Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 773-798, December 2020.
http://repository.uwyo.edu/ela

SPECTRAL ANALYSIS OF SADDLE–POINT MATRICES FROM OPTIMIZATION PROBLEMS WITH ELLIPTIC PDE CONSTRAINTS[*]

FABIO DURASTANTE[†] AND ISABELLA FURCI[‡]

Abstract. The main focus of this paper is the characterization and exploitation of the asymptotic spectrum of the saddle–point matrix sequences arising from the discretization of optimization problems constrained by elliptic partial differential equations. They uncover the existence of an hidden structure in these matrix sequences, namely, they show that these are indeed an example of Generalized Locally Toeplitz (GLT) sequences. They show that this enables a sharper characterization of the spectral properties of such sequences than the one that is available by using only the fact that they deal with saddle–point matrices. Finally, they exploit it to propose an optimal preconditioner strategy for the GMRES, and Flexible–GMRES methods.

Key words. Saddle–point matrices, Optimal control, GLT theory, Preconditioning.

AMS subject classifications. 62M15, 65F08, 15B05.

1. Introduction.

Linear systems with saddle–point matrices arises in a wide context of applications and have attracted a great deal of attention [5, 2]. In general form, they can be simply stated as the family of linear systems where the left–hand side is given by block–matrices of the form

$$\mathcal{A}_N = \begin{bmatrix} A & B_1^T \\ B_2 & -C \end{bmatrix}, \qquad A \in \mathbb{R}^{q\times q}, \quad B_1, B_2 \in \mathbb{R}^{p\times q}, \quad C \in \mathbb{R}^{p\times p}. \tag{1.1}$$

We are interested here in the analysis of their spectral properties in the very specific context of the discretized version of optimal constraint problems [33]

$$\begin{cases} \displaystyle\min_{y,u} J(y,u) = & \frac{1}{2}\|y-y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)}, \\ & e(y,u)=0, \quad \text{in } \Omega, \\ \text{such that} & y = f, \quad \text{on } \partial\Omega_D, \\ & \frac{\partial y}{\partial \mathbf{n}} = g, \quad \text{on } \partial\Omega_N, \end{cases} \tag{1.2}$$

where, $\alpha > 0$ is a fixed constant that acts as a Tikhonov regularization parameter, J is a cost functional, $\Omega \subset \mathbb{R}^d$ is the domain of both the state y and the control u, and $\partial\Omega_D$ and $\partial\Omega_N$ are two disjoint sets that represent the Dirichlet and Neumann boundary respectively and have the whole boundary as union.

Spectral properties of the general case (1.1) have been indeed thoroughly analyzed [26, 22, 6, 3, 23, 18, 31, 7] under several hypotheses on the blocks of $\mathcal{A}_N$, e.g., $B_1 = B_2 = B$, C semipositive definite, A symmetric and positive definite, and so on. The goal of the latter works has been to provide a sharp localization bounds for their spectrum, and exploit them to devise efficient iterative solvers for such problems. Here we focus on

[*] Received by the editors on December 23, 2019. Accepted for publication on October 1, 2020. Handling Editor: James G. Nagy. Corresponding Author: Isabella Furci. This work was partially supported by INdAM-GNCS project "Tecniche innovative per problemi di algebra lineare" (2018), and by the Tor Vergata University "MISSION: SUSTAINABILITY" project "NUMnoSIDS", CUP E86C18000530005.

[†] Istituto per le Applicazioni del Calcolo "Mauro Picone". Consiglio Nazionale delle Ricerche, Napoli, Italy (f.durastante@na.iac.cnr.it).

[‡] Department of Mathematics and Informatics. University of Wuppertal, Wuppertal, Germany (furci@uni-wuppertal.de).

a less general objective, i.e., we intend to exploit finer information on the structure of the blocks of (1.1), a knowledge coming from the coupling of the source problem (1.2) and its discretization, to give an asymptotic description of the spectrum of the matrices $\{\mathcal{A}_N\}_N$. Specifically, we show that the saddle–point form of $\mathcal{A}_N$ obtained from (1.1) hides inside another structure, namely, that the sequence of matrices $\{\mathcal{A}_N\}_N$ is a Generalized Locally Toeplitz (GLT) sequence [28, 16]. This enables us to obtain a sharper localization of its asymptotic spectrum. Furthermore, we use this characterization to suggest an effective preconditioning strategy for such problems. We stress that an approach of this type has already been exploited for both the saddle–point matrices obtained from a two–dimensional linear elasticity–type problem in [11], and partially explored in [10, 12] for a constrained optimization problem where the constraints $e(y, u)$ were Fractional Differential Equations.

The paper is therefore divided as follows: In Section 2, we describe the discrete form of (1.2) fully specifying the sequence of matrices $\{\mathcal{A}_N\}_N$. In Section 3, we recall the essential tools needed for working with GLT sequences and apply them to our problem, while in Section 4, we exploit them to devise an efficient preconditioning strategy. In Section 5, we substantiate our claims with some numerical examples, and give conclusions in Section 6.

2. From the continuous problem to the saddle–point sequence $\{\mathcal{A}_N\}_N$.

The first point we need to answer is how we obtain the sequence of saddle–point matrices from (1.2), indeed a way of doing so is going through its Langrangian formulation. Thus, we find the Lagrangian of (1.2) as

$$\mathcal{L}(y, u, p) = J(y, u) - \langle p, e(y, u)\rangle_{W^*,W}, \tag{2.3}$$

where $e(y, u)$ represents the PDE constraint as an operator between the Banach spaces $Y \times U$ and W, and p is the Adjoint status between the space W and its dual W^* acting as Lagrange multiplier. Indeed, a solution for the original constrained optimization problem (1.2) is a stationary point for the Lagrangian (2.3). To obtain such stationary point $(\hat{y}, \hat{u}, \hat{p}) \in Y \times U \times W^*$ we require that the Gâteaux derivative with respect to each of the variables of (2.3) is zero, i.e.,

$$\begin{aligned}
\mathcal{L}'_y(\hat{y}, \hat{u}, \hat{p})\mathfrak{h} &= J'_y(\hat{y}, \hat{u})\mathfrak{h} - \langle \hat{p}, e'_y(\hat{y}, \hat{u})\mathfrak{h}\rangle_{W^*,W} = 0, && \forall\, \mathfrak{h} \in Y,\\
\mathcal{L}'_u(\hat{y}, \hat{u}, \hat{p})\mathfrak{w} &= J'_u(\hat{y}, \hat{u})\mathfrak{w} - \langle \hat{p}, e'_u(\hat{y}, \hat{u})\mathfrak{w}\rangle_{W^*,W} = 0, && \forall\, \mathfrak{w} \in U,\\
\mathcal{L}'_p(\hat{y}, \hat{u}, \hat{p}) &= e(\hat{y}, \hat{u}) = 0.
\end{aligned}$$

These are called, in general, the first order optimality conditions or the Karush-Kuhn-Tucker conditions (KKT-conditions) for Problem (1.2). Finally, for obtaining such characterization we have to fully specify the operator $e(y, u)$, and consequently all the functional spaces Y, U, and W. The prototypical elliptic problem in this class is represented by the Poisson distributed control

$$\begin{cases}
\displaystyle\min_{y,u} J(y, u) = & \dfrac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \dfrac{\alpha}{2}\|u\|^2_{L^2(\Omega)},\\
 & -\nabla^2 y = u + z, \quad \text{in } \Omega,\\
\quad\text{such that} & y = f, \quad \text{on } \partial\Omega_D,\\
 & \frac{\partial y}{\partial \mathbf{n}} = g, \quad \text{on } \partial\Omega_N,
\end{cases} \tag{2.4}$$

where z represents the forcing term.

The KKT conditions for problem (2.4) are expressed as

$$\begin{aligned}
&\begin{cases} -\nabla^2 y = u + z, & \text{in } \Omega,\\ y = f, & \text{on } \partial\Omega_D,\\ \frac{\partial y}{\partial \mathbf{n}} = g, & \text{on } \partial\Omega_N. \end{cases} && \text{(State equation)}\\
&\begin{cases} -\nabla^2 p = y - y_d, & \text{in } \Omega,\\ y = 0, & \text{on } \partial\Omega_D,\\ \frac{\partial y}{\partial \mathbf{n}} = 0, & \text{on } \partial\Omega_N. \end{cases} && \text{(Adjoint equation)}\\
&\alpha u + p = 0. && \text{(Gradient condition)}
\end{aligned} \tag{2.5}$$

By posing $\hat{p} = -p$ and choosing $v \in H_0^1(\Omega)$ we can rewrite conditions (2.5) in weak form as:

$$\begin{aligned}
\int_\Omega \nabla u \cdot \nabla v \, dx &= \int_\Omega uv \, dx, + \int_\Omega zv \, dx,\\
\int_\Omega \nabla \hat{p} \cdot \nabla v \, dx &= \int_\Omega (y_d - y) v \, dx,\\
\alpha \int_\Omega uv \, dx - \int_\Omega \hat{p} v \, dx &= 0.
\end{aligned} \tag{2.6}$$

Finally, the sequence $\{\mathcal{A}_N\}$ is obtained by fixing a Finite Element (FEM) approximation of the optimality system (2.6). This means fixing a space $V_{0,\mathbf{n}}(\Omega_\mathbf{n})$ with $V_{0,\mathbf{n}} = \operatorname{Span}\{\phi_1, \ldots, \phi_{N(\mathbf{n})}\} \subset H_0^1(\Omega)$ over a mesh $\Omega_\mathbf{n}$ on the domain Ω thus obtaining the linear system

$$\bar{\mathcal{A}}_N \mathbf{x} \equiv \left[\begin{array}{cc|c} \bar{M} & O & \bar{K}^T \\ O & \alpha \bar{M} & -\bar{M} \\ \hline \bar{K} & -\bar{M} & O \end{array}\right] \begin{bmatrix}\mathbf{y}\\ \mathbf{u}\\ \mathbf{p}\end{bmatrix} = \begin{bmatrix} M\mathbf{y}_d \\ \mathbf{0} \\ \mathbf{z}\end{bmatrix} \equiv \bar{\mathbf{b}}, \tag{2.7}$$

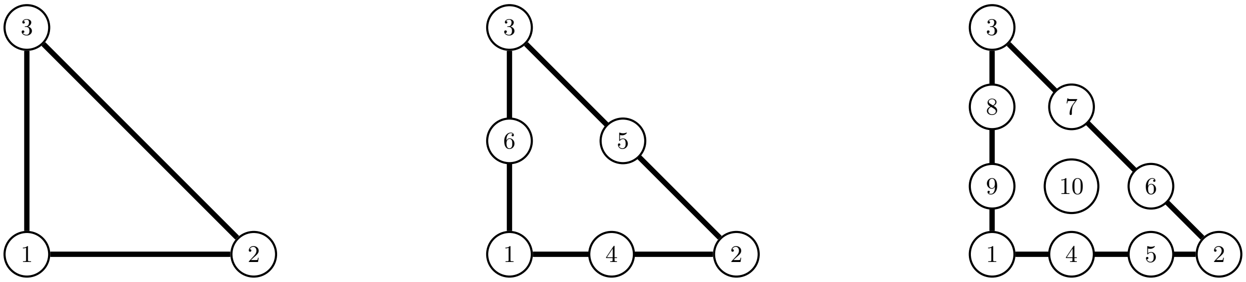
where

$$(\bar{M})_{i,j} = \int_{\tau_h} \phi_i \phi_j d\mathbf{x}, \qquad (\bar{K})_{i,j} = \int_{\tau_h} \nabla\phi_i \cdot \nabla\phi_j d\mathbf{x}, \tag{2.8}$$

are the usual (scaled) mass and stiffness matrices, and O is the zero matrix of order $N(\mathbf{n}) = n_1 n_2 \cdots n_d$.

2.1. Triangular Lagrangian elements. To completely specify the linear system (2.7) we need to precise both the mesh $\Omega_{N(\mathbf{n})}$ and the basis functions $\{\phi_j\}_{j=1}^{N(\mathbf{n})}$, i.e., chose the element defining our discretization. We focus here on nodal Lagrangian elements [9, Chapter 5] of degree p. These are built starting from $\mathbb{P}_p$, the vector space of polynomials $q(x_1, x_2)$ with scalar coefficients of $\mathbb{R}^2$ in $\mathbb{R}$ of degree less than or equal to p,

$$\mathbb{P}_p = \left\{ q(x_1, x_2) = \sum_{0 \leq i+j \leq p} c_{i,j} x_1^i x_2^j, \;\; c_{i,j} \in \mathbb{R} \right\}.$$



(a) $p = 1$ (b) $p = 2$ (c) $p = 3$

Figure 1: Nodes $\mathbf{N}_i$ for the linear ($p = 1$), quadratic ($p = 2$) and cubic ($p = 3$) Lagrange polynomials on a triangle.

That is indeed a vector space of dimension $\dim \mathbb{P}_p = \frac{1}{2}(p+1)(p+2)$. Then an homogeneous triangulation $\Omega_{N(\mathbf{n})}$ of the unit square domain $\Omega = [0,1]^2$ is considered, i.e., a mesh consisting in 2D triangular cells τ_h with straight sides, and a lattice Σ_p of nodes $\{\mathbf{N}_i\}_{i=1}^{\dim \mathbb{P}_p}$ on each triangle; see Figure 1.

By this construction, every polynomial $q \in \mathbb{P}_p$ is uniquely determined by its values at the points $\{\mathbf{N}_i\}_{i=1}^{\dim \mathbb{P}_p}$. The finite element method for triangular Lagrange $\mathbb{P}_p$ elements is then built on the discrete finite dimensional space

$$V_{\mathbf{n}}^p = \{v \in \mathcal{C}^0(\Omega)\, v|_{\tau_h} \in \mathbb{P}_p, \ \ \tau_h \in \Omega_{N(\mathbf{n})}\} \subset H^1,$$

and its subspace

$$V_{0,\mathbf{n}}^p = \{v \in V_{\mathbf{n}}^p, \ \ v = 0 \text{ on } \partial\Omega\} \subset H_0^1.$$

We call degrees of freedom of a function $v \in V_{\mathbf{n}}^p$ the set of the values of v at the nodes $\mathbf{N}_j$ on the entire mesh, then the space $V_{0,\mathbf{n}}^p$ has exactly the dimension corresponding to the number of internal degrees of freedom, i.e., excluding the nodes on $\partial\Omega$. For our model grid we find that the degrees of freedom are $N(\mathbf{n}) = n_1 n_2 = (pn_x + 1)(pn_y + 1)$, where n_x and n_y are the number of elements in the x and y direction, respectively. Thus, the dimension N of the matrix in (2.8) will be equal to $3N(\mathbf{n})$. The matrices (2.8) are then constructed by means of the opportune Gauss quadrature formulas, and in terms of the Lagrange basis functions $\{\phi_i\}_{i=1}^{N(\mathbf{n})}$. For all the discussion, and computation in the paper we deal with the matrices generated for such elements by the FEniCS library (v.2018.1.0) [1, 21].

3. Spectral analysis of the resulting sequence of saddle point matrices. This section is devoted to the attainment of a characterization of the spectra of a suitable scaling $\{\mathcal{A}_N\}_N$ of the sequence of matrices $\{\bar{\mathcal{A}}_N\}_N$ in (2.7). Specifically, we are going to answer to the following questions,

Q1 can we individuate some (possibly sharp) intervals containing the spectrum with respect to N?
Q2 For a given N how many eigenvalues are in each interval?
Q3 What is the relation between the condition number of a suitably preconditioned matrix sequence and the value of the regularization parameter α?

As we mentioned in the introduction, there exist classical localization results for the eigenvalues of a sym-

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 773-798, December 2020.
http://repository.uwyo.edu/ela

metric saddle–point matrix, like the $\mathcal{A}_N$ in (1.1).

Theorem 1 (Rusten and Winther [26]). *Given $\mathcal{A}_N$ in (1.1), assume A is symmetric and positive definite, $B_1 = B_2 = B$ has full rank, and $C = 0$. Let μ_1 and μ_n denote the largest and smallest eigenvalues of A, and let σ_1 and σ_m denote the largest and smallest singular values of B. Then the spectrum of $\mathcal{A}_N$ is contained in*

$$I^- \cup I^+,$$

where

$$I^- = \left[\frac{1}{2}\left(\mu_n - \sqrt{\mu_n^2 + 4\sigma_1^2}\right); \frac{1}{2}\left(\mu_1 - \sqrt{\mu_1^2 + 4\sigma_m^2}\right)\right], \qquad I^+ = \left[\mu_n; \frac{1}{2}\left(\mu_1 + \sqrt{\mu_1^2 + 4\sigma_1^2}\right)\right].$$

This bound is indeed very general and versatile, since it requires only information on the symmetry/definiteness of the diagonal blocks, and on the rank of the extradiagonal ones. It can be used to obtain an estimate of the condition number of $\mathcal{A}_N$ as function of N in a straightforward way. To this end, an even sharper result can be obtained by means of [3, Theorem 1(c)] that permits to characterize exactly the eigenvalues with the largest and the smallest module. Nevertheless, by exploiting further information on the blocks, we show that finer answers to our question are indeed possible. Specifically, we are going to individuate three disjoint intervals I_0^-, I_1^+, and I_2^+ containing the spectrum of the scaled version of $\bar{\mathcal{A}}_N$, we show that this choice is not arbitrary, and that it stems directly from the structure of the problem, and the selection of the discretization scheme.

In Section 3.1, we start recalling the tools we use, and then we deploy them to achieve these results in Section 3.2.

3.1. Background and definitions. Throughout this paper, we use the following notation. Let $\mathbb{C}^{s\times s}$ be the linear space of the complex $s \times s$ matrices and let $\mathbf{f} : G \to \mathbb{C}^{s\times s}$, with $G \subseteq \mathbb{R}^\ell$, $\ell \geq 1$, measurable set. We say that $\mathbf{f}$ belongs to $L^1(G)$ (resp. is measurable) if all its components $f_{ij} : G \to \mathbb{C}$, $i, j = 1, \dots, s$, belong to $L^1(G)$ (resp. are measurable). We denote by $\mathcal{I}_d$ the d-dimensional cube $(-\pi, \pi)^d$ and define $L^1(d, s)$ as the linear space of d-variate functions $\mathbf{f} : \mathcal{I}_d \to \mathbb{C}^{s\times s}$, $\mathbf{f} \in L^1(\mathcal{I}_d)$.

Moreover, we indicate by $\{\mathcal{A}_N\}_{\mathbf{n}\in\mathbb{N}^d}$, or simply $\{\mathcal{A}_N\}_{\mathbf{n}}$, the matrix sequence whose elements are the matrices $\mathcal{A}_N$ of dimensions $N \times N = N(s,\mathbf{n}) \times N(s,\mathbf{n})$, with $N(s,\mathbf{n}) = sN(\mathbf{n}) = sn_1n_2\cdots n_d$, $\mathbf{n} = (n_1, n_2, \dots, n_d)$.

Definition 1. Let the Fourier coefficients of a given function $\mathbf{f} \in L^1(d, s)$ be defined as

$$\hat{\mathbf{f}}_{\mathbf{j}} := \frac{1}{(2\pi)^d}\int_{\mathcal{I}_d} \mathbf{f}(\boldsymbol{\theta})\, \mathrm{e}^{-\iota\,\langle \mathbf{j}\,,\,\boldsymbol{\theta}\rangle}\, \mathrm{d}\boldsymbol{\theta} \in \mathbb{C}^{s\times s}, \qquad \mathbf{j} = (j_1, \dots, j_d) \in \mathbb{Z}^d, \quad \iota^2 = -1, \tag{3.9}$$

where $\langle \mathbf{j}, \boldsymbol{\theta}\rangle = \sum_{t=1}^d j_t\theta_t$ and the integrals in (3.9) are computed componentwise.

Then, the $\mathbf{n}$*th Toeplitz matrix* associated with $\mathbf{f}$ is the matrix of order $N(s,\mathbf{n})$ given by

$$T_{\mathbf{n}}(\mathbf{f}) = \sum_{\mathbf{j}=-(\mathbf{n}-\mathbf{e})}^{\mathbf{n}-\mathbf{e}} J_{n_1}^{j_1} \otimes \cdots \otimes J_{n_d}^{j_d} \otimes \hat{\mathbf{f}}_{\mathbf{j}}. \tag{3.10}$$

where $\mathbf{e} = (1, \dots, 1) \in \mathbb{N}^d$, $\mathbf{j} = (j_1, \dots, j_d) \in \mathbb{N}^d$ and $J_{n_\xi}^{j_\xi}$ is the $n_\xi \times n_\xi$ matrix whose (i, l)th entry equals 1 if $(i - l) = j_\xi$ and 0 otherwise.

The set $\{T_{\mathbf{n}}(\mathbf{f})\}_{\mathbf{n}}$ (with $\mathbf{n} \in \mathbb{N}^d$) is called the family of d-level Toeplitz matrices generated by $\mathbf{f}$, that in turn is referred to as the generating function or the symbol of $\{T_{\mathbf{n}}(\mathbf{f})\}_{\mathbf{n}}$.

Moreover, from (3.9), the symbol can be expressed via the Fourier series

$$\mathbf{f}(\boldsymbol{\theta}) = \sum_{\mathbf{j}=-\boldsymbol{\infty}}^{\boldsymbol{\infty}} \hat{\mathbf{f}}_{\mathbf{j}} e^{\iota\langle \mathbf{j},\boldsymbol{\theta}\rangle}. \tag{3.11}$$

In order to deal with low–rank/small–norm perturbations and to show that they do not affect the symbol of a Toeplitz sequence, we introduce the definition of spectral distribution in the sense of the eigenvalues and of the singular values for a generic matrix-sequence $\{\mathcal{A}_N\}_{\mathbf{n}\in\mathbb{N}^v}$, $v \geq 1$, and then the notion of GLT algebra.

Definition 2. Let $\mathbf{f} : G \to \mathbb{C}^{s\times s}$ be a measurable function, defined on a measurable set $G \subset \mathbb{R}^\ell$ with $\ell \geq 1$, $0 < \mu_\ell(G) < \infty$. Let $\mathcal{C}_0(\mathbb{K})$ be the set of continuous functions with compact support over $\mathbb{K} \in \{\mathbb{C}, \mathbb{R}_0^+\}$ and let $\{\mathcal{A}_N\}_{\mathbf{n}\in\mathbb{N}^v}$, $v \geq 1$, be a sequence of matrices with eigenvalues $\lambda_j(\mathcal{A}_N)$, $j = 1,\ldots,N$, and singular values $\sigma_j(\mathcal{A}_N)$, $j = 1,\ldots,N$.

- $\{\mathcal{A}_N\}_{\mathbf{n}\in\mathbb{N}^v}$ is *distributed as the pair* $(\boldsymbol{f}, G)$ *in the sense of the eigenvalues,* in symbols

$$\{\mathcal{A}_N\}_{\mathbf{n}\in\mathbb{N}^v} \sim_\lambda (\mathbf{f}, G),$$

if the following limit relation holds for all $F \in \mathcal{C}_0(\mathbb{C})$:

$$\lim_{\mathbf{n}\to\infty} \frac{1}{N}\sum_{j=1}^{N} F(\lambda_j(\mathcal{A}_N)) = \frac{1}{\mu_\ell(G)}\int_G \frac{\sum_{i=1}^{s} F\Big(\Big(\lambda^{(i)}(\mathbf{f})\Big)(\boldsymbol{\theta})\Big)}{s}\mathrm{d}\boldsymbol{\theta}. \tag{3.12}$$

- $\{\mathcal{A}_N\}_{\mathbf{n}\in\mathbb{N}^v}$ is *distributed as the pair* $(\boldsymbol{f}, G)$ *in the sense of the singular values,* in symbols

$$\{\mathcal{A}_N\}_{\mathbf{n}\in\mathbb{N}^v} \sim_\sigma (\mathbf{f}, G),$$

if the following limit relation holds for all $F \in \mathcal{C}_0(\mathbb{R}_0^+)$:

$$\lim_{\mathbf{n}\to\infty} \frac{1}{N}\sum_{j=1}^{N} F(\sigma_j(\mathcal{A}_N)) = \frac{1}{\mu_\ell(G)}\int_G \frac{\sum_{i=1}^{s} F\Big(\Big(\sigma^{(i)}(\mathbf{f})\Big)(\boldsymbol{\theta})\Big)}{s}\mathrm{d}\boldsymbol{\theta}. \tag{3.13}$$

In this setting the expression $\mathbf{n} \to \infty$ means that every component of the vector $\mathbf{n}$ tends to infinity, that is, $\min_{i=1,\ldots,v} n_i \to \infty$.

Remark 1. We denote by $\lambda^{(1)}(\mathbf{f}),\ldots,\lambda^{(s)}(\mathbf{f})$ and by $\sigma^{(1)}(\mathbf{f}),\ldots,\sigma^{(s)}(\mathbf{f})$ the eigenvalues and the singular values of a $s \times s$ matrix-valued function $\mathbf{f}$, respectively. If $\mathbf{f}$ is smooth enough, an informal interpretation of the limit relation (3.12) (resp. (3.13)) is that when the matrix-size of $\mathcal{A}_N$ is sufficiently large, then N/s eigenvalues (resp. singular values) of $\mathcal{A}_N$ can be approximated by a sampling of $\lambda^{(1)}(\mathbf{f})$ (resp. $\sigma^{(1)}(\mathbf{f})$) on a uniform equispaced grid of the domain G. Analogously each following N/s eigenvalues (resp. singular values) can be approximated by an equispaced sampling of the relative $\lambda^{(j)}(\mathbf{f})$ (resp. $\sigma^{(j)}(\mathbf{f})$), $j = 2,\ldots,s$, in the domain.

Remark 2. To perform the sampling in Remark 1 computing a closed analytical expression of any of the eigenvalue functions of $\mathbf{f}$ is not the most effective procedure. It is costly and, essentially, useless since for $q = 1, \dots, s$, we can provide an "exact" evaluation of $\lambda^{(q)}(\mathbf{f})$ at the grid points $\{\boldsymbol{\theta}_{\mathbf{n}} = (\theta_1^{(j)}, \theta_2^{(k)})\}_{j,k=0}^{n-1}$ without actually computing the analytical expression. Indeed the "exact" evaluation for $d = 2$ case is achieved by

1. sampling $\mathbf{f}$ at $\boldsymbol{\theta}_{\mathbf{n}-\mathbf{e}} = (\theta_{n-1}^{(j)}, \theta_{n-1}^{(k)})$, $j, k = 0, \dots, n-1$, and thus, obtain n^2 $s \times s$ matrices, $A_{j,k}$, $j, k = 0, \dots, n-1$;
2. for each $j, k = 0, \dots, n-1$, compute the s eigenvalues of $A_{j,k}$, $\lambda_q(A_{j,k})$, $q = 1, \dots, s$;
3. for a fixed $q = 1, \dots, s$, the evaluation of $\lambda^{(q)}(\mathbf{f})$ at $\boldsymbol{\theta}_{\mathbf{n}-\mathbf{e}}$, $j, k = 0, \dots, n-1$, is given by $\lambda_q(A_{j,k})$, $j, k = 0, \dots, n-1$.

3.1.1. Spectral analysis of Hermitian (block) Toeplitz sequences: distribution results. We collect here some classical results concerning the distribution of Hermitian (block) Toeplitz sequences from [19, 32], that we will use extensively in the following.

Theorem 2 (Grenander and Szegő [19]). *Let $f \in L^1(d, 1)$ be a real-valued function with $d \geq 1$. Then,*

$$\{T_{\mathbf{n}}(f)\}_{\mathbf{n}\in\mathbb{N}^d} \sim_\lambda (f, \mathcal{I}_d).$$

In the case where $\mathbf{f}$ is a Hermitian matrix-valued function, according to Tilli [32], the previous theorem can be extended as follows:

Theorem 3 (Tilli [32]). *Let $\boldsymbol{f} \in L^1(d, s)$ be a Hermitian matrix-valued function with $d \geq 1, s \geq 2$. Then,*

$$\{T_{\mathbf{n}}(\boldsymbol{f})\}_{\mathbf{n}\in\mathbb{N}^d} \sim_\lambda (\boldsymbol{f}, \mathcal{I}_d).$$

Remark 3. If $\{T_{\mathbf{n}}(\mathbf{f})\}_{\mathbf{n}\in\mathbb{N}^d}$ is such that each $T_{\mathbf{n}}(\mathbf{f})$ is symmetric with real symmetric blocks, then the symbol has the additional property that

$$\mathbf{f}(\pm\theta_1, \dots, \pm\theta_d) \equiv \mathbf{f}(\theta_1, \dots, \theta_d), \quad \forall(\theta_1, \dots, \theta_d) \in \mathcal{I}_d^+ = [0, \pi]^d,$$

and therefore, Theorem 3 can be restated as

$$\{T_{\mathbf{n}}(\mathbf{f})\}_{\mathbf{n}\in\mathbb{N}^d} \sim_\lambda (\mathbf{f}, \mathcal{I}_d^+).$$

3.1.2. GLT sequences: operative features. We list here some properties and operative features from the theory of GLT sequences in their block form; refer to [29, 15, 17] for a full account of the GLT theory.

GLT1 Each GLT sequence has a singular value symbol $\mathbf{f}(\mathbf{x}, \boldsymbol{\theta})$ for $(\mathbf{x}, \boldsymbol{\theta}) \in [0, 1]^d \times [-\pi, \pi]^d$ according to the second Item in Definition 2 with $\ell = 2d$. If the sequence is Hermitian, then the distribution also holds in the eigenvalue sense. If $\{\mathcal{A}_N\}_N$ has a GLT symbol $\mathbf{f}(\mathbf{x}, \boldsymbol{\theta})$ we will write $\{\mathcal{A}_N\}_N \sim_{\text{GLT}} \mathbf{f}(\mathbf{x}, \boldsymbol{\theta})$.

GLT2 The set of GLT sequences form a $*$-algebra, i.e., it is closed under linear combinations, products, inversion (whenever the symbol is singular, at most, in a set of zero Lebesgue measure), and conjugation. Hence, the sequence obtained via algebraic operations on a finite set of given GLT sequences is still a GLT sequence and its symbol is obtained by performing the same algebraic manipulations on the corresponding symbols of the input GLT sequences.

GLT3 Every Toeplitz sequence generated by an $L^1(d,s)$ function $\mathbf{f}=\mathbf{f}(\boldsymbol{\theta})$ is a GLT sequence and its symbol is $\mathbf{f}$, with the specifications reported in item **GLT1**. We note that the function $\mathbf{f}$ does not depend on the space variables $\mathbf{x}\in[0,1]^d$.

GLT4 Every sequence which is distributed as the constant zero in the singular value sense is a GLT sequence with symbol 0. In particular:

- every sequence in which the rank divided by the size tends to zero, as the matrix size tends to infinity;
- every sequence in which the trace-norm (i.e., sum of the singular values) divided by the size tends to zero, as the matrix size tends to infinity.

GLT5 If $\{\mathcal{A}_N\}_N \sim_{\rm GLT} \kappa$ and the matrices $\mathcal{A}_N$ are such that $\mathcal{A}_N=\mathcal{X}_N+\mathcal{Y}_n$, where

- every $\mathcal{X}_N$ is Hermitian,
- the spectral norms of $\mathcal{X}_N$ and $\mathcal{Y}_N$ are uniformly bounded with respect to N,
- the trace-norm of $\mathcal{Y}_N$ divided by the matrix size N converges to 0,

then the distribution holds in the eigenvalue sense.

We highlight that from the previous properties follows that a sequence of Toeplitz matrices is, up to low-rank corrections, a GLT sequence whose symbol is not affected by the low-rank perturbation.

Theorem 4. *[16, Section 8.4] Let $\{A_N\}_N$ be a sequence of Hermitian matrices such that $\{A_N\}_N \sim_{GLT} \kappa$, and let $\{P_N\}_N$ be a sequence of Hermitian positive definite matrices such that $\{P_N\}_N \sim_{GLT} \xi$ and $\xi\neq 0$ a.e. Then*

$$\{P_N^{-1}A_N\}_N \sim_{\rm GLT} \xi^{-1}\kappa, \qquad \{P_N^{-1}A_N\}_N \sim_{\sigma,\,\lambda} (\xi^{-1}\kappa,\mathcal{I}^d).$$

3.2. Spectral analysis of the sequence $\{\mathcal{A}_N\}_N$. We can now use the introduced tools to perform the spectral analysis of the matrix sequence $\{\bar{\mathcal{A}}_N\}_N$, assuming that $n=n_1=n_2$, $p=1$. For studying it is easier to consider the equivalent distribution given by the following symmetric diagonal scaling

$$\mathcal{A}_N=\mathcal{D}_N^{(1)}\bar{\mathcal{A}}_N\mathcal{D}_N^{(2)}=\begin{bmatrix} h^4M & O & K^T\\ O & \alpha M & -M\\ K & -M & O\end{bmatrix}, \qquad h=\frac{1}{n+1}, \tag{3.14}$$

with

$$\mathcal{D}_N^{(1)}=\begin{bmatrix} h^2I_{n^2} & O & O\\ O & I_{n^2} & O\\ O & O & I_{n^2}\end{bmatrix}, \qquad \mathcal{D}_N^{(2)}=\begin{bmatrix} I_{n^2} & O & O\\ O & \frac{1}{h^2}I_{n^2} & O\\ O & O & \frac{1}{h^2}I_{n^2}\end{bmatrix}.$$

From the discretization of Section 2, the elements of the matrix $\bar{M}$ depend on n as $1/(n+1)^2$. Hence, the effect of the proposed scaling permits to eliminate the dependence of h^2 of the elements in $\bar{M}$, which, for n large, would make the matrix $\mathcal{A}_N$ ill-conditioned.

In particular the matrices $M=\frac{1}{h^2}\bar{M}=T_{\mathbf{n}}(m)$, $K=\bar{K}=T_{\mathbf{n}}(\kappa)$ are $n^2\times n^2$ bi-level Toeplitz matrices with generating functions

$$m(\theta_1,\theta_2)=\frac{\cos(\theta_1)}{6}+\frac{\cos(\theta_2)}{6}+\frac{1}{6}\cos(\theta_1+\theta_2)+\frac{1}{2} \tag{3.15}$$

and

$$\kappa(\theta_1,\theta_2)=-2\cos(\theta_1)-2\cos(\theta_2)+4. \tag{3.16}$$

We stress that in this case the matrices M and K are real and symmetric. A property that we will exploit in the theoretical analysis, nevertheless we keep the notation K^T for the (1,3) block of the matrix $\mathcal{A}_N$ for two reasons. On one side, for being consistent with the continuous setting, in which the adjoint is usually explicitly expressed. On the other, to keep the analogy with Section 3.3 in which we will discuss the usage of the advection-diffusion equation as constraint.

Theorem 5. *The matrix sequence $\{\mathcal{A}_N\}_N$ in (3.14) is distributed in the sense of the Eigenvalues as*

$$\boldsymbol{f}(\theta_1,\theta_2) = \hat{\boldsymbol{f}}_{(0,0)} + 2\hat{\boldsymbol{f}}_{(0,-1)}\left(\cos\theta_1+\cos\theta_2\right) + 2\hat{\boldsymbol{f}}_{(-1,-1)}\left(\cos(\theta_1+\theta_2)\right), \tag{3.17}$$

i.e., $\{\mathcal{A}_N\}_N \sim_\lambda (\boldsymbol{f}, [0,\pi]^2)$, where

$$\hat{\boldsymbol{f}}_{(0,0)} = \begin{bmatrix} 0 & 0 & 4 \\ 0 & \frac{\alpha}{2} & -\frac{1}{2} \\ 4 & -\frac{1}{2} & 0 \end{bmatrix}, \quad \hat{\boldsymbol{f}}_{(1,1)} = \hat{\boldsymbol{f}}_{(-1,-1)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{\alpha}{12} & -\frac{1}{12} \\ 0 & -\frac{1}{12} & 0 \end{bmatrix},$$
$$\hat{\boldsymbol{f}}_{(-1,0)} = \hat{\boldsymbol{f}}_{(0,-1)} = \hat{\boldsymbol{f}}_{(0,1)} = \hat{\boldsymbol{f}}_{(1,0)} = \begin{bmatrix} 0 & 0 & -1 \\ 0 & \frac{\alpha}{12} & -\frac{1}{12} \\ -1 & -\frac{1}{12} & 0 \end{bmatrix}. \tag{3.18}$$

Proof. Let $\mathbf{e}_i$, $i = 1,\dots,N$ be the ith column of the identity matrix of size N, we can define a proper $N\times N$ permutation matrix, $\Pi = [P_1|P_2|P_3]$, $P_l \in \mathbb{R}^{N\times n^2}$, $l = 1,2,3$, such that the kth column of P_l $l = 1,2,3$, is $e_{l+3(k-1)}$. The matrix Π transforms $\mathcal{A}_N$ as

$$B_N = \Pi\mathcal{A}_N\Pi^T = T_{\mathbf{n}}(\mathbf{f}) + E_{\mathbf{n}}, \tag{3.19}$$

where

- $T_{\mathbf{n}}(\mathbf{f})$ is the bi-level 3×3 block Toeplitz $T_{\mathbf{n}}(\mathbf{f}) = \left[\hat{\mathbf{f}}_{\mathbf{i}-\mathbf{j}}\right]_{\mathbf{i},\mathbf{j}=\mathbf{e}}^{\mathbf{n}} \in \mathbb{C}^{N\times N}$ generated by $\mathbf{f} : [-\pi,\pi]^2 \to \mathbb{C}^{3\times 3}$ as in (3.11),
- $E_{\mathbf{n}}$ is a small-norm matrix, with $||E_{\mathbf{n}}|| < C$, C constant depending on the bandwidths of B_N and $N^{-1}\|E_{\mathbf{n}}\|_1 \to 0$.

This is a congruence transformation, thus if we find the distribution of the sequence $\{B_N\}_N$, we found also the distribution for the sequence $\{\mathcal{A}_N\}_N$. Let us observe that the nonzero entries of $T_{\mathbf{n}}(\mathbf{f}) = [\hat{\mathbf{f}}_{\mathbf{i}-\mathbf{j}}]_{\mathbf{i},\mathbf{j}=\mathbf{e}}^{\mathbf{n}}$ correspond to the indexes $\mathbf{i} = (i_1,i_2), \mathbf{j} = (j_1,j_2)$ satisfying

$$\{|i_1-j_1| + |i_2-j_2| \le 1\} \cup \{i_1 = i_2 = j_1 = j_2 = 1\} \cup \{i_1 = i_2 = j_1 = j_2 - 1\},$$

as shown in equation (3.20), for $\mathbf{n} = (3,3)$ we find $T_{\mathbf{n}}(\mathbf{f})$

$$\left[\begin{array}{ccc|ccc|ccc}
\hat{\mathbf{f}}_{(0,0)} & \hat{\mathbf{f}}_{(0,-1)} & 0 & \hat{\mathbf{f}}_{(-1,0)} & \hat{\mathbf{f}}_{(-1,-1)} & 0 & 0 & 0 & 0 \\
\hat{\mathbf{f}}_{(0,1)} & \hat{\mathbf{f}}_{(0,0)} & \hat{\mathbf{f}}_{(0,-1)} & 0 & \hat{\mathbf{f}}_{(-1,0)} & \hat{\mathbf{f}}_{(-1,-1)} & 0 & 0 & 0 \\
0 & \hat{\mathbf{f}}_{(0,1)} & \hat{\mathbf{f}}_{(0,0)} & 0 & 0 & \hat{\mathbf{f}}_{(-1,0)} & 0 & 0 & 0 \\
\hline
\hat{\mathbf{f}}_{(1,0)} & 0 & 0 & \hat{\mathbf{f}}_{(0,0)} & \hat{\mathbf{f}}_{(0,-1)} & 0 & \hat{\mathbf{f}}_{(-1,0)} & \hat{\mathbf{f}}_{(-1,-1)} & 0 \\
\hat{\mathbf{f}}_{(1,1)} & \hat{\mathbf{f}}_{(1,0)} & 0 & \hat{\mathbf{f}}_{(0,1)} & \hat{\mathbf{f}}_{(0,0)} & \hat{\mathbf{f}}_{(0,-1)} & 0 & \hat{\mathbf{f}}_{(-1,0)} & \hat{\mathbf{f}}_{(-1,-1)} \\
0 & \hat{\mathbf{f}}_{(1,1)} & \hat{\mathbf{f}}_{(1,0)} & 0 & \hat{\mathbf{f}}_{(0,1)} & \hat{\mathbf{f}}_{(0,0)} & 0 & 0 & \hat{\mathbf{f}}_{(-1,0)} \\
\hline
0 & 0 & 0 & \hat{\mathbf{f}}_{(1,0)} & 0 & 0 & \hat{\mathbf{f}}_{(0,0)} & \hat{\mathbf{f}}_{(0,-1)} & 0 \\
0 & 0 & 0 & \hat{\mathbf{f}}_{(1,1)} & \hat{\mathbf{f}}_{(1,0)} & 0 & \hat{\mathbf{f}}_{(0,1)} & \hat{\mathbf{f}}_{(0,0)} & \hat{\mathbf{f}}_{(0,-1)} \\
0 & 0 & 0 & 0 & \hat{\mathbf{f}}_{(1,1)} & \hat{\mathbf{f}}_{(1,0)} & 0 & \hat{\mathbf{f}}_{(0,1)} & \hat{\mathbf{f}}_{(0,0)}
\end{array}\right] \tag{3.20}$$

Therefore, from (3.11), the generating function $\mathbf{f}$ is given by the finite sum

$$\begin{aligned}\mathbf{f}(\theta_1,\theta_2) = & \hat{\mathbf{f}}_{(0,0)} + \hat{\mathbf{f}}_{(-1,0)}e^{-\mathbf{i}\theta_1} + \hat{\mathbf{f}}_{(0,-1)}e^{-\mathbf{i}\theta_2} + \hat{\mathbf{f}}_{(1,0)}e^{\mathbf{i}\theta_1} + \hat{\mathbf{f}}_{(0,1)}e^{\mathbf{i}\theta_2} + \\ & + \hat{\mathbf{f}}_{(-1,-1)}e^{-\mathbf{i}(\theta_1+\theta_2)} + \hat{\mathbf{f}}_{(1,1)}e^{\mathbf{i}(\theta_1+\theta_2)},\end{aligned} \tag{3.21}$$

where $\hat{\mathbf{f}}_{(0,0)}, \hat{\mathbf{f}}_{(-1,0)}, \hat{\mathbf{f}}_{(0,-1)}, \hat{\mathbf{f}}_{(1,0)}, \hat{\mathbf{f}}_{(0,1)}, \hat{\mathbf{f}}_{(1,1)}, \hat{\mathbf{f}}_{(-1,-1)} \in \mathbb{R}^{3\times 3}$, that is $\mathbf{f}$ is a linear trigonometric polynomial in the variables θ_1 and θ_2 with matrix coefficients from (3.18). Moreover, using the equalities in (3.18), the symbol in (3.21) can be readily simplified as

$$\begin{aligned}\mathbf{f}(\theta_1,\theta_2) = & \hat{\mathbf{f}}_{(0,0)} + \hat{\mathbf{f}}_{(0,-1)}e^{-\mathbf{i}\theta_1} + \hat{\mathbf{f}}_{(0,-1)}e^{-\mathbf{i}\theta_2} + \hat{\mathbf{f}}_{(0,-1)}e^{\mathbf{i}\theta_1} + \hat{\mathbf{f}}_{(0,-1)}e^{\mathbf{i}\theta_2} + \\ & + \hat{\mathbf{f}}_{(-1,-1)}e^{-\mathbf{i}(\theta_1+\theta_2)} + \hat{\mathbf{f}}_{(-1,-1)}e^{\mathbf{i}(\theta_1+\theta_2)} \\ = & \hat{\mathbf{f}}_{(0,0)} + \hat{\mathbf{f}}_{(0,-1)}(e^{-\mathbf{i}\theta_1} + e^{\mathbf{i}\theta_1} + e^{-\mathbf{i}\theta_2} + e^{\mathbf{i}\theta_2}) + \hat{\mathbf{f}}_{(-1,-1)}(e^{-\mathbf{i}(\theta_1+\theta_2)} + e^{\mathbf{i}(\theta_1+\theta_2)}) \\ = & \hat{\mathbf{f}}_{(0,0)} + 2\hat{\mathbf{f}}_{(0,-1)}\left(\cos\theta_1 + \cos\theta_2\right) + 2\hat{\mathbf{f}}_{(-1,-1)}\left(\cos(\theta_1+\theta_2)\right).\end{aligned}$$

Note, from the latter, that

$$\mathbf{f}^T(\theta_1,\theta_2) = \mathbf{f}(\theta_1,\theta_2),$$

thus $\mathbf{f}$ is a symmetric matrix-valued function which implies that $T_{\mathbf{n}}(\mathbf{f})$ is a symmetric matrix. By Theorem 3, we conclude that

$$\{T_{\mathbf{n}}(\mathbf{f})\}_{\mathbf{n}} \sim_{\lambda} (\mathbf{f}, [-\pi,\pi]^2). \tag{3.22}$$

While, from **GLT3**, we know that $\{T_{\mathbf{n}}(\mathbf{f})\}_{\mathbf{n}}$ is a GLT sequence with symbol $\mathbf{f}$. Moreover, let us observe that $\{E_{\mathbf{n}}\}$ is a zero–distributed sequence hence $\{E_{\mathbf{n}}\}_{\mathbf{n}} \sim_{\sigma} (\mathbf{0}, \mathcal{I}_2^+)$. Indeed, $E_{\mathbf{n}}$ is the permutation of a matrix that in block position (1,1) collects all the terms that contains the scaling h^4, deriving from the (1,1) block of $\mathcal{A}_N$, and 0 anywhere else. Then it can be written as $E_{\mathbf{n}} = h^4\tilde{E}_{\mathbf{n}}$.

Since the trace norm $\|\cdot\|_1$ of $\tilde{E}_{\mathbf{n}}$ is equal to a constant C independent on $\mathbf{n}$, we have

$$\lim_{\mathbf{n}\to\infty} N^{-1}||E_{\mathbf{n}}||_1 = \lim_{\mathbf{n}\to\infty} N^{-1}\sum_{i=1}^{N}\sigma_i(E_{\mathbf{n}}) \le \lim_{\mathbf{n}\to\infty} N^{-1}\sigma_{\max}(E_{\mathbf{n}})N = 0,$$

and hence, the zero–distribution follows from **GLT4**. In addition, from **GLT1** and the fact that $E_{\mathbf{n}}$ is Hermitian, $\{E_{\mathbf{n}}\}_{\mathbf{n}} \sim_\lambda (\mathbf{0}, \mathcal{I}_2^+)$.

The conclusion of the theorem is then achieved by applying **GLT2** and (3.22), since this proves that $\{T_{\mathbf{n}}(\mathbf{f}) + E_{\mathbf{n}}\}_{\mathbf{n}\in\mathbb{N}^2}$ is a GLT sequence with symbol $\mathbf{f}$, i.e., $\{\mathcal{A}_N\}_N \sim_{\mathrm{GLT}} \mathbf{f}$. Consequently, by recalling that $T_{\mathbf{n}}(\mathbf{f}) + E_{\mathbf{n}}$ is real symmetric for every $\mathbf{n}$ and using **GLT1**, we deduce that the distribution result holds in the sense of the eigenvalues

$$\{B_N\}_N \sim_\lambda (\mathbf{f}, [-\pi, \pi]^2). \tag{3.23}$$

Furthermore, since each B_N is symmetric and its blocks are symmetric and real, then $\mathbf{f}$ is such that $\mathbf{f}(\pm\theta_1, \pm\theta_2) \equiv \mathbf{f}(\theta_1, \theta_2)$, $\forall(\theta_1, \theta_2) \in [0, \pi]^2$, and therefore, (3.23) can be rephrased as

$$\{B_N\}_N \sim_\lambda (\mathbf{f}, \mathcal{I}_2^+). \tag{3.24}$$ □

We can now find a first answer to the questions *Q1* and *Q2*. For N sufficiently large, let

$$\lambda_1(B_N) \le \lambda_2(B_N) \le \cdots \le \lambda_N(B_N).$$

be the eigenvalues of B_N from (3.19), i.e., of $\mathcal{A}_N$. By Remark 1, with $s = 3$, and equation (3.24), we discover that $N/3 = n^2$ eigenvalues of B_N, up to a number of outliers infinitesimal in the dimension, can be approximated by a sampling of $\lambda^{(1)}(\mathbf{f})$ on an opportune grid (see the following discussion). The next $N/3$ on the second one and the last n^2 on the sampling of $\lambda^{(3)}(\mathbf{f})$. Moreover, obtaining the following proposition, as a specialized version of Theorem 1, is straightforward.

Proposition 1. *Let $m_i = \operatorname{ess\,inf}_{\mathcal{I}_2^+} \lambda^{(i)}(\boldsymbol{f}(\boldsymbol{\theta}))$ and $M_i = \operatorname{ess\,sup}_{\mathcal{I}_2^+} \lambda^{(i)}(\boldsymbol{f}(\boldsymbol{\theta}))$ be the essential infimum and essential supremum of $\lambda^{(i)}(\boldsymbol{f}(\boldsymbol{\theta}))$ respectively, for $i = 1, 2, 3$. Then, for N sufficiently large, the spectrum $\lambda(\mathcal{A}_N)$ of the matrix sequence $\{\mathcal{A}_N\}_N$ is contained in three intervals*

$$\begin{aligned}
\lambda(\mathcal{A}_N) \subset I_0^- \cup I_1^+ \cup I_2^+ =& (\operatorname*{ess\,inf}_{\mathcal{I}_2^+} \lambda^{(1)}(\boldsymbol{f}(\boldsymbol{\theta})), \operatorname*{ess\,sup}_{\mathcal{I}_2^+} \lambda^{(1)}(\boldsymbol{f}(\boldsymbol{\theta}))] \\
&\cup (\operatorname*{ess\,inf}_{\mathcal{I}_2^+} \lambda^{(2)}(\boldsymbol{f}(\boldsymbol{\theta})), \operatorname*{ess\,sup}_{\mathcal{I}_2^+} \lambda^{(2)}(\boldsymbol{f}(\boldsymbol{\theta}))] \\
&\cup [\operatorname*{ess\,inf}_{\mathcal{I}_2^+} \lambda^{(3)}(\boldsymbol{f}(\boldsymbol{\theta})), \operatorname*{ess\,sup}_{\mathcal{I}_2^+} \lambda^{(3)}(\boldsymbol{f}(\boldsymbol{\theta}))) \\
=& (m_1, M_1] \cup (m_2, M_2] \cup [m_3, M_3),
\end{aligned}$$

for $\mathcal{I}_2^+ = [0, \pi]^2$.

Proof. From the definition of $\mathbf{f}$ in (3.17), $\forall\,(\theta_1, \theta_2) \in [0, \pi]^2$, and matching with the classical analysis for saddle–point matrices in Theorem 1, we find

$$\left(\lambda^{(1)}(\mathbf{f})\right)(\theta_1, \theta_2) < 0 \le \left(\lambda^{(2)}(\mathbf{f})\right)(\theta_1, \theta_2) < \left(\lambda^{(3)}(\mathbf{f})\right)(\theta_1, \theta_2), \tag{3.25}$$

i.e.,

$$M_1 < m_2, \qquad M_2 < m_3. \tag{3.26}$$

and

$$\begin{aligned}
&\operatorname*{ess\,sup}_{\mathcal{I}_2^+} \lambda^{(1)}(\mathbf{f}(\boldsymbol{\theta})) \le \operatorname*{ess\,inf}_{\mathcal{I}_2^+} \lambda^{(2)}(\mathbf{f}(\boldsymbol{\theta})), \\
&\operatorname*{ess\,sup}_{\mathcal{I}_2^+} \lambda^{(2)}(\mathbf{f}(\boldsymbol{\theta})) \le \operatorname*{ess\,inf}_{\mathcal{I}_2^+} \lambda^{(3)}(\mathbf{f}(\boldsymbol{\theta})).
\end{aligned} \tag{3.27}$$

From [27, Theorem 2.3], we know that the thesis holds true for $T_{\mathbf{n}}(\mathbf{f})$ and, from the relation $\{\mathcal{A}_N\}_N \sim_\lambda (\mathbf{f}, [0,\pi]^2)$ of Theorem 5, we have that asymptotically the inclusion in (3.27) is valid, also involving the small norm correction. □

To deliver an actual numerical estimate for these bounds what we need is a reasonable approximation of the eigenvalue functions $\lambda^{(l)}(\mathbf{f})$, $l = 1,2,3$, following the procedure from Remark 2 and exploiting Theorem 5, we define the following equispaced grid on $\mathcal{I}_2^+$

$$\boldsymbol{\theta}_{\mathbf{n}-\mathbf{e}} = \left\{ (\theta_{n-1}^{(j)}, \theta_{n-1}^{(k)}) = \left(\frac{j\pi}{n}, \frac{k\pi}{n} \right), \quad j,k = 0,\dots,n-1 \right\},$$

and consider the following n^2 Hermitian matrices of size 3×3

$$A_{j,k} := \mathbf{f}(\theta_{n-1}^{(j)}, \theta_{n-1}^{(k)}), \quad j,k = 0,\dots,n-1. \tag{3.28}$$

Ordering in ascending way the eigenvalues of $A_{j,k}$

$$\lambda_1(A_{j,k}) \le \lambda_2(A_{j,k}) \le \lambda_3(A_{j,k}), \quad j,k = 0,\dots,n-1,$$

for any $l = 1,2,3$, an evaluation of $\lambda^{(l)}(\mathbf{f})$ at $(\theta_1^{(j)}, \theta_2^{(k)})$ is given by $\lambda_l(A_{j,k})$, $j,k = 1,\dots,n$. For a fixed l, we denote the vector of all eigenvalues $\lambda_l(A_{j,k})$, $j,k = 0,\dots,n-1$ as $\mathbf{P}_l^{(n)}$, i.e.,

$$\mathbf{P}_l^{(n)} := [\lambda_l(A_{0,0}), \lambda_l(A_{0,1}), \dots, \lambda_l(A_{n-1,n-1})],$$

and by $\mathbf{P}^{(n)}$ the vector of all eigenvalues $\lambda_l(A_{j,k})$, $j,k = 0,\dots,n-1$ varying l, i.e.,

$$\mathbf{P}^{(n)} := [\lambda_1(A_{0,0}), \dots, \lambda_1(A_{n-1,n-1}), \dots, \lambda_3(A_{0,0}), \dots, \lambda_3(A_{n-1,n-1})].$$

Note that, refining the grid by increasing n, we can provide the evaluation of the eigenvalue functions of $\mathbf{f}$ in a larger number of grid points: numerical evidences of this fact are reported in Figure 2, in which we compare the approximation of $\lambda^{(l)}(\mathbf{f})$ on $\boldsymbol{\theta}_{\mathbf{n}}$, $n = 5,6$ contained in $\mathbf{P}_l^{(n)}$ (ordered in ascending way) with the approximation of the same eigenvalue function on a grid that is twice as fine $\boldsymbol{\theta}_{\mathbf{2n}-\mathbf{e}}$, $n = 5,6$ contained in $\mathbf{P}_l^{(2n)}$ (ordered in ascending way as well) for every $l = 1,2,3$.

Then, for n sufficiently large, if we order in ascending way $\mathbf{P}_l^{(n)}$, its extremes satisfy the following relations

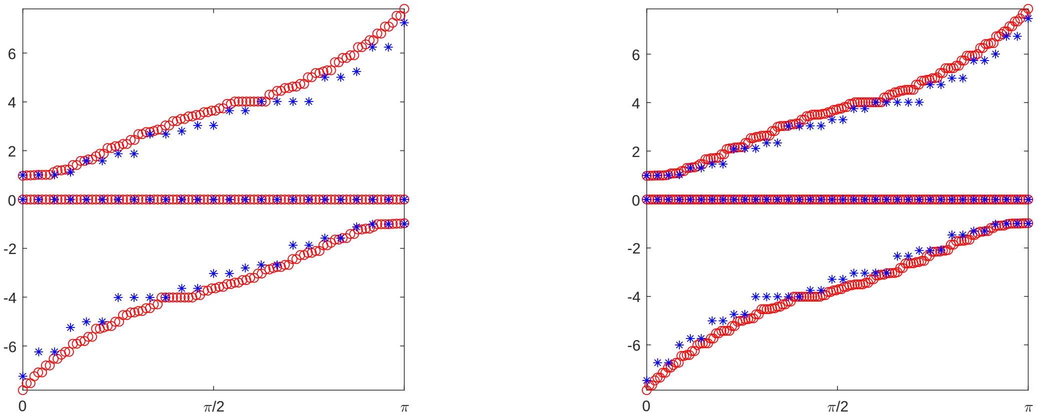
$$(\mathbf{P}_l^{(n)})_1 \approx m_l, \quad (\mathbf{P}_l^{(n)})_{n^2} \approx M_l, \qquad l = 1,2,3,$$

and we can can compute a satisfactory approximation of the $\{m_l, M_l\}_{l=1}^3$ from Proposition 1, e.g., by setting $n = 3 \cdot 10^3$, and $\alpha =$ `1.0e-04`, we obtain the following approximations:

$$\begin{aligned} \{m_1, M_1\} &\approx \{-8.006939205138657, -0.971179393341684\}, \\ \{m_2, M_2\} &\approx \{0, 0.00006086664699\}, \\ \{m_3, M_3\} &\approx \{0.971268643759555, 8.006939262908668\}. \end{aligned}$$

This clearly matches with the fact that the matrix–valued symbol is analytically singular in $(0,0)$, i.e.,

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



(a) $n = 5$

(b) $n = 6$

Figure 2: Comparison between the evaluation of the eigenvalue functions $\lambda^{(l)}(\mathbf{f})$, $l = 1, 2, 3$, ordered in ascending way, on the grid $\boldsymbol{\theta}_{\mathbf{n-e}}$ contained in $P_l^{(n)}$ ($\circ$) and the corresponding evaluation on the grid twice as fine $\boldsymbol{\theta}_{\mathbf{2n-e}}$ contained in $P_l^{(2n)}$ ($*$). Each 'curve' refers to a different value of l. The parameter n equals 5 and 6 in subplots (a) and (b), respectively.

hence $m_2 = 0$, nevertheless we stress again that this is not in contradiction with the fact that $\mathcal{A}_N$ is non singular.

In conclusion, we can exploit Remark 1, to provide an answer to *Q2* determining how many eigenvalues are asymptotically contained in each of the three blocks. According to the relations (3.24), (3.26) we expect the eigenvalues of B_N to verify

$$\begin{aligned}
\#\left\{i \,:\, \lambda_i(B_N) \in (m1, M_1]\right\} &= \frac{3n^2}{3} + o(3n^2),\\
\#\left\{i \,:\, \lambda_i(B_N) \in (m_2, M_2]\right\} &= \frac{3n^2}{3} + o(3n^2),\\
\#\left\{i \,:\, \lambda_i(B_N) \in [m_3, M_3)\right\} &= \frac{3n^2}{3} + o(3n^2).
\end{aligned} \tag{3.29}$$

and then to identify 3 blocks

$$\begin{aligned}
\mathrm{Bl}_1 &= \left[\lambda_1(B_N), \ldots, \lambda_{n^2}(B_N)\right],\\
\mathrm{Bl}_2 &= \left[\lambda_{n^2+1}(B_N), \ldots, \lambda_{2n^2}(B_N)\right],\\
\mathrm{Bl}_3 &= \left[\lambda_{2n^2+1}(B_N), \ldots, \lambda_{3n^2}(B_N)\right].
\end{aligned}$$

Correspondingly, we can split the vector $\mathbf{P}^{(n)}$ containing the sampling of the eigenvalue functions on $\boldsymbol{\theta}_{\mathbf{n-e}}$

as follows

$$\begin{aligned}
\mathrm{Eval}_1 &= [(\mathbf{P}^{(n)})_1, \dots, (\mathbf{P}^{(n)})_{n^2}],\\
\mathrm{Eval}_2 &= [(\mathbf{P}^{(n)})_{n^2+1}, \dots, (\mathbf{P}^{(n)})_{2n^2}],\\
\mathrm{Eval}_3 &= [(\mathbf{P}^{(n)})_{2n^2+1}, \dots, (\mathbf{P}^{(n)})_{3n^2}].
\end{aligned}$$

We stress again that (3.29) allows for a number of outliers that is infinitesimal in the dimension N.

For example, for $\mathbf{n} = (n, n) = (40, 40)$ $(N = 4800)$, approximately $\frac{3n^2}{3} = 1600$ eigenvalues should be in each block, by a straightforward numerical check one obtains

$$\begin{aligned}
\#\left\{i \,:\, \lambda_i(B_N) \in (m1_{,}M_1]\right\} &= 1600,\\
\#\left\{i \,:\, \lambda_i(B_N) \in (m_2, M_2]\right\} &= 1421,\\
\#\left\{i \,:\, \lambda_i(B_N) \in [m_3, M_3)\right\} &= 1600.
\end{aligned} \tag{3.30}$$

Therefore, we expect from that a certain number of eigenvalues of B_N are in none of the blocks; in the example the effective 1421 eigenvalues against the expected 1600 in the second block. This is confirmed again by Figure 3 in which we highlight represent in blue the whole spectrum of B_N and highlight in black the outliers not belonging to the blocks. On the other hand, such a phenomenon is in line with (3.29), since

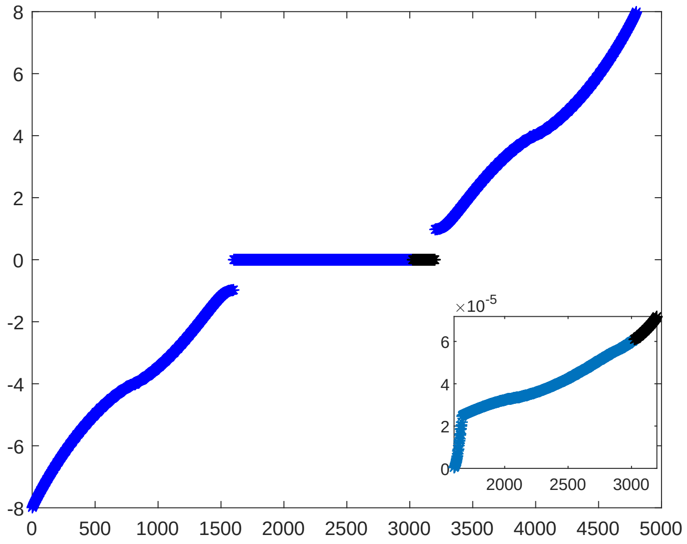


Figure 3: Eigenvalues of B_N for $\mathbf{n} = (n, n) = (40, 40)$ $(*)$ together with the eigenvalues of B_N satisfying one of the relations (3.29) $(*)$, for $\alpha =$ `1.0e-04`.

the order of what is missing/exceeding is infinitesimal in the dimension N. As an example, in Table 1 we compare the actual number of eigenvalues of B_N contained in the second interval $(m_2, M_2]$ with the expected number n^2. In such way, we succeed in counting the outliers of B_N in $(m_2, M_2]$, whose cardinality behaves as $O(\sqrt{3n^2})$. A further and more natural evidence of relation (3.24) can be obtained by comparing block by block the eigenvalues of B_N with the sampling of the eigenvalue functions of $\mathbf{f}$, that is comparing Bl_1,

n	$\#\{\lambda \in (m_2, M_2]\}$	n^2	$\#\{\lambda \notin (m_2, M_2]\}$	$\#\{\lambda \notin (m_2, M_2]\}/\sqrt{3n^2}$
10	74	100	26	0.086
20	353	400	47	0.039
40	1421	1600	179	0.037
80	5694	6400	706	0.036

Table 1: Comparison of the effective number of eigenvalues of B_N contained in the second interval $(m_2, M_2]$ with the expected number n^2.

Bl_2, Bl_3, with Eval_1, Eval_2, Eval_3, respectively. Indeed we want to compare the eigenvalues of B_N (properly ordered) with the evaluation of $\lambda^{(l)}(\mathbf{f})$ $l = 1, 2, 3$ at $\boldsymbol{\theta}_{\mathbf{n}-\mathbf{e}}$, using the values that are present in the blocks of $\mathbf{P}^{(n)}$.

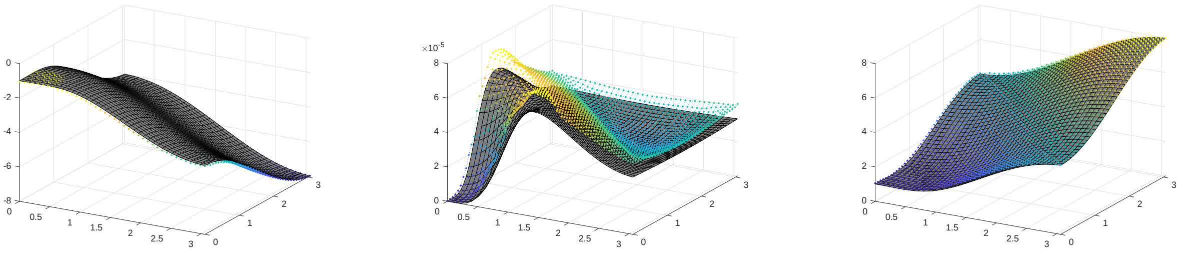
More precisely, we compare the elements of Eval_t with the elements of Bl_t by means of the following matching algorithm:

- save the couples $(\theta_{n-1}^{(j_t)}, \theta_{n-1}^{(k_t)})$ of $\boldsymbol{\theta}_{\mathbf{n}-\mathbf{e}}$ to which the elements of Eval_t are associated with;
- for a fixed $\lambda \in \mathrm{Bl}_t$ find $\tilde{\eta} \in \mathrm{Eval}_t$ such that

$$\tilde{\eta} = \arg \min_{\eta \in \mathrm{Eval}_t} \|\lambda - \eta\|;$$

- associate λ to the couple $(\theta_{n-1}^{(j_t)}, \theta_{n-1}^{(k_t)})$ corresponding to $\tilde{\eta}$.

Making use of the previous algorithm, in Figure 4, we compare the eigenvalues of B_N with $\lambda^{(l)}(\mathbf{f})$, $l = 1, 2, 3$ displayed as a mesh on $\boldsymbol{\theta}_{\mathbf{n}-\mathbf{e}}$, for $n = 40$. The eigenvalues of B_N mimic, up to some outliers shown in the Figure 4b, the sampling of the eigenvalue functions, numerically confirming the result given in Theorem 5.



(a) $\lambda^{(1)}(\mathbf{f})$ (b) $\lambda^{(2)}(\mathbf{f})$ (c) $\lambda^{(3)}(\mathbf{f})$

Figure 4: Comparison between the eigenvalues of B_N and $\lambda^{(l)}(\mathbf{f})$, $l = 1, 2, 3$ displayed as a mesh on $\boldsymbol{\theta}_{\mathbf{n}-\mathbf{e}}$, when $n = 40$.

3.3. From Poisson to advection-diffusion equations. We have built the whole construction using as constraint the Poisson differential equation, this is not restrictive since the analysis can be transparently extended to encompass constraints given by a generic elliptic differential equations, i.e.,

$$-\nabla^2 y + \mathbf{c}\cdot\nabla y + ry = z. \tag{3.31}$$

The matrix sequence (2.7) maintains the same 3×3 block structure, but with a different (1,3) and (3,1) block $\bar{Z}$. The latter, whenever $\mathbf{c}=(c_1,c_2)\neq 0$, is no longer symmetric since the new constraint is no longer self–adjoint. Specifically, the new block $\bar{Z}$ can be decomposed into the sum of three terms,

$$\bar{Z} = \bar{K} + \bar{V} + \gamma\bar{M}, \qquad (\bar{V})_{i,j} = \int_{\tau_h} (\mathbf{c}\cdot\nabla\phi_i)\phi_j d\mathbf{x},$$

with $V \neq V^T$. Therefore, the relative scaled version is given by

$$\mathcal{S}_N = \mathcal{D}_N^{(1)}\bar{\mathcal{S}}_N\mathcal{D}_N^{(2)} = \begin{bmatrix} h^4 M & O & Z^T \\ O & \alpha M & -M \\ Z & -M & O \end{bmatrix}, \qquad Z = K + hV + h^2 M. \tag{3.32}$$

By means of a GLT perturbation argument from Section 3.1, and exploiting the analysis in [17, Section 7.4] for the presence of lower order differential terms, we can obtain again a characterization of the eigenvalues of $\mathcal{S}_N$ in (3.32) that is analogous to the one we gave in Theorem 5.

Proposition 2. *The matrix sequence $\{\mathcal{S}_N\}_N$ from (3.32) is distributed in the eigenvalue sense as the matrix–valued function* $\mathbf{f}$ *from Theorem 5.*

Proof. Follows from Theorem 5, the techniques adopted in its proof, and from **GLT5** applied to $\mathcal{S}_N = \mathcal{A}_N + \mathcal{Y}_N$, where

□

$$\mathcal{Y}_N = \begin{bmatrix} O & O & hV^T + h^2 M \\ O & O & O \\ hV + h^2 M & O & O \end{bmatrix}.$$

4. An optimal preconditioning strategy.

In this section, we analyze an effective procedure to precondition the GMRES method for the solution of the systems (3.14), and (3.32). There exist indeed many preconditioners for the linear systems of saddle–point type exploiting their block structure, see, e.g, the review [5] the comparisons in [2], and, more specifically, the approaches described in [4, 24, 25, 20]. What we present here belongs to this class, and is built with the objective of obtaining algorithmic scalability, i.e., independence of the number of iteration from h, and optimality with respect to the parameter α, i.e., independence of the number of iteration also with respect to it. To achieve this kind of results the classical techniques can be broadly divided into three classes, the case of definite Hermitian preconditioners for which it is possible to retrieve a cluster of the eigenvalue sense from a cluster of the singular values [30, 24, 4], that allows also for the use of the MINRES method; the case of the indefinite Hermitian preconditioners, and non Hermitian preconditioner [25, 20]. We focus here on the last approach, while benefiting both from the spectral distribution of the sequence $\{T_{\mathbf{n}}(m)\}_{\mathbf{n}}$ and $\{T_{\mathbf{n}}(\kappa)\}_{\mathbf{n}}$ of the Sections 3.2, 3.3, and from the block form of the matrices $\mathcal{A}_N$ and $\mathcal{S}_N$. Specifically, we propose the following preconditioner

$$\mathcal{P}_N \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{z}_3 \end{bmatrix} = \begin{bmatrix} O & \alpha K^T & O \\ O & \alpha M & -M \\ K & -M & O \end{bmatrix} \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{z}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{bmatrix}. \tag{4.33}$$

This is clearly an indefinite, and non Hermitian matrix, nevertheless, the linear systems involving it can be easily solved by the following back–substitution procedure:

1. Solve $\alpha K^T \mathbf{z}_2 = \mathbf{r}_1$;
2. Solve $M\mathbf{z}_3 = \alpha M\mathbf{z}_2 - \mathbf{r}_2$;
3. Solve $K\mathbf{z}_1 = \mathbf{r}_3 + M\mathbf{z}_2$.

We stress that this does not require the approximation of any of the possible Schur complements of $\mathcal{A}_N$ ($\mathcal{S}_N$), thus greatly simplifying the construction of the preconditioner. Moreover, we are going to prove now that this choice provides a strong cluster at 1 for the eigenvalues of the preconditioned linear system while obtaining also the independence of α. We obtain this result in two steps by means of the GLT theory showing that the matrix sequence $\{\mathcal{P}_N^{-1}\mathcal{A}_N\}_N$ is distributed in the sense of the eigenvalues as $\mathbf{1}$. First, in Proposition 3, we show that the eigenvalues of the preconditioned matrix ${\mathcal{P}_N}^{-1}\mathcal{A}_N$ are either 1, or the generalized eigenvalues of an auxiliary problem, then, in Lemma 1, we prove that the matrix sequence associated to the latter is indeed distributed in the eigenvalue sense as the function $\mathbf{1}$, thus obtaining that the eigenvalues of the preconditioned system are strictly clustered at 1.

Proposition 3. *Let $\mathcal{A}_N$ ($\mathcal{S}_N$) be the coefficient matrix in (3.14) (respectively in (3.32)), and let $\mathcal{P}_N$ be the associated preconditioner from (4.33). Then, the eigenvalues of the preconditioned matrix $\mathcal{P}_N^{-1}\mathcal{A}_N$ are*

- *$\lambda_j = 1$ for $j = 1, \ldots, 2N(\mathbf{n})$,*
- *λ_j for $j = 2N(\mathbf{n})+1, \ldots, N(3,\mathbf{n})$ given by the solution of the generalized eigenvalue problem*

$$\left(\frac{h^4}{\alpha}M + K^T M^{-1} K\right)\boldsymbol{x}_1 = \lambda K^T M^{-1} K \boldsymbol{x}_1,$$

with $\boldsymbol{x}_1 \neq \boldsymbol{0} \in \mathbb{R}^{N(\mathbf{n})}$.

Proof. For each n, λ is an eigenvalue of the matrix $\mathcal{P}_N^{-1}\mathcal{A}_N$ if $(\lambda, \mathbf{x})$ is an eigenpair of the eigenvalue problem

$$\mathcal{A}_N\mathbf{x} = \lambda\mathcal{P}_N\mathbf{x},$$

with

$$\mathbf{x} = \begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\\ \mathbf{x}_3\end{bmatrix} \neq \mathbf{0} \in \mathbb{R}^{N(3,\mathbf{n})}.$$

That is $(\lambda, \mathbf{x})$ is solution of

$$\begin{bmatrix} h^4M & O & K^T\\ O & \alpha M & -M\\ K & -M & O\end{bmatrix}\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\\ \mathbf{x}_3\end{bmatrix} = \lambda\begin{bmatrix} O & \alpha K^T & O\\ O & \alpha M & -M\\ K & -M & O\end{bmatrix}\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\\ \mathbf{x}_3\end{bmatrix}.$$

It is clear from the second and the third "block" equations that $(1, \mathbf{x})$ is an eigenpair for the latter problem for all the vectors in the $N(2,\mathbf{n})$ subspace of $\mathbb{R}^{N(3,\mathbf{n})}$

$$\left\{\mathbf{x} = \begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\\ \mathbf{x}_3\end{bmatrix} \text{ s.t. } \mathbf{x}_3 = \alpha\mathbf{x}_2 - h^4K^{-T}M\mathbf{x}_1,\ \ \forall\, \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^{N(\mathbf{n})}\right\}.$$

Otherwise, if $\lambda \neq 1$, from the third "block" equation

$$(1-\lambda)K\mathbf{x}_1 = (1-\lambda)M\mathbf{x}_2,$$

follows

$$\mathbf{x}_2 = M^{-1}K\mathbf{x}_1.$$

And thus, by substitution, we easily find

$$\mathbf{x}_3 = \alpha M^{-1}K\mathbf{x}_1,$$

and thus, the remaining eigenpairs are given by the solution of

$$\left(\frac{h^4}{\alpha}M + K^TM^{-1}K\right)\mathbf{x}_1 = \lambda K^TM^{-1}K\mathbf{x}_1.$$

□

Lemma 1. *The matrix sequence*

$$\left\{\left(K^TM^{-1}K\right)^{-1}\left(\frac{h^4}{\alpha}M + K^TM^{-1}K\right)\right\}_{\mathbf{n}},$$

associated to the generalized eigenvalue problem

$$\left(\frac{h^4}{\alpha}M + K^TM^{-1}K\right)\boldsymbol{x}_1 = \lambda K^TM^{-1}K\boldsymbol{x}_1,$$

is distributed in the eigenvalue sense as $\mathbf{1}$ *over* $\mathcal{I}_2^+$.

Proof. The statement is equivalent to

$$\left\{(T_{\mathbf{n}}(\kappa)T_{\mathbf{n}}^{-1}(m)T_{\mathbf{n}}(\kappa))^{-1}\left(\frac{h^4}{\alpha}T_{\mathbf{n}}(m) + T_{\mathbf{n}}(\kappa)T_{\mathbf{n}}^{-1}(m)T_{\mathbf{n}}(\kappa)\right)\right\}_{\mathbf{n}} \sim_\lambda (\mathbf{1}, \mathcal{I}_2^+),$$

since, from (3.15) and (3.16), we have that M and K are the symmetric and positive definite matrices $T_{\mathbf{n}}(m)$ and $T_{\mathbf{n}}(\kappa)$, respectively.

Moreover, the sequence $\left\{\frac{h^4}{\alpha}T_{\mathbf{n}}(m)\right\}_{\mathbf{n}}$ is distribuited in the singular value sense as 0 over $\mathcal{I}_2^+$. Hence, from property **GLT4** plus properties **GLT2**–**GLT3**, we have that the following GLT results hold:

$$\left\{\frac{h^4}{\alpha}T_{\mathbf{n}}(m)\right\}_{\mathbf{n}} \sim_{GLT} \mathbf{0},$$

and

$$\{T_{\mathbf{n}}(m)\}_{\mathbf{n}} \sim_{GLT} m, \qquad \{T_{\mathbf{n}}(\kappa)\}_{\mathbf{n}} \sim_{GLT} \kappa,$$
$$\{T_{\mathbf{n}}^{-1}(m)\}_{\mathbf{n}} \sim_{GLT} \frac{1}{m}, \qquad \{T_{\mathbf{n}}^{-1}(\kappa)\}_{\mathbf{n}} \sim_{GLT} \frac{1}{\kappa}.$$

Exploiting again **GLT 2**–**GLT4**, we obtain that

$$\{T_{\mathbf{n}}(\kappa)T_{\mathbf{n}}^{-1}(m)T_{\mathbf{n}}(\kappa)\}_{\mathbf{n}} \sim_{GLT} \frac{m}{\kappa^2}$$

and

$$\left\{\frac{h^4}{\alpha}T_{\mathbf{n}}(m) + T_{\mathbf{n}}(\kappa)T_{\mathbf{n}}^{-1}(m)T_{\mathbf{n}}(\kappa)\right\}_{\mathbf{n}} \sim_{GLT} \frac{\kappa^2}{m}.$$

Since the matrix $T_{\mathbf{n}}(\kappa)T_{\mathbf{n}}^{-1}(m)T_{\mathbf{n}}(\kappa)$ is positive definite, then Theorem 4 implies

$$\left\{(T_{\mathbf{n}}(\kappa)T_{\mathbf{n}}^{-1}(m)T_{\mathbf{n}}(\kappa))^{-1}\left(\frac{h^4}{\alpha}T_{\mathbf{n}}(m)+T_{\mathbf{n}}(\kappa)T_{\mathbf{n}}^{-1}(m)T_{\mathbf{n}}(\kappa)\right)\right\}_{\mathbf{n}} \sim_{GLT,\sigma,\lambda} (\mathbf{1},\mathcal{I}_2^+)$$

and, hence, the thesis. □

Remark 4. Let us stress that the conclusion in Lemma 1 is again an asymptotic result for $h \to 0$ that is then valid for a fixed value of the parameter α. Furthermore, it permits also an answer to *Q3* characterizing the condition number of the preconditioned matrix sequence. Specifically, if we let X be the matrix of the generalized eigenvectors for the pencil (K, M), i.e., if X is an invertible matrix such that

$$KX = MXD, \text{ with } \begin{array}{l} X^TKX = \operatorname{diag}(d_{\mathbf{1}}^{(K)},\ldots,d_{\mathbf{n}}^{(K)}) \equiv D^{(K)}, \\ X^TMX = \operatorname{diag}(d_{\mathbf{1}}^{(M)},\ldots,d_{\mathbf{n}}^{(M)}) \equiv D^{(M)}, \\ D = \operatorname{diag}\left(\frac{d_{\mathbf{1}}^{(K)}}{d_{\mathbf{1}}^{(M)}},\ldots,\frac{d_{\mathbf{n}}^{(K)}}{d_{\mathbf{n}}^{(M)}}\right), \end{array}$$

then we find

$$\begin{aligned}\left(K^TM^{-1}K\right)^{-1}\left(\frac{h^4}{\alpha}M+K^TM^{-1}K\right) &= \left(X^{-T}D^{(K)}X^{-1}XD^{(M)^{-1}}X^TX^{-T}D^{(K)}X^{-1}\right)^{-1} \\ &\quad \left(\frac{h^4}{\alpha}X^{-T}D^{(M)}X^{-1}+X^{-T}D^{(K)}X^{-1}XD^{(M)^{-1}}X^TX^{-T}D^{(K)}X^{-1}\right) \\ &= XD^{(M)}(D^{(K)})^{-2}\left(\frac{h^4}{\alpha}D^{(M)}+(D^{(K)})^2(D^{(M)})^{-1}\right)X^{-1}.\end{aligned}$$

It is then straightforward to use (3.15) and (3.16) to estimate the maximum eigenvalues of the generalized eigenvalue problem in Proposition 3 as an $O(\alpha^{-1})$. This means that the asymptotic regime described in Lemma 1 is evident whenever h^4 becomes smaller than the fixed value of α of the given problem.

We can now answer to question *Q3* for both the matrix sequences $\{\mathcal{P}_N^{-1}\mathcal{A}_N\}_N$, and $\{\mathcal{P}_N^{-1}S_N\}_N$ of Subsection 3.3, where in the definition of the preconditioner (4.33) Z plays the same role of K.

Theorem 6. *The matrix sequences* $\{\mathcal{P}_N^{-1}\mathcal{A}_N\}_N \sim_\lambda (\mathbf{1},\mathcal{I}_2^+)$, $\{\mathcal{P}_N^{-1}\mathcal{S}_N\}_N \sim_\lambda (\mathbf{1},\mathcal{I}_2^+)$ *independently of* α.

Moreover, an analogous spectral result to Theorem 6 can be given for the sequence $\{\mathcal{P}_{\mathrm{BCT}}^{-1}\mathcal{A}_N\}_N$ (respectively, $\{\mathcal{P}_{\mathrm{BCT}}^{-1}\mathcal{S}_N\}_N$), for

$$\mathcal{P}_{\mathrm{BCT}} = \begin{bmatrix} O & O & K^T \\ O & \alpha M & -M \\ K & -M & O \end{bmatrix}.$$

Theorem 7. *The matrix sequences* $\{\mathcal{P}_{BCT}^{-1}\mathcal{A}_N\}_N \sim_\lambda (\mathbf{1},\mathcal{I}_2^+)$, $\{\mathcal{P}_{BCT}^{-1}\mathcal{S}_N\}_N \sim_\lambda (\mathbf{1},\mathcal{I}_2^+)$ *independently of* α.

Proof. The proof follows the proofs of the Proposition 3 and Lemma 1, replacing the expression of $\mathcal{P}_N$ with that of $\mathcal{P}_{\mathrm{BCT}}$. □

This is indeed an example of a block–counter–triangular preconditioner in the style of [4].

Remark 5. The preconditioner proposed in [4] takes the lower anti–triangular part of a different permutation of the system matrix $\mathcal{A}_N$, and considers also a different scaling. By this approach, the term that is dropped out in the preconditioner is not a correction of "small" norm, and this makes a substantial difference in the performances of the two approaches. Specifically, comparing the results of Proposition 3, with [4, Theorem 3.1], it is straightforward to observe that in the latter case it is not possible to infer a cluster of the eigenvalues of the preconditioned system, specifically, for the rearranged system

$$\tilde{\mathcal{P}}_{\mathrm{BCT}}^{-1}\tilde{\mathcal{A}}_N = \begin{bmatrix} O & O & -M \\ O & h^4M & K^T \\ -M & K & O \end{bmatrix}^{-1} \begin{bmatrix} \alpha M & O & -M \\ O & h^4M & K^T \\ -M & K & O \end{bmatrix}.$$

The non-unit eigenvalues are the one of the matrix sequence $\{I + \alpha h^{-4}M^{-1}KM^{-1}K^T\}_N$, for which the clustering at one cannot be concluded. Similar observation can be made also for the null–space based block anti–triangular preconditioners [24] arising from the block anti–triangular factorization of the saddle–point matrix. Furthermore, one could consider the preconditioner which neglects the (3,2) block of $\bar{\mathcal{A}}_N$, avoiding the reordering and the scaling. This would bring to the case where the non-unit eigenvalues are the solution of the following generalized eigenvalue problem

$$\begin{bmatrix} h^2M & O & K^T \\ O & \alpha h^2M & -h^2M \\ K & -h^2M & O \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix} = \lambda \begin{bmatrix} h^2M & O & K^T \\ O & \alpha h^2M & -h^2M \\ K & O & O \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix},$$

and, then, we have a behavior analogous to the case with preconditioner $\tilde{\mathcal{P}}_{\mathrm{BCT}}$ (i.e., the absence of a provable cluster of the preconditioned sequence). Precisely, the non-unit eigenvalues are of the form $\lambda_i = 1+\mu_i$, where, μ_i are the reciprocal of the eigenvalues of the matrix sequence $\{\frac{\alpha}{h^4}M^{-1}KM^{-1}K^T\}_N$.

4.1. Approximate iterative solution of the auxiliary linear systems. The application of the proposed preconditioners requires the solution of auxiliary linear systems with the matrices K ,K^T, and M or, respectively, Z, Z^T, and M obtained from (4.33). In both cases we are dealing with very common linear systems for which there exist highly efficient and specific solvers, e.g., fast Poisson solvers, multigrid methods of geometric, and algebraic type, inner–outer Krylov solver with incomplete factorization preconditioner, and several combinations of all the previous. Potentially, any optimal preconditioner for these matrices could be included in the present framework without spoiling the overall construction, the actual choice is indeed a matter of computational framework; see, e.g., [8, Chapter 3.8]. For the solution of the systems involving the mass matrix M a straightforward solution is using the unpreconditioned CG method or its preconditioned version. In the latter case, we use either a modified incomplete Cholesky factorization with drop–tolerance `1e-2` or a standard algebraic multigrid. We stress that the solution of the system involving the stiffness matrix can be machine-dependent; see, e.g., Figure 5. We easily observe that the fastest solution with the required accuracy for the system involving the $K = T_{\mathbf{n}}(k)$ is obtained by using the PCG with a standard AMG preconditioner. On the other hand, for the non symmetric case we can use the BiCGstab method together with a modified incomplete LU factorization of Crout type. Nevertheless, as we discuss in the next Section 5, the time–efficiency in the auxiliary solve it is not so crucial, observe that already the direct method gives acceptable results under this aspect. What really matters is the combination of the achieved accuracy of the auxiliary solve with the presence, and the possible accumulation, of the α factor in the right–hand side of the auxiliary linear systems. This will cause for their solution by a direct method to return better performances for the lowest value of α.

5. Numerical examples. In this section, we test the application of the preconditioners analyzed in Section 4 on some test problems. All the numerical tests are made on a laptop running Linux with 8 Gb

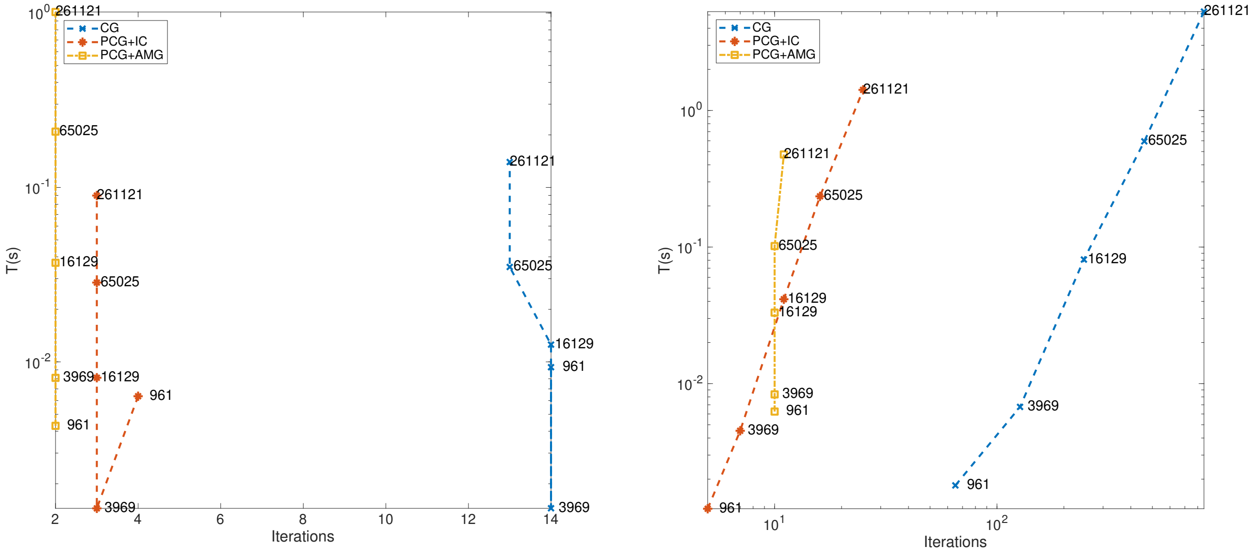


Figure 5: Comparison of solving routines for the auxiliary linear systems, on the left we compare the solution for the system involving the mass matrix M, while on the right the comparison is for the Hermitian stiffness matrix K. The comparisons do not take into account the building time for the various preconditioner since it is then distributed among the repeated solution. The maximum number of allowed PCG iterations is the size of the problem, while the stopping criterion on the relative residual is set to a tolerance of `1e-8`.

memory and CPU Intel® Core$^{\text{TM}}$ i7–4710HQ CPU with clock 2.50 GHz and MATLAB version 9.4.0.813654 (R2018a). We recall again that all the relevant matrices and right–hand sides are generated by means of the FEniCS library (v.2018.1.0) [1, 21]; see again Section 2 for the details.

We test the solution procedure with the un–restarted GMRES method set to achieve a tolerance on the residual of `tol = 1e-6`, and a maximum number of iteration `maxit = 100`, and measure the number of iterations, and the timings in second. As test problem we consider an instance of a Poisson control problem (2.4), and one with the diffusion–advection–reaction constraint from Section 3.3.

Poisson. The first test problem is an instance of the Poisson control problem (2.4), in which we want to obtain the desired state,

$$y_d(x_1, x_2) = -\sin(8\pi x_1)\sin(8\pi x_2) + \sin(\pi x_1)\sin(\pi x_2),$$

while using the forcing term

$$z(x_1, x_2) = 2\pi^2 \sin(\pi x_1) + \frac{1}{128\pi^2}\sin(8\pi x_1)\sin(8\pi x_2).$$

We test the solution for regularization parameter $\alpha = \texttt{1.0e-03}, \texttt{1.0e-06}, \texttt{1.0e-09}$, and collect the results in Table 2. The approximate preconditioners are applied inside the Flexible–GMRES method as discussed in Section 4.1. What we observe is that the approximate solution are at an advantage for the higher value of α, while perform poorly for the smallest $\alpha = \texttt{1.0e-09}$. We stress that this effect is more connected to the behavior of the accuracy in the computation of the Krylov vectors inside the FGMRES method, than to the

		GMRES						FGMRES+PCG+IC			
		I_N		$\mathcal{P}_N$		$\mathcal{P}_{\mathrm{BCT}}$		$\mathcal{P}_N$		$\mathcal{P}_{\mathrm{BCT}}$	
α	N	IT	T(s)	IT	T(s)	IT	T(s)	IT	T(s)	IT	T(s)
1.0e-03	147	†	-	3	3.0e-03	3	**2.5e-03**	3	4.4e-03	3	4.5e-03
	675	†	-	3	6.4e-03	3	**3.7e-03**	3	4.7e-03	3	4.6e-03
	2883	†	-	3	1.0e-02	3	9.9e-03	3	**7.3e-03**	3	**7.3e-03**
	11907	†	-	2	3.1e-02	2	3.0e-02	2	**2.3e-02**	2	**2.3e-02**
	48387	†	-	2	2.1e-01	2	1.7e-01	2	**1.5e-01**	2	**1.5e-01**
	195075	†	-	2	9.4e-01	2	9.0e-01	2	**7.3e-01**	2	7.4e-01
	783363	†	-	1	2.1e+00	1	**2.0e+00**	1	2.9e+00	1	2.9e+00
1.0e-06	147	†	-	15	4.3e-03	15	4.1e-03	15	**1.5e-03**	15	**1.5e-03**
	675	†	-	14	1.3e-02	14	**1.2e-02**	14	1.7e-02	14	1.7e-02
	2883	†	-	9	3.0e-02	9	3.0e-02	10	**2.4e-02**	10	**2.4e-02**
	11907	†	-	6	1.0e-01	6	9.3e-02	6	**7.0e-02**	6	7.4e-02
	48387	†	-	4	3.1e-01	4	3.0e-01	4	2.5e-01	4	**2.1e-01**
	195075	86	3.8e+00	2	8.7e-01	2	8.4e-01	2	7.8e-01	2	**7.6e-01**
	783363	80	3.0e+01	2	**4.3e+00**	2	**4.3e+00**	2	4.5e+00	2	4.6e+00
1.0e-09	147	†	-	27	9.6e-03	27	8.7e-03	27	**3.2e-03**	27	3.4e-03
	675	†	-	54	6.3e-02	54	**5.8e-02**	54	7.9e-02	54	7.9e-02
	2883	†	-	52	2.0e-01	52	2.1e-01	52	**1.6e-01**	52	**1.6e-01**
	11907	†	-	33	5.8e-01	33	6.1e-01	33	**5.2e-01**	33	5.5e-01
	48387	†	-	20	**1.5e+00**	20	**1.5e+00**	43	4.8e+00	42	4.8e+00
	195075	86	**2.8e+00**	33	1.3e+01	33	1.3e+01	37	3.0e+01	36	2.9e+01
	783363	80	3.0e+01	33	**2.5e+01**	33	**2.5e+01**	†	-	†	-

Table 2: Poisson Control Problem. We compare both the number of iterations, and the solution time for the various preconditioners. Best timings are highlighted in bold face. When the method fails to converge, i.e., the method reaches the maximum number of iterations, a † is reported. The inner tolerance for the PCG is set to 1e-8.

optimal behavior of the auxiliary problems. Secondarily, what we observe is indeed the optimal behavior with respect to the iteration discussed in Theorem 7. Indeed, the preconditioning routine becomes asymptotically better with the size of the problem, i.e., we get fewer iteration for bigger problems. Moreover, the decreasing of the α introduces just a latency effect in the solution, i.e., the asymptotic regimes kicks in for slightly bigger problems when α is smaller, we stress that this is exactly the phenomenon described in Remark 4 regarding the asymptotic relation between the value of h going to zero, and the value of α being fixed independently of h. To overcome this limitation, one could decouple the system by neglecting the matrix $\alpha\bar{M}^{-1}$, i.e., the

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 773-798, December 2020.
http://repository.uwyo.edu/ela

(2, 2) block in (2.7), thus obtaining the preconditioner

$$\mathcal{P}_D = \left[\begin{array}{cc|c} \bar{M} & O & \bar{K}^T \\ O & O & -\bar{M} \\ \hline \bar{K} & -\bar{M} & O \end{array}\right]. \tag{5.34}$$

By computation analogous to the one in Remark 5, we find that the non-unit eigenvalues for this preconditioner are the ones of the matrix sequence $\left\{I + \frac{\alpha}{h^4} M^{-1} K M^{-1} K^T\right\}_N$. The non-unit eigenvalues tend to cluster at one whenever $\alpha h^{-4} \propto \alpha N^4$ goes to zero. This means that $\mathcal{P}_D$ is efficient for small values of α and moderate values of N and worsen for diverging values of N (keeping fixed α), indeed this is confirmed by the numerical test in Table 3.

		GMRES preconditioned by $\mathcal{P}_D$						
	N	147	675	2883	11907	48387	195075	783363
α = 1.0e-09	IT	4	5	6	†	†	†	†
	T(s)	1.0e-02	5.4e-03	1.7e-02	-	-	-	-

Table 3: Poisson Control Problem. We report both the number of iterations, and the solution time for the $\mathcal{P}_D$ preconditioner in (5.34), compare these entries with the last block of rows of Table 2.

Diffusion–Convection–Reaction. The second case we consider is the problem (1.2) in which the constraint $e(y, u)$ is given by the Equation (3.31), with coefficients $r = 1$, and $\mathbf{c} = (2, 3)$. The desired state is given by the sum of the two impulses

$$y_d(x_1, x_2) = \frac{0.5}{0.07\sqrt{2\pi}} e^{-\frac{(x_1-0.2)^2+(x_2-0.2)^2}{2(0.07)^2}} + \frac{0.8}{0.05\sqrt{2\pi}} e^{-\frac{(x_1-0.6)^2+(x_2-0.6)^2}{2(0.05)^2}},$$

while the forcing term is given by

$$z(x_1, x_2) = \sin(\pi x_1)\sin(\pi x_2).$$

We test the solution for regularization parameter $\alpha = \texttt{1.0e-03}, \texttt{1.0e-06}, \texttt{1.0e-09}$, and collect the results in Table 4. The results are completely analogous to the one for the Poisson case. We observe a higher number of iteration that is due to the fact that we are using an asymptotic argument both for the sequence $\mathcal{S}_N$, and for its block; see Proposition 2, and the discussion in Remark 4 for the asymptotic relationship between h, and α.

6. Conclusions and future developments. In this paper, we have produced a characterization for the saddle–point matrices arising from the application of the discretize–then–optimize approach to quadratic optimization problems with elliptic PDE constraints highlighting the presence of an hidden Generalized Locally Toeplitz structure, i.e., we have proposed an analysis that is sharper and more informative than the one that can be obtained by looking only at the saddle–point structure. We have produced a localization of

		GMRES						FGMRES PCG/BiCGstab+IC/ILU			
		I_N		$\mathcal{P}_N$		$\mathcal{P}_{\text{BCT}}$		$\mathcal{P}_N$		$\mathcal{P}_{\text{BCT}}$	
α	N	IT	T(s)	IT	T(s)	IT	T(s)	IT	T(s)	IT	T(s)
1.0e-03	147	†	-	5	8.7e-01	5	**4.7e-03**	5	7.7e-03	5	7.0e-03
	675	†	-	5	9.6e-03	5	8.7e-03	5	7.1e-03	5	**6.4e-03**
	2883	†	-	4	7.2e-02	4	2.7e-02	4	1.4e-01	4	**9.5e-03**
	11907	†	-	3	8.6e-01	3	8.9e-02	3	4.8e-02	3	**3.4e-02**
	48387	†	-	3	1.1e+00	3	4.4e-01	3	3.9e-01	3	**2.5e-01**
	195075	†	-	2	1.7e+00	2	1.7e+00	2	1.7e+00	2	**1.1e+00**
	783363	†	-	2	8.5e+00	2	8.9e+00	2	**7.1e+00**	2	7.4e+00
1.0e-06	147	†	-	24	1.5e-02	24	**1.4e-02**	24	2.9e-02	24	2.4e-02
	675	†	-	26	4.7e-02	26	4.6e-02	26	3.6e-02	27	**3.3e-02**
	2883	†	-	24	1.7e-01	24	1.6e-01	24	7.2e-02	25	**6.0e-02**
	11907	†	-	22	6.9e-01	22	7.0e-01	22	4.0e-01	24	**2.9e-01**
	48387	†	-	19	2.8e+00	19	2.8e+00	19	2.5e+00	22	**1.9e+00**
	195075	†	-	17	1.4e+01	17	1.4e+01	17	1.4e+01	18	**1.2e+01**
	783363	†	-	14	5.9e+01	14	6.1e+01	14	7.9e+01	14	**5.9e+01**
1.0e-09	147	†	-	38	3.8e-02	38	**3.5e-02**	38	4.2e-02	38	4.4e-02
	675	†	-	73	1.5e-01	73	1.6e-01	73	**1.2e-01**	87	1.4e-01
	2883	†	-	84	6.5e-01	73	6.5e-01	86	**3.3e-01**	73	3.7e-01
	11907	†	-	94	3.5e+00	94	3.4e+00	97	2.1e+00	97	**1.9e+00**
	48387	†	-	87	1.4e+01	87	1.4e+01	87	1.2e+01	87	**1.1e+01**
	195075	†	-	77	6.8e+01	77	6.8e+01	†	-	†	-
	783363	†	-	66	5.2e+02	66	5.4e+02	†	-	†	-

Table 4: Diffusion–Convection–Reaction Control Problem. We compare both the number of iterations, and the solution time for the various preconditioners. Best timings are highlighted in bold face. When the method fails to converge, i.e., the method reaches the maximum number of iterations, a † is reported. The tolerances for the inner solvers are set to `1e-8`.

the spectrum in three intervals, up to a number of outliers infinitesimal in the dimension of the problem, and used this characterization to produce an asymptotically optimal preconditioner, i.e., a preconditioner that is independent of the value of the regularization parameter α, and whose performance increases for finer grids.

We plan to extend this analysis in order that it can cover more general constraints, i.e., we would like to discuss also the case of sparse optimization, and bounded controls. Moreover, the GLT spectral analysis techniques we are using have been recently extended for becoming tools for the fast and reliable computation of generalized eigenvalues see, e.g., [13, 14], since we have analyzed the structure of the eigenvectors of our preconditioned problems (Proposition 3), we plan to investigate the possible application of deflation techniques to further accelerate our iterative methods.

Acknowledgment. We are thankful to Prof. S. Serra–Capizzano for the insightful discussions on the spectral distribution results, and to the referee whose suggestions were extremely helpful in improving the presentation of the material.

REFERENCES

[1] M.S. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M.E. Rognes, and G.N. Wells. *The FEniCS Project Version 1.5*, Archive of Numerical Software, 3, 2015. Available at https://doi.org/10.11588/ans.2015.100.20553, http://nbn-resolving.de/urn:nbn:de:bsz:16-ans-205530.

[2] O. Axelsson, S. Farouq, and M. Neytcheva; Comparison of preconditioned Krylov subspace iteration methods for PDE-constrained optimization problems: Poisson and convection-diffusion control. *Numer. Algorithms*, 73:631–663, 2016. Available at https://doi.org/10.1007/s11075-016-0111-1.

[3] O. Axelsson and M. Neytcheva. Eigenvalue estimates for preconditioned saddle point matrices. *Numer. Linear Algebra Appl.*, 13:339–360, 2006. Available at https://doi.org/10.1002/nla.469.

[4] Z.-Z. Bai. Block preconditioners for elliptic PDE-constrained optimization problems. *Computing*, 91:379–395, 2011. Available at https://doi.org/10.1007/s00607-010-0125-9.

[5] M. Benzi, G.H. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Numer.*, 14:1–137, 2005. Available at https://doi.org/10.1017/S0962492904000212.

[6] M. Benzi and V. Simoncini. On the eigenvalues of a class of saddle point matrices. *Numer. Math.*, 103:173–196, 2006. Available at https://doi.org/10.1007/s00211-006-0679-9.

[7] L. Bergamaschi. On eigenvalue distribution of constraint-preconditioned symmetric saddle point matrices. *Numer. Linear Algebra Appl.*, 19:754–772, 2012. Available at https://doi.org/10.1002/nla.806.

[8] D. Bertaccini and F. Durastante. *Iterative Methods and Preconditioning for Large and Sparse Linear Systems with Applications.* Monographs and Research Notes in Mathematics, CRC Press, Boca Raton, FL, 2018.

[9] D. Braess. *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*, third edition (translated from the German by Larry L. Schumaker). Cambridge University Press, Cambridge, 2007. Available at https://doi.org/10.1017/CBO9780511618635.

[10] S. Cipolla and F. Durastante. Fractional PDE constrained optimization: An optimize-then-discretize approach with L-BFGS and approximate inverse preconditioning. *Appl. Numer. Math.*, 123:43–57, 2018. Available at https://doi.org/10.1016/j.apnum.2017.09.001.

[11] 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. *Comput. Math. Appl.*, 74:1015–1028, 2017. Available at https://doi.org/10.1016/j.camwa.2017.05.024.

[12] F. Durastante and S. Cipolla. Fractional PDE constrained optimization: box and sparse constrained problems. In: *Numerical Methods for Optimal Control Problems*, Springer International Publishing, Cham, Chapter 6, 111–135, 2018. Available at https://doi.org/10.1007/978-3-030-01959-4_6.

[13] S.-E. Ekström, I. Furci, and S. Serra-Capizzano. Exact formulae and matrix-less eigensolvers for block banded symmetric Toeplitz matrices. *BIT*, 58:937–968, 2018. Available at https://doi.org/10.1007/s10543-018-0715-z.

[14] S.-E. Ekström, C. Garoni, and S. Serra-Capizzano. Are the eigenvalues of banded symmetric Toeplitz matrices known in almost closed form? *Exp. Math.*, 27:478–487, 2018. Available at https://doi.org/10.1080/10586458.2017.1320241.

[15] C. Garoni, M. Mazza, and S. Serra-Capizzano. Block generalized locally toeplitz sequences: From the theory to the applications. *Axioms*, 7, 2018. Available at https://doi.org/10.3390/axioms7030049.

[16] C. Garoni and S. Serra-Capizzano. *Generalized Locally Toeplitz Sequences: Theory and Applications*, Vol. I. Springer, Cham, 2017. Available at https://doi.org/10.1007/978-3-319-53679-8.

[17] C. Garoni and S. Serra-Capizzano. *Generalized Locally Toeplitz Sequences: Theory and Applications*, Vol. II. Springer, Cham, 2018. Available at https://doi.org/10.1007/978-3-030-02233-4.

[18] N.I.M. Gould and V. Simoncini. Spectral analysis of saddle point matrices with indefinite leading blocks. *SIAM J. Matrix Anal. Appl.*, 31:1152–1171, 2009. Available at https://doi.org/10.1137/080733413.

[19] U. Grenander and G. Szegő. *Toeplitz Forms and Their Applications*, second edition. Chelsea Publishing Co., New York, 1984.

[20] Y.-F. Ke and C.-F. Ma. Some preconditioners for elliptic PDE-constrained optimization problems. *Comput. Math. Appl.*, 75:2795–2813, 2018. Available at https://doi.org/10.1016/j.camwa.2018.01.009.

[21] A. Logg, B.K. Ølgaard, M.E. Rognes, and G.N. Wells. FFC: The FEniCS Form Compiler. In: A. Logg, K.-A. Mardal, and G.N. Wells (editors), *Automated Solution of Differential Equations by the Finite Element Method*, Lecture Notes in Computational Science and Engineering, Vol. 84, Chapter 11, Springer, 2012.

[22] M.F. Murphy, G.H. Golub, and A.J. Wathen. A note on preconditioning for indefinite linear systems. *SIAM J. Sci. Comput.*, 21:1969–1972, 2000. Available at https://doi.org/10.1137/S1064827599355153, https://doi.org/10.1137/S1064827599355153.

[23] I. Perugia and V. Simoncini. Block-diagonal and indefinite symmetric preconditioners for mixed finite element formulations. *Numer. Linear Algebra Appl.*, (Preconditioning Techniques for Large Sparse Matrix Problems in Industrial Applications, Minneapolis, MN, 1999) 7:585–616, 2000. Available at https://doi.org/10.1002/1099-1506(200010/12)7:7/8⟨585::AID-NLA214⟩3.3.CO;2-6.

[24] J. Pestana and A.J. Wathen. The antitriangular factorization of saddle point matrices. *SIAM J. Matrix Anal. Appl.*, 35:339–353, 2014. Available at https://doi.org/10.1137/130934933.

[25] T. Rees and M. Stoll. Block-triangular preconditioners for PDE-constrained optimization. *Numer. Linear Algebra Appl.*, 17:977–996, 2010. Available at https://doi.org/10.1002/nla.693.

[26] T. Rusten and R. Winther. A preconditioned iterative method for saddlepoint problems. *SIAM J. Matrix Anal. Appl.*, (Iterative Methods in Numerical Linear Algebra, Copper Mountain, CO, 1990) 13:887–904, 1992.

[27] S. Serra-Capizzano. Asymptotic results on the spectra of block Toeplitz preconditioned matrices. *SIAM J. Matrix Anal. Appl.*, 20:31–44, 1999. Available at https://doi.org/10.1137/S0895479896310160.

[28] S. Serra-Capizzano. Generalized locally Toeplitz sequences: Spectral analysis and applications to discretized partial differential equations. *Linear Algebra Appl.*, (Special issue on Structured Matrices: Analysis, Algorithms and Applications, Cortona, 2000) 366:371–402, 2003. Available at https://doi.org/10.1016/S0024-3795(02)00504-9.

[29] S. Serra-Capizzano. The GLT class as a generalized Fourier analysis and applications. *Linear Algebra Appl.*, 419:180–233, 2006. Available at https://doi.org/10.1016/j.laa.2006.04.012.

[30] S. Serra-Capizzano, D. Bertaccini, and G.H. Golub. How to deduce a proper eigenvalue cluster from a proper singular value cluster in the nonnormal case. *SIAM J. Matrix Anal. Appl.*, 27:82–86, 2005. Available at https://doi.org/10.1137/040608027.

[31] D. Sesana and V. Simoncini. Spectral analysis of inexact constraint preconditioning for symmetric saddle point matrices. *Linear Algebra Appl.*, 438:2683–2700, 2013. Available at https://doi.org/10.1016/j.laa.2012.11.022.

[32] P. Tilli A note on the spectral distribution of Toeplitz matrices. *Linear Multilinear Algebra*, 45:147–159, 1998. Available at https://doi.org/10.1080/03081089808818584.

[33] F. Tröltzsch. *Optimal Control of Partial Differential Equations: Theory, Methods and Applications* (translated from the 2005 German original by Jürgen Sprekels). Graduate Studies in Mathematics, Vol. 112, American Mathematical Society, Providence, RI, 2010. Available at https://doi.org/10.1090/gsm/112.