Linear Algebra and its Applications 667 (2023) 10–43

Contents lists available at ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

Spectral and norm estimates for matrix-sequences arising from a finite difference approximation of elliptic operators

Armando Coco [a,c], Sven-Erik Ekström [b,*], Giovanni Russo [c], Stefano Serra-Capizzano [b,d], Santina Chiara Stissi [c,e]

[a] *School of Engineering, Computing and Mathematics, Oxford Brookes University, OX33 1HX, Oxford, UK*
[b] *Department of Information Technology, Division of Scientific Computing, Uppsala University, Lägerhyddsv. 1, Box 337, SE-751 05, Uppsala, Sweden*
[c] *Department of Mathematics and Informatics, Catania University, Viale A. Doria 6, 95125 Catania, Italy*
[d] *Department of Science and high Technology, Insubria University, via Valleggio 11, 22100 Como, Italy*
[e] *Istituto Nazionale di Geofisica e Vulcanologia, Piazza Roma 2, 95125 Catania, Italy*

ARTICLE INFO

Article history:
Received 19 August 2021
Accepted 5 March 2023
Available online 9 March 2023
Submitted by E. Tyrtyshnikov

MSC:
15A18
15A69
15B05
65N06
65N12
65N15

Keywords:
Toeplitz matrix
Generating function and spectral symbol

ABSTRACT

When approximating elliptic problems by using specialized approximation techniques, we obtain large structured matrices whose analysis provides information on the stability of the method. Here we provide spectral and norm estimates for matrix-sequences arising from the approximation of the Laplacian via ad hoc finite differences. The analysis involves several tools from matrix theory and in particular from the setting of Toeplitz operators and Generalized Locally Toeplitz matrix-sequences. Several numerical experiments are conducted, which confirm the correctness of the theoretical findings.

* Corresponding author.
E-mail address: sven-erik.ekstrom@it.uu.se (S.-E. Ekström).

https://doi.org/10.1016/j.laa.2023.03.005

1. Introduction

In the numerical approximation of elliptic differential equations, by using specialized approximation techniques, we obtain large structured matrices whose analysis provides information on the stability of the method. Here we provide spectral and norm estimates for matrix-sequences arising from the approximation of the Laplacian via ad hoc finite differences that is from the Coco–Russo method [7].

The analysis involves several tools from matrix theory and in particular from the setting of Toeplitz operators and Generalized Locally Toeplitz (GLT) matrix-sequences. Several numerical experiments are conducted, which confirm the theoretical findings.

The paper is organized as follows. Subsection 1.1 contains a motivation and a description of the Coco–Russo method, together with a brief account on the related literature. Subsection 1.2 contains the definition of the used matrix norms and the necessary tools from the Toeplitz technology, while Section 2 contains the matrix formulation in $1\mathbb{D}$ in the language of Toeplitz structures, the analysis of the norm estimates in $1\mathbb{D}$, together with related numerical experiments and a preliminary discussion on the spectral features of the involved matrix-sequences. Section 3 contains more details on the $2\mathbb{D}$ method, on its matrix formulation, on the basic tools of the GLT theory, and on spectral results of distributional type in $1\mathbb{D}$ and in $2\mathbb{D}$. Numerical examples in $2\mathbb{D}$ for variable coefficients and non-rectangular domains are presented in connection with the GLT theory, while a discussion on the more challenging case of the norm estimates in $2\mathbb{D}$ is also provided. A conclusion section ends the paper with a mention to a few open problems.

1.1. Method description and motivation

The design of numerical methods to solve Partial Differential Equations (PDE) on complex-shaped domains is obtaining an increasing interest in the scientific community. One of the bottlenecks of modern computer simulations is the modeling of physical processes around $3\mathbb{D}$ complex-shaped objects through PDE. Finite Element Methods (FEM) are well-established approaches to solve PDE and supported by rigorous theoretical analysis developed in the last decades to prove the convergence and accuracy order of the method when the grid size approaches zero.

However, some critical limitations are commonly associated in literature with FEM, especially when applied to curved boundaries. In particular, the generation of elements to conform highly varying curvatures of the boundary might become cumbersome, especially if the domain changes its shape over time. Also, the design of a balanced partition of the mesh for parallel FEM is unhandy. For these reasons, approaches based on Finite Difference Methods (FDM) where the domain is immersed into a fixed grid are increasing their popularity in literature, since they do not require any mesh generation effort and at the same time allow for a natural design of parallel solvers.

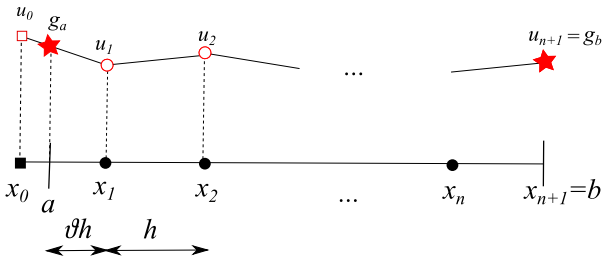

Fig. 1. Discretization of the 1𝔻 domain. Full black circles are the interior grid points, while the full square is the ghost point. Boundary values are indicated with stars. Linear extrapolation is used to define the ghost value u_0 from u_1 and the left boundary value g_a.

On the other hand, FDM are commonly based on heuristic approaches and convergence and stability analysis are not sufficiently developed in literature, especially for the case of curved boundaries.

The Immersed Boundary Method proposed by Peskin in [17] and further developed by LeVeque and Li in [14] is a pioneer approach based on FDM for general domains immersed on fixed grids.

A more recent approach is the Ghost-Fluid Method proposed by Fedkiw et al. in [8] and further extended to higher accuracy by Gibou et al. in [11,12], where the values on grid nodes just outside the domain (ghost points) are obtained by accurate extrapolations of the boundary condition from inside values.

In [7], the authors present a highly efficient and accurate ghost-point method to solve a Poisson equation on a complex-shaped domain, modeled by a level-set function. Several numerical tests were presented to confirm the accuracy order and the efficiency of the multigrid solver. However, a theoretical analysis was missing. The method has been extended to several applications, such as compressible fluids in moving domains [5] or volcanology [6].

In this paper we present a technique to prove the stability of the Coco–Russo method [7] and the convergence to the predicted order of accuracy.

We start from the 1𝔻 problem. Consider the elliptic boundary-value problem:

$$-\Delta u = f \text{ on } \Omega = (a, b), \tag{1}$$

$$u(a) = g_a, \quad u(b) = g_b, \tag{2}$$

and a one-dimensional uniform grid $\mathcal{G}_h = \{x_0, x_1, \ldots, x_{n+1}\}$ with a constant spatial step $h = x_i - x_{i-1}$, for $i = 1, \ldots, n+1$. Then, $x_i = x_0 + i\,h$. Let $x_0 < a < x_1$ and $x_{n+1} = b$ (see Fig. 1).

The elliptic equation $-\Delta u = f$ is discretized by central differences on x_i for $i = 1, \ldots, n$ and the boundary condition on $x = b = x_{n+1}$ is included in the internal discretization (this is the so called *eliminated boundary condition* approach):

$$\frac{-u_{i-1} + 2u_i - u_{i+1}}{h^2} = f_i \text{ for } i = 1, \ldots, n-1, \tag{3}$$

$$\frac{2u_n - u_{n-1}}{h^2} = f_n + \frac{g_b}{h^2}. \tag{4}$$

The boundary condition on $x = a$ is approximated by $q(a) = g_a$, where $q(x)$ is the polynomial of degree $s-1$ that interpolates u on the grid points $u_0, u_1, \ldots, u_{s-1}$. We call s the stencil size for the boundary condition on $x = a$.

The discretization of the boundary condition can be represented as:

$$\sum_{i=0}^{s-1} c_i u_i = g_a. \tag{5}$$

For $s = 2$ we have

$$\vartheta u_0 + (1-\vartheta)u_1 = g_a, \tag{6}$$

where $\vartheta = (x_1 - a)/h$. The grid point x_0 is called ghost point and u_0 is the ghost value.

Although we can follow a similar technique for the boundary condition on $x = a$ to the one that we adopted for $x = b$ (i.e. we can solve (6) for u_0 and substitute its value into the internal equation for x_1), we keep a non-eliminated boundary condition approach in order to develop a theoretical analysis that can be straightforwardly extended to higher dimensional cases, where the eliminated approach is impractical.

The discretized problem is then a linear system $A_h \mathbf{u}_h = \mathbf{f}_h$ where $A_h \in \mathbb{R}^{(n+1)\times(n+1)}$:

$$A_h \mathbf{u}_h = \begin{bmatrix} \vartheta & 1-\vartheta & 0 & \ldots & \ldots & 0 \\ -\frac{1}{h^2} & \frac{2}{h^2} & -\frac{1}{h^2} & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & -\frac{1}{h^2} & \frac{2}{h^2} & -\frac{1}{h^2} \\ 0 & 0 & \ldots & 0 & -\frac{1}{h^2} & \frac{2}{h^2} \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ \vdots \\ u_{n-1} \\ u_n \end{bmatrix} = \begin{bmatrix} g_a \\ f_1 \\ f_2 \\ \vdots \\ f_{n-1} \\ f_n + \frac{g_b}{h^2} \end{bmatrix} = \mathbf{f}_h, \tag{7}$$

where $h = (b - x_0)/(n+1)$.

1.2. Matrix theoretic notations, Toeplitz structures, and related tools

For $X \in \mathbb{C}^{m\times m}$, by $\|X\|_p$ we indicate the matrix norm induced by the l^p vector norm $\|\mathbf{y}\|_p = \left[\sum_{j=1}^m |y_j|^p\right]^{1/p}$ for $p \in [1, \infty)$ and $\|\mathbf{y}\|_\infty = \max_{1\le j\le m} |y_j|$, with $\mathbf{y} \in \mathbb{C}^m$. A further useful class of matrix norms is that of the Schatten p-norms (see [3] and references therein), where

$$\|X\|_{S,p} = \left[\sum_{j=1}^m \sigma_j(X)^p\right]^{1/p}$$

if $p \in [1, \infty)$ and $\|X\|_{S,\infty} = \sigma_1(X) = \|X\|_2$, with $\sigma_1(X) \geq \cdots \geq \sigma_m(X) \geq 0$ being the singular values of X. In other terms, the Schatten p-norm of a matrix can be viewed as the l^p norm of the vector of its singular values. The latter, in view of the singular value decomposition [3], implies that all the Schatten p-norms are unitarily invariant, that is $\|PXQ\|_{S,p} = \|X\|_{S,p}$ for every $X, P, Q \in \mathbb{C}^{m\times m}$, P, Q unitary matrices, and for all $p \in [1, \infty]$.

Let T_n be a Toeplitz matrix of order n and let $\omega < n$ be a positive integer

$$T_n = \begin{bmatrix} a_0 & \cdots & a_{-\omega} & & & \\ \vdots & \ddots & & \ddots & & \\ a_\omega & & \ddots & & \ddots & \\ & \ddots & & \ddots & & a_{-\omega} \\ & & \ddots & & \ddots & \vdots \\ & & & a_\omega & \cdots & a_0 \end{bmatrix}, \tag{8}$$

where the coefficients a_k, $k = -\omega, \ldots, \omega$, are complex numbers.

Let $f \in L^1(-\pi, \pi)$ and let $T_n(f)$ be the Toeplitz matrix generated by f i.e. $(T_n(f))_{s,t} = a_{s-t}(f)$, $s, t = 1, \ldots, n$, with f indicated as *generating function* of $\{T_n(f)\}$ and with $a_k(f)$ being the k-th Fourier coefficient of f that is

$$a_k(f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \mathrm{e}^{-\mathbf{i}ks}\, \mathrm{d}s, \quad \mathbf{i}^2 = -1, \ k \in \mathbb{Z}. \tag{9}$$

With these notations the matrix reported in (8) can be written as $T_n = T_n(f)$, where the generating function is $f(s) = \sum_{j=-\omega}^{\omega} a_j \mathrm{e}^{\mathbf{i}js}$. It is worth noticing that study of the generating function gives plenty of information on the spectrum of $T_n(f)$ for any fixed n, and also asymptotically as the matrix-size n diverges to infinity (see [9] and [10] for the multilevel setting). For instance, if f is real-valued almost everywhere (a.e.), then $T_n(f)$ is Hermitian for all n. Furthermore, when f is real-valued and even a.e., the matrix $T_n(f)$ is (real) symmetric for all n, while f real-valued and nonnegative a.e., but not identically zero a.e., implies that $T_n(f)$ is Hermitian positive definite for all n: in such a setting the considered matrix-sequence could be ill-conditioned and indeed if f is nonnegative and bounded with essential supremum equal to $M > 0$ and a unique zero of order $\alpha > 0$, then the maximal eigenvalue converges monotonically from below to M, whereas the minimal eigenvalues converges to zero monotonically from above with a speed dictated by α, that is the minimal eigenvalue is asymptotical to $n^{-\alpha}$. In many practical applications we remind that it is required to solve numerically linear systems of Toeplitz kind and of (very) large dimensions and hence several specialized techniques of iterative type, such as preconditioned Krylov methods and ad hoc multigrid procedures have been designed; we refer the interested reader to the books [4,16] and to the references therein. We recall that such types of large Toeplitz linear systems emerge from specific applications

involving e.g. the numerical solution of (integro-) differential equations and of problems with Markov chains.

2. Matrix formulation and notation in 1𝔻

The linear system to solve is (7), and we can decompose the matrix $A_h \in \mathbb{R}^{(n+1)\times(n+1)}$ as follows

$$A_h = \frac{1}{h^2}\begin{bmatrix} 2 & -1 & 0 & \dots & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -1 & 2 & -1 \\ 0 & 0 & \dots & 0 & -1 & 2 \end{bmatrix} + \frac{1}{h^2}\begin{bmatrix} \vartheta h^2 - 2 & (1-\vartheta)h^2 + 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

$$= \frac{1}{h^2}T_{n+1}(2-2\cos(s)) + \frac{1}{h^2}\begin{bmatrix} \mathbf{v}_h^{\mathrm{T}} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \tag{10}$$

$$= S_{n+1} + \frac{1}{h^2}\mathbf{e}_1\mathbf{v}_h^{\mathrm{T}}, \tag{11}$$

where $T_{n+1}(f)$ in (10) is the Toeplitz matrix generated by f according to (9), with $f(s) = 2 - 2\cos(s)$ so that, in the matrix in (8), we have $\alpha = 2$, $a_0 = 2, a_1 = a_{-1} = -1$. Furthermore, we have defined S_{n+1} in (11) as

$$S_{n+1} = \frac{1}{h^2}T_{n+1}(2-2\cos(s)).$$

For this matrix everything is known and in fact

$$T_{n+1}(2-2\cos(s)) = QDQ$$

with Q real symmetric and orthogonal and

$$Q = Q_{n+1} = \left(\sqrt{\frac{2}{n+2}}\sin\left(\frac{lt\pi}{n+2}\right)\right)_{l,t=1}^{n+1}, \quad D = \mathrm{diag}_{l=1,\dots,n+1}\left(4\sin^2\left(\frac{l\pi}{2(n+2)}\right)\right).$$

Hence its conditioning $\kappa_2(\cdot)$ in spectral norm (the one induced by the Euclidean vector norm) is exactly known and it is equal to

$$\kappa_2(S_{n+1}) = \sin^2\left(\frac{(n+1)\pi}{2(n+2)}\right)\sin^{-2}\left(\frac{\pi}{2(n+2)}\right) \approx \frac{4}{\pi^2}n^2,$$

where $a_n \approx b_n$ means $a_n = b_n(1 + o(1))$ and where, in our setting, a even more precise relation can be derived, that is $\kappa_2(S_{n+1}) = \frac{4}{\pi^2}n^2 + O(1)$. Since everything is known regarding the term S_{n+1} our idea is to reduce the analysis as much as possible to information concerning the matrix S_{n+1} and its inverse and to this end the application of the Sherman–Morrison–Woodbury is appropriate.

The Sherman–Morrison–Woodbury formula states that for and invertible square matrix A, column vectors $\mathbf{u}$ and $\mathbf{v}$, and $1 + \mathbf{v}^{\mathrm{T}}A^{-1}\mathbf{u} \neq 0$

$$\left(A + \mathbf{u}\mathbf{v}^{\mathrm{T}}\right)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{u}\mathbf{v}^{\mathrm{T}}A^{-1}}{1 + \mathbf{v}^{\mathrm{T}}A^{-1}\mathbf{u}}. \tag{12}$$

and thus we can obtain in our setting defined above in (12) with $A = S_{n+1}$ and $\mathbf{u} = \frac{\mathbf{e}_1}{h^2}$ and $\mathbf{v} = \mathbf{v}_h$.

$$\left(S_{n+1} + \frac{1}{h^2}\mathbf{e}_1\mathbf{v}_h^{\mathrm{T}}\right)^{-1} = S_{n+1}^{-1} - \frac{S_{n+1}^{-1}\frac{1}{h^2}\mathbf{e}_1\mathbf{v}_h^{\mathrm{T}}S_{n+1}^{-1}}{1 + \mathbf{v}_h^{\mathrm{T}}S_{n+1}^{-1}\frac{\mathbf{e}_1}{h^2}}.$$

or

$$\begin{aligned} A_h^{-1} &= S_{n+1}^{-1} - \frac{\frac{1}{h^2}S_{n+1}^{-1}\mathbf{e}_1\mathbf{v}_h^{\mathrm{T}}S_{n+1}^{-1}}{1 + \frac{1}{h^2}\mathbf{v}_h^{\mathrm{T}}S_{n+1}^{-1}\mathbf{e}_1} \\ &= S_{n+1}^{-1} - R_{n+1}. \end{aligned} \tag{13}$$

Our goal is to estimate quite accurately $\|A_h^{-1}\|_p$ with $p \in [1, \infty]$ and with $\|\cdot\|_p$ being the induced matrix norm introduced in Section 1.2. We concentrate our efforts in the case where $p = 1, 2, \infty$, since the other estimates can be obtained via classical interpolation techniques.

We start by estimating $\|S_{n+1}^{-1}\|_1, \|S_{n+1}^{-1}\|_\infty, \|R_{n+1}\|_1, \|R_{n+1}\|_\infty$. The latter are used for giving quite precise bounds on $\|A_h^{-1}\|_1$ and $\|A_h^{-1}\|_\infty$. It should be noticed that for symmetric matrices the norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$ coincide, but our final structures of interest A_h^{-1} are non symmetric, due to the non symmetry of the one-rank correction induced by the boundary conditions. Of course, this difficulty becomes very heavy in the case of multi-dimensional domains of non-rectangular type.

The estimate for $\|A_h^{-1}\|_2$ can be obtained by a direct check, but it essentially follows from the estimates on $\|A_h^{-1}\|_1$ and $\|A_h^{-1}\|_\infty$, by means of the inequality $\|A_h^{-1}\|_2 \leq \sqrt{\|A_h^{-1}\|_1\|A_h^{-1}\|_\infty}$.

2.1. Estimating $\|S_{n+1}^{-1}\|_p$ with $p = 1, \infty$

We have $S_{n+1}^{-1} = h^2T_{n+1}^{-1}$ where $T_{n+1} = T_{n+1}(2 - 2\cos(s))$ and the inverse $\left(T_{n+1}^{-1}\right)_{r,c} = t_r^{(c)}$

$$T_{n+1}^{-1} = \begin{bmatrix} t_1^{(1)} & t_1^{(2)} & \dots & t_1^{(n+1)} \\ t_2^{(1)} & t_2^{(2)} & \dots & t_2^{(n+1)} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n+1}^{(1)} & t_{n+1}^{(2)} & \dots & t_{n+1}^{(n+1)} \end{bmatrix} = \left[\mathbf{t}^{(1)} \ \mathbf{t}^{(2)} \ \dots \ \mathbf{t}^{(n+1)} \right].$$

The components of the inverse T_{n+1}^{-1}, $t_r^{(c)}$, are defined by, for a fixed column c,

$$t_r^{(c)} = \frac{(n+2-c)r}{n+2}, \quad r = 1, \dots, c-1, \text{ for } c > 1, \tag{14}$$

$$t_r^{(c)} = \frac{(n+2-r)c}{n+2}, \quad r = c \dots, n+1, \tag{15}$$

and symmetrically for a fixed row r

$$t_r^{(c)} = \frac{(n+2-r)c}{n+2}, \quad c = 1, \dots, r-1, \text{ for } r > 1, \tag{16}$$

$$t_r^{(c)} = \frac{(n+2-c)r}{n+2}, \quad c = r, \dots, n+1. \tag{17}$$

All terms of S_{n+1}^{-1} (and T_{n+1}^{-1}) are positive and real, and they are symmetric. Hence by using the explicit expressions of the considered norms, we find

$$\begin{aligned} \|S_{n+1}^{-1}\|_\infty &= \max_r \left\{ \sum_{c=1}^{n+1} \left(S_{n+1}^{-1}\right)_{r,c} \right\} = \max_r \left\{ h^2 \sum_{c=1}^{n+1} \left(T_{n+1}^{-1}\right)_{r,c} \right\} \\ &= \max_c \left\{ h^2 \sum_{r=1}^{n+1} \left(T_{n+1}^{-1}\right)_{r,c} \right\} = \max_c \left\{ \sum_{r=1}^{n+1} \left(S_{n+1}^{-1}\right)_{r,c} \right\} = \|S_{n+1}^{-1}\|_1. \end{aligned} \tag{18}$$

By the formulas above, the highest row sum for the matrices T_{n+1}^{-1} with $n+1$ even is obtained for the row index $r = (n+1)/2$ (or $r = (n+1)/2 + 1$, they are equal). For odd $n+1$, the highest row sum appears for the row index $r = (n+2)/2$.

Thus, for $n+1$ even

$$\begin{aligned} \|T_{n+1}^{-1}\|_\infty &= \sum_{c=1}^{n+1} t_{(n+1)/2}^{(c)} = \sum_{c=1}^{(n-1)/2} \frac{(n+2-(n+1)/2)c}{n+2} + \sum_{c=(n+1)/2}^{n+1} \frac{(n+2-c)(n+1)/2}{n+2} \\ &= \frac{(n+1)^2 + 2(n+1)}{8} = \frac{1+2h}{8h^2} \end{aligned}$$

and for $n+1$ odd

$$\|T_{n+1}^{-1}\|_\infty = \sum_{c=1}^{n+1} t_{(n+2)/2}^{(c)} = \sum_{c=1}^{n/2} \frac{(n+2-(n+2)/2)c}{n+2} + \sum_{c=(n+2)/2}^{n+1} \frac{(n+2-c)(n+2)/2}{n+2}$$

$$= \frac{(n+2)^2}{8} = \frac{(1+h)^2}{8h^2}$$

Consequently, for $n+1$ even, we deduce

$$\|S_{n+1}^{-1}\|_\infty = h^2\|T_{n+1}^{-1}\|_\infty = \frac{1+2h}{8}, \tag{19}$$

and for $n+1$ odd, we have

$$\|S_{n+1}^{-1}\|_\infty = h^2\|T_{n+1}^{-1}\|_\infty = \frac{1+2h+h^2}{8}. \tag{20}$$

As a conclusion, for all $n+1$ and using the symmetry and (19) and (20), we obtain that

$$\|S_{n+1}^{-1}\|_\infty = \|S_{n+1}^{-1}\|_1 \leq \frac{1+2h+h^2}{8}, \tag{21}$$

and the limit as the matrix-size tends to infinity, i.e. $h \to 0$, is $\|S_{n+1}^{-1}\|_\infty = \|S_{n+1}^{-1}\|_1 \to \frac{1}{8}$.

2.2. Estimating $\|R_{n+1}\|_p$ for $p = 1, \infty$

Since $S_{n+1}^{-1} = h^2 T_{n+1}^{-1}$ and $T_{n+1}^{-1}\mathbf{e}_1 = \mathbf{t}^{(1)}$, we find that

$$\begin{aligned} R_{n+1} &= \frac{\frac{1}{h^2} S_{n+1}^{-1}\mathbf{e}_1\mathbf{v}_h^{\mathrm{T}} S_{n+1}^{-1}}{1 + \frac{1}{h^2}\mathbf{v}_h^{\mathrm{T}} S_{n+1}^{-1}\mathbf{e}_1} \\ &= \frac{T_{n+1}^{-1}\mathbf{e}_1\mathbf{v}_h^{\mathrm{T}} S_{n+1}^{-1}}{1 + \mathbf{v}_h^{\mathrm{T}} T_{n+1}^{-1}\mathbf{e}_1} \\ &= \frac{\mathbf{t}^{(1)}\mathbf{v}_h^{\mathrm{T}} S_{n+1}^{-1}}{1 + \mathbf{v}_h^{\mathrm{T}}\mathbf{t}^{(1)}}. \end{aligned} \tag{22}$$

Moreover we have from (15) that the components of $\mathbf{t}^{(1)}$ are

$$t_r^{(1)} = \frac{n+2-r}{n+2} = 1 - \frac{hr}{1+h} = \frac{1-h(r-1)}{1+h}, \quad r = 1, \ldots, n+1, \tag{23}$$

and we have from (10)

$$\mathbf{v}_h^{\mathrm{T}} = \begin{bmatrix} \vartheta h^2 - 2 & (1-\vartheta)h^2 + 1 & 0 & \ldots & 0 \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & 0 & \ldots & 0 \end{bmatrix}.$$

Thus,

$$\mathbf{v}_h^{\mathrm{T}}\mathbf{t}^{(1)} = v_1 t_1^{(1)} + v_2 t_2^{(1)} = (\vartheta h^2 - 2)\frac{1}{1+h} + ((1-\vartheta)h^2 + 1)\frac{1-h}{1+h}$$
$$= \frac{(\vartheta - 1)h^3 + h^2 - h - 1}{1+h},$$

and

$$1 + \mathbf{v}_h^{\mathrm{T}}\mathbf{t}^{(1)} = \frac{h^2\left((\vartheta - 1)h + 1\right)}{1+h}. \tag{24}$$

Also,

$$\mathbf{v}_h^{\mathrm{T}} S_{n+1}^{-1} = h^2 \mathbf{v}_h^{\mathrm{T}} T_{n+1}^{-1}$$
$$= h^2 \left[v_1 t_1^{(1)} + v_2 t_2^{(1)} \;\; v_1 t_1^{(2)} + v_2 t_2^{(2)} \;\; \dots \;\; v_1 t_1^{(c)} + v_2 t_2^{(c)} \;\; \dots \;\; v_1 t_1^{(n+1)} + v_2 t_2^{(n+1)} \right],$$

and thus the components of the row vector $\left(\mathbf{v}_h^{\mathrm{T}} S_{n+1}^{-1}\right)_c$ are

$$\left(\mathbf{v}_h^{\mathrm{T}} S_{n+1}^{-1}\right)_c = h^2 \left(v_1 t_1^{(c)} + v_2 t_2^{(c)} \right),$$

and the components of the matrix $\left(\mathbf{t}^{(1)}\mathbf{v}_h^{\mathrm{T}} S_{n+1}^{-1}\right)_{r,c}$ are

$$\left(\mathbf{t}^{(1)}\mathbf{v}_h^{\mathrm{T}} S_{n+1}^{-1}\right)_{r,c} = t_r^{(1)} h^2 \left(v_1 t_1^{(c)} + v_2 t_2^{(c)} \right), \tag{25}$$

where $t_r^{(1)}$ is defined in (23), and

$$t_1^{(c)} = \frac{1 - h(c-1)}{1+h}, \quad c = 1, \dots, n+1, \tag{26}$$

$$t_2^{(c)} = \begin{cases} \frac{1-h}{1+h}, & c = 1, \\ \frac{2-2h(c-1)}{1+h}, & c > 1, \end{cases} \tag{27}$$

are defined in (16) and (17). Therefore, for $c = 1$

$$\left(\mathbf{t}^{(1)}\mathbf{v}_h^{\mathrm{T}} S_{n+1}^{-1}\right)_{r,1} = t_r^{(1)} h^2 \left(v_1 t_1^{(1)} + v_2 t_2^{(1)} \right) = t_r^{(1)} h^2 \mathbf{v}_h^{\mathrm{T}}\mathbf{t}^{(1)}$$
$$= \frac{1 - h(r-1)}{1+h} h^2 \frac{(\vartheta - 1)h^3 + h^2 - h - 1}{1+h},$$
$$= \frac{h^2(1 - h(r-1))((\vartheta - 1)h^3 + h^2 - h - 1)}{(h+1)^2}, \tag{28}$$

and for $c > 1$

$$\begin{aligned}\left(\mathbf{t}^{(1)}\mathbf{v}_h^{\mathrm{T}}S_{n+1}^{-1}\right)_{r,c} &= t_r^{(1)}h^2\left(v_1t_1^{(c)}+v_2t_2^{(c)}\right) = t_r^{(1)}h^2\mathbf{v}_h^{\mathrm{T}}\mathbf{t}^{(c)}\\ &= \frac{1-h(r-1)}{1+h}h^2\Bigg(\left(\vartheta h^2-2\right)\frac{1-h(c-1)}{1+h}\\ &\quad+\left((1-\vartheta)h^2+1\right)\frac{2-2h(c-1)}{1+h}\Bigg)\\ &= \frac{h^4(2-\vartheta)(1-h(c-1))(1-h(r-1))}{(h+1)^2}. \end{aligned} \tag{29}$$

Thus, we can now define the components of $(R_{n+1})_{r,c} = \left(\frac{\mathbf{t}^{(1)}\mathbf{v}_h^{\mathrm{T}}S_{n+1}^{-1}}{1+\mathbf{v}_h^{\mathrm{T}}\mathbf{t}^{(1)}}\right)_{r,c}$, defined in (22), since we have (24), (28), and (29). For $c=1$ we have

$$\begin{aligned}(R_{n+1})_{r,1} &= \frac{h^2(1-h(r-1))((\vartheta-1)h^3+h^2-h-1)}{(h+1)^2}\Bigg/\frac{h^2\left((\vartheta-1)h+1\right)}{1+h}\\ &= \frac{h(r-1)-1}{h(\vartheta-1)+1}-\frac{h^2(h(r-1)-1)}{h+1}\end{aligned} \tag{30}$$

and for $c>1$

$$\begin{aligned}(R_{n+1})_{r,c} &= \frac{h^4(2-\vartheta)(1-h(c-1))(1-h(r-1))}{(h+1)^2}\Bigg/\frac{h^2\left((\vartheta-1)h+1\right)}{1+h}\\ &= \frac{h^2(2-\vartheta)(1-(c-1)h)(1-h(r-1))}{(h+1)(h(\vartheta-1)+1)}.\end{aligned} \tag{31}$$

Looking at the matrix structures, we deduce that the norms $\|R_{n+1}\|_1$ and $\|R_{n+1}\|_\infty$ are attained by looking at the l^1 norm of first column and of the first row, respectively.

Now we compute $\|R_{n+1}\|_1$,

$$\begin{aligned}\|R_{n+1}\|_1 &= \sum_{r=1}^{n+1}\left|(R_{n+1})_{r,1}\right|\\ &= \sum_{r=1}^{n+1}\left|\frac{h(r-1)-1}{h(\vartheta-1)+1}-\frac{h^2(h(r-1)-1)}{h+1}\right|\\ &= \sum_{r=1}^{n+1}\left|\frac{(1-h(r-1))(h^3(\vartheta-1)+h^2-h-1)}{(h+1)(h(\vartheta-1)+1)}\right|\\ &= \frac{-h^3(\vartheta-1)-h^2+h+1}{(h+1)(h(\vartheta-1)+1)}\sum_{r=1}^{n+1}(1+h-hr))\\ &= \frac{-h^3(\vartheta-1)-h^2+h+1}{(h+1)(h(\vartheta-1)+1)}\left((n+1)(1+h)-h\frac{(n+1)(n+2)}{2}\right)\end{aligned}$$

$$= \frac{-h^3(\vartheta-1)-h^2+h+1}{(h+1)(h(\vartheta-1)+1)}\left(\frac{1+h}{h}-\frac{1+h}{2h}\right)$$
$$= \frac{h^3(1-\vartheta)-h^2+h+1}{2h(h(\vartheta-1)+1)}, \tag{32}$$

since $0<\vartheta<1$.

We now compute $\|R_{n+1}\|_\infty$, by taking into account that all coefficients are positive except the first in the first column. From (30) and (31) we find

$$(R_{n+1})_{1,1} = \frac{h^2}{h+1} - \frac{1}{h(\vartheta-1)+1}, \tag{33}$$
$$(R_{n+1})_{1,c} = \frac{h^2(2-\vartheta)(1-(c-1)h)}{(h+1)(h(\vartheta-1)+1)}, \quad c=2,\dots,n+1. \tag{34}$$

Thus,

$$\begin{aligned}
\|R_{n+1}\|_\infty &= -(R_{n+1})_{1,1} + \sum_{c=2}^{n+1}(R_{n+1})_{1,c}\\
&= -\left(\frac{h^2}{h+1}-\frac{1}{h(\vartheta-1)+1}\right) + \sum_{c=2}^{n+1}\frac{h^2(2-\vartheta)(1-(c-1)h)}{(h+1)(h(\vartheta-1)+1)}\\
&= -\left(\frac{h^2}{h+1}-\frac{1}{h(\vartheta-1)+1}\right) + \frac{h^2(2-\vartheta)}{(h+1)(h(\vartheta-1)+1)}\sum_{c=2}^{n+1}(1+h-ch)\\
&= -\left(\frac{h^2}{h+1}-\frac{1}{h(\vartheta-1)+1}\right) + \frac{h^2(2-\vartheta)}{(h+1)(h(\vartheta-1)+1)}\\
&\quad\times\left((n+1-1)(1+h)-h\left(\frac{(n+1)(n+2)}{2}-1\right)\right)\\
&= -\left(\frac{h^2}{h+1}-\frac{1}{h(\vartheta-1)+1}\right) + \frac{h^2(2-\vartheta)}{(h+1)(h(\vartheta-1)+1)}\\
&\quad\times\left(\frac{1-h^2}{h}-\frac{1+h-2h^2}{2h}\right)\\
&= -\left(\frac{h^2}{h+1}-\frac{1}{h(\vartheta-1)+1}\right) + \frac{1}{2}\frac{h(2-\vartheta)(1-h)}{(h+1)(h(\vartheta-1)+1)}\\
&= \frac{1}{2}\frac{h(2-\vartheta)(1-h)-2h^2(h(\vartheta-1)+1)+2(h+1)}{(h+1)(h(\vartheta-1)+1)}\\
&\to 1 \text{ as } h\to 0.
\end{aligned}$$

2.3. Estimating $\|A_h^{-1}\|_p$ for $p = 1, \infty$

By looking at the formal expressions of the involved matrices, we infer that $\|A_h^{-1}\|_1$ is reached by taking into consideration the first column, thus since $A_h^{-1} = S_{n+1}^{-1} - R_{n+1}$, S_{n+1}^{-1} and R_{n+1} positive and negative respectively, we can just compute the norm directly for A_h^{-1}. The sum of the positive elements of the first column of T_{n+1}^{-1} is equal to $\frac{n+1}{2}$, and thus the sum for the first column of S_{n+1}^{-1} is $\frac{h^2(n+1)}{2} = \frac{h}{2}$, and the sum of components $-(R_{n+1})_{r,1}$ is given in (32), that is

$$\begin{aligned}\|A_h^{-1}\|_1 &= \frac{h}{2} + \frac{h^3(1-\vartheta) - h^2 + h + 1}{2h(h(\vartheta-1)+1)} \\ &= \frac{h^3(\vartheta-1) + h^2 + h^3(1-\vartheta) - h^2 + h + 1}{2h(h(\vartheta-1)+1)} \\ &= \frac{h+1}{2h(h(\vartheta-1)+1)}. \end{aligned} \tag{35}$$

Now we compute $\|A_h^{-1}\|_\infty$. $A_h^{-1} = S_{n+1}^{-1} - R_{n+1}$. From (17) we infer

$$\left(S_{n+1}^{-1}\right)_{1,c} = h^2 t_1^{(c)} = h^2 \frac{1 + h(1-c)}{1+h}, \quad c = 1, \dots, n+1,$$

and by using (33) we find

$$\begin{aligned}\left(A_h^{-1}\right)_{1,1} &= (S_{n+1}^{-1})_{1,1} - (R_{n+1})_{1,1} \\ &= \frac{h^2}{1+h} - \left(\frac{h^2}{h+1} - \frac{1}{h(\vartheta-1)+1}\right) \\ &= \frac{1}{h(\vartheta-1)+1}, \end{aligned} \tag{36}$$

while the use of (34) leads to

$$\begin{aligned}\left(A_h^{-1}\right)_{1,c} &= (S_{n+1}^{-1})_{1,c} - (R_{n+1})_{1,c} \\ &= h^2 \frac{1+h(1-c)}{1+h} - \frac{h^2(2-\vartheta)(1-(c-1)h)}{(h+1)(h(\vartheta-1)+1)} \\ &= \frac{h^2(1+h(1-c))}{1+h}\left(1 - \frac{2-\vartheta}{h(\vartheta-1)+1}\right) \\ &= \frac{h^2(1+h(1-c))(\vartheta-1)}{h(\vartheta-1)+1}. \end{aligned} \tag{37}$$

Since $\left(A_h^{-1}\right)_{1,1}$ in (36) is always positive and $\left(A_h^{-1}\right)_{1,c}$ of (37) is always negative we have

$$\begin{aligned}
\|A_h^{-1}\|_\infty &= \left(A_h^{-1}\right)_{1,1} - \sum_{c=2}^{n+1} \left(A_h^{-1}\right)_{1,c} \\
&= \frac{1}{h(\vartheta-1)+1} - \sum_{c=2}^{n+1} \frac{h^2(1+h(1-c))(\vartheta-1)}{h(\vartheta-1)+1} \\
&= \frac{1}{h(\vartheta-1)+1} - \frac{h^2(\vartheta-1)}{h(\vartheta-1)+1} \sum_{c=2}^{n+1} (1+h-hc) \\
&= \frac{1}{h(\vartheta-1)+1} - \frac{h^2(\vartheta-1)}{h(\vartheta-1)+1} \left((n+1-1)(1+h) - h\left(\frac{(n+1)(n+2)}{2} - 1\right)\right) \\
&= \frac{1}{h(\vartheta-1)+1} - \frac{h^2(\vartheta-1)}{h(\vartheta-1)+1} \left(\frac{1-h^2}{h} - \frac{1+h-2h^2}{2h}\right) \\
&= \frac{2-h(\vartheta-1)(1-h)}{2(h(\vartheta-1)+1)}.
\end{aligned} \tag{38}$$

As a conclusion we deduce from (35) and (38)

$$\begin{aligned}
\|A_h^{-1}\|_2 &\le \sqrt{\|A_h^{-1}\|_1 \|A_h^{-1}\|_\infty} \\
&= \sqrt{\frac{h+1}{2h(h(\vartheta-1)+1)} \frac{2-h(\vartheta-1)(1-h)}{2(h(\vartheta-1)+1)}} \\
&= \frac{1}{2(h(\vartheta-1)+1)} \sqrt{\frac{h+1}{h}(2-h(\vartheta-1)(1-h))} \\
&= \frac{1}{2(h(\vartheta-1)+1)} \sqrt{\frac{2(h+1)+(h^2-1)h(\vartheta-1)}{h}} \\
&= \frac{1}{2(h(\vartheta-1)+1)} \sqrt{\frac{2}{h}+2+(h^2-1)(\vartheta-1)}.
\end{aligned}$$

In order to make a comparison, we recall that we know the exact asymptotical behavior of $\|S_{n+1}^{-1}\|_2$, with S_{n+1} being the pure Toeplitz counterpart of A_h, as reported below

$$\|S_{n+1}^{-1}\|_2 = \frac{1}{\lambda_{\min}(S_{n+1})} = \frac{h^2}{4\sin^2\left(\frac{\pi}{2(n+2)}\right)} = \left(\frac{h}{2\sin\left(\frac{\pi h}{2(1+h)}\right)}\right)^2 \overset{h\to 0}{\to} \frac{1}{\pi^2}. \tag{39}$$

2.4. Spectral results: comments

Here we give a short discussion on few items that, for some aspects, will be considered in more detail in Section 3 and for other aspects will be listed as open problems in the conclusion Section 4.

- The estimates for $\|A_h^{-1}\|_p$ are tight and the growth is like $n^{1/p}$: however the numerical growth of the error seems to be bounded by a constant independently of p. The reason relies on the vectors for which the norm is attained. Such vectors should be concentrated on the first component and this is quite unphysical and it is not observed in practice.
- Even if A_h and its inverse are not symmetric we can prove the spectrum of the related matrix-sequence is clustered along a real positive interval, using the results of the GLT technology reported in Subsection 3.2 (see also [2,13]): we refer to Subsection 3.4 where the analysis is performed both in $1\mathbb{D}$ and $2\mathbb{D}$.
- Regarding the estimates of $\|A_h^{-1}\|_p$, the $2\mathbb{D}$ case (and generically the $d\mathbb{D}$ case) is more difficult, but we can take advantage of the one dimensional case and from a clever tensor structure of the problem when the domain is rectangular (hyper-rectangular in the $d\mathbb{D}$ case).
- When the domain is generic a possibility is given by embedding techniques already exploited in the distributional setting via the GLT approach (see [18,19]).

2.5. Numerical tests in $1\mathbb{D}$

We consider the $1\mathbb{D}$ problem (1) with $a = 0$ and $b = \pi$. We choose $f = -\sin(x)$, $g_a = 0$, and $g_b = 0$ so that $\mathbf{u} = -\sin(x_i)$ is the exact solution in points x_i.

We perform several tests varying the value of $\vartheta \in [0, 1]$, in order to establish whether the convergence of the method depends on the choice of ϑ. In practice, we choose ϑ and n and we compute h and x_0 accordingly:

$$h = \frac{b-a}{n+\vartheta}, \qquad x_0 = b - (n+1)h.$$

The numerical error $\mathbf{e}_h = \mathbf{u} - \mathbf{u}_h$ satisfies the following equation:

$$A_h \mathbf{e}_h = \tau_h,$$

where τ_h is the consistency error:

$$\tau_h = \mathbf{f}_h - A_h \mathbf{u}.$$

Consider the $p-$norm:

$$\|\tau_h\|_{L^p} \approx \left(h \sum_{i=0}^{n} |\tau_h(x_i)|^p \right)^{\frac{1}{p}}, \tag{40}$$

$$\|\mathbf{e}_h\|_{L^p} \approx \left(h \sum_{i=0}^{n} |\mathbf{e}_h(x_i)|^p \right)^{\frac{1}{p}}. \tag{41}$$

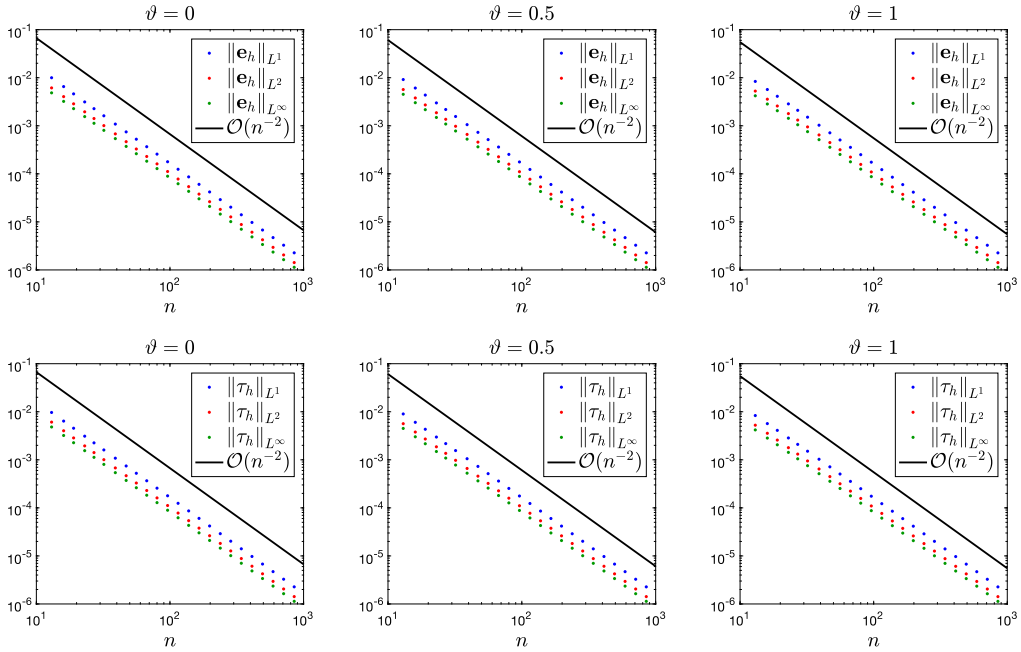

Fig. 2. The dots represent the p - norm of the numerical error $\mathbf{e}_h$ (top) and consistency error τ_h (bottom) for different values of n (horizontal axis) and ϑ: $\vartheta = 0$ (left), $\vartheta = 0.5$ (middle), $\vartheta = 1$ (right). The solid line is a reference for second-order decay.

In Fig. 2 we show that:

$$\|\tau_h\|_{L^p}, \|\mathbf{e}_h\|_{L^p} \approx O(h^2), \text{ for } p = 1, 2, \infty, \tag{42}$$

confirming that the method is second-order consistent and accurate.

We complete the analysis showing the behavior of the spectral radius of the matrix A_h^{-1}. Fig. 3 shows how the smallest eigenvalue (in absolute value) of the matrix A_h changes in relation to n (left panel) and in relation to ϑ (right panel).

Fig. 3 shows that the smallest eigenvalue (in absolute value) of the matrix A_h essentially does not depend on the value of ϑ and it approaches a constant value when n goes to infinity.

Since $\|\mathbf{e}_h\|_{L^p} \leq \left\|A_h^{-1}\right\|_p \|\tau_h\|_{L^p}$, we can conclude that $\|\mathbf{e}_h\|_{L^p} \approx O(h^2)$ and $\left\|A_h^{-1}\right\|_p \|\tau_h\|_{L^p} \approx O(h^{2-\frac{1}{p}})$, as predicted in the first item of Subsection 2.4.

3. Problem formulation in 2𝔻 and related analysis

The section is organized into three parts: first we introduce the d-level notation and the d-level Toeplitz matrices in Subsection 3.1, secondly we define the notion of spectral and singular value distribution and the $*$-algebra of Generalized Locally Toeplitz matrix-sequences in Subsection 3.2, then we describe the matrices arising in the approximation

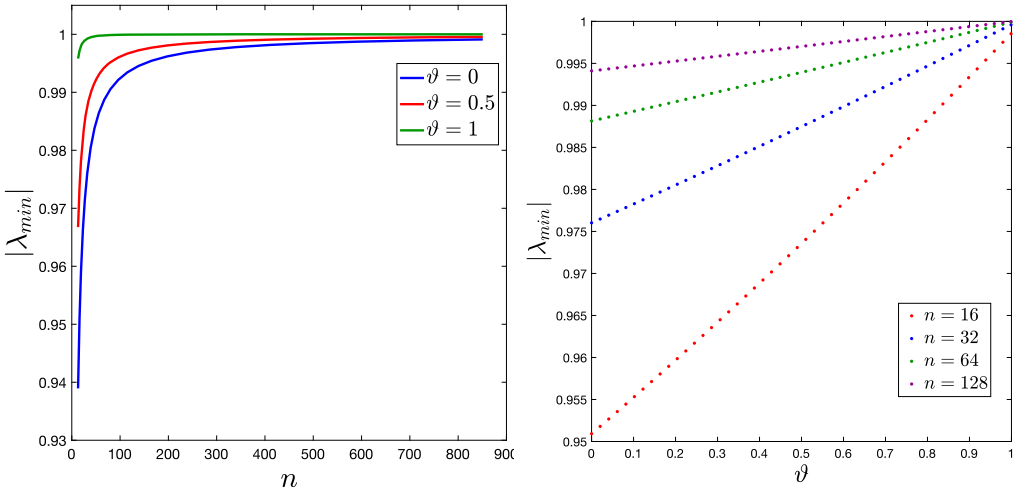

Fig. 3. Smallest eigenvalue in absolute value (vertical axis) for different values of n (horizontal axis, left plot) or for different values of ϑ (horizontal axis, right plot).

of a Dirichlet problem by the Coco–Russo method in Subsection 3.3, and finally we give a spectral analysis of the resulting matrix-sequences in Subsection 3.4.

3.1. Multilevel notation: the case of multilevel Toeplitz and diagonal sampling matrices

We start by introducing the multi-index notation, which is useful in our context. A multi-index $\mathbf{i} \in \mathbb{Z}^d$, also called a d-index, is simply a (row) vector in $\mathbb{Z}^d$; its components are denoted by $i_1, \ldots, i_d$.

- $\mathbf{0}$, $\mathbf{1}$, $\mathbf{2}$, ... are the vectors of all zeros, all ones, all twos, ... (their size will be clear from the context).
- For any positive d-index $\mathbf{m} \in \mathbb{N}_+^d$, we set $N(\mathbf{m}) = \prod_{j=1}^d m_j$ and we write $\mathbf{m} \to \infty$ to indicate that $\min(\mathbf{m}) \to \infty$.
- If $\mathbf{h}, \mathbf{k}$ are d-indices, $\mathbf{h} \le \mathbf{k}$ means that $h_r \le k_r$ for all $r = 1, \ldots, d$.
- The standard lexicographic ordering is assumed uniformly

$$\left[\ldots \left[\left[(j_1, \ldots, j_d) \right]_{j_d = h_d, \ldots, k_d} \right]_{j_{d-1} = h_{d-1}, \ldots, k_{d-1}} \ldots \right]_{j_1 = h_1, \ldots, k_1}. \tag{43}$$

For instance, in the case $d = 2$ the ordering is the following: (h_1, h_2), $(h_1, h_2 + 1)$, ..., (h_1, k_2), $(h_1 + 1, h_2)$,

Multilevel Toeplitz Matrices.

We now briefly summarize the definition and few relevant properties of multilevel Toeplitz matrices, that we will employ in the analysis of the 2𝔻 setting. Given $\mathbf{n} \in \mathbb{N}^d$, a matrix of the form

$$[a_{\mathbf{i}-\mathbf{j}}]_{\mathbf{i},\mathbf{j}=\mathbf{e}}^{\mathbf{n}} \in \mathbb{C}^{N(\mathbf{n})\times N(\mathbf{n})}$$

with $\mathbf{e}$ vector of all ones, with entries $a_{\mathbf{k}} \in \mathbb{C}$, $\mathbf{k} = -(\mathbf{n}-\mathbf{e}), \ldots, \mathbf{n}-\mathbf{e}$, is called a multilevel Toeplitz matrix, or, more precisely, a d-level Toeplitz matrix. Let $\phi : [-\pi,\pi]^d \to \mathbb{C}^{r\times r}$ a matrix-valued function in which each entry belongs to $L^1([-\pi,\pi]^d)$. We denote the Fourier coefficients of the generating function ϕ as

$$\hat{\phi}_{\mathbf{k}} = \frac{1}{(2\pi)^d} \int\limits_{[-\pi,\pi]^d} \phi(\boldsymbol{s}) e^{-\hat{\imath}(\mathbf{k},\boldsymbol{s})}\, \mathrm{d}\boldsymbol{s} \in \mathbb{C}, \quad \mathbf{k} \in \mathbb{Z}^d,$$

where the integrals are computed component-wise and $(\mathbf{k},\boldsymbol{s}) = k_1 s_1 + \ldots + k_d s_d$. For every $\mathbf{n} \in \mathbb{N}^d$, the $\mathbf{n}$-th Toeplitz matrix associated with ϕ is defined as

$$T_{\mathbf{n}}(\phi) := [\hat{\phi}_{\mathbf{i}-\mathbf{j}}]_{\mathbf{i},\mathbf{j}=\mathbf{e}}^{\mathbf{n}}$$

or, equivalently, as

$$T_{\mathbf{n}}(\phi) = \sum_{|j_1|<n_1} \cdots \sum_{|j_d|<n_d} \hat{\phi}_{(j_1,\ldots,j_d)} [J_{n_1}^{(j_1)} \otimes \ldots \otimes J_{n_d}^{(j_d)}], \tag{44}$$

where $\otimes$ denotes the (Kronecker) tensor product of matrices, while $J_m^{(l)}$ is the matrix of order m whose (i,j) entry equals 1 if $i-j=l$ and zero otherwise. We call $\{T_{\mathbf{n}}(\phi)\}_{\mathbf{n}\in\mathbb{N}^d}$ the family of (multilevel block) Toeplitz matrices associated with ϕ, which, in turn, is called the generating function of $\{T_{\mathbf{n}}(\phi)\}_{\mathbf{n}\in\mathbb{N}^d}$.

Multilevel Diagonal Sampling Matrices. For $n \in \mathbb{N}$ and $a : [0,1] \to \mathbb{C}$, we define the diagonal sampling matrix $D_n(a)$ as the diagonal matrix

$$D_n(a) = \underset{i=1,\ldots,n}{\mathrm{diag}}\, a\Big(\frac{i}{n}\Big) = \begin{bmatrix} a(\frac{1}{n}) & & & \\ & a(\frac{2}{n}) & & \\ & & \ddots & \\ & & & a(1) \end{bmatrix} \in \mathbb{C}^{n\times n}.$$

For $\mathbf{n} \in \mathbb{N}^d$ and $a : [0,1]^d \to \mathbb{C}$, we define the multilevel diagonal sampling matrix $D_{\mathbf{n}}(a)$ as the diagonal matrix

$$D_{\mathbf{n}}(a) = \underset{\mathbf{i}=\mathbf{1},\ldots,\mathbf{n}}{\mathrm{diag}}\, a\Big(\frac{\mathbf{i}}{\mathbf{n}}\Big) \in \mathbb{C}^{N(\mathbf{n})\times N(\mathbf{n})},$$

with the lexicographical ordering (43) as discussed at the beginning of the subsection.

3.2. GLT matrix-sequences: operative features

We start with the definition of distribution in the sense of the eigenvalues (spectral distribution) and in the sense of the singular values (singular value distribution) for a

given matrix-sequence. Then we give the operative feature of the $*$-algebra of matrix-sequences.

Definition 1. Let $\{A_n\}_n$ be a sequence of matrices, with A_n of size d_n, and let $f : D \subset \mathbb{R}^t \to \mathbb{C}$ be a measurable function defined on a set D with $0 < \mu_t(D) < \infty$.

- We say that $\{A_n\}_n$ has a (asymptotic) singular value distribution described by f, and we write $\{A_n\}_n \sim_\sigma f$, if

$$\lim_{n\to\infty} \frac{1}{d_n} \sum_{i=1}^{d_n} F(\sigma_i(A_n)) = \frac{1}{\mu_t(D)} \int_D F(|f(\mathbf{x})|)\ \mathrm{d}\mathbf{x}, \quad \forall\, F \in C_c(\mathbb{R}). \tag{45}$$

- We say that $\{A_n\}_n$ has a (asymptotic) spectral (or eigenvalue) distribution described by f, and we write $\{A_n\}_n \sim_\lambda f$, if

$$\lim_{n\to\infty} \frac{1}{d_n} \sum_{i=1}^{d_n} F(\lambda_i(A_n)) = \frac{1}{\mu_t(D)} \int_D F(f(\mathbf{x}))\ \mathrm{d}\mathbf{x}, \quad \forall\, F \in C_c(\mathbb{C}). \tag{46}$$

If $\{A_n\}_n$ has both a singular value and an eigenvalue distribution described by f, then we write $\{A_n\}_n \sim_{\sigma,\lambda} f$.

The symbol f contains spectral/singular value information briefly described informally as follows. With reference to (46), assuming that d_n is large enough and f is at least Riemann integrable, except possibly for a small number of outliers, the eigenvalues of A_n are approximately formed by the samples of f over a uniform grid in D, so that the range of f is a (weak) cluster for the eigenvalues of $\{A_n\}_n$. It is then clear that the symbol f provides a 'compact' and a quite accurate description of the spectrum of the matrices A_n for n large enough. Relation (45) has the same meaning when talking of the singular values of A_n and by replacing f with $|f|$.

A d-level ($d \geq 1$ integer) GLT matrix-sequence $\{A_n\}_n$ is nothing more than a matrix–sequence endowed with a measurable function $\kappa : [0,1]^d \times [-\pi,\pi]^d \to \mathbb{C}$ called *symbol* characterizing the distributional properties of its singular values, and, under certain hypothesis, of its spectrum. For a complete overview of the theory we refer to the books [9, 10], while here we recall only the operative features we need for our restricted setting. Since we have already introduced the multilevel Toeplitz and diagonal matrix-sequences, the only other class we need is that of zero–distributed matrix-sequences, whose definition depends on Definition 1.

Definition 2. [Zero–distributed sequence] A matrix-sequence $\{Z_n\}_n$ such that $\{Z_n\}_n \sim_\sigma 0$ is referred to as a zero-distributed sequence. In other words, $\{Z_n\}_n$ is zero-distributed if and only if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} F(\sigma_i(Z_n)) = F(0), \qquad \forall\, F \in C_c(\mathbb{R}).$$

In a different language, more common in the context of preconditioning and of the convergence analysis of (preconditioned) Krylov methods, a zero–distributed matrix-sequence is a sequence of matrices showing a (weak) clustering at zero in the sense of the singular values (see e.g. [9,21] and references therein).

With the notation introduced in Section 1.2 regarding the Schatten p-norm, with $\|\cdot\|_{S,\infty} = \|\cdot\|_2$ the spectral norm and with $\|\cdot\|_{S,1}$ the trace norm (i.e. the sum of all singular values), the following result holds true [9].

Theorem 3.

GLT 1. *If* $\{A_n\}_n \sim_{\text{GLT}} \kappa$ *then* $\{A_n\}_n \sim_\sigma \kappa$. *If* $\{A_n\}_n \sim_{\text{GLT}} \kappa$ *and the matrices* A_n *are Hermitian then* $\{A_n\}_n \sim_\lambda \kappa$.

GLT 2. *If* $\{A_n\}_n \sim_{\text{GLT}} \kappa$ *and* $A_n = X_n + Y_n$, *where*
- *every* X_n *is Hermitian,*
- $\|X_n\|_{S,\infty}, \|Y_n\|_{S,\infty} \le C$ *for some constant* C *independent of* n,
- $n^{-1}\|Y_n\|_{S,1} \to 0$,

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

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

GLT 4. *If* $\{A_n\}_n \sim_{\text{GLT}} \kappa$ *and* $\{B_n\}_n \sim_{\text{GLT}} \xi$ *then*
- $\{A_n^*\}_n \sim_{\text{GLT}} \overline{\kappa}$,
- $\{\alpha A_n + \beta B_n\}_n \sim_{\text{GLT}} \alpha\kappa + \beta\xi$ *for all* $\alpha, \beta \in \mathbb{C}$,
- $\{A_n B_n\}_n \sim_{\text{GLT}} \kappa\xi$.

GLT 5. *If* $\{A_n\}_n \sim_{\text{GLT}} \kappa$ *and* $\kappa \neq 0$ *a.e. then* $\{A_n^\dagger\}_n \sim_{\text{GLT}} \kappa^{-1}$.

A more general and more advanced result regarding item **GLT2** can be found in [2,13], even if for our purposes item **GLT2** is sufficient in our setting.

3.3. Coco–Russo method in $2\mathbb{D}$

We consider the following Dirichlet problem:

$$\begin{cases} -u_{xx} - u_{yy} = f, & \text{in } \Omega, \\ u = g, & \text{in } \partial\Omega, \end{cases} \tag{47}$$

where $\Omega \subseteq [0,1]^2$, $f, g : \Omega \to \mathbb{R}$ are assigned functions and $u : \Omega \to \mathbb{R}$ is the unknown function. We will consider several geometries in the following sections: rectangular, L-shaped and circular domains.

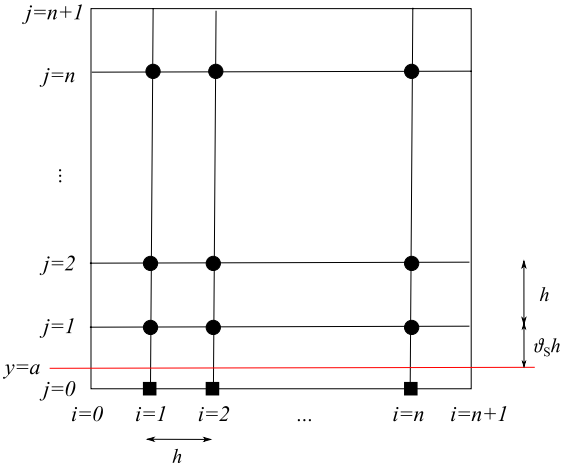

Fig. 4. Discretization of the 2𝔻 domain. Full circles are the n^2 inside grid points, while full squares are the n ghost points. Linear extrapolation is used to define the ghost values $u_{i,0}$ from $u_{i,1}$ and the boundary values $g(x_i, 0)$, for $i = 1, \ldots, n$.

3.3.1. Rectangular domain

We consider $\Omega = [0,1] \times [a,1] \subset [0,1]^2$. The square $[0,1]^2$ is discretized through a uniform Cartesian grid with $(n+2)^2$ grid points $(x_i, y_j) = (i\,h, j\,h)$, for $i, j = 0, \ldots, n+1$, where $h = 1/(n+1)$. As in the 1𝔻 case, let $0 < a < h$ and call $\vartheta_S = (y_1 - a)/h$ (see Fig. 4). The subscript S stands for *south*, since the boundary $y = a$ is the bottom side of the domain. A similar approach can be followed in the other cases.

The elliptic equation $-\Delta u = f$ of problem (47) is discretized by central finite difference on internal grid points, with eliminated boundary conditions on the boundaries $x = 0$, $x = 1$ and $y = 1$. Then, for $2 \le i \le n-1$ and $1 \le j \le n-1$ we have:

$$\frac{4u_{ij} - (u_{i-1\,j} + u_{i+1\,j} + u_{i\,j-1} + u_{i\,j+1})}{h^2} = f_{ij}, \tag{48}$$

while for $i = 1$ and $j = 1, \ldots, n-1$ we eliminate the boundary condition on $x = 0$:

$$\frac{4u_{1\,j} - (u_{2\,j} + u_{1\,j-1} + u_{1\,j+1})}{h^2} = f_{ij} + \frac{g(0, y_j)}{h^2}. \tag{49}$$

Similarly, we eliminate the boundary conditions on $x = 1$ and $y = 1$. The boundary condition on $y = a$ is discretized by linear interpolation:

$$\vartheta_S u_{i,0} + (1 - \vartheta_S) u_{i,1} = g(x_i, a), \qquad \text{for } i = 1, \ldots, n.$$

Overall, there are n^2 inside grid points (x_i, y_j) for $i, j = 1, \ldots, n$ and n ghost points for $(x_i, 0)$ for $i = 1 \ldots, n$.

Using a total lexicographical order, the matrix of coefficients that we obtain is a 2-level matrix with the following structure:

$$A_h = \left(\begin{array}{c|c|c|c|c|c} \vartheta_S \mathbb{I}_n & (1-\vartheta_S)\mathbb{I}_n & & & & \\ \hline B & G & B & & & \\ \hline & B & G & B & & \\ \hline & & \ddots & \ddots & \ddots & \\ \hline & & & B & G & B \\ \hline & & & & B & G \end{array}\right), \tag{50}$$

where

$$\vartheta_S \mathbb{I}_n \in \mathbb{R}^{n\times n},$$
$$(1-\vartheta_S)\mathbb{I}_n \in \mathbb{R}^{n\times n},$$
$$B = -\frac{1}{h^2}\mathbb{I}_n \in \mathbb{R}^{n\times n},$$

and

$$G = \frac{1}{h^2}\begin{pmatrix} 4 & -1 & & & \\ -1 & 4 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 4 & -1 \\ & & & -1 & 4 \end{pmatrix} \in \mathbb{R}^{n\times n},$$

A_h has n blocks of G, so $A_h \in \mathbb{R}^{n(n+1)\times n(n+1)}$.

3.3.2. L-shaped domain

We now consider the case of an L-shaped domain $\Omega \subset [0,1]^2$ (Fig. 5). The square $[0,1]^2$ is discretized as in the previous case. Let $0 < a_x, a_y < h$, $1-h < c_x, c_y < 1$, and $i_x h < b_x < (i_x+1)h$, $j_y h < b_y < (j_y+1)h$ for some i_x, j_y such that $0 < i_x, j_y < n+1$. The number of inside and ghost points depends on the position of the extremes b_x and b_y in the domain, in particular, we denote with Ω_I the set of inside points and with Ω_G the set of ghost points. Similar to the rectangular case, by discretizing the elliptic problem we obtain a linear system of $|\Omega_I| + |\Omega_G|$ equations and $|\Omega_I| + |\Omega_G|$ unknowns. The $|\Omega_I|$ equations are obtained discretizing the elliptic equation on the inside points:

$$\frac{-u_{i,j-1} - u_{i,j+1} + 4u_{i,j} - u_{i-1,j} - u_{i+1,j}}{h^2} = f_{i,j}, \quad (i,j) : (x_i, y_j) \in \Omega_I$$

The $|\Omega_G|$ equations for the ghost points are obtained from the discretization of the boundary conditions on $y = a_y$, $y = b_y$, $y = c_y$, $x = a_x$, $x = b_x$ and $x = c_x$ by linear interpolations:

$$\vartheta_S u_{i,0} + (1-\vartheta_S)u_{i,1} = g(x_i, a_y), \ i = 1, \dots, n,$$
$$\vartheta_{N_1} u_{i,n+1} + (1-\vartheta_{N_1})u_{i,n} = g(x_i, c_y), \ i = 1, \dots, i_x,$$

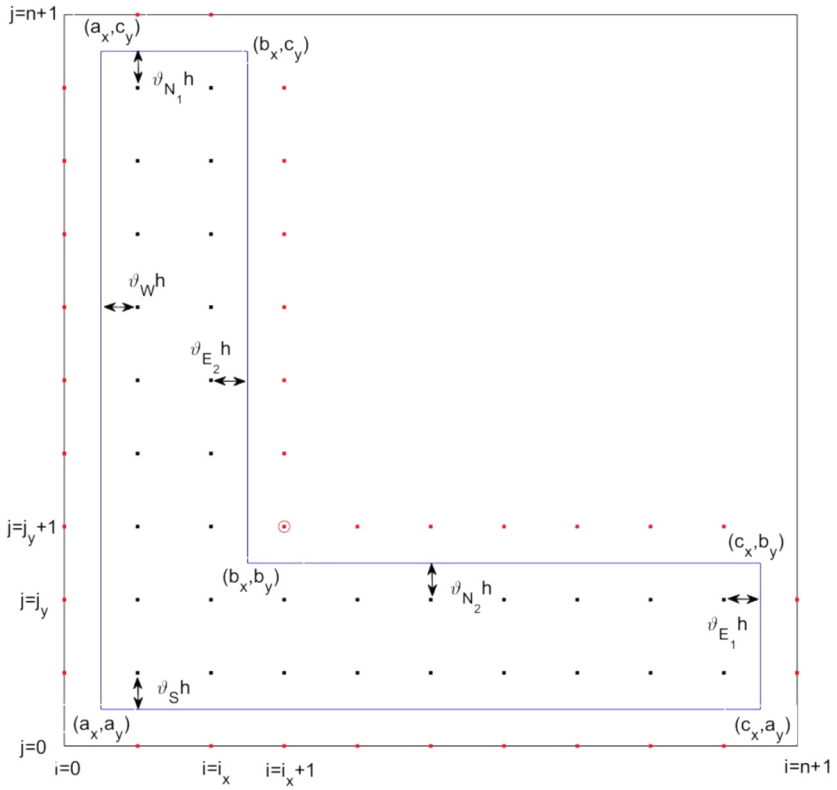

Fig. 5. Discretization of the *L*-shaped domain (Section 3.3.2). The black dots are the inside grid points, while the red dots are the ghost grid points.

$$\vartheta_{\mathrm{N}_2} u_{i,j_y+1} + (1-\vartheta_{\mathrm{N}_2}) u_{i,j_y} = g(x_i, b_y),\ i = i_x+2, \dots, n,$$
$$\vartheta_{\mathrm{W}} u_{0,j} + (1-\vartheta_{\mathrm{W}}) u_{1,j} = g(a_x, y_j),\ j = 1, \dots, n,$$
$$\vartheta_{\mathrm{E}_1} u_{n+1,j} + (1-\vartheta_{\mathrm{E}_1}) u_{n,j} = g(c_x, y_j),\ j = 1, \dots, j_y,$$
$$\vartheta_{\mathrm{E}_2} u_{i_x+1,j} + (1-\vartheta_{\mathrm{E}_2}) u_{i_x,j} = g(b_x, y_j),\ j = j_y+2, \dots, n,$$

where $\vartheta_{\mathrm{S}} = \frac{y_1 - a_y}{h}$, $\vartheta_{\mathrm{N}_1} = \frac{c_y - y_n}{h}$, $\vartheta_{\mathrm{N}_2} = \frac{b_y - y_{j_y}}{h}$, $\vartheta_{\mathrm{W}} = \frac{x_1 - a_x}{h}$, $\vartheta_{\mathrm{E}_1} = \frac{c_x - x_n}{h}$, and $\vartheta_{\mathrm{E}_2} = \frac{b_x - x_{i_x}}{h}$. The ghost point at the corner (red circle in Fig. 5) of the domain is an exception, as it belongs to the stencil of two different inside points and for which the discretized equation becomes:

$$(\vartheta_{\mathrm{E}_2} + \vartheta_{\mathrm{N}_2}) u_{i_x+1,j_y+1} + (1-\vartheta_{\mathrm{E}_2}) u_{i_x,j_y+1} + (1-\vartheta_{\mathrm{N}_2}) u_{i_x+1,j_y} = g(b_x, y_{j_y+1}) + g(x_{i_x+1}, b_y).$$

Using a total lexicographical order the matrix A_h is a block matrix with the following structure:

$$
\mathrm{A}_h = \left(\begin{array}{c|c|c|c|c|c|c|c|c|c|c|c}
\vartheta_{\mathrm{S}}\mathbb{I}_n & \mathrm{I}_{\mathrm{S}} & & & & & & & & & & \\ \hline
\tilde{\mathrm{B}}_1 & \mathrm{G}_1 & \mathrm{B}_1 & & & & & & & & & \\ \hline
 & \mathrm{B}_1 & \mathrm{G}_1 & \mathrm{B}_1 & & & & & & & & \\ \hline
 & & \ddots & \ddots & \ddots & & & & & & & \\ \hline
 & & & \mathrm{B}_1 & \mathrm{G}_1 & \mathrm{B}_1 & & & & & & \\ \hline
 & & & & \mathrm{B}_1 & \mathrm{G}_1 & \mathrm{B}_1^* & & & & & \\ \hline
 & & & & & \mathrm{BG}_1 & \mathrm{BG}_2 & \mathrm{BG}_3 & & & & \\ \hline
 & & & & & & \mathrm{B}_2^* & \mathrm{G2} & \mathrm{B}_2 & & & \\ \hline
 & & & & & & & \mathrm{B}_2 & \mathrm{G}_2 & \mathrm{B}_2 & & \\ \hline
 & & & & & & & & \ddots & \ddots & \ddots & \\ \hline
 & & & & & & & & & \mathrm{B}_2 & \mathrm{G}_2 & \tilde{\mathrm{B}}_2 \\ \hline
 & & & & & & & & & & \mathrm{I}_{\mathrm{N}_1} & \vartheta_{\mathrm{N}_1}\mathbb{I}_{n_x}
\end{array}\right)
$$

$$
\in \mathbb{R}^{(|\Omega_I|+|\Omega_G|)\times(|\Omega_I|+|\Omega_G|)},
$$

where, denoting with $\underline{0}_a$ a zero column vector of size a, $\underline{0}_b^{\mathrm{T}}$ a zero row vector of size b, $\underline{0}_{c,d}$ a zero matrix with c rows and d columns, n_x the number of ghost points in the top boundary, we have:

$$
\vartheta_{\mathrm{S}}\mathbb{I}_n \in \mathbb{R}^{n\times n},\ \vartheta_{\mathrm{N}_1}\mathbb{I}_{n_x} \in \mathbb{R}^{n_x\times n_x},
$$

$$
\mathrm{I}_{\mathrm{S}} = \left(\underline{0}_n \middle| (1-\vartheta_{\mathrm{S}})\mathbb{I}_n \middle| \underline{0}_n\right) \in \mathbb{R}^{n\times(n+2)},
$$

$$
\mathrm{I}_{\mathrm{N}_1} = \left(\underline{0}_{n_x} \middle| (1-\vartheta_{\mathrm{N}_1})\mathbb{I}_{n_x} \middle| \underline{0}_{n_x}\right) \in \mathbb{R}^{n_x\times(n_x+2)},
$$

$$
\tilde{\mathrm{B}}_1 = \left(\begin{array}{c} \underline{0}_n^{\mathrm{T}} \\ \hline -\frac{1}{h^2}\mathbb{I}_n \\ \hline \underline{0}_n^{\mathrm{T}} \end{array}\right) \in \mathbb{R}^{(n+2)\times n},\ \tilde{\mathrm{B}}_2 = \left(\begin{array}{c} \underline{0}_{n_x}^{\mathrm{T}} \\ \hline -\frac{1}{h^2}\mathbb{I}_{n_x} \\ \hline \underline{0}_{n_x}^{\mathrm{T}} \end{array}\right) \in \mathbb{R}^{(n_x+2)\times n_x},
$$

$$
\mathrm{B}_1 = \left(\begin{array}{c|c|c} 0 & \underline{0}_n^{\mathrm{T}} & 0 \\ \hline \underline{0}_n & -\frac{1}{h^2}\mathbb{I}_n & \underline{0}_n \\ \hline 0 & \underline{0}_n^{\mathrm{T}} & 0 \end{array}\right) \in \mathbb{R}^{(n+2)\times(n+2)},\ \mathrm{B}_2 = \left(\begin{array}{c|c|c} 0 & \underline{0}_{n_x}^{\mathrm{T}} & 0 \\ \hline \underline{0}_{n_x} & -\frac{1}{h^2}\mathbb{I}_{n_x} & \underline{0}_{n_x} \\ \hline 0 & \underline{0}_{n_x}^{\mathrm{T}} & 0 \end{array}\right) \in \mathbb{R}^{(n_x+2)\times(n_x+2)},
$$

$$
\mathrm{B}_1^* = \left(\begin{array}{c|c} 0 & \underline{0}_n^{\mathrm{T}} \\ \hline \underline{0}_n & -\frac{1}{h^2}\mathbb{I}_n \\ \hline 0 & \underline{0}_n^{\mathrm{T}} \end{array}\right) \in \mathbb{R}^{(n+2)\times(n+1)},\ \mathrm{B}_2^* = \left(\begin{array}{c|c} \underline{0}_{n_x}^{\mathrm{T}} & \underline{0}_{n+1-n_x}^{\mathrm{T}} \\ \hline -\frac{1}{h^2}\mathbb{I}_{n_x} & \underline{0}_{n_x,n+1-n_x} \\ \hline \underline{0}_{n_x}^{\mathrm{T}} & \underline{0}_{n+1-n_x}^{\mathrm{T}} \end{array}\right) \in \mathbb{R}^{(n_x+2)\times(n+1)},
$$

$$
\mathrm{G}_1 = \begin{pmatrix} \vartheta_{\mathrm{W}} & 1-\vartheta_{\mathrm{W}} & & & \\ -\frac{1}{h^2} & \frac{4}{h^2} & -\frac{1}{h^2} & & \\ & \ddots & \ddots & \ddots & \\ & & -\frac{1}{h^2} & \frac{4}{h^2} & -\frac{1}{h^2} \\ & & & 1-\vartheta_{\mathrm{E}_1} & \vartheta_{\mathrm{E}_1} \end{pmatrix} \in \mathbb{R}^{(n+2)\times(n+2)},
$$

$$\mathrm{G}_2 = \begin{pmatrix} \vartheta_{\mathrm{W}} & 1-\vartheta_{\mathrm{W}} & & & \\ -\frac{1}{h^2} & \frac{4}{h^2} & -\frac{1}{h^2} & & \\ & \ddots & \ddots & \ddots & \\ & & -\frac{1}{h^2} & \frac{4}{h^2} & -\frac{1}{h^2} \\ & & & 1-\vartheta_{\mathrm{E}_2} & \vartheta_{\mathrm{E}_2} \end{pmatrix} \in \mathbb{R}^{(n_x+2)\times(n_x+2)},$$

$$\mathrm{BG}_1 = \left(\begin{array}{c|c|c} 0 & \underline{0}_{n_x}^{\mathrm{T}} & 0 \\ \hline \underline{0}_{n_x} & -\frac{1}{h^2}\mathbb{I}_{n_x} & \underline{0}_{n_x,n+1-n_x} \\ \hline \underline{0}_{n-n_x} & \underline{0}_{n-n_x,n_x} & (1-\vartheta_{\mathrm{N}_2})\mathbb{I}_{n+1-n_x} \end{array}\right) \in \mathbb{R}^{(n+1)\times(n+2)},$$

$$\mathrm{BG}_3 = \left(\begin{array}{c|c|c} 0 & \underline{0}_{n_x}^{\mathrm{T}} & 0 \\ \hline \underline{0}_{n_x} & -\frac{1}{h^2}\mathbb{I}_{n_x} & \underline{0}_{n_x} \\ \hline \underline{0}_{n-n_x} & \underline{0}_{n-n_x,n_x} & \underline{0}_{n-n_x} \end{array}\right) \in \mathbb{R}^{(n+1)\times(n+1)},$$

$$\mathrm{BG}_2 = \left(\begin{array}{c|c} A & \underline{0}_{n_x+1,n-n_x-1} \\ \hline B & C \end{array}\right) \in \mathbb{R}^{(n+1)\times(n+1)},$$

with

$$\mathrm{A} = \begin{pmatrix} \vartheta_{\mathrm{W}} & 1-\vartheta_{\mathrm{W}} & & & \\ -\frac{1}{h^2} & \frac{4}{h^2} & -\frac{1}{h^2} & & \\ & \ddots & \ddots & \ddots & \\ & & -\frac{1}{h^2} & \frac{4}{h^2} & -\frac{1}{h^2} \end{pmatrix} \in \mathbb{R}^{(n_x+1)\times(n_x+2)},$$

$$\mathrm{B} = \left(\begin{array}{c|cc} \underline{0}_{n_x}^{\mathrm{T}} & 1-\vartheta_{\mathrm{E}_2} & \vartheta_{\mathrm{E}_2}+\vartheta_{\mathrm{N}_2} \\ \hline \underline{0}_{n-n_x-1,n_x} & \multicolumn{2}{c}{\underline{0}_{n-n_x-1,2}} \end{array}\right) \in \mathbb{R}^{(n-n_x)\times(n_x+2)},$$

$$\mathrm{C} = \left(\begin{array}{c} \underline{0}_{n-n_x-1}^{\mathrm{T}} \\ \hline \vartheta_{\mathrm{N}_2}\mathbb{I}_{n-n_x-1} \end{array}\right) \in \mathbb{R}^{(n-n_x)\times(n-n_x-1)}.$$

In particular, the matrices G_1 and G_2 appear in A_h, respectively, n_y and $n-n_y-1$ times, where n_y is the number of ghost points on the side E_1 of the L-shaped domain.

3.3.3. Circular domain

Finally, we consider the case of a circular domain $\Omega \subset [0,1]^2$ (Fig. 6). The reader can find more details on the discretization of the elliptic problem on general shapes in [7]. After determining the inside and ghost points, the discretization of the elliptic equation on the inside points is carried out as in the case of the L-shaped domain. The $|\Omega_G|$ equations for the ghost points are obtained from the discretization of the boundary conditions imposed on the boundary. Therefore, for each ghost point, we determine the orthogonal projection $P = (x_P, y_P)$ on the boundary and impose the Dirichlet boundary condition on P by bilinear interpolation (Fig. 7):

$$\vartheta_x\vartheta_y u_{i,j-1} + (1-\vartheta_x)\vartheta_y u_{i+1,j-1} + \vartheta_x(1-\vartheta_y)u_{i,j} + (1-\vartheta_x)(1-\vartheta_y)u_{i+1,j} = g(x_P, y_P),$$

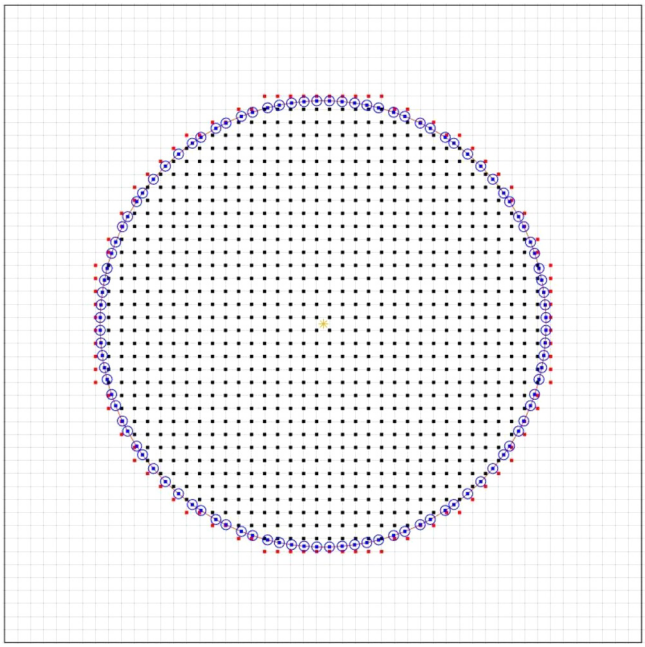

Fig. 6. Discretization of the circular domain (Section 3.3.3). The black dots are the inside grid points, the red dots are the ghost grid points, the blue circles are the projection of each ghost point on the boundary. In these points we impose the boundary conditions.

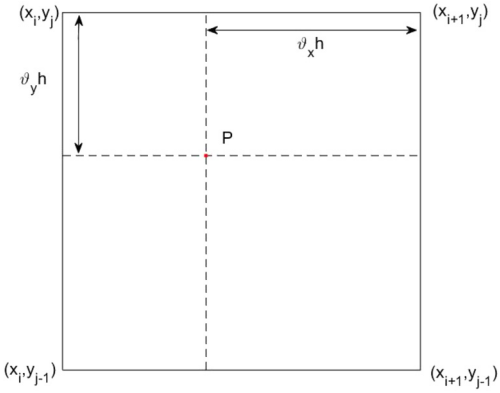

Fig. 7. Bilinear interpolation on the point P.

where $\vartheta_x = \frac{x_{i+1}-x_P}{h}$ and $\vartheta_y = \frac{y_P-y_j}{h}$.

3.4. Spectral analysis in $\mathbb{1D}$ *and in* $\mathbb{2D}$

Having in mind the notations of Subsection 3.1, the matrix A_h can be decomposed in the following way

$$A_h = \frac{1}{h^2}\left[T_{\mathbf{n}}(f) + X_{\mathbf{n}}\right] \tag{51}$$

where $\mathbf{n} = (n+1, n)$, the size of A_h is $N(\mathbf{n}) = n(n+1)$,

$$T_{\mathbf{n}}(f) = T_{n+1}(2 - 2\cos(s)) \otimes \mathbb{I}_n + \mathbb{I}_{n+1} \otimes T_n(2 - 2\cos(s)), \tag{52}$$

$T_k(2 - 2\cos(s))$ is a Toeplitz matrix, already used in the $1\mathbb{D}$ case in Section 2, and

$$X_{\mathbf{n}} = \left[\begin{array}{c|c|c} T_n\left(h^2\vartheta_S - 4 + 2\cos(s)\right) & T_n\left(h^2(1-\vartheta_S) + 1\right) & \mathbf{0}_{n\times n(n-1)} \\ \hline \mathbf{0}_{n^2\times n} & \mathbf{0}_{n^2\times n} & \mathbf{0}_{n^2\times n(n-1)} \end{array}\right]. \tag{53}$$

Of course, taking into account relation (44) with $d = 2$ and (52), the function f is bivariate and can be written as

$$f(s_1, s_2) = 4 - 2\cos(s_1) - 2\cos(s_2).$$

Therefore by using item **GLT1** in Theorem 3 we have

$$\{T_{\mathbf{n}}(f)\} \sim_{\text{GLT}} f$$

in the sense of Subsection 3.2, so that

$$\{T_{\mathbf{n}}(f)\} \sim_{\sigma} f,$$

according to Definition 1. Furthermore, since $T_{\mathbf{n}}(f)$ is Hermitian (in fact real symmetric) for any choice of the partial sizes, thanks to item **GLT1**, we deduce $\{T_{\mathbf{n}}(f)\} \sim_{\lambda} f$ as well.

Now, taking into account Definition 2, it is easy to see that $\{X_{\mathbf{n}}\}$ is a zero–distributed matrix-sequence, simply because its rank is bounded by n and hence the number of nonzero singular values is at most $n = o(n(n+1))$ with $N(\mathbf{n}) = n(n+1)$ being the size of $X_{\mathbf{n}}$. Therefore by item **GLT3**

$$\{X_{\mathbf{n}}\} \sim_{\text{GLT}} 0,$$

so that $\{h^2 A_h\} \sim_{\text{GLT}} f$ by item **GLT4**, since both $\{T_{\mathbf{n}}(f)\}, \{X_{\mathbf{n}}\}$ are GLT matrix-sequences and $h^2 A_h = T_{\mathbf{n}}(f) + X_{\mathbf{n}}$ for any choice of the partial sizes. Then, again by item **GLT1** we deduce

$$\{h^2 A_h\} \sim_{\sigma} f.$$

However, $X_{\mathbf{n}}$ is non-Hermitian and therefore we cannot apply item **GLT1** for concluding $\{h^2 A_h\} \sim_{\lambda} f$. However, this can be done by using item **GLT2**, as proven in the following lines both in $1\mathbb{D}$ and in $2\mathbb{D}$.

Theorem 4. *With the notations used so far in* $1\mathbb{D}$ *we have*

$$\{h^2 A_h\} \sim_\lambda 2 - 2\cos(s), \tag{54}$$

while in $2\mathbb{D}$ *we have*

$$\{h^2 A_h\} \sim_\lambda 4 - 2\cos(s_1) - 2\cos(s_2). \tag{55}$$

Proof. In $1\mathbb{D}$ we recall the identity

$$h^2 A_h = T_n(2 - 2\cos(s)) + \mathbf{e}_1 \mathbf{v}_h^{\mathrm{T}}.$$

Since $\mathbf{e}_1 \mathbf{v}_h^{\mathrm{T}}$ is a rank one matrix, it has a unique nonzero singular value so that

$$\|\mathbf{e}_1 \mathbf{v}_h^{\mathrm{T}}\|_{S,1} = \|\mathbf{e}_1 \mathbf{v}_h^{\mathrm{T}}\|_{S,\infty} = \|\mathbf{v}_h^{\mathrm{T}}\|_2,$$

and hence a trivial computation shows that

$$\lim_{n\to\infty} \frac{\|\mathbf{e}_1 \mathbf{v}_h^{\mathrm{T}}\|_{S,1}}{n} = 0.$$

Therefore, by item **GLT2**, we infer that both the GLT matrix-sequences $\{h^2 A_h\}, \{T_n(2 - 2\cos(s))\}$ share the same eigenvalue distribution function $2 - 2\cos(s)$, which is the GLT symbol, so that (54) is proven.

In $2\mathbb{D}$, according to the 2-level notation, we remind that

$$h^2 A_h = T_{\mathbf{n}}(4 - 2\cos(s_1) - 2\cos(s_2)) + X_{\mathbf{n}}, \quad \mathbf{n} = (n+1, n).$$

Now in the light of (53) we deduce that

$$\|X_{\mathbf{n}}\|_{S,1} \le \|T_n\left(h^2\vartheta_S - 4 + 2\cos(s)\right)\|_{S,1} + \|T_n\left(h^2(1-\vartheta_S) + 1\right)\|_{S,1}.$$

Now, using the fact that $2\pi\|T_n(g)\|_{S,1} \le n\int_{[-\pi,\pi]} |g(s)|\,\mathrm{d}s$ (see [20]), we obtain

$$\|X_{\mathbf{n}}\|_{S,1} \le n2\pi \int_{[-\pi,\pi]} |h^2\vartheta_S - 4 + 2\cos(s)|\,\mathrm{d}s + n(h^2(1-\vartheta_S) + 1)$$

and, as in the $1\mathbb{D}$ setting, if we divide by the size of $X_{\mathbf{n}}$ i.e. $n(n+1)$ we find

$$\lim_{\mathbf{n}\to\infty} \frac{\|X_{\mathbf{n}}\|_{S,1}}{n(n+1)} = 0.$$

Consequently, again by item **GLT2**, we deduce that both the GLT matrix-sequences $\{h^2 A_h\}, \{T_{\mathbf{n}}(4 - 2\cos(s_1) - 2\cos(s_2))\}$ share the same eigenvalue distribution function $4 - 2\cos(s_1) - 2\cos(s_2)$, which is the GLT symbol, and hence (55) is proven. •

The previous result shows a spectral distribution as nonnegative functions both in $1\mathbb{D}$ and $2\mathbb{D}$. More precisely, looking at the range of the spectral symbols, we deduce that $[0,4]$ is a cluster for the eigenvalues of $\{h^2A_h\}$ in $1\mathbb{D}$, while $[0,8]$ is a cluster for the eigenvalues of $\{h^2A_h\}$ in $2\mathbb{D}$.

This is nontrivial (and somehow unexpected), given the fact that the related corrections are non-Hermitian and possess only strictly negative eigenvalues and zero eigenvalues.

3.5. Numerical results and discussion in a general setting

We consider the more general case of a variable coefficient case

$$-\nabla \cdot (a\,\nabla u) = f \text{ in } \Omega \tag{56}$$

where Ω is a segment in $1\mathbb{D}$ and a rectangular, L-shaped or circular domain in $2\mathbb{D}$ (described in Sections 3.3.1, 3.3.2 and 3.3.3, respectively). Eq. (56) is discretized by the usual central finite difference method. In particular, the $1\mathbb{D}$ discretization (3) becomes

$$\frac{-a_L u_{i-1} + (a_L + a_R)u_i - a_R u_{i+1}}{h^2} = f_i \text{ for } i = 1, \ldots, n-1,$$

where $a_L = (a_{i-1} + a_i)/2$ and $a_R = (a_{i+1} + a_i)/2$, while the $2\mathbb{D}$ discretization (48) becomes

$$\frac{(a_L + a_R + a_B + a_T)u_{ij} - a_L\, u_{i-1\,j} - a_R\, u_{i+1\,j} - a_B\, u_{i\,j-1} - a_T\, u_{i\,j+1}}{h^2} = f_{ij}.$$

where $a_L = (a_{i-1,j} + a_{ij})/2$, $a_R = (a_{i+1,j} + a_{ij})/2$, $a_B = (a_{i,j-1} + a_{ij})/2$ and $a_T = (a_{i,j+1} + a_{ij})/2$. In the $1\mathbb{D}$ setting we consider the variable coercive coefficient $a(x) = 1 + x^2$. We construct the matrices h^2A_h from (7) for several values of h and we denote by N the matrix size. We plot in the left panel of Fig. 8 the eigenvalues of the matrices and of the function $a(x)(2 - 2\cos(s))$ sampled in m equispaced points in x on $[0,1]$ and in m equispaced points in s on $[0,\pi]$, with m^2 much larger than N: both sampling and eigenvalues sorted in nondecreasing way. When dealing with the $2\mathbb{D}$ cases we consider the coercive variable coefficient $a(x,y) = 1 + x^2 + y$. We construct the matrices h^2A_h for different values of h and we denote by N the matrix size. We plot in the right panel of Fig. 8 the eigenvalues of the matrix for the rectangular case described in Section 3.3.1 and of the function $a(x,y)(4 - 2\cos(s))$ sampled in $m = m_1, m_2$ equispaced points in x, y on $[0,1]$ and in $m = m_1, m_2$ equispaced points in s_1, s_2 on $[0,\pi]$, with m^4 much larger than N: both sampling and eigenvalues sorted in nondecreasing way.

We repeat the previous procedure on a L-shaped (left panel of Fig. 9) domain in $2\mathbb{D}$ and on a curved domain in $2\mathbb{D}$ (a circle, centered in $(0.5, 0.5)$ with radius $r = 0.35$, right panel of Fig. 9). While the values of h are the same for all $2\mathbb{D}$ tests (right panel of Fig. 8 and both panels of Fig. 9), the values of N differ due to the different sizes of the domains,

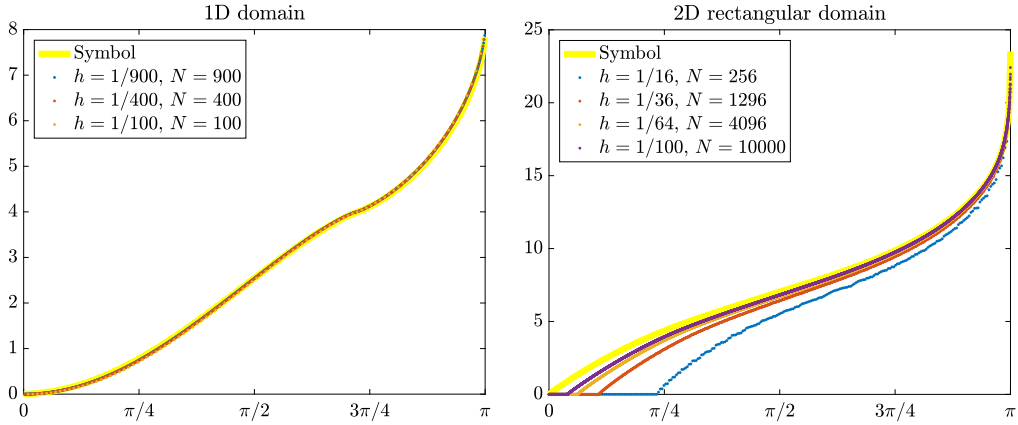

Fig. 8. GLT symbols (yellow line) and eigenvalues of $h^2 A_h$ (dots) for different values of h and matrix sizes N (Section 3.5). Left: $1\mathbb{D}$ problem in $\Omega = [a, b]$ (Eq. (7) and Fig. 1). Right: $2\mathbb{D}$ problem (56) on a rectangular domain (Section 3.3.1).

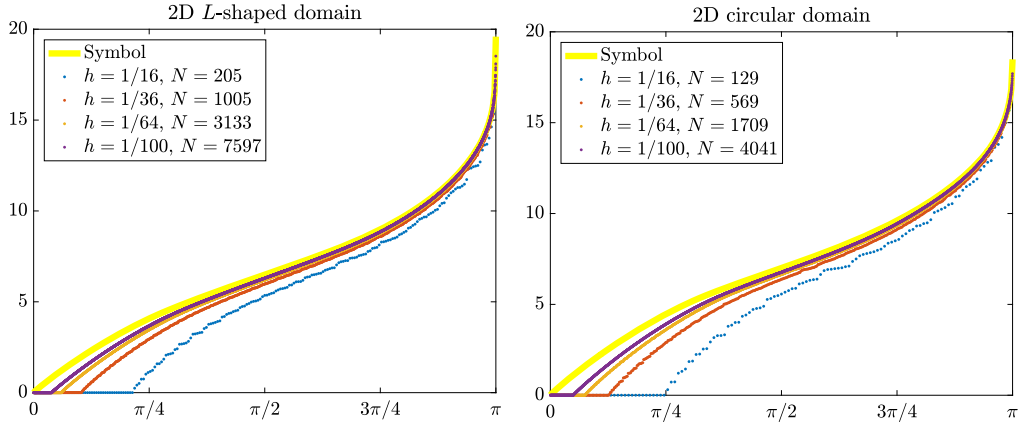

Fig. 9. GLT symbols (yellow line) and eigenvalues of $h^2 A_h$ (dots) for different values of h and matrix sizes N (Section 3.5). Left: $2\mathbb{D}$ problem (56) on a L-shaped domain (Section 3.3.2). Right: $2\mathbb{D}$ problem (56) on a circular domain (Section 3.3.3).

since it is the sum of internal and ghost points only and does not include the external points. In all the previous situations, we observe a good matching (see Figs. 8 and 9) between the global eigenvalue distribution of $\{h^2 A_h\}$ and the GLT symbol

$$a(x)(2 - 2\cos(s)), \qquad x \in [0, 1],\ s \in [0, \pi], \tag{57}$$

in $1\mathbb{D}$ and

$$a(x, y)(4 - 2\cos(s_1) - 2\cos(s_2)), \quad x, y \in \Omega,\ s_1, s_2 \in [0, \pi], \tag{58}$$

with Ω being a $2\mathbb{D}$ domain. Since the rows related to ghost points have a different scaling (see the beginning of Section 2 in $1\mathbb{D}$ and, for example, equation (50) in $2\mathbb{D}$ (rectangular case), we observe that N_G eigenvalues of $h^2 A_h$ scale with h^2, where N_G is the number of ghost points. We call them *ghost eigenvalues*, while the remaining eigenvalues are called *internal eigenvalues*. Due to the non-decreasing order, the ghost eigenvalues are the smallest ones and their contribution to the spectrum is asymptotically negligible, since $N_G/N \to 0$ as $N \to \infty$. To better understanding the contribution of the ghost eigenvalues, we run an additional test in a rectangular domain similar to Fig. 4, with the difference that we place ghost points on all sides of the rectangle. We denote by ϑ_S, ϑ_N, ϑ_W and ϑ_E the values of ϑ on the bottom, top, left and right side, respectively. We choose four different combinations of $(\vartheta_S, \vartheta_N, \vartheta_W, \vartheta_E)$: $(0.1, 0.1, 0.1, 0.1)$, $(0.5, 0.5, 0.5, 0.5)$, $(0.8, 0.8, 0.8, 0.8)$ and $(0.1, 0.5, 0.8, 1)$. For each combination we plot in Fig. 10 the GLT symbol in yellow and the eigenvalues (blue dots for internal eigenvalues and red dots for ghost eigenvalues, upper-left 2×2 plots). In addition, due to the different scaling described above, we multiply the ghost eigenvalues by h^{-2} (scaled eigenvalues) and re-order all eigenvalues in a non-decreasing order (upper-right 2×2 plots). With this smart normalization choice, as well evident in Fig. 10, the agreement between the eigenvalues and the GLT symbol becomes more evident. Of course, when considering the internal eigenvalues only, the agreement with GLT symbol is, as expected, even stronger (lower-right 2×2 plots). Finally, we plot the scaled eigenvalues only (lower-left 2×2 plots), observing that they are bounded away from zero independently of h. From a theoretical point of view, it should be noticed that the derivations for obtaining (57) and (58) are essentially the same as in Theorem 4, using the whole GLT machinery. The idea behind was presented in $2\mathbb{D}$ in [18] and formalized under the notion of reduced GLT matrix-sequences in [19]. Very recently Barbarino gave a systematic treatment of the reduced GLT matrix-sequences and the tools introduced in [1] are exactly those needed for proving formally that the symbols in (57) and (58) are simultaneously the GLT and spectral symbols, for the various considered cases, even if the involved matrices are globally non symmetric. In the present setting the key observation is that the rows responsible for the non symmetry show a cardinality of $O(1) = o(N)$ in $1\mathbb{D}$, of $O(N^{1/2}) = o(N)$ in $2\mathbb{D}$, and of $O(N^{(d-1)/d}) = o(N)$ for a generic dimensionality d, with N indicating the actual matrix-size. Since the related corrections have bounded spectral norm independent of the finesse parameter, the use of item **GLT2** allows to conclude also the spectral distribution, exactly as in Theorem 4.

4. Conclusions

We have provided spectral and norm estimates for matrix-sequences arising from the approximation of the Laplacian via the Coco–Russo method and we have validated them with a few numerical experiments. The analysis has involved several tools from matrix theory and in particular from the setting of Toeplitz operators and Generalized Locally Toeplitz matrix-sequences. A formal proof in the setting involving variable coefficients

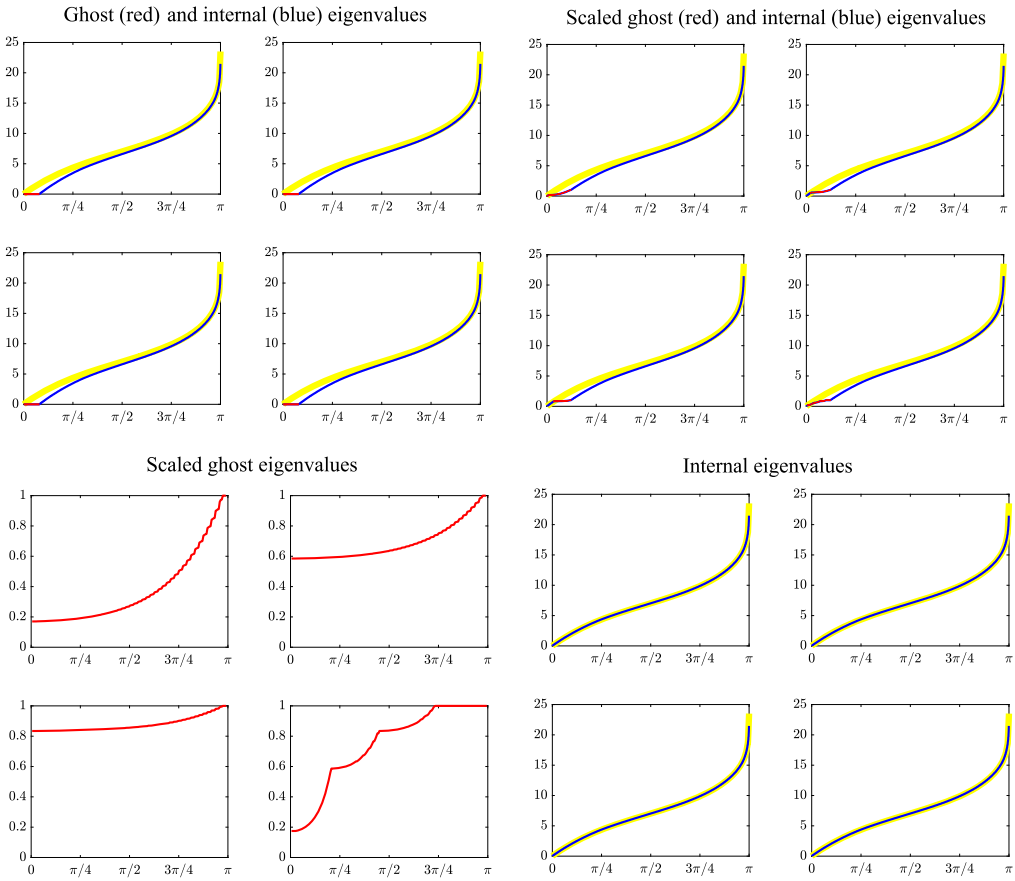

Fig. 10. GLT symbols (yellow line) and internal (blue dots) and ghost (red dots) eigenvalues of $h^2 A_h$ for a rectangular domain with ghost points on all fours sides (Section 3.5). The figure is divided in four subplots: ghost and internal eigenvalues (upper-left), scaled ghost and internal eigenvalues (upper-right), scaled ghost eigenvalues (lower-left), internal eigenvalues (lower-right). Each of these subplots is composed by 2×2 plots related to four combinations of $(\vartheta_{\mathrm{S}}, \vartheta_{\mathrm{N}}, \vartheta_{\mathrm{W}}, \vartheta_{\mathrm{E}})$: $(0.1, 0.1, 0.1, 0.1)$ (upper-left plots of each subplot), $(0.5, 0.5, 0.5, 0.5)$ (upper-right plots of each subplot), $(0.8, 0.8, 0.8, 0.8)$ (lower-left plots of each subplot) and $(0.1, 0.5, 0.8, 1)$ (lower-right plots of each subplot).

and non square domains has not been given in detail in Section 3.5, but the numerics strongly support, as expected, the conclusions. However, as previously observed, both cases of variable coefficients and non square domains can be handled from a spectral view point using the full power of the GLT machinery, as already indicated in Section 3.5, where the main proof steps have been mentioned. In particular, when considering variable coefficients, the use of the diagonal sampling matrix-sequences allows to remain in GLT $*$-algebra, while the case of non square domains can be treated using the reduced GLT theory (see page 398-399 in [18], Subsection 3.1.4 in [19], and the meticulous study in [1]).

More involved is the case of the norm estimates of the inverse even in the case of a square in $2\mathbb{D}$. Below we present an idea in this direction.

Actually the decomposition (53) suggests, as in the 1$\mathbb{D}$ setting, the use of the Sherman–Morrison–Woodbury formula: we can set $A = T_{\mathbf{n}}(f)$, $X_{\mathbf{n}} = UCV$, $\mathbf{n} = (n+1, n)$, so that

$$
\begin{aligned}
U &= \left[\frac{\mathbb{I}_n}{\mathbf{0}_{n^2\times n}}\right] \in \mathbb{R}^{n(n+1)\times n}\\
C &= \mathbb{I}_n \in \mathbb{R}^{n\times n}\\
V &= \left[T_n\left(h^2\vartheta_S - 4 + 2\cos(s)\right) | T_n\left(h^2(1-\vartheta_S)+1\right) | \mathbf{0}_{n\times n(n-1)}\right] \in \mathbb{R}^{n\times n(n+1)}\\
&= [V_1|V_2|\mathbf{0}_{n\times n(n-1)}].
\end{aligned}
$$

Hence

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

and thus $A_h^{-1} = h^2(A+UCV)^{-1} = h^2\left(A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}\right)$, with $C^{-1} = C = \mathbb{I}_n$.

The previous reasoning can be useful and promising, since the entries of the inverse of $A = T_{\mathbf{n}}(f)$, $f(s_1, s_2) = 4 - 2\cos(s_1) - 2\cos(s_2)$, are explicitly known (see [15]). However technical difficulties remain due to the complicate expression of the entries of $T_{\mathbf{n}}^{-1}(f)$: this task will be the subject of future investigations.

Declaration of competing interest

The is no competing interest.

Acknowledgements

Giovanni Russo, Stefano Serra-Capizzano and Armando Coco are grateful to GNCS INdAM for the support in the present research. Giovanni Russo acknowledges support from the Italian Ministry of Education, University and Research (MIUR), PRIN Project 2017 (No. 2017KKJP4X entitled Innovative numerical methods for evolutionary partial differential equations and applications). Furthermore, the work of Stefano Serra-Capizzano was funded from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No. 955701. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Belgium, France, Germany, Switzerland.

References

[1] G. Barbarino, A systematic approach to reduced GLT, BIT 62 (2022) 681–743.
[2] G. Barbarino, S. Serra-Capizzano, Non-Hermitian perturbations of Hermitian matrix-sequences and applications to the spectral analysis of the numerical approximation of partial differential equations, Numer. Linear Algebra Appl. 27 (3) (2020) e2286.

[3] R. Bhatia, Matrix Analysis, 1997 edition, Graduate Texts in Mathematics, Springer, New York, NY, Nov. 1996.
[4] R.H.-F. Chan, X.-Q. Jin, An Introduction to Iterative Toeplitz Solvers, SIAM, 2007.
[5] A. Chertock, A. Coco, A. Kurganov, G. Russo, A second-order finite-difference method for compressible fluids in domains with moving boundaries, Commun. Comput. Phys. 23 (2018) 230–263.
[6] A. Coco, G. Currenti, C. Del Negro, G. Russo, A second order finite-difference ghost-point method for elasticity problems on unbounded domains with applications to volcanology, Commun. Comput. Phys. 16 (4) (2014) 983–1009.
[7] A. Coco, G. Russo, Finite-difference ghost-point multigrid methods on Cartesian grids for elliptic problems in arbitrary domains, J. Comput. Phys. 241 (2013) 464–501.
[8] R.P. Fedkiw, T. Aslam, B. Merriman, S. Osher, A non-oscillatory Eulerian approach to interfaces in multimaterial flows (the ghost fluid method), J. Comput. Phys. 152 (2) (1999) 457–492.
[9] C. Garoni, S. Serra-Capizzano, Generalized Locally Toeplitz Sequences: Theory and Applications, vol. 1, Springer, 2017.
[10] C. Garoni, S. Serra-Capizzano, Generalized Locally Toeplitz Sequences: Theory and Applications, vol. 2, Springer, 2018.
[11] F. Gibou, R.P. Fedkiw, A fourth order accurate discretization for the Laplace and heat equations on arbitrary domains, with applications to the Stefan problem, J. Comput. Phys. 202 (2) (2005) 577–601.
[12] F. Gibou, R.P. Fedkiw, L.-T. Cheng, M. Kang, A second-order-accurate symmetric discretization of the Poisson equation on irregular domains, J. Comput. Phys. 176 (1) (2002) 205–227.
[13] L. Golinskii, S. Serra-Capizzano, The asymptotic properties of the spectrum of nonsymmetrically perturbed Jacobi matrix sequences, J. Approx. Theory 144 (1) (2007) 84–102.
[14] R.J. LeVeque, Z. Li, The immersed interface method for elliptic equations with discontinuous coefficients and singular sources, SIAM J. Numer. Anal. 31 (4) (1994) 1019–1044.
[15] G. Meurant, A review on the inverse of symmetric tridiagonal and block tridiagonal matrices, SIAM J. Matrix Anal. Appl. 13 (3) (1992) 707–728.
[16] M.K. Ng, Iterative Methods for Toeplitz Systems, Numerical Mathematics and Scientific Computation, 2004.
[17] C.S. Peskin, Numerical analysis of blood flow in the heart, J. Comput. Phys. 25 (3) (1977) 220–252.
[18] S. Serra-Capizzano, Generalized locally Toeplitz sequences: spectral analysis and applications to discretized partial differential equations, Linear Algebra Appl. 366 (2003) 371–402.
[19] S. Serra-Capizzano, The GLT class as a generalized Fourier analysis and applications, Linear Algebra Appl. 419 (1) (2006) 180–233.
[20] S. Serra-Capizzano, P. Tilli, On unitarily invariant norms of matrix-valued linear positive operators, J. Inequal. Appl. 7 (3) (2002) 309–330.
[21] E. Tyrtyshnikov, N. Zamarashkin, Spectra of multilevel Toeplitz matrices: advanced theory via simple matrix relationships, Linear Algebra Appl. 270 (1–3) (1998) 15–27.