min\{\lambda_j(P_n^\tau), \lambda_{j-1}(P_n^\tau), \ldots, \lambda_{j-R_{f,g}}(P_n^\tau)\} \\
&= \lambda_{j-R_{f,g}}(P_n^\tau).
\end{aligned}$$

By exploiting the previous inequality, relations (22) and (24), we obtain for $j = R_{f,g}+1, \ldots, n - R_{f,g}$

$$r\left(\frac{(j-s)\pi}{n+1}\right) = \lambda_{j-s}(P_n^\tau) \leq \lambda_j(\mathcal{P}_n(f,g)) \leq \lambda_{j+s}(P_n^\tau) = r\left(\frac{(j+s)\pi}{n+1}\right), \quad (27)$$

where $s = R_{f,g}$.
The function r is a RCTP on $[0,\pi]$ and a monotone increasing function so we have, $\forall n$ and $\forall\, j = s+1, \ldots, n-s$,

$$\lambda_j(\mathcal{P}_n(f,g)) - r\left(\frac{j\pi}{n+1}\right) \leq r\left(\frac{(j+s)\pi}{n+1}\right) - r\left(\frac{j\pi}{n+1}\right) = r'(\bar{\theta})\frac{s\pi}{n+1} \leq ||r'||_\infty s\pi h, \quad (28)$$

with $\bar{\theta} \in \left(\frac{j\pi}{n+1}, \frac{(j+s)\pi}{n+1}\right)$ and

$$\lambda_j(\mathcal{P}_n(f,g)) - r\left(\frac{j\pi}{n+1}\right) \geq r\left(\frac{(j-s)\pi}{n+1}\right) - r\left(\frac{j\pi}{n+1}\right) \geq -||r'||_\infty s\pi h. \quad (29)$$

By setting $C = ||r'||_\infty s\pi$, for $s+1 \leq j \leq n-s$, we obtain

$$\left|\lambda_j(\mathcal{P}_n(f,g)) - r\left(\frac{j\pi}{n+1}\right)\right| \leq Ch. \quad (30)$$

Furthermore, from [11] $\forall\, j = 1, \ldots, n$, we know that

$$m_r \leq \lambda_j(\mathcal{P}_n(f,g)) \leq M_r,$$

where

$$m_r = \min_{\theta \in [0,\pi]} r(\theta); \quad m_r = \max_{\theta \in [0,\pi]} r(\theta),$$

with strict inequalities that is $m_r < \lambda_j(\mathcal{P}_n(f,g)) < M_r$ if $m_r < M_r$, while the case $m_r = M_r$ is in fact trivial. Hence, for $n - s < j \leq n$

$$\left|r\left(\frac{j\pi}{n+1}\right) - \lambda_j(\mathcal{P}_n(f,g))\right| \leq \left|r\left(\frac{j\pi}{n+1}\right) - r\left(\frac{n\pi}{n+1}\right)\right| \leq |r'(\bar{\theta})|\left|\frac{(n-j)\pi}{n+1}\right|,$$

where $\bar{\theta} \in (\frac{j\pi}{n+1}, \frac{n\pi}{n+1})$. If $n - s < j \leq n$ then $|n - j| < s$, so that

$$\left| r\left(\frac{j\pi}{n+1}\right) - \lambda_j(\mathcal{P}_n(f,g)) \right| \leq ||r'||_\infty s\pi h = Ch.$$

For $1 \leq j < s + 1$

$$\left| r\left(\frac{j\pi}{n+1}\right) - \lambda_j(\mathcal{P}_n(f,g)) \right| \leq \left| r\left(\frac{j\pi}{n+1}\right) - r\left(\frac{\pi}{n+1}\right) \right| \leq \left| r'\left(\bar{\theta}\right) \right| \left| \frac{(j-1)\pi}{n+1} \right|,$$

where $\bar{\theta} \in (\frac{\pi}{n+1}, \frac{j\pi}{n+1})$. If $1 \leq j < s + 1$ then $|j - 1| < s$, so

$$\left| r\left(\frac{j\pi}{n+1}\right) - \lambda_j(\mathcal{P}_n(f,g)) \right| \leq ||r'||_\infty s\pi h = Ch.$$

Hence,

$$\left| \lambda_j(\mathcal{P}_n(f,g)) - r\left(\frac{j\pi}{n+1}\right) \right| \leq Ch \quad \forall\, j\, \forall\, n.$$

□

Here, we present a second proof of the previous theorem.

Proof We adopt the very same notation used for the first proof. First, we notice that the low rank matrices $H_n(f)$ and $H_n(g)$ are also Hermitian matrices because $T_n(f)$, $T_n(g)$, $\tau_n(f)$, and $\tau_n(g)$ are Hermitian matrices. Let $\mathbf{x}_i$ and $\lambda_i\left(\mathcal{P}_n(f,g)\right)$ be a pair eigenvector and eigenvalue of $\mathcal{P}_n(f,g)$. Then we can write

$$\mathcal{P}_n(f,g)\mathbf{x}_i = \lambda_i\left(\mathcal{P}_n(f,g)\right)\mathbf{x}_i.$$

By multiplying the previous equation from the left by the matrix $T_n(g) = \tau_n(g) + H_n(g)$, we obtain

$$\left(\tau_n(f) + H_n(f)\right)\mathbf{x}_i = \lambda_i\left(\mathcal{P}_n(f,g)\right)\left(\tau_n(g) + H_n(g)\right)\mathbf{x}_i,$$

which is equivalent to

$$\left(\tau_n(f) + H_n(f) - \lambda_i\left(\mathcal{P}_n(f,g)\right)H_n(g)\right)\mathbf{x}_i = \lambda_i\left(\mathcal{P}_n(f,g)\right)\tau_n(g)\mathbf{x}_i.$$

Finally, by setting $\mathbf{y}_i = \tau_n^{1/2}(g)\mathbf{x}_i$ and by multiplying from the left by the matrix $\tau_n^{-1/2}(g)$, we have

$$\tau^{-1/2}(g)\left(\tau_n(f) + H_n(f) - \lambda_i\left(\mathcal{P}_n(f,g)\right)H_n(g)\right)\tau^{-1/2}(g)\mathbf{y}_i = \lambda_i\left(\mathcal{P}_n(f,g)\right)\mathbf{y}_i. \tag{31}$$

Equation (31) tells us that $\lambda_i\left(\mathcal{P}_n(f,g)\right)$ is also the eigenvalue of

$$\tau_n^{-1/2}(g)\left(\tau_n(f) + H_n(f) - \lambda_i\left(\mathcal{P}_n(f,g)\right)H_n(g)\right)\tau_n^{-1/2}(g).$$

We can write

$$\tau_n^{-1/2}(g)\left(\tau_n(f) + H_n(f) - \lambda_i\left(\mathcal{P}_n(f,g)\right)H_n(g)\right)\tau_n^{-1/2}(g)$$

as

$$\begin{aligned}\tau_n^{-1/2}(g)\tau_n(f)\tau_n^{-1/2}(g) &+ \tau_n^{-1/2}(g)\left(H_n(f) - \lambda_i\left(\mathcal{P}_n(f,g)\right)H_n(g)\right)\tau_n^{-1/2}(g) \\ &= \tau_n(f/g) + \tau_n^{-1/2}(g)\left(H_n(f) - \lambda_i\left(\mathcal{P}_n(f,g)\right)H_n(g)\right)\tau_n^{-1/2}(g).\end{aligned} \tag{32}$$

Notice that the rank of any linear combination of $H_n(f)$ and $H_n(g)$ is $R_{f,g} = \max\{\text{rank}(H_n(f)), \text{rank}(H_n(g))\}$ and the argument is the special Hankel structure of $H_n(f)$ and $H_n(g)$. As a conclusion, from the expression above, using the MinMax characterization and the interlacing theorem for Hermitian matrices, we write

$$\lambda_{i-R_{f,g}}(\tau_n(f/g)) \leq \lambda_i\left(\mathcal{P}_n(f,g)\right) \leq \lambda_{i+R_{f,g}}(\tau_n(f/g)), \tag{33}$$

where $i \in \{R_{f,g}-1, \cdots, n-R_{f,g}\}$, which leads again to the proof of Theorem 1. □

Remark With regard to Theorem 1, the case where r is bounded and nonmonotone is even easier. If we consider $\hat{r}$, the monotone nondecreasing rearrangement of r on $[0, \pi]$, taking into account that the derivative of r has at most a finite number S of sign changes, we deduce that $\hat{r}$ is Lipschitz continuous and its Lipschitz constant is bounded by $\|r'\|_\infty$ (notice that $\hat{r}$ is not necessarily continuously differentiable, but the derivative of $\hat{r}$ has at most S points of discontinuity). Furthermore, the eigenvalues of $\tau_n(r)$ are exactly given

$$r\left(\frac{j\pi}{n+1}\right)$$

so that, by ordering these values nondecreasingly, we deduce that they coincide with $\hat{r}(x_{j,h})$, with $x_{j,h}$ of the form $\frac{j\pi}{n+1}(1+o(1))$. With these premises, the proof follows exactly the same steps as in Theorem 1, using the MinMax characterization and the interlacing theorem for Hermitian matrices.

References

1. Barrera, M., Grudsky, S.M.: Asymptotics of eigenvalues for pentadiagonal symmetric Toeplitz matrices. Oper. Theory Adv. Appl. **259**, 51–77 (2017)
2. Bhatia, R.: Matrix Analysis Graduate Texts in Mathematics, vol. 169. Springer, New York (1997)
3. Bini, D., Capovani, M.: Spectral and computational properties of band symmetric Toeplitz matrices. Linear Algebra Appl. **52–53**, 99–126 (1983)
4. Bogoya, J.M., Böttcher, A., Grudsky, S.M., Maximenko, E.A.: Eigenvalues of Hermitian Toeplitz matrices with smooth simple-loop symbols. J. Math. Anal. Appl. **422**, 1308–1334 (2015)
5. Bogoya, J.M., Grudsky, S.M., Maximenko, E.A.: Eigenvalues of Hermitian Toeplitz matrices generated by simple-loop symbols with relaxed smoothness. Oper. Theory Adv. Appl. **259**, 179–212 (2017)
6. Böttcher, A., Grudsky, S.M., Maximenko, E.A.: Inside the eigenvalues of certain Hermitian Toeplitz band matrices. J. Comput. Appl. Math. **233**, 2245–2264 (2010)
7. Böttcher, A., Silbermann, B.: Introduction to Large Truncated Toeplitz Matrices. Springer (1999)
8. Brezinski, C., Redivo Zaglia, M.: Extrapolation Methods: Theory and Practice. Elsevier Science Publishers B.V, North-Holland (1991)
9. Chan, R.H., Ng, M.: Conjugate gradient methods for Toeplitz systems. SIAM Rev. **38-3**, 427–482 (1996)
10. Chan, R.H., Tang, P.: Fast band–Toeplitz preconditioners for Hermitian Toeplitz systems. SIAM J. Sci. Comput. **15**, 164–171 (1994)
11. Di Benedetto, F., Fiorentino, G., Serra, S.: C.G. Preconditioning for Toeplitz matrices. Comput. Math. Appl. **25-6**, 33–45 (1993)
12. Ekström, S.-E., Serra-Capizzano, S.: Eigenvalues and Eigenvectors of Banded Toeplitz Matrices and the Related Symbols. Technical report, 2017-010, Department of Information Technology Uppsala University (2017)

13. Ekström S.-E., Garoni C., Serra-Capizzano S.: Are the eigenvalues of banded symmetric Toeplitz matrices known in almost closed form? Exp. Math., in press (2017). https://doi.org/10.1080/10586458.2017.1320241
14. Garoni, C., Serra-Capizzano, S.: Generalized Locally Toeplitz Sequences: Theory and Applications, vol. I. Springer (2017)
15. Huckle, T., Serra-Capizzano, S., Tablino-Possio, C.: Preconditioning strategies for non-Hermitian Toeplitz linear systems. Numer. Linear Algebra Appl. **12-2/3**, 211–220 (2005)
16. Huckle, T., Serra-Capizzano, S., Tablino-Possio, C.: Preconditioning strategies for Hermitian indefinite Toeplitz linear systems. SIAM. J. Sci. Comput. **25-5**, 1633–1654 (2004)
17. Serra-Capizzano, S.: New PCG based algorithms for the solution of Hermitian Toeplitz systems. Calcolo **32**, 53–176 (1995)
18. Serra-Capizzano, S.: Optimal, quasi-optimal and superlinear band-Toeplitz preconditioners for asymptotically ill-conditioned positive definite Toeplitz systems. Math. Comp. **66-218**, 651–665 (1997)
19. Serra-Capizzano, S.: An ergodic theorem for classes of preconditioned matrices. Linear Algebra Appl. **282-1/3**, 161–183 (1998)
20. Stoer, J., Bulirsch, R.: Introduction to Numerical Analysis, 3rd edn. Springer (2002)

Paper III [redacted]

Numer Algor
https://doi.org/10.1007/s11075-018-0508-0

ORIGINAL PAPER

A matrix-less and parallel interpolation–extrapolation algorithm for computing the eigenvalues of preconditioned banded symmetric Toeplitz matrices

Sven-Erik Ekström[1] · **Carlo Garoni**[2,3]

Received: 22 August 2017 / Accepted: 4 March 2018

Abstract In the past few years, Bogoya, Böttcher, Grudsky, and Maximenko obtained the precise asymptotic expansion for the eigenvalues of a Toeplitz matrix $T_n(f)$, under suitable assumptions on the generating function f, as the matrix size n goes to infinity. On the basis of several numerical experiments, it was conjectured by Serra-Capizzano that a completely analogous expansion also holds for the eigenvalues of the preconditioned Toeplitz matrix $T_n(u)^{-1}T_n(v)$, provided $f = v/u$ is monotone and further conditions on u and v are satisfied. Based on this expansion, we here propose and analyze an interpolation–extrapolation algorithm for computing the eigenvalues of $T_n(u)^{-1}T_n(v)$. The algorithm is suited for parallel implementation and it may be called "matrix-less" as it does not need to store the entries of the matrix. We illustrate the performance of the algorithm through numerical experiments and we also present its generalization to the case where $f = v/u$ is non-monotone.

Keywords Preconditioned Toeplitz matrices · Eigenvalues · Asymptotic eigenvalue expansion · Polynomial interpolation · Extrapolation

Mathematics Subject Classification (2010) 15B05 · 65F15 · 65D05 · 65B05

✉ Sven-Erik Ekström
sven-erik.ekstrom@it.uu.se

Carlo Garoni
carlo.garoni@usi.ch; carlo.garoni@uninsubria.it

[1] Department of Information Technology, Division of Scientific Computing, Uppsala University, ITC, Lägerhyddsv. 2, Hus 2, P.O. Box 337, SE-751 05 Uppsala, Sweden

[2] Institute of Computational Science, University of Italian Switzerland (USI), Via Giuseppe Buffi 13, 6900 Lugano, Switzerland

[3] Department of Science and High Technology, University of Insubria, Via Valleggio 11, 22100 Como, Italy

Published online: 24 March 2018

1 Introduction

A matrix of the form

$$\left[a_{i-j}\right]_{i,j=1}^{n} = \begin{bmatrix} a_0 & a_{-1} & \cdots & \cdots & a_{-(n-1)} \\ a_1 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & a_{-1} \\ a_{n-1} & \cdots & \cdots & a_1 & a_0 \end{bmatrix},$$

whose entries are constant along each diagonal, is called a Toeplitz matrix. Given a function $g : [-\pi, \pi] \to \mathbb{C}$ belonging to $L^1([-\pi, \pi])$, the nth Toeplitz matrix associated with g is defined as

$$T_n(g) = \left[\hat{g}_{i-j}\right]_{i,j=1}^{n},$$

where the numbers $\hat{g}_k$ are the Fourier coefficients of g,

$$\hat{g}_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\theta) \mathrm{e}^{-\mathrm{i}k\theta} \mathrm{d}\theta, \qquad k \in \mathbb{Z}.$$

We refer to $\{T_n(g)\}_n$ as the Toeplitz sequence generated by g, which in turn is called the generating function of $\{T_n(g)\}_n$. It is not difficult to see that, whenever g is real, $T_n(g)$ is Hermitian for all n. Moreover, if g is real non-negative and not almost everywhere equal to zero in $[-\pi, \pi]$, then $T_n(g)$ is Hermitian positive definite for all n; see [9, 14]. In the case where g is a real cosine trigonometric polynomial (RCTP), that is, a function of the form

$$g(\theta) = \hat{g}_0 + 2 \sum_{k=1}^{m} \hat{g}_k \cos(k\theta), \qquad \hat{g}_0, \hat{g}_1, \ldots, \hat{g}_m \in \mathbb{R}, \qquad m \in \mathbb{N},$$

the nth Toeplitz matrix generated by g is the real symmetric banded matrix given by

$$T_n(g) = \begin{bmatrix} \hat{g}_0 & \hat{g}_1 & \cdots & \hat{g}_m & & & & & \\ \hat{g}_1 & \ddots & \ddots & & \ddots & & & & \\ \vdots & \ddots & \ddots & \ddots & & \ddots & & & \\ \hat{g}_m & & \ddots & \ddots & \ddots & & \ddots & & \\ & \ddots & & \ddots & \ddots & \ddots & & \ddots & \\ & & \boxed{\hat{g}_m \cdots \hat{g}_1 \; \hat{g}_0 \; \hat{g}_1 \cdots \hat{g}_m} & & & & & & \\ & & & \ddots & & \ddots & \ddots & \ddots & \\ & & & & \ddots & & \ddots & \ddots & \ddots \\ & & & & & \ddots & & \ddots & \ddots \\ & & & & & & \ddots & & \ddots \\ & & & & & & & \hat{g}_m \cdots \hat{g}_1 & \hat{g}_0 \end{bmatrix}.$$

The numerical approximation of the eigenvalues of real symmetric banded Toeplitz matrices is a problem that has been faced by several authors; see, e.g., Arbenz [2], Badía and Vidal [3], Bini and Pan [5], the authors and Serra-Capizzano [13], and Trench [16–20]. Less attention has been devoted to the numerical approximation of the eigenvalues of preconditioned banded symmetric Toeplitz matrices of the form $T_n(u)^{-1}T_n(v)$, with u, v being RCTPs. Yet, this problem is worthy of consideration as noted in [4, Section 1]. Some algorithms to solve it have been proposed in [1, 4]. For general discussions on the various algorithmic proposals for solving eigenvalue problems related to banded Toeplitz matrices, we refer the reader [2, Section 1] and [4, Section 1].

In this paper, we propose a new algorithm for the numerical approximation of the eigenvalues of preconditioned banded symmetric Toeplitz matrices. The algorithm relies on the following conjecture, which has been formulated by Serra-Capizzano in [1], on the basis of several numerical experiments.

Conjecture 1 Let u, v be RCTPs, with $u > 0$ on $(0, \pi)$, and suppose that $f = v/u$ is monotone increasing over $(0, \pi)$. Set $X_n = T_n(u)^{-1}T_n(v)$ for all n. Then, for every integer $\alpha \geq 0$, every n and every $j = 1, \ldots, n$, the following asymptotic expansion holds:

$$\lambda_j(X_n) = f(\theta_{j,n}) + \sum_{k=1}^{\alpha} c_k(\theta_{j,n})h^k + E_{j,n,\alpha}, \tag{1}$$

where:

- The eigenvalues of X_n are arranged in non-decreasing order, $\lambda_1(X_n) \leq \ldots \leq \lambda_n(X_n)$.[1]
- $\{c_k\}_{k=1,2,\ldots}$ is a sequence of functions from $(0, \pi)$ to $\mathbb{R}$ which depends only on u, v.
- $h = \frac{1}{n+1}$ and $\theta_{j,n} = \frac{j\pi}{n+1} = j\pi h$.
- $E_{j,n,\alpha} = O(h^{\alpha+1})$ is the remainder (the error), which satisfies the inequality $|E_{j,n,\alpha}| \leq C_\alpha h^{\alpha+1}$ for some constant C_α depending only on α, u, v.

In the case where $u = 1$ identically, Conjecture 1 was originally formulated and supported through numerical experiments in [13]. In the case where $u = 1$ identically and v satisfies some additional assumptions, Conjecture 1 was formally proved by Bogoya, Böttcher, Grudsky, and Maximenko in a sequence of recent papers [6, 8, 10].

Assuming Conjecture 1, in Section 2 of this paper, we describe and analyze a new algorithm for computing the eigenvalues of $X_n = T_n(u)^{-1}T_n(v)$; and in Section 3, we illustrate its performance through numerical experiments. The algorithm, which is suited for *parallel implementation* and may be called *matrix-less* as it does not need to store the entries of X_n, combines the extrapolation procedure proposed in [1, 13]—which allows the computation of *some* of the eigenvalues of X_n—with an appropriate interpolation process, thus allowing the simultaneous computation of *all* the eigenvalues of X_n. In Section 4, we provide a generalization of the

[1] Note that the eigenvalues of X_n are real, because X_n is similar to the symmetric matrix $T_n(u)^{-1/2}T_n(v)T_n(u)^{-1/2}$.

algorithm to the case where $f = v/u$ is non-monotone; this generalization is based on another conjecture which is analogous to Conjecture 1 and which will be discussed later on. In Section 5, we draw conclusions and suggest possible future lines of research.

2 The algorithm

Throughout this paper, we associate with each positive integer $n \in \mathbb{N} = \{1, 2, 3, \ldots\}$ the stepsize $h = \frac{1}{n+1}$ and the grid points $\theta_{j,n} = j\pi h$, $j = 1, \ldots, n$. For notational convenience, we will always denote a positive integer and the associated stepsize in a similar way, in the sense that if the positive integer is denoted by n, the associated stepsize is denoted by h; if the positive integer is denoted by n_j, the associated stepsize is denoted by h_j; etc. Throughout this section, we make the following assumptions:

- u, v, f are as in Conjecture 1.
- $n, n_1, \alpha \in \mathbb{N}$ are fixed parameters and $X_n = T_n(u)^{-1} T_n(v)$.
- $n_k = 2^{k-1}(n_1 + 1) - 1$ for $k = 2, \ldots, \alpha$.
- $j_k = 2^{k-1} j_1$ for $j_1 = 1, \ldots, n_1$ and $k = 2, \ldots, \alpha$. Note that $j_k = j_k(j_1)$ depends not only on k but also on j_1, though we hide the dependence on j_1 for notational simplicity. Note also that j_k is the index in $\{1, \ldots, n_k\}$ such that $\theta_{j_k,n_k} = \theta_{j_1,n_1}$. Hence, the grid $\{\theta_{j_k,n_k} : \ j_1 = 1, \ldots, n_1\}$ is the same as the grid $\{\theta_{j_1,n_1} : \ j_1 = 1, \ldots, n_1\}$ for all $k = 2, \ldots, \alpha$.

A graphical representation of the grids $\{\theta_{1,n_k}, \ldots, \theta_{n_k,n_k}\}$, $k = 1, \ldots, \alpha$, is reported in Fig. 1 for $n_1 = 5$ and $\alpha = 4$. For each "level" $k = 2, \ldots, \alpha$, the corresponding red circles highlight the subgrid $\{\theta_{j_k,n_k} : \ j_1 = 1, \ldots, n_1\}$ which coincides with the coarsest grid $\{\theta_{j_1,n_1} : \ j_1 = 1, \ldots, n_1\}$.

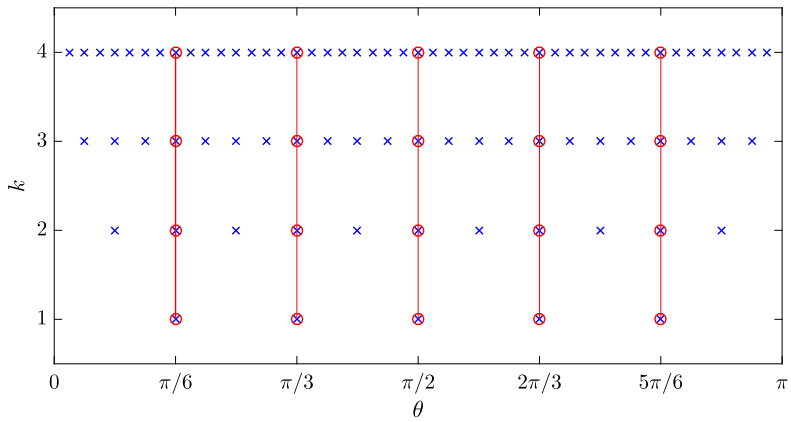


Fig. 1 Representation of the grids $\{\theta_{1,n_k}, \ldots, \theta_{n_k,n_k}\}$, $k = 1, \ldots, \alpha$, for $n_1 = 5$ and $\alpha = 4$

2.1 Description and formulation of the algorithm

The algorithm we are going to describe is designed for computing the eigenvalues of X_n in the case where n is large with respect to $n_1, \ldots, n_\alpha$, so that the computation of the eigenvalues of X_n is hard from a computational viewpoint but the computation of the eigenvalues of $X_{n_1}, \ldots, X_{n_\alpha}$—which is required in the algorithm—can be efficiently performed by any standard eigensolver (e.g., MATLAB's `eig` function); see also Remark 1 below. The algorithm is composed of two phases: a first phase where we invoke extrapolation procedures from [1, 13] and a second phase where local interpolation techniques are employed.

Extrapolation For each fixed $j_1 = 1, \ldots, n_1$, we apply α times the expansion (1) with $n = n_1, n_2, \ldots, n_\alpha$ and $j = j_1, j_2, \ldots, j_\alpha$. Since $\theta_{j_1,n_1} = \theta_{j_2,n_2} = \ldots = \theta_{j_\alpha,n_\alpha}$ (by definition of $j_2, \ldots, j_\alpha$), we obtain

$$\begin{cases} E_{j_1,n_1,0} &= c_1(\theta_{j_1,n_1})h_1 + c_2(\theta_{j_1,n_1})h_1^2 + \ldots + c_\alpha(\theta_{j_1,n_1})h_1^\alpha + E_{j_1,n_1,\alpha} \\ E_{j_2,n_2,0} &= c_1(\theta_{j_1,n_1})h_2 + c_2(\theta_{j_1,n_1})h_2^2 + \ldots + c_\alpha(\theta_{j_1,n_1})h_2^\alpha + E_{j_2,n_2,\alpha} \\ \quad\vdots & \\ E_{j_\alpha,n_\alpha,0} &= c_1(\theta_{j_1,n_1})h_\alpha + c_2(\theta_{j_1,n_1})h_\alpha^2 + \ldots + c_\alpha(\theta_{j_1,n_1})h_\alpha^\alpha + E_{j_\alpha,n_\alpha,\alpha} \end{cases} \tag{2}$$

where

$$E_{j_k,n_k,0} = \lambda_{j_k}(X_{n_k}) - f(\theta_{j_1,n_1}), \qquad k = 1, \ldots, \alpha,$$

and

$$|E_{j_k,n_k,\alpha}| \le C_\alpha h_k^{\alpha+1}, \qquad k = 1, \ldots, \alpha. \tag{3}$$

Let $\tilde{c}_1(\theta_{j_1,n_1}), \ldots, \tilde{c}_\alpha(\theta_{j_1,n_1})$ be the approximations of $c_1(\theta_{j_1,n_1}), \ldots, c_\alpha(\theta_{j_1,n_1})$ obtained by removing all the errors $E_{j_1,n_1,\alpha}, \ldots, E_{j_\alpha,n_\alpha,\alpha}$ in (2) and by solving the resulting linear system:

$$\begin{cases} E_{j_1,n_1,0} &= \tilde{c}_1(\theta_{j_1,n_1})h_1 + \tilde{c}_2(\theta_{j_1,n_1})h_1^2 + \ldots + \tilde{c}_\alpha(\theta_{j_1,n_1})h_1^\alpha \\ E_{j_2,n_2,0} &= \tilde{c}_1(\theta_{j_1,n_1})h_2 + \tilde{c}_2(\theta_{j_1,n_1})h_2^2 + \ldots + \tilde{c}_\alpha(\theta_{j_1,n_1})h_2^\alpha \\ \quad\vdots & \\ E_{j_\alpha,n_\alpha,0} &= \tilde{c}_1(\theta_{j_1,n_1})h_\alpha + \tilde{c}_2(\theta_{j_1,n_1})h_\alpha^2 + \ldots + \tilde{c}_\alpha(\theta_{j_1,n_1})h_\alpha^\alpha \end{cases} \tag{4}$$

Note that this way of computing approximations for $c_1(\theta_{j_1,n_1}), \ldots, c_\alpha(\theta_{j_1,n_1})$ was already proposed in [1, 13], and it is completely analogous to the Richardson extrapolation procedure that is employed in the context of Romberg integration to accelerate the convergence of the trapezoidal rule [15, Section 3.4]. In this regard, the asymptotic expansion (1) plays here the same role as the Euler–Maclaurin summation formula [15, Section 3.3]. For more advanced studies on extrapolation methods, we refer the reader to [11]. The next theorem shows that the approximation error $|c_k(\theta_{j_1,n_1}) - \tilde{c}_k(\theta_{j_1,n_1})|$ is $O(h_1^{\alpha-k+1})$.

Theorem 1 There exists a constant A_α depending only on α, u, v such that, for $j_1 = 1, \ldots, n_1$ and $k = 1, \ldots, \alpha$,

$$|c_k(\theta_{j_1,n_1}) - \tilde{c}_k(\theta_{j_1,n_1})| \leq A_\alpha h_1^{\alpha-k+1}. \tag{5}$$

Proof See Appendix A. □

Interpolation Fix an index $j \in \{1, \ldots, n\}$. To compute an approximation of $\lambda_j(X_n)$ through the expansion (1), we would need the value $c_k(\theta_{j,n})$ for each $k = 1, \ldots, \alpha$. Of course, $c_k(\theta_{j,n})$ is not available in practice, but we can approximate it by interpolating in some way the values $\tilde{c}_k(\theta_{j_1,n_1})$, $j_1 = 1, \ldots, n_1$. For example, we may define $\tilde{c}_k(\theta)$ as the interpolation polynomial of the data $(\theta_{1,n_1}, \tilde{c}_k(\theta_{1,n_1})), \ldots, (\theta_{n_1,n_1}, \tilde{c}_k(\theta_{n_1,n_1}))$—so that $\tilde{c}_k(\theta)$ is expected to be an approximation of $c_k(\theta)$ over the whole interval $(0, \pi)$—and take $\tilde{c}_k(\theta_{j,n})$ as an approximation to $c_k(\theta_{j,n})$. It is known, however, that interpolation over a large number of uniform nodes is not advisable as it may give rise to spurious oscillations (Runge's phenomenon [12, p. 78]). It is therefore better to adopt another kind of approximation. An alternative could be the following: we approximate $c_k(\theta)$ by the spline function $\tilde{c}_k(\theta)$ which is linear on each interval $[\theta_{j_1,n_1}, \theta_{j_1+1,n_1}]$ and takes the value $\tilde{c}_k(\theta_{j_1,n_1})$ at θ_{j_1,n_1} for all $j_1 = 1, \ldots, n_1$. This strategy removes for sure any spurious oscillation, yet it is not accurate. In particular, it does not preserve the accuracy of approximation at the nodes θ_{j_1,n_1} established in Theorem 1, i.e., there is no guarantee that $|c_k(\theta) - \tilde{c}_k(\theta)| \leq B_\alpha h_1^{\alpha-k+1}$ for $\theta \in (0, \pi)$ or $|c_k(\theta_{j,n}) - \tilde{c}_k(\theta_{j,n})| \leq B_\alpha h_1^{\alpha-k+1}$ for $j = 1, \ldots, n$, with B_α being a constant depending only on α, u, v. As proved in Theorem 2, a local approximation strategy that preserves the accuracy (5), at least if $c_k(\theta)$ is sufficiently smooth, is the following: let $\theta^{(1)}, \ldots, \theta^{(\alpha-k+1)}$ be $\alpha - k + 1$ points of the grid $\{\theta_{1,n_1}, \ldots, \theta_{n_1,n_1}\}$ which are closest to the point $\theta_{j,n}$,[2] and let $\tilde{c}_{k,j}(\theta)$ be the interpolation polynomial of the data $(\theta^{(1)}, \tilde{c}_k(\theta^{(1)})), \ldots, (\theta^{(\alpha-k+1)}, \tilde{c}_k(\theta^{(\alpha-k+1)}))$; then, we approximate $c_k(\theta_{j,n})$ by $\tilde{c}_{k,j}(\theta_{j,n})$. Note that, by selecting $\alpha - k + 1$ points from $\{\theta_{1,n_1}, \ldots, \theta_{n_1,n_1}\}$, we are implicitly assuming that $n_1 \geq \alpha - k + 1$.

Theorem 2 Let $1 \leq k \leq \alpha$, and suppose $n_1 \geq \alpha - k + 1$ and $c_k \in C^{\alpha-k+1}([0, \pi])$. For $j = 1, \ldots, n$, if $\theta^{(1)}, \ldots, \theta^{(\alpha-k+1)}$ are $\alpha - k + 1$ points of $\{\theta_{1,n_1}, \ldots, \theta_{n_1,n_1}\}$ which are closest to $\theta_{j,n}$, and if $\tilde{c}_{k,j}(\theta)$ is the interpolation polynomial of the data $(\theta^{(1)}, \tilde{c}_k(\theta^{(1)})), \ldots, (\theta^{(\alpha-k+1)}, \tilde{c}_k(\theta^{(\alpha-k+1)}))$, then

$$|c_k(\theta_{j,n}) - \tilde{c}_{k,j}(\theta_{j,n})| \leq B_\alpha h_1^{\alpha-k+1} \tag{6}$$

for some constant B_α depending only on α, u, v.

Proof See Appendix A. □

[2] These $\alpha - k + 1$ points are uniquely determined by $\theta_{j,n}$ except in the case where $\theta_{j,n}$ coincides with either a grid point θ_{j_1,n_1} or the midpoint between two consecutive grid points θ_{j_1,n_1} and θ_{j_1+1,n_1}.

Formulation of the algorithm We are now ready to formulate our algorithm for computing the eigenvalues of X_n. As we shall see in Remark 4, the algorithm is suited for *parallel implementation*. Since it does not even need to store the entries of X_n, it may be called *matrix-less*. It can be used for computing either a specific eigenvalue $\lambda_j(X_n)$, a subset of the eigenvalues of X_n, or the whole spectrum of X_n. A plain (non-parallel) MATLAB implementation of this algorithm is reported in Appendix B.

Algorithm 1 Given two RCTPs u, v (with $u > 0$ on $(0, \pi)$ and $f = v/u$ monotone increasing over $(0, \pi)$ as in Conjecture 1), three integers $n, n_1, \alpha \in \mathbb{N}$ with $n_1 \geq \alpha$, and $S \subseteq \{1, \ldots, n\}$, we compute an approximation of the eigenvalues $\{\lambda_j(X_n) : j \in S\}$ as follows:

1. For $j_1 = 1, \ldots, n_1$ compute $\tilde{c}_1(\theta_{j_1,n_1}), \ldots, \tilde{c}_\alpha(\theta_{j_1,n_1})$ by solving (4)
2. For $j \in S$
   - For $k = 1, \ldots, \alpha$
     - Determine $\alpha - k + 1$ points $\theta^{(1)}, \ldots, \theta^{(\alpha-k+1)} \in \{\theta_{1,n_1}, \ldots, \theta_{n_1,n_1}\}$ which are closest to $\theta_{j,n}$
     - Compute $\tilde{c}_{k,j}(\theta_{j,n})$, where $\tilde{c}_{k,j}(\theta)$ is the interpolation polynomial of $(\theta^{(1)}, \tilde{c}_k(\theta^{(1)})), \ldots, (\theta^{(\alpha-k+1)}, \tilde{c}_k(\theta^{(\alpha-k+1)}))$
   - Compute $\tilde{\lambda}_j(X_n) = f(\theta_{j,n}) + \sum_{k=1}^{\alpha} \tilde{c}_{k,j}(\theta_{j,n}) h^k$
3. Return $\{\tilde{\lambda}_j(X_n) : j \in S\}$ as an approximation to $\{\lambda_j(X_n) : j \in S\}$

Remark 1 Algorithm 1 is specifically designed for computing the eigenvalues of X_n in the case where the matrix size n is quite large. When applying this algorithm, it is implicitly assumed that n_1 and α are small (much smaller than n), so that each $n_k = 2^{k-1}(n_1 + 1) - 1$ is small as well and the computation of the eigenvalues of X_{n_k}—which is required in the first step—can be efficiently performed by any standard eigensolver (e.g., MATLAB's `eig` function).

Remark 2 A careful evaluation shows that the computational cost of Algorithm 1 is bounded by

$$C(\alpha^2 n_1 + \alpha^3 |S|) + \sum_{k=1}^{\alpha} C_{\text{eig}}(n_k),$$

where $|S|$ is the cardinality of S, C is a constant depending only on f, and $C_{\text{eig}}(n_k)$ is the cost for computing the eigenvalues of X_{n_k}.

Remark 3 Algorithm 1 can be optimized in several ways. For example, if $S = \{j\}$, so that only the jth eigenvalue $\lambda_j(X_n)$ must be computed, then in the first step one can just compute the values $\tilde{c}_1(\theta_{j_1,n_1}), \ldots, \tilde{c}_\alpha(\theta_{j_1,n_1})$ for $\theta_{j_1,n_1} \in \{\theta^{(1)}, \ldots, \theta^{(\alpha)}\}$, where $\theta^{(1)}, \ldots, \theta^{(\alpha)}$ are α points in $\{\theta_{1,n_1}, \ldots, \theta_{n_1,n_1}\}$ which are closest to $\theta_{j,n}$. Indeed, only these values are needed in the second step. A similar consideration applies in the case where only the extremal eigenvalues of X_n must be computed, and also in the case where S is a small subset of $\{1, \ldots, n\}$ of the form $\{j, \ldots, j + r\}$, with $r \ll n$.

Remark 4 Suppose $|S| = n$ and consider the ideal situation where we have n processors. Then, the jth processor can compute the jth eigenvalue $\lambda_j(X_n)$ independently of the others. In view of Remark 3, the jth processor can act as follows:

- In the first step of the algorithm, it computes only the values $\tilde{c}_1(\theta_{j_1,n_1}), \ldots, \tilde{c}_\alpha(\theta_{j_1,n_1})$ for $\theta_{j_1,n_1} \in \{\theta^{(1)}, \ldots, \theta^{(\alpha)}\}$, where $\theta^{(1)}, \ldots, \theta^{(\alpha)}$ are α points in $\{\theta_{1,n_1}, \ldots, \theta_{n_1,n_1}\}$ which are closest to $\theta_{j,n}$.
- It performs the second step of the algorithm for the index j only.

It is clear that such a parallel implementation is very fast as the computation of all the eigenvalues of X_n takes the same time as the computation of one eigenvalue only. A similar consideration also applies in the case where $|S| < n$ and we have $|S|$ processors, each of which has to compute only one of the requested $|S|$ eigenvalues. In a more realistic situation, we will not have a number of processors equal to $|S|$ if $|S|$ is large. Instead, we will have p processors with $p \ll |S|$. In this case, we can divide S into p different subsets $S_1, \ldots, S_p$ of approximately the same cardinality and assign to the ith processor the computation of the eigenvalues corresponding to S_i, $i = 1, \ldots, p$. When doing so, it is advisable that each S_i is constructed so that the "positions" $\theta_{j,n}$ of the related eigenvalues $\lambda_j(X_n)$ are close to each other, because in this way each processor will have the possibility to perform a reduced form of the first step of the algorithm, in analogy with what has been explained above for the case $p = |S|$. For example, if $|S| = n$ and n is a multiple of p, then we can assign to the ith processor the computation of the eigenvalues $\lambda_j(X_n)$ for $j = (i-1)(n/p) + 1, \ldots, i(n/p)$, so that in the first step of the algorithm the ith processor will only have to compute $\tilde{c}_1(\theta_{j_1,n_1}), \ldots, \tilde{c}_\alpha(\theta_{j_1,n_1})$ for θ_{j_1,n_1} in a neighborhood of the interval $[\theta_{(i-1)(n/p)+1,n}, \theta_{i(n/p),n}]$.

2.2 Error estimate

Theorem 3 Assume that Conjecture 1 holds. Suppose $n \geq n_1 \geq \alpha$ and $c_k \in C^{\alpha-k+1}([0,\pi])$ for $k = 1, \ldots, \alpha$. Let $(\tilde{\lambda}_1(X_n), \ldots, \tilde{\lambda}_n(X_n))$ be the approximation of $(\lambda_1(X_n), \ldots, \lambda_n(X_n))$ computed by Algorithm 1. Then, there exists a constant D_α depending only on α, u, v such that, for $j = 1, \ldots, n$,

$$|\lambda_j(X_n) - \tilde{\lambda}_j(X_n)| \leq D_\alpha h_1^\alpha h.$$

Proof By (1) and Theorem 2,

$$\begin{aligned}
|\lambda_j(X_n) - \tilde{\lambda}_j(X_n)| &= \left| f(\theta_{j,n}) + \sum_{k=1}^{\alpha} c_k(\theta_{j,n}) h^k + E_{j,n,\alpha} - f(\theta_{j,n}) - \sum_{k=1}^{\alpha} \tilde{c}_{k,j}(\theta_{j,n}) h^k \right| \\
&= \left| \sum_{k=1}^{\alpha} (c_k(\theta_{j,n}) - \tilde{c}_{k,j}(\theta_{j,n})) h^k + E_{j,n,\alpha} \right| \\
&\leq B_\alpha \sum_{k=1}^{\alpha} h_1^{\alpha-k+1} h^k + C_\alpha h^{\alpha+1} \leq D_\alpha h_1^\alpha h,
\end{aligned}$$

where $D_\alpha = (\alpha+1)\max(B_\alpha, C_\alpha)$. □

Remark 5 The error estimate provided in Theorem 3 suggests that the eigenvalue approximations provided by Algorithm 1 improve as n increases, i.e., as h decreases. Numerical experiments reveal that this is in fact the case (see Example 2 below).

Remark 6 Theorem 3 shows that, for any fixed $\alpha \geq 1$, the numerical eigenvalues computed by Algorithm 1 converge like h_1^α to the exact eigenvalues as n_1 grows. In practice, it is advisable to fix α and increase n_1 until a proper stopping criterion is reached. The other way (fix n_1 and increase α) is not advisable as the constant D_α in Theorem 3 apparently grows very quickly with α (see Example 1 below) and, consequently, there is no guarantee on the convergence of the algorithm as α grows (see Example 5 below).

3 Numerical experiments

In this section, we illustrate through numerical examples the performance of Algorithm 1. Numerical experiments have been performed with MATLAB R2015b (64 bit) on a platform with 4GB RAM, using an Intel® Celeron® Processor N2820 (up to 2.39 GHz, 1 MB L2 cache). The CPU times for Algorithm 1 refer to the plain MATLAB implementation reported in Appendix B. In what follows, the symbol $\varepsilon_{j,n}$ denotes the error $|\lambda_j(X_n) - \tilde{\lambda}_j(X_n)|$, which occurs when approximating the exact eigenvalue $\lambda_j(X_n)$ with the corresponding numerical eigenvalue $\tilde{\lambda}_j(X_n)$ computed by Algorithm 1. The inputs u, v, n, n_1, α with which Algorithm 1 is applied are specified in each example.

Example 1 Let

$$\begin{aligned} u(\theta) &= 1, \\ v(\theta) &= 6 - 8\cos(\theta) + 2\cos(2\theta). \end{aligned}$$

Note that $f(\theta) = v(\theta)/u(\theta) = v(\theta)$ is monotone increasing on $(0, \pi)$. Suppose we want to approximate the eigenvalues of $X_n = T_n(u)^{-1}T_n(v) = T_n(f)$ for $n = 5000$. Let $\tilde{\lambda}_j(X_n)$ be the approximation of $\lambda_j(X_n)$ obtained by applying Algorithm 1 with $n_1 = 10$ and $\alpha = 7$. In Fig. 2, we plot the errors $\varepsilon_{j,n}$ versus $\theta_{j,n}$ for $j = 1, \ldots, n$. We note that the largest errors are attained when either $\theta_{j,n} \approx 0$ or $\theta_{j,n} \approx \pi$. As highlighted also in Example 3 below, this is probably due to two concomitant factors:

- The errors $\varepsilon_{j,n}$ are supposed to be smaller for $\theta_{j,n} \in [\theta_{1,n_1}, \theta_{n_1,n_1}] = [\pi/11, 10\pi/11]$, because in this case the approximations $\tilde{c}_{k,j}(\theta_{j,n})$ computed by Algorithm 1 for the values $c_k(\theta_{j,n})$ are expected to be more accurate as the interpolation polynomial $\tilde{c}_{k,j}(\theta)$ is evaluated inside the convex hull of the interpolation nodes.
- $\theta = 0$ and $\theta = \pi$ are the two points on $[0, \pi]$ where f' vanishes, which means that the monotonicity of f is "weak" around these points (recall that Algorithm 1 works under the assumption that f is monotone as in Conjecture 1).

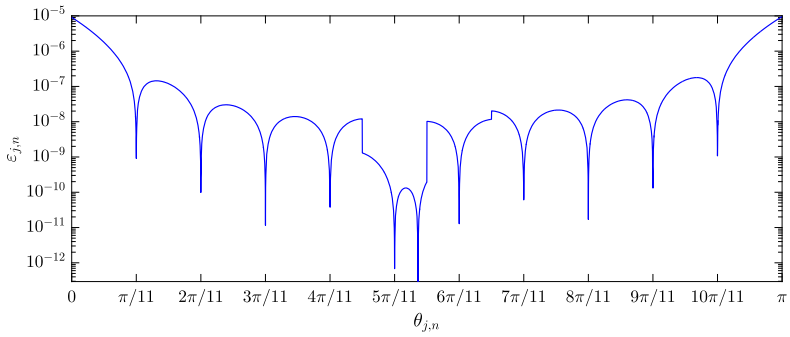


Fig. 2 Example 1: errors $\varepsilon_{j,n}$ versus $\theta_{j,n}$ for $j = 1, \ldots, n$ in the case where $u(\theta) = 1$, $v(\theta) = 6 - 8\cos(\theta) + 2\cos(2\theta)$, $n = 5000$, $n_1 = 10$, and $\alpha = 7$

In reference to the previous discussion, we note that the maximum error for $\theta_{j,n} \in [\theta_{1,n_1}, \theta_{n_1,n_1}]$ is given by

$$\max\{\varepsilon_{j,n} : \theta_{j,n} \in [\theta_{1,n_1}, \theta_{n_1,n_1}]\} \approx 1.7803 \cdot 10^{-7},$$

which is about two order of magnitude less than

$$\max_{j=1,\ldots,n} \varepsilon_{j,n} \approx 9.5167 \cdot 10^{-6}.$$

A careful look at Fig. 2 shows that, aside from the exceptional minimum attained inside the interval $(5\pi/11, 6\pi/11)$, the local minima of $\varepsilon_{j,n}$ are attained when $\theta_{j,n}$ is approximately equal to some of the grid points θ_{j_1,n_1}, $j_1 = 1, \ldots, n_1$. This is no surprise, because for $\theta_{j,n} = \theta_{j_1,n_1}$ we have $\tilde{c}_{k,j}(\theta_{j,n}) = \tilde{c}_k(\theta_{j_1,n_1})$ and $c_k(\theta_{j,n}) = c_k(\theta_{j_1,n_1})$, which means that the error of the approximation $\tilde{c}_{k,j}(\theta_{j,n}) \approx c_k(\theta_{j,n})$ reduces to the error of the approximation $\tilde{c}_k(\theta_{j_1,n_1}) \approx c_k(\theta_{j_1,n_1})$; that is, we are not introducing further error due to the interpolation process. To conclude, we make the following observation: for α, u, v as in this example, Theorem 3 yields

$$D_\alpha \geq \frac{\max_{j=1,\ldots,n} \varepsilon_{j,n}}{h_1^\alpha h} \approx 9.2745 \cdot 10^5 > \alpha^\alpha = 8.23543 \cdot 10^5.$$

This suggests that, unfortunately, the best constant D_α for which the error estimate of Theorem 3 is satisfied grows very quickly with α.

Example 2 Let u, v, f be as in Example 1. Suppose we want to approximate the eigenvalues of $X_n = T_n(u)^{-1}T_n(v) = T_n(f)$ for $n = 10000$. Let $\tilde{\lambda}_j(X_n)$ be the approximation of $\lambda_j(X_n)$ obtained by applying Algorithm 1 with $n_1 = 10$ and $\alpha = 7$ as in Example 1. In Fig. 3, we plot the errors $\varepsilon_{j,n}$ versus $\theta_{j,n}$ for $j = 1, \ldots, n$. We note that the errors in Fig. 3 are smaller than in Fig. 2. This shows that the eigenvalue approximations provided by Algorithm 1 improve as n increases (see also Remark 5).

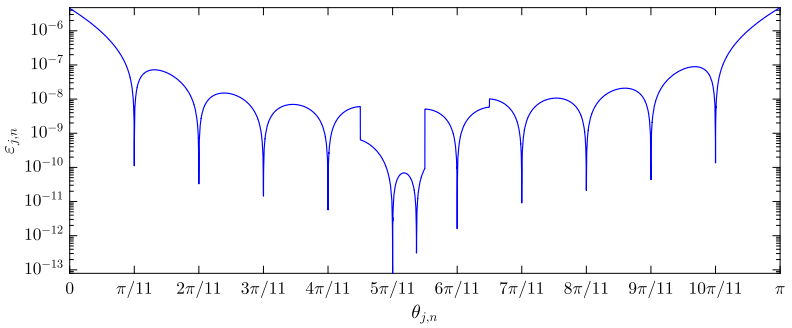


Fig. 3 Example 2: errors $\varepsilon_{j,n}$ versus $\theta_{j,n}$ for $j = 1, \ldots, n$ in the case where $u(\theta) = 1$, $v(\theta) = 6 - 8\cos(\theta) + 2\cos(2\theta)$, $n = 10000$, $n_1 = 10$, and $\alpha = 7$

Example 3 Let

$$\begin{aligned} u(\theta) &= 1, \\ v(\theta) &= -\frac{1}{4} - \frac{1}{2}\cos(\theta) + \frac{1}{4}\cos(2\theta) - \frac{1}{12}\cos(3\theta). \end{aligned}$$

Note that $f(\theta) = v(\theta)/u(\theta) = v(\theta)$ is monotone increasing on $(0, \pi)$. Suppose we want to approximate the eigenvalues of $X_n = T_n(u)^{-1}T_n(v) = T_n(f)$ for $n = 10000$. Let $\tilde{\lambda}_j^{(m)}(X_n)$ be the approximation of $\lambda_j(X_n)$ obtained by applying Algorithm 1 with $n_1 = 10 \cdot 2^{m-1}$ and $\alpha = 5$. In Fig. 4, we plot the errors $\varepsilon_{j,n}^{(m)} = |\lambda_j(X_n) - \tilde{\lambda}_j^{(m)}(X_n)|$ versus $\theta_{j,n}$ for $j = 1, \ldots, n$ and $m = 1, 2, 3, 4$. We see from the figure that, as m increases, the error decreases rather quickly everywhere except in a neighborhood of the point $\theta = \pi/3$ where f' vanishes. Actually, the three points of $[0, \pi]$ where f' vanishes are 0, $\pi/3$, π, and these are precisely the points around which the error is higher than elsewhere. We remark that, as in Examples 1 and 2, the error $\varepsilon_{j,n}^{(m)}$ attains its local minima when $\theta_{j,n}$ is approximately equal to some of the nodes $\theta_{1,n_1}, \ldots, \theta_{n_1,n_1}$.

Example 4 Let

$$\begin{aligned} u(\theta) &= 1, \\ v(\theta) &= \frac{301}{400} - \cos(\theta) + \frac{1}{5}\cos(2\theta) + \frac{1}{10}\cos(3\theta) - \frac{1}{20}\cos(4\theta) + \frac{1}{400}\cos(6\theta). \end{aligned}$$

Note that $f(\theta) = v(\theta)/u(\theta) = v(\theta)$ is monotone increasing on $(0, \pi)$ and $f'(\theta) = 0$ only for $\theta = 0, \pi$.[3] Suppose we want to approximate the eigenvalues of $X_n = T_n(u)^{-1}T_n(v) = T_n(f)$ for $n = 10000$. Let $\tilde{\lambda}_j^{(m)}(X_n)$ be the approximation of $\lambda_j(X_n)$ obtained by applying Algorithm 1 with $n_1 = 25 \cdot 2^{m-1}$ and $\alpha = 5$. In Fig. 5, we plot the errors $\varepsilon_{j,n}^{(m)}$ versus $\theta_{j,n}$ for $j = 1, \ldots, n$ and $m = 1, 2, 3, 4$. Considerations analogous to those of Example 3 apply also in this case.

[3]Note that we always have $g'(0) = g'(\pi) = 0$ whenever $g(\theta)$ is an RCTP.

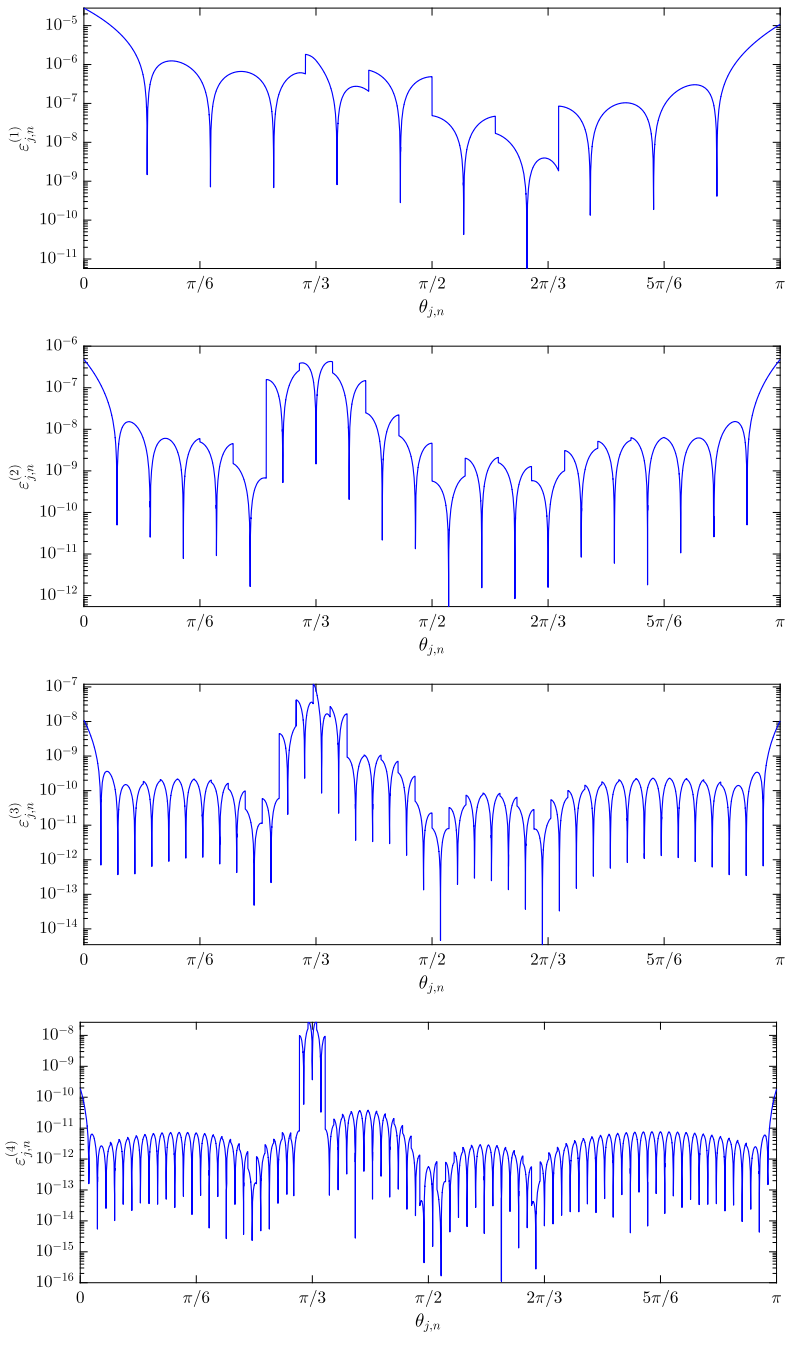


Fig. 4 Example 3: errors $\varepsilon_{j,n}^{(m)}$ versus $\theta_{j,n}$ for $j = 1, \ldots, n$, in the case where $u(\theta) = 1$, $v(\theta) = -\frac{1}{4} - \frac{1}{2}\cos(\theta) + \frac{1}{4}\cos(2\theta) - \frac{1}{12}\cos(3\theta)$, $n = 10000$, $n_1 = 10 \cdot 2^{m-1}$, and $\alpha = 5$

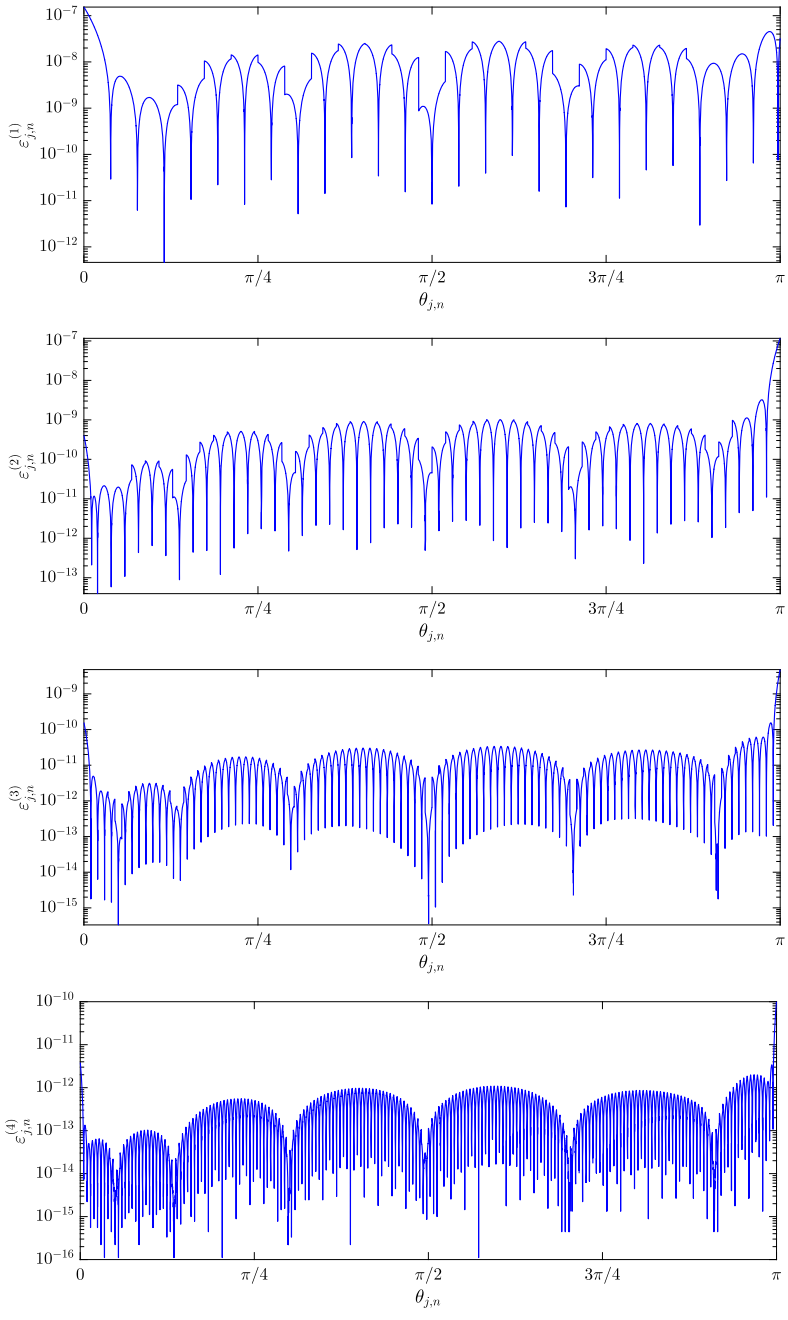


Fig. 5 Example 4: errors $\varepsilon_{j,n}^{(m)}$ versus $\theta_{j,n}$ for $j = 1, \ldots, n$, in the case where $u(\theta) = 1$, $v(\theta) = \frac{301}{400} - \cos(\theta) + \frac{1}{5}\cos(2\theta) + \frac{1}{10}\cos(3\theta) - \frac{1}{20}\cos(4\theta) + \frac{1}{400}\cos(6\theta)$, $n = 10000$, $n_1 = 25 \cdot 2^{m-1}$, and $\alpha = 5$

Example 5 Let u, v, f as in Example 4. Suppose we want to approximate the eigenvalues of $X_n = T_n(u)^{-1}T_n(v) = T_n(f)$ for $n = 10000$. Let $\tilde{\lambda}_j^{(m)}(X_n)$ be the approximation of $\lambda_j(X_n)$ obtained by applying Algorithm 1 with $n_1 = 25$ and $\alpha = 4 + m$. In Fig. 6, we plot the errors $\varepsilon_{j,n}^{(m)}$ versus $\theta_{j,n}$ for $j = 1, \ldots, n$ and $m = 1, 2, 3, 4$. By comparing Fig. 5 with Fig. 6, we see that the strategy of keeping n_1 fixed and increasing α is much less efficient than the strategy of keeping α fixed and increasing n_1. Indeed, while in Fig. 5 the error $\varepsilon_{j,n}^{(m)}$ decreases approximately in a uniform way by one order of magnitude as m increases, this is not observed in Fig. 6. Note also that the computational cost of Algorithm 1 for $n_1 = 25 \cdot 2^{m-1}$ and $\alpha = 5$ (as in Fig. 5) is essentially the same as the cost of Algorithm 1 for $n_1 = 25$ and $\alpha = 4 + m$ (as in Fig. 6), because the main task of the algorithm in both cases is the computation of the eigenvalues of X_{n_α}, and in both cases n_α is approximately equal to $25 \cdot 2^{m+3}$. The bad behavior of Algorithm 1 when increasing α finds an explanation in the fact that, as observed in Example 1, the constant D_α appearing in the error estimate of Theorem 3 apparently grows very quickly with α.

Example 6 Let

$$\begin{aligned} u(\theta) &= 3 + 2\cos(\theta), \\ v(\theta) &= 2 - \cos(\theta) - \cos(2\theta). \end{aligned}$$

Note that $f(\theta) = v(\theta)/u(\theta) = 1 - \cos(\theta)$ is monotone increasing on $(0, \pi)$ and $f'(\theta) = 0$ only for $\theta = 0, \pi$. Suppose we want to approximate the eigenvalues of $X_n = T_n(u)^{-1}T_n(v)$ for $n = 5000$. Let $\tilde{\lambda}_j^{(m)}(X_n)$ be the approximation of $\lambda_j(X_n)$ obtained by applying Algorithm 1 with $n_1 = 50 \cdot 2^{m-1}$ and $\alpha = 4$. The graph of the errors $\varepsilon_{j,n}^{(m)}$ versus $\theta_{j,n}$ is shown in Fig. 7 for $j = 1, \ldots, n$ and $m = 1, 2, 3, 4$. Table 1 compares the CPU times for computing the eigenvalues of X_n by using MATLAB's `eig` function and Algorithm 1.

Example 7 This example is suggested by the cubic B-spline isogeometric analysis discretization of second-order eigenvalue problems [14, Section 10.7.3]. Let

$$\begin{aligned} u(\theta) &= 1208 + 1191\cos(\theta) + 120\cos(2\theta) + \cos(3\theta), \\ v(\theta) &= 40 - 15\cos(\theta) - 24\cos(2\theta) - \cos(3\theta). \end{aligned}$$

It can be shown that $u(\theta) > 0$ on $(0, \pi)$,

$$f(\theta) = \frac{v(\theta)}{u(\theta)} = \frac{40 - 15\cos(\theta) - 24\cos(2\theta) - \cos(3\theta)}{1208 + 1191\cos(\theta) + 120\cos(2\theta) + \cos(3\theta)}$$

is monotone increasing on $(0, \pi)$, and $f'(\theta) = 0$ only for $\theta = 0, \pi$. Suppose we want to approximate the eigenvalues of $X_n = T_n(u)^{-1}T_n(v)$ for $n = 5000$. Let $\tilde{\lambda}_j^{(m)}(X_n)$ be the approximation of $\lambda_j(X_n)$ obtained by applying Algorithm 1 with $n_1 = 50 \cdot 2^{m-1}$ and $\alpha = 4$. The graph of the errors $\varepsilon_{j,n}^{(m)}$ versus $\theta_{j,n}$ is shown in Fig. 8 for $j = 1, \ldots, n$ and $m = 1, 2, 3, 4$. The CPU times are reported in Table 2.

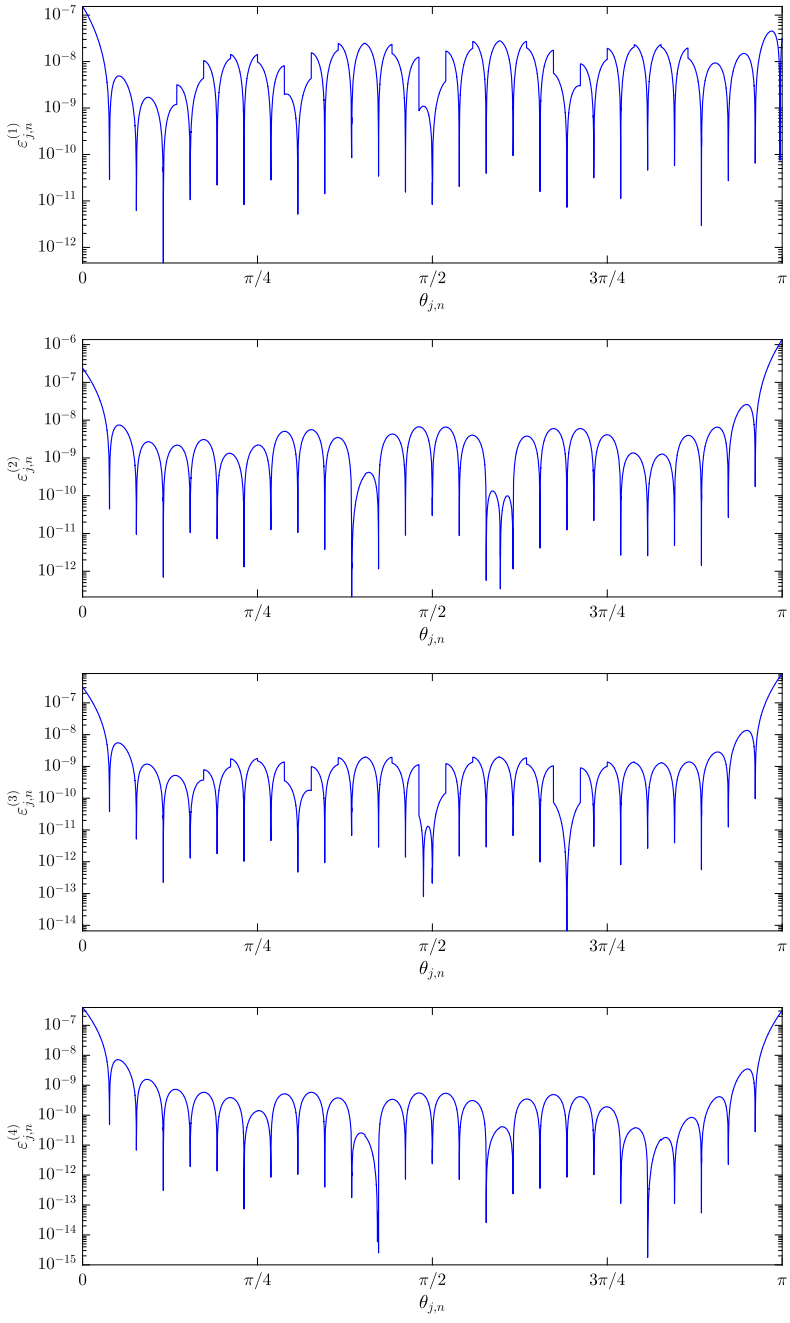


Fig. 6 Example 5: errors $\varepsilon_{j,n}^{(m)}$ versus $\theta_{j,n}$ for $j = 1, \ldots, n$, in the case where $u(\theta) = 1$, $v(\theta) = \frac{301}{400} - \cos(\theta) + \frac{1}{5}\cos(2\theta) + \frac{1}{10}\cos(3\theta) - \frac{1}{20}\cos(4\theta) + \frac{1}{400}\cos(6\theta)$, $n = 10000$, $n_1 = 25$, and $\alpha = 4 + m$

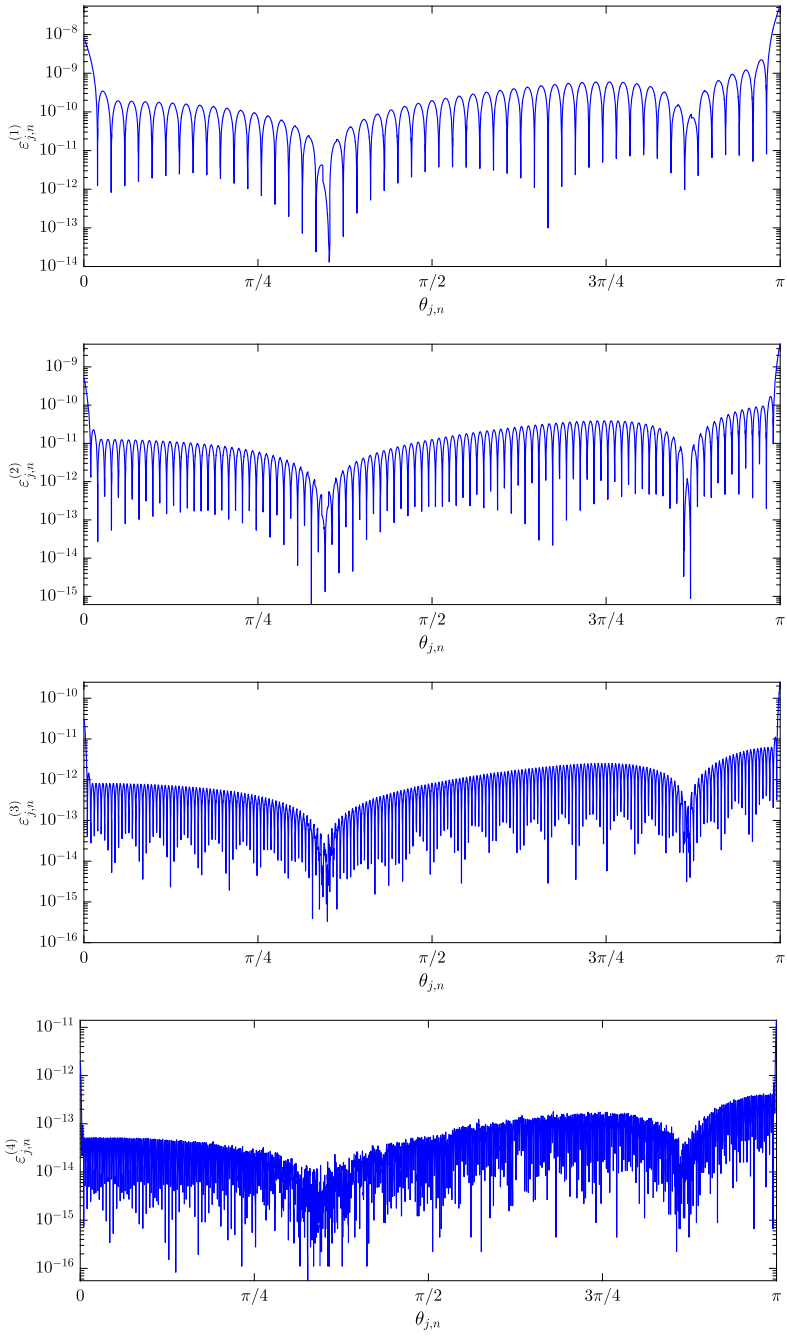


Fig. 7 Example 6: errors $\varepsilon_{j,n}^{(m)}$ versus $\theta_{j,n}$ for $j = 1, \ldots, n$, in the case where $u(\theta) = 3 + 2\cos(\theta)$, $v(\theta) = 2 - \cos(\theta) - \cos(2\theta)$, $n = 5000$, $n_1 = 50 \cdot 2^{m-1}$, and $\alpha = 4$

Table 1 Example 6 (Fig. 7): CPU times for computing the eigenvalues of X_n in the case where $u(\theta) = 3 + 2\cos(\theta)$, $v(\theta) = 2 - \cos(\theta) - \cos(2\theta)$, and $n = 5000$

Method	CPU time
Algorithm 1 with $n_1 = 50$ and $\alpha = 4$	1.81 s
Algorithm 1 with $n_1 = 100$ and $\alpha = 4$	7.14 s
Algorithm 1 with $n_1 = 200$ and $\alpha = 4$	32.45 s
Algorithm 1 with $n_1 = 400$ and $\alpha = 4$	144.08 s
MATLAB's `eig` function	694.76 s

Example 8 Let

$$\begin{aligned} u(\theta) &= 8 - 3\cos(\theta) - 4\cos(2\theta) - \cos(3\theta), \\ v(\theta) &= \frac{35}{2} - 12\cos(\theta) - 6\cos(2\theta) + \frac{1}{2}\cos(4\theta). \end{aligned}$$

It can be shown that $u(\theta) > 0$ on $(0, \pi)$,

$$f(\theta) = \frac{v(\theta)}{u(\theta)} = 2 - \cos(\theta)$$

is monotone increasing on $(0, \pi)$, and $f'(\theta) = 0$ only for $\theta = 0, \pi$. Suppose we want to approximate the smallest five eigenvalues of $X_n = T_n(u)^{-1}T_n(v)$ for $n = 5000$. Let $\tilde{\lambda}_j(X_n)$ be the approximations of $\lambda_j(X_n)$ obtained by applying Algorithm 1 with $n_1 = 100$ and $\alpha = 4$. Table 3 shows the errors $\varepsilon_{j,n}$ for $j = 1, \ldots, 5$, whereas Table 4 compares the CPU times for computing the eigenvalues of X_n by using Algorithm 1, MATLAB's `eig` function, and MATLAB's `eigs` function (applied to the generalized eigenvalue problem $T_n(v)\mathbf{x} = \lambda\, T_n(u)\mathbf{x}$ with $T_n(v)$ and $T_n(u)$ allocated as sparse matrices through MATLAB's `sparse` command).

4 Generalization to the non-monotone case

With reference to Conjecture 1, suppose that the function $f = v/u$ is monotone decreasing on $(0, \pi)$. Then, $-f = -v/u$ is monotone increasing on $(0, \pi)$ and, moreover, $T_n(u)^{-1}T_n(v) = -T_n(u)^{-1}T_n(-v)$. This immediately implies that Algorithm 1 allows one to compute the eigenvalues of $T_n(u)^{-1}T_n(v)$ even in the case where $f = v/u$ is monotone decreasing on $(0, \pi)$: it suffices to apply the algorithm with $X_n = T_n(u)^{-1}T_n(-v)$. Some limitations on the applicability of Algorithm 1 arise when f is non-monotone on $(0, \pi)$. This is precisely the case we are going to investigate in this section. We begin by formulating the following conjecture.

Conjecture 2 Let u, v be RCTPs, with $u > 0$ on $(0, \pi)$, and suppose that $f = v/u$ restricted to the interval $I \subseteq (0, \pi)$ is monotone and $f^{-1}(f(I)) = I$. Set $X_n = T_n(u)^{-1}T_n(v)$ for all n. Then, for every integer $\alpha \geq 0$, every n and every $j = 1, \ldots, n$ such that $\theta_{j,n} \in I$, the following asymptotic expansion holds:

$$\lambda_{\rho_n(j)}(X_n) = f(\theta_{j,n}) + \sum_{k=1}^{\alpha} c_k(\theta_{j,n})h^k + E_{j,n,\alpha}, \tag{7}$$

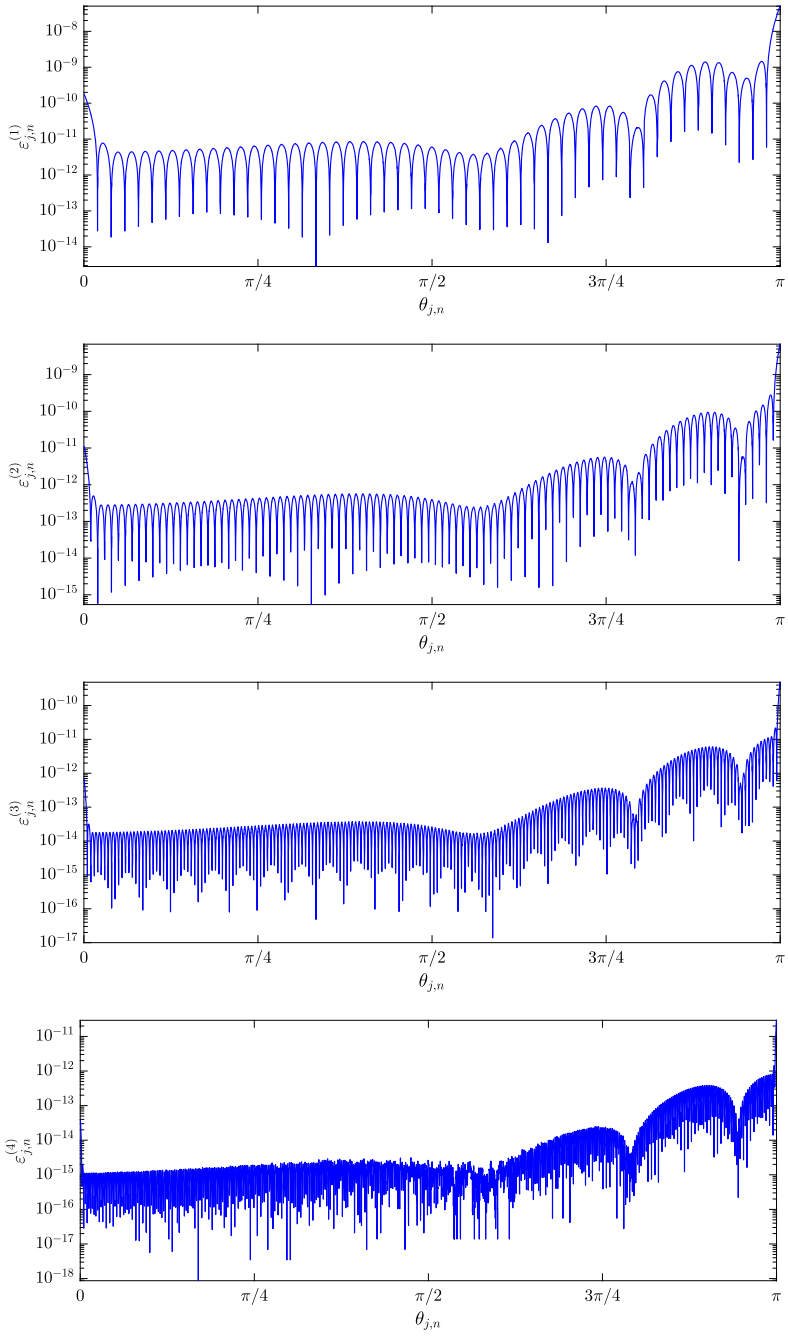


Fig. 8 Example 7: errors $\varepsilon_{j,n}^{(m)}$ versus $\theta_{j,n}$ for $j = 1, \ldots, n$, in the case where $u(\theta) = 1208 + 1191\cos(\theta) + 120\cos(2\theta) + \cos(3\theta)$, $v(\theta) = 40 - 15\cos(\theta) - 24\cos(2\theta) - \cos(3\theta)$, $n = 5000$, $n_1 = 50 \cdot 2^{m-1}$, and $\alpha = 4$

Table 2 Example 7 (Fig. 8): CPU times for computing the eigenvalues of X_n in the case where $u(\theta) = 1208 + 1191\cos(\theta) + 120\cos(2\theta) + \cos(3\theta)$, $v(\theta) = 40 - 15\cos(\theta) - 24\cos(2\theta) - \cos(3\theta)$, and $n = 5000$

Method	CPU time
Algorithm 1 with $n_1 = 50$ and $\alpha = 4$	1.69 s
Algorithm 1 with $n_1 = 100$ and $\alpha = 4$	2.77 s
Algorithm 1 with $n_1 = 200$ and $\alpha = 4$	18.30 s
Algorithm 1 with $n_1 = 400$ and $\alpha = 4$	280.27 s
MATLAB's `eig` function	1265.55 s

where:

- The eigenvalues of X_n are arranged in non-decreasing order, $\lambda_1(X_n) \leq \ldots \leq \lambda_n(X_n)$.
- $\rho_n = \sigma_n^{-1}$ is the inverse of σ_n, where σ_n is a permutation of $\{1, \ldots, n\}$ such that $f(\theta_{\sigma_n(1),n}) \leq \ldots \leq f(\theta_{\sigma_n(n),n})$.
- $\{c_k\}_{k=1,2,\ldots}$ is a sequence of functions from I to $\mathbb{R}$ which depends only on u, v.
- $h = \frac{1}{n+1}$ and $\theta_{j,n} = \frac{j\pi}{n+1} = j\pi h$.
- $E_{j,n,\alpha} = O(h^{\alpha+1})$ is the error, which satisfies the inequality $|E_{j,n,\alpha}| \leq C_\alpha h^{\alpha+1}$ for some constant C_α depending only on α, u, v.

Conjecture 2 is clearly an extension of Conjecture 1. Indeed, in the case where f is monotone increasing on $(0, \pi)$, if we take $I = (0, \pi)$ and we note that both σ_n and ρ_n reduce to the identity on $\{1, \ldots, n\}$, we see that Conjecture 2 reduces to Conjecture 1. Conjecture 2 is based on the numerical experiments carried out in [1, 13]. In the case where $u = 1$ identically, it was already formulated in [13]. In the case where $u = 1$ identically and $\alpha = 0$, it can be formally proved by adapting the argument used by Bogoya, Böttcher, Grudsky, and Maximenko in the proof of [7, Theorem 1.6].

In the situation described in Conjecture 2, we propose the following natural modification of Algorithm 1 for computing the eigenvalues of X_n corresponding to the the interval I (that is, the eigenvalues $\lambda_{\rho_n(j)}(X_n)$ corresponding to points $\theta_{j,n} \in I$). In what follows, for any integer n_1, we denote by $n_1(I)$ the cardinality of $\{\theta_{1,n_1}, \ldots, \theta_{n_1,n_1}\} \cap I$.

Algorithm 2 With the notation introduced in Conjecture 2, given two RCTPs u, v (with $u > 0$ on $(0, \pi)$ and $f = v/u$ such that f restricted to the interval $I \subseteq (0, \pi)$ is monotone and $f^{-1}(f(I)) = I$), three integers $n, n_1, \alpha \in \mathbb{N}$ with $n_1(I) \geq \alpha$ and

Table 3 Example 8: errors $\varepsilon_{j,n}$ for $j = 1, \ldots, 5$, in the case where $u(\theta) = 8 - 3\cos(\theta) - 4\cos(2\theta) - \cos(3\theta)$, $v(\theta) = \frac{35}{2} - 12\cos(\theta) - 6\cos(2\theta) + \frac{1}{2}\cos(4\theta)$, $n = 5000$, $n_1 = 100$, and $\alpha = 4$

j	1	2	3	4	5
$\varepsilon_{j,n}$	$1.56 \cdot 10^{-6}$	$1.42 \cdot 10^{-6}$	$1.47 \cdot 10^{-6}$	$1.34 \cdot 10^{-6}$	$1.39 \cdot 10^{-6}$

Table 4 Example 8: CPU times for computing the smallest five eigenvalues of X_n in the case where $u(\theta) = 8 - 3\cos(\theta) - 4\cos(2\theta) - \cos(3\theta)$, $v(\theta) = \frac{35}{2} - 12\cos(\theta) - 6\cos(2\theta) + \frac{1}{2}\cos(4\theta)$, and $n = 5000$

Method	CPU time
Algorithm 1 with $n_1 = 100$ and $\alpha = 4$	1.13 s
MATLAB's `eig` function	346.21 s
MATLAB's `eigs` function	Does not converge

$S \subseteq I$, we compute approximations of the eigenvalues $\{\lambda_{\rho_n(j)}(X_n) : \theta_{j,n} \in S\}$ as follows:

1. For $j_1 = 1, \ldots, n_1$ such that $\theta_{j_1,n_1} \in I$ compute $\tilde{c}_1(\theta_{j_1,n_1}), \ldots, \tilde{c}_\alpha(\theta_{j_1,n_1})$ by solving the linear system

$$\begin{cases} E_{j_1,n_1,0} = \tilde{c}_1(\theta_{j_1,n_1})h_1 + \tilde{c}_2(\theta_{j_1,n_1})h_1^2 + \ldots + \tilde{c}_\alpha(\theta_{j_1,n_1})h_1^\alpha \\ E_{j_2,n_2,0} = \tilde{c}_1(\theta_{j_1,n_1})h_2 + \tilde{c}_2(\theta_{j_1,n_1})h_2^2 + \ldots + \tilde{c}_\alpha(\theta_{j_1,n_1})h_2^\alpha \\ \quad\vdots \\ E_{j_\alpha,n_\alpha,0} = \tilde{c}_1(\theta_{j_1,n_1})h_\alpha + \tilde{c}_2(\theta_{j_1,n_1})h_\alpha^2 + \ldots + \tilde{c}_\alpha(\theta_{j_1,n_1})h_\alpha^\alpha \end{cases} \tag{8}$$

   where $n_k = 2^{k-1}(n_1 + 1) - 1$, $j_k = 2^{k-1} j_1$, and

$$E_{j_k,n_k,0} = \lambda_{\rho_{n_k}(j_k)}(X_{n_k}) - f(\theta_{j_1,n_1}), \qquad k = 1, \ldots, \alpha.$$

2. For $j = 1, \ldots, n$ such that $\theta_{j,n} \in S$
   - For $k = 1, \ldots, \alpha$
     - Determine $\alpha - k + 1$ points $\theta^{(1)}, \ldots, \theta^{(\alpha-k+1)} \in \{\theta_{1,n_1}, \ldots, \theta_{n_1,n_1}\} \cap I$ which are closest to $\theta_{j,n}$
     - Compute $\tilde{c}_{k,j}(\theta_{j,n})$, where $\tilde{c}_{k,j}(\theta)$ is the interpolation polynomial of $(\theta^{(1)}, \tilde{c}_k(\theta^{(1)})), \ldots, (\theta^{(\alpha-k+1)}, \tilde{c}_k(\theta^{(\alpha-k+1)}))$
   - Compute $\tilde{\lambda}_{\rho_n(j)}(X_n) = f(\theta_{j,n}) + \sum_{k=1}^{\alpha} \tilde{c}_{k,j}(\theta_{j,n})h^k$
3. Return $\{\tilde{\lambda}_{\rho_n(j)}(X_n) : \theta_{j,n} \in S\}$ as an approximation to $\{\lambda_{\rho_n(j)}(X_n) : \theta_{j,n} \in S\}$

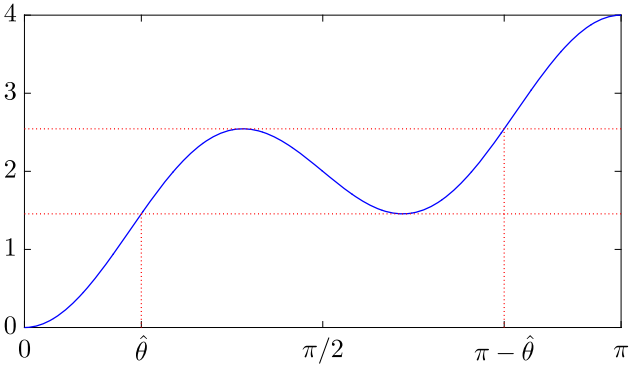


Fig. 9 Example 9: graph of $f(\theta) = v(\theta)/u(\theta) = 2 - \cos(\theta) - \cos(3\theta)$ over $(0, \pi)$

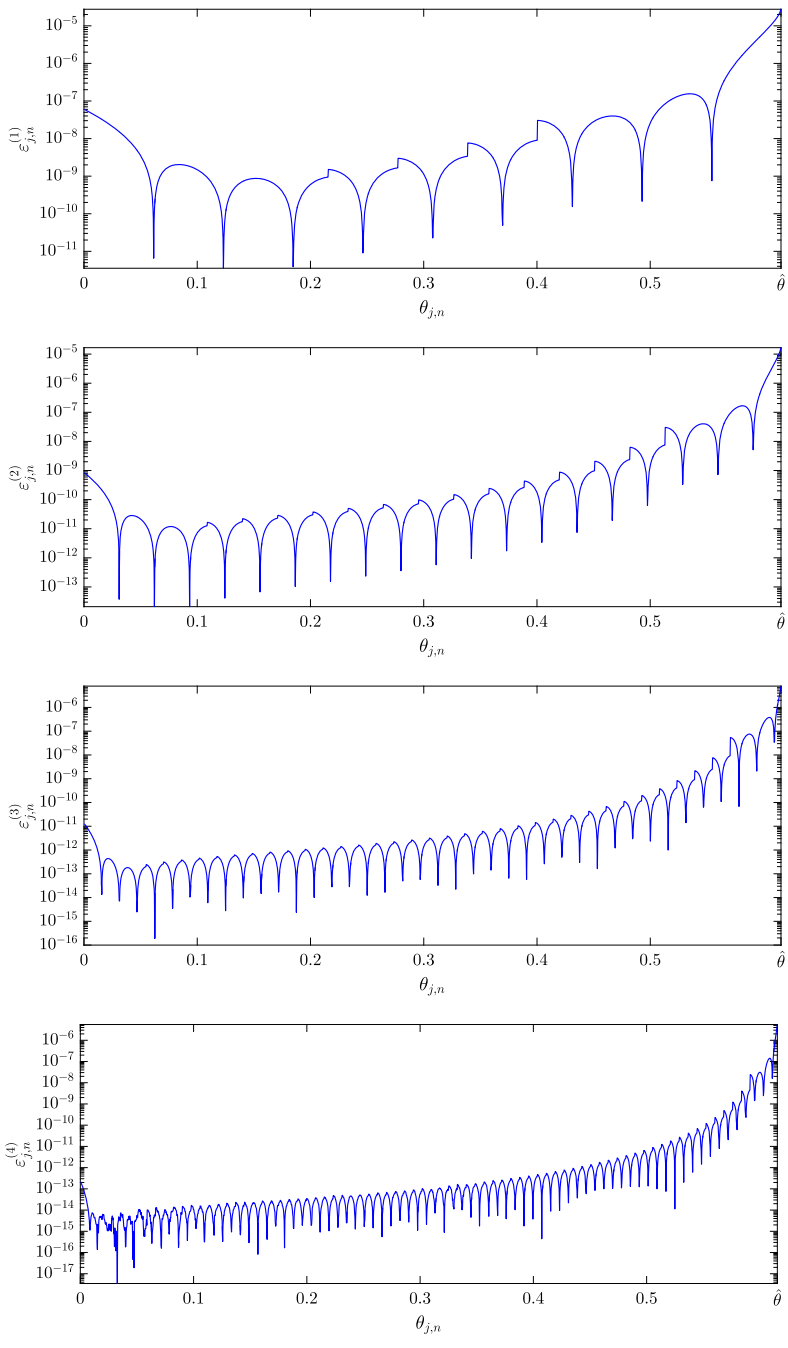


Fig. 10 Example 9: errors $\varepsilon_{j,n}^{(m)}$ versus $\theta_{j,n}$ for $\theta_{j,n} \in I = (0, \hat{\theta})$, in the case where $u(\theta) = 1$, $v(\theta) = 2 - \cos(\theta) - \cos(3\theta)$, $n = 10000$, $n_1 = 50 \cdot 2^{m-1}$, and $\alpha = 5$

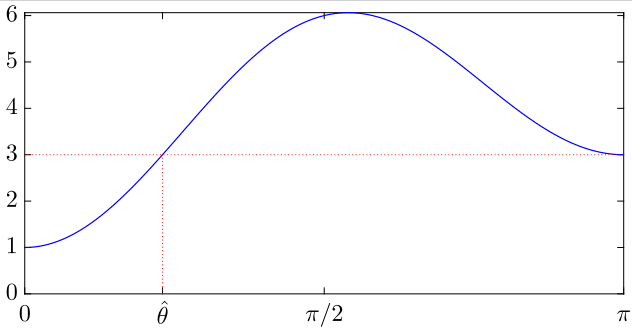


Fig. 11 Example 10: graph of $f(\theta) = v(\theta)/u(\theta) = 4 - \cos(\theta) - 2\cos(2\theta)$ over $(0, \pi)$

Example 9 Let

$$\begin{aligned} u(\theta) &= 1, \\ v(\theta) &= 2 - \cos(\theta) - \cos(3\theta). \end{aligned}$$

The graph of $f(\theta) = v(\theta)/u(\theta) = v(\theta)$ is depicted in Fig. 9. The hypotheses of Conjecture 2 are satisfied with either $I = (0, \hat{\theta})$ or $I = (\pi - \hat{\theta}, \pi)$, where $\hat{\theta} = 0.61547970867038...$ To fix the ideas, let $I = (0, \hat{\theta})$. Note that any permutation σ_n which sorts the samples $f(\theta_{1,n}), \ldots, f(\theta_{n,n})$ in non-decreasing order is such that $\sigma_n(j) = j$ whenever $\theta_{j,n} \in I$. As a consequence, $\rho_n(j) = j$ whenever $\theta_{j,n} \in I$. Set $X_n = T_n(u)^{-1}T_n(v) = T_n(f)$ and let $\{\tilde{\lambda}_j^{(m)}(X_n) : \theta_{j,n} \in I\}$ be the approximation of $\{\lambda_j(X_n) : \theta_{j,n} \in I\}$ obtained for $n = 10000$ by applying Algorithm 2 with $n_1 = 50 \cdot 2^{m-1}$, $\alpha = 5$, and $S = I$. The graph of the errors $\varepsilon_{j,n}^{(m)} = |\lambda_j(X_n) - \tilde{\lambda}_j^{(m)}(X_n)|$ versus $\theta_{j,n}$ is shown in Fig. 10 for $\theta_{j,n} \in I$ and $m = 1, 2, 3, 4$. We note that the error $\varepsilon_{j,n}^{(m)}$ tends to increase as $\theta_{j,n}$ moves toward $\hat{\theta}$, that is, as $\theta_{j,n}$ approaches to exit the interval I over which f satisfies the assumptions of Conjecture 2. Moreover, in a neighborhood of $\hat{\theta}$, the error decreases very slowly. This phenomenon is related to the fact that the expansion (7) does not hold in $[\hat{\theta}, \pi - \hat{\theta}]$ and, in fact, the errors $E_{j,n,0} = \lambda_{\rho_n(j)}(X_n) - f(\theta_{j,n})$ have a wild behavior inside this interval; see [13, Fig. 7].

Example 10 Let

$$\begin{aligned} u(\theta) &= 2 + \cos(3\theta), \\ v(\theta) &= 8 - 3\cos(\theta) - \frac{9}{2}\cos(2\theta) + 4\cos(3\theta) - \frac{1}{2}\cos(4\theta) - \cos(5\theta). \end{aligned}$$

The graph of $f(\theta) = v(\theta)/u(\theta) = 4 - \cos(\theta) - 2\cos(2\theta)$ is depicted in Fig. 11. The hypotheses of Conjecture 2 are satisfied with $I = (0, \hat{\theta})$, where $\hat{\theta} = 0.72273424781341...$ Any permutation σ_n which sorts the samples $f(\theta_{1,n}), \ldots, f(\theta_{n,n})$ in non-decreasing order is such that $\sigma_n(j) = j$ whenever $\theta_{j,n} \in I$. As a consequence, $\rho_n(j) = j$ whenever $\theta_{j,n} \in I$. Set $X_n = T_n(u)^{-1}T_n(v)$ and let $\{\tilde{\lambda}_j^{(m)}(X_n) : \theta_{j,n} \in I\}$ be the approximation of $\{\lambda_j(X_n) : \theta_{j,n} \in I\}$ obtained for $n = 5000$ by applying Algorithm 2 with $n_1 = 25 \cdot 2^{m-1}$, $\alpha = 5$, and $S = I$. The graph of the errors $\varepsilon_{j,n}^{(m)} = |\lambda_j(X_n) - \tilde{\lambda}_j^{(m)}(X_n)|$ versus $\theta_{j,n}$ is shown in Fig. 12 for

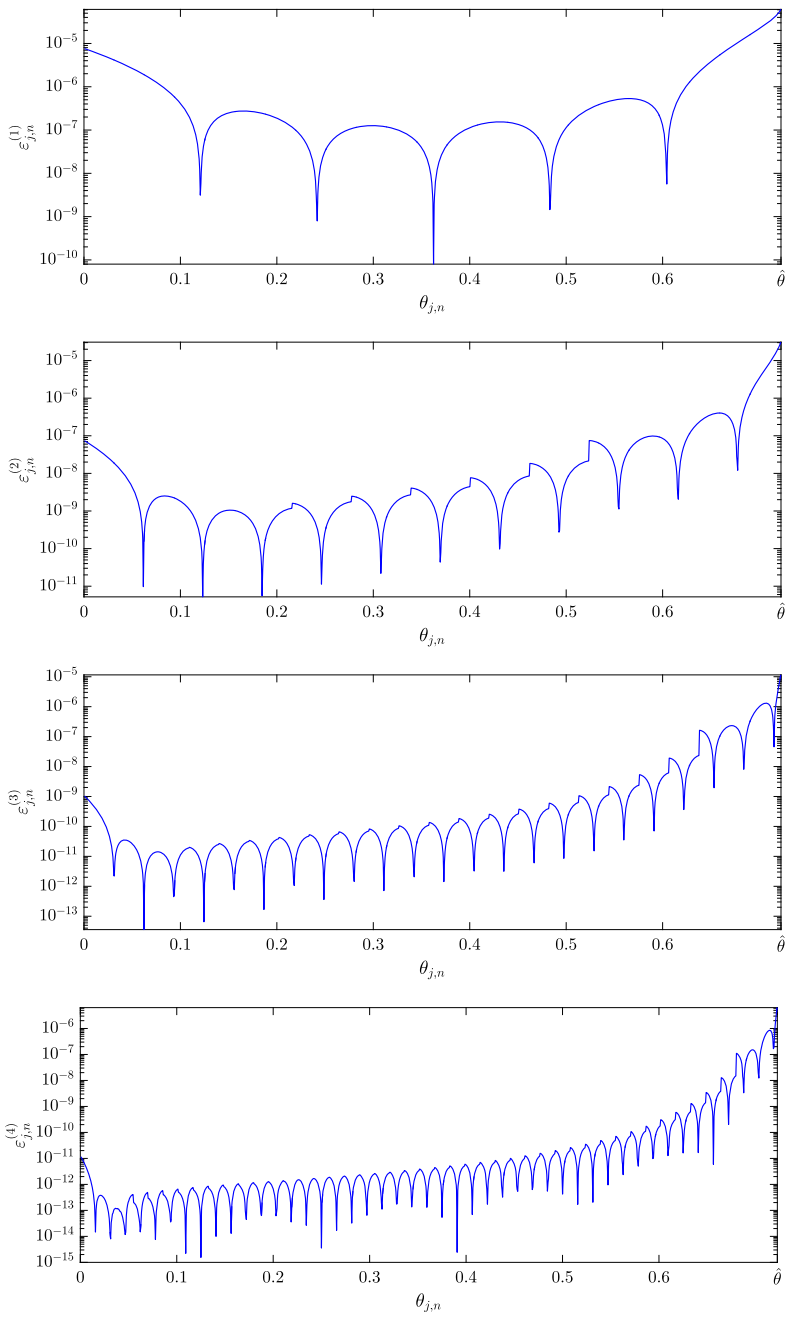


Fig. 12 Example 10: errors $\varepsilon_{j,n}^{(m)}$ versus $\theta_{j,n}$ for $\theta_{j,n} \in I = (0, \hat{\theta})$, in the case where $u(\theta) = 2 + \cos(3\theta)$, $v(\theta) = 8 - 3\cos(\theta) - \frac{9}{2}\cos(2\theta) + 4\cos(3\theta) - \frac{1}{2}\cos(4\theta) - \cos(5\theta)$, $n = 5000$, $n_1 = 25 \cdot 2^{m-1}$, and $\alpha = 5$

$\theta_{j,n} \in I$ and $m = 1, 2, 3, 4$. Considerations analogous to those in Example 10 apply also in this case.

5 Conclusions and perspectives

We have proposed and analyzed a matrix-less parallel interpolation–extrapolation algorithm for computing the eigenvalues of preconditioned banded symmetric Toeplitz matrices of the form $T_n(u)^{-1}T_n(v)$, where u, v are RCTPs, $u > 0$ on $(0, \pi)$, and $f = v/u$ is monotone on $(0, \pi)$. We have illustrated the performance of the algorithm through numerical experiments, and we have presented its generalization to the case where $f = v/u$ is non-monotone. We conclude by suggesting two possible future lines of research:

- Algorithm 1, as well as its generalized version for the non-monotone case (Algorithm 2), is based on a local interpolation strategy, as described in Section 2.1. An interesting topic for future research could be the following: try another kind of approximation (for example, an higher-order spline approximation) to see whether this reduces the errors and accelerates the convergence of both these algorithms.
- Understand whether an asymptotic eigenvalue expansion analogous to (7) holds without the hypothesis that f restricted to some interval $I \subseteq (0, \pi)$ is monotone and satisfies $f^{-1}(f(I)) = I$. Such a result would eliminate any limitation in the applicability of Algorithm 2 (provided that the latter is properly modified according to the new expansion).

Funding Information The research of Sven-Erik Ekström is cofinanced by the Graduate School in Mathematics and Computing (FMB) and Uppsala University. Carlo Garoni is a Marie-Curie fellow of the Italian INdAM (Istituto Nazionale di Alta Matematica) under grant agreement PCOFUND-GA-2012-600198.

Appendix A

This appendix collects the proofs of Theorems 1 and 2.

Proof of Theorem 1 We follow the argument in [1, Section 2]. Equations (2) and (4) can be rewritten as

$$A(h_1, \dots, h_1)\mathbf{c}(j_1) = \mathbf{E}_0(j_1) - \mathbf{E}_\alpha(j_1) \tag{9}$$

$$A(h_1, \dots, h_1)\tilde{\mathbf{c}}(j_1) = \mathbf{E}_0(j_1), \tag{10}$$

where

$$\mathbf{c}(j_1) = \begin{bmatrix} c_1(\theta_{j_1,n_1}) \\ \vdots \\ c_\alpha(\theta_{j_1,n_1}) \end{bmatrix}, \qquad \tilde{\mathbf{c}}(j_1) = \begin{bmatrix} \tilde{c}_1(\theta_{j_1,n_1}) \\ \vdots \\ \tilde{c}_\alpha(\theta_{j_1,n_1}) \end{bmatrix},$$
$$\mathbf{E}_0(j_1) = \begin{bmatrix} E_{j_1,n_1,0} \\ \vdots \\ E_{j_\alpha,n_\alpha,0} \end{bmatrix}, \qquad \mathbf{E}_\alpha(j_1) = \begin{bmatrix} E_{j_1,n_1,\alpha} \\ \vdots \\ E_{j_\alpha,n_\alpha,\alpha} \end{bmatrix}, \tag{11}$$

and

$$A(h_1,\ldots,h_\alpha) = \mathrm{diag}(h_1,\ldots,h_\alpha)\, V(h_1,\ldots,h_\alpha), \tag{12}$$

with $V(h_1,\ldots,h_\alpha)$ being the Vandermonde matrix associated with the nodes $h_1,\ldots,h_\alpha$,

$$V(h_1,\ldots,h_\alpha) = \begin{bmatrix} 1 & h_1 & h_1^2 & \cdots & h_1^{\alpha-1} \\ 1 & h_2 & h_2^2 & \cdots & h_2^{\alpha-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & h_\alpha & h_\alpha^2 & \cdots & h_\alpha^{\alpha-1} \end{bmatrix}.$$

By (9), (10), and (12), we have

$$\tilde{\mathbf{c}}(j_1) - \mathbf{c}(j_1) = A(h_1,\ldots,h_\alpha)^{-1}\mathbf{E}_\alpha(j_1) = V(h_1,\ldots,h_\alpha)^{-1}\mathbf{F}_\alpha(j_1),$$

where

$$\mathbf{F}_\alpha(j_1) = \mathrm{diag}(h_1,\ldots,h_\alpha)^{-1}\mathbf{E}_\alpha(j_1) = \begin{bmatrix} E_{j_1,n_1,\alpha}/h_1 \\ \vdots \\ E_{j_\alpha,n_\alpha,\alpha}/h_\alpha \end{bmatrix}.$$

Note that, by (3),

$$|(\mathbf{F}_\alpha(j_1))_k| = |E_{j_k,n_k,\alpha}/h_k| \le C_\alpha h_k^\alpha, \qquad k = 1,\ldots,\alpha. \tag{13}$$

The inverse of $V(h_1,\ldots,h_\alpha)$ is explicitly given by

$$(V(h_1,\ldots,h_\alpha)^{-1})_{ij} = \begin{cases} (-1)^{\alpha-i} \dfrac{\displaystyle\sum_{\substack{1\le k_1<\ldots<k_{\alpha-i}\le\alpha \\ k_1,\ldots,k_{\alpha-i}\neq j}} h_{k_1}\cdots h_{k_{\alpha-i}}}{\displaystyle\prod_{\substack{1\le k\le\alpha \\ k\neq j}} (h_j - h_k)}, & 1 \le i < \alpha, \\ \dfrac{1}{\displaystyle\prod_{\substack{1\le k\le\alpha \\ k\neq j}} (h_j - h_k)}, & i = \alpha. \end{cases} \tag{14}$$

Taking into account (13) and the equation $h_k = 2^{1-k}h_1$ for $k = 1, \ldots, \alpha$, we obtain the following:

- For $i = \alpha$,

$$
\begin{aligned}
&|\tilde{c}_\alpha(\theta_{j_1,n_1}) - c_\alpha(\theta_{j_1,n_1})| = |(\tilde{\mathbf{c}}(j_1) - \mathbf{c}(j_1))_\alpha| \\
&= \left| \sum_{j=1}^{\alpha} (V(h_1, \ldots, h_\alpha)^{-1})_{\alpha j} (\mathbf{F}_\alpha(j_1))_j \right| \\
&\le \sum_{j=1}^{\alpha} \frac{|(\mathbf{F}_\alpha(j_1))_j|}{\prod\limits_{\substack{1 \le k \le \alpha \\ k \ne j}} |h_j - h_k|} \le \sum_{j=1}^{\alpha} \frac{C_\alpha h_j^\alpha}{h_j^{\alpha-1} \prod\limits_{\substack{1 \le k \le \alpha \\ k \ne j}} |1 - h_k/h_j|} \\
&= C_\alpha h_1 \sum_{j=1}^{\alpha} \frac{2^{1-j}}{\prod\limits_{\substack{1 \le k \le \alpha \\ k \ne j}} |1 - 2^{j-k}|} = A(\alpha) h_1,
\end{aligned}
$$

  with $A(\alpha)$ depending only on α, u, v just like C_α.

- For $1 \le i < \alpha$,

$$
\begin{aligned}
&|\tilde{c}_i(\theta_{j_1,n_1}) - c_i(\theta_{j_1,n_1})| = |(\tilde{\mathbf{c}}(j_1) - \mathbf{c}(j_1))_i| \\
&= \left| \sum_{j=1}^{\alpha} (V(h_1, \ldots, h_\alpha)^{-1})_{ij} (\mathbf{F}_\alpha(j_1))_j \right| \\
&\le \sum_{j=1}^{\alpha} \frac{|(\mathbf{F}_\alpha(j_1))_j| \sum\limits_{\substack{1 \le k_1 < \ldots < k_{\alpha-i} \le \alpha \\ k_1, \ldots, k_{\alpha-i} \ne j}} h_{k_1} \cdots h_{k_{\alpha-i}}}{\prod\limits_{\substack{1 \le k \le \alpha \\ k \ne j}} |h_j - h_k|} \\
&\le \sum_{j=1}^{\alpha} \frac{C_\alpha h_j^\alpha \sum\limits_{\substack{1 \le k_1 < \ldots < k_{\alpha-i} \le \alpha \\ k_1, \ldots, k_{\alpha-i} \ne j}} h_{k_1} \cdots h_{k_{\alpha-i}}}{h_j^{\alpha-1} \prod\limits_{\substack{1 \le k \le \alpha \\ k \ne j}} |1 - h_k/h_j|} \\
&= C_\alpha h_1^{\alpha-i+1} \sum_{j=1}^{\alpha} \frac{2^{1-j} \sum\limits_{\substack{1 \le k_1 < \ldots < k_{\alpha-i} \le \alpha \\ k_1, \ldots, k_{\alpha-i} \ne j}} 2^{1-k_1} \cdots 2^{1-k_{\alpha-i}}}{\prod\limits_{\substack{1 \le k \le \alpha \\ k \ne j}} |1 - 2^{j-k}|} \\
&= A(\alpha, i) h_1^{\alpha-i+1},
\end{aligned}
$$

  with $A(\alpha, i)$ depending only on α, i, u, v.

In conclusion, Theorem 1 is proved with $A_\alpha = \max_{i=1,\dots,\alpha} A(\alpha, i)$, where $A(\alpha, \alpha) = A(\alpha)$. □

Proof of Theorem 2 Let $L_1, \dots, L_{\alpha-k+1}$ be the Lagrange polynomials associated with the nodes $\theta^{(1)}, \dots, \theta^{(\alpha-k+1)}$,

$$L_r(\theta) = \prod_{\substack{s=1 \\ s\neq r}}^{\alpha-k+1} \frac{\theta - \theta^{(s)}}{\theta^{(r)} - \theta^{(s)}}, \qquad r = 1, \dots, \alpha - k + 1.$$

The interpolation polynomial of the data $(\theta^{(1)}, \tilde{c}_k(\theta^{(1)})), \dots, (\theta^{(\alpha-k+1)}, \tilde{c}_k(\theta^{(\alpha-k+1)}))$ is

$$\tilde{c}_{k,j}(\theta) = \sum_{r=1}^{\alpha-k+1} \tilde{c}_k(\theta^{(r)}) L_r(\theta)$$

and the interpolation polynomial of the data $(\theta^{(1)}, c_k(\theta^{(1)})), \dots, (\theta^{(\alpha-k+1)}, c_k(\theta^{(\alpha-k+1)}))$ is

$$p(\theta) = \sum_{r=1}^{\alpha-k+1} c_k(\theta^{(r)}) L_r(\theta).$$

Considering that $\theta^{(1)}, \dots, \theta^{(\alpha-k+1)}$ are $\alpha - k + 1$ points from $\{\theta_{1,n_1}, \dots, \theta_{n_1,n_1}\}$ which are closest to $\theta_{j,n}$, the length of the smallest interval I containing the nodes $\theta^{(1)}, \dots, \theta^{(\alpha-k+1)}$ and the point $\theta_{j,n}$ is bounded by $(\alpha - k + 1)\pi h_1$. Hence, by Theorem 1, for all $\theta \in I$ we have

$$\begin{aligned} |\tilde{c}_{k,j}(\theta) - p(\theta)| &\le \sum_{r=1}^{\alpha-k+1} |\tilde{c}_{k,j}(\theta^{(r)}) - c_k(\theta^{(r)})| \prod_{\substack{s=1 \\ s\neq r}}^{\alpha-k+1} \frac{|\theta - \theta^{(s)}|}{|\theta^{(r)} - \theta^{(s)}|} \\ &\le \sum_{r=1}^{\alpha-k+1} A_\alpha h_1^{\alpha-k+1} \prod_{\substack{s=1 \\ s\neq r}}^{\alpha-k+1} \frac{(\alpha - k + 1)\pi h_1}{\pi h_1} \\ &= A_\alpha h_1^{\alpha-k+1} (\alpha - k + 1)^{\alpha-k+1}. \end{aligned} \tag{15}$$

Since $c_k \in C^{\alpha-k+1}([0, \pi])$ by assumption, from interpolation theory we know that for every $\theta \in I$ there exists $\xi(\theta) \in I$ such that

$$c_k(\theta) - p(\theta) = \frac{c_k^{(\alpha-k+1)}(\xi(\theta))}{(\alpha - k + 1)!} \prod_{r=1}^{\alpha-k+1} (\theta - \theta^{(r)});$$

see, e.g., [12, Theorem 3.1.1]. Thus, for all $\theta \in I$, we have

$$\begin{aligned} |c_k(\theta) - p(\theta)| &\le \frac{|c_k^{(\alpha-k+1)}(\xi(\theta))|}{(\alpha-k+1)!} \prod_{r=1}^{\alpha-k+1} |\theta - \theta^{(r)}| \\ &\le \frac{\|c_k^{(\alpha-k+1)}\|_\infty}{(\alpha-k+1)!} \prod_{r=1}^{\alpha-k+1} (\alpha-k+1)\pi h_1 \\ &= \frac{(\alpha-k+1)^{\alpha-k+1}\pi^{\alpha-k+1}\|c_k^{(\alpha-k+1)}\|_\infty}{(\alpha-k+1)!} h_1^{\alpha-k+1}. \end{aligned} \tag{16}$$

From (15) and (16) we obtain

$$|c_k(\theta) - \tilde{c}_{k,j}(\theta)| \le B(k,\alpha) h_1^{\alpha-k+1} \le B_\alpha h_1^{\alpha-k+1}, \qquad \theta \in I, \tag{17}$$

where

$$B(k,\alpha) = \frac{(\alpha-k+1)^{\alpha-k+1}\pi^{\alpha-k+1}\|c_k^{(\alpha-k+1)}\|_\infty}{(\alpha-k+1)!} + A_\alpha(\alpha-k+1)^{\alpha-k+1}$$

and $B_\alpha = \max_{i=1,\ldots,\alpha} B(i,\alpha)$. Since $\theta_{j,n} \in I$, it is clear that (6) follows from (17). □

Appendix B

This appendix provides a plain MATLAB implementation of Algorithm 1.

```
function lambdaS = eigs_preconditioned_toeplitz(n,cu,cv,n1,alpha,S)
% INPUT
%        n: positive integer (size of X_n = T_n(u)^(-1) * T_n(v))
%       cu: row vector of the coefficients of the trigonometric polynomial
%           u(t) = cu(1)+2*cu(2)*cos(t)+...+2*cu(end)*cos((end-1)*t)
%       cv: row vector of the coefficients of the trigonometric polynomial
%           v(t) = cv(1)+2*cv(2)*cos(t)+...+2*cv(end)*cos((end-1)*t)
%       n1: positive integer (number of points of the coarsest grid
%           theta_{j1,n1} = j1*pi/(n1+1), j1=1,...,n1)
%    alpha: positive integer (number of coefficients c_k(theta)
%           to be approximated on the coarsest grid by the tilde c_k(theta))
%        S: row vector containing the indices corresponding to the
%           eigenvalues of X_n to be computed; the indices should be sorted
%           in increasing order, and it is understood that the eigenvalues
%           of X_n are sorted in increasing order as well
% OUTPUT
%  lambdaS: row vector of length length(S) containing the approximations
%           of the eigenvalues of X_n corresponding to the indices S
%           computed by using Algorithm 1 with n1 and alpha as inputs
% FURTHER SPECIFICATIONS
%   This Matlab function works under the same assumptions as in this paper,
%   i.e., u(t), v(t), f(t)=v(t)/u(t) should be as in Conjecture 1 and n1
%   should be greater or equal to alpha
% EXAMPLE (CORRESPONDING TO EXAMPLE 8 OF THIS PAPER)
%   n = 5000; cu = [8, -1.5, -2, -0.5]; cv = [17.5, -6, -3, 0, 0.25];
%   n1 = 100; alpha = 4; S = 1:5;
%   lambdaS = eigs_preconditioned_toeplitz(n,cu,cv,n1,alpha,S)

lu = length(cu); lv = length(cv);
```

```
u = @(t)cu(1)+sum(2*cu(2:lu).*cos((1:lu-1)*t));
v = @(t)cv(1)+sum(2*cv(2:lv).*cos((1:lv-1)*t));
f = @(t) arrayfun(@(t)v(t)./u(t),t);

nn = zeros(1,alpha); hh = zeros(1,alpha);
for k = 1:alpha
    nn(k) = 2^(k-1)*(n1+1)-1;
    hh(k) = 1/(nn(k)+1);
end

A = zeros(alpha);
for i = 1:alpha
    for j = 1:alpha
        A(i,j) = hh(i)^j;
    end
end

E = zeros(alpha,n1);
j1 = 1:n1;
theta = j1*pi*hh(1);
TTu = toeplitz( [cu, sparse(1, nn(alpha) - lu)] );
TTv = toeplitz( [cv, sparse(1, nn(alpha) - lv)] );
for k = 1:alpha
    eigX = sort(eig(full(TTv(1:nn(k),1:nn(k))),full(TTu(1:nn(k),1:nn(k)))));
    jk = 2^(k-1)*j1;
    E(k,:) = eigX(jk)' - f(theta);
end

c_tilde = A\E;

lS = length(S);
lambdaS = zeros(1,lS);
h = 1/(n+1);
t = S*pi*h;
for j = 1:lS
    ell = t(j)*(n1+1)/pi;
    poly_evals = zeros(1,alpha);
    for k = 1:alpha
        indices = localization(ell,alpha-k+1);
        if indices(1)<1
           indices = indices - indices(1) + 1;
        end
        if indices(end)>n1
            indices = indices - indices(end) + n1;
        end
        tt = indices*pi*hh(1);
        poly_evals(k) = polyval(polyfit(tt,c_tilde(k,indices),alpha-k),t(j));
    end
    lambdaS(j) = polyval([poly_evals(end:-1:1) f(t(j))],h);
end

end

function u = localization(x,m)

% INPUT
%        x: real number
%        m: natural number >= 1
% OUTPUT
%        u: row vector of length m such that u(1),...,u(m) are m integers
%           that are closest to x (which are not uniquely determined
%           in some cases)
```

```
b = mod(m,2);
v = (m + b)/2;
fx = floor(x);
cx = ceil(x);

if x - fx <= cx - x
   u = (fx - v + 1):(fx + v - b);
else
   u = (cx - v + b):(cx + v - 1);
end

end
```

References

1. Ahmad, F., Al-Aidarous, E.S., Alrehaili, D.A., Ekström, S.-E., Furci, I., Serra-Capizzano, S.: Are the eigenvalues of preconditioned banded symmetric Toeplitz matrices known in almost closed form? Numer. Alg. (in press). https://doi.org/10.1007/s11075-017-0404-z
2. Arbenz, P.: Computing the eigenvalues of banded symmetric Toeplitz matrices. SIAM J. Sci. Stat. Comput. **12**, 743–754 (1991)
3. Badía, J.M., Vidal, A.M.: Parallel algorithms to compute the eigenvalues and eigenvectors of symmetric Toeplitz matrices. Parallel Algorithms Appl. **13**, 75–93 (2000)
4. Bini, D., Di Benedetto, F.: Solving the generalized eigenvalue problem for rational Toeplitz matrices. SIAM J. Matrix Anal. Appl. **11**, 537–552 (1990)
5. Bini, D., Pan, V.: Efficient algorithms for the evaluation of the eigenvalues of (block) banded Toeplitz matrices. Math. Comput. **50**, 431–448 (1988)
6. Bogoya, J.M., Böttcher, A., Grudsky, S.M., Maximenko, E.A.: Eigenvalues of Hermitian Toeplitz matrices with smooth simple-loop symbols. J. Math. Anal. Appl. **422**, 1308–1334 (2015)
7. Bogoya, J.M., Böttcher, A., Grudsky, S.M., Maximenko, E.A.: Maximum norm versions of the Szegő and Avram–Parter theorems for Toeplitz matrices. J. Approx. Theory **196**, 79–100 (2015)
8. Bogoya, J.M., Grudsky, S.M., Maximenko, E.A.: Eigenvalues of Hermitian Toeplitz matrices generated by simple-loop symbols with relaxed smoothness. Oper. Theory Adv. Appl. **259**, 179–212 (2017)
9. Böttcher, A., Silbermann, B.: Introduction to Large Truncated Toeplitz Matrices. Springer, New York (1999)
10. Böttcher, A., Grudsky, S.M., Maximenko, E.A.: Inside the eigenvalues of certain Hermitian Toeplitz band matrices. J. Comput. Appl. Math. **233**, 2245–2264 (2010)
11. Brezinski, C., Redivo Zaglia, M.: Extrapolation Methods: Theory and Practice. North-Holland, Elsevier Science Publishers B.V., Amsterdam (1991)
12. Davis, P.J.: Interpolation and Approximation. Dover, New York (1975)
13. Ekström, S.-E., Garoni, C., Serra-Capizzano, S.: Are the eigenvalues of banded symmetric Toeplitz matrices known in almost closed form? Exper. Math. (in press). https://doi.org/10.1080/10586458.2017.1320241
14. Garoni, C., Serra-Capizzano, S.: Generalized Locally Toeplitz Sequences: Theory and Applications, vol. I. Springer, Cham (2017)
15. Stoer, J., Bulirsch, R.: Introduction to Numerical Analysis, 3rd edn. Springer, New York (2010)
16. Trench, W.F.: On the eigenvalue problem for Toeplitz band matrices. Linear Algebra Appl. **64**, 199–214 (1985)
17. Trench, W.F.: Characteristic polynomials of symmetric rationally generated Toeplitz matrices. Linear Multilinear Algebra **21**, 289–296 (1987)
18. Trench, W.F.: Numerical solution of the eigenvalue problem for symmetric rationally generated Toeplitz matrices. SIAM J. Matrix Anal. Appl. **9**, 291–303 (1988)
19. Trench, W.F.: Numerical solution of the eigenvalue problem for Hermitian Toeplitz matrices. SIAM J. Matrix Anal. Appl. **10**, 135–146 (1989)
20. Trench, W.F.: Numerical solution of the eigenvalue problem for efficiently structured Hermitian matrices. Linear Algebra Appl. **154–156**, 415–432 (1991)

Paper IV

Are the eigenvalues of the B-spline IgA approximation of $-\Delta u = \lambda u$ known in almost closed form?

Sven-Erik Ekström

Uppsala University, Department of Information Technology, Division of Scientific Computing, ITC, Lägerhyddsv. 2, Hus 2, P.O. Box 337, SE-751 05 Uppsala, Sweden. Email: sven-erik.ekstrom@it.uu.se.

Isabella Furci

University of Insubria, Department of Science and High Technology, Via Valleggio 11, 22100 Como, Italy. Email: ifurci@uninsubria.it.

Carlo Garoni

University of Italian Switzerland (USI), Institute of Computational Science, Via Giuseppe Buffi 13, 6900 Lugano, Switzerland. Email: carlo.garoni@usi.ch.

University of Insubria, Department of Science and High Technology, Via Valleggio 11, 22100 Como, Italy. Email: carlo.garoni@uninsubria.it.

Stefano Serra-Capizzano

University of Insubria, Department of Science and High Technology, Via Valleggio 11, 22100 Como, Italy. Email: stefano.serrac@uninsubria.it.

Uppsala University, Department of Information Technology, Division of Scientific Computing, ITC, Lägerhyddsv. 2, Hus 2, P.O. Box 337, SE-751 05 Uppsala, Sweden. Email: stefano.serra@it.uu.se.

October 2, 2017

Abstract

We consider the B-spline Isogeometric Analysis (IgA) approximation of the Laplacian eigenvalue problem $-\Delta u = \lambda u$ over the d-dimensional hypercube $(0,1)^d$. By using tensor-product arguments, we show that the eigenvalue–eigenvector structure of the resulting discretization matrix is completely determined by the eigenvalue–eigenvector structure of the matrix $L_n^{[p]}$ arising from the IgA approximation based on B-splines of degree p of the unidimensional problem $-u'' = \lambda u$. Here, n is the mesh fineness parameter and the size of $L_n^{[p]}$ is $N(n,p) = n+p-2$. In previous works, it was established that the normalized sequence $\{n^{-2}L_n^{[p]}\}_n$ enjoys an asymptotic spectral distribution described by a function $e_p(\theta)$, the so-called spectral symbol. The contributions of this paper can be summarized as follows.

1. For $p = 1$ and $p = 2$ we show that $L_n^{[p]}$ belongs to a matrix algebra associated with a fast unitary sine transform, and we compute eigenvalues and eigenvectors of $L_n^{[p]}$. In both cases, the eigenvalues are given by $e_p(\theta_{j,n})$, $j = 1, \ldots, n+p-2$, where $\theta_{j,n} = j\pi/n$.
2. For $p \geq 3$, we provide numerical evidence of a precise asymptotic expansion for the eigenvalues of $n^{-2}L_n^{[p]}$, excluding the largest $n_p^{\text{out}} = p-2+\text{mod}(p,2)$ eigenvalues (the so-called outliers). More precisely, we numerically show that for every $p \geq 3$, every integer $\alpha \geq 0$, every n, and every $j = 1, \ldots, N(n,p) - n_p^{\text{out}}$,

$$\lambda_j(n^{-2}L_n^{[p]}) = e_p(\theta_{j,n}) + \sum_{k=1}^{\alpha} c_k^{[p]}(\theta_{j,n})h^k + E_{j,n,\alpha}^{[p]},$$

where:

- the eigenvalues of $n^{-2}L_n^{[p]}$ are arranged in ascending order, $\lambda_1(n^{-2}L_n^{[p]}) \leq \ldots \leq \lambda_{n+p-2}(n^{-2}L_n^{[p]})$;
- $\{c_k^{[p]}\}_{k=1,2,\ldots}$ is a sequence of functions from $[0,\pi]$ to $\mathbb{R}$ which depends only on p;
- $h = 1/n$ and $\theta_{j,n} = j\pi h$ for $j = 1, \ldots, n$;
- $E_{j,n,\alpha}^{[p]} = O(h^{\alpha+1})$ is the remainder, which satisfies $|E_{j,n,\alpha}^{[p]}| \leq C_\alpha^{[p]} h^{\alpha+1}$ for some constant $C_\alpha^{[p]}$ depending only on α and p.

We also provide a proof of this expansion for $\alpha = 0$ and $j = 1, \ldots, N(n,p) - (4p-2)$, where $4p-2$ represents a theoretical estimate of the number of outliers n_p^{out}.

3. We show through numerical experiments that, for $p \geq 3$ and $k \geq 1$, there exists a point $\theta(p,k) \in (0,\pi)$ such that $c_k^{[p]}(\theta)$ vanishes on $[0,\theta(p,k)]$. Moreover, as it is suggested by the numerics of this paper, the infimum of $\theta(p,k)$ over all $k \geq 1$, say y_p, is strictly positive, and the equation $\lambda_j(n^{-2}L_n^{[p]}) = e_p(\theta_{j,n})$ holds numerically whenever $\theta_{j,n} < \theta(p)$, where $\theta(p)$ is a point in $(0, y_p]$ which grows with p.
4. For $p \geq 3$, based on the asymptotic expansion in the above item 2, we propose an interpolation–extrapolation algorithm for computing the eigenvalues of $L_n^{[p]}$, excluding the n_p^{out} outliers. The performance of the algorithm is illustrated through numerical experiments. Note that, by the previous item 3, the algorithm is actually not necessary for computing the eigenvalues corresponding to points $\theta_{j,n} < \theta(p)$.

Keywords: Laplacian eigenvalue problem, isogeometric analysis, B-splines, mass and stiffness matrices, eigenvalues and eigenvectors, asymptotic eigenvalue expansion, polynomial interpolation, extrapolation

2010 MSC: 65N25, 65N30, 41A15, 65F15, 65D05, 65B05

1 Introduction

1.1 Problem setting

Consider the one-dimensional Laplacian eigenvalue problem

$$\begin{cases} -u''(x) = \lambda u(x), & x \in (0,1), \\ u(0) = u(1) = 0. \end{cases} \tag{1}$$

The corresponding weak formulation reads as follows: find eigenvalues $\lambda \in \mathbb{R}^+$ and eigenfunctions $u \in H_0^1(0,1)$ such that, for all $v \in H_0^1(0,1)$,

$$\mathrm{a}(u,v) = \lambda(u,v),$$

where

$$\mathrm{a}(u,v) = \int_0^1 u'(x)v'(x)\mathrm{d}x, \qquad (u,v) = \int_0^1 u(x)v(x)\mathrm{d}x.$$

In the Galerkin method, we choose a finite-dimensional vector space $\mathscr{W} \subset H_0^1(0,1)$, we set $N = \dim \mathscr{W}$, and we look for approximations of the exact eigenpairs

$$\lambda_j = j^2\pi^2, \qquad u_j(x) = \sin(j\pi x), \qquad j \geq 1, \tag{2}$$

by solving the following Galerkin problem: find $\lambda_{j,\mathscr{W}} \in \mathbb{R}^+$ and $u_{j,\mathscr{W}} \in \mathscr{W}$, for $j = 1,\dots,N$, such that, for all $v \in \mathscr{W}$,

$$\mathrm{a}(u_{j,\mathscr{W}}, v) = \lambda_{j,\mathscr{W}}(u_{j,\mathscr{W}}, v). \tag{3}$$

Assuming the numerical eigenvalues $\lambda_{j,\mathscr{W}}$ are arranged in ascending order, the pair $(\lambda_{j,\mathscr{W}}, u_{j,\mathscr{W}})$ is taken as an approximation of the pair (λ_j, u_j) for all $j = 1,\dots,N$. The numbers $\lambda_{j,\mathscr{W}}/\lambda_j - 1$, $j = 1,\dots,N$, are referred to as the (relative) eigenvalue errors. If $\{\varphi_1,\dots,\varphi_N\}$ is a basis of $\mathscr{W}$, in view of the canonical identification between each $v \in \mathscr{W}$ and its coefficient vector with respect to $\{\varphi_1,\dots,\varphi_N\}$, solving the Galerkin problem (3) is equivalent to solving the generalized eigenvalue problem

$$K\mathbf{u}_{j,\mathscr{W}} = \lambda_{j,\mathscr{W}} M\mathbf{u}_{j,\mathscr{W}}, \tag{4}$$

where $\mathbf{u}_{j,\mathscr{W}}$ is the coefficient vector of $u_{j,\mathscr{W}}$ with respect to $\{\varphi_1,\dots,\varphi_N\}$ and

$$K = [\mathrm{a}(\varphi_j,\varphi_i)]_{i,j=1}^N = \left[\int_0^1 \varphi_j'(x)\varphi_i'(x)\mathrm{d}x\right]_{i,j=1}^N, \tag{5}$$

$$M = [(\varphi_j,\varphi_i)]_{i,j=1}^N = \left[\int_0^1 \varphi_j(x)\varphi_i(x)\mathrm{d}x\right]_{i,j=1}^N. \tag{6}$$

The matrices K and M are referred to as the stiffness matrix and the mass matrix, respectively. Both K and M are always symmetric positive definite, regardless of the chosen basis functions $\varphi_1,\dots,\varphi_N$. Moreover, it is clear from (4) that the numerical eigenvalues $\lambda_{j,\mathscr{W}}$, $j = 1,\dots,N$, are just the eigenvalues of the matrix

$$L = M^{-1}K. \tag{7}$$

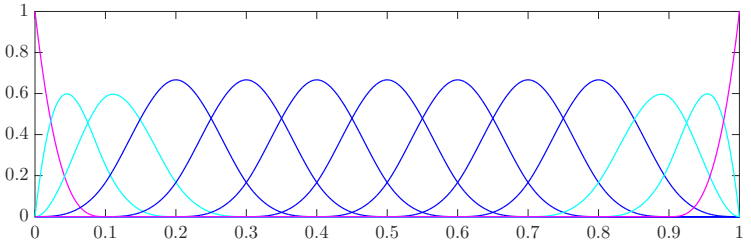


Figure 1: Cubic B-splines $\{N_{1,[3]},\dots,N_{n+3,[3]}\}$ for the knot sequence $\big\{0,0,0,0,\frac{1}{n},\frac{2}{n},\dots,\frac{n-1}{n},1,1,1,1\big\}$ $(n=10)$.

Now, for $p,n\geq 1$ let

$$N_{i,[p]},\qquad i=1,\dots,n+p, \tag{8}$$

be the B-splines of degree $p\geq 1$ and smoothness $C^{p-1}(\mathbb{R})$ defined over the knot sequence

$$\underbrace{0,\dots,0}_{p+1},\frac{1}{n},\frac{2}{n},\dots,\frac{n-1}{n},\underbrace{1,\dots,1}_{p+1}.$$

The B-splines (8) form a basis for the spline space

$$\mathscr{V}_{n,[p]}=\big\{v\in C^{p-1}[0,1]:\ v_{|\left[\frac{i}{n},\frac{i+1}{n}\right)}\in\mathbb{P}_p \text{ for } i=0,\dots,n-1\big\},$$

where $\mathbb{P}_p$ is the space of polynomials of degree at most p. Except for the first and the last one, all the other B-splines vanish on the boundary of $[0,1]$. In particular, the B-splines

$$N_{i+1,[p]},\qquad i=1,\dots,n+p-2, \tag{9}$$

form a basis for the space

$$\mathscr{W}_{n,[p]}=\big\{v\in\mathscr{V}_{n,[p]}:\ v(0)=v(1)=0\big\}.$$

We refer the reader to Figure 1 for the graphs of the B-splines (8) corresponding to the degree $p=3$. For more on B-splines, including the precise definition of the functions (8), see [11, 19].

In the Isogeometric Analysis (IgA) approximation of (1) based on uniform B-splines of degree $p\geq 1$, we look for approximations of the exact eigenpairs (2) by using the Galerkin method described above, in which the basis functions $\varphi_1,\dots,\varphi_N$ are chosen as the B-splines $N_{2,[p]},\dots,N_{n+p-1,[p]}$ and, consequently, the vector space $\mathscr{W}$ is equal to $\mathscr{W}_{n,[p]}$. The resulting stiffness and mass matrices (5)–(6) are given by

$$K_n^{[p]}=\left[\int_0^1 N'_{j+1,[p]}(x)N'_{i+1,[p]}(x)\mathrm{d}x\right]_{i,j=1}^{n+p-2}, \tag{10}$$

$$M_n^{[p]}=\left[\int_0^1 N_{j+1,[p]}(x)N_{i+1,[p]}(x)\mathrm{d}x\right]_{i,j=1}^{n+p-2}, \tag{11}$$

and the numerical eigenvalues $\lambda_{j,n}^{[p]}$, $j=1,\dots,n+p-2$, are the eigenvalues of the matrix

$$L_n^{[p]}=(M_n^{[p]})^{-1}K_n^{[p]}. \tag{12}$$

For more details on IgA, we refer the reader to [10].

Let $\phi_{[q]}$ be the B-spline of degree $q\geq 0$ corresponding to the knot sequence $\{0,1,\dots,q+1\}$. The function $\phi_{[q]}$ is usually referred to as the cardinal B-spline of degree q and it is recursively defined as follows [11]:

$$\begin{aligned}\phi_{[0]}(t)&=\chi_{[0,1)}(t),\qquad t\in\mathbb{R},\\ \phi_{[q]}(t)&=\frac{t}{q}\phi_{[q-1]}(t)+\frac{q+1-t}{q}\phi_{[q-1]}(t-1),\qquad t\in\mathbb{R},\qquad q\geq 1,\end{aligned}$$

where $\chi_{[0,1)}$ is the characteristic (indicator) function of the interval $[0,1)$. Let

$$f_p : [0,\pi] \to \mathbb{R}, \qquad f_p(\theta) = -\phi''_{[2p+1]}(p+1) - 2\sum_{k=1}^{p} \phi''_{[2p+1]}(p+1-k)\cos(k\theta), \qquad p \geq 1, \tag{13}$$

$$g_p : [0,\pi] \to \mathbb{R}, \qquad g_p(\theta) = \phi_{[2p+1]}(p+1) + 2\sum_{k=1}^{p} \phi_{[2p+1]}(p+1-k)\cos(k\theta), \qquad p \geq 0, \tag{14}$$

$$e_p : [0,\pi] \to \mathbb{R}, \qquad e_p(\theta) = \frac{f_p(\theta)}{g_p(\theta)}, \qquad p \geq 1. \tag{15}$$

It was proved in [15, Section 3] that [1]

$$f_p(\theta) = (2-2\cos(\theta))g_{p-1}(\theta), \qquad \theta \in [0,\pi], \qquad p \geq 1, \tag{16}$$

$$g_p(\theta) > 0, \qquad \theta \in [0,\pi], \qquad p \geq 0, \tag{17}$$

so in particular the function $e_p(\theta)$ is well-defined. It turns out that $e_p(\theta)$ is also monotone increasing over $[0,\pi]$; see Appendix A. From the analysis in [16, Section 10.7], we know that the three sequences of matrices $\{n^{-1}K_n^{[p]}\}_n$, $\{nM_n^{[p]}\}_n$, $\{n^{-2}L_n^{[p]}\}_n$ have an asymptotic spectral distribution (in the Weyl sense) described by the functions $f_p(\theta)$, $g_p(\theta)$, $e_p(\theta)$, respectively; that is, for any sufficiently large n, up to a small number of outliers, the eigenvalues of $n^{-1}K_n^{[p]}$ (resp., $nM_n^{[p]}$, $n^{-2}L_n^{[p]}$) are approximately given by the samples of $f_p(\theta)$ (resp., $g_p(\theta)$, $e_p(\theta)$) over some uniform grid in $[0,\pi]$. This is illustrated in Figure 2 for the matrix $n^{-2}L_n^{[p]}$ and for $p = 1,\ldots,6$. For more details on the spectral distribution of a sequence of matrices, see [16, Section 3.1].

1.2 Contributions of this work

The main contributions of this work can be summarized as follows. Throughout this paper, we will use the notations $n_p^{\text{out}} = p - 2 + \text{mod}(p,2)$ and $N(n,p) = n + p - 2$.

1. For $p = 1$ and $p = 2$, we compute eigenvalues and eigenvectors of $L_n^{[p]}$. In both cases, the eigenvalues are given by $e_p(\theta_{j,n})$ for $j = 1,\ldots,n+p-2$, where $\theta_{j,n} = j\pi/n$. The exact computation of eigenvalues and eigenvectors is made possible by the fact that the matrices $K_n^{[p]}$, $M_n^{[p]}$, $L_n^{[p]}$ belong to the same matrix algebra, which is the tau algebra $\tau_{n-1}(0,0)$ for $p = 1$ and the algebra $\tau_n(-1,-1)$ for $p = 2$ (we are using the notations of [7]). It is worth noting that both these algebras are related to fast unitary sine transforms [7], which implies that many numerical linear algebra computations involving the matrices $K_n^{[p]}$, $M_n^{[p]}$, $L_n^{[p]}$ (matrix-vector products, solutions of linear systems, inversions, etc.) are stable and fast.
2. For $p \geq 3$, we provide numerical evidence of a precise asymptotic expansion for the eigenvalues of $n^{-2}L_n^{[p]}$. Such an expansion, which obviously begins with the spectral distribution function $e_p(\theta)$, is in force for the whole of the spectrum except for the largest n_p^{out} eigenvalues (the so-called outliers; see Figure 2). To be more precise, we show through numerical experiments that for every $p \geq 3$, every integer $\alpha \geq 0$, every n, and every $j = 1,\ldots,N(n,p) - n_p^{\text{out}} = n - \text{mod}(p,2)$, we have

$$\lambda_j(n^{-2}L_n^{[p]}) = e_p(\theta_{j,n}) + \sum_{k=1}^{\alpha} c_k^{[p]}(\theta_{j,n})h^k + E_{j,n,\alpha}^{[p]}, \tag{18}$$

   where:
   - the eigenvalues of $n^{-2}L_n^{[p]}$ are arranged in ascending order, $\lambda_1(n^{-2}L_n^{[p]}) \leq \ldots \leq \lambda_{n+p-2}(n^{-2}L_n^{[p]})$;
   - $\{c_k^{[p]}\}_{k=1,2,\ldots}$ is a sequence of functions from $[0,\pi]$ to $\mathbb{R}$ which depends only on p;
   - $h = \frac{1}{n}$ and $\theta_{j,n} = \frac{j\pi}{n} = j\pi h$ for $j = 1,\ldots,n$;
   - $E_{j,n,\alpha}^{[p]} = O(h^{\alpha+1})$ is the remainder (the error), which satisfies the inequality $|E_{j,n,\alpha}^{[p]}| \leq C_\alpha^{[p]} h^{\alpha+1}$ for some constant $C_\alpha^{[p]}$ depending only on α and p.

[1] Note that in [15] the function $g_p(\theta)$ is denoted by $h_p(\theta)$.

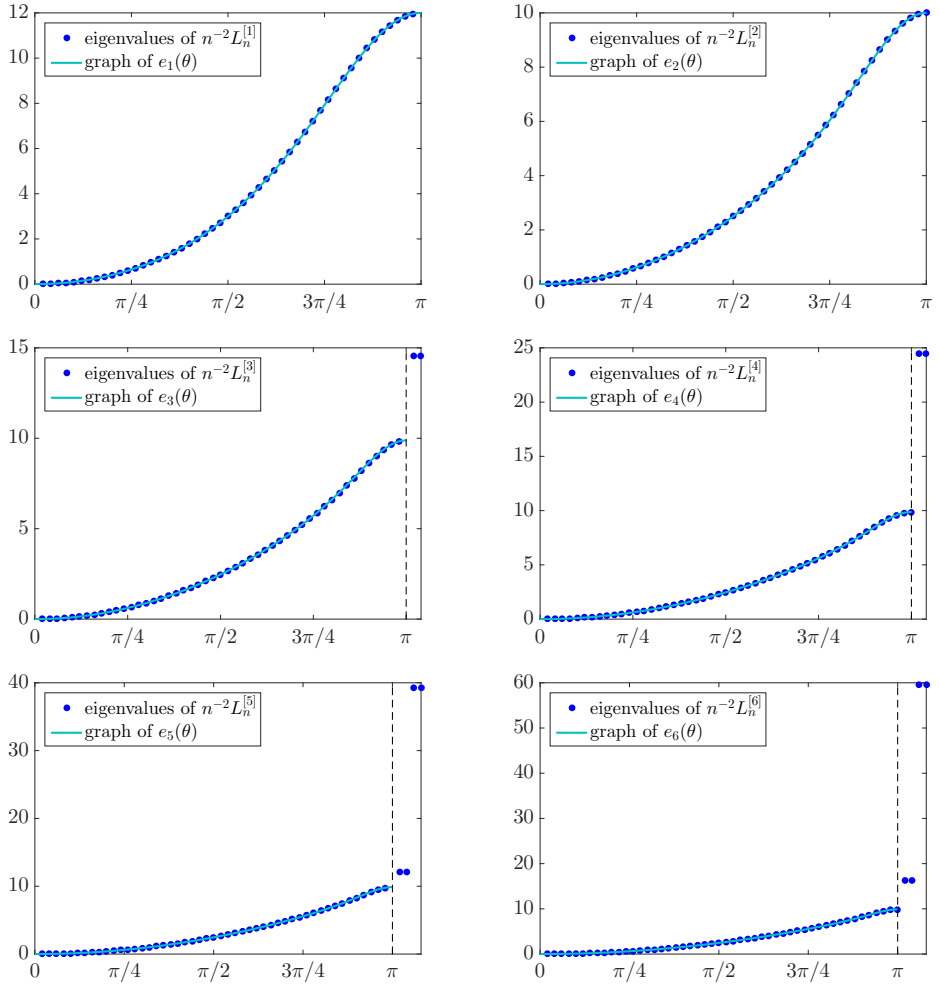


Figure 2: Comparison between the eigenvalues of $n^{-2}L_n^{[p]}$ and the graph of $e_p(\theta)$ for $n = 50$ and $p = 1, \ldots, 6$. The eigenvalues of $n^{-2}L_n^{[p]}$ are sorted in ascending order and are represented by the thick dots placed at the points $(\theta_{j,n}, \lambda_j(n^{-2}L_n^{[p]}))$, $j = 1, \ldots, n - \mathrm{mod}(p, 2)$, where $\theta_{j,n} = j\pi/n$. The eigenvalues $\lambda_j(n^{-2}L_n^{[p]})$ for $j > n - \mathrm{mod}(p, 2)$ are the so-called outliers and are positioned outside the domain $[0, \pi]$.

We refer the reader to Appendix B for a proof of the expansion (18) for $\alpha = 0$ and $j = 1, \dots, N(n,p) - (4p-2)$, where $4p-2$ represents an estimate, solely based on interlacing/rank-correction arguments, of the actual number of outliers n_p^{out}. We note that (18) is formally the same as the expansions for the eigenvalues of Toeplitz and preconditioned Toeplitz matrices, which have been conjectured and validated through numerical experiments in [1, 14]. In the case of Toeplitz matrices, the eigenvalue expansion has also been proved by Bogoya, Böttcher, Grudsky, and Maximenko in a sequel of recent papers [4, 5, 6]. Furthermore, basic eigenvalue expansions (and related extrapolation techniques) have been used in [9, 25] in the context of finite element approximations of differential problems. In the light of these considerations, the expansion (18) is not completely unexpected, because $n^{-2}L_n^{[p]}$ is 'almost' a preconditioned Toeplitz matrix as $n^{-2}L_n^{[p]} = (nM_n^{[p]})^{-1}(n^{-1}K_n^{[p]})$ and $nM_n^{[p]}$, $n^{-1}K_n^{[p]}$ are Toeplitz matrices, up to low rank corrections. To be precise, let $T_m(a)$ be the Toeplitz matrix of size m generated by the function $a \in L^1(-\pi,\pi)$, that is,

$$T_m(a) = [a_{i-j}]_{i,j=1}^m = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & \cdots & a_{-(m-1)} \\ a_1 & \ddots & \ddots & \ddots & & \vdots \\ a_2 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & a_{-2} \\ \vdots & & \ddots & \ddots & \ddots & a_{-1} \\ a_{m-1} & \cdots & \cdots & a_2 & a_1 & a_0 \end{bmatrix},$$

where the numbers $a_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} a(\theta)\mathrm{e}^{-\mathrm{i}k\theta}\mathrm{d}\theta$, $k \in \mathbb{Z}$, are the Fourier coefficients of a. Then,

$$n^{-1}K_n^{[p]} = T_{n+p-2}(f_p) + R_n^{[p]}, \tag{19}$$

$$nM_n^{[p]} = T_{n+p-2}(g_p) + S_n^{[p]}, \tag{20}$$

where f_p, g_p are defined in (13)–(14) and

$$(R_n^{[p]})_{ij} = 0, \quad 2p \le i \le n-p-1 \quad \Longrightarrow \quad \operatorname{rank}(R_n^{[p]}) \le 4p-2, \tag{21}$$

$$(S_n^{[p]})_{ij} = 0, \quad 2p \le i \le n-p-1 \quad \Longrightarrow \quad \operatorname{rank}(S_n^{[p]}) \le 4p-2; \tag{22}$$

see [15, Subsection 4.1].

3. We show through numerical experiments that, for $p \ge 3$ and $k \ge 1$, there exists a point $\theta(p,k) \in (0,\pi)$ such that $c_k^{[p]}(\theta)$ vanishes over $[0, \theta(p,k)]$. Moreover, as it is suggested by the numerics of this paper, it is very likely that $y_p = \inf_{k\ge 1}\theta(p,k) > 0$ for all $p \ge 3$. This is consistent with another crucial numerical observation, namely the fact that, for all $p \ge 3$, the equation $\lambda_j(n^{-2}L_n^{[p]}) = e_p(\theta_{j,n})$ holds numerically whenever $\theta_{j,n} < \theta(p)$, with $\theta(p)$ being a point in $(0, y_p]$. In addition, $\theta(p)$ apparently grows with p, i.e., the portion of the spectrum of $\lambda_j(n^{-2}L_n^{[p]})$ which is exactly described by $e_p(\theta)$, at least from a numerical viewpoint, increases with p.
4. For $p \ge 3$, based on the expansion (18) and drawing inspiration from [13], we propose an interpolation–extrapolation algorithm for computing the eigenvalues of $L_n^{[p]}$, excluding the n_p^{out} outliers. The performance of the algorithm is illustrated through numerical experiments. Note that we actually need to compute only the eigenvalues of $L_n^{[p]}$ corresponding in the expansion (18) to points $\theta_{j,n} \ge \theta(p)$, because whenever $\theta_{j,n} < \theta(p)$ we numerically have $\lambda_j(L_n^{[p]}) = n^2 e_p(\theta_{j,n})$ by the previous item 3.
5. We present a detailed extension of the whole analysis to the general d-dimensional setting, in which problem (1) is replaced by (31). By using tensor-product arguments, we show that the eigenvalue–eigenvector structure of the matrix arising from the IgA approximation of (31) is completely determined by the eigenvalue–eigenvector structure of the matrix $L_n^{[p]}$. In short, the analysis of $L_n^{[p]}$ is enough also to cover the multidimensional case.

1.3 Organization of the paper

The paper is organized as follows. In Section 2 we compute eigenvalues and eigenvectors of the matrix $L_n^{[p]}$ for $p = 1$ and $p = 2$. In Section 3, assuming the asymptotic eigenvalue expansion (18), we present our interpolation–extrapolation algorithm for computing the eigenvalues of $L_n^{[p]}$ for $p \ge 3$, excluding the n_p^{out} outliers. In Section 4

we provide numerical experiments in support of both the asymptotic eigenvalue expansion (18) and the properties described in item 3 of Subsection 1.2. Moreover, we numerically illustrate the performance of the interpolation–extrapolation algorithm presented in Section 3. In Section 5 we extend the whole analysis carried out in Sections 2–4 to the multidimensional setting by showing through appropriate tensor-product arguments that the multidimensional case reduces to the unidimensional case. Finally, in Section 6 we draw conclusions and outline future lines of research.

2 Eigenvalues and eigenvectors of $L_n^{[p]}$ for $p = 1$ and $p = 2$

In this section we compute the exact spectral decomposition of the matrix $L_n^{[p]}$ for $p = 1$ and $p = 2$. As a preliminary step, we recall some properties of the matrix algebras $\tau_n(\epsilon, \phi)$ introduced in [7] for $\epsilon, \phi \in \{0, 1, -1\}$. It will turn out that $K_n^{[1]}, M_n^{[1]}, L_n^{[1]}$ belong to $\tau_{n-1}(0, 0)$ and $K_n^{[2]}, M_n^{[2]}, L_n^{[2]}$ belong to $\tau_n(-1, -1)$, and this will be the key for computing eigenvalues and eigenvectors of both $L_n^{[1]}$ and $L_n^{[2]}$.

2.1 The matrix algebras $\tau_m(\epsilon, \phi)$ for $\epsilon, \phi \in \{0, 1, -1\}$

Following [7], for any $m \geq 2$ and any $\epsilon, \phi \in \{0, 1, -1\}$ we define the tridiagonal matrix

$$H_m(\epsilon, \phi) = \begin{bmatrix} \epsilon & 1 & 0 & \cdots & 0 \\ 1 & 0 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 & 1 \\ 0 & \cdots & 0 & 1 & \phi \end{bmatrix} = T_m(2\cos(\theta)) + \epsilon \mathbf{e}_1 \mathbf{e}_1^T + \phi \mathbf{e}_m \mathbf{e}_m^T,$$

where $\mathbf{e}_i$ is the ith vector of the canonical basis of $\mathbb{R}^m$. Since $H_m(\epsilon, \phi)$ is real and symmetric, it can be decomposed as

$$H_m(\epsilon, \phi) = Q_m(\epsilon, \phi) D_m(\epsilon, \phi) Q_m(\epsilon, \phi)^T,$$

where $Q_m(\epsilon, \phi)$ is a real unitary matrix and $D_m(\epsilon, \phi)$ is a real diagonal matrix. The matrix algebra generated by $H_m(\epsilon, \phi)$ is denoted by $\tau_m(\epsilon, \phi)$ and is given by

$$\tau_m(\epsilon, \phi) = \{Q_m(\epsilon, \phi) D_m Q_m(\epsilon, \phi)^T : \ D_m \text{ is a diagonal matrix of size } m\}.$$

It turns out that the matrix $Q_m(\epsilon, \phi)$ is a fast trigonometric transform such that the matrix-vector product $Q_m(\epsilon, \phi)\mathbf{v}$ can be computed in $O(m \log m)$ operations. Moreover, the diagonal entries of the matrix $D_m(\epsilon, \phi)$ (i.e., the eigenvalues of $H_m(\epsilon, \phi)$) are equal to the samples of the function $2\cos(\theta)$ at a uniform grid in $[0, \pi]$.

The cases of interest in this paper are $\epsilon = \phi = 0$ and $\epsilon = \phi = -1$. For $\epsilon = \phi = 0$, the matrix algebra $\tau_m(0, 0)$ is the so-called tau algebra, which was originally introduced in [3]. In this case, the sampling grid is

$$\frac{j\pi}{m+1}, \qquad j = 1, \ldots, m,$$

and we have

$$D_m(0,0) = \underset{j=1,\ldots,m}{\text{diag}} \left[2\cos\Big(\frac{j\pi}{m+1}\Big) \right],$$
$$Q_m(0,0) = \sqrt{\frac{2}{m+1}} \left[\sin\Big(\frac{ij\pi}{m+1}\Big) \right]_{i,j=1}^{m}.$$

For $\epsilon = \phi = -1$, the sampling grid is

$$\frac{j\pi}{m}, \qquad j = 1, \ldots, m,$$

and we have

$$D_m(-1,-1) = \underset{j=1,\dots,m}{\operatorname{diag}}\left[2\cos\Big(\frac{j\pi}{m}\Big)\right],$$
$$Q_m(-1,-1) = \sqrt{\frac{2}{m}}\left[k_j \sin\Big(\frac{(2i-1)j\pi}{2m}\Big)\right]_{i,j=1}^{m}, \qquad k_j = \begin{cases} 1/\sqrt{2}, & \text{if } j=m,\\ 1, & \text{otherwise.}\end{cases}$$

For more details on the matrix algebras $\tau_m(\epsilon,\phi)$ we refer the reader to [7].

2.2 Eigenvalues and eigenvectors of $L_n^{[p]}$ for $p = 1$

In the case $p = 1$, the stiffness and mass matrices $K_n^{[1]}$ and $M_n^{[1]}$ have size $n-1$ and a direct computation shows that

$$n^{-1}K_n^{[1]} = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix} = T_{n-1}(f_1) = 2I_{n-1} - H_{n-1}(0,0),$$
$$nM_n^{[1]} = \frac{1}{6}\begin{bmatrix} 4 & 1 & & & \\ 1 & 4 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 4 & 1 \\ & & & 1 & 4 \end{bmatrix} = T_{n-1}(g_1) = \frac{2}{3}I_{n-1} + \frac{1}{6}H_{n-1}(0,0),$$

where I_m is the $m \times m$ identity matrix and f_1, g_1 are given by (13)–(14) for $p = 1$, i.e.,

$$f_1(\theta) = 2 - 2\cos(\theta),$$
$$g_1(\theta) = \frac{2}{3} + \frac{1}{3}\cos(\theta).$$

It follows that both $K_n^{[1]}$ and $M_n^{[1]}$ belong to the tau algebra $\tau_{n-1}(0,0)$. Moreover, based on the results of Subsection 2.1, we have

$$n^{-1}K_n^{[1]} = 2I_{n-1} - H_{n-1}(0,0) = Q_{n-1}(0,0)\left(\underset{j=1,\dots,n-1}{\operatorname{diag}}\left[f_1\Big(\frac{j\pi}{n}\Big)\right]\right)Q_{n-1}(0,0)^T,$$
$$nM_n^{[1]} = \frac{2}{3}I_{n-1} + \frac{1}{6}H_{n-1}(0,0) = Q_{n-1}(0,0)\left(\underset{j=1,\dots,n-1}{\operatorname{diag}}\left[g_1\Big(\frac{j\pi}{n}\Big)\right]\right)Q_{n-1}(0,0)^T.$$

Given the algebra structure of $\tau_{n-1}(0,0)$, we obtain

$$n^{-2}L_n^{[1]} = (nM_n^{[1]})^{-1}(n^{-1}K_n^{[1]}) = Q_{n-1}(0,0)\left(\underset{j=1,\dots,n-1}{\operatorname{diag}}\left[e_1\Big(\frac{j\pi}{n}\Big)\right]\right)Q_{n-1}(0,0)^T,$$

where

$$e_1(\theta) = \frac{f_1(\theta)}{g_1(\theta)} = \frac{6(1-\cos(\theta))}{2+\cos(\theta)},$$

as defined by (15) for $p = 1$. In particular, $L_n^{[1]}$ belongs to the tau algebra $\tau_{n-1}(0,0)$ just like $K_n^{[1]}$ and $M_n^{[1]}$, and the eigenvalues and eigenvectors of $L_n^{[1]}$ are given by

$$n^2 e_1\Big(\frac{j\pi}{n}\Big), \qquad j = 1,\dots,n-1,$$
$$\sqrt{\frac{2}{n}}\left[\sin\Big(\frac{ij\pi}{n}\Big)\right]_{i=1}^{n-1}, \qquad j = 1,\dots,n-1.$$

2.3 Eigenvalues and eigenvectors of $L_n^{[p]}$ for $p = 2$

In the case $p = 2$, the stiffness and mass matrices $K_n^{[2]}$ and $M_n^{[2]}$ have size n and a direct computation shows that

$$n^{-1}K_n^{[2]} = \frac{1}{6}\begin{bmatrix} 8 & -1 & -1 & & & & \\ -1 & 6 & -2 & -1 & & & \\ -1 & -2 & 6 & -2 & -1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & -1 & -2 & 6 & -2 & -1 \\ & & & -1 & -2 & 6 & -1 \\ & & & & -1 & -1 & 8 \end{bmatrix} = T_n(f_2) + R_n^{[2]},$$

$$nM_n^{[2]} = \frac{1}{120}\begin{bmatrix} 40 & 25 & 1 & & & & \\ 25 & 66 & 26 & 1 & & & \\ 1 & 26 & 66 & 26 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & 26 & 66 & 26 & 1 \\ & & & 1 & 26 & 66 & 25 \\ & & & & 1 & 25 & 40 \end{bmatrix} = T_n(g_2) + S_n^{[2]},$$

where f_2, g_2 are given by (13)–(14) for $p = 2$, i.e.,

$$f_2(\theta) = 1 - \frac{2}{3}\cos(\theta) - \frac{1}{3}\cos(2\theta),$$
$$g_2(\theta) = \frac{11}{20} + \frac{13}{30}\cos(\theta) + \frac{1}{60}\cos(2\theta),$$

and $R_n^{[2]}, S_n^{[2]}$ are matrices of rank 4 given by

$$R_n^{[2]} = \frac{1}{6}\begin{bmatrix} 2 & 1 & & & & \\ 1 & & & & & \\ & & & & & \\ & & & & & 1 \\ & & & & 1 & 2 \end{bmatrix},$$

$$S_n^{[2]} = \frac{1}{120}\begin{bmatrix} -26 & -1 & & & & \\ -1 & & & & & \\ & & & & & \\ & & & & & -1 \\ & & & & -1 & -26 \end{bmatrix}.$$

We note that both $n^{-1}K_n^{[2]}$ and $nM_n^{[2]}$ are of the form

$$A_n(a,b,c) = T_n(a + 2b\cos(\theta) + 2c\cos(2\theta)) + R_n(b,c), \qquad R_n(b,c) = -\begin{bmatrix} b & c & & & \\ c & & & & \\ & & & & c \\ & & & c & b \end{bmatrix}. \tag{23}$$

Indeed,

$$n^{-1}K_n^{[2]} = A_n\Big(1, -\frac{1}{3}, -\frac{1}{6}\Big),$$
$$nM_n^{[2]} = A_n\Big(\frac{11}{20}, \frac{13}{60}, \frac{1}{120}\Big).$$

Now, any matrix of the form (23) is a polynomial in $H_n(-1,-1)$, and precisely

$$A_n(a,b,c) = (a - 2c)I_n + bH_n(-1,-1) + cH_n(-1,-1)^2.$$

It follows that $A_n(a,b,c)$ belongs to the matrix algebra $\tau_n(-1,-1)$. Moreover, based on the results of Subsection 2.1, we have

$$A_n(a,b,c) = Q_n(-1,-1)\left(\operatorname*{diag}_{j=1,\dots,n}\left[a+2b\cos\Big(\frac{j\pi}{n}\Big)+2c\cos\Big(\frac{2j\pi}{n}\Big)\right]\right)Q_n(-1,-1)^T.$$

In particular, $K_n^{[2]}$ and $M_n^{[2]}$ belong to $\tau_n(-1,-1)$ and

$$n^{-1}K_n^{[2]} = Q_n(-1,-1)\left(\operatorname*{diag}_{j=1,\dots,n}\left[f_2\Big(\frac{j\pi}{n}\Big)\right]\right)Q_n(-1,-1)^T,$$
$$nM_n^{[2]} = Q_n(-1,-1)\left(\operatorname*{diag}_{j=1,\dots,n}\left[g_2\Big(\frac{j\pi}{n}\Big)\right]\right)Q_n(-1,-1)^T.$$

Given the algebra structure of $\tau_n(-1,-1)$, we obtain

$$n^{-2}L_n^{[2]} = (nM_n^{[2]})^{-1}(n^{-1}K_n^{[2]}) = Q_n(-1,-1)\left(\operatorname*{diag}_{j=1,\dots,n}\left[e_2\Big(\frac{j\pi}{n}\Big)\right]\right)Q_n(-1,-1)^T,$$

where

$$e_2(\theta) = \frac{f_2(\theta)}{g_2(\theta)} = \frac{20(3-2\cos(\theta)-\cos(2\theta))}{33+26\cos(\theta)+\cos(2\theta)},$$

as defined by (15) for $p=2$. In particular, $L_n^{[2]}$ belongs to the algebra $\tau_n(-1,-1)$ just like $K_n^{[2]}$ and $M_n^{[2]}$, and the eigenvalues and eigenvectors of $L_n^{[2]}$ are given by

$$n^2 e_2\Big(\frac{j\pi}{n}\Big), \qquad j=1,\dots,n,$$
$$\sqrt{\frac{2}{n}}\left[k_j\sin\Big(\frac{(2i-1)j\pi}{2n}\Big)\right]_{i=1}^{n}, \qquad k_j=\begin{cases}1/\sqrt{2}, & \text{if } j=n,\\ 1, & \text{otherwise,}\end{cases} \qquad j=1,\dots,n.$$

Remark 1. In a recent work [24], Tani proposed a preconditioner based on the fast sine transform $Q_n(-1,-1)$ for solving linear systems arising from the IgA discretization of unidimensional differential problems. For the case $p=2$, the performance of the preconditioner was extremely good: just one Krylov iteration! The theoretical explanation of such an excellent behavior lies precisely in the exact spectral decompositions obtained in this subsection, where it is shown that $Q_n(-1,-1)$ diagonalizes simultaneously the three matrices $K_n^{[2]}, M_n^{[2]}, L_n^{[2]}$. Note that decompositions of this kind can also be used for accelerating the convergence of recently proposed iterative solvers for IgA linear systems, such as multigrid-based and preconditioned Krylov-based methods; see [12, 18] and the references therein.

3 Algorithm for computing the eigenvalues of $L_n^{[p]}$ for $p \geq 3$

Assuming the expansion (18) and drawing inspiration from [13], in this section we propose an interpolation–extrapolation algorithm for computing the eigenvalues of $L_n^{[p]}$, excluding the n_p^{out} outliers. In what follows, for each positive integer $n \in \mathbb{N} = \{1,2,3,\dots\}$ and each $p \geq 3$ we define $n^{[p]} = n - \text{mod}(p,2)$. Moreover, with each positive integer n we associate the stepsize $h = \frac{1}{n}$ and the grid points $\theta_{j,n} = j\pi h$, $j=1,\dots,n$. For notational convenience, unless otherwise stated, we will always denote a positive integer and the associated stepsize in the same way. For example, if the positive integer is n, the associated stepsize is h; if the positive integer is n_1, the associated stepsize is h_1; if the positive integer is $\bar{n}$, the associated stepsize is $\bar{h}$; etc. Throughout this section, we make the following assumptions.

- $p \geq 3$ and $n, n_1, \alpha \in \mathbb{N}$ are fixed parameters.
- $n_k = 2^{k-1}n_1$ for $k = 1,\dots,\alpha$.
- $j_k = 2^{k-1}j_1$ for $j_1 = 1,\dots,n_1$ and $k = 1,\dots,\alpha$; j_k is the index in $\{1,\dots,n_k\}$ such that $\theta_{j_k,n_k} = \theta_{j_1,n_1}$.

A graphical representation of the grids $\{\theta_{1,n_k},\dots,\theta_{n_k,n_k}\}$, $k=1,\dots,\alpha$, is reported in Figure 3 for $n_1=5$ and $\alpha=4$.

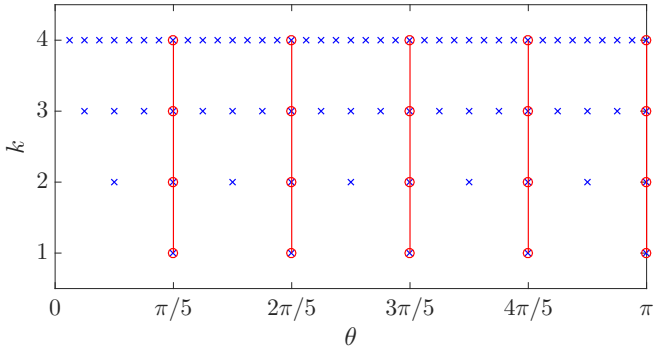


Figure 3: Representation of the grids $\{\theta_{1,n_k},\dots,\theta_{n_k,n_k}\}$, $k=1,\dots,\alpha$, for $n_1=5$ and $\alpha=4$.

For each fixed $j_1=1,\dots,n_1^{[p]}$ we apply α times the expansion (18) with $n=n_1,n_2,\dots,n_\alpha$ and $j=j_1,j_2,\dots,j_\alpha$. Since $\theta_{j_1,n_1}=\theta_{j_2,n_2}=\dots=\theta_{j_\alpha,n_\alpha}$ (by definition of $j_2,\dots,j_\alpha$), we obtain

$$\begin{cases} E^{[p]}_{j_1,n_1,0} = c_1^{[p]}(\theta_{j_1,n_1})h_1 + c_2^{[p]}(\theta_{j_1,n_1})h_1^2 + \dots + c_\alpha^{[p]}(\theta_{j_1,n_1})h_1^\alpha + E^{[p]}_{j_1,n_1,\alpha} \\ E^{[p]}_{j_2,n_2,0} = c_1^{[p]}(\theta_{j_1,n_1})h_2 + c_2^{[p]}(\theta_{j_1,n_1})h_2^2 + \dots + c_\alpha^{[p]}(\theta_{j_1,n_1})h_2^\alpha + E^{[p]}_{j_2,n_2,\alpha} \\ \quad\vdots \\ E^{[p]}_{j_\alpha,n_\alpha,0} = c_1^{[p]}(\theta_{j_1,n_1})h_\alpha + c_2^{[p]}(\theta_{j_1,n_1})h_\alpha^2 + \dots + c_\alpha^{[p]}(\theta_{j_1,n_1})h_\alpha^\alpha + E^{[p]}_{j_\alpha,n_\alpha,\alpha} \end{cases} \tag{24}$$

where

$$E^{[p]}_{j_k,n_k,0} = \lambda_{j_k}(n_k^{-2}L^{[p]}_{n_k}) - e_p(\theta_{j_1,n_1}), \qquad k=1,\dots,\alpha,$$

and

$$|E^{[p]}_{j_k,n_k,\alpha}| \le C_\alpha^{[p]} h_k^{\alpha+1}, \qquad k=1,\dots,\alpha. \tag{25}$$

Let $\tilde{c}_1^{[p]}(\theta_{j_1,n_1}),\dots,\tilde{c}_\alpha^{[p]}(\theta_{j_1,n_1})$ be the approximations of $c_1^{[p]}(\theta_{j_1,n_1}),\dots,c_\alpha^{[p]}(\theta_{j_1,n_1})$ obtained by removing all the errors $E^{[p]}_{j_1,n_1,\alpha},\dots,E^{[p]}_{j_\alpha,n_\alpha,\alpha}$ in (24) and by solving the resulting linear system:

$$\begin{cases} E^{[p]}_{j_1,n_1,0} = \tilde{c}_1^{[p]}(\theta_{j_1,n_1})h_1 + \tilde{c}_2^{[p]}(\theta_{j_1,n_1})h_1^2 + \dots + \tilde{c}_\alpha^{[p]}(\theta_{j_1,n_1})h_1^\alpha \\ E^{[p]}_{j_2,n_2,0} = \tilde{c}_1^{[p]}(\theta_{j_1,n_1})h_2 + \tilde{c}_2^{[p]}(\theta_{j_1,n_1})h_2^2 + \dots + \tilde{c}_\alpha^{[p]}(\theta_{j_1,n_1})h_2^\alpha \\ \quad\vdots \\ E^{[p]}_{j_\alpha,n_\alpha,0} = \tilde{c}_1^{[p]}(\theta_{j_1,n_1})h_\alpha + \tilde{c}_2^{[p]}(\theta_{j_1,n_1})h_\alpha^2 + \dots + \tilde{c}_\alpha^{[p]}(\theta_{j_1,n_1})h_\alpha^\alpha \end{cases} \tag{26}$$

Note that this way of computing approximations for $c_1^{[p]}(\theta_{j_1,n_1}),\dots,c_\alpha^{[p]}(\theta_{j_1,n_1})$ is completely analogous to the Richardson extrapolation procedure that is employed in the context of Romberg integration to accelerate the convergence of the trapezoidal rule [23, Section 3.4]. In this regard, the asymptotic expansion (18) plays here the same role as the Euler–Maclaurin summation formula [23, Section 3.3]. For more advanced studies on extrapolation methods, we refer the reader to Brezinski and Redivo-Zaglia [8]. The next theorem shows that the approximation error $|c_k^{[p]}(\theta_{j_1,n_1}) - \tilde{c}_k^{[p]}(\theta_{j_1,n_1})|$ is $O(h_1^{\alpha-k+1})$.

Theorem 1. *There exists a constant $A_\alpha^{[p]}$ depending only on α and p such that, for $j_1=1,\dots,n_1^{[p]}$ and $k=1,\dots,\alpha$,*

$$|c_k^{[p]}(\theta_{j_1,n_1}) - \tilde{c}_k^{[p]}(\theta_{j_1,n_1})| \le A_\alpha^{[p]} h_1^{\alpha-k+1}. \tag{27}$$

Proof. It is a straightforward adaptation of the proof of [13, Theorem 1]. □

Now, fix an index $j \in \{1, \dots, n^{[p]}\}$. To compute an approximation of $\lambda_j(n^{-2}L_n^{[p]})$ through the expansion (18) we would need the value $c_k^{[p]}(\theta_{j,n})$ for each $k = 1, \dots, \alpha$. Of course, $c_k^{[p]}(\theta_{j,n})$ is not available in practice, but we can approximate it by interpolating in some way the values $\tilde{c}_k^{[p]}(\theta_{j_1,n_1})$, $j_1 = 1, \dots, n_1^{[p]}$. For example, we may define $\tilde{c}_k^{[p]}(\theta)$ as the interpolation polynomial of the data $(\theta_{j_1,n_1}, \tilde{c}_k^{[p]}(\theta_{j_1,n_1}))$, $j_1 = 1, \dots, n_1^{[p]}$, — so that $\tilde{c}_k^{[p]}(\theta)$ is expected to be an approximation of $c_k^{[p]}(\theta)$ over the whole interval $[0, \pi]$ — and take $\tilde{c}_k^{[p]}(\theta_{j,n})$ as an approximation to $c_k^{[p]}(\theta_{j,n})$. It is known, however, that interpolation over a large number of uniform nodes is not advisable as it may give rise to spurious oscillations (Runge's phenomenon). It is therefore better to adopt another kind of approximation. An alternative could be the following: we approximate $c_k^{[p]}(\theta)$ by the spline function $\tilde{c}_k^{[p]}(\theta)$ which is linear on each interval $[\theta_{j_1,n_1}, \theta_{j_1+1,n_1}]$ and takes the value $\tilde{c}_k^{[p]}(\theta_{j_1,n_1})$ at θ_{j_1,n_1} for all $j_1 = 1, \dots, n_1^{[p]}$. This strategy removes for sure any spurious oscillation, yet it is not accurate. In particular, it does not preserve the accuracy of approximation at the nodes θ_{j_1,n_1} established in Theorem 1, i.e., there is no guarantee that $|c_k^{[p]}(\theta) - \tilde{c}_k^{[p]}(\theta)| \le B_\alpha^{[p]} h_1^{\alpha-k+1}$ for $\theta \in [0, \pi]$ or $|c_k^{[p]}(\theta_{j,n}) - \tilde{c}_k^{[p]}(\theta_{j,n})| \le B_\alpha^{[p]} h_1^{\alpha-k+1}$ for $j = 1, \dots, n^{[p]}$, with $B_\alpha^{[p]}$ being a constant depending only on α and p. As proved in Theorem 2, a local approximation strategy that preserves the accuracy (27), at least if $c_k^{[p]}(\theta)$ is sufficiently smooth, is the following: let $\theta^{(1)}, \dots, \theta^{(\alpha-k+1)}$ be $\alpha - k + 1$ points of the grid $\{\theta_{1,n_1}, \dots, \theta_{n_1^{[p]},n_1}\}$ which are closest to the point $\theta_{j,n}$,[2] and let $\tilde{c}_{k,j}^{[p]}(\theta)$ be the interpolation polynomial of the data $(\theta^{(1)}, \tilde{c}_k^{[p]}(\theta^{(1)})), \dots, (\theta^{(\alpha-k+1)}, \tilde{c}_k^{[p]}(\theta^{(\alpha-k+1)}))$; then, we approximate $c_k^{[p]}(\theta_{j,n})$ by $\tilde{c}_{k,j}^{[p]}(\theta_{j,n})$. Note that, by selecting $\alpha - k + 1$ points from $\{\theta_{1,n_1}, \dots, \theta_{n_1^{[p]},n_1}\}$, we are implicitly assuming that $n_1^{[p]} \ge \alpha - k + 1$.

Theorem 2. *Let $p \ge 3$ and $1 \le k \le \alpha$, and suppose $n_1^{[p]} \ge \alpha - k + 1$ and $c_k^{[p]} \in C^{\alpha-k+1}[0, \pi]$. For $j = 1, \dots, n^{[p]}$, if $\theta^{(1)}, \dots, \theta^{(\alpha-k+1)}$ are $\alpha - k + 1$ points of $\{\theta_{1,n_1}, \dots, \theta_{n_1^{[p]},n_1}\}$ which are closest to $\theta_{j,n}$, and if $\tilde{c}_{k,j}^{[p]}(\theta)$ is the interpolation polynomial of the data $(\theta^{(1)}, \tilde{c}_k^{[p]}(\theta^{(1)})), \dots, (\theta^{(\alpha-k+1)}, \tilde{c}_k^{[p]}(\theta^{(\alpha-k+1)}))$, then*

$$|c_k^{[p]}(\theta_{j,n}) - \tilde{c}_{k,j}^{[p]}(\theta_{j,n})| \le B_\alpha^{[p]} h_1^{\alpha-k+1} \tag{28}$$

for some constant $B_\alpha^{[p]}$ depending only on α and p.

Proof. It is a straightforward adaptation of the proof of [13, Theorem 2]. □

We are now ready to formulate our algorithm for computing the eigenvalues of $L_n^{[p]}$, excluding the outliers.

Algorithm 1. Given $p \ge 3$ and $n, n_1, \alpha \in \mathbb{N}$ with $n_1^{[p]} \ge \alpha$, we compute approximations of the eigenvalues $\lambda_j(L_n^{[p]})$ for $j = 1, \dots, n^{[p]}$ as follows.

1. For $j_1 = 1, \dots, n_1^{[p]}$ compute $\tilde{c}_1^{[p]}(\theta_{j_1,n_1}), \dots, \tilde{c}_\alpha^{[p]}(\theta_{j_1,n_1})$ by solving (26).
2. For $j = 1, \dots, n^{[p]}$
   - for $k = 1, \dots, \alpha$
     - determine $\alpha - k + 1$ points $\theta^{(1)}, \dots, \theta^{(\alpha-k+1)} \in \{\theta_{1,n_1}, \dots, \theta_{n_1^{[p]},n_1}\}$ which are closest to $\theta_{j,n}$;
     - compute $\tilde{c}_{k,j}^{[p]}(\theta_{j,n})$, where $\tilde{c}_{k,j}^{[p]}(\theta)$ is the interpolation polynomial of the data

$$(\theta^{(1)}, \tilde{c}_k^{[p]}(\theta^{(1)})), \ \dots, \ (\theta^{(\alpha-k+1)}, \tilde{c}_k^{[p]}(\theta^{(\alpha-k+1)}));$$

   - compute $\tilde{\lambda}_j(n^{-2}L_n^{[p]}) = e_p(\theta_{j,n}) + \sum_{k=1}^{\alpha} \tilde{c}_{k,j}^{[p]}(\theta_{j,n}) h^k$ and $\tilde{\lambda}_j(L_n^{[p]}) = n^2 \tilde{\lambda}_j(n^{-2}L_n^{[p]})$.
3. Return $(\tilde{\lambda}_1(L_n^{[p]}), \dots, \tilde{\lambda}_{n^{[p]}}(L_n^{[p]}))$ as an approximation to $(\lambda_1(L_n^{[p]}), \dots, \lambda_{n^{[p]}}(L_n^{[p]}))$.

Remark 2. Algorithm 1 is specifically designed for computing the eigenvalues of $L_n^{[p]}$ in the case where n is quite large. When applying this algorithm, it is implicitly assumed that n_1 and α are small (much smaller than n), so that each $n_k = 2^{k-1} n_1$ is small as well and the computation of the eigenvalues of $L_{n_k}^{[p]}$ — which is required in the first step — can be efficiently performed by any standard eigensolver (e.g., the MATLAB `eig` function).

[2] These $\alpha - k + 1$ points are uniquely determined by $\theta_{j,n}$ except in the following two cases: (a) $\theta_{j,n}$ coincides with a grid point θ_{j_1,n_1} and $\alpha - k + 1$ is even; (b) $\theta_{j,n}$ coincides with the midpoint between two consecutive grid points $\theta_{j_1,n_1}, \theta_{j_1+1,n_1}$ and $\alpha - k + 1$ is odd.

The last theorem of this section provides an estimate for the approximation error made by Algorithm 1.

Theorem 3. *Let $p \geq 3$, $n^{[p]} \geq n_1^{[p]} \geq \alpha$ and $c_k^{[p]} \in C^{\alpha-k+1}[0,\pi]$ for $k = 1,\dots,\alpha$. Let $(\tilde{\lambda}_1(L_n^{[p]}),\dots,\tilde{\lambda}_{n^{[p]}}(L_n^{[p]}))$ be the approximation of $(\lambda_1(L_n^{[p]}),\dots,\lambda_{n^{[p]}}(L_n^{[p]}))$ computed by Algorithm 1. Then, there exists a constant $D_\alpha^{[p]}$ depending only on α and p such that, for $j = 1,\dots,n^{[p]}$,*

$$|\lambda_j(L_n^{[p]}) - \tilde{\lambda}_j(L_n^{[p]})| \leq D_\alpha^{[p]} n h_1^\alpha. \tag{29}$$

Proof. By (18) and Theorem 2,

$$\begin{aligned}
|\lambda_j(n^{-2}L_n^{[p]}) - \tilde{\lambda}_j(n^{-2}L_n^{[p]})| &= \left| e_p(\theta_{j,n}) + \sum_{k=1}^{\alpha} c_k^{[p]}(\theta_{j,n})h^k + E_{j,n,\alpha}^{[p]} - e_p(\theta_{j,n}) - \sum_{k=1}^{\alpha} \tilde{c}_{k,j}^{[p]}(\theta_{j,n})h^k \right| \\
&\leq \sum_{k=1}^{\alpha} |c_k^{[p]}(\theta_{j,n}) - \tilde{c}_{k,j}^{[p]}(\theta_{j,n})|h^k + |E_{j,n,\alpha}^{[p]}| \\
&\leq B_\alpha^{[p]} \sum_{k=1}^{\alpha} h_1^{\alpha-k+1}h + C_\alpha^{[p]} h^{\alpha+1} \leq D_\alpha^{[p]} h_1^\alpha h,
\end{aligned}$$

where $D_\alpha^{[p]} = (\alpha+1)\max(B_\alpha^{[p]}, C_\alpha^{[p]})$. Multiplying both sides by n^2 we get the thesis. □

Note that the error estimate provided by Theorem 3 seems disappointing, due to the presence of the large factor n in the right-hand side of (29). However, one should take into account that (29) is an absolute error estimate which, moreover, is uniform in j. Considering that the largest non-outlier eigenvalue of $L_n^{[p]}$, namely $\lambda_{n^{[p]}}(L_n^{[p]})$, diverges to ∞ with the same asymptotic speed as n^2, from (29) we obtain the approximate inequality

$$\frac{|\lambda_{n^{[p]}}(L_n^{[p]}) - \tilde{\lambda}_{n^{[p]}}(L_n^{[p]})|}{|\lambda_{n^{[p]}}(L_n^{[p]})|} \leq D_\alpha^{[p]} h_1^\alpha h,$$

which is a good relative error estimate. We refer the reader to Subsection 4.2 for several numerical illustrations of the actual performance of Algorithm 1.

4 Numerical experiments

This section is composed of two subsections. In Subsection 4.1 we implement the program described in items 2 and 3 of Subsection 1.2. In other words, we validate through numerical experiments the expansion (18) for $p \geq 3$; we numerically show, for $p \geq 3$ and $k \geq 1$, the existence of a point $\theta(p,k) \in (0,\pi)$ such that $c_k^{[p]}(\theta)$ vanishes over $[0,\theta(p,k)]$; and we provide numerical evidence of the fact that the infimum $y_p = \inf_{k\geq 1}\theta(p,k)$ is strictly positive and the equation $\lambda_j(n^{-2}L_n^{[p]}) = e_p(\theta_{j,n})$ holds numerically whenever $\theta_{j,n} < \theta(p)$, with $\theta(p)$ being a point in $(0, y_p]$. In Subsection 4.2 we illustrate the numerical performance of Algorithm 1.

4.1 Numerical experiments in support of the eigenvalue expansion

Fix $p \geq 3$ and $\alpha \in \mathbb{N}$. As in Section 3, for every $n_1 \in \mathbb{N}$ we set

$$\begin{aligned}
n_k &= 2^{k-1}n_1, \qquad k = 1,\dots,\alpha, \\
j_k &= 2^{k-1}j_1, \qquad k = 1,\dots,\alpha, \qquad j_1 = 1,\dots,n_1.
\end{aligned}$$

In the hypothesis that the expansion (18) holds, we can follow the derivation of Section 3 until Theorem 1 and we conclude that, for each $k = 1,\dots,\alpha$ and $j_1 = 1,\dots,n_1^{[p]}$, the value $\tilde{c}_k^{[p]}(\theta_{j_1,n_1})$ computed by solving the linear system (26) converges to the value $c_k^{[p]}(\theta_{j_1,n_1})$ as $n_1 \to \infty$ with the same asymptotic speed as $h_1^{\alpha-k+1}$. In other words, in the hypothesis that the expansion (18) holds, if we plot the values $\tilde{c}_k^{[p]}(\theta_{j_1,n_1})$ versus the points θ_{j_1,n_1} for $j_1 = 1,\dots,n_1^{[p]}$, the resulting picture should converge as $n_1 \to \infty$ to the graph of a function from $[0,\pi]$ to $\mathbb{R}$, which is, by definition, $c_k^{[p]}(\theta)$. The next examples show that this is in fact the case, thus providing a validation of the expansion (18). The examples also support the following conjectures:

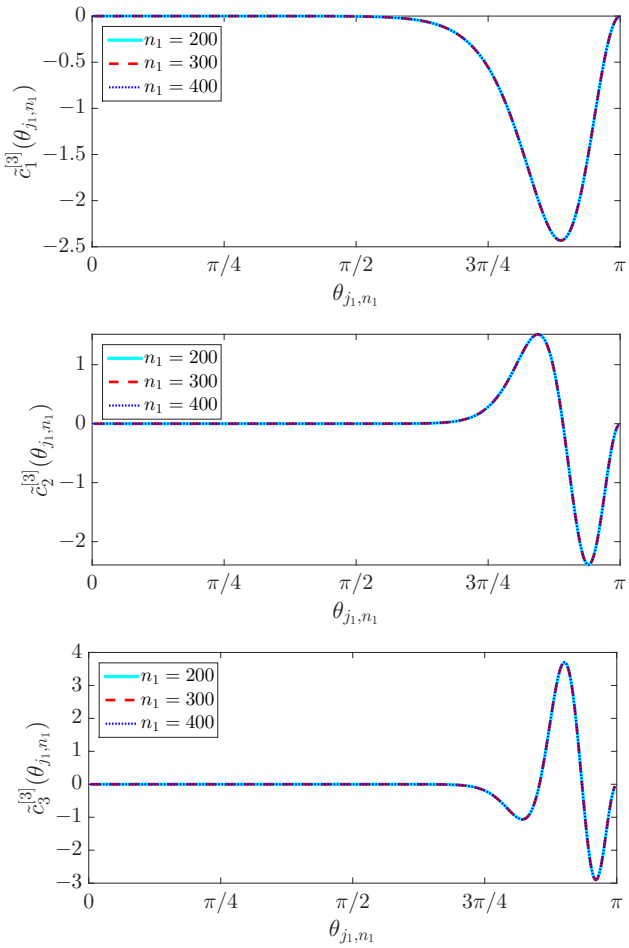


Figure 4: Example 1, $p = 3$: graph of the pairs $(\theta_{j_1,n_1}, \tilde{c}_k^{[3]}(\theta_{j_1,n_1}))$, $j_1 = 1, \ldots, n_1 - 1$, for $n_1 = 200, 300, 400$ and $k = 1, 2, 3$.

n_1	200	300	400	500	600
$\theta_{n_1}^{(\varepsilon)}(3,1)$	$\frac{86\pi}{200} \approx 1.3509$	$\frac{129\pi}{300} \approx 1.3509$	$\frac{172\pi}{400} \approx 1.3509$	$\frac{214\pi}{500} \approx 1.3446$	$\frac{257\pi}{600} \approx 1.3456$
$\theta_{n_1}^{(\varepsilon)}(3,2)$	$\frac{115\pi}{200} \approx 1.8064$	$\frac{172\pi}{300} \approx 1.8012$	$\frac{229\pi}{400} \approx 1.7986$	$\frac{286\pi}{500} \approx 1.7970$	$\frac{343\pi}{600} \approx 1.7959$
$\theta_{n_1}^{(\varepsilon)}(3,3)$	$\frac{126\pi}{200} \approx 1.9792$	$\frac{188\pi}{300} \approx 1.9687$	$\frac{251\pi}{400} \approx 1.9713$	$\frac{313\pi}{500} \approx 1.9666$	$\frac{377\pi}{600} \approx 1.9740$

Table 1: Example 1, $p = 3$: values $\theta_{n_1}^{(\varepsilon)}(3,k)$ for $k = 1, 2, 3$ and $n_1 = 200, 300, 400, 500, 600$, computed with the threshold $\varepsilon = 0.0005$.

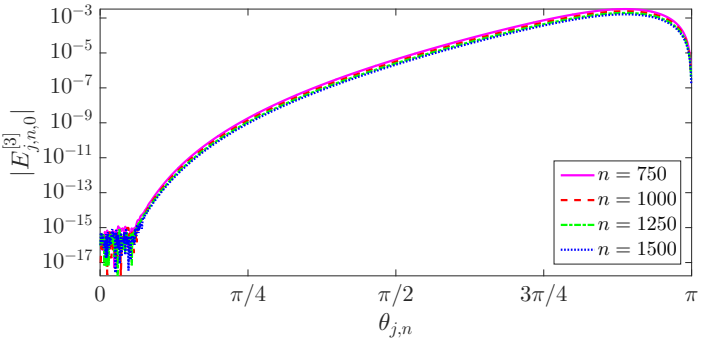


Figure 5: Example 1, $p = 3$: errors $|E^{[3]}_{j,n,0}|$ versus $\theta_{j,n}$ for $j = 1, \ldots, n-1$ and $n = 750, 1000, 1250, 1500$.

n	750	1000	1250	1500
j	58	80	101	123
$\theta_{j,n}$	0.2429	0.2513	0.2538	0.2576

Table 2: Example 1, $p = 3$: first index j such that $|E^{[3]}_{j,n,0}| > 10^{-14}$ and corresponding grid point $\theta_{j,n}$, for $n = 750, 1000, 1250, 1500$.

- the limit function $c^{[p]}_k(\theta)$ vanishes over an interval $[0, \theta(p,k)]$ with $\theta(p,k) \in (0, \pi)$;
- $y_p = \inf_{k \geq 1} \theta(p,k) > 0$;
- $\lambda_j(n^{-2} L^{[p]}_n) = e_p(\theta_{j,n})$ numerically whenever $\theta_{j,n} < \theta(p)$, where $\theta(p)$ is a point in $(0, y_p]$ which grows with p.

Example 1. Fix $p = 3$ and let $\alpha = 3$. In Figure 4 we plot the pairs

$$(\theta_{j_1,n_1}, \tilde{c}^{[3]}_k(\theta_{j_1,n_1})), \qquad j_1 = 1, \ldots, n^{[3]}_1 = n_1 - 1, \tag{30}$$

for $n_1 = 200, 300, 400$ and $k = 1, 2, 3$. We note that, for each fixed k, the graph of the pairs (30) is essentially the same for all the considered values of n_1. In other words, this graph converges to the graph of a function $c^{[3]}_k(\theta)$ as $n_1 \to \infty$, and the convergence is essentially reached already for $n_1 = 200$, at least from the point of view of graphical visualization. Moreover, the limit function $c^{[3]}_k(\theta)$ is apparently zero over an interval $[0, \theta(3,k)]$, where $\theta(3,k) \in (0, \pi)$. An ε-approximation of $\theta(3,k)$ is obtained as the limit of $\theta^{(\varepsilon)}_{n_1}(3,k)$ for $n_1 \to \infty$, where

$$\theta^{(\varepsilon)}_{n_1}(3,k) = \max\big\{\theta_{j_1,n_1} : \ 1 \leq j_1 \leq n_1 - 1, \ |\tilde{c}^{[3]}_k(\theta_{i_1,n_1})| \leq \varepsilon \text{ for all } i_1 < j_1\big\}$$

and ε is a fixed threshold. Table 1 shows the values $\theta^{(\varepsilon)}_{n_1}(3,k)$ computed for $k = 1, 2, 3$ and $n_1 = 200, 300, 400, 500, 600$ with the fixed threshold $\varepsilon = 0.0005$. Both Figure 4 and Table 1 suggest that $\theta(3,k)$ grows with k. In particular, we may expect that

$$y_3 = \inf_{k \geq 1} \theta(3,k) = \theta(3,1) > 0.$$

In Figure 5 we plot the errors $|E^{[3]}_{j,n,0}| = |\lambda_j(n^{-2} L^{[3]}_n) - e_3(\theta_{j,n})|$ versus the points $\theta_{j,n}$ for $j = 1, \ldots, n^{[3]} = n - 1$ and $n = 750, 1000, 1250, 1500$. For the same values of n, in Table 2 we record the first index j such that $|E^{[3]}_{j,n,0}| > 10^{-14}$ and the corresponding grid point $\theta_{j,n}$. From Figure 5 and Table 2 we immediately see that a nontrivial portion of the spectrum of $n^{-2} L^{[3]}_n$ is exactly approximated, at least from a numerical viewpoint, by the spectral distribution function $e_3(\theta)$. Moreover, the points $\theta_{j,n}$ shown in Table 2 apparently form a monotone increasing sequence; the limit of this sequence as $n \to \infty$, say $\theta(3) \approx 0.2576$, is a point such that the equation $\lambda_i(n^{-2} L^{[3]}_n) = e_3(\theta_{i,n})$ holds numerically whenever $\theta_{i,n} < \theta(3)$. In other words, the ratio $\theta(3)/\pi \approx 0.082$ represents the portion of the spectrum of $n^{-2} L^{[3]}_n$ which is exactly described by $e_3(\theta)$, at least numerically.

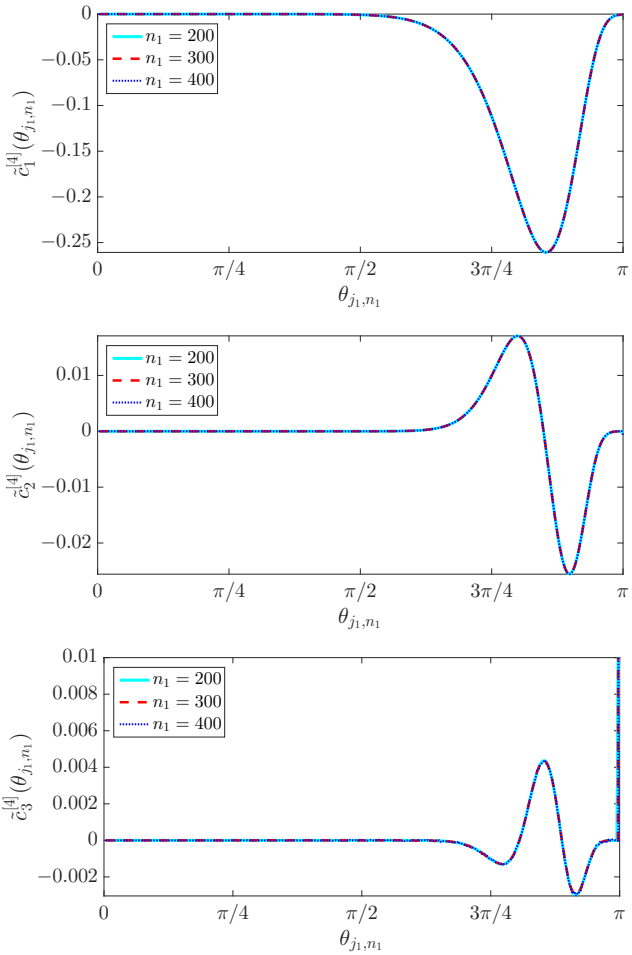


Figure 6: Example 2, $p = 4$: graph of the pairs $(\theta_{j_1,n_1}, \tilde{c}_k^{[4]}(\theta_{j_1,n_1}))$, $j_1 = 1, \ldots, n_1$, for $n_1 = 200, 300, 400$ and $k = 1, 2, 3$.

n_1	200	300	400	500	600
$\theta_{n_1}^{(\varepsilon)}(4,1)$	$\frac{97\pi}{200} \approx 1.5237$	$\frac{146\pi}{300} \approx 1.5289$	$\frac{194\pi}{400} \approx 1.5237$	$\frac{242\pi}{500} \approx 1.5205$	$\frac{291\pi}{600} \approx 1.5237$
$\theta_{n_1}^{(\varepsilon)}(4,2)$	$\frac{129\pi}{200} \approx 2.0263$	$\frac{194\pi}{300} \approx 2.0316$	$\frac{258\pi}{400} \approx 2.0263$	$\frac{322\pi}{500} \approx 2.0232$	$\frac{387\pi}{600} \approx 2.0263$
$\theta_{n_1}^{(\varepsilon)}(4,3)$	$\frac{145\pi}{200} \approx 2.2777$	$\frac{217\pi}{300} \approx 2.2724$	$\frac{289\pi}{400} \approx 2.2698$	$\frac{362\pi}{500} \approx 2.2745$	$\frac{434\pi}{600} \approx 2.2724$

Table 3: Example 2, $p = 4$: values $\theta_{n_1}^{(\varepsilon)}(4,k)$ for $k = 1, 2, 3$ and $n_1 = 200, 300, 400, 500, 600$, computed with the threshold $\varepsilon = 0.0005$.

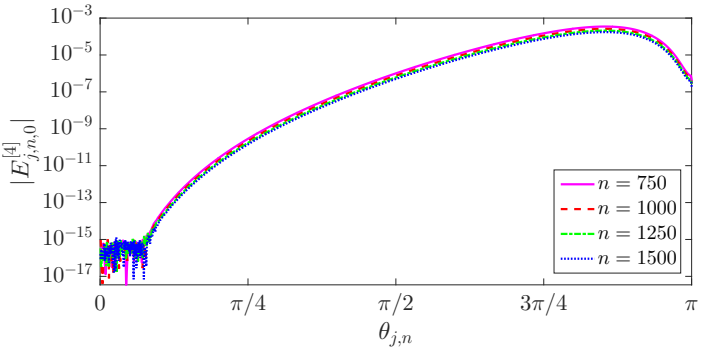


Figure 7: Example 2, $p = 4$: errors $|E^{[4]}_{j,n,0}|$ versus $\theta_{j,n}$ for $j = 1, \ldots, n$ and $n = 750, 1000, 1250, 1500$.

n	750	1000	1250	1500
j	71	97	123	152
$\theta_{j,n}$	0.2974	0.3047	0.3091	0.3183

Table 4: Example 2, $p = 4$: first index j such that $|E^{[4]}_{j,n,0}| > 10^{-14}$ and corresponding grid point $\theta_{j,n}$, for $n = 750, 1000, 1250, 1500$.

Example 2. In this example we verbatim repeat for the case $p = 4$ what we have done in Example 1 for $p = 3$. For the sake of brevity, we do not include here any comment and we limit to report the exact analogs of Figure 4, Table 1, Figure 5, and Table 2 in Figure 6, Table 3, Figure 7, and Table 4.

Example 3. A comparison between Table 2 and Table 4 shows that the portion of the spectrum of $n^{-2}L_n^{[p]}$ which is exactly described by $e_p(\theta)$, at least from a numerical viewpoint, grows from $\theta(3)/\pi \approx 0.082$ for $p = 3$ to $\theta(4)/\pi \approx 0.101$ for $p = 4$. Actually, this spectrum portion increases more and more with p, i.e., $\theta(p)$ grows with p; see Figure 8.

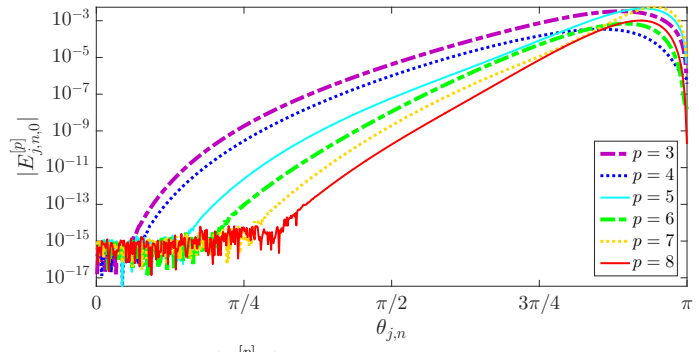


Figure 8: Example 3: errors $|E^{[p]}_{j,n,0}|$ versus $\theta_{j,n}$ for $j = 1, \ldots, n$ and $p = 3, \ldots, 8$, with $n = 750$.

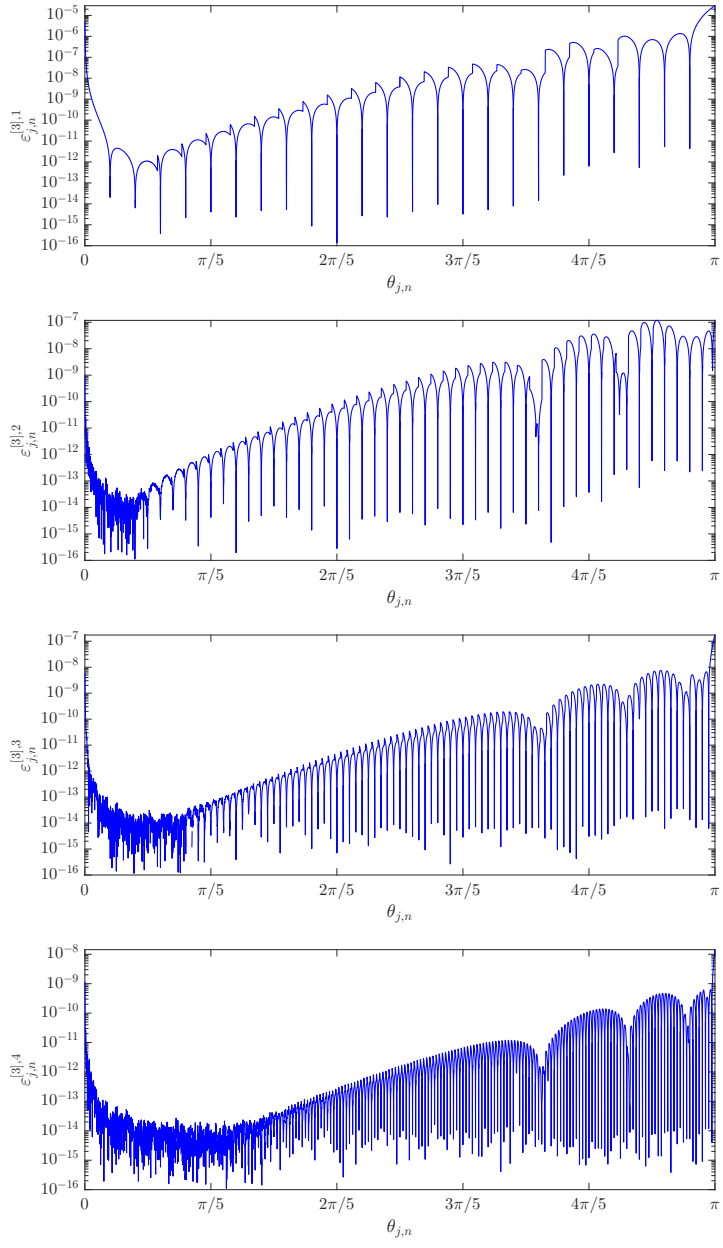


Figure 9: Example 4, $p = 3$: errors $\varepsilon_{j,n}^{[3],m}$ versus $\theta_{j,n}$ for $j = 1, \ldots, n-1$, in the case where $n = 5000$, $n_1 = 25 \cdot 2^{m-1}$, and $\alpha = 4$.

4.2 Numerical experiments illustrating the performance of Algorithm 1

Example 4. Let $p = 3$. Suppose we want to approximate the eigenvalues of $L_n^{[3]}$ (excluding the $n_3^{\text{out}} = 2$ outliers) for $n = 5000$. Let $\tilde{\lambda}_j^{(m)}(L_n^{[3]})$ be the approximation of $\lambda_j(L_n^{[3]})$ obtained by applying Algorithm 1 with $n_1 = 25 \cdot 2^{m-1}$ and $\alpha = 4$. In Figure 9 we plot the relative errors

$$\varepsilon_{j,n}^{[3],m} = \frac{|\lambda_j(L_n^{[3]}) - \tilde{\lambda}_j^{(m)}(L_n^{[3]})|}{|\lambda_j(L_n^{[3]})|}$$

versus $\theta_{j,n}$ for $j = 1, \dots, n^{[3]} = n-1$ and $m = 1, \dots, 4$. We see from the figure that the errors decrease rather quickly as m increases. A careful consideration of Figure 9 also reveals that, aside from the exceptional minima attained in a neighborhood of $\theta = 0$,[3] the local minima of $\varepsilon_{j,n}^{[3],m}$ are attained when $\theta_{j,n}$ is approximately equal to some of the coarse grid points $\theta_{j_1,n_1},\ j_1 = 1, \dots, n_1$. This is no surprise, because for $\theta_{j,n} = \theta_{j_1,n_1}$ we have $\tilde{c}_{k,j}^{[3]}(\theta_{j,n}) = \tilde{c}_k^{[3]}(\theta_{j_1,n_1})$ and $c_k^{[3]}(\theta_{j,n}) = c_k^{[3]}(\theta_{j_1,n_1})$, which means that the error of the approximation $\tilde{c}_{k,j}^{[3]}(\theta_{j,n}) \approx c_k^{[3]}(\theta_{j,n})$ reduces to the error of the approximation $\tilde{c}_k^{[3]}(\theta_{j_1,n_1}) \approx c_k^{[3]}(\theta_{j_1,n_1})$; that is, we are not introducing further error due to the interpolation process.

Example 5. Let $p = 4$. Suppose we want to approximate the eigenvalues of $L_n^{[4]}$ (excluding the $n_4^{\text{out}} = 2$ outliers) for $n = 5000$. Let $\tilde{\lambda}_j^{(m)}(L_n^{[4]})$ be the approximation of $\lambda_j(L_n^{[4]})$ obtained by applying Algorithm 1 with $n_1 = 10 \cdot 2^{m-1}$ and $\alpha = 5$. In Figure 10 we plot the relative errors

$$\varepsilon_{j,n}^{[4],m} = \frac{|\lambda_j(L_n^{[4]}) - \tilde{\lambda}_j^{(m)}(L_n^{[4]})|}{|\lambda_j(L_n^{[4]})|},$$

versus $\theta_{j,n}$ for $j = 1, \dots, n^{[4]} = n$ and $m = 1, \dots, 4$. Considerations analogous to those of Example 4 apply also in this case.

5 Extension to the multidimensional setting

We present in this section the extension to the multidimensional setting of the analysis carried out in the previous sections. In what follows, we will systematically use the multi-index notation and the properties of tensor products as described in [17, Subsections 2.1.1 and 2.6.1]. If $w_i : D_i \to \mathbb{C},\ i = 1, \dots, d$, are arbitrary functions, we will denote by $w_1 \otimes \cdots \otimes w_d : D_1 \times \cdots \times D_d \to \mathbb{C}$ the tensor-product function

$$(w_1 \otimes \cdots \otimes w_d)(\xi_1, \dots, \xi_d) = \prod_{i=1}^{d} w_i(\xi_i), \qquad (\xi_1, \dots, \xi_d) \in D_1 \times \cdots \times D_d.$$

5.1 Problem setting

Consider the d-dimensional Laplacian eigenvalue problem

$$\begin{cases} -\Delta u(\mathbf{x}) = \lambda u(\mathbf{x}), & \mathbf{x} \in (0,1)^d, \\ u(\mathbf{x}) = 0, & \mathbf{x} \in \partial((0,1)^d). \end{cases} \tag{31}$$

The corresponding weak formulation reads as follows: find eigenvalues $\lambda \in \mathbb{R}^+$ and eigenfunctions $u \in H_0^1((0,1)^d)$ such that, for all $v \in H_0^1((0,1)^d)$,

$$\mathrm{a}(u, v) = \lambda(u, v),$$

where

$$\mathrm{a}(u, v) = \int_{(0,1)^d} \nabla u(\mathbf{x}) \cdot \nabla v(\mathbf{x}) \mathrm{d}\mathbf{x}, \qquad (u, v) = \int_{(0,1)^d} u(\mathbf{x}) v(\mathbf{x}) \mathrm{d}\mathbf{x}.$$

[3] These minima, as well as the highly oscillatory behavior of the error around $\theta = 0$, are probably due to the fact that $e_3(\theta)$ provides a numerically exact description of the spectrum of $n^{-2} L_n^{[3]}$ around $\theta = 0$; see also Example 1.

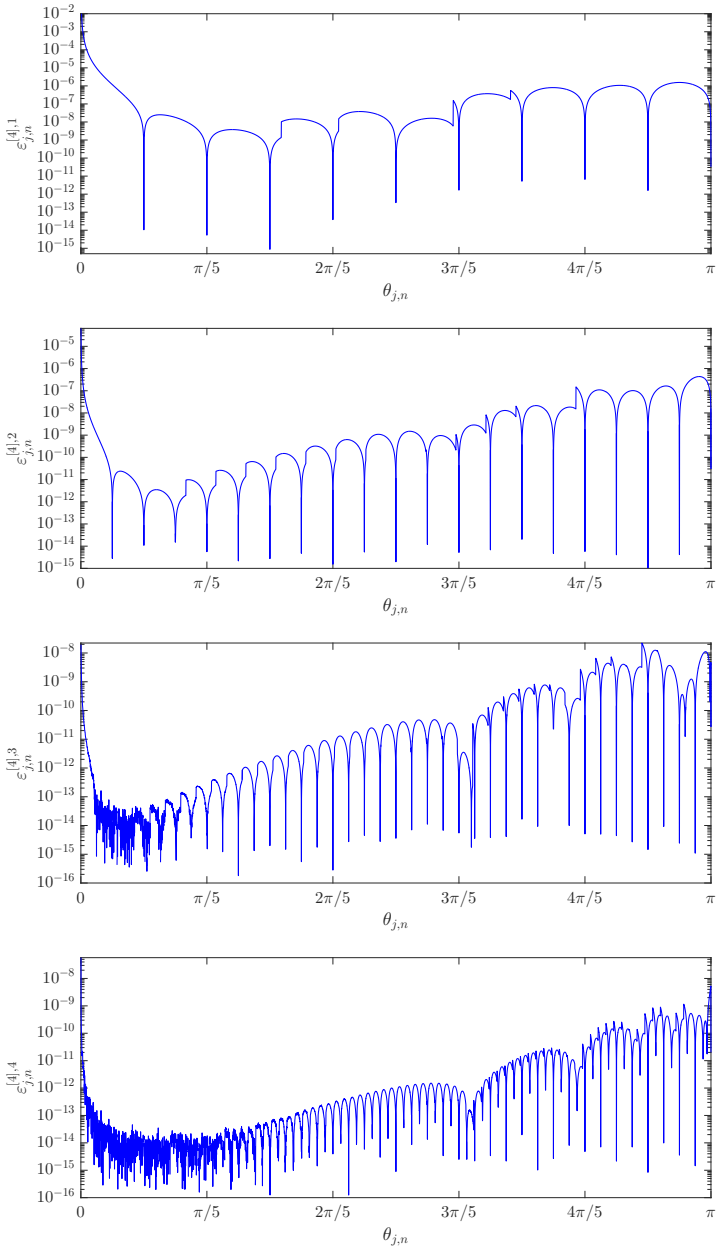


Figure 10: Example 5, $p = 4$: errors $\varepsilon_{j,n}^{[4],m}$ versus $\theta_{j,n}$ for $j = 1, \ldots, n$, in the case where $n = 5000$, $n_1 = 10 \cdot 2^{m-1}$, and $\alpha = 5$.

In the 'tensor-product version' of the Galerkin method, we choose d finite-dimensional vector spaces $\mathscr{W}_1, \dots, \mathscr{W}_d \subset H_0^1(0,1)$ and we set

$$\mathscr{W} = \mathscr{W}_1 \otimes \cdots \otimes \mathscr{W}_d = \mathrm{span}(w_1 \otimes \cdots \otimes w_d : \ w_1 \in \mathscr{W}_1, \dots, w_d \in \mathscr{W}_d) \subset H_0^1((0,1)^d).$$

Then, we define $N_s = \dim \mathscr{W}_s$ for $s = 1, \dots, d$ and $\boldsymbol{N} = (N_1, \dots, N_d)$, and we look for approximations of the exact eigenpairs

$$\lambda_{\boldsymbol{j}} = \sum_{i=1}^{d} j_i^2 \pi^2, \qquad u_{\boldsymbol{j}}(\mathbf{x}) = \prod_{i=1}^{d} \sin(j_i \pi x_i), \qquad \boldsymbol{j} = (j_1, \dots, j_d) \in \mathbb{N}^d, \tag{32}$$

by solving the following Galerkin problem: find $\lambda_{\boldsymbol{j},\mathscr{W}} \in \mathbb{R}^+$ and $u_{\boldsymbol{j},\mathscr{W}} \in \mathscr{W}$, for $\boldsymbol{j} = \mathbf{1}, \dots, \boldsymbol{N}$, such that, for all $v \in \mathscr{W}$,

$$\mathrm{a}(u_{\boldsymbol{j},\mathscr{W}}, v) = \lambda_{\boldsymbol{j},\mathscr{W}} (u_{\boldsymbol{j},\mathscr{W}}, v). \tag{33}$$

If $\{\varphi_{1,[s]}, \dots, \varphi_{N_s,[s]}\}$ is a basis of $\mathscr{W}_s$ for $s = 1, \dots, d$, then

$$\varphi_{\boldsymbol{i}} = \varphi_{i_1,[1]} \otimes \cdots \otimes \varphi_{i_d,[d]}, \qquad \boldsymbol{i} = \mathbf{1}, \dots, \boldsymbol{N},$$

is a basis of $\mathscr{W}$, and in view of the canonical identification between each $v \in \mathscr{W}$ and its coefficient vector with respect to $\{\varphi_{\mathbf{1}}, \dots, \varphi_{\boldsymbol{N}}\}$, solving the Galerkin problem (33) is equivalent to solving the generalized eigenvalue problem

$$K \mathbf{u}_{\boldsymbol{j},\mathscr{W}} = \lambda_{\boldsymbol{j},\mathscr{W}} M \mathbf{u}_{\boldsymbol{j},\mathscr{W}}, \tag{34}$$

where $\mathbf{u}_{\boldsymbol{j},\mathscr{W}}$ is the coefficient vector of $u_{\boldsymbol{j},\mathscr{W}}$ with respect to $\{\varphi_{\mathbf{1}}, \dots, \varphi_{\boldsymbol{N}}\}$,

$$K = [\mathrm{a}(\varphi_{\boldsymbol{j}}, \varphi_{\boldsymbol{i}})]_{\boldsymbol{i},\boldsymbol{j}=\mathbf{1}}^{\boldsymbol{N}} = \left[\int_{(0,1)^d} \nabla \varphi_{\boldsymbol{j}}(\mathbf{x}) \cdot \nabla \varphi_{\boldsymbol{i}}(\mathbf{x}) \mathrm{d}\mathbf{x} \right]_{\boldsymbol{i},\boldsymbol{j}=\mathbf{1}}^{\boldsymbol{N}} = \sum_{r=1}^{d} \left(\bigotimes_{s=1}^{r-1} M^{(s)} \right) \otimes K^{(r)} \otimes \left(\bigotimes_{s=r+1}^{d} M^{(s)} \right), \tag{35}$$

$$M = [(\varphi_{\boldsymbol{j}}, \varphi_{\boldsymbol{i}})]_{\boldsymbol{i},\boldsymbol{j}=\mathbf{1}}^{\boldsymbol{N}} = \left[\int_{(0,1)^d} \varphi_{\boldsymbol{j}}(\mathbf{x}) \varphi_{\boldsymbol{i}}(\mathbf{x}) \mathrm{d}\mathbf{x} \right]_{\boldsymbol{i},\boldsymbol{j}=\mathbf{1}}^{\boldsymbol{N}} = \bigotimes_{s=1}^{d} M^{(s)}, \tag{36}$$

and

$$K^{(s)} = \left[\int_0^1 \varphi'_{j,[s]}(x) \varphi'_{i,[s]}(x) \mathrm{d}x \right]_{i,j=1}^{N_s}, \qquad s = 1, \dots, d,$$

$$M^{(s)} = \left[\int_0^1 \varphi_{j,[s]}(x) \varphi_{i,[s]}(x) \mathrm{d}x \right]_{i,j=1}^{N_s}, \qquad s = 1, \dots, d.$$

The matrices K and M are, respectively, the stiffness matrix and the mass matrix. Both K and M are always symmetric positive definite, regardless of the basis functions $\varphi_{\mathbf{1}}, \dots, \varphi_{\boldsymbol{N}}$. Moreover, it is clear from (34) that the numerical eigenvalues $\lambda_{\boldsymbol{j},\mathscr{W}}$, $\boldsymbol{j} = \mathbf{1}, \dots, \boldsymbol{N}$, are just the eigenvalues of the matrix

$$L = M^{-1} K = \sum_{r=1}^{d} \left(\bigotimes_{s=1}^{r-1} I_{N_s} \right) \otimes (M^{(r)})^{-1} K^{(r)} \otimes \left(\bigotimes_{s=r+1}^{d} I_{N_s} \right). \tag{37}$$

In the IgA approximation of (31) based on uniform tensor-product B-splines of degree $\boldsymbol{p} = (p_1, \dots, p_d)$, we look for approximations of the exact eigenpairs (32) by using the tensor-product version of the Galerkin method described above, in which the basis functions $\varphi_{1,[s]}, \dots, \varphi_{N_s,[s]}$ are chosen as the B-splines $N_{2,[p_s]}, \dots, N_{n_s+p_s-1,[p_s]}$ for $s = 1, \dots, d$, where the functions $N_{i_s+1,[p_s]}$, $i_s = 1, \dots, n_s + p_s - 2$, are defined in (8) for $n = n_s$ and $p = p_s$. Setting $\boldsymbol{n} = (n_1, \dots, n_d)$, the resulting stiffness and mass matrices (35)–(36) are given by

$$K_{\boldsymbol{n}}^{[\boldsymbol{p}]} = \sum_{r=1}^{d} \left(\bigotimes_{s=1}^{r-1} M_{n_s}^{[p_s]} \right) \otimes K_{n_r}^{[p_r]} \otimes \left(\bigotimes_{s=r+1}^{d} M_{n_s}^{[p_s]} \right), \tag{38}$$

$$M_{\boldsymbol{n}}^{[\boldsymbol{p}]} = \bigotimes_{s=1}^{d} M_{n_s}^{[p_s]}, \tag{39}$$

and the numerical eigenvalues $\lambda_{\boldsymbol{j},\boldsymbol{n}}^{[\boldsymbol{p}]}$, $\boldsymbol{j} = \mathbf{1}, \ldots, \boldsymbol{n}+\boldsymbol{p}-\mathbf{2}$, are the eigenvalues of the matrix

$$L_{\boldsymbol{n}}^{[\boldsymbol{p}]} = (M_{\boldsymbol{n}}^{[\boldsymbol{p}]})^{-1} K_{\boldsymbol{n}}^{[\boldsymbol{p}]} = \sum_{r=1}^{d} \Bigg(\bigotimes_{s=1}^{r-1} I_{n_s+p_s-2} \Bigg) \otimes L_{n_r}^{[p_r]} \otimes \Bigg(\bigotimes_{s=r+1}^{d} I_{n_s+p_s-2} \Bigg), \tag{40}$$

where the matrices $K_n^{[p]}, M_n^{[p]}, L_n^{[p]}$ are defined in (10)–(12) for all $p, n \geq 1$.

5.2 Eigenvalue–eigenvector structure of $L_{\boldsymbol{n}}^{[\boldsymbol{p}]}$

We now show that the eigenvalue–eigenvector structure of $L_{\boldsymbol{n}}^{[\boldsymbol{p}]}$ is determined by the eigenvalue–eigenvector structure of the matrices $L_n^{[p]}$ for $p \in \{p_1, \ldots, p_d\}$. It will immediately follow that the eigenvalues and eigenvectors of $L_{\boldsymbol{n}}^{[\boldsymbol{p}]}$ are explicitly known for $\mathbf{1} \leq \boldsymbol{p} \leq \mathbf{2}$, due to the results of Section 2. Moreover, the interpolation–extrapolation algorithm devised in Section 3 for computing the eigenvalues of $L_n^{[p]}$ also allows the computation of the eigenvalues of $L_{\boldsymbol{n}}^{[\boldsymbol{p}]}$.

For $p, n \geq 1$, let

$$L_n^{[p]} = V_n^{[p]} D_n^{[p]} (V_n^{[p]})^{-1}, \qquad D_n^{[p]} = \operatorname*{diag}_{j=1,\ldots,n+p-2} \lambda_j(L_n^{[p]}), \tag{41}$$

be a spectral decomposition of $L_n^{[p]}$. Note that such a decomposition exists because $L_n^{[p]}$ is diagonalizable, due to the similarity equation

$$L_n^{[p]} = (M_n^{[p]})^{-1} K_n^{[p]} = (M_n^{[p]})^{-1/2} \big[(M_n^{[p]})^{-1/2} K_n^{[p]} (M_n^{[p]})^{-1/2} \big] (M_n^{[p]})^{1/2}.$$

It follows from (41) and the properties of tensor products that

$$\begin{aligned} L_{\boldsymbol{n}}^{[\boldsymbol{p}]} &= \sum_{r=1}^{d} \Bigg(\bigotimes_{s=1}^{r-1} I_{n_s+p_s-2} \Bigg) \otimes L_{n_r}^{[p_r]} \otimes \Bigg(\bigotimes_{s=r+1}^{d} I_{n_s+p_s-2} \Bigg), \\ &= \Bigg(\bigotimes_{s=1}^{d} V_{n_s}^{[p_s]} \Bigg) \Bigg[\sum_{r=1}^{d} \Bigg(\bigotimes_{s=1}^{r-1} I_{n_s+p_s-2} \Bigg) \otimes D_{n_r}^{[p_r]} \otimes \Bigg(\bigotimes_{s=r+1}^{d} I_{n_s+p_s-2} \Bigg) \Bigg] \Bigg(\bigotimes_{s=1}^{d} V_{n_s}^{[p_s]} \Bigg)^{-1}, \end{aligned} \tag{42}$$

which is a spectral decomposition of $L_{\boldsymbol{n}}^{[\boldsymbol{p}]}$. More explicitly, let $\mathbf{v}_{1,n}^{[p]}, \ldots, \mathbf{v}_{n+p-2,n}^{[p]}$ be the columns of $V_n^{[p]}$, i.e., the eigenvectors of $L_n^{[p]}$,

$$L_n^{[p]} \mathbf{v}_{j,n}^{[p]} = \lambda_j(L_n^{[p]}) \mathbf{v}_{j,n}^{[p]}, \qquad j = 1, \ldots, n+p-2,$$

and let

$$\mathbf{v}_{\boldsymbol{j},\boldsymbol{n}}^{[\boldsymbol{p}]} = \bigotimes_{s=1}^{d} \mathbf{v}_{j_s,n_s}^{[p_s]}, \qquad \boldsymbol{j} = \mathbf{1}, \ldots, \boldsymbol{n}+\boldsymbol{p}-\mathbf{2}. \tag{43}$$

Then, we can rewrite (42) as

$$L_{\boldsymbol{n}}^{[\boldsymbol{p}]} \mathbf{v}_{\boldsymbol{j},\boldsymbol{n}}^{[\boldsymbol{p}]} = \lambda_{\boldsymbol{j}}(L_{\boldsymbol{n}}^{[\boldsymbol{p}]}) \mathbf{v}_{\boldsymbol{j},\boldsymbol{n}}^{[\boldsymbol{p}]}, \qquad \boldsymbol{j} = \mathbf{1}, \ldots, \boldsymbol{n}+\boldsymbol{p}-\mathbf{2},$$

where

$$\lambda_{\boldsymbol{j}}(L_{\boldsymbol{n}}^{[\boldsymbol{p}]}) = \sum_{r=1}^{d} \lambda_{j_r}(L_{n_r}^{[p_r]}), \qquad \boldsymbol{j} = \mathbf{1}, \ldots, \boldsymbol{n}+\boldsymbol{p}-\mathbf{2}. \tag{44}$$

In other words, the eigenvalue–eigenvector pairs of $L_{\boldsymbol{n}}^{[\boldsymbol{p}]}$ are

$$(\lambda_{\boldsymbol{j}}(L_{\boldsymbol{n}}^{[\boldsymbol{p}]}), \mathbf{v}_{\boldsymbol{j},\boldsymbol{n}}^{[\boldsymbol{p}]}), \qquad \boldsymbol{j} = \mathbf{1}, \ldots, \boldsymbol{n}+\boldsymbol{p}-\mathbf{2},$$

with $\mathbf{v}_{\boldsymbol{j},\boldsymbol{n}}^{[\boldsymbol{p}]}$ and $\lambda_{\boldsymbol{j}}(L_{\boldsymbol{n}}^{[\boldsymbol{p}]})$ defined as in (43) and (44), respectively.

6 Conclusions and perspectives

We have considered the B-spline IgA approximation of the d-dimensional Laplacian eigenvalue problem (31). Through tensor-product arguments, we have shown that the eigenvalue–eigenvector structure of the resulting discretization matrix $L_{\boldsymbol{n}}^{[\boldsymbol{p}]}$ is completely determined by the eigenvalue–eigenvector structure of the matrix $L_n^{[p]}$ arising from the B-spline IgA approximation of the unidimensional eigenproblem (1). As for the matrix $L_n^{[p]}$, we implemented the program detailed in items 1 to 4 of Subsection 1.2. We conclude this work by suggesting a few possible future lines of research.

- Prove or disprove the existence of a proper matrix algebra containing the matrices $K_n^{[p]}, M_n^{[p]}, L_n^{[p]}$ for $p \geq 3$.
- Provide a formal proof of the asymptotic eigenvalue expansion (18). Considering that the eigenvalue expansion (18) is strongly connected with the eigenvalue expansion for preconditioned Toeplitz matrices [1], a proof of the former may suggest the way to prove the latter, and vice versa. Insights on how to perform these proofs might be gained from the works of Bogoya, Böttcher, Grudsky, and Maximenko [4, 5, 6], where a completely analogous eigenvalue expansion was proved for Toeplitz matrices.
- By the results of [1, 4, 5, 6, 14], Toeplitz and preconditioned Toeplitz matrices possess asymptotic eigenvalue expansions completely analogous to (18). The matrices arising from the discretization of a linear Partial Differential Equation (PDE) by a linear Numerical Method (NM) — hereinafter referred to as PDE discretization matrices — usually have a Toeplitz or Toeplitz-related structure (for example, a locally or generalized locally Toeplitz structure [16, 17, 21, 22]). A natural question is then the following: do we have asymptotic expansions also for the eigenvalues of PDE discretization matrices? This paper has provided a positive answer in the case where the PDE is the Laplacian eigenproblem (31) and the NM is the B-spline IgA. It is clear, however, that the previous question opens the doors to a series of possible future researches, whose purpose is not only to ascertain the existence of an asymptotic eigenvalue expansion for PDE discretization matrices, but also to exploit this expansion (if any) for computing the eigenvalues themselves through fast interpolation–extrapolation procedures (such as Algorithm 1).

7 Acknowledgements

The research of Sven-Erik Ekström is cofinanced by the Graduate School in Mathematics and Computing (FMB) and Uppsala University. The research of Isabella Furci and Stefano Serra-Capizzano is partially supported by the Italian INdAM – GNCS (Istituto Nazionale di Alta Matematica – Gruppo Nazionale per il Calcolo Scientifico). Carlo Garoni is an INdAM Marie-Curie fellow under grant agreement PCOFUND-GA-2012-600198.

A Monotonicity of $e_p(\theta)$

The following theorem has been proved by direct computation using MATLAB and MAPLE.

Theorem 4. *For $p = 1, \ldots, 30$, the function e_p defined in (15) is monotone increasing over $[0, \pi]$.*

Theorem 4 immediately leads to the following conjecture.

Conjecture 1. *For any $p \geq 1$, the function e_p defined in (15) is monotone increasing over $[0, \pi]$.*

Throughout this paper we have implicitly assumed Conjecture 1. The same will be done in Appendix B, where Conjecture 1 will be tacitly exploited to prove Theorem 5.

B Proof of the eigenvalue expansion for $\alpha = 0$

This appendix is devoted to the proof of the following theorem, that is, the expansion (18) for $\alpha = 0$ and $j = 1, \ldots, N(n,p) - (4p-2)$.

Theorem 5. *For every $p \geq 3$, every n, and every $j = 1, \ldots, N(n,p) - (4p-2) = n - 3p$, we have*

$$\lambda_j(n^{-2} L_n^{[p]}) = e_p(\theta_{j,n}) + E_{j,n,0}^{[p]}, \tag{45}$$

where:

- *the eigenvalues of* $n^{-2}L_n^{[p]}$ *are arranged in ascending order,* $\lambda_1(n^{-2}L_n^{[p]}) \le \ldots \le \lambda_{n+p-2}(n^{-2}L_n^{[p]})$*;*
- e_p *is the function defined in* (15)*;*
- $h = \frac{1}{n}$ *and* $\theta_{j,n} = \frac{j\pi}{n} = j\pi h$ *for* $j = 1, \ldots, n$*;*
- $|E_{j,n,0}^{[p]}| \le C^{[p]}h$ *for some constant* $C^{[p]}$ *depending only on* p*.*

Proof. Throughout this proof, we will use the simplified notations $N = N(p,n)$ and $\rho = 4p-2$. Moreover, we will write $V \subseteq_{\text{sp.}} \mathbb{C}^N$ to indicate that V is a vector subspace of $\mathbb{C}^N$. If A is an $N \times N$ matrix and $V \subseteq_{\text{sp.}} \mathbb{C}^N$, the symbol $A(V)$ will denote the subspace of $\mathbb{C}^N$ defined as $\{A\mathbf{x} : \mathbf{x} \in V\}$. Note that $A(V)$ has the same dimension as V whenever A is invertible.

We know from [20, Section 3] that

$$T_N(f_p) = \tau_N(f_p) + H_N(f_p), \tag{46}$$
$$T_N(g_p) = \tau_N(g_p) + H_N(g_p), \tag{47}$$

where, for any cosine trigonometric polynomial $\psi(\theta) = \psi_0 + 2\sum_{k=1}^{p}\psi_k\cos(k\theta)$,

- $\tau_N(\psi)$ is the tau matrix of order N generated by ψ, that is, the matrix in $\tau_N(0,0)$ defined as

$$\tau_N(\psi) = Q_N(0,0)\Big(\operatorname*{diag}_{j=1,\ldots,N}\psi\Big(\frac{j\pi}{N+1}\Big)\Big)Q_N(0,0);$$

- $H_N(\psi)$ is the Hankel matrix defined as

$$H_N(\psi) = \begin{bmatrix} \psi_2 & \psi_3 & \cdots & \psi_p & & & & \\ \psi_3 & & \iddots & & & & & \\ \vdots & \iddots & & & & & & \\ \psi_p & & & & & & & \\ & & & & & & \psi_p & \\ & & & & & \iddots & \vdots & \\ & & & & \iddots & & \psi_3 & \\ & & & \psi_p & \cdots & \psi_3 & \psi_2 & \end{bmatrix}.$$

Considering that $(H_N(f_p))_{ij} = (H_N(g_p))_{ij} = 0$ for $2p \le i \le N-2p+1 = n-p-1$, in view of (19)–(22) we have

$$n^{-1}K_n^{[p]} = \tau_N(f_p) + \hat{R}_N^{[p]}, \tag{48}$$
$$nM_n^{[p]} = \tau_N(g_p) + \hat{S}_N^{[p]}, \tag{49}$$

where the rank corrections $\hat{R}_N^{[p]} = H_N(f_p) + R_N^{[p]}$ and $\hat{S}_N^{[p]} = H_N(g_p) + S_N^{[p]}$ satisfy

$$(\hat{R}_n^{[p]})_{ij} = 0, \quad 2p \le i \le n-p-1 \quad \Longrightarrow \quad \operatorname{rank}(\hat{R}_N^{[p]}) \le \rho, \tag{50}$$
$$(\hat{S}_n^{[p]})_{ij} = 0, \quad 2p \le i \le n-p-1 \quad \Longrightarrow \quad \operatorname{rank}(\hat{S}_N^{[p]}) \le \rho. \tag{51}$$

Since $M_n^{[p]}$ is symmetric positive definite and $L_n^{[p]} = (M_n^{[p]})^{-1}K_n^{[p]}$ is similar to $(M_n^{[p]})^{-1/2}K_n^{[p]}(M_n^{[p]})^{-1/2}$, by the minimax principle for the eigenvalues of Hermitian matrices [2] we have, for every $j = 1, \ldots, N$,

$$\begin{aligned}\lambda_j(n^{-2}L_n^{[p]}) &= \lambda_j(n^{-2}(M_n^{[p]})^{-1/2}K_n^{[p]}(M_n^{[p]})^{-1/2}) \\ &= \max_{\substack{V \subseteq_{\text{sp.}} \mathbb{C}^N \\ \dim V = N-j+1}} \min_{\substack{\mathbf{x} \in V \\ \mathbf{x} \neq \mathbf{0}}} \frac{n^{-2}\mathbf{x}^*(M_n^{[p]})^{-1/2}K_n^{[p]}(M_n^{[p]})^{-1/2}\mathbf{x}}{\mathbf{x}^*\mathbf{x}} \\ &= \max_{\substack{V \subseteq_{\text{sp.}} \mathbb{C}^N \\ \dim V = N-j+1}} \min_{\substack{\mathbf{y} \in (M_n^{[p]})^{-1/2}(V) \\ \mathbf{y} \neq \mathbf{0}}} \frac{n^{-2}\mathbf{y}^*K_n^{[p]}\mathbf{y}}{\mathbf{y}^*M_n^{[p]}\mathbf{y}}\end{aligned}$$

$$= \max_{\substack{U\subseteq_{\mathrm{sp.}}\mathbb{C}^N\\ \dim U=N-j+1}} \min_{\substack{\mathbf{y}\in U\\ \mathbf{y}\neq\mathbf{0}}} \frac{\mathbf{y}^*(n^{-1}K_n^{[p]})\mathbf{y}}{\mathbf{y}^*(nM_n^{[p]})\mathbf{y}}. \tag{52}$$

Let F be the subspace of $\mathbb{C}^N$ generated by the union of the nonzero columns of $\hat{R}_n^{[p]}$ and $\hat{S}_n^{[p]}$. By (50)–(51), we have $\dim F \le \rho$ and, consequently, $\dim F^\perp \ge N-\rho$. Moreover, if U is any subspace of $\mathbb{C}^N$ such that $\dim U = u$, we have $\dim(U\cap F^\perp) = \dim U + \dim F^\perp - \dim(U+F^\perp) \ge u + (N-\rho) - N = u-\rho$. Thus, for $j=1,\dots,N-\rho$, from (48)–(49) and (52) we obtain

$$\begin{aligned}
\lambda_j(n^{-2}L_n^{[p]}) &\le \max_{\substack{U\subseteq_{\mathrm{sp.}}\mathbb{C}^N\\ \dim U=N-j+1}} \min_{\substack{\mathbf{y}\in U\cap F^\perp\\ \mathbf{y}\neq\mathbf{0}}} \frac{\mathbf{y}^*(\tau_N(f_p)+\hat{R}_n^{[p]})\mathbf{y}}{\mathbf{y}^*(\tau_N(g_p)+\hat{S}_n^{[p]})\mathbf{y}}\\
&= \max_{\substack{U\subseteq_{\mathrm{sp.}}\mathbb{C}^N\\ \dim U=N-j+1}} \min_{\substack{\mathbf{y}\in U\cap F^\perp\\ \mathbf{y}\neq\mathbf{0}}} \frac{\mathbf{y}^*\tau_N(f_p)\mathbf{y}}{\mathbf{y}^*\tau_N(g_p)\mathbf{y}}\\
&\le \max_{\substack{W\subseteq_{\mathrm{sp.}}\mathbb{C}^N\\ \dim W\ge N-(j+\rho)+1}} \min_{\substack{\mathbf{y}\in W\\ \mathbf{y}\neq\mathbf{0}}} \frac{\mathbf{y}^*\tau_N(f_p)\mathbf{y}}{\mathbf{y}^*\tau_N(g_p)\mathbf{y}}\\
&= \max_{\substack{W\subseteq_{\mathrm{sp.}}\mathbb{C}^N\\ \dim W\ge N-(j+\rho)+1}} \min_{\substack{\mathbf{x}\in(\tau_N(g_p))^{1/2}(W)\\ \mathbf{x}\neq\mathbf{0}}} \frac{\mathbf{x}^*(\tau_N(g_p))^{-1/2}\tau_N(f_p)(\tau_N(g_p))^{-1/2}\mathbf{x}}{\mathbf{x}^*\mathbf{x}}\\
&= \max_{\substack{V\subseteq_{\mathrm{sp.}}\mathbb{C}^N\\ \dim V\ge N-(j+\rho)+1}} \min_{\substack{\mathbf{x}\in V\\ \mathbf{x}\neq\mathbf{0}}} \frac{\mathbf{x}^*\tau_N(e_p)\mathbf{x}}{\mathbf{x}^*\mathbf{x}}\\
&= \max_{\substack{V\subseteq_{\mathrm{sp.}}\mathbb{C}^N\\ \dim V=N-(j+\rho)+1}} \min_{\substack{\mathbf{x}\in V\\ \mathbf{x}\neq\mathbf{0}}} \frac{\mathbf{x}^*\tau_N(e_p)\mathbf{x}}{\mathbf{x}^*\mathbf{x}}\\
&= \lambda_{j+\rho}(\tau_N(e_p)) = e_p\Big(\frac{(j+\rho)\pi}{N+1}\Big),
\end{aligned} \tag{53}$$

where the last equality is due to the monotonicity of e_p; see Appendix A. Similarly, using again the minimax

principle for Hermitian matrices, for $j = \rho+1, \dots, N$ we obtain

$$\begin{aligned}
\lambda_j(n^{-2}L_n^{[p]}) &= \lambda_j(n^{-2}(M_n^{[p]})^{-1/2}K_n^{[p]}(M_n^{[p]})^{-1/2}) \\
&= \min_{\substack{V\subseteq_{\mathrm{sp.}}\mathbb{C}^N \\ \dim V = j}} \max_{\substack{\mathbf{x}\in V \\ \mathbf{x}\neq\mathbf{0}}} \frac{n^{-2}\mathbf{x}^*(M_n^{[p]})^{-1/2}K_n^{[p]}(M_n^{[p]})^{-1/2}\mathbf{x}}{\mathbf{x}^*\mathbf{x}} \\
&= \min_{\substack{V\subseteq_{\mathrm{sp.}}\mathbb{C}^N \\ \dim V = j}} \max_{\substack{\mathbf{y}\in(M_n^{[p]})^{-1/2}(V) \\ \mathbf{y}\neq\mathbf{0}}} \frac{n^{-2}\mathbf{y}^*K_n^{[p]}\mathbf{y}}{\mathbf{y}^*M_n^{[p]}\mathbf{y}} \\
&= \min_{\substack{U\subseteq_{\mathrm{sp.}}\mathbb{C}^N \\ \dim U = j}} \max_{\substack{\mathbf{y}\in U \\ \mathbf{y}\neq\mathbf{0}}} \frac{\mathbf{y}^*(n^{-1}K_n^{[p]})\mathbf{y}}{\mathbf{y}^*(nM_n^{[p]})\mathbf{y}} \\
&\geq \min_{\substack{U\subseteq_{\mathrm{sp.}}\mathbb{C}^N \\ \dim U = j}} \max_{\substack{\mathbf{y}\in U\cap F^\perp \\ \mathbf{y}\neq\mathbf{0}}} \frac{\mathbf{y}^*(\tau_N(f_p)+\hat{R}_n^{[p]})\mathbf{y}}{\mathbf{y}^*(\tau_N(g_p)+\hat{S}_n^{[p]})\mathbf{y}} \\
&= \min_{\substack{U\subseteq_{\mathrm{sp.}}\mathbb{C}^N \\ \dim U = j}} \max_{\substack{\mathbf{y}\in U\cap F^\perp \\ \mathbf{y}\neq\mathbf{0}}} \frac{\mathbf{y}^*\tau_N(f_p)\mathbf{y}}{\mathbf{y}^*\tau_N(g_p)\mathbf{y}} \\
&\geq \min_{\substack{W\subseteq_{\mathrm{sp.}}\mathbb{C}^N \\ \dim W \geq j-\rho}} \max_{\substack{\mathbf{y}\in W \\ \mathbf{y}\neq\mathbf{0}}} \frac{\mathbf{y}^*\tau_N(f_p)\mathbf{y}}{\mathbf{y}^*\tau_N(g_p)\mathbf{y}} \\
&= \min_{\substack{W\subseteq_{\mathrm{sp.}}\mathbb{C}^N \\ \dim W \geq j-\rho}} \max_{\substack{\mathbf{x}\in(\tau_N(g_p))^{1/2}(W) \\ \mathbf{x}\neq\mathbf{0}}} \frac{\mathbf{x}^*(\tau_N(g_p))^{-1/2}\tau_N(f_p)(\tau_N(g_p))^{-1/2}\mathbf{x}}{\mathbf{x}^*\mathbf{x}} \\
&= \min_{\substack{V\subseteq_{\mathrm{sp.}}\mathbb{C}^N \\ \dim V \geq j-\rho}} \max_{\substack{\mathbf{x}\in V \\ \mathbf{x}\neq\mathbf{0}}} \frac{\mathbf{x}^*\tau_N(e_p)\mathbf{x}}{\mathbf{x}^*\mathbf{x}} \\
&= \min_{\substack{V\subseteq_{\mathrm{sp.}}\mathbb{C}^N \\ \dim V = j-\rho}} \max_{\substack{\mathbf{x}\in V \\ \mathbf{x}\neq\mathbf{0}}} \frac{\mathbf{x}^*\tau_N(e_p)\mathbf{x}}{\mathbf{x}^*\mathbf{x}} \\
&= \lambda_{j-\rho}(\tau_N(e_p)) = e_p\Big(\frac{(j-\rho)\pi}{N+1}\Big).
\end{aligned} \tag{54}$$

Putting together (53) and (54), we get

$$e_p\Big(\frac{(j-\rho)\pi}{N+1}\Big) \leq \lambda_j(n^{-2}L_n^{[p]}) \leq e_p\Big(\frac{(j+\rho)\pi}{N+1}\Big), \qquad j = \rho+1, \dots, N-\rho. \tag{55}$$

From (55) we immediately obtain

$$\begin{aligned}
\Big|\lambda_j(n^{-2}L_n^{[p]}) - e_p\Big(\frac{j\pi}{N+1}\Big)\Big| &\leq \max\Big(\Big|e_p\Big(\frac{(j-\rho)\pi}{N+1}\Big) - e_p\Big(\frac{j\pi}{N+1}\Big)\Big|, \Big|e_p\Big(\frac{(j+\rho)\pi}{N+1}\Big) - e_p\Big(\frac{j\pi}{N+1}\Big)\Big|\Big) \\
&\leq \|e_p'\|_\infty \frac{\rho\pi}{N+1} \leq \|e_p'\|_\infty \rho\pi h, \qquad j = \rho+1, \dots, N-\rho.
\end{aligned} \tag{56}$$

Moreover, since the eigenvalues of $n^{-2}L_n^{[p]}$ are positive (because of the similarity between $L_n^{[p]}$ and the symmetric positive definite matrix $(M_n^{[p]})^{-1/2}K_n^{[p]}(M_n^{[p]})^{-1/2}$) and $e_p(0) = 0 = \min_{\theta\in[0,\pi]} e_p(\theta)$ (by (16)–(17)), for $j = 1, \dots, \rho$

we have

$$\left|\lambda_j(n^{-2}L_n^{[p]}) - e_p\Big(\frac{j\pi}{N+1}\Big)\right| = \begin{cases} \lambda_j(n^{-2}L_n^{[p]}) - e_p\Big(\frac{j\pi}{N+1}\Big), & \text{if } \lambda_j(n^{-2}L_n^{[p]}) - e_p\Big(\frac{j\pi}{N+1}\Big) \geq 0, \\ e_p\Big(\frac{j\pi}{N+1}\Big) - \lambda_j(n^{-2}L_n^{[p]}), & \text{otherwise,} \end{cases}$$

$$\leq \begin{cases} \lambda_{\rho+1}(n^{-2}L_n^{[p]}) - e_p\Big(\frac{j\pi}{N+1}\Big), & \text{if } \lambda_j(n^{-2}L_n^{[p]}) - e_p\Big(\frac{j\pi}{N+1}\Big) \geq 0, \\ e_p\Big(\frac{j\pi}{N+1}\Big), & \text{otherwise,} \end{cases}$$

$$\leq \begin{cases} \left|\lambda_{\rho+1}(n^{-2}L_n^{[p]}) - e_p\Big(\frac{(\rho+1)\pi}{N+1}\Big)\right| + e_p\Big(\frac{(\rho+1)\pi}{N+1}\Big) - e_p\Big(\frac{j\pi}{N+1}\Big), & \text{if } \lambda_j(n^{-2}L_n^{[p]}) - e_p\Big(\frac{j\pi}{N+1}\Big) \geq 0, \\ e_p\Big(\frac{\rho\pi}{N+1}\Big) - e_p(0), & \text{otherwise,} \end{cases}$$

$$\leq \begin{cases} \|e_p'\|_\infty \rho\pi h + \|e_p'\|_\infty \rho\pi h, & \text{if } \lambda_j(n^{-2}L_n^{[p]}) - e_p\Big(\frac{j\pi}{N+1}\Big) \geq 0, \\ \|e_p'\|_\infty \rho\pi h, & \text{otherwise,} \end{cases}$$

$$\leq 2\|e_p'\|_\infty \rho\pi h. \tag{57}$$

Combining (56) and (57), we obtain

$$\left|\lambda_j(n^{-2}L_n^{[p]}) - e_p\Big(\frac{j\pi}{N+1}\Big)\right| \leq 2\|e_p'\|_\infty \rho\pi h, \qquad j = 1, \ldots, N-\rho. \tag{58}$$

To conclude the proof, we note that the stepsizes $h = \frac{1}{n}$ and $H = \frac{1}{N+1}$ are such that

$$0 < h - H = \frac{N+1-n}{n(N+1)} = \frac{p-1}{n(n+p-1)} < \frac{p}{n^2}$$

and, consequently, the grid points $\theta_{j,n} = j\pi h$ and $\Theta_{j,n} = j\pi H$ satisfy

$$0 < \theta_{j,n} - \Theta_{j,n} < \frac{p\pi}{n}, \qquad j = 1, \ldots, n.$$

Thus, the inequality (58) yields the thesis (45) with

$$\begin{aligned} |E_{j,n,0}^{[p]}| = |\lambda_j(n^{-2}L_n^{[p]}) - e_p(\theta_{j,n})| &\leq |\lambda_j(n^{-2}L_n^{[p]}) - e_p(\Theta_{j,n})| + |e_p(\Theta_{j,n}) - e_p(\theta_{j,n})| \\ &\leq 2\|e_p'\|_\infty \rho\pi h + \|e_p'\|_\infty p\pi h = C^{[p]}h, \qquad j = 1, \ldots, N-\rho, \end{aligned}$$

where $C^{[p]} = (2\rho + p)\pi\|e_p'\|_\infty$. □

References

[1] Ahmad F., Al-Aidarous E. S., Alrehaili D. A., Ekström S.-E., Furci I., Serra-Capizzano S. *Are the eigenvalues of preconditioned banded symmetric Toeplitz matrices known in almost closed form?* Numer. Alg. (in press) http://dx.doi.org/10.1007/s11075-017-0404-z.

[2] Bhatia R. *Matrix Analysis.* Springer (1997).

[3] Bini D., Capovani M. *Spectral and computational properties of band symmetric Toeplitz matrices.* Linear Algebra Appl. 52–53 (1983) 99–126.

[4] Bogoya J. M., Böttcher A., Grudsky S. M., Maximenko E. A. *Eigenvalues of Hermitian Toeplitz matrices with smooth simple-loop symbols.* J. Math. Anal. Appl. 422 (2015) 1308–1334.

[5] Bogoya J. M., Grudsky S. M., Maximenko E. A. *Eigenvalues of Hermitian Toeplitz matrices generated by simple-loop symbols with relaxed smoothness.* Oper. Theory Adv. Appl. 259 (2017) 179–212.

[6] Böttcher A., Grudsky S. M., Maximenko E. A. *Inside the eigenvalues of certain Hermitian Toeplitz band matrices.* J. Comput. Appl. Math. 233 (2010) 2245–2264.

[7] Bozzo E., Di Fiore C. *On the use of certain matrix algebras associated with discrete trigonometric transforms in matrix displacement decomposition.* SIAM J. Matrix Anal. Appl. 16 (1995) 312–326.

[8] Brezinski C., Redivo-Zaglia M. *Extrapolation Methods: Theory and Practice.* North-Holland, Elsevier Science Publishers B.V. (1991).

[9] Chen H., Jia S., Xie H. *Postprocessing and higher order convergence for the mixed finite element approximations of the eigenvalue problem.* Appl. Numer. Math. 61 (2011) 615–629.

[10] Cottrell J. A., Hughes T. J. R., Bazilevs Y. *Isogeometric Analysis: Toward Integration of CAD and FEA.* John Wiley & Sons (2009).

[11] De Boor C. *A Practical Guide to Splines.* Revised Edition, Springer (2001).

[12] Donatelli M., Garoni C., Manni C., Serra-Capizzano S., Speleers H. *Robust and optimal multi-iterative techniques for IgA Galerkin linear systems.* Comput. Methods Appl. Mech. Engrg. 284 (2015) 230–264.

[13] Ekström S.-E., Garoni C. *An interpolation–extrapolation algorithm for computing the eigenvalues of preconditioned banded symmetric Toeplitz matrices.* Technical Report 2017-015, Department of Information Technology, Uppsala University (2017).

[14] Ekström S.-E., Garoni C., Serra-Capizzano S. *Are the eigenvalues of banded symmetric Toeplitz matrices known in almost closed form?* Exper. Math. (in press) http://dx.doi.org/10.1080/10586458.2017.1320241.

[15] Garoni C., Manni C., Pelosi F., Serra-Capizzano S., Speleers H. *On the spectrum of stiffness matrices arising from isogeometric analysis.* Numer. Math. 127 (2014) 751–799.

[16] Garoni C., Serra-Capizzano S. *Generalized Locally Toeplitz Sequences: Theory and Applications, Volume I.* Springer (2017).

[17] Garoni C., Serra-Capizzano S. *Generalized Locally Toeplitz Sequences: Theory and Applications.* Technical Report 2017-002, Department of Information Technology, Uppsala University (2017). Preliminary version of: Garoni C., Serra-Capizzano S. *Generalized Locally Toeplitz Sequences: Theory and Applications, Volume II.* In preparation for Springer.

[18] Sangalli G., Tani M. *Isogeometric preconditioners based on fast solvers for the Sylvester equation.* SIAM J. Sci. Comput. 38 (2016) A3644–A3671.

[19] Schumaker L. L. *Spline Functions: Basic Theory.* Third Edition, Cambridge Mathematical Library (2007).

[20] Serra-Capizzano S. *On the extreme spectral properties of Toeplitz matrices generated by L^1 functions with several minima/maxima.* BIT 36 (1996) 135–142.

[21] Serra-Capizzano S. *Generalized locally Toeplitz sequences: spectral analysis and applications to discretized partial differential equations.* Linear Algebra Appl. 366 (2003) 371–402.

[22] Serra-Capizzano S. *The GLT class as a generalized Fourier analysis and applications.* Linear Algebra Appl. 419 (2006) 180–233.

[23] Stoer J., Bulirsch R. *Introduction to Numerical Analysis.* Third Edition, Springer (2002).

[24] Tani M. *FFT-based fast diagonalization methods for Galerkin IgA.* Private Communication (2017).

[25] Yin X., Xie H., Jia S., Gao S. *Asymptotic expansions and extrapolations of eigenvalues for the Stokes problem by mixed finite element methods.* J. Comput. Appl. Math. 215 (2008) 127–141.

Paper V

Received: 15 May 2017 | Revised: 30 October 2017 | Accepted: 14 November 2017
DOI: 10.1002/nla.2137

RESEARCH ARTICLE

WILEY

Eigenvalues and eigenvectors of banded Toeplitz matrices and the related symbols

S.-E. Ekström[1] | S. Serra-Capizzano[1,2]

[1]Department of Information Technology, Division of Scientific Computing, ITC, Uppsala University, Lägerhyddsv. 2, SE-751 05 Uppsala, Sweden

[2]Department of Science and High Technology, Insubria University, via Valleggio 11, Como 22100, Italy

Correspondence
S.-E. Ekström, Department of Information Technology, Division of Scientific Computing, ITC, Uppsala University, Lägerhyddsv. 2, P.O. Box 337, SE-751 05 Uppsala, Sweden.
Email: sven-erik.ekstrom@it.uu.se

Funding information
Graduate School in Mathematics and Computing (FMB); Uppsala University; Istituto Nazionale di Alta Matematica-Gruppo Nazionale di Calcolo Scientifico

Summary

It is known that for a tridiagonal Toeplitz matrix, having on the main diagonal the constant a_0 and on the two first off-diagonals the constants a_1 (lower) and a_{-1} (upper), which are all complex values, there exist closed form formulas, giving the eigenvalues of the matrix and a set of associated eigenvectors. For example, for the 1D discrete Laplacian, this triple is $(a_0, a_1, a_{-1}) = (2, -1, -1)$. In the first part of this article, we consider a tridiagonal Toeplitz matrix of the same form $(a_0, a_\omega, a_{-\omega})$, but where the two off-diagonals are positioned ω steps from the main diagonal instead of only one. We show that its eigenvalues and eigenvectors can also be identified in closed form and that interesting connections with the standard Toeplitz symbol are identified. Furthermore, as numerical evidences clearly suggest, it turns out that the eigenvalue behavior of a general banded symmetric Toeplitz matrix with real entries can be described qualitatively in terms of the symmetrically sparse tridiagonal case with real a_0, $a_\omega = a_{-\omega}$, $\omega = 2, 3, \ldots$, and also quantitatively in terms of those having monotone symbols. A discussion on the use of such results and on possible extensions complements the paper.

KEYWORDS

eigensolver, generating function and spectral symbol, Toeplitz matrix

1 | INTRODUCTION

Let A_n be a Toeplitz matrix of order n and let $\omega < n$ be a positive integer, as follows:

$$A_n = \begin{bmatrix} a_0 & \cdots & a_{-\omega} & & & \\ \vdots & \ddots & & \ddots & & \\ a_\omega & & \ddots & & \ddots & \\ & \ddots & & \ddots & & a_{-\omega} \\ & & \ddots & & \ddots & \vdots \\ & & & a_\omega & \cdots & a_0 \end{bmatrix}, \tag{1}$$

with the coefficients a_k, $k = -\omega, \ldots, \omega$, being complex numbers.

Let $f \in L^1(-\pi, \pi)$ and let $T_n(f)$ be the Toeplitz matrix generated by f, that is, $(T_n(f))_{s,t} = \hat{f}_{s-t}$, $s, t = 1, \ldots, n$, with f being the generating function of $\{T_n(f)\}$ and with $\hat{f}_k$ being the kth Fourier coefficient of f, that is,

$$\hat{f}_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) \quad e^{-\mathbf{i}k\theta} \mathrm{d}\theta, \quad \mathbf{i}^2 = -1, \quad k \in \mathbb{Z}. \tag{2}$$

If f is real valued, then several spectral properties are known (localization, extremal behavior, and collective distribution; see other works[1,2] and references therein), and f is also the spectral symbol of $\{T_n(f)\}$ in the Weyl sense.[1,3–5] If f is complex valued, then the same type of information is transferred to the singular values, whereas the eigenvalues can have a "wild" behavior[6] in some cases. According to the notation above, our setting is very special because by direct computation, the generating function of the Toeplitz matrix in (1) is the trigonometric polynomial $f(\theta) = \sum_{k=-\omega}^{\omega} a_k e^{\mathbf{i}k\theta}$, that is, $A_n = T_n(f)$.

In this paper, we are interested in quantitative estimates of the eigenvalues of A_n. Indeed, in the the band symmetric Toeplitz setting, quantitative estimates are already available in the relevant literature. In fact, using an embedding argument in the Tau algebra (the set of matrices diagonalized by a sine transform[7]), we are led to the conclusion that the jth eigenvalue $\lambda_j(A_n) = \lambda_{j,n}$, $A_n = T_n(f)$, $a_k = a_{-k} \in \mathbb{R}$, $k = 1, \dots, \omega$, can be approximated by the value $f(\theta_{\sigma(j),n})$, σ proper permutation, with an error bounded by $K_f h$, where K_f is a constant depending on f, but independent of h and j (see other works[7–10] and references therein).

The following notation is used throughout this paper. Given a positive integer n and the grid points $\theta_{j,n} = \frac{j\pi}{n+1}$, $j = 1, \dots, n$, the full grid is denoted by the following:

$$\theta_n = \{\theta_{j,n} : j = 1, \dots, n\}.$$

In the same manner, the new gridding defined in Section 2 is denoted by $\tilde{\theta}_n$. When adding a third subscript r, we mean the r:th repetition of j:th grid point, that is, $\tilde{\theta}_{r,j,n}$ is the same for all r with fixed j and n. More specifically, we will use grids of the following form:

$$\tilde{\theta}_n^{(s)} = \left\{\tilde{\theta}_{r,j,n} : \tilde{\theta}_{r,j,n} = \tilde{\theta}_{j,n},\ r = 1, \dots, n/\alpha_s, j = 1, \dots, \alpha_s\right\},$$

such that α_s divides n and $s = 1, 2$. By $\lambda_n, \mu_n, \nu_n, \xi_n$, we denote the ordered sets of eigenvalues in nondecreasing order, of the unsorted eigenvalues using the new grid, of the unsorted eigenvalue approximations from the standard grid and the standard symbol, and of the related approximations in nondecreasing order, respectively.

Here, taking into account the notation above, we furnish more precise estimates in some cases and discuss the general setting, as explained in the following.

More specifically, in Section 2, we consider the special case where $a_0, a_\omega, a_{-\omega} \in \mathbb{C}$, $a_k = 0$ for $k \neq 0, \pm\omega$ (the nontrivial setting is when $a_\omega a_{-\omega} \neq 0$). Under such assumptions, starting from the generating function $f(\theta) = a_0 + a_\omega e^{\mathbf{i}\omega\theta} + a_{-\omega} e^{-\mathbf{i}\omega\theta}$ and from the grid $\tilde{\theta}_n = \{\tilde{\theta}_{j,n} : j = 1, \dots, n\}$ described in Section 2.1, we give the closed form expression of the eigenvalues and eigenvectors in Section 2.2: a new simplified symbol emerges because the eigenvalues $\mu_n = \{\mu_{j,n}\}$, where $j = 1, \dots, n$, are exactly given as $\mu_{j,n} = g\left(\tilde{\theta}_{j,n}\right)$, with $\tilde{\theta}_n$, a proper grid, on $[0, \pi]$ and $g(\theta) = a_0 + 2\sqrt{a_\omega a_{-\omega}}\cos(\theta)$, where the new symbol $g(\theta)$ is different from the generating function $f(\theta) = a_0 + a_\omega e^{\mathbf{i}\omega\theta} + a_{-\omega} e^{-\mathbf{i}\omega\theta}$ and does not depend on ω, whereas the grid $\tilde{\theta}_n$ contains the information on ω. Finally, in Section 2.3, we discuss few relationships between the symbol g and the generating function f, in terms of the concepts of rearrangement (see, for example, other works[11] and references therein) and of spectral symbol in the Weyl sense.

In Section 3, we impose real symmetry to the matrices (1) and consider different cases. More in detail in Section 3.1, we assume that the only nonzero real coefficients of (1) are a_0 and $a_\omega = a_{-\omega}$. We compare the true eigenvalues $\lambda_{j,n}$, $j = 1, \dots, n$, sorted in a nondecreasing order, with the generating function $f(\theta) = a_0 + 2a_\omega \cos(\omega\theta)$ evaluated at the grid given by the points $\frac{j\pi}{n+1}$, which does not lead to an exact representation (except for $\omega = 1$). A closed form symbol and grid for the exact evaluation of the eigenvalues are reported in Theorem 1, and in comparison with the given representation, the accuracy of the algorithm in the work of Ekström et al.[9] is examined.

For any given sequence of indices n, where $\beta = \mathrm{mod}(n, \omega)$, $\beta = 0, 1, \dots, \omega - 1$, we show numerically that ω different "error modes" emerge, and hence, in total, ω^2 different "error modes" can be observed for a symbol of the type $f(\theta) = a_0 + 2a_\omega \cos(\omega\theta)$.

We show that each error mode $s = 0, \dots, \omega - 1$, of a given β, has the following form:

$$E^{\{s\}}_{j_\omega, n_\omega + \eta} = \lambda_{j_s,n} - f\left(\theta_{\sigma_n(j_s),n}\right) = \sum_{k=1}^{\infty} c_{k,s}\left(\theta_{\sigma_n(j_s),n}\right) h^k, \qquad h = \frac{1}{n+1}$$

and present analytical and numerical results regarding $c_{k,s}(\theta)$; see (45) and (46) for the formal definition of all variables.

On the other hand, when considering the finite-difference approximation of the operators $(-1)^q \frac{\partial^{2q}}{\partial x^{2q}}$, $q \geq 1$,we obtain Toeplitz matrices $T_n(f)$ with $f(\theta) = (2 - 2\cos(\theta))^q$ (the case of $q = 1$ coincides with $a_0 = 2$, $a_\omega = a_{-\omega} = -1$, $\omega = 1$). In such a case with $q > 1$, and more generally for monotone symbols f, the error below has the following form:

$$E_{j,n} = \lambda_{j,n} - f\left(\theta_{j,n}\right) = \sum_{k=1}^{\infty} c_k\left(\theta_{j,n}\right) h^k, \qquad h = \frac{1}{n+1}, \tag{3}$$

with $\theta_{j,n} = j\pi h, j = 1, \dots, n$, and $c_k(\theta)$, $k = 1, 2, \dots$, higher order symbols (regarding (3), see the algorithmic proposals and related numerics in the work of Ekström et al.,[9] the analysis in the work of Bogoya et al.,[12] and extensions to preconditioned and differential problems in other works[13,14]).

The functions $c_{k,s}(\theta)$ and $c_k(\theta)$ can be approximated, and a scheme is presented for performing such computations. When f is a cosine trigonometric polynomial monotone on $[0, \pi]$, it is worthwhile to mention that in other works,[15,16] expansions as in (3) are in part formally proven: however, one of the assumptions, that is, the positivity of the second derivative at zero (see page 310, line 3, in the work of Bogoya et al.[15]), excludes the important case of finite-difference approximations of (high-order) differential operators considered because $f(\theta) = (2 - 2\cos(\theta))^q$. However, even if some of the functions c_k can become discontinuous in this setting, as shown in the work of Ekström et al.,[9] the given expansions can be exploited for designing fast eigensolvers also for large matrix sizes.

In Section 3.2, we analyze the case of the general matrices in (1) with a_k being real, $a_k = a_{-k}$, $k = 1, \dots, \omega$. We consider the features and behavior of the error of the eigenvalue approximation using the symbol, because in this setting, a grid and a function giving the exact eigenvalues are not known. However, we show numerically that the eigenvalue behavior of a general banded symmetric Toeplitz matrix with real entries can be described, qualitatively in terms of the symmetrically sparse tridiagonal case with real a_0, $a_\omega = a_{-\omega}$, $\omega = 2, 3, \dots$, and also quantitatively in terms of those having monotone symbols as those related to the classical finite-difference discretization of the operators $(-1)^q \frac{\partial^{2q}}{\partial x^{2q}}$, $q \in \mathbb{N}$, $q \neq 0, 1$.

Some conclusions and possible directions for extending the current results are given in Section 4.

2 | EXACT EIGENVALUES AND EIGENVECTORS OF SYMMETRICALLY SPARSE TRIDIAGONAL, COMPLEX-VALUED TOEPLITZ MATRICES AND THE RELATED SYMBOLS

Let A_n be a Toeplitz matrix of order n and with the following nonzero structure:

$$A_n = \begin{bmatrix} a_0 & \overbrace{0 & \cdots & 0}^{\omega-1} & a_{-\omega} & & \\ 0 & a_0 & \ddots & \ddots & \ddots & \ddots & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & a_{-\omega} \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ a_\omega & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ & \ddots & \ddots & \ddots & \ddots & a_0 & 0 \\ & & a_\omega & 0 & \cdots & 0 & a_0 \end{bmatrix}, \tag{4}$$

and let the constant coefficients $a_0, a_\omega, a_{-\omega}$ be either real or complex. The constants a_ω and $a_{-\omega}$ are located on the $\omega, -\omega$ off-diagonals, respectively. The standard generating function of the matrix $A_n = T_n(f)$ is defined as follows:

$$f(\theta) = a_0 + a_\omega e^{\mathbf{i}\omega\theta} + a_{-\omega} e^{-\mathbf{i}\omega\theta}, \tag{5}$$

which is also the symbol of the sequence of matrices $\{A_n = T_n(f)\}$ in the Weyl sense.[1,3–5] Notably, when $a_\omega a_{-\omega} \neq 0$, the matrix A_n can be symmetrized in the sense that there exists a diagonal invertible matrix D_n such that

$$A_n^{\text{sym}} = D_n A_n D_n^{-1} = \begin{bmatrix} a_0 & \overbrace{0 & \cdots & 0}^{\omega-1} & \sqrt{a_\omega a_{-\omega}} & & \\ 0 & a_0 & \ddots & \ddots & \ddots & \ddots & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \sqrt{a_\omega a_{-\omega}} \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \sqrt{a_\omega a_{-\omega}} & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ & \ddots & \ddots & \ddots & \ddots & a_0 & 0 \\ & & \sqrt{a_\omega a_{-\omega}} & 0 & \cdots & 0 & a_0 \end{bmatrix}. \tag{6}$$

Therefore, A_n and A_n^{sym} are similar and share the same eigenvalues, where $A_n^{sym} = T_n(g_\omega)$ with

$$g_\omega(\theta) = a_0 + 2\sqrt{a_\omega a_{-\omega}}\cos(\omega\theta). \tag{7}$$

For the particular case $\omega = 1$, by defining the equidistant grid as follows:

$$\theta_{j,n} = \frac{j\pi}{n+1} = j\pi h, \quad j = 1, \ldots, n, \quad h = \frac{1}{n+1}, \tag{8}$$

the jth eigenvalue $\mu_{j,n}$[7,17–21] of A_n is known in closed form and is expressed as follows:

$$\mu_{j,n} = a_0 + 2\sqrt{a_\omega a_{-\omega}}\cos\left(\theta_{j,n}\right), \quad j = 1, \ldots, n. \tag{9}$$

We notice that $\mu_{j,n} = g(\theta_{j,n})$ with $g(\theta) = g_1(\theta) = a_0 + 2\sqrt{a_1 a_{-1}}\cos(\theta)$, for g_ω with $\omega = 1$ given in Equation (7). Furthermore, for the eigenvalue $\mu_{j,n}$, a corresponding eigenvector $\mathbf{x}_{j,n} = [x_1^{(j,n)}, \ldots, x_n^{(j,n)}]^{\mathrm{T}}$ has components given as follows:

$$x_k^{(j,n)} = \left(\sqrt{\frac{a_\omega}{a_{-\omega}}}\right)^k \sin\left(k\theta_{j,n}\right), \quad k = 1, \ldots, n. \tag{10}$$

It is worth noticing that the operations of square root mentioned above and used in the rest of the paper have to be handled carefully: when we write $\sqrt{\alpha/\beta}$, $\sqrt{\alpha\beta}$, we mean $\sqrt{\rho(\alpha)/\rho(\beta)}e^{\mathbf{i}\frac{\omega(\alpha)-\omega(\beta)}{2}}$, $\sqrt{\rho(\alpha)\rho(\beta)}e^{\mathbf{i}\frac{\omega(\alpha)+\omega(\beta)}{2}}$, respectively, with $\gamma = \rho(\gamma)e^{\mathbf{i}\omega(\gamma)}$, $\gamma \in \{\alpha, \beta\}$, $\rho(\gamma) \geq 0$, $\omega(\gamma) \in [0, 2\pi)$. In this way, for instance, $\sqrt{(-1)(-1)} = -1$ and, for example, without this formal convention, the formulae derived from Theorem 2.4 in the book by Böttcher et al.,[17] for the association eigenvalue–eigenvector, are simply false.

We introduce now a new sampling grid, $\tilde{\theta}_n$, which gives the exact eigenvalues $\mu_{j,n}$ for any $a_0, a_\omega, a_{-\omega} \in \mathbb{C}$ and $\omega \in \mathbb{N}, \omega < n$, in (9), and we introduce a modified version of (10) for expressing the corresponding eigenvectors $\mathbf{x}_{j,n}, j = 1, \ldots, n$.

2.1 | The new sampling grid

We start by introducing a new grid $\tilde{\theta}_n$, defined in the subsequent scheme. We first define β as the remainder of the Euclidean division of n by ω, that is,

$$\beta = n - \omega n_\omega, \quad n_\omega = \frac{n-\beta}{\omega}, \quad 0 \leq \beta < n, \quad n, \omega, \beta, n_\omega \in \mathbb{N}, \tag{11}$$

or in other words, β is the modulus operator applied to the pair (n, ω), $\beta = \mathrm{mod}(n, \omega)$, and n_ω is the quotient, which will be used as a "new" n in the subsequent definition of the new grid. We construct two separate grids, each with a standard equidistant sampling, expressed as follows:

$$\theta_{j_1,n_\omega} = \frac{j_1\pi}{n_\omega+1}, \quad j_1 = 1, \ldots, n_\omega, \tag{12}$$

$$\theta_{j_2,n_\omega+1} = \frac{j_2\pi}{n_\omega+2}, \quad j_2 = 1, \ldots, n_\omega+1. \tag{13}$$

We know that there might be multiple eigenvalues of multiplicity greater than one, and thus, we might need to repeat the same grid point several times. More specifically, we set the following gridpoints:

$$\tilde{\theta}^{(1)}_{r_1,j_1,n_\omega(\omega-\beta)} = \theta_{j_1,n_\omega}, \qquad r_1 = 1, \ldots, \omega-\beta, \quad j_1 = 1, \ldots, n_\omega, \tag{14}$$

$$\tilde{\theta}^{(2)}_{r_2,j_2,(n_\omega+1)\beta} = \theta_{j_2,n_\omega+1}, \qquad r_2 = 1, \ldots, \beta, \quad j_2 = 1, \ldots, n_\omega+1, \tag{15}$$

which is the same as writing that the grid points in (12) are repeated $\omega - \beta$ times and the grid points in (13) are repeated β times. Now, define the following two grids:

$$\tilde{\theta}^{(1)}_{n_\omega(\omega-\beta)} = \left\{ \left\{ \tilde{\theta}^{(1)}_{r_1,j_1,n_\omega(\omega-\beta)} \right\}_{r_1=1}^{\omega-\beta} \right\}_{j_1=1}^{n_\omega}, \tag{16}$$

$$\tilde{\theta}^{(2)}_{(n_\omega+1)\beta} = \left\{ \left\{ \tilde{\theta}^{(2)}_{r_2,j_2,(n_\omega+1)\beta} \right\}_{r_2=1}^{\beta} \right\}_{j_2=1}^{n_\omega+1}. \tag{17}$$

The full sampling grid $\tilde{\theta}_n$ is finally given by the union of the two grids (16) and (17), that is,

$$\tilde{\theta}_n = \tilde{\theta}^{(1)}_{n_\omega(\omega-\beta)} \bigcup \tilde{\theta}^{(2)}_{(n_\omega+1)\beta}. \tag{18}$$

For examples of concrete constructions of these grids, refer to the work of Ekström et al.[22]

2.2 | Eigenvalues and eigenvectors described by the new sampling grid

We start with the main results regarding symmetrically sparse tridiagonal (SST) Toeplitz matrices.

Theorem 1. *The eigenvalues of a SST Toeplitz matrix with center diagonal a_0 and two off-diagonals a_ω and $a_{-\omega}$ at off-diagonal $-\omega$ and ω, as in (4), are given by the following:*

$$\mu_{j,n} = g\left(\tilde{\theta}_{j,n}\right) = a_0 + 2\sqrt{a_\omega a_{-\omega}}\cos\left(\tilde{\theta}_{j,n}\right), \quad j = 1, \dots, n, \tag{19}$$

where $\tilde{\theta}_{j,n}$ is the jth component of the grid $\tilde{\theta}_n$ defined in (18).

Remark 1. By $\mu_n^{(1)}$ and $\mu_n^{(2)}$, we denote the set of eigenvalues given by the symbol evaluations of grids $\tilde{\theta}^{(1)}_{n_\omega(\omega-\beta)}$ and $\tilde{\theta}^{(2)}_{(n_\omega+1)\beta}$ given in (16) and (17), respectively. Assume $a_\omega a_{-\omega} \geq 0$, so that $g(\cdot)$ is real valued; let $\lambda_{j,n}$ be the eigenvalues $\mu_{j,n}$ in Theorem 1 sorted in a nondecreasing order, and let π_n be a permutation of $\{1, \dots, n\}$, which sorts the samples $g(\tilde{\theta}_{1,n}), \dots, g(\tilde{\theta}_{n,n})$ in nondecreasing order, that is, $g(\tilde{\theta}_{\pi_n(1),n}) \leq \cdots \leq g(\tilde{\theta}_{\pi_n(n),n})$. Then,

$$\lambda_{j,n} = g\left(\tilde{\theta}_{\pi_n(j),n}\right) \quad j = 1, \dots, n.$$

Theorem 2. *Given a SST Toeplitz matrix with center diagonal a_0 and two off-diagonals a_ω and $a_{-\omega}$ at off-diagonal $-\omega$ and ω, as in (4), the following statements concerning its eigenvalues and eigenvectors hold.*

For each eigenvalue given by $\mu^{(1)}_{r_1,j_1,n_\omega(\omega-\beta)} = g\left(\tilde{\theta}^{(1)}_{r_1,j_1,n_\omega(\omega-\beta)}\right) = g(\theta_{j_1,n_\omega})$ with $j_1 = 1, \dots, n_\omega$, and $r_1 = 1, \dots, \omega-\beta$, we define a corresponding eigenvector $\mathbf{x}^{(1)}_{r_1,j_1,n} = \left[x_1^{(r_1,j_1,n)}, \cdots, x_n^{(r_1,j_1,n)}\right]^{\mathrm{T}}$, with the following components:

$$x^{(r_1,j_1,n)}_{\omega(k_1-1)+r_1+\beta} = \left(\sqrt{\frac{a_\omega}{a_{-\omega}}}\right)^{k_1} \sin\left(k_1\theta_{j_1,n_\omega}\right), \quad k_1 = 1, \dots, n_\omega, \tag{20}$$

and all nondefined components of $\mathbf{x}_{r_1,j_1,n}$ equal to zero.

For each eigenvalue $\mu^{(2)}_{r_2,j_2,(n_\omega+1)\beta} = g\left(\tilde{\theta}^{(2)}_{r_2,j_2,(n_\omega+1)\beta}\right) = g(\theta_{j_2,n_\omega+1})$ with $j_2 = 1, \dots, n_\omega+1$, and $r_2 = 1, \dots, \beta$, we can define a corresponding eigenvector $\mathbf{x}^{(2)}_{r_2,j_2,n} = \left[x_1^{(r_2,j_2,n)}, \dots, x_n^{(r_2,j_2,n)}\right]^{\mathrm{T}}$, where the components are as follows:

$$x^{(r_2,j_2,n)}_{\omega(k_2-1)+r_2} = \left(\sqrt{\frac{a_\omega}{a_{-\omega}}}\right)^{k_2} \sin\left(k_2\theta_{j_2,n_\omega+1}\right), \quad k_2 = 1, \cdots, n_\omega+1, \tag{21}$$

and all nondefined components of $\mathbf{x}_{r_2,j_2,n}$ are equal to zero.

Remark 2. To save memory and evaluations, taking into account (12) and (13), the steps to construct $\tilde{\theta}_n$ can be skipped, as long as the information concerning multiple eigenvalues is stored. Note that if a grid is desired with all $\theta \in \tilde{\theta}_n$ unique in $[0, \pi]$, one can modify the set $\tilde{\theta}_n$ in (18) as follows: take $\theta \in \tilde{\theta}_n/\omega$ and then shift each grid point by appropriate multiples of π/ω. Then, also the symbol reported in Theorem 1 has to be modified, and instead of $g(\theta) = g_1(\theta)$, we use the generating function of the symmetrized matrix A_n^{sym}, that is, $g_\omega(\theta) = a_0 + 2\sqrt{a_\omega a_{-\omega}}\cos(\omega\theta)$.

Proof of Theorem 1 and Theorem 2. The proof for $\omega > 1$ follows the same ideas as for the case $\omega = 1$ presented in the work of Böttcher et al.[17] We start by observing that the matrix A_n in (4) has the standard symbol as follows:

$$f(\theta) = a_0 + a_\omega e^{\mathbf{i}\omega\theta} + a_{-\omega}e^{-\mathbf{i}\omega\theta}.$$

By assuming $a_\omega \neq 0$ and $a_{-\omega} \neq 0$, and defining $\gamma = \sqrt{a_{-\omega}/a_\omega}$, we consider the matrix B_n defined as follows:

$$B_n = \begin{bmatrix} 0 & \overbrace{0 & \cdots & 0}^{\omega-1} & \gamma^2 & & \\ 0 & 0 & \ddots & \ddots & \ddots & \ddots & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \gamma^2 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 1 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ & \ddots & \ddots & \ddots & \ddots & 0 & 0 \\ & & 1 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Thus, B_n has the following symbol:

$$f_B(\theta) = e^{\mathbf{i}\omega\theta} + \gamma^2 e^{-\mathbf{i}\omega\theta} = e^{\mathbf{i}\omega\theta} + \frac{a_{-\omega}}{a_\omega} e^{-\mathbf{i}\omega\theta}.$$

Following the general framework, because $f(\theta) = a_0 + a_\omega f_B(\theta)$, it is sufficient to show that B_n has the eigenvalues as follows:

$$\mu^{(1)}_{r_1,j_1,n_\omega(\omega-\beta)} = 2\gamma\cos\left(\theta_{j_1,n_\omega}\right), \qquad r_1 = 1,\ \dots,\omega-\beta, \quad j_1 = 1,\ \dots,n_\omega, \tag{22}$$

$$\mu^{(2)}_{r_2,j_2,(n_\omega+1)\beta} = 2\gamma\cos\left(\theta_{j_2,n_\omega+1}\right), \qquad r_2 = 1,\ \dots,\beta, \quad j_2 = 1,\cdots,n_\omega+1, \tag{23}$$

and that the following corresponding eigenvectors:

$$\mathbf{x}^{(1)}_{r_1,j_1,n} = \left[x_1^{(r_1,j_1,n)},\cdots,x_n^{(r_1,j_1,n)}\right]^{\mathrm{T}}, \tag{24}$$

$$\mathbf{x}^{(2)}_{r_2,j_2,n} = \left[x_1^{(r_2,j_2,n)},\ \dots,x_n^{(r_2,j_2,n)}\right]^{\mathrm{T}}, \tag{25}$$

have components of the following form:

$$x^{(r_1,j_1,n)}_{\omega(k_1-1)+r_1+\beta} = \gamma^{-k_1}\sin\left(k_1\theta_{j_1,n_\omega}\right), \qquad k_1 = 1,\ \dots,n_\omega, \tag{26}$$

$$x^{(r_2,j_2,n)}_{\omega(k_2-1)+r_2} = \gamma^{-k_2}\sin\left(k_2\theta_{j_2,n_\omega+1}\right), \qquad k_2 = 1,\cdots,n_\omega+1, \tag{27}$$

respectively. Because $B_n\mathbf{x} = \mu\mathbf{x}$ for a given eigenpair $(\mu,\mathbf{x})$, for all k relationships, (28)–(32) must hold true. For $\omega \le n/2$,

$$\gamma^2 x_{\omega+k} = \mu x_k, \qquad k = 1,\ \dots,\omega, \tag{28}$$

$$x_k + \gamma^2 x_{2\omega+k} = \mu x_{\omega+k}, \qquad k = 1,\ \dots,n-2\omega, \tag{29}$$

$$x_{n+1-(\omega+k)} = \mu x_{n+1-k}, \qquad k = 1,\ \dots,\omega. \tag{30}$$

For $n/2 < \omega < n$,

$$\gamma^2 x_{\omega+k} = \mu x_k, \qquad k = 1,\ \dots,n-\omega, \tag{31}$$

$$x_{n+1-(\omega+k)} = \mu x_{n+1-k}, \qquad k = 1,\ \dots,n-\omega. \tag{32}$$

First, we show that Equations (28) and (31) are satisfied. For $\mathbf{x}^{(1)}_{r_1,j_1,n}$ in (24), the nonzero components have indices of the form $\omega(k_1-1)+r_1+\beta, k_1 = 1,\ \dots\ ,n_\omega$ (as seen in (26)). For $k_1 = 1$, we have $r_1+\beta$, and for $k_2 = 2$, we have $\omega+r_1+\beta$, which are the only two nonzero components that match (28) and (31). More specifically, we observe the following:

$$x^{(r_1,j_1,n)}_{\omega+r_1+\beta} = \mu^{(1)}_{r_1,j_1,n_\omega(\omega-\beta)} x^{(r_1,j_1,n)}_{r_1+\beta}, \tag{33}$$

or, explicitly,

$$\gamma^2\gamma^{-2}\sin\left(2\theta_{j_1,n_\omega}\right) = 2\gamma\cos\left(\theta_{j_1,n_\omega}\right)\gamma^{-1}\sin\left(\theta_{j_1,n_\omega}\right), \tag{34}$$

that is, $\sin\left(2\theta_{j_1,n_\omega}\right) = 2\cos\left(\theta_{j_1,n_\omega}\right)\sin\left(\theta_{j_1,n_\omega}\right)$, which is true, owing to the trigonometric identity as follows:

$$\sin(2\gamma_1) = 2\cos(\gamma_1)\sin(\gamma_1). \tag{35}$$

For $\mathbf{x}^{(2)}_{r_2,j_2,n}$ in (25), we observe the same behavior as for $\mathbf{x}^{(1)}_{r_1,j_1,n}$ in (24) above, but the relation analogous to (33) is now as follows:

$$x^{(r_2,j_2,n)}_{\omega+r_2} = \mu^{(2)}_{r_2,j_2,(n_\omega+1)\beta} x^{(r_2,j_2,n)}_{r_2}.$$

Namely, it is the same as (34), except for the fact that $\theta_{j_2,n_\omega+1}$ replaces θ_{j_1,n_ω}.

Secondly, we show that (29) is true. For $\mathbf{x}^{(1)}_{r_1,j_1,n}$ in (24), the nonzero components have indices of the form $\omega(k_1-1)+r_1+\beta, k_1 = 1,\ \dots\ ,n_\omega$ (as seen in (26)). For k_1, k_1+1, k_1+2, with $k_1 = 1,\ \dots,k^{r_1,j_1}_{\max}$, where $k^{r_1,j_1}_{\max} \le (n-r_1-\beta-\omega)/\omega, k^{r_1,j_1}_{\max} \in \mathbb{N}$, we find all nonzero terms of (29) expressed as follows:

$$x^{(r_1,j_1,n)}_{\omega(k_1-1)+r_1+\beta} + \gamma^2 x^{(r_1,j_1,n)}_{\omega(k_1+1)+r_1+\beta} = \mu^{(1)}_{r_1,j_1,n_\omega(\omega-\beta)} x^{(r_1,j_1,n)}_{\omega k_1+r_1+\beta}.$$

Explicitly, we deduce the following:

$$\gamma^{-(\omega(k_1-1)+r_1+\beta)}\sin\left((\omega(k_1-1)+r_1+\beta)\,\theta_{j_1,n_\omega}\right)$$
$$+\gamma^2\gamma^{-(\omega(k_1+1)+r_1+\beta)}\sin\left((\omega(k_1+1)+r_1+\beta)\,\theta_{j_1,n_\omega}\right)=2\gamma\cos\left(\theta_{j_1,n_\omega}\right)\gamma^{-(\omega k_1+r_1+\beta)}\sin\left((\omega k_1+r_1+\beta)\,\theta_{j_1,n_\omega}\right),$$

or

$$\sin\left((\omega(k_1-1)+r_1+\beta)\,\theta_{j_1,n_\omega}\right)+\sin\left((\omega(k_1+1)+r_1+\beta)\,\theta_{j_1,n_\omega}\right)=2\cos\left(\theta_{j_1,n_\omega}\right)\sin\left((\omega k_1+r_1+\beta)\,\theta_{j_1,n_\omega}\right),$$

which is satisfied because of the trigonometric identity as follows:

$$\sin(\gamma_1)+\sin(\gamma_2)=2\cos\left(\frac{\gamma_1-\gamma_2}{2}\right)\sin\left(\frac{\gamma_1+\gamma_2}{2}\right).$$

For $\mathbf{x}^{(2)}_{r_2,j_2,n}$ in (25), for $k_2=1,\cdots,k^{r_2,j_2}_{\max}$, where $k^{r_2,j_2}_{\max}\le(n-r_2-\omega)/\omega, k^{r_2,j_2}_{\max}\in\mathbb{N}$, taking into account (29), we find the following:

$$x^{(r_2,j_2,n)}_{\omega(k_2-1)+r_2}+\gamma^2x^{(r_2,j_2,n)}_{\omega(k_2+1)+r_1}=\mu^{(2)}_{r_2,j_2,(n_\omega+1)\beta}x^{(r_2,j_2,n)}_{\omega k_2+r_2},$$

and this is proven for the case $\mu^{(1)}_{r_1,j_1,n_\omega(\omega-\beta)}$ and $\mathbf{x}^{(1)}_{r_1,j_1,n}$ described above.

As last step, we show that the relationships in (30) and (32) are true. For $\mathbf{x}^{(1)}_{r_1,j_1,n}$ in (24), the nonzero components have indices of the form $\omega(k_1-1)+r_1+\beta, k_1=1,\ldots,n_\omega$ (as seen in (26)). For $k_1=n_\omega$, we find $n+r_1-\omega$, and for $k_2=n_\omega-1$, we have $n+r_1-2\omega$, which are the only two nonzero components that match (30) and (32), namely,

$$x^{(r_1,j_1,n)}_{n+r_1-2\omega}=\mu^{(1)}_{r_1,j_1,n_\omega(\omega-\beta)}x^{(r_1,j_1,n)}_{n+r_1-\omega}. \tag{36}$$

More in detail, we infer that

$$\begin{aligned}\gamma^{-(n_\omega-1)}\sin\left((n_\omega-1)\theta_{j_1,n_\omega}\right)&=2\gamma\cos\left(\theta_{j_1,n_\omega}\right)\gamma^{-n_\omega}\sin\left(n_\omega\theta_{j_1,n_\omega}\right),\\ \sin\left((n_\omega-1)\theta_{j_1,n_\omega}\right)&=2\cos\left(\theta_{j_1,n_\omega}\right)\sin\left(n_\omega\theta_{j_1,n_\omega}\right),\\ \sin\left((n_\omega-1)\frac{j_1\pi}{n_\omega+1}\right)&=2\cos\left(\frac{j_1\pi}{n_\omega+1}\right)\sin\left(n_\omega\frac{j_1\pi}{n_\omega+1}\right).\end{aligned} \tag{37}$$

Furthermore, because

$$\begin{aligned}\sin\left((n_\omega-1)\frac{j_1\pi}{n_\omega+1}\right)&=\sin\left(j_1\pi-2\frac{j_1\pi}{n_\omega+1}\right)=(-1)^{j_1+1}\sin\left(2\frac{j_1\pi}{n_\omega+1}\right),\\ \sin\left(n_\omega\frac{j_1\pi}{n_\omega+1}\right)&=\sin\left(j_1\pi-\frac{j_1\pi}{n_\omega+1}\right)=(-1)^{j_1+1}\sin\left(\frac{j_1\pi}{n_\omega+1}\right),\end{aligned}$$

we deduce that relation (37) is equivalent to $\sin\left(2\theta_{j_1,n_\omega}\right)=2\cos\left(\theta_{j_1,n_\omega}\right)\sin\left(\theta_{j_1,n_\omega}\right)$, which is an identity, because of the basic relation in (35). Equivalently, the latter is true for $\mu^{(2)}_{r_2,j_2,(n_\omega+1)\beta}$ in (23) and for $\mathbf{x}^{(2)}_{r_2,j_2,n}$ in (25). □

2.3 | The real symmetric SST Toeplitz case: the generating function and a simplified distribution function

We now consider the previous results from the point of view of spectral distributions in the sense of Weyl. First, we introduce some notations and definitions concerning the general sequences of matrices. For any function F defined on the complex field and for any matrix A_n of size d_n, by the symbol $\Sigma_\lambda(F,A_n)$, we denote the following means:

$$\frac{1}{d_n}\sum_{j=1}^{d_n}F\left[\lambda_j(A_n)\right].$$

Moreover, given a sequence $\{A_n\}$ of matrices of size d_n with $d_n<d_{n+1}$ and given a Lebesgue-measurable function ψ defined over a measurable set $K\subset\mathbb{R}^\nu$, $\nu\in\mathbb{N}^+$, of finite e positive Lebesgue measure $\mu(K)$, we say that $\{A_n\}$ is distributed as (ψ,K) in the sense of the eigenvalues if for any continuous F with bounded support, the following limit relation holds:

$$\lim_{n\to\infty}\Sigma_\lambda(F,A_n)=\frac{1}{\mu(K)}\int_K F(\psi)\,\mathrm{d}\mu. \tag{38}$$

In this case, we write in short $\{A_n\} \sim_\lambda (\psi, K)$. In Remark 3, we provide an informal meaning of the notion of eigenvalue distribution.

Remark 3. The informal meaning behind the above definition is the following. If ψ is continuous, n is large enough, and

$$\left\{\mathbf{x}_j^{(m_n)}, \quad j = 1, \dots, d_n\right\}$$

is an equispaced grid on K, then a suitable ordering $\lambda_j(A_n), j = 1, \dots, d_n$, of the eigenvalues of A_n is such that the pairs $\left\{\left(\mathbf{x}_j^{(d_n)}, \lambda_j(A_n)\right), \quad j = 1, \cdots, m_n\right\}$ reconstruct approximately the hypersurface as follows:

$$\{(\mathbf{x}, \psi(\mathbf{x})), \ \mathbf{x} \in K\}.$$

In other words, the spectrum of A_n "behaves" like a uniform sampling of ψ over K. For instance, if $\nu = 1$, $d_n = n$, and $K = [a, b]$, then the eigenvalues of A_n are approximately equal to $\psi(a + j(b-a)/n), j = 1, \dots, n$, for n large enough. Analogously, if $\nu = 2$, $d_n = n^2$, and $K = [a_1, b_1] \times [a_2, b_2]$, then the eigenvalues of A_n are approximately equal to $\psi(a_1 + j(b_1 - a_1)/n, a_2 + k(b_2 - a_2)/n), j, k = 1, \dots, n$, for n large enough.

Let f be a complex-valued (Lebesgue) integrable function, defined over $Q = (-\pi, \pi)$, and let us consider the sequence $\{T_n(f)\}$ with $T_n(f) = \left(\hat{f}_{j-k}\right)_{j,k=1}^n$, $\hat{f}_s, s \in \mathbb{Z}$ being the Fourier coefficients of f defined as in (2). The asymptotic distribution of eigenvalues and singular values of a sequence of Toeplitz matrices has been thoroughly studied in the last century (for example, see other works[1,23] and the references reported therein). The starting point of this theory, which contains many extensions and other results, is a famous theorem of Szegő,[3] which we report in the version given by Tyrtyshnikov et al.[23]

Theorem 3. *If f is integrable over Q, and if $\{T_n(f)\}$ is the sequence of Toeplitz matrices generated by f, it then holds that*

$$\{T_n^*(f)T_n(f)\} \sim_\lambda \left(|f|^2, Q\right). \tag{39}$$

Moreover, if f is also real valued, then each matrix $T_n(f)$ is Hermitian and

$$\{T_n(f)\} \sim_\lambda (f, Q). \tag{40}$$

However, a simple remark has to be added. The symbol in the Weyl sense is far from unique, and in fact, any rearrangement is still a symbol. A simple case is given by standard Toeplitz sequences $\{T_n(f)\}$, with f real valued, and even that is $f(\theta) = f(-\theta)$ almost everywhere, $\theta \in Q$. In that case, relation (40) has the following form:

$$\lim_{n\to\infty} \Sigma_\lambda(F, T_n(f)) = \frac{1}{2\pi}\int_{-\pi}^{\pi} F(f(\theta))\mathrm{d}\theta = \frac{1}{\pi}\int_0^{\pi} F(f(\theta))\,\mathrm{d}\theta, \tag{41}$$

due to the even character of f, and hence, $\{T_n(f)\} \sim_\lambda (f, Q_+)$, $Q_+ = (0, \pi)$. In fact, the grid points are not searched in the big interval Q but in the restricted interval Q_+ (see Remark 3).

However, formula (19) in Theorem 1 seems to be confusing, because the generating function is $g_\omega(\theta) = a_0 + 2a_\omega \cos(\omega\theta)$, whereas the eigenvalues result in an equispaced sampling of the function $a_0 + 2|a_\omega| \cos(\theta)$. Because Theorem 3 tells one that $\{T_n(g_\omega)\} \sim_\lambda (g_\omega, Q)$, whereas our explicit computation implies $\{T_n(g_\omega)\} \sim_\lambda (g_1, Q_+)$, it follows that g_1 on Q_+ is a rearrangement of g_ω on Q. Indeed, the latter is true, as demonstrated in the following simple derivations:

$$\begin{aligned}
\int_{-\pi}^{\pi} F(g_\omega(\theta))\,\mathrm{d}\theta &= \int_0^{2\pi} F(g_\omega(\theta))\,\mathrm{d}\theta \\
&= \omega\int_0^{2\pi/\omega} F(g_\omega(\theta))\,\mathrm{d}\theta \\
&= \omega\int_0^{2\pi} F(g_\omega(s/\omega))\,\mathrm{d}s/\omega \\
&= \int_0^{2\pi} F(g_1(s))\,\mathrm{d}s = 2\int_0^{\pi} F(g_1(s))\,\mathrm{d}s.
\end{aligned}$$

By the way, the fact that g_1 has exactly two branches, one monotonically increasing on $(0, \pi/2)$ and the other monotonically decreasing on $(\pi/2, \pi)$, represents a qualitative confirmation of the fact that the grid $\tilde{\theta}_n$ in (18), for the exact eigenvalue formulae, is obtained by the merging of exactly two distinct grids, $\tilde{\theta}^{(1)}_{n_\omega(\omega-\beta)}$ and $\tilde{\theta}^{(2)}_{(n_\omega+1)\beta}$, independently of the parameter ω.

3 | THE REAL SYMMETRIC SST CASE AND ITS USE IN THE GENERAL SYMMETRIC BANDED TOEPLITZ CASE

Let A_n be a Toeplitz matrix of order n and let $\hat{\omega} < n$ be a positive integer as follows:

$$A_n = \begin{bmatrix} a_0 & a_1 & \cdots & a_{\hat{\omega}} & & & \\ a_1 & a_0 & \ddots & & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & & \ddots & \\ a_{\hat{\omega}} & & \ddots & \ddots & \ddots & & \ddots \\ & \ddots & & \ddots & \ddots & \ddots & a_{\hat{\omega}} \\ & & \ddots & & \ddots & \ddots & \vdots \\ & & & \ddots & & \ddots & a_0 \; a_1 \\ & & & & a_{\hat{\omega}} & \cdots & a_1 \; a_0 \end{bmatrix}, \tag{42}$$

where the coefficients a_k, $k = 0, \ldots, \hat{\omega}$, are real numbers.

We now show that the behavior of the spectrum of such matrices can be qualitatively described via the spectral behavior of two different types of matrices: matrices of the form in (4) with different $\omega = 2, \ldots, \hat{\omega}$ and with $a_0, a_\omega = a_{-\omega}$, real numbers, and matrices of the form (42) with monotone generating function f on $[0, \pi]$, as the case of $f(\theta) = (2 - 2\cos(\theta))^2$. We observe that the case $f(\theta) = (2 - 2\cos(\theta))^2$ corresponds to the choice of $q = 2$ with $a_0 = 6, a_1 = -4, a_2 = 1$ and that for such a case, an expansion similar to that in (3) holds. We remind that expansions as in (3) are observed in other works[9,15] (and formally proven under mild assumptions[15]) for the general case, in which the generating function is a monotone cosine polynomial in $[0, \pi]$.

In Section 3.1, we compare the generating function $g_\omega(\theta) = 2 - 2\cos(\omega\theta)$ with the spectrum of matrices of the form in (4) with different $\omega = 2, \ldots, q$ and with $a_0, a_\omega = a_{-\omega}$, real numbers, by proving the expansions in (44).

In Section 3.2, for a general matrix of the form (42), we show numerical evidences that a qualitative comparison between the eigenvalues and the generating function is described either by an expansion like (3), characterizing the monotone case, or by an expansion like (44), characterizing the purely oscillatory case as $g_\omega(\theta) = 2 - 2\cos(\omega\theta)$, $\omega = 2, \ldots, q$. From a computational viewpoint, as explained by Ekström et al.,[9] the crucial observation is that such a qualitative behavior turns out to be the theoretical key for designing fast extrapolation-type algorithms for computing eigenvalues of large matrices of the form reported in (42).

3.1 | The real symmetric SST Toeplitz case: eigenvalues and generating function

Typically, a correct symbol and grid combination, which together exactly samples the eigenvalues of a given matrix, is not known, but the error can be reconstructed in some cases; see the work of Ekström et al.[9]

When approximating the eigenvalues for the standard nonmonotone symbol as follows:

$$f(\theta) = g_\omega(\theta) = 2 - 2\cos(\omega\theta), \tag{43}$$

with $1 < \omega$ fixed with respect to n, and sampling $g_\omega(\cdot)$ at the standard equispaced grid of (8), we obtain the exact eigenvalues plus an error. This error can be expressed analytically, because the eigenvalues are given by Theorem 1. Subsequently, we furnish an expression for the expansion of such an error (refer also to the work of Ekström et al.[9] for similar expansions in the monotone case).

We begin by defining the permutations $\pi_n, \sigma_n : \{1, \ldots, n\} \to \{1, \ldots, n\}$ such that $g(\tilde{\theta}_{\pi_n(1),n}) \leq \cdots \leq g(\tilde{\theta}_{\pi_n(n),n})$, $f(\theta_{\sigma_n(1),n}) \leq \cdots \leq f(\theta_{\sigma_n(n),n})$. We denote $\mu_{j,n} = g(\tilde{\theta}_{j,n})$, $\lambda_{j,n} = g(\tilde{\theta}_{\pi_n(j),n})$, and $\nu_{j,n} = f(\theta_{j,n})$, $\xi_{j,n} = g(\theta_{\sigma_n(j),n})$.

The error for (43) with sampling grid (8) to approximate the eigenvalues after sorting is thus

$$E_{j,n} = g\big(\tilde{\theta}_{\pi_n(j),n}\big) - f\big(\theta_{\sigma_n(j),n}\big) = \lambda_{j,n} - \xi_{j,n}. \tag{44}$$

This error is shown, for example, in Figure 1(a)–(c) in light gray for $\omega = 3$. At first glance, this error can seem chaotic, but it is clear numerically that in this case, and for any $1 < \omega < n$, there will be ω^2 different "error modes"; ω different modes for any fixed $\beta = \mathrm{mod}(n, \omega) \in \{0, \ldots, \omega - 1\}$. Indeed, for each β, we will denote the different error modes by $s = 0, \ldots, \omega - 1$. In Figure 1(a)–(c), these modes are shown for $\beta = 0, 1, 2$, $s = 0$ yellow (dotted), $s = 1$ blue (solid), and $s = 2$ red (dashed). Each error mode for a given n and β is given by the indices $j_s \in I_s, s = 0, \ldots, \omega - 1$, where $I_s = \{s, s + \omega, s + 2\omega, \ldots\}$ (except for $s = 0$ where $I_0 = \{\omega, 2\omega, \ldots\}$), and the union of all I_s is the whole set of indices $\{1, \ldots, n\}$. In other words, $s = \mathrm{mod}(j, \omega)$ for $j = 1, \ldots, n$, and for $s = 0$, we have $j_0 = j_\omega \omega$, $j_\omega = 1, \ldots, n_\omega$ and $s > 0$,

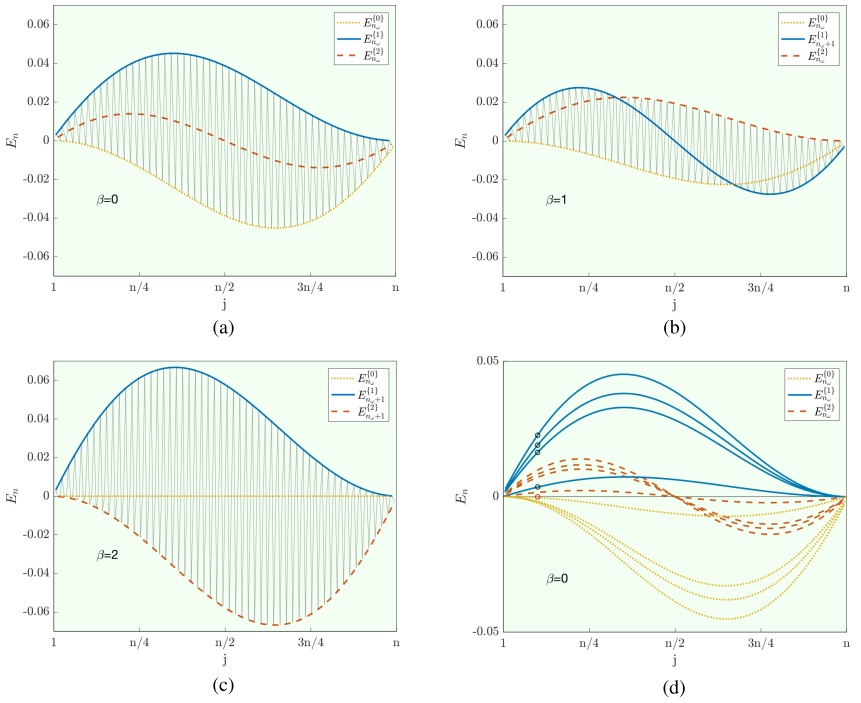


FIGURE 1 Errors for eigenvalue approximations for matrices of different sizes with standard symbol $g_3(\theta) = 2 - 2\cos(3\theta)$ and grids $\theta_{j,n} = j\pi h, j = 1, \dots, n, h = 1/(n+1)$. For each $\beta = \text{mod}(n, \omega) = \text{mod}(n, 3)$, there exist $\omega = 3$ different error modes $E^{\{i\}}_{n_\omega+\eta}, i = 0, 1, 2$, represented in yellow (dotted), blue (solid), and red (dashed). In gray, we show the errors not separated into different error modes. In panel (d), the error reduction for $g_3(\theta) = 2 - 2\cos(3\theta)$ for $\bar{\theta} = \pi/10$ is reported, by using the algorithm presented by Ekström et al.[9] (a) n = 159, $\beta = 0$. (b) n = 160, $\beta = 1$. (c) n = 161, $\beta = 2$. (d) Estimation of $c_{k,0}, k = 1, 2, 3; \bar{\theta} = \pi/10, \beta = 0$

$j_s = s + (j_\omega - 1)\omega, j_\omega = 1, \dots n_\omega + \eta$, where $n_\omega = (n - \beta)/\omega$ and $\eta = 1$ for $s = 1, \dots, \beta$, and otherwise, $\eta = 0$. In this setting, there exist functions $c_{k,s}(\cdot), s = 0, 1, \dots, \omega - 1, k \geq 1$ for which the following error:

$$E_{j_s,n} = g\big(\tilde{\theta}_{\pi_n(j_s),n}\big) - f\big(\theta_{\sigma_n(j_s),n}\big) = \lambda_{j_s,n} - \xi_{j_s,n} = \lambda^{\{s\}}_{j_\omega,n_\omega+\eta} - \xi^{\{s\}}_{j_\omega,n_\omega+\eta} = E^{\{s\}}_{j_\omega,n_\omega+\eta} \tag{45}$$

has the following form:

$$E^{\{s\}}_{j_\omega,n_\omega+\eta} = \sum_{k=1}^{\infty} c_{k,s}\left(\theta_{\sigma_n(j_s),n}\right) h^k, \quad h \frac{1}{n+1}. \tag{46}$$

We will refer to the functions $c_{k,s}(\theta), k = 1, 2, \dots, s = 0, 1, \dots, \omega - 1$ as higher order symbols.

Example 1. As a demonstrative example, we will look at the symbol $f_3(\theta) = 2 - 2\cos(3\theta)$. We have $n = 12$, and because $\omega = 3$, we have $\beta = 0$ and $n_\omega = 4$. Because $\beta = 0$ is the simplest case where $\tilde{\theta}_n = \tilde{\theta}^{(1)}_n$, which consists of $\theta_{n_\omega} = \theta_4$ repeated $\omega - \beta = 3$ times, we have the following:

$$\theta_{j_1,n_\omega} = \frac{j_1\pi}{n_\omega + 1} \quad j_1 = 1, \dots, n_\omega, \quad \theta_{j,n} = \frac{j\pi}{n+1}, \quad j = 1, \dots, n.$$

In the following table, the different evaluations are reported.

j	1	2	3	4	5	6	7	8	9	10	11	12
$f_3(\theta_{j,n}) = \nu_{j,12}$	$\nu_{1,12}$	$\nu_{2,12}$	$\nu_{3,12}$	$\nu_{4,12}$	$\nu_{5,12}$	$\nu_{6,12}$	$\nu_{7,12}$	$\nu_{8,12}$	$\nu_{9,12}$	$\nu_{10,12}$	$\nu_{11,12}$	$\nu_{12,12}$
$g(\tilde{\theta}_{j,n}) = \mu_{j,12}$	$\mu_{1,4}$	$\mu_{1,4}$	$\mu_{1,4}$	$\mu_{2,4}$	$\mu_{2,4}$	$\mu_{2,4}$	$\mu_{3,4}$	$\mu_{3,4}$	$\mu_{3,4}$	$\mu_{4,4}$	$\mu_{4,4}$	$\mu_{4,4}$.

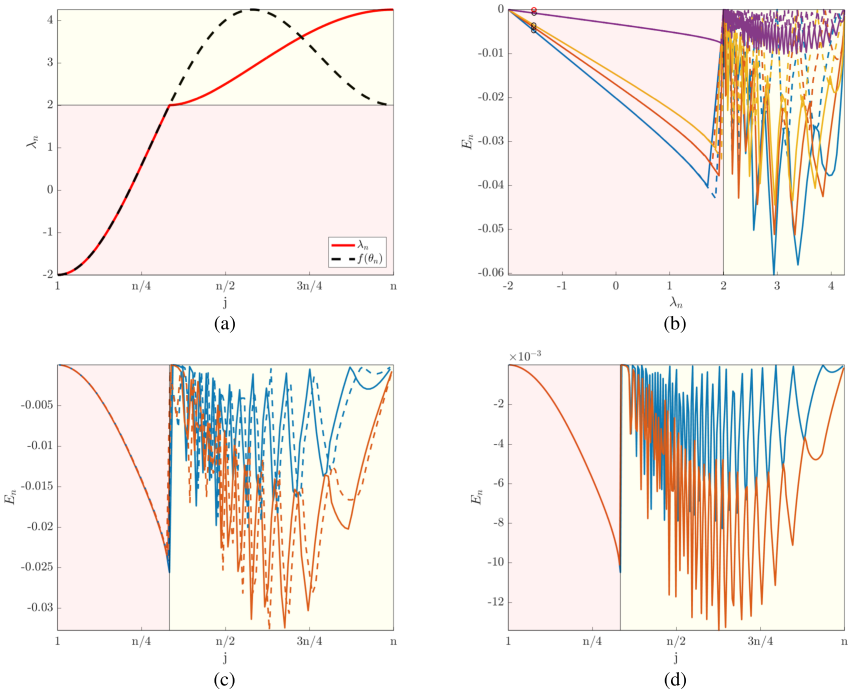


FIGURE 2 Eigenvalues, symbol, and errors for matrices with standard symbol $f(\theta) = 2 - 2\cos(\theta) - 2\cos(2\theta)$ and grids $\theta_{j,n} = j\pi h, j = 1, \ldots, n, h = 1/(n+1)$. (a) True eigenvalues (sorted, solid in red). Sampling of the symbol (unsorted, dashed in black). (b) Errors for different n. Reduction of error for $\bar{\theta} = \pi/10$. (c) Errors for n 200 (solid) and $n = 202$ (dashed). (d) Errors for $n = 500$

Sorting the evaluations of $g(\tilde{\theta}_{j,n})$ in nondecreasing order, that is, $g(\tilde{\theta}_{\pi_n(j),n})$, we will have the true eigenvalues as follows:

$$\lambda_{12} = \left\{\mu_{4,4}, \mu_{4,4}, \mu_{4,4}, \mu_{3,4}, \mu_{3,4}, \mu_{3,4}, \mu_{2,4}, \mu_{2,4}, \mu_{2,4}, \mu_{1,4}, \mu_{1,4}, \mu_{1,4}\right\}.$$

By splitting the eigenvalues by the different indices in order to separate the error modes, we obtain the following:

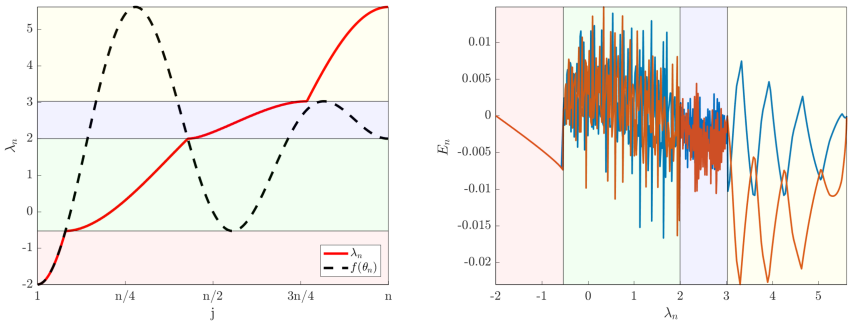
$$\begin{aligned}
\lambda_4^{\{0\}} &= \left\{\mu_{4,4}, \mu_{3,4}, \mu_{2,4}, \mu_{1,4}\right\} = \left\{\lambda_{j_0,12}\right\}, && j_0 = 3, 6, 9, 12, \quad s = \mathrm{mod}(j_0, \omega) = 0,\\
\lambda_4^{\{1\}} &= \left\{\mu_{4,4}, \mu_{3,4}, \mu_{2,4}, \mu_{1,4}\right\} = \left\{\lambda_{j_1,12}\right\}, && j_1 = 1, 4, 7, 10, \quad s = \mathrm{mod}(j_1, \omega) = 1,\\
\lambda_4^{\{2\}} &= \left\{\mu_{4,4}, \mu_{3,4}, \mu_{2,4}, \mu_{1,4}\right\} = \left\{\lambda_{j_2,12}\right\}, && j_2 = 2, 5, 8, 11, \quad s = \mathrm{mod}(j_2, \omega) = 2.
\end{aligned}$$

Sorting the evaluations of $f(\theta_{j,n})$ in a nondecreasing order, that is, $f(\theta_{\sigma_n(j),n})$, we will have the approximations of the eigenvalues as follows:

$$\xi_{12} = \left\{\nu_{9,12}, \nu_{8,12}, \nu_{1,12}, \nu_{10,12}, \nu_{7,12}, \nu_{2,12}, \nu_{11,12}, \nu_{6,12}, \nu_{3,12}, \nu_{12,12}, \nu_{5,12}, \nu_{4,12}\right\}.$$

By splitting the approximations of the eigenvalues by the different indices for separating the error modes, we find the following:

$$\begin{aligned}
\xi_4^{\{0\}} &= \left\{\nu_{1,12}, \nu_{2,12}, \nu_{3,12}, \nu_{4,12}\right\} = \left\{\xi_{j_0,12}\right\}, && j_0 = 3, 6, 9, 12, \quad s = \mathrm{mod}(j_0, \omega) = 0,\\
\xi_4^{\{1\}} &= \left\{\nu_{9,12}, \nu_{10,12}, \nu_{11,12}, \nu_{12,12}\right\} = \left\{\xi_{j_1,12}\right\}, && j_1 = 1, 4, 7, 10, \quad s = \mathrm{mod}(j_1, \omega) = 1,\\
\xi_4^{\{2\}} &= \left\{\nu_{8,12}, \nu_{7,12}, \nu_{6,12}, \nu_{5,12}\right\} = \left\{\xi_{j_2,12}\right\}, && j_2 = 2, 5, 8, 11, \quad s = \mathrm{mod}(j_2, \omega) = 2.
\end{aligned}$$



(a) (b)

FIGURE 3 Eigenvalues, symbol, and errors for a matrix with standard symbol $f(\theta) = 2 - 2\cos(3\theta) - 2\cos(4\theta)$ and grids $\theta_{j,n} = j\pi h, j = 1, \ldots, n, h = 1/(n+1)$. (a) True eigenvalues (sorted, solid in red). Sampling of the symbol (unsorted, dashed in black). (b) Errors for $n = 1000$

Hence, we have ω different error modes for $\omega = 3$ and $\beta = 0$, which are given by the following:

$$E^{\{0\}}_{j_\omega,n_\omega} = g\big(\theta_{n_\omega+1-j_\omega,n_\omega}\big) - f_3\big(\theta_{j_\omega,n}\big) = g\big(\theta_{5-j_\omega,4}\big) - f_3\big(\theta_{j_\omega,12}\big), \qquad j_\omega = 1, \ldots, 4, \tag{47}$$

$$E^{\{1\}}_{j_\omega,n_\omega} = g\big(\theta_{n_\omega+1-j_\omega,n_\omega}\big) - f_3\big(\theta_{j_\omega+2n_\omega,n}\big) = g\big(\theta_{5-j_\omega,4}\big) - f_3\big(\theta_{j_\omega+8,12}\big), \qquad j_\omega = 1, \ldots, 4, \tag{48}$$

$$E^{\{2\}}_{j_\omega,n_\omega} = g(\theta_{n_\omega+1-j_\omega,n_\omega}) - f_3(\theta_{2n_\omega+1-j_\omega,n}) = g(\theta_{5-j_\omega,4}) - f_3(\theta_{9-j_\omega,12}), \qquad j_\omega = 1, \ldots, 4, \tag{49}$$

because $\eta = 0$ in (46) for all $s = 0, 1, 2$, and because $\beta = 0$. Using the algorithm presented by Ekström et al.,[9] we look at a specific eigenvalue of interest $\bar{\theta} = \pi/10$. By this, we mean that for a matrix of size n, the index of the eigenvalue of interest, when they are sorted in nondecreasing order, $\bar{j}$, is found by $\pi/10 = \bar{j}\pi/(n+1)$. The error is then specifically $E_{\bar{j},n} = \lambda_{\bar{j},n} - \xi_{\bar{j},n}$ or $E^{\{1\}}_{\bar{j}_\omega,n_\omega}$ because $\beta = 0$ for all n of interest in this example. We look specifically at the pairs $(j_1, n_1) = (16, 159), (j_2, n_2) = (19, 189), (j_3, n_3) = (22, 219)$, and $(j, n) = (100, 999)$, which are presented in Figure 1(d). The light green background indicates that the derivative of the symbol changes sign two times in the region. Other examples of a different number of sign changes are presented in Figures 2 and 3. Because $s = \text{mod}(j, \omega) = 1$, the error will have the following expression:

$$E^{\{1\}}_{j_\omega,n_\omega} = g\big(\theta_{n_\omega+1-j_\omega,n_\omega}\big) - f_3\big(\theta_{j_\omega+2n_\omega,n}\big), \qquad j_\omega = 1, \ldots, n_\omega, \tag{50}$$

given by (48). We now look at a specific j_ω, namely $\bar{j}_\omega = (n_\omega + 7)/10$. Hence, the pairs for each error mode are instead $(\bar{j}_\omega, n_\omega)$, that is, $(6, 53), (7, 63), (8, 73)$, and $(34, 333)$. Explicitly, we obtain the following:

$$E^{\{1\}}_{\bar{j}_\omega,n_\omega} = g\big(\theta_{n_\omega+1-\bar{j}_\omega,n_\omega}\big) - f_3\big(\theta_{\bar{j}_\omega+2n_\omega,n}\big) = \sum_{k=1}^{\infty} c_{k,1}\left(\bar{\theta}\right) h^k, \quad h = \frac{1}{n+1}, \tag{51}$$

and we can analytically express the constants $c_{k,1}(\bar{\theta})$. More in detail, we have the following:

$$\begin{aligned} E^{\{1\}}_{\bar{j}_\omega,n_\omega} &= g\big(\theta_{n_\omega+1-\bar{j}_\omega,n_\omega}\big) - f_3\big(\theta_{\bar{j}_\omega+2n_\omega,n}\big) \\ &= g\left(\frac{3\pi}{10}\frac{3n_\omega+1}{n_\omega+1}\right) - f_3\left(\frac{7\pi}{10}\right) \\ &= 2\cos\left(\frac{\pi}{10}\right) - 2\cos\left(\pi\frac{\bar{j}_\omega}{n_\omega+1}\right). \end{aligned} \tag{52}$$

Explicitly, the errors in this example in Figure 1(d), denoted by black circles, are as follows:

$$E^{\{1\}}_{6,53} = 2\cos\left(\frac{\pi}{10}\right) - 2\cos\left(\frac{6\pi}{54}\right), \quad E^{\{1\}}_{7,63} = 2\cos\left(\frac{\pi}{10}\right) - 2\cos\left(\frac{7\pi}{64}\right),$$
$$E^{\{1\}}_{8,73} = 2\cos\left(\frac{\pi}{10}\right) - 2\cos\left(\frac{8\pi}{74}\right), \quad E^{\{1\}}_{34,333} = 2\cos\left(\frac{\pi}{10}\right) - 2\cos\left(\frac{34\pi}{334}\right),$$

and the latter relations are verified numerically to machine precision. The red circle in Figure 1(d) shows the error after applying the algorithm of Ekström et al.[9]: it reduces from $3.518 \cdot 10^{-3}$ to $-2.826 \cdot 10^{-8}$. By reformulating (52), we deduce the following:

$$E^{\{1\}}_{\tilde{j}_\omega, n_\omega} = 2\cos\left(\frac{\pi}{10}\right) - 2\cos\left(\frac{\pi}{10} + \frac{9\pi h}{5(1+2h)}\right), \tag{53}$$

and by the Taylor expansion of the error (53), we derive exactly the constants $c_{k,1}$ in (51), that is,

$$\begin{aligned}
E^{\{1\}}_{\tilde{j}_\omega, n_\omega} &= 2\cos\left(\frac{\pi}{10}\right) - \left(2\cos\left(\frac{\pi}{10}\right) + 2\sum_{k=1}^{\infty} \frac{\cos^{(k)}(\pi/10)}{k!}\left(\frac{9\pi h}{5(1+2h)}\right)^k\right) \\
&= -2\sum_{k=1}^{\infty} \frac{\cos^{(k)}(\pi/10)}{k!}\left(\frac{9\pi}{5}\right)^k h^k \left(\frac{1}{1+2h}\right)^k \\
&= -2\sum_{k=1}^{\infty} \frac{\cos^{(k)}(\pi/10)}{k!}\left(\frac{9\pi}{5}\right)^k h^k \left(\sum_{l=0}^{\infty}(-2h)^l\right)^k \\
&= -2\sum_{k=1}^{\infty} \frac{\cos^{(k)}(\pi/10)}{k!}\left(\frac{9\pi}{5}\right)^k \left(\sum_{l=0}^{\infty}(-2)^l h^{l+1}\right)^k \\
&= 2\sin(\pi/10)\left(\frac{9\pi}{5}\right)\sum_{l=0}^{\infty}(-2)^l h^{l+1} + \\
&\quad + \cos(\pi/10)\left(\frac{9\pi}{5}\right)^2\left(\sum_{l=0}^{\infty}(-2)^l h^{l+1}\right)^2 - \\
&\quad - \frac{\sin(\pi/10)}{3}\left(\frac{9\pi}{5}\right)^3\left(\sum_{l=0}^{\infty}(-2)^l h^{l+1}\right)^3 - \\
&\quad - \underbrace{2\sum_{k=4}^{\infty} \frac{\cos^{(k)}(\pi/10)}{k!}\left(\frac{9\pi}{5}\right)^k \left(\sum_{l=0}^{\infty}(-2)^l h^{l+1}\right)^k}_{\mathcal{O}(h^4)}.
\end{aligned} \tag{54}$$

By expanding the expression in (53) up to a $\mathcal{O}(h^4)$ term, we deduce precise representations for $c_{k,1}, k = 1, 2, 3$, that is,

$$\begin{aligned}
E^{\{1\}}_{\tilde{j}_\omega, n_\omega} &= 2\sin(\pi/10)\left(\frac{9\pi}{5}\right)\underbrace{\left(h - 2h^2 + 4h^3 + \sum_{l=3}^{\infty}(-2)^l h^{l+1}\right)}_{h-2h^2+4h^3+\mathcal{O}(h^4)} + \\
&= + \cos(\pi/10)\left(\frac{9\pi}{5}\right)^2\underbrace{\left(h - 2h^2 + \sum_{l=3}^{\infty}(-2)^l h^{l+1}\right)^2}_{h^2-4h^3+\mathcal{O}(h^4)} - \\
&\quad - \frac{\sin(\pi/10)}{3}\left(\frac{9\pi}{5}\right)^3\underbrace{\left(h + \sum_{l=2}^{\infty}(-2)^l h^{l+1}\right)^3}_{h^3+\mathcal{O}(h^4)} + \mathcal{O}(h^4).
\end{aligned}$$

TABLE 1 Analytical $c_{k,1}(\bar{\theta})$, and the corresponding approximation $\tilde{c}_{k,1}(\bar{\theta})$, for m different coarse matrices in algorithm from Ekström et al.[9] for $g_3(\theta) = 2 - 2\cos(3\theta)$, $\bar{\theta} = \pi/10$

	$m = 1$	$m = 2$	$m = 3$	$m = 4$
	159	**159,189**	**159,189,219**	**159,189,219,249**
$c_{1,1}(\bar{\theta})$	3.49489987	3.49489987	3.49489987	3.49489987
$\tilde{c}_{1,1}(\bar{\theta})$	3.63644656	3.49891734	3.49495321	3.49490028
$c_{2,1}(\bar{\theta})$		23.42262738	23.42262738	23.42262738
$\tilde{c}_{2,1}(\bar{\theta})$		22.00467555	23.39212062	23.42229454
$c_{3,1}(\bar{\theta})$			−126.29647972	−126.29647972
$\tilde{c}_{3,1}(\bar{\theta})$			−120.50951417	−126.19491717
$E^{\{1\}}_{34,333}$	$3.51819657 \cdot 10^{-3}$	$3.51819657 \cdot 10^{-3}$	$3.51819657 \cdot 10^{-3}$	$3.51819657 \cdot 10^{-3}$
$\sum_{k=1}^{m} \tilde{c}_{k,1}(\bar{\theta})h^k$	$3.63644656 \cdot 10^{-3}$	$3.52092202 \cdot 10^{-3}$	$3.51822482 \cdot 10^{-3}$	$3.51819673 \cdot 10^{-3}$
$E^{\{1\}}_{34,333} - \sum_{k=1}^{m} \tilde{c}_{k,1}(\bar{\theta})h^k$	$-1.18249995 \cdot 10^{-4}$	$-2.72544868 \cdot 10^{-6}$	$-2.82554797 \cdot 10^{-8}$	$-0.16133076 \cdot 10^{-9}$

Thus, we have the following:

$$E^{\{1\}}_{\tilde{j}_\omega, n_\omega} = \underbrace{2\sin(\pi/10)\left(\frac{9\pi}{5}\right)}_{c_{1,1}(\bar{\theta})\approx 3.49489987} h + \underbrace{\left(-4\sin(\pi/10)\left(\frac{9\pi}{5}\right) + \cos(\pi/10)\left(\frac{9\pi}{5}\right)^2\right)}_{c_{2,1}(\bar{\theta})\approx 23.42262738} h^2 + \underbrace{\left(8\sin(\pi/10)\left(\frac{9\pi}{5}\right) - 4\cos(\pi/10)\left(\frac{9\pi}{5}\right)^2 - \frac{\sin(\pi/10)}{3}\left(\frac{9\pi}{5}\right)^3\right)}_{c_{3,1}(\bar{\theta})\approx -126.29647972} h^3 + \sum_{k=4}^{\infty} c_{k,1}(\bar{\theta})h^k. \tag{55}$$

Note that the explicit expressions of (55) can be derived for any combination of n, ω, and $\bar{\theta}$, but the computation will be more complicated if $\beta > 0$ because also $\tilde{\theta}^{(2)}$ has to be considered.

In Table 1, we show the results using the algorithm of Ekström et al.[9] to approximate m different constants $c_{k,1}(\bar{\theta})$ with the same number of different coarse matrices. As m increases, $\tilde{c}_{k,1}(\bar{\theta})$ converges to $c_{k,1}(\bar{\theta})$ as expected. Using the analytical expression of $c_{k,1}(\bar{\theta})$ in (55), we have $\sum_{k=1}^{3} c_{k,1}(\bar{\theta})h^k = 3.51819620 \cdot 10^{-3}$, and thus, the error after the error reduction is $E^{\{1\}}_{34,333} - \sum_{k=1}^{m} c_{k,1}(\bar{\theta})h^k = 3.67020511 \cdot 10^{-10}$.

In Table 2, we show the results obtained when using the algorithm by Ekström et al.[9] for nonmonotone cases $g_\omega(\theta) = 2 - 2\cos(\omega\theta)$ for $\omega = 2, 3, 4$: the goal is to reduce the error of the eigenvalue approximation when considering the largest matrix. The errors for $m = 0, 1, 2, 3$ different coarse matrices used to approximate the constants $c_{k,1}(\bar{\theta})$, $k = 1, \dots, m$, are presented. For $g_2(\theta)$, the coarse matrices have sizes belonging to $\{149, 189, 209\}$, and the largest size is $n = 9999$; for $g_3(\theta)$, the coarse matrices have sizes belonging to $\{159, 189, 219\}$ and $n = 10009$; for $g_4(\theta)$, the coarse matrices have sizes belonging to $\{169, 209, 249\}$ and $n = 10009$. The errors behave as expected, and hence, the algorithm taken from the work of Ekström et al.[9] can also be used for these specific nonmonotone examples, although in this setting, a numerical computation is not necessary because the exact eigenvalues can be evaluated exactly by exploiting the symbol and sampling the grid described in Section 2.

3.2 | The general symmetric banded case: conjectures and numerics

As we have seen in the previous subsection, given a positive integer $\omega \geq 2$ and the nonmonotone symbol $f(\theta) = g_\omega(\theta) = 2 - 2\cos(\omega\theta)$, and evaluating it at a equidistant grid such as $\theta_{j,n} = j\pi h, j = 1, \dots, n, h = 1/(n+1)$, numerical tests show that the error $E_n = \lambda_n - \xi_n$ can be separated into ω different types of error modes for each $\beta = \text{mod}(n, \omega)$. That is, for each $\beta = \text{mod}(n, \omega)$, there are ω disjoint subgrids of the original grid (see Figure 1 for $\omega = 3$ and the related caption). For a given n and β, each error mode is obtained by the indices $j \in I_s, s = 0, \dots, \omega - 1$, where $I_0 = \{\omega, 2\omega, 3\omega, \dots\}$ and for $s > 0$, $I_s = \{s, s + \omega, s + 2\omega, \dots\}$, and the union of all I_s forms the whole set of indices $\{1, \dots, n\}$.

TABLE 2 Errors for eigenvalue approximations for matrices with standard symbol $g\omega(\theta) = 2 - 2cos(\omega\theta)$, $\bar{\theta} = \pi/10$

		$E^{(1)}_{j_\omega,n_\omega} - \sum_{k=1}^{m} \tilde{c}_{k,1}(\bar{\theta})h^k$		
$g_\omega(\theta)$	$E^{(1)}_{j_\omega,n_\omega}$	$m = 1$	$m = 2$	$m = 3$
$g_2(\theta)$	$-3.88581714 \cdot 10^{-5}$	$4.32478954 \cdot 10^{-6}$	$-5.21177503 \cdot 10^{-8}$	$-1.12193334 \cdot 10^{-9}$
$g_3(\theta)$	$34.97240870 \cdot 10^{-5}$	$-13.92056931 \cdot 10^{-6}$	$-38.76938472 \cdot 10^{-8}$	$-5.03491210 \cdot 10^{-9}$
$g_4(\theta)$	$65.96546126 \cdot 10^{-5}$	$-7.93740842 \cdot 10^{-6}$	$-127.70747416 \cdot 10^{-8}$	$-50.14789443 \cdot 10^{-9}$

The latter remark induces the conjecture that the number of the different expansions is related to the number of sign changes of the derivative of the generating function in the basic interval $(0, \pi)$, that is, a formula of the type as follows:

$$\lambda_{j,n} = f(\theta_{\sigma_n(j),n}) + \sum_{k=1}^{m} c_{k,s}\left(\theta_{\sigma_n(j),n}\right) h^k + O(h^{m+1}), \quad j \in I_s, \quad s = 0, \ldots, \omega - 1, \tag{56}$$

may hold. In Figure 2, we see a clarifying example of the nonmonotone error given by the function $f(\theta) = 2 - 2\cos(\theta) - 2\cos(2\theta)$.

In Figure 2(a), we show the true eigenvalues (sorted, solid in red) and the sampling of the symbol (unsorted, dashed in black). The two different regions displayed in light colors (red on bottom and yellow on top) represent the different number of sign changes in the derivative of the symbol $f(\theta)$ inside the region (zero and one). These different regions will give rise to different features in the behavior of the errors.

The approximation error of the function possesses the same monotone behavior as the one observed for $(2 - 2\cos(\theta))^2$, when using, for example, the grid $(j-1)\pi/(n-1)$ instead of the exact $j\pi/(n+1)$, in the interval $[0, \pi/3]$ with $f(\pi/3) = 2$, and almost the behavior typical of $2 - 2\cos(2\theta)$ in the interval $[\pi/3, \pi]$ with $f(\pi/3) = f(\pi) = 2$. Indeed, for the eigenvalues belonging to $(-2, 2]$, $-2 = f(0) = \min f$, $2 = f(\pi/3)$, as represented in the light red regions of Figure 2, the behavior of the error is like the one related with a monotone function that (56) with $\omega = 1$ holds. For the eigenvalues belonging to $(2, 17/4)$, $2 = f(\pi/3) = f(\pi)$, $17/4 = \max f$, as represented in the light yellow regions in Figure 2, the behavior of the error behaves almost like the one displayed in (56) with $\omega = 2$, because the sign of the derivative changes once.

In Figure 2(b), we present a visualization of error reduction for $f(\theta) = 2 - 2\cos(\theta) - 2\cos(2\theta)$, $\bar{\theta} = \pi/10$ with the algorithm presented by Ekström et al.[9] The largest matrix dimension is $n = 669$, whereas the coarse grids have sizes belonging to $\{109, 129, 149\}$. The black circles represent the error of symbol approximation on the corresponding grids, and the red circle is the error on the fine grid after reduction using the coarse errors. The error is reduced from $-7.899 \cdot 10^{-4}$ to $-9.959 \cdot 10^{-11}$. Note that here, the x-axis is ordered by the size of the true eigenvalues. The error on the left region (light red) behaves like a monotone symbol, whereas the right region (light yellow) behaves, in general terms, as a symbol of the form g_ω but with a slight shift.

As seen in Figures 2(c)–(d), the local change is somewhat drastic with a small change of n, but the general structure of the error remains as n increases. In Figure 2(c), we see the errors for $n = 200$ (solid) and $n = 202$ (dashed). Assuming two error modes for each n, note the rather large "shift" of the error curve just increasing n by a factor two. Note also that the x-axis is ordered by n and not by the size of the true eigenvalues. In Figure 2(d), we see the errors for $n = 500$ assuming two error modes. Note that the general regularity of the error in the large eigenvalues (right part of the figure) is comparable to $n = 200$ and $n = 202$ shown in Figure 2(c). In other words, the global error behavior is still regular in a weaker sense and should be investigated formally.

In Figure 3, we report the case of the error using the standard grid on the symbol $f(\theta) = 2 - 2\cos(3\theta) - 2\cos(4\theta)$. In Figure 3(a), the true eigenvalues (sorted, solid red) and the sampling of the symbol (unsorted, dashed black) are shown. Clearly, four different regions are present, colored in light red, green, blue, and yellow, depending on the number of sign changes of the derivative of the symbol in the region (zero, two, three, and one). These different regions will give rise to different characteristics of the behavior of the errors.

The error $E_{j,n} = \lambda_{j,n} - f(\theta_{\sigma_n j,n})$, for $n = 1000$, was plotted as if there are two error modes, that is, $j_1 = 1, 3, 5, \ldots$ (blue) and $j_2 = 2, 4, 6, \ldots$ (red). The light red (first) region shows the error behaving as in the monotone case, that is, the error can be reconstructed in the manner presented by Ekström et al.[9] The light yellow (fourth) part shows a clear regularity when representing the error in two sets (blue and red). On the other hand, when increasing n, we do not only decrease the error in the region but also maintain the error function, and we also change the number of "peaks", as previously demonstrated in Figure 2. In the light red region, the error behaves like a monotone symbol, and the error can be efficiently reconstructed by the same techniques as described in Section 3.1 and in Figures 1 and 2. The light green (second) and

blue (third) regions show "chaotic" behavior, resulting from the "naive" ordering of the approximated eigenvalues. Again, this behavior deserves a further study.

4 | CONCLUSIONS AND FUTURE WORK

The paper contains two types of theoretical results and a numerical part.

The first result concerns the fact that for the SST Toeplitz matrices as in (4), with $a_0, a_\omega, a_{-\omega} \in \mathbb{C}$, $0 < \omega < n$, the eigenvalues and the eigenvectors have a closed form expression. In particular, the formula for the eigenvalues $\mu_{j,n}$ in Theorem 1 is expressed in an elegant and compact way, because there exist a grid $\tilde{\theta}_n$, the one defined in (18), and the simple function $g(\theta) = a_0 + 2\sqrt{a_\omega a_{-\omega}}\cos(\theta)$ such that

$$\mu_{j,n} = g\left(\tilde{\theta}_{j,n}\right), \quad j = 1, \ldots, n.$$

Furthermore, using basic changes of variable in the integral representation of the distribution results, we show clear relationships between the symbol g and the standard generating functions of the matrices A_n, A_n^{sym}, that is, $f_\omega(\theta) = a_0 + a_\omega e^{\mathbf{i}\omega\theta} + a_{-\omega}e^{-\mathbf{i}\omega\theta}$, $g_\omega(\theta) = a_0 + 2\sqrt{a_\omega a_{-\omega}}\cos(\omega\theta)$, respectively. Also, a closed form formula for the corresponding eigenvectors is presented in Theorem 2.

The second result regards three banded Toeplitz matrices (4), with $a_0, a_\omega, a_{-\omega} \in \mathbb{R}$, $0 < \omega < n$: here, we show that an asymptotic expansion of the eigenvalues holds, with respect to the standard generating function and the usual grid (see formula (44)). The latter extends a similar asymptotic expansion holding for the eigenvalues of general symmetric real Toeplitz matrices, having polynomial cosine generating function, which is monotone on $[0, \pi]$ (see formula (3) and other works[9,15,16]): an important example of such matrices is represented by the finite-difference discretization of the operators $(-1)^q \partial^{2q}/\partial x^{2q}$, whose generating function is $(2 - 2\cos(\theta))^q$, $q \geq 1$.

The final part concerns a conjecture supported by numerical tests in which it is shown that for a generic banded real symmetric Toeplitz matrix, the eigenvalue $\lambda_{j,n}$ compared with $f\left(\theta_{\sigma_n j,n}\right)$ either shows an expansion similar to formula (44) if $\lambda_{j,n} \in [m, M]$ and $f'(\theta)$ has ω changes of sign for $f(\theta) \in [m, M]$ or shows an expansion like formula (3) if $\lambda_{j,n} \in [m, M]$ and $f(\theta) \in [m, M]$ is monotone.

The latter gives the ground for extrapolation techniques[24] for computing the eigenvalues of large banded real symmetric Toeplitz matrices in a fast way. Of course, also the multidimensional and the block cases should be considered and explored in future works, owing to their importance in the numerical approximation of (systems of) partial differential equations.

ACKNOWLEDGEMENTS

The authors would like to thank their families, colleagues, and friends for fruitful discussions and insights. The research of the first author is funded by the Graduate School in Mathematics and Computing (FMB) and Uppsala University, and the second author is supported by the Italian Group of Scientific Computing INDAM-GNCS.

ORCID

S.-E. Ekström http://orcid.org/0000-0002-7875-7543
S. Serra-Capizzano http://orcid.org/0000-0001-9477-109X

REFERENCES

1. Böttcher A, Silbermann B. Introduction to large truncated Toeplitz matrices. New York: Springer-Verlag; 1999.
2. Serra-Capizzano S. On the extreme eigenvalues of Hermitian (block) Toeplitz matrices. Linear Algebra Appl. 1998;270:109–129.
3. Grenander U, Szegő G. Toeplitz forms and their applications. 2nd ed. New York: Chelsea; 1984.
4. Tilli P. A note on the spectral distribution of Toeplitz matrices. Linear Multil Algebra. 1998;45(2/3):147–159.
5. Tyrtyshnikov E. A unifying approach to some old and new theorems on distribution and clustering. Linear Algebra Appl. 1996;232:1–43.
6. Serra-Capizzano S, Bertaccini D, Golub G. How to deduce a proper eigenvalue cluster from a proper singular value cluster in the nonnormal case. SIAM J Matrix Anal Appl. 2005;27(1):82–86.
7. Bini D, Capovani M. Spectral and computational properties of band symmetric Toeplitz matrices. Linear Algebra Appl. 1983;52/53:99–126.
8. Bogoya J, Böttcher A, Maximenko E. From convergence in distribution to uniform convergence. Bol Soc Mat Mex. 2016;3(22):695–710.

9. Ekström S-E, Garoni C, Serra-Capizzano S. Are the eigenvalues of banded symmetric Toeplitz matrices known in almost closed form? Exp Math. 2017. https://doi.org/10.1080/10586458.2017.1320241
10. Serra-Capizzano S. On the extreme spectral properties of Toeplitz matrices generated by L1 functions with several minima/maxima. BIT. 1996;36(1): 135–142.
11. Stein EM, Weiss G. Introduction to Fourier analysis on Euclidean spaces. Princeton, NJ: Princeton University Press; 1971.
12. Bogoya J, Grudsky S, Maximenko E. Eigenvalues of Hermitian Toeplitz matrices generated by simple-loop symbols with relaxed smoothness. Oper Theory Adv Appl. 2017;259:179–212.
13. Ahmad F, Al-Aidarous E, Alrehaili D, Ekström S-E, Furci I, Serra-Capizzano S. Are the eigenvalues of preconditioned banded symmetric Toeplitz matrices known in almost closed form? Numer Algoritm 1–27, 2017. In press.
14. Ekström S-E, Furci I, Serra-Capizzano S. Are the Eigenvalues of the B-spline IgA approximation of $-\Delta u = \lambda u$ known in almost closed form? 2017. TR Division of Scientific Computing, IT Department, Uppsala University. http://www.it.uu.se/research/publications/reports/2017-016
15. Bogoya J, Böttcher A, Grudsky S, Maximenko E. Eigenvalues of Hermitian Toeplitz matrices with smooth simple-loop symbols. J Math Anal Appl. 2015;422(2):1308–1334.
16. Böttcher A, Grudsky S, Maximenko E. Inside the eigenvalues of certain Hermitian Toeplitz band matrices. J Comput Appl Math. 2010;233(9): 2245–2264.
17. Böttcher A, Grudsky S. Spectral properties of banded Toeplitz matrices. Philadelphia: SIAM; 2005.
18. Elliott JF. The characteristic roots of certain real symmetric matrices [master's thesis]. Knoxville: University of Tennessee; 1953. http://trace.tennessee.edu/utk_gradthes/2384
19. Gantmakher FR, Krein MG. Ostsilliatsionnye matritsy i iadra i malye kolebaniia mekhanicheskikh sistem (Oscillation matrices and kernels and small vibrations of mechanical systems). Soviet Union: Gostekhizdat; 1950.
20. Noschese S, Pasquini L, Reichel L. Tridiagonal Toeplitz matrices: Properties and novel applications. Numer Linear Algebra Appl. 2013;20(2):302–326.
21. Smith GD. Numerical solution of partial differential equations: Finite difference methods. 2nd ed. Oxford: Clarendon Press; 1978.
22. Ekström S-E, Serra-Capizzano S. Eigenvalues and eigenvectors of banded Toeplitz matrices and the related symbols. 2017. TR Division of Scientific Computing, IT Department, Uppsala University. http://www.it.uu.se/research/publications/reports/2017-010/
23. Tyrtyshnikov E, Zamarashkin N. Spectra of multilevel Toeplitz matrices: Advanced theory via simple matrix relationships. Linear Algebra Appl. 1998;270:15–27.
24. Brezinski C, Redivo Zaglia M. Extrapolation methods theory and practice, studies in computational mathematics. 2nd ed. Amsterdam: North-Holland Publishing Co; 1991.

How to cite this article: Ekström S-E, Serra-Capizzano S. Eigenvalues and eigenvectors of banded Toeplitz matrices and the related symbols. *Numer Linear Algebra Appl*. 2018;e2137. https://doi.org/10.1002/nla.2137

Colophon

This book is set in Bembo, a typeface based on an old-style Roman face that was used for Cardinal Pietro Bembo's tract De Aetna in 1495. Bembo was cut by Franceso Griffo (1450–1518) in the early sixteenth century for the Italian Renaissance printer and publisher Aldus Manutius (1449–1515). The typeface used to set the math of this book, is Computer Modern Roman, developed by Donald E. Knuth, and finalized for TEX and METAFONT in 1992. Computer Modern Roman is based on Monotype Modern 8a, and created for scientific and mathematical writing. `Bitstream Vera Sans Mono`, designed by Jim Lyles, was created for technical text in 2003, and is used to set the monospace parts. For Cyrillic script Times New Roman, developed by Stanley Morison in 1931 for the newspaper The Times, is used. The type used on the cover of this book is Berling antikva, the first Antiqua typeface developed in Sweden, and was designed by Karl-Erik Forsberg in 1951 for Berlingska Stilgjuteriet. Special glyphs are created with `fontforge`. This book is typeset using XƎLATEX. Final Art by Finn Ljunggren and Göran Wallby. Latin by Harald Nilsson. This book is printed and bound by KpH Trycksaksbolaget AB, Uppsala 2018.

Experiments is mainly made using MATLAB [50] and JULIA v.0.6.2 [10] (with BANDEDMATRICES.JL [56], ELEMENTAL.JL [53, 59], and LINEARALGEBRA.JL [54]), on an Apple MACBOOK (Retina, 12-inch, Early 2015) 1,3 GHz Intel Core M, 8 GB 1600 MHz DDR3. Both Wolfram Alpha (`https://www.wolframalpha.com`) and The On-Line Encyclopedia of Integer Sequences (OEIS) (`https://oeis.org`) [69] have been extensively utilized in this work.

The figures in this book are created using MATLAB, Ti*k*Z [71], `inkscape` [72], and MATHEMATICA [80]. On the front cover of this book is the seal of Uppsala University, represented by fifty symbols, and different choices of bandwidths. On the back cover of this book are the elements of a generic banded Toeplitz matrix. The figure on the spine of the cover, and on the bottom of this page, is an illustration by François Desprez from the 1565 edition of "Les Songes Drolatiques de Pantagruel, ou sont contenues plusieurs figures de l'invention de maistre François Rabelais: & derniere oeuvre d'iceluy, pour la recreation des bons esprits", and is kindly vectorized by Michael Olsson.

Quicumque scriptor scribit
Laeti ut scribunt scribae