MATHEMATICS OF COMPUTATION
Volume 86, Number 305, May 2017, Pages 1343–1373
http://dx.doi.org/10.1090/mcom/3143
Article electronically published on August 3, 2016

SPECTRAL ANALYSIS AND SPECTRAL SYMBOL OF MATRICES IN ISOGEOMETRIC GALERKIN METHODS

CARLO GARONI, CARLA MANNI, STEFANO SERRA-CAPIZZANO, DEBORA SESANA, AND HENDRIK SPELEERS

Abstract. A linear full elliptic second-order Partial Differential Equation (PDE), defined on a d-dimensional domain Ω, is approximated by the isogeometric Galerkin method based on uniform tensor-product B-splines of degrees $(p_1, \dots, p_d)$. The considered approximation process leads to a d-level stiffness matrix, banded in a multilevel sense. This matrix is close to a d-level Toeplitz structure if the PDE coefficients are constant and the physical domain Ω is the hypercube $(0,1)^d$ without using any geometry map. In such a simplified case, a detailed spectral analysis of the stiffness matrices has already been carried out in a previous work. In this paper, we complete the picture by considering non-constant PDE coefficients and an arbitrary domain Ω, parameterized with a non-trivial geometry map. We compute and study the spectral symbol of the related stiffness matrices. This symbol describes the asymptotic eigenvalue distribution when the fineness parameters tend to zero (so that the matrix-size tends to infinity). The mathematical tool used for computing the symbol is the theory of Generalized Locally Toeplitz (GLT) sequences.

1. Introduction

Isogeometric Analysis (IgA) is a paradigm for the analysis of problems governed by Partial Differential Equations (PDEs); see [5]. Its goal is to improve the connection between numerical simulation and Computer Aided Design (CAD) systems. In its original formulation, the main idea in IgA is to use directly the geometry provided by CAD systems and to approximate the unknown solutions of differential equations by the same type of functions. Tensor-product B-splines and their rational extension, the so-called NURBS, are the dominant technology in CAD systems used in engineering, and thus also in IgA.

In this paper we consider the following linear full elliptic second-order PDE, with non-constant coefficients and homogeneous Dirichlet boundary conditions:

$$\begin{cases} -\nabla \cdot K\nabla u + \boldsymbol{\alpha} \cdot \nabla u + \gamma u = \mathrm{f}, & \text{in } \Omega, \\ u = 0, & \text{on } \partial\Omega, \end{cases} \tag{1.1}$$

where Ω is a bounded open domain in $\mathbb{R}^d$ with Lipschitz boundary, $K : \Omega \to \mathbb{R}^{d\times d}$ is a Symmetric Positive Definite (SPD) matrix of functions in $C(\overline{\Omega})$, $\boldsymbol{\alpha} : \Omega \to \mathbb{R}^d$ is

Received by the editor February 9, 2015 and, in revised form, October 10, 2015.
2010 *Mathematics Subject Classification.* Primary 15A18, 15B05, 41A15, 15A69, 65N30.
Key words and phrases. Spectral distribution, symbol, Galerkin method, B-splines, isogeometric analysis.
This work was partially supported by INdAM-GNCS Gruppo Nazionale per il Calcolo Scientifico, by the MIUR 'Futuro in Ricerca 2013' Programme through the project DREAMS, and by Donation KAW 2013.0341 from the Knut & Alice Wallenberg Foundation in collaboration with the Royal Swedish Academy of Sciences, supporting Swedish research in mathematics.

a vector of functions in $L^\infty(\Omega)$, $\gamma \in L^\infty(\Omega)$, $\gamma \geq 0$, and $\mathrm{f} \in L^2(\Omega)$. We focus on the isogeometric Galerkin approximation of (1.1) using a geometry map $\mathbf{G} : [0,1]^d \to \overline{\Omega}$ and uniform tensor-product B-splines of arbitrary degrees $\boldsymbol{p} := (p_1, \ldots, p_d)$.

This paper is devoted to the study of the asymptotic spectral (and singular value) distribution of the resulting Galerkin B-spline IgA stiffness matrices when the fineness parameters tend to zero so that the related matrix-size tends to infinity. Our results extend those obtained in [11] and generalized in [10, Chapter 4], which address the simplified case where K is the identity matrix, $\Omega = (0,1)^d$, and $\mathbf{G}$ is the identity map. In more detail, in this paper we prove the following:

a) a spectral distribution exists and is compactly described by a symbol f;
b) the symbol f has a canonical structure incorporating:
   b1) the approximation technique, identified by a finite set of polynomials in the Fourier variables $\boldsymbol{\theta} := (\theta_1, \ldots, \theta_d) \in [-\pi, \pi]^d$;
   b2) the geometry, identified by the map $\mathbf{G}$ in the parametric variables $\hat{\mathbf{x}} := (\hat{x}_1, \ldots, \hat{x}_d)$ defined on the reference domain $[0,1]^d$;
   b3) the coefficients of the higher-order differential operator of the PDE, namely K, in the physical variables $\mathbf{x} := (x_1, \ldots, x_d)$ defined on the physical domain Ω;
c) the symbol f is the same as in the isogeometric collocation setting [9], up to a determinant factor $|\det(J_{\mathbf{G}})|$, being $J_{\mathbf{G}}$ the Jacobian matrix of $\mathbf{G}$;
d) when K is the identity matrix, $\Omega = (0,1)^d$ and $\mathbf{G}$ is the identity map, the symbol f reduces, as expected, to the one in [10, Section 4.5].

The picture described in item b) is intrinsic to the approximation of PDEs by any local method, such as Finite Differences (FDs) and Finite Elements (FEs). Actually, the formal structure of the symbol is essentially the same when considering different techniques to approximate the same problem; see [2, 23, 24] and references therein, with particular attention to [23, Section 2] and [24, Question 3.1]. Even the appearance of the determinant factor mentioned in item c) is expected, because it was already observed when passing from FDs (whose philosophy is analogous in the collocation framework) to FEs in the Galerkin context. Note that the determinant factor helps in keeping the IgA Galerkin matrices less ill-conditioned than the IgA collocation matrices, also when the map $\mathbf{G}$ is (nearly) singular.

Although the formal structure of the symbol is shared by different approximation techniques, some of its analytic features are not so common. For instance, if one of the components of $\boldsymbol{p}$, say p_i, is large, then the symbol f has 'numerical zeros' at the points $(\hat{\mathbf{x}}, \boldsymbol{\theta}) \in [0,1]^d \times [-\pi, \pi]^d$ where $\theta_i = \pi$. More precisely, if $\theta_i = \pi$, the value $f(\hat{\mathbf{x}}, \boldsymbol{\theta})$ tends to 0 exponentially as $p_i \to \infty$. The latter information implies that small eigenvalues are related to high-frequency eigenvectors, and this non-canonical source of ill-conditioning is responsible for the slowdown of all the standard multigrid and preconditioning techniques when one of the p_i grows. Nevertheless, quite recently, we have found a way to exploit the spectral information provided by the symbol f for designing algorithms with convergence speed independent of the fineness parameters and also substantially independent of the approximation parameters $\boldsymbol{p}$; see [6–8].

We want to emphasize that, besides the identification of the symbol f for the Galerkin B-spline IgA stiffness matrices, another important aspect of this paper is the mathematical technique used in our derivation. As explained in Section 4, this technique is quite general and can also be applied in contexts other than Galerkin

B-spline IgA. It consists of the following mathematical tools. We use the theory of Generalized Locally Toeplitz (GLT) sequences, which goes back to the pioneering work by Tilli [26] and is developed in [12, 23, 24]. Moreover, we implicitly use the concept of approximating class of sequences (a.c.s.), which was introduced in [20] and allows one to derive the singular value and eigenvalue distribution of a complicated sequence of matrices (matrix-sequence) starting from those of simpler matrix-sequences; see [15, 20]. Finally, we exploit general results, derived in [16] and generalized in [13], which allow one to determine the spectral distribution of arbitrary (non-Hermitian) perturbed versions of sequences formed by Hermitian matrices under certain conditions on the perturbation matrices.

An alternative way to obtain the main result of this paper (Theorem 4.1) could be a comparison technique between the Galerkin B-spline IgA stiffness matrices considered herein and the B-spline IgA collocation matrices analyzed in [9] and [10, Chapter 5]. We prefer, however, the approach discussed in the previous paragraph due to its intrinsic generality. On this concern, we remark that the continuity assumption on the components of K is not a limitation of our technique, but is merely intended to avoid too many technicalities in the proof of Theorem 4.1. Actually, continuity can be relaxed as explained in Remark 4.3 and Section 6.

The paper is organized as follows. In Section 2 we introduce the notation and definitions used throughout the paper; we also report some basic results. In Section 3 we describe the isogeometric Galerkin approximation of the full elliptic problem (1.1) based on uniform tensor-product B-splines. Sections 4 and 5 contain our main results: the computation of the spectral and singular value distribution of the Galerkin B-spline IgA stiffness matrices, the identification of the corresponding symbol f, and the study of its properties. Section 6 is devoted to a critical discussion regarding the generality and applicability of our analysis, corroborated by a few numerical illustrations. We end in Section 7 with some concluding remarks.

2. Preliminaries

2.1. **Multi-index notation and matrix-sequences.** Throughout this paper, we use the multi-index notation. When discretizing a linear PDE over a d-dimensional domain by means of a numerical method, the resulting discretization matrices often have a d-level structure (see [28, Section 6] for the corresponding definition). The multi-index notation is a powerful tool that allows one to give a compact expression of these matrices, treating the dimensionality parameter d as any other parameter involved in the discretization process. In this way, the dependence of the matrix structure on d is highlighted and a compact presentation is possible.

A multi-index $\boldsymbol{m} \in \mathbb{Z}^d$, also called a d-index, is simply a (row) vector in $\mathbb{Z}^d$; its components are denoted by $m_1, \ldots, m_d$. We indicate by $\mathbf{0}, \mathbf{1}, \mathbf{2}$ the vectors consisting of all zeros, all ones, all twos, respectively (their size will be clear from the context). For any d-index $\boldsymbol{m}$, we set $N(\boldsymbol{m}) := \prod_{i=1}^{d} m_i$ and we write $\boldsymbol{m} \to \infty$ to indicate that $\min_{i=1,\ldots,d} m_i \to \infty$. Inequalities between multi-indices must be interpreted in the componentwise sense. For instance, $\boldsymbol{j} \le \boldsymbol{k}$ means that $j_i \le k_i$ for every i. If $\boldsymbol{j}, \boldsymbol{k}$ are d-indices such that $\boldsymbol{j} \le \boldsymbol{k}$, the multi-index range $\boldsymbol{j}, \ldots, \boldsymbol{k}$ is the set $\{\boldsymbol{i} \in \mathbb{Z}^d : \boldsymbol{j} \le \boldsymbol{i} \le \boldsymbol{k}\}$. We assume for this set the standard lexicographic ordering:

$$\left[\ldots \left[\left[(i_1, \ldots, i_d) \right]_{i_d = j_d, \ldots, k_d} \right]_{i_{d-1} = j_{d-1}, \ldots, k_{d-1}} \ldots \right]_{i_1 = j_1, \ldots, k_1}. \tag{2.1}$$

For instance, in the case $d = 2$, this ordering is

$$(j_1, j_2),\ (j_1, j_2+1),\ \ldots,\ (j_1, k_2),\ (j_1+1, j_2),\ (j_1+1, j_2+1),\ \ldots,\ (j_1+1, k_2),$$
$$\ldots, \ldots, \ldots,\ (k_1, j_2),\ (k_1, j_2+1),\ \ldots,\ (k_1, k_2).$$

When a d-index $\boldsymbol{i}$ varies in a multi-index range $\boldsymbol{j}, \ldots, \boldsymbol{k}$ (this is often written as $\boldsymbol{i} = \boldsymbol{j}, \ldots, \boldsymbol{k}$), it is always assumed that $\boldsymbol{i}$ varies from $\boldsymbol{j}$ to $\boldsymbol{k}$ following the specific ordering (2.1). In particular, if $\boldsymbol{m} \in \mathbb{N}^d$ and $\mathbf{x} := [x_{\boldsymbol{i}}]_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{m}}$, then $\mathbf{x}$ is a vector of length $N(\boldsymbol{m})$ whose components $x_{\boldsymbol{i}}$, $\boldsymbol{i} = \mathbf{1}, \ldots, \boldsymbol{m}$, are ordered in accordance with (2.1): the first component is $x_{\mathbf{1}} = x_{(1,\ldots,1,1)}$, the second is $x_{(1,\ldots,1,2)}$, and so on until the last, which is $x_{\boldsymbol{m}} = x_{(m_1,\ldots,m_d)}$. Similarly, if $X := [x_{\boldsymbol{ij}}]_{\boldsymbol{i},\boldsymbol{j}=\mathbf{1}}^{\boldsymbol{m}}$, then X is an $N(\boldsymbol{m}) \times N(\boldsymbol{m})$ matrix whose entries are indexed by two d-indices $\boldsymbol{i}, \boldsymbol{j}$, both varying in $\mathbf{1}, \ldots, \boldsymbol{m}$ according to (2.1). Operations involving d-indices that have no meaning in the vector space $\mathbb{Z}^d$ must always be interpreted in the componentwise sense. For instance, $\boldsymbol{jk} := (j_1k_1, \ldots, j_dk_d)$, $\boldsymbol{j}/\boldsymbol{k} := (j_1/k_1, \ldots, j_d/k_d)$, etc.

Throughout this paper, by sequence of matrices (or matrix-sequence) we mean a sequence of the form $\{X_{\boldsymbol{m}}\}_n$, where

- n varies in some infinite subset of $\mathbb{N}$;
- $\boldsymbol{m} = \boldsymbol{m}(n)$ is a d-index in $\mathbb{N}^d$ which depends on n, and $\boldsymbol{m} \to \infty$ as $n \to \infty$;
- $X_{\boldsymbol{m}}$ is a square matrix of size $N(\boldsymbol{m})$.

The multi-index $\boldsymbol{m}$ that parameterizes a matrix-sequence is always assumed to be a d-index.

2.2. **Preliminaries on matrix analysis.** For all $X \in \mathbb{C}^{m\times m}$, the singular values of X are denoted by $\sigma_j(X)$, $j = 1, \ldots, m$, and the eigenvalues of X by $\lambda_j(X)$, $j = 1, \ldots, m$. If $X, Y \in \mathbb{C}^{m\times m}$, the notation $X \geq Y$ means that X, Y are Hermitian and $X - Y$ is positive semi-definite. The ∞-norm and the spectral norm (2-norm) of both vectors and matrices are denoted by $\|\cdot\|_\infty$ and $\|\cdot\|$, respectively. We recall that

$$\|X\| \leq \sqrt{\|X\|_\infty \|X^T\|_\infty}, \quad \forall X \in \mathbb{C}^{m\times m}. \tag{2.2}$$

For $X \in \mathbb{C}^{m\times m}$, we denote by $\|X\|_1$ the trace-norm (or Schatten 1-norm) of X, i.e., the sum of all the singular values of X. Since $\operatorname{rank}(X)$ is the number of non-zero singular values of X and $\|X\|$ equals the largest singular value of X, we have

$$\|X\|_1 \leq \operatorname{rank}(X)\|X\| \leq m\|X\|, \quad \forall X \in \mathbb{C}^{m\times m}. \tag{2.3}$$

Let X, Y be matrices of any dimension, say $X \in \mathbb{C}^{k\times m}$ and $Y \in \mathbb{C}^{r\times s}$. Then, the tensor (Kronecker) product $X \otimes Y$ is defined by

$$X \otimes Y := [x_{ij}Y]_{\substack{i=1,\ldots,k \\ j=1,\ldots,m}} = \begin{bmatrix} x_{11}Y & \cdots & x_{1m}Y \\ \vdots & & \vdots \\ x_{k1}Y & \cdots & x_{km}Y \end{bmatrix} \in \mathbb{C}^{kr\times ms}.$$

Tensor products possess a lot of nice algebraic properties. One of them is the associativity, which allows one to omit parentheses in expressions like $X_1 \otimes X_2 \otimes \cdots \otimes X_d$. Another property is the following [10, p. 15]: let $X_1, \ldots, X_d, Y_1, \ldots, Y_d$ be matrices such that $X_i, Y_i \in \mathbb{C}^{m_i\times m_i}$ for all $i = 1, \ldots, d$; then

$$\operatorname{rank}(X_1 \otimes \cdots \otimes X_d - Y_1 \otimes \cdots \otimes Y_d) \leq N(\boldsymbol{m}) \sum_{i=1}^{d} \frac{\operatorname{rank}(X_i - Y_i)}{m_i}, \tag{2.4}$$

where $\boldsymbol{m} := (m_1, \ldots, m_d)$.

Given two matrices X, Y of the same dimension, the componentwise (Hadamard) product of X and Y is denoted by $X \circ Y$. If $X, Y, Z \in \mathbb{C}^{m\times m}$ are such that Z is Hermitian positive semi-definite and $X \geq Y$, then

$$X \circ Z \geq Y \circ Z; \tag{2.5}$$

see, e.g., the proof of [9, Theorem 5.2] or [10, Lemma 1.6].

2.3. **Spectral distribution and spectral symbol.** We denote by $C_c(\mathbb{C})$ the space of continuous functions $F : \mathbb{C} \to \mathbb{C}$ with bounded support and by μ_q the Lebesgue measure in $\mathbb{R}^q$.

Definition 2.1. Let $\{X_{\boldsymbol{m}}\}_n$ be a matrix-sequence, and let $f : D \to \mathbb{C}$ be a measurable function defined on a set $D \subset \mathbb{R}^q$, with $0 < \mu_q(D) < \infty$. We say that $\{X_{\boldsymbol{m}}\}_n$ is distributed like f in the sense of the singular values and we write $\{X_{\boldsymbol{m}}\}_n \sim_\sigma f$, if, for all $F \in C_c(\mathbb{C})$,

$$\lim_{n\to\infty} \frac{1}{N(\boldsymbol{m})} \sum_{j=1}^{N(\boldsymbol{m})} F(\sigma_j(X_{\boldsymbol{m}})) = \frac{1}{\mu_q(D)} \int_D F(|f(s_1,\dots,s_q)|)\,\mathrm{d}s_1\cdots\mathrm{d}s_q.$$

In this case, f is referred to as the singular value symbol of $\{X_{\boldsymbol{m}}\}_n$. Similarly, we say that $\{X_{\boldsymbol{m}}\}_n$ is distributed like f in the sense of the eigenvalues and we write $\{X_{\boldsymbol{m}}\}_n \sim_\lambda f$, if, for all $F \in C_c(\mathbb{C})$,

$$\lim_{n\to\infty} \frac{1}{N(\boldsymbol{m})} \sum_{j=1}^{N(\boldsymbol{m})} F(\lambda_j(X_{\boldsymbol{m}})) = \frac{1}{\mu_q(D)} \int_D F(f(s_1,\dots,s_q))\,\mathrm{d}s_1\cdots\mathrm{d}s_q.$$

In this case, f is referred to as the eigenvalue (or spectral) symbol of $\{X_{\boldsymbol{m}}\}_n$.

Definition 2.1 can be interpreted as follows. If f is continuous, $\{X_{\boldsymbol{m}}\}_n \sim_\lambda f$, n is large enough, and $\{\mathbf{s}_j^{(n)},\ j = 1,\dots,N(\boldsymbol{m})\}$ is an equispaced grid on D, then the eigenvalues of $X_{\boldsymbol{m}}$ can be ordered, say $\lambda_j(X_{\boldsymbol{m}})$, $j = 1,\dots,N(\boldsymbol{m})$, such that the pairs $\{(\mathbf{s}_j^{(n)}, \lambda_j(X_{\boldsymbol{m}})),\ j = 1,\dots,N(\boldsymbol{m})\}$ reconstruct approximately the hypersurface $\{(\mathbf{s}, f(\mathbf{s})),\ \mathbf{s} \in D\}$. In other words, the spectrum of $X_{\boldsymbol{m}}$ behaves approximately like a uniform sampling of f over D. For instance, if f is continuous, $q = 1$, $N(\boldsymbol{m}) = n$, and $D = [a, b]$, then the eigenvalues of $X_{\boldsymbol{m}}$ are approximately equal to $f(a + j(b-a)/n)$, $j = 1,\dots,n$, for n large enough. Analogously, if f is continuous, $q = 2$, $N(\boldsymbol{m}) = n^2$, and $D = [a_1, b_1] \times [a_2, b_2]$, then the eigenvalues of $X_{\boldsymbol{m}}$ are approximately equal to $f(a_1 + j_1(b_1 - a_1)/n, a_2 + j_2(b_2 - a_2)/n)$, $j_1, j_2 = 1,\dots,n$, for n large enough.

Remark 2.2. In the context of PDEs defined on a d-dimensional domain Ω and approximated by a local method (such as FDs, FEs and IgA), the number q appearing in Definition 2.1 is generally equal to $2d$ and the variables can be grouped into two sets having different meanings. More precisely, $(s_1,\dots,s_q) = (\hat{\mathbf{x}}, \boldsymbol{\theta}) \in [0,1]^d \times [-\pi,\pi]^d$, where $\hat{\mathbf{x}} := (\hat{x}_1,\dots,\hat{x}_d)$ are the parametric variables related to the physical domain Ω via the geometry map (possibly the identity map), and where $\boldsymbol{\theta} := (\theta_1,\dots,\theta_d)$ are the Fourier variables intrinsically related to the vibrations of the differential operator via the numerical approximation. There is an exception: if the matrix K in (1.1) is constant and the geometry map is an affine transformation, then the number of variables appearing in the distribution reduces to d because only the Fourier variables will show up; see, e.g., [11].

2.4. **Toeplitz and diagonal sampling matrices.** Given $\boldsymbol{m} \in \mathbb{N}^d$ and a function $f : [-\pi,\pi]^d \to \mathbb{C}$ belonging to $L^1([-\pi,\pi]^d)$, the d-level Toeplitz matrix $T_{\boldsymbol{m}}(f)$ associated with f is defined as follows [4, 28]:

$$T_{\boldsymbol{m}}(f) := \big[\hat{f}_{\boldsymbol{i}-\boldsymbol{j}}\big]_{\boldsymbol{i},\boldsymbol{j}=\boldsymbol{1}}^{\boldsymbol{m}},$$

where

$$\hat{f}_{\boldsymbol{\ell}} := \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} f(\boldsymbol{\theta})\mathrm{e}^{-\mathrm{i}\,\boldsymbol{\ell}\cdot\boldsymbol{\theta}}\mathrm{d}\boldsymbol{\theta}, \quad \boldsymbol{\ell} \in \mathbb{Z}^d,$$

are the Fourier coefficients of f, and $\boldsymbol{\ell}\cdot\boldsymbol{\theta} := \sum_{i=1}^d \ell_i\theta_i$. The function f is referred to as the generating function of the Toeplitz family $\{T_{\boldsymbol{m}}(f)\}_{\boldsymbol{m}\in\mathbb{N}^d}$. If f is real, then $T_{\boldsymbol{m}}(f)$ is Hermitian for all $\boldsymbol{m} \in \mathbb{N}^d$. The Toeplitz operator $T_{\boldsymbol{m}}(\cdot) : L^1([-\pi,\pi]^d) \to \mathbb{C}^{N(\boldsymbol{m})\times N(\boldsymbol{m})}$ is linear, so for $a, b \in \mathbb{C}$ and $f, g \in L^1([-\pi,\pi]^d)$ we have

$$T_{\boldsymbol{m}}(af + bg) = a\,T_{\boldsymbol{m}}(f) + b\,T_{\boldsymbol{m}}(g).$$

Given $f_i : E_i \to \mathbb{C}$, $i = 1,\dots,d$, the tensor-product function $f_1 \otimes \cdots \otimes f_d : E_1 \times \cdots \times E_d \to \mathbb{C}$ is defined by

$$(f_1 \otimes \cdots \otimes f_d)(\zeta_1,\dots,\zeta_d) := f_1(\zeta_1)\cdots f_d(\zeta_d), \quad (\zeta_1,\dots,\zeta_d) \in E_1 \times \cdots \times E_d.$$

For $f_1,\dots,f_d \in L^1([-\pi,\pi])$ and $\boldsymbol{m} \in \mathbb{N}^d$, we have

$$T_{\boldsymbol{m}}(f_1 \otimes \cdots \otimes f_d) = T_{m_1}(f_1) \otimes \cdots \otimes T_{m_d}(f_d). \tag{2.6}$$

Given $\boldsymbol{m} \in \mathbb{N}^d$ and $a : [0,1]^d \to \mathbb{C}$, the d-level diagonal sampling matrix $D_{\boldsymbol{m}}(a)$ associated with a is the $N(\boldsymbol{m}) \times N(\boldsymbol{m})$ diagonal matrix defined by

$$D_{\boldsymbol{m}}(a) := \operatorname*{diag}_{\boldsymbol{j}=\boldsymbol{1},\dots,\boldsymbol{m}} a\left(\frac{\boldsymbol{j}}{\boldsymbol{m}}\right),$$

where $\boldsymbol{j}$ varies from $\boldsymbol{1}$ to $\boldsymbol{m}$ following the lexicographic ordering (2.1), as explained in Section 2.1.

2.5. **GLT sequences.** A GLT sequence $\{X_{\boldsymbol{m}}\}_n$ is a specific matrix-sequence equipped with a measurable function $\chi : [0,1]^d \times [-\pi,\pi]^d \to \mathbb{C}$, the so-called kernel.[1] In the context of PDEs defined on a d-dimensional domain Ω and approximated by a local method, the resulting sequence of discretization matrices $\{X_{\boldsymbol{m}}\}_n$ is often a GLT sequence whose singular value and spectral symbol coincide with the kernel χ; see [12, 23, 24]. For this reason and in view of Remark 2.2, a point of $[0,1]^d \times [-\pi,\pi]^d$ will be denoted by $(\hat{\mathbf{x}}, \boldsymbol{\theta})$, where $\hat{\mathbf{x}} := (\hat{x}_1,\dots,\hat{x}_d)$ are the parametric variables and $\boldsymbol{\theta} := (\theta_1,\dots,\theta_d)$ are the Fourier variables.

The formal definition of GLT sequences (see [12, Section 5.1]) is quite cumbersome. Elaborating such a definition here would require the preliminary introduction of the a.c.s. notion, as well as the concept of locally Toeplitz operator and separable locally Toeplitz sequences. Instead of providing all this machinery, it is more useful to overview the main properties of GLT sequences that will be needed throughout this paper. Actually, these properties give a better idea of the meaning (and the potential) of GLT sequences than the definition itself. In the following, we write $\{X_{\boldsymbol{m}}\}_n \sim_{\mathrm{GLT}} \chi$ to indicate that $\{X_{\boldsymbol{m}}\}_n$ is a GLT sequence with kernel χ.

GLT 1. If $\{X_{\boldsymbol{m}}\}_n \sim_{\mathrm{GLT}} \chi$, then $\{X_{\boldsymbol{m}}\}_n \sim_\sigma \chi$. If, moreover, the matrices $X_{\boldsymbol{m}}$ are Hermitian, then $\{X_{\boldsymbol{m}}\}_n \sim_\lambda \chi$.

[1] Usually, χ is referred to as the symbol. However, to avoid confusion with the singular value and spectral symbol, in this paper we use the word kernel.

GLT 2. It holds that:
- $\{T_{\boldsymbol{m}}(f)\}_n \sim_{\text{GLT}} \chi(\hat{\mathbf{x}}, \boldsymbol{\theta}) := f(\boldsymbol{\theta})$ if $f \in L^1([-\pi,\pi]^d)$;
- $\{D_{\boldsymbol{m}}(a)\}_n \sim_{\text{GLT}} \chi(\hat{\mathbf{x}}, \boldsymbol{\theta}) := a(\hat{\mathbf{x}})$ if $a : [0,1]^d \to \mathbb{C}$ is Riemann-integrable;
- $\{Z_{\boldsymbol{m}}\}_n \sim_{\text{GLT}} \chi(\hat{\mathbf{x}}, \boldsymbol{\theta}) := 0$ if and only if $\{Z_{\boldsymbol{m}}\}_n \sim_\sigma 0$.

GLT 3. If $X_{\boldsymbol{m}} := \sum_{i=1}^r c_i \prod_{j=1}^{q_i} X_{\boldsymbol{m}}^{(i,j)}$, where $r, q_1, \ldots, q_r \in \mathbb{N}$, $c_1, \ldots, c_r \in \mathbb{C}$, and $\{X_{\boldsymbol{m}}^{(i,j)}\}_n \sim_{\text{GLT}} \chi_{ij}$, then $\{X_{\boldsymbol{m}}\}_n \sim_{\text{GLT}} \chi := \sum_{i=1}^r c_i \prod_{j=1}^{q_i} \chi_{ij}$.

The above properties can be found in [12, Section 5.6]. **GLT 1** states the main asymptotic distribution results for GLT sequences. **GLT 2** lists the fundamental examples of GLT sequences, from which one can construct, via **GLT 3**, many other GLT sequences. Note that, in **GLT 2**, $\{T_{\boldsymbol{m}}(f)\}_n$ (resp., $\{D_{\boldsymbol{m}}(a)\}_n$) is a generic matrix-sequence extracted from the family $\{T_{\boldsymbol{m}}(f)\}_{\boldsymbol{m}\in\mathbb{N}^d}$ (resp., $\{D_{\boldsymbol{m}}(a)\}_{\boldsymbol{m}\in\mathbb{N}^d}$). From **GLT 3** we see that the GLT sequences form an algebra. More precisely, for any fixed sequence of d-indices $\{\boldsymbol{m} = \boldsymbol{m}(n)\}_n \subseteq \mathbb{N}^d$ such that $\boldsymbol{m} \to \infty$ as $n \to \infty$, the set

$$\mathscr{A} := \big\{\{X_{\boldsymbol{m}}\}_n : \ \{X_{\boldsymbol{m}}\}_n \sim_{\text{GLT}} \chi \text{ for some measurable } \chi : [0,1]^d \times [-\pi,\pi]^d \to \mathbb{C}\big\}$$

is an algebra on the complex field $\mathbb{C}$, the so-called GLT algebra [12, Section 5.5].

Any matrix-sequence $\{Z_{\boldsymbol{m}}\}_n$ such that $\{Z_{\boldsymbol{m}}\}_n \sim_\sigma 0$ is referred to as a zero-distributed sequence. A matrix-sequence $\{Z_{\boldsymbol{m}}\}_n$ is zero-distributed if and only if $Z_{\boldsymbol{m}} = R_{\boldsymbol{m}} + S_{\boldsymbol{m}}$ for all n, where

$$\lim_{n\to\infty} \|S_{\boldsymbol{m}}\| = \lim_{n\to\infty} \frac{\operatorname{rank}(R_{\boldsymbol{m}})}{N(\boldsymbol{m})} = 0;$$

see [12, Theorem 2.5].

Remark 2.3. The multilevel version of Szegö's theorem [27, 29] is a consequence of **GLT 1** – **GLT 2**. Indeed, it follows from **GLT 1** – **GLT 2** that $\{T_{\boldsymbol{m}}(f)\}_n \sim_\sigma f$ for all $f \in L^1([-\pi,\pi]^d)$ and that $\{T_{\boldsymbol{m}}(f)\}_n \sim_\lambda f$ for all real-valued $f \in L^1([-\pi,\pi]^d)$, where we recall that the matrices $T_{\boldsymbol{m}}(f)$ are Hermitian if f is real. Therefore, using the notation of Definition 2.1, we see that in the case of pure d-level Toeplitz sequences we have $q = d$, $(s_1, \ldots, s_d) = (\theta_1, \ldots, \theta_d)$, and the parametric variables are not present. This is not surprising as Toeplitz matrices model space-invariant phenomena. For this reason, the notions 'space-invariant' and 'Toeplitz' are used as synonyms in the image processing community (see [22] and references therein). On the other hand, when considering sequences of d-level diagonal sampling matrices of the form $\{D_{\boldsymbol{m}}(a)\}_n$, a direct computation shows that $\{D_{\boldsymbol{n}}(a)\}_n \sim_\sigma a$ and $\{D_{\boldsymbol{n}}(a)\}_n \sim_\lambda a$ for all Riemann-integrable functions $a : [0,1]^d \to \mathbb{C}$. Hence, also for this kind of sequence we have $q = d$, but this time $(s_1, \ldots, s_d) = (\hat{x}_1, \ldots, \hat{x}_d)$ and the Fourier variables are not present.

3. Isogeometric Galerkin B-spline approximation

3.1. **Isogeometric Galerkin methods.** The weak form of (1.1) consists of finding $u \in H_0^1(\Omega)$ such that

$$a(u,v) = \mathrm{F}(v), \quad \forall v \in H_0^1(\Omega),$$

where $a(u,v) := \int_\Omega \big((\nabla u)^T K \nabla v + (\nabla u)^T \boldsymbol{\alpha}\, v + \gamma u v\big)$ and $\mathrm{F}(v) := \int_\Omega \mathrm{f} v$. In the standard Galerkin method, we look for an approximation $u_{\not\#}$ of u by choosing a

finite dimensional approximation space $\mathscr{W} \subset H_0^1(\Omega)$ and by solving the following (Galerkin) problem: find $u_{\mathscr{W}} \in \mathscr{W}$ such that

$$a(u_{\mathscr{W}}, v) = \mathrm{F}(v), \quad \forall v \in \mathscr{W}.$$

If $\{\varphi_1, \dots, \varphi_N\}$ is a basis of $\mathscr{W}$, then $u_{\mathscr{W}} = \sum_{j=1}^N u_j \varphi_j$ for a unique vector $\mathbf{u} := (u_1, \dots, u_N)^T$, and, by linearity, the computation of $u_{\mathscr{W}}$ is equivalent to solving the linear system

$$A\mathbf{u} = \mathbf{f},$$

where

$$A := [a(\varphi_j, \varphi_i)]_{i,j=1}^N = \left[\int_\Omega \left((\nabla \varphi_j)^T K \nabla \varphi_i + (\nabla \varphi_j)^T \boldsymbol{\alpha}\, \varphi_i + \gamma \varphi_j \varphi_i \right) \right]_{i,j=1}^N \tag{3.1}$$

is the stiffness matrix and $\mathbf{f} := [\mathrm{F}(\varphi_i)]_{i=1}^N$.

Suppose that the physical domain Ω can be described by a global geometry function $\mathbf{G} : [0,1]^d \to \overline{\Omega}$, which is invertible and satisfies $\mathbf{G}(\partial([0,1]^d)) = \partial\overline{\Omega}$. Let

$$\{\hat{\varphi}_1, \dots, \hat{\varphi}_N\} \tag{3.2}$$

be a set of basis functions defined on the reference (parametric) domain $[0,1]^d$ and vanishing on the boundary $\partial([0,1]^d)$. In the Galerkin IgA approach, the approximation space is chosen as $\mathscr{W} := \langle \varphi_i : \ i = 1, \dots, N \rangle$, with

$$\varphi_i(\mathbf{x}) := \hat{\varphi}_i(\mathbf{G}^{-1}(\mathbf{x})) = \hat{\varphi}_i(\hat{\mathbf{x}}), \quad \mathbf{x} = \mathbf{G}(\hat{\mathbf{x}}). \tag{3.3}$$

The resulting stiffness matrix A is given by (3.1), with the basis functions φ_i defined in (3.3). Assuming that $\mathbf{G}$ and $\hat{\varphi}_i$, $i = 1, \dots, N$, are sufficiently regular, we can apply standard differential calculus to obtain the following expression for A in terms of $\mathbf{G}$ and $\hat{\varphi}_i$, $i = 1, \dots, N$:

$$A = \left[\int_{[0,1]^d} \left((\nabla \hat{\varphi}_j)^T K_{\mathbf{G}} \nabla \hat{\varphi}_i + (\nabla \hat{\varphi}_j)^T (J_{\mathbf{G}})^{-1} \boldsymbol{\alpha}(\mathbf{G})\, \hat{\varphi}_i + \gamma(\mathbf{G}) \hat{\varphi}_j \hat{\varphi}_i \right) |\det(J_{\mathbf{G}})| \right]_{i,j=1}^N, \tag{3.4}$$

where

$$K_{\mathbf{G}} := (J_{\mathbf{G}})^{-1} K(\mathbf{G}) (J_{\mathbf{G}})^{-T}, \tag{3.5}$$

and $J_{\mathbf{G}}$ is the Jacobian matrix of $\mathbf{G}$, i.e.,

$$J_{\mathbf{G}} := \left[\frac{\partial G_i}{\partial \hat{x}_j} \right]_{i,j=1}^d = \left[\frac{\partial x_i}{\partial \hat{x}_j} \right]_{i,j=1}^d.$$

In the context of IgA, the geometry map $\mathbf{G}$ is expressed in terms of the functions $\hat{\varphi}_i$, as in [9, Eq. (2.6)], and the functions $\hat{\varphi}_i$ themselves are usually tensor-product B-splines or NURBS. In this paper, the role of the $\hat{\varphi}_i$ will be played by tensor-product B-splines over uniform knot sequences. Furthermore, we do not confine ourselves to the isoparametric approach, because we allow the geometry map $\mathbf{G}$ to be any sufficiently smooth function from $[0,1]^d$ to $\overline{\Omega}$, not necessarily expressed in terms of B-splines.

3.2. **B-spline basis functions and IgA Galerkin matrices.** We now detail the explicit construction of our basis functions $\hat{\varphi}_i$.

For $p, n \geq 1$, consider the uniform knot sequence

$$t_1 = \cdots = t_{p+1} = 0 < t_{p+2} < \cdots < t_{p+n} < 1 = t_{p+n+1} = \cdots = t_{2p+n+1},$$

where

$$t_{i+p+1} := \frac{i}{n}, \quad i = 0, \ldots, n.$$

The B-splines of degree p on this knot sequence are denoted by

$$N_{i,[p]} : [0,1] \to \mathbb{R}, \quad i = 1, \ldots, n+p,$$

and are defined recursively as follows [3]: for $1 \leq i \leq n + 2p$,

$$N_{i,[0]}(x) := \begin{cases} 1, & \text{if } x \in [t_i, t_{i+1}), \\ 0, & \text{elsewhere;} \end{cases}$$

for $1 \leq k \leq p$ and $1 \leq i \leq n + 2p - k$,

$$N_{i,[k]}(x) := \frac{x - t_i}{t_{i+k} - t_i} N_{i,[k-1]}(x) + \frac{t_{i+k+1} - x}{t_{i+k+1} - t_{i+1}} N_{i+1,[k-1]}(x),$$

where we assume that a fraction with zero denominator is zero. We know from [3] that the functions $N_{1,[p]}, \ldots, N_{n+p,[p]}$ form a basis for the spline space

$$\left\{ s \in C^{p-1}([0,1]) : \ s_{|[\frac{i}{n}, \frac{i+1}{n})} \in \mathbb{P}_p, \ i = 0, \ldots, n-1 \right\},$$

where $\mathbb{P}_p$ is the space of polynomials of degree less than or equal to p. Moreover, the functions $N_{i,[p]}$ are non-negative, they form a partition of unity on $[0,1]$, and they have a local support, namely

$$\operatorname{supp}(N_{i,[p]}) = [t_i, t_{i+p+1}], \quad i = 1, \ldots, n+p.$$

Finally,

$$N_{i,[p]}(0) = N_{i,[p]}(1) = 0, \quad i = 2, \ldots, n+p-1.$$

Let $\boldsymbol{p} := (p_1, \ldots, p_d)$ and $\boldsymbol{n} := (n_1, \ldots, n_d)$ be multi-indices such that $p_i, n_i \geq 1$, $i = 1, \ldots, d$. The tensor-product B-splines $N_{\boldsymbol{i},[\boldsymbol{p}]} : [0,1]^d \to \mathbb{R}$ are defined by

$$N_{\boldsymbol{i},[\boldsymbol{p}]} := N_{i_1,[p_1]} \otimes \cdots \otimes N_{i_d,[p_d]}, \quad \boldsymbol{i} = \boldsymbol{2}, \ldots, \boldsymbol{n+p-1}. \tag{3.6}$$

In the framework of IgA based on (uniform) B-splines, the functions $\hat{\varphi}_1, \ldots, \hat{\varphi}_N$ in (3.2) are chosen as the tensor-product B-splines in (3.6). In this setting, $N = N(\boldsymbol{n} + \boldsymbol{p} - \boldsymbol{2})$. We adopt for the tensor-product B-splines (3.6) the lexicographic ordering (2.1).[2] This ordering is followed when assembling the stiffness matrix (3.4), which from now on will be denoted by $A^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}$ in order to emphasize its dependence on $\boldsymbol{p}, \boldsymbol{n}$ and the geometry map $\mathbf{G}$. In multi-index notation, we have

$$A^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}} := \left[\int_{[0,1]^d} \left((\nabla N_{\boldsymbol{j}+\boldsymbol{1},[\boldsymbol{p}]})^T K_{\mathbf{G}} \nabla N_{\boldsymbol{i}+\boldsymbol{1},[\boldsymbol{p}]} + (\nabla N_{\boldsymbol{j}+\boldsymbol{1},[\boldsymbol{p}]})^T (J_{\mathbf{G}})^{-1} \boldsymbol{\alpha}(\mathbf{G}) \, N_{\boldsymbol{i}+\boldsymbol{1},[\boldsymbol{p}]} \right.\right.$$
$$\left.\left. + \gamma(\mathbf{G}) N_{\boldsymbol{j}+\boldsymbol{1},[\boldsymbol{p}]} N_{\boldsymbol{i}+\boldsymbol{1},[\boldsymbol{p}]} \right) |\det(J_{\mathbf{G}})| \right]_{\boldsymbol{i},\boldsymbol{j}=\boldsymbol{1}}^{\boldsymbol{n+p-2}}.$$

Note that

$$A^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}} = K^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}} + R^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}, \tag{3.7}$$

[2] Although the lexicographic ordering is common in the literature, an alternative ordering has been used in the related papers [6–9, 11].

where

$$K_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]} := \left[\int_{[0,1]^d} (\nabla N_{\boldsymbol{j}+\mathbf{1},[\boldsymbol{p}]})^T K_{\mathbf{G}} |\det(J_{\mathbf{G}})| \nabla N_{\boldsymbol{i}+\mathbf{1},[\boldsymbol{p}]}\right]_{\boldsymbol{i},\boldsymbol{j}=\mathbf{1}}^{\boldsymbol{n}+\boldsymbol{p}-\mathbf{2}} \tag{3.8}$$

is the matrix resulting from the discretization of the diffusive term in (1.1), and

$$R_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]} := \left[\int_{[0,1]^d} \big((\nabla N_{\boldsymbol{j}+\mathbf{1},[\boldsymbol{p}]})^T (J_{\mathbf{G}})^{-1} \boldsymbol{\alpha}(\mathbf{G})\, N_{\boldsymbol{i}+\mathbf{1},[\boldsymbol{p}]} + \gamma(\mathbf{G}) N_{\boldsymbol{j}+\mathbf{1},[\boldsymbol{p}]} N_{\boldsymbol{i}+\mathbf{1},[\boldsymbol{p}]}\big) |\det(J_{\mathbf{G}})|\right]_{\boldsymbol{i},\boldsymbol{j}=\mathbf{1}}^{\boldsymbol{n}+\boldsymbol{p}-\mathbf{2}} \tag{3.9}$$

is the matrix resulting from the discretization of the terms in (1.1) with lower-order derivatives. The matrix $R_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}$ can be regarded as a 'residual term'. Indeed, we shall see that the norm of $R_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}$ is negligible with respect to the norm of the diffusion matrix $K_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}$ when the discretization parameters $\boldsymbol{n}$ are large.

4. Spectral distribution

4.1. **The symbol of the normalized IgA Galerkin matrices.** Let $\mathbb{Q}_+^d := \{\boldsymbol{r} := (r_1,\dots,r_d) \in \mathbb{Q}^d :\ r_i > 0,\ i = 1,\dots,d\}$ and fix $\boldsymbol{\nu} := (\nu_1,\dots,\nu_d) \in \mathbb{Q}_+^d$. From now on, we assume that $n_j = \nu_j n$ for each $j = 1,\dots,d$, i.e., $\boldsymbol{n} = \boldsymbol{\nu} n$. The discretization parameter n is assumed to vary in an infinite subset of $\mathbb{N}$ such that $\boldsymbol{n} = \boldsymbol{\nu} n \in \mathbb{N}^d$.

In our main result (Theorem 4.1), we consider the sequence of normalized matrices $\{n^{d-2} A_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}\}_n$, and we compute its singular value and eigenvalue distribution in the sense of Definition 2.1. The proof of Theorem 4.1 relies on the theory of GLT sequences [12, 23, 24], and, more precisely, it is based on the properties **GLT 1**–**GLT 3** described in Section 2.5. We shall show, in particular, that $\{n^{d-2} A_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}\}_n$ is a GLT sequence.

We start with some preliminary definitions. For $p \geq 0$, let $\phi_{[p]}$ be the cardinal B-spline of degree p, which is defined recursively over the uniform knot sequence $\{0, 1, \dots, p+1\}$ as follows [3]:

$$\phi_{[0]}(t) := \begin{cases} 1, & \text{if } t \in [0,1), \\ 0, & \text{elsewhere,} \end{cases}$$
$$\phi_{[p]}(t) := \frac{t}{p}\phi_{[p-1]}(t) + \frac{p+1-t}{p}\phi_{[p-1]}(t-1), \quad p \geq 1.$$

The cardinal B-spline $\phi_{[p]}$ is a non-negative piecewise polynomial of degree p and of class $C^{p-1}(\mathbb{R})$; its support is the interval $[0, p+1]$.

Let us now define the following $d \times d$ symmetric matrix $H_{\boldsymbol{p}}$, whose components are continuous functions in the Fourier variables $\boldsymbol{\theta} := (\theta_1,\dots,\theta_d) \in [-\pi,\pi]^d$:

$$(H_{\boldsymbol{p}})_{ij} := \begin{cases} \big(\bigotimes_{r=1}^{i-1} h_{p_r}\big) \otimes f_{p_i} \otimes \big(\bigotimes_{r=i+1}^{d} h_{p_r}\big), & \text{if } i = j, \\ \big(\bigotimes_{r=1}^{i-1} h_{p_r}\big) \otimes g_{p_i} \otimes \big(\bigotimes_{r=i+1}^{j-1} h_{p_r}\big) \otimes g_{p_j} \otimes \big(\bigotimes_{r=j+1}^{d} h_{p_r}\big), & \text{if } i < j, \\ \big(\bigotimes_{r=1}^{j-1} h_{p_r}\big) \otimes g_{p_j} \otimes \big(\bigotimes_{r=j+1}^{i-1} h_{p_r}\big) \otimes g_{p_i} \otimes \big(\bigotimes_{r=i+1}^{d} h_{p_r}\big), & \text{if } i > j, \end{cases} \tag{4.1}$$

where h_p, g_p, $f_p : [-\pi,\pi] \to \mathbb{R}$ are defined for all $p \geq 1$ by

$$h_p(\theta) := \phi_{[2p+1]}(p+1) + 2\sum_{k=1}^{p} \phi_{[2p+1]}(p+1-k)\cos(k\theta), \tag{4.2}$$

$$g_p(\theta) := -2\sum_{k=1}^{p} \dot{\phi}_{[2p+1]}\,(p+1-k)\sin(k\theta), \tag{4.3}$$

$$f_p(\theta) := -\ddot{\phi}_{[2p+1]}(p+1) - 2\sum_{k=1}^{p} \ddot{\phi}_{[2p+1]}(p+1-k)\cos(k\theta), \tag{4.4}$$

and $\phi_{[p]}$, $\dot{\phi}_{[p]}$, $\ddot{\phi}_{[p]}$ are, respectively, the cardinal B-spline of degree p, its first derivative and its second derivative. The (i,j)-th entry of the matrix $H_{\boldsymbol{p}}$ is related to the second-order partial derivative $-\frac{\partial^2}{\partial\hat{x}_i\partial\hat{x}_j}$.

Theorem 4.1. *Consider the differential problem* (1.1)*, and let* $\mathbf{G}$ *be a regular geometry map, i.e.,* $\mathbf{G} \in C^1([0,1]^d)$ *and* $\det(J_{\mathbf{G}}) \neq 0$ *in* $[0,1]^d$*. Then, the sequence of normalized IgA Galerkin matrices* $\{n^{d-2}A^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}\}_n$*, with* $\boldsymbol{n} = \boldsymbol{\nu}n$*, is distributed, in the sense of both the singular values and the eigenvalues, like the function* $f^{(\boldsymbol{\nu})}_{\mathbf{G},\boldsymbol{p}} : [0,1]^d \times [-\pi,\pi]^d \to \mathbb{R}$*,*

$$f^{(\boldsymbol{\nu})}_{\mathbf{G},\boldsymbol{p}}(\hat{\mathbf{x}},\boldsymbol{\theta}) := \frac{\boldsymbol{\nu}\left(|\det(J_{\mathbf{G}}(\hat{\mathbf{x}}))|\, K_{\mathbf{G}}(\hat{\mathbf{x}}) \circ H_{\boldsymbol{p}}(\boldsymbol{\theta})\right)\boldsymbol{\nu}^T}{N(\boldsymbol{\nu})}. \tag{4.5}$$

In formulas, $\{n^{d-2}A^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}\}_n \sim_\sigma f^{(\boldsymbol{\nu})}_{\mathbf{G},\boldsymbol{p}}$ *and* $\{n^{d-2}A^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}\}_n \sim_\lambda f^{(\boldsymbol{\nu})}_{\mathbf{G},\boldsymbol{p}}$.

Before going into the details of the proof of Theorem 4.1, we first outline the main idea in Section 4.2. The technique is quite general and offers an abstract framework for the computation of the singular value and eigenvalue distribution of matrix-sequences coming from a PDE discretization. It can also be applied in contexts other than Galerkin B-spline IgA. For instance, we can mention FDs [23, Section 6], FEs [2,14], and isogeometric collocation methods [9]. Since this technique is based on the theory of GLT sequences, we shall refer to it as the 'GLT analysis'.

Remark 4.2. The differential problem (1.1) is formulated under the assumption that K is SPD. The proof of Theorem 4.1 does not require that K is positive definite; only the symmetry of K is needed. Nevertheless, the positive definiteness of K is necessary for the ellipticity of (1.1).

Remark 4.3. In order to simplify the presentation of the proof of Theorem 4.1, the components of K are required to be continuous. Actually, the continuity requirement can be relaxed by assuming that the components of K are only piecewise continuous. With a little more effort, one could also extend the validity of Theorem 4.1 to the case where the components of $K(\mathbf{G})$ are Riemann integrable, taking into consideration that Riemann integrability is the classical assumption under which the singular value and eigenvalue distribution results for GLT sequences have been formulated; see, e.g., [23, Theorems 4.1–4.8].

Remark 4.4. The real challenge is to show that Theorem 4.1 holds even if the components of K are only in $L^\infty(\Omega)$, as in the usual formulation of problem (1.1); see [18, Section 3.4]. The extension of the proof to the L^∞ case would require the application of the Lusin theorem [19], which is used in function theory for

approximating a measurable function by a continuous function. We refer the reader to [21, Theorem 6.3] for an application of the Lusin theorem in the context of FDs under homogenization. The extension of Theorem 4.1 to the L^∞ case is an interesting topic for further investigation.

Remark 4.5. The proof of Theorem 4.1 does not really require that $\mathbf{G}$ is regular. By looking at the theory of GLT sequences, it is enough that the inverse of its Jacobian is made by functions ψ such that ψ_M is Riemann integrable for every positive number M, where, given a real-valued function ψ and $M > 0$, we define ψ_M such that

$$\psi_M(x) := \begin{cases} \psi(x), & \text{if } \psi(x) \in [-M, M], \\ M, & \text{if } \psi(x) > M, \\ -M, & \text{if } \psi(x) < -M. \end{cases}$$

Remark 4.6. The concept of geometry map is also known in other contexts; see, for instance, the homotopy map techniques [1, Sections 4.4 and 4.10] and their use when designing Finite Volume (FV) methods [30]. It is likely that the theory of GLT sequences (already applied for the spectral analysis of FD, FE, IgA discretization matrices) can be applied in connection with FVs as well.

4.2. **The idea for proving Theorem 4.1.** The main idea consists of the following six steps, which will be detailed afterwards. Here, we just outline the key ingredients so as to emphasize better the generality of the technique we are going to use. As mentioned before, this technique is referred to as the GLT analysis.

In the following, for any two sequences of real numbers $\{a_n\}_n$ and $\{b_n\}_n$, the notation $a_n = O(b_n)$ means that there exists a constant C, independent of n, such that $|a_n| \le C|b_n|$ for all sufficiently large n. On the other hand, $a_n = o(b_n)$ means $\lim_{n\to\infty} a_n/b_n = 0$. A sequence of real numbers $\{a_n\}_n$ is said to be bounded away from 0 (resp., ∞) if there exists a positive constant c such that $|a_n| \ge c$ (resp., $|a_n| \le c$) for all n. Note that the terminology 'bounded away from ∞' is equivalent to 'uniformly bounded with respect to n'.

Step 1. We recall from (3.9) that $R^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}$ is the matrix resulting from the discretization of the terms in (1.1) with lower-order derivatives. Since n^{d-2} is the 'correct' normalization factor, which keeps the spectral norm of the diffusion matrix $n^{d-2}K^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}$ bounded away from 0 and ∞, we can show that $\|n^{d-2}R^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}\| \to 0$ as $n \to \infty$, and more precisely $\|n^{d-2}R^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}\| = O(n^{-1})$. This implies that $\{n^{d-2}R^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}\}_n$ is zero-distributed, and **GLT 2** gives $\{n^{d-2}R^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}\}_n \sim_{\text{GLT}} 0$. Hence, by (3.7) and **GLT 3**, $\{n^{d-2}A^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}\}_n \sim_{\text{GLT}} f^{(\boldsymbol{\nu})}_{\mathbf{G},\boldsymbol{p}}$ if and only if $\{n^{d-2}K^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}\}_n \sim_{\text{GLT}} f^{(\boldsymbol{\nu})}_{\mathbf{G},\boldsymbol{p}}$. In this way, we have reduced the analysis of the sequence $\{n^{d-2}A^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}\}_n$ to the analysis of its strictly diffusive part $\{n^{d-2}K^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}\}_n$.

Step 2. Let $[L^1([0,1]^d)]^{d\times d}$ be the space of functions $L : [0,1]^d \to \mathbb{R}^{d\times d}$ such that $L_{ij} \in L^1([0,1]^d)$ for all $i,j = 1,\dots,d$. Let us consider the operator $\mathscr{L}^{[\boldsymbol{p}]}_{\boldsymbol{n}}(\cdot) : [L^1([0,1]^d)]^{d\times d} \to \mathbb{R}^{N(\boldsymbol{n}+\boldsymbol{p}-\boldsymbol{2})\times N(\boldsymbol{n}+\boldsymbol{p}-\boldsymbol{2})}$,

$$\mathscr{L}^{[\boldsymbol{p}]}_{\boldsymbol{n}}(L) := \left[\int_{[0,1]^d} (\nabla N_{\boldsymbol{j}+\boldsymbol{1},[\boldsymbol{p}]})^T L\, \nabla N_{\boldsymbol{i}+\boldsymbol{1},[\boldsymbol{p}]}\right]_{\boldsymbol{i},\boldsymbol{j}=\boldsymbol{1}}^{\boldsymbol{n}+\boldsymbol{p}-\boldsymbol{2}}. \tag{4.6}$$

By (3.8), we have

$$K_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]} = \mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(K_{\mathbf{G}}|\det(J_{\mathbf{G}})|).$$

In Steps 3–5, we use the linearity of $\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(\cdot)$ as well as the algebra structure of GLT sequences to prove that $\{n^{d-2}K_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}\}_n \sim_{\text{GLT}} f_{\mathbf{G},\boldsymbol{p}}^{(\nu)}$.

Step 3. Take $L = E_{st}$ in (4.6), where E_{st} is the $d \times d$ matrix having 1 in position (s,t) and 0 elsewhere. Note that if $\mathbf{G}$ is the identity map and we take $K = E_{st}$ in (1.1), then we are 'selecting' the second-order partial derivative $-\frac{\partial^2 u}{\partial x_s \partial x_t} = -\frac{\partial^2 u}{\partial \hat{x}_s \partial \hat{x}_t}$, which is a separable differential operator.[3] With a direct computation, one can show that $n^{d-2}\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(E_{st})$ is a d-level Toeplitz matrix up to a low-rank correction. Indeed,

$$n^{d-2}\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(E_{st}) = \frac{\nu_s\nu_t}{N(\boldsymbol{\nu})}\, T_{\boldsymbol{n}+\boldsymbol{p}-\mathbf{2}}((H_{\boldsymbol{p}})_{st}) + R_{\boldsymbol{n}+\boldsymbol{p}-\mathbf{2}}, \tag{4.7}$$

where $\operatorname{rank}(R_{\boldsymbol{n}+\boldsymbol{p}-\mathbf{2}}) = O(n^{d-1}) = O(N(\boldsymbol{n}+\boldsymbol{p}-\mathbf{2})/n)$. In (4.7), $H_{\boldsymbol{p}}(\boldsymbol{\theta})$ is the $d\times d$ symmetric matrix in the Fourier variables $\boldsymbol{\theta}$, whose (i,j)-th entry comes from the discretization of the second-order partial derivative $-\frac{\partial^2}{\partial \hat{x}_i \partial \hat{x}_j}$. In our specific case, this matrix is given by (4.1). Since $\operatorname{rank}(R_{\boldsymbol{n}+\boldsymbol{p}-\mathbf{2}})/N(\boldsymbol{n}+\boldsymbol{p}-\mathbf{2}) \to 0$, the sequence $\{R_{\boldsymbol{n}+\boldsymbol{p}-\mathbf{2}}\}_n$ is zero-distributed. Hence, from (4.7) and **GLT 2** – **GLT 3** we obtain

$$\{n^{d-2}\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(E_{st})\}_n \sim_{\text{GLT}} \frac{\nu_s\nu_t}{N(\boldsymbol{\nu})}(H_{\boldsymbol{p}}(\boldsymbol{\theta}))_{st}. \tag{4.8}$$

If L is a constant matrix, say $L = \sum_{s,t=1}^{d} L_{st}E_{st}$, then (4.8) and **GLT 3** yield

$$\{n^{d-2}\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(L)\}_n \sim_{\text{GLT}} \sum_{s,t=1}^{d} L_{st}\frac{\nu_s\nu_t}{N(\boldsymbol{\nu})}(H_{\boldsymbol{p}}(\boldsymbol{\theta}))_{st} = \frac{\boldsymbol{\nu}(L\circ H_{\boldsymbol{p}}(\boldsymbol{\theta}))\boldsymbol{\nu}^T}{N(\boldsymbol{\nu})}.$$

Step 4. Let us pass to the variable coefficient case and, inspired by Step 3, let us take $L = aE_{st}$ in (4.6), where $a \in C([0,1]^d)$. Note that if $\mathbf{G}$ is the identity map and we take $K = aE_{st}$ in (1.1), then we are 'selecting' $-\frac{\partial}{\partial \hat{x}_s}(a\frac{\partial u}{\partial \hat{x}_t})$, which coincides with the second-order separable differential operator $-a\frac{\partial^2 u}{\partial \hat{x}_s \partial \hat{x}_t}$ up to a term with a lower-order derivative, namely $-\frac{\partial a}{\partial \hat{x}_s}\frac{\partial u}{\partial \hat{x}_t}$. Given the local support of the basis functions $N_{\boldsymbol{i},[\boldsymbol{p}]}$, $\boldsymbol{i} = \mathbf{2},\dots,\boldsymbol{n}+\boldsymbol{p}-\mathbf{1}$, and the fact that $\operatorname{supp}(N_{\boldsymbol{i},[\boldsymbol{p}]})$ is located near the point $\boldsymbol{i}/\boldsymbol{n} := (i_1/n_1,\dots,i_d/n_d)$, it can be shown that

$$n^{d-2}\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(aE_{st}) = n^{d-2}D_{\boldsymbol{n}+\boldsymbol{p}-\mathbf{2}}(a)\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(E_{st}) + Q_{\boldsymbol{n}+\boldsymbol{p}-\mathbf{2}}. \tag{4.9}$$

[3] We say that a differential operator is separable if it is obtained by multiplying a given function with a product of partial derivatives. The general separable differential operator can be written as

$$a\frac{\partial^{r_1+\dots+r_d}u}{\partial x_1^{r_1}\cdots\partial x_d^{r_d}}.$$

An example of a non-separable differential operator is the Laplacian, which however can be written, like any other differential operator, as a sum of separable differential operators:

$$\Delta u = \sum_{k=1}^{d}\frac{\partial^2 u}{\partial x_k^2}.$$

What we are going to see is that a separable differential operator gives rise to a GLT sequence (or, more precisely, to a separable locally Toeplitz sequence [12, Section 4.2]). As a consequence, an arbitrary differential operator (a sum of separable differential operators) gives rise to a sum of GLT sequences, which is again a GLT sequence by **GLT 3**.

The matrix $Q_{\boldsymbol{n}+\boldsymbol{p}-\boldsymbol{2}}$ is a sparse matrix whose components are $O(\omega_a(n^{-1}))$ as well as its spectral norm. Here, $\omega_a(\cdot)$ stands for the modulus of continuity of the function a. Since $\omega_a(n^{-1}) \to 0$, the sequence $\{Q_{\boldsymbol{n}+\boldsymbol{p}-\boldsymbol{2}}\}_n$ is zero-distributed. Using (4.8)–(4.9) and **GLT 2** – **GLT 3** we get

$$\{n^{d-2}\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(aE_{st})\}_n \sim_{\text{GLT}} \frac{\nu_s\nu_t}{N(\boldsymbol{\nu})} a(\hat{\mathbf{x}})(H_{\boldsymbol{p}}(\boldsymbol{\theta}))_{st}. \tag{4.10}$$

Step 5. To obtain the relation $\{n^{d-2}K_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}\}_n \sim_{\text{GLT}} f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}$, we invoke the linearity of $\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(\cdot)$ and **GLT 3**. Since

$$K_{\mathbf{G}}|\det(J_{\mathbf{G}})| = \sum_{s,t=1}^{d} (K_{\mathbf{G}}|\det(J_{\mathbf{G}})|)_{st}E_{st},$$

with $(K_{\mathbf{G}}|\det(J_{\mathbf{G}})|)_{st} \in C([0,1]^d)$ for every $s,t = 1,\dots,d$, we have

$$n^{d-2}\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(K_{\mathbf{G}}|\det(J_{\mathbf{G}})|) = \sum_{s,t=1}^{d} n^{d-2}\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}((K_{\mathbf{G}}|\det(J_{\mathbf{G}})|)_{st}E_{st}).$$

Hence, by (4.10) and **GLT 3**,

$$\begin{aligned}\{n^{d-2}\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(K_{\mathbf{G}}|\det(J_{\mathbf{G}})|)\}_n &\sim_{\text{GLT}} \sum_{s,t=1}^{d} \frac{\nu_s\nu_t}{N(\boldsymbol{\nu})}(\,K_{\mathbf{G}}(\hat{\mathbf{x}})|\det(J_{\mathbf{G}}(\hat{\mathbf{x}}))|\,)_{st}(H_{\boldsymbol{p}}(\boldsymbol{\theta}))_{st}\\ &= \frac{\boldsymbol{\nu}\,(K_{\mathbf{G}}(\hat{\mathbf{x}})|\det(J_{\mathbf{G}}(\hat{\mathbf{x}}))|\circ H_{\boldsymbol{p}}(\boldsymbol{\theta}))\,\boldsymbol{\nu}^T}{N(\boldsymbol{\nu})}\\ &= f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}(\hat{\mathbf{x}},\boldsymbol{\theta}).\end{aligned}$$

Step 6. From Step 1 and Step 5 it follows that $\{n^{d-2}A_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}\}_n \sim_{\text{GLT}} f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}$. Hence, by **GLT 1**, $\{n^{d-2}A_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}\}_n \sim_\sigma f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}$, and, if the matrices $n^{d-2}A_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}$ are symmetric (this happens when $\boldsymbol{\alpha} = \mathbf{0}$), then $\{n^{d-2}A_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}\}_n \sim_\lambda f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}$. The eigenvalue distribution $\{n^{d-2}A_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}\}_n \sim_\lambda f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}$ continues to hold even if the matrices $n^{d-2}A_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}$ are not symmetric. Indeed,

a) $\{n^{d-2}K_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}\}_n \sim_\lambda f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}$ by **GLT 1**, because $\{n^{d-2}K_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}\}_n \sim_{\text{GLT}} f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}$ and the matrices $K_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}$ are symmetric;

b) $\|n^{d-2}K_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}\|$ is uniformly bounded with respect to n, due to the correct normalization factor n^{d-2} (see Step 1);

c) $\|n^{d-2}R_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}\| \to 0$ by Step 1, so $\|n^{d-2}R_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}\|_1 = o(N(\boldsymbol{n}+\boldsymbol{p}-\boldsymbol{2}))$ by (2.3).

Therefore, all the assumptions of [16, Theorem 3.4] or of its generalized version [13, Theorem 3.3] are satisfied, and it follows that $\{n^{d-2}A_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}\}_n \sim_\lambda f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}$.

4.3. **Proof of Step 1.** We show that $\|n^{d-2}R_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}\| = O(n^{-1})$. We also prove that $\|n^{d-2}K_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}\|$ is uniformly bounded with respect to n: this is asserted in Step 1 and is used in Step 6, item b). In the following, C denotes a generic constant independent of n.

By hypothesis, $\mathbf{G}$ is regular, so the components of $(J_{\mathbf{G}})^{-1}$ are continuous and bounded over $[0,1]^d$. Moreover, the coefficients γ and α_i, $i = 1,\dots,d$, in (1.1) are

in $L^\infty(\Omega)$. Hence, by (3.9), for all $\boldsymbol{i},\boldsymbol{j}=\boldsymbol{2},\dots,\boldsymbol{n}+\boldsymbol{p}-\boldsymbol{1}$ we have

$$\begin{aligned}
&\left|(R^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}})_{\boldsymbol{i}-\boldsymbol{1},\boldsymbol{j}-\boldsymbol{1}}\right| \\
&\le \int_{[0,1]^d} \left(\left|(\nabla N_{\boldsymbol{j},[\boldsymbol{p}]})^T (J_{\mathbf{G}})^{-1} \boldsymbol{\alpha}(\mathbf{G}) N_{\boldsymbol{i},[\boldsymbol{p}]}\right| + \left|\gamma(\mathbf{G}) N_{\boldsymbol{j},[\boldsymbol{p}]} N_{\boldsymbol{i},[\boldsymbol{p}]}\right| \right) |\det(J_{\mathbf{G}})| \\
&\le C\left[\int_{[0,1]^d} \sum_{r=1}^{d} \left|\frac{\partial N_{\boldsymbol{j},[\boldsymbol{p}]}}{\partial \hat{x}_r}\right| \left|N_{\boldsymbol{i},[\boldsymbol{p}]}\right| + \int_{[0,1]^d} \left|N_{\boldsymbol{j},[\boldsymbol{p}]}\right| \left|N_{\boldsymbol{i},[\boldsymbol{p}]}\right|\right] \\
&= C\left[\sum_{r=1}^{d} \int_0^1 |N'_{j_r,[p_r]}||N_{i_r,[p_r]}| \prod_{\substack{s=1\\ s\neq r}}^{d} \int_0^1 |N_{j_s,[p_s]}||N_{i_s,[p_s]}| + \prod_{s=1}^{d} \int_0^1 |N_{j_s,[p_s]}||N_{i_s,[p_s]}|\right].
\end{aligned}$$

From the non-negativity of the B-splines $N_{i,[p]}$, the results in [11, proof of Lemma 8] and the relation $\boldsymbol{n}=\boldsymbol{\nu} n$, we get

$$\left|(R^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}})_{\boldsymbol{i}-\boldsymbol{1},\boldsymbol{j}-\boldsymbol{1}}\right| \le C\left[\sum_{r=1}^{d} 2 \prod_{\substack{s=1\\ s\neq r}}^{d} \frac{1}{n_s} + \prod_{s=1}^{d} \frac{1}{n_s}\right] = O(n^{-d+1}).$$

In view of the local support property of the B-splines, $\operatorname{supp}(N_{i,[p]}) = [t_i, t_{i+p+1}]$, it is clear that $(R^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}})_{\boldsymbol{i}-\boldsymbol{1},\boldsymbol{j}-\boldsymbol{1}} = 0$ if $\|\boldsymbol{i}-\boldsymbol{j}\|_\infty > \|\boldsymbol{p}\|_\infty$. It follows that $R^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}$ is a sparse matrix, consisting of at most $(2\|\boldsymbol{p}\|_\infty+1)^d$ non-zero components in each row and column with a value $O(n^{-d+1})$. Thus, by (2.2), $\|R^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}\| = O(n^{-d+1})$ and $\|n^{d-2} R^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}\| = O(n^{-1})$.

To prove that $\|n^{d-2} K^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}\|$ is uniformly bounded with respect to n, we can follow the same pattern as for the proof of $\|n^{d-2} R^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}\| = O(n^{-1})$. By (3.8), for all $\boldsymbol{i},\boldsymbol{j}=\boldsymbol{2},\dots,\boldsymbol{n}+\boldsymbol{p}-\boldsymbol{1}$ we have

$$\begin{aligned}
&\left|(K^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}})_{\boldsymbol{i}-\boldsymbol{1},\boldsymbol{j}-\boldsymbol{1}}\right| \le \int_{[0,1]^d} \left|(\nabla N_{\boldsymbol{j},[\boldsymbol{p}]})^T K_{\mathbf{G}} \nabla N_{\boldsymbol{i},[\boldsymbol{p}]}\right| |\det(J_{\mathbf{G}})| \\
&\le C \int_{[0,1]^d} \sum_{r,s=1}^{d} \left|\frac{\partial N_{\boldsymbol{j},[\boldsymbol{p}]}}{\partial \hat{x}_r}\right| \left|\frac{\partial N_{\boldsymbol{i},[\boldsymbol{p}]}}{\partial \hat{x}_s}\right| \\
&= C\left[\sum_{r=1}^{d} \int_0^1 |N'_{j_r,[p_r]}||N'_{i_r,[p_r]}| \prod_{\substack{t=1\\ t\neq r}}^{d} \int_0^1 |N_{j_t,[p_t]}||N_{i_t,[p_t]}| \right. \\
&\qquad \left. + \sum_{\substack{r,s=1\\ r\neq s}}^{d} \int_0^1 |N'_{j_r,[p_r]}||N_{i_r,[p_r]}| \int_0^1 |N_{j_s,[p_s]}||N'_{i_s,[p_s]}| \prod_{\substack{t=1\\ t\neq r,s}}^{d} \int_0^1 |N_{j_t,[p_t]}| \, |N_{i_t,[p_t]}|\right] \\
&\le C\left[\sum_{r=1}^{d} 4 p_r n_r \prod_{\substack{t=1\\ t\neq r}}^{d} \frac{1}{n_t} + \sum_{\substack{r,s=1\\ r\neq s}}^{d} 2\cdot 2 \prod_{\substack{t=1\\ t\neq r,s}}^{d} \frac{1}{n_t}\right] = O(n^{-d+2}),
\end{aligned}$$

where in the last inequality we used again the non-negativity of the B-splines $N_{i,[p]}$ and the results in [11, proof of Lemma 8]. By means of the local support of $N_{i,[p]}$ and by following the same argument used for $R^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}$, it can be shown that $K^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}$ is

a sparse matrix, with at most $(2\|\boldsymbol{p}\|_\infty+1)^d$ non-zero components in each row and column. Thus, $\|n^{d-2}K^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}\| = O(1)$ is uniformly bounded with respect to n.

Finally, we note that in Step 2 there is nothing to prove.

4.4. **Proof of Step 3.** Let $1 \le s,t \le d$. We show that the rank of

$$R_{\boldsymbol{n}+\boldsymbol{p}-\mathbf{2}} := n^{d-2}\mathscr{L}^{[\boldsymbol{p}]}_{\boldsymbol{n}}(E_{st}) - \frac{\nu_s\nu_t}{N(\boldsymbol{\nu})}T_{\boldsymbol{n}+\boldsymbol{p}-\mathbf{2}}((H_{\boldsymbol{p}})_{st})$$

is $O(n^{d-1})$ as $n\to\infty$, with $H_{\boldsymbol{p}}$ as in (4.1). By following the construction of the matrices in [10, Eqs. (3.17), (4.38) and (5.17)], one can show that

$$\mathscr{L}^{[\boldsymbol{p}]}_{\boldsymbol{n}}(E_{st}) = \begin{cases} \big(\bigotimes_{r=1}^{s-1}\frac{1}{n_r}M^{[p_r]}_{n_r}\big)\otimes n_s K^{[p_s]}_{n_s}\otimes\big(\bigotimes_{r=s+1}^{d}\frac{1}{n_r}M^{[p_r]}_{n_r}\big), & \text{if } s=t,\\ -\big(\bigotimes_{r=1}^{s-1}\frac{1}{n_r}M^{[p_r]}_{n_r}\big)\otimes H^{[p_s]}_{n_s}\otimes\big(\bigotimes_{r=s+1}^{t-1}\frac{1}{n_r}M^{[p_r]}_{n_r}\big)\otimes H^{[p_t]}_{n_t}\otimes\big(\bigotimes_{r=t+1}^{d}\frac{1}{n_r}M^{[p_r]}_{n_r}\big), & \text{if } s<t,\\ -\big(\bigotimes_{r=1}^{t-1}\frac{1}{n_r}M^{[p_r]}_{n_r}\big)\otimes H^{[p_t]}_{n_t}\otimes\big(\bigotimes_{r=t+1}^{s-1}\frac{1}{n_r}M^{[p_r]}_{n_r}\big)\otimes H^{[p_s]}_{n_s}\otimes\big(\bigotimes_{r=s+1}^{d}\frac{1}{n_r}M^{[p_r]}_{n_r}\big), & \text{if } s>t, \end{cases}$$

where the matrices $M^{[p]}_n$, $H^{[p]}_n$, $K^{[p]}_n$ are defined for all $p,n\ge 1$ as follows:

$$nK^{[p]}_n := \left[\int_0^1 N'_{j+1,[p]}N'_{i+1,[p]}\right]_{i,j=1}^{n+p-2},$$
$$H^{[p]}_n := \left[\int_0^1 N'_{j+1,[p]}N_{i+1,[p]}\right]_{i,j=1}^{n+p-2},$$
$$\frac{1}{n}M^{[p]}_n := \left[\int_0^1 N_{j+1,[p]}N_{i+1,[p]}\right]_{i,j=1}^{n+p-2}.$$

Following the analysis in [11] (see in particular [11, Theorem 7, Eqs. (48)–(50) and Eqs. (65)–(68)]), we have that $M^{[p]}_n$, $H^{[p]}_n$, $K^{[p]}_n$ are small-rank perturbations of the Toeplitz matrices $T_{n+p-2}(h_p)$, $\mathrm{i}\,T_{n+p-2}(g_p)$, $T_{n+p-2}(f_p)$, where the functions h_p, g_p, f_p are defined in (4.2)–(4.4). More precisely,

$$\operatorname{rank}(M^{[p]}_n - T_{n+p-2}(h_p)) \le c_p,$$
$$\operatorname{rank}(H^{[p]}_n - \mathrm{i}\,T_{n+p-2}(g_p)) \le c_p,$$
$$\operatorname{rank}(K^{[p]}_n - T_{n+p-2}(f_p)) \le c_p,$$

with c_p a constant depending only on p; for instance, we can take $c_p = 4p-2$ or $c_p = 4p-4$, according to either [11, Eqs. (66), (68)] or [10, Eqs. (4.84), (4.86)]. Furthermore, we observe that, in the case $s=t$,

$$\frac{\nu_s^2}{N(\boldsymbol{\nu})}T_{\boldsymbol{n}+\boldsymbol{p}-\mathbf{2}}((H_{\boldsymbol{p}})_{ss}) = \frac{\nu_s^2}{N(\boldsymbol{\nu})}T_{\boldsymbol{n}+\boldsymbol{p}-\mathbf{2}}(h_{p_1}\otimes\cdots\otimes h_{p_{s-1}}\otimes f_{p_s}\otimes h_{p_{s+1}}\otimes\cdots\otimes h_{p_d}).$$

Recalling the relation $\boldsymbol{n} = \boldsymbol{\nu}n$, we have

$$n^{d-2}\mathscr{L}^{[\boldsymbol{p}]}_{\boldsymbol{n}}(E_{ss}) = \frac{\nu_s^2}{N(\boldsymbol{\nu})}M^{[p_1]}_{n_1}\otimes\cdots\otimes M^{[p_{s-1}]}_{n_{s-1}}\otimes K^{[p_s]}_{n_s}\otimes M^{[p_{s+1}]}_{n_{s+1}}\otimes\cdots\otimes M^{[p_d]}_{n_d}.$$

By invoking (2.6) and (2.4), we see that the rank of $R_{\boldsymbol{n+p-2}}$ in the case $s = t$ is bounded from above by

$$N(\boldsymbol{n}+\boldsymbol{p}-\boldsymbol{2})\sum_{i=1}^{d}\frac{c_{p_i}}{n_i+p_i-2} = O(n^{d-1}).$$

The same argument shows that $\operatorname{rank}(R_{\boldsymbol{n+p-2}}) = O(n^{d-1})$ even in the case $s \neq t$.

4.5. **Proof of Step 4.** Let $1 \le s, t \le d$. We show that, whenever $a : [0,1]^d \to \mathbb{R}$ is continuous,

$$\|n^{d-2}\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(aE_{st}) - n^{d-2}D_{\boldsymbol{n+p-2}}(a)\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(E_{st})\| = O(\omega_a(n^{-1})). \tag{4.11}$$

In the following, C denotes a generic constant independent of n.

For every $\boldsymbol{i}, \boldsymbol{j} = \boldsymbol{2}, \dots, \boldsymbol{n}+\boldsymbol{p}-\boldsymbol{1}$,

$$\begin{aligned}
&\left|(\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(aE_{st}) - D_{\boldsymbol{n+p-2}}(a)\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(E_{st}))_{\boldsymbol{i-1},\boldsymbol{j-1}}\right| \\
&= \left|(\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(aE_{st}))_{\boldsymbol{i-1},\boldsymbol{j-1}} - (D_{\boldsymbol{n+p-2}}(a))_{\boldsymbol{i-1},\boldsymbol{i-1}}(\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(E_{st}))_{\boldsymbol{i-1},\boldsymbol{j-1}}\right| \\
&= \left|\int_{[0,1]^d}\left[a(\hat{\mathbf{x}}) - a\left(\frac{\boldsymbol{i}-\boldsymbol{1}}{\boldsymbol{n}+\boldsymbol{p}-\boldsymbol{2}}\right)\right]\frac{\partial N_{\boldsymbol{j},[\boldsymbol{p}]}}{\partial \hat{x}_s}\frac{\partial N_{\boldsymbol{i},[\boldsymbol{p}]}}{\partial \hat{x}_t}\right| \\
&\le \max_{\hat{\mathbf{x}}\,\in\,\operatorname{supp}(N_{\boldsymbol{i},[\boldsymbol{p}]})}\left|a(\hat{\mathbf{x}}) - a\left(\frac{\boldsymbol{i}-\boldsymbol{1}}{\boldsymbol{n}+\boldsymbol{p}-\boldsymbol{2}}\right)\right|\int_{[0,1]^d}\left|\frac{\partial N_{\boldsymbol{j},[\boldsymbol{p}]}}{\partial \hat{x}_s}\right|\left|\frac{\partial N_{\boldsymbol{i},[\boldsymbol{p}]}}{\partial \hat{x}_t}\right|. \qquad (4.12)
\end{aligned}$$

Since $\operatorname{supp}(N_{\boldsymbol{i},[\boldsymbol{p}]}) = [t_{i_1}, t_{i_1+p_1+1}] \times \cdots \times [t_{i_d}, t_{i_d+p_d+1}]$ and

$$\left|\hat{x} - \frac{i-1}{n+p-2}\right| \le Cn^{-1}, \quad \hat{x} \in [t_i, t_{i+p+1}],$$

we see that

$$\max_{\hat{\mathbf{x}}\,\in\,\operatorname{supp}(N_{\boldsymbol{i},[\boldsymbol{p}]})}\left\|\hat{\mathbf{x}} - \frac{\boldsymbol{i}-\boldsymbol{1}}{\boldsymbol{n}+\boldsymbol{p}-\boldsymbol{2}}\right\|_\infty \le C\left(\min_{r=1,\dots,d} n_r\right)^{-1} \le Cn^{-1}.$$

Therefore, from the definition of modulus of continuity, i.e.,

$$\omega_a(\delta) := \max_{\substack{\hat{\mathbf{x}},\hat{\mathbf{y}}\in[0,1]^d \\ \|\hat{\mathbf{x}}-\hat{\mathbf{y}}\|_\infty \le \delta}} |a(\hat{\mathbf{x}}) - a(\hat{\mathbf{y}})|, \quad \delta \ge 0,$$

we obtain

$$\max_{\hat{\mathbf{x}}\,\in\,\operatorname{supp}(N_{\boldsymbol{i},[\boldsymbol{p}]})}\left|a(\hat{\mathbf{x}}) - a\left(\frac{\boldsymbol{i}-\boldsymbol{1}}{\boldsymbol{n}+\boldsymbol{p}-\boldsymbol{2}}\right)\right| \le C\omega_a(n^{-1}).$$

Moreover, we know from the proof of Step 1 that

$$\int_{[0,1]^d}\left|\frac{\partial N_{\boldsymbol{j},[\boldsymbol{p}]}}{\partial \hat{x}_s}\right|\left|\frac{\partial N_{\boldsymbol{i},[\boldsymbol{p}]}}{\partial \hat{x}_t}\right| \le Cn^{-d+2}.$$

Hence, from (4.12) we obtain

$$\left|(\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(aE_{st}) - D_{\boldsymbol{n+p-2}}(a)\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(E_{st}))_{\boldsymbol{i-1},\boldsymbol{j-1}}\right| \le C\omega_a(n^{-1})n^{-d+2}$$

and

$$\left|(n^{d-2}\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(aE_{st}) - n^{d-2}D_{\boldsymbol{n+p-2}}(a)\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(E_{st}))_{\boldsymbol{i-1},\boldsymbol{j-1}}\right| \le C\omega_a(n^{-1}),$$

for every $\boldsymbol{i}, \boldsymbol{j} = \boldsymbol{2}, \dots, \boldsymbol{n}+\boldsymbol{p}-\boldsymbol{1}$. In addition, the matrix $n^{d-2}\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(aE_{st}) - n^{d-2}D_{\boldsymbol{n+p-2}}(a)\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(E_{st})$, like the matrices $R_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}$ and $K_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}$ already considered in the proof of Step 1, has the entry 0 in each position $(\boldsymbol{i}, \boldsymbol{j})$ whenever $\|\boldsymbol{i}-\boldsymbol{j}\|_\infty > \|\boldsymbol{p}\|_\infty$, due to the local support of the B-splines. It follows that the number of

non-zero components of $n^{d-2}\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(aE_{st}) - n^{d-2}D_{\boldsymbol{n}+\boldsymbol{p}-\boldsymbol{2}}(a)\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(E_{st})$ in each row and column is at most $(2\|\boldsymbol{p}\|_\infty + 1)^d$, and so

$$\left\|n^{d-2}\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(aE_{st}) - n^{d-2}D_{\boldsymbol{n}+\boldsymbol{p}-\boldsymbol{2}}(a)\mathscr{L}_{\boldsymbol{n}}^{[\boldsymbol{p}]}(E_{st})\right\| \le C\omega_a(n^{-1}).$$

This proves (4.11).

Finally, we note that in Steps 5–6 there is nothing to prove.

4.6. **Final remarks.** Assume that $K := [\kappa_{ij}]_{i,j=1}^d$ and u are sufficiently smooth, say $\kappa_{ij} \in C^1(\Omega) \cap C(\overline{\Omega})$ for all $i, j = 1, \dots, d$, and $u \in C^2(\Omega) \cap C(\overline{\Omega})$. In this case, our model problem (1.1) can be reformulated as follows:

$$\begin{cases} -\mathbf{1}(K \circ Hu)\mathbf{1}^T + \boldsymbol{\beta} \cdot \nabla u + \gamma u = \mathrm{f}, & \text{in } \Omega, \\ u = 0, & \text{on } \partial\Omega, \end{cases} \tag{4.13}$$

where Hu is the Hessian of u,

$$(Hu)_{ij} := \frac{\partial^2 u}{\partial x_i \partial x_j},$$

and $\boldsymbol{\beta}$ collects the coefficients of the first-order derivatives in (1.1):

$$\beta_j := \alpha_j - \sum_{i=1}^d \frac{\partial \kappa_{ij}}{\partial x_i}.$$

For any $u : \overline{\Omega} \to \mathbb{R}$, consider the corresponding function defined on the parametric domain $[0,1]^d$ by

$$\hat{u} : [0,1]^d \to \mathbb{R}, \quad \hat{u}(\hat{\mathbf{x}}) := u(\mathbf{x}), \quad \mathbf{x} = \mathbf{G}(\hat{\mathbf{x}}).$$

In other words, $\hat{u} := u(\mathbf{G})$. Then, u satisfies (4.13) if and only if $\hat{u}$ satisfies the corresponding transformed problem

$$\begin{cases} -\mathbf{1}(K_{\mathbf{G}} \circ H\hat{u})\mathbf{1}^T + \boldsymbol{\beta}_{\mathbf{G}} \cdot \nabla \hat{u} + \gamma(\mathbf{G})\hat{u} = \mathrm{f}(\mathbf{G}), & \text{in } (0,1)^d, \\ \hat{u} = 0, & \text{on } \partial((0,1)^d), \end{cases} \tag{4.14}$$

where $H\hat{u}$ is the Hessian of $\hat{u}$,

$$(H\hat{u})_{ij} = \frac{\partial^2 \hat{u}}{\partial \hat{x}_i \partial \hat{x}_j},$$

$K_{\mathbf{G}}$ is given in (3.5), and $\boldsymbol{\beta}_{\mathbf{G}}$ is the transformed advection coefficient of the PDE, whose expression in terms of K, $\boldsymbol{\beta}$, $\mathbf{G}$ is complicated and hence not reported here.

Now assume that $\Omega = (0,1)^d$ and take $\mathbf{G}$ equal to the identity map. In this case, the problems (4.13) and (4.14) are the same, with $\mathbf{x} = \hat{\mathbf{x}}$, $u = \hat{u}$, $K_{\mathbf{G}} = K$, and the symbol (4.5) reduces to

$$\frac{\boldsymbol{\nu}(K(\hat{\mathbf{x}}) \circ H_{\boldsymbol{p}}(\boldsymbol{\theta}))\boldsymbol{\nu}^T}{N(\boldsymbol{\nu})}.$$

In particular, if we discretize equally in each direction by choosing $\boldsymbol{\nu} = \mathbf{1}$, then the symbol is

$$\mathbf{1}(K(\hat{\mathbf{x}}) \circ H_{\boldsymbol{p}}(\boldsymbol{\theta}))\mathbf{1}^T.$$

It is quite remarkable that the main operator in the Hörmander sense [17], namely $-\mathbf{1}(K \circ H\hat{u})\mathbf{1}^T$, has a discrete spectral counterpart $\mathbf{1}(K(\hat{\mathbf{x}}) \circ H_{\boldsymbol{p}}(\boldsymbol{\theta}))\mathbf{1}^T$ which looks formally the same. The similarity becomes even more evident if we recall (from Step 3 in Section 4.2) that $H_{\boldsymbol{p}}(\boldsymbol{\theta})$ is a matrix of trigonometric polynomials in

the Fourier variables, whose components are related to the discretization of the components of

$$-H := \left[-\frac{\partial^2}{\partial \hat{x}_i \partial \hat{x}_j}\right]_{i,j=1}^{d},$$

i.e., the opposite of the Hessian operator in the parametric variables $\hat{\mathbf{x}}$.[4]

In the case of an arbitrary domain Ω described by a non-trivial geometry map $\mathbf{G}$, the symbol is

$$\frac{\boldsymbol{\nu}(|\det(J_{\mathbf{G}}(\hat{\mathbf{x}}))|\, K_{\mathbf{G}}(\hat{\mathbf{x}}) \circ H_{\boldsymbol{p}}(\boldsymbol{\theta}))\boldsymbol{\nu}^T}{N(\boldsymbol{\nu})},$$

and when $\boldsymbol{\nu} = \mathbf{1}$ it reduces to

$$\mathbf{1}(|\det(J_{\mathbf{G}}(\hat{\mathbf{x}}))|\, K_{\mathbf{G}}(\hat{\mathbf{x}}) \circ H_{\boldsymbol{p}}(\boldsymbol{\theta}))\mathbf{1}^T.$$

Even in this case, the symbol preserves the formal structure of the higher-order differential operator $-\mathbf{1}(K_{\mathbf{G}} \circ H\hat{u})\mathbf{1}^T$ associated with the problem (4.14). We see, however, the appearance of the determinant factor $|\det(J_{\mathbf{G}})|$. This factor is not present when our PDE is approximated by isogeometric collocation methods [9], in which case the resulting symbol is given by

$$\boldsymbol{\nu}(K_{\mathbf{G}}(\hat{\mathbf{x}}) \circ H_{\boldsymbol{p}}(\boldsymbol{\theta}))\boldsymbol{\nu}^T,$$

and is formally identical to the main operator $-\mathbf{1}(K_{\mathbf{G}} \circ H\hat{u})\mathbf{1}^T$, especially when $\boldsymbol{\nu} = \mathbf{1}$. The determinant factor $|\det(J_{\mathbf{G}})|$ appearing in the context of isogeometric Galerkin methods is interpreted by saying that the symbol of the (normalized) isogeometric Galerkin matrices is less ill-conditioned than the symbol of the (normalized) isogeometric collocation matrices. Indeed, since $|\det(J_{\mathbf{G}})|$ appears in the numerator of the symbol, it helps in keeping the conditioning of the IgA Galerkin matrices moderate when the geometry map $\mathbf{G}$ is (nearly) singular. A geometry map $\mathbf{G}$ is (nearly) singular when $\det(J_{\mathbf{G}})$ is (nearly) zero at one or more points. We refer the reader to [25, Section 5.2] for the analysis of the map effect on the conditioning in the collocation setting.

We end this section with the following observation. Assume that $\Omega = (0,1)^d$, $\mathbf{G}$ is the identity map, and $K = I$ is the identity matrix. In this case, the function $f^{(\boldsymbol{\nu})}_{\mathbf{G},\boldsymbol{p}}$ is independent of the $\hat{\mathbf{x}}$-variables. Therefore, we can consider its restriction to the domain $[-\pi,\pi]^d$, i.e., the function $f^{(\boldsymbol{\nu})}_{\boldsymbol{p}} : [-\pi,\pi]^d \to \mathbb{R}$,

$$f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}(\boldsymbol{\theta}) := \frac{1}{N(\boldsymbol{\nu})} \sum_{i=1}^{d} \nu_i^2 (h_{p_1} \otimes \cdots \otimes h_{p_{i-1}} \otimes f_{p_i} \otimes h_{p_{i+1}} \otimes \cdots \otimes h_{p_d})(\boldsymbol{\theta}). \tag{4.15}$$

A direct application of Definition 2.1 shows that $f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}$ is an eigenvalue and singular value symbol for the normalized sequence $\{n^{d-2} A^{[\boldsymbol{p}]}_{\boldsymbol{n}}\}_n$, where $\boldsymbol{n} = \boldsymbol{\nu} n$ and $A^{[\boldsymbol{p}]}_{\boldsymbol{n}}$ is the matrix resulting from the Galerkin B-spline IgA approximation of (1.1) when $\Omega = (0,1)^d$, $\mathbf{G}$ is the identity map, and $K = I$. The symbol $f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}$ is exactly the one appearing in [10, Chapter 4], where the problem (1.1) was considered over the hypercube $(0,1)^d$ in the constant coefficient case ($K = I$); see also [11] for the original analysis in the univariate and bivariate settings.

[4] The matrix $H_{\boldsymbol{p}}(\boldsymbol{\theta})$ is sometimes referred to as 'the symbol of the (negative) Hessian operator'. With a careful consideration of the results presented in this paper, the reader could guess the origin of this terminology, which, however, is not completely rigorous from the mathematical viewpoint.

5. Properties of the symbol

The formal structure of the symbol $f^{(\boldsymbol{\nu})}_{\mathbf{G},\boldsymbol{p}}$, discussed in the previous section, is actually not surprising, because it is analogous to the structure of the symbols arising in the context of other approximation techniques, such as FDs and FEs. The determinant factor $|\det(J_{\mathbf{G}})|$, appearing in the context of isogeometric Galerkin methods, was already observed in the FE setting (see [2]), but it is not present in the framework of isogeometric collocation methods (see [9]) and FDs (see [23, Section 6] and [25]). Although the formal structure of the symbol $f^{(\boldsymbol{\nu})}_{\mathbf{G},\boldsymbol{p}}$ is shared by different approximation techniques, certain analytic features are not so common. In this section, we investigate the properties of the symbol $f^{(\boldsymbol{\nu})}_{\mathbf{G},\boldsymbol{p}}$, for which we heavily rely on results obtained in [9].

We first recall that the functions h_p, g_p, f_p in (4.2)–(4.4) coincide with the functions h_{2p+1}, g_{2p+1}, f_{2p+1} defined in [9, Eqs. (3.7)–(3.9)], which have been deeply studied in [9]. From the properties of these functions, we get the following results. The first result concerns the symbol $f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}$ introduced in (4.15) and can be deduced from [9, Lemmas 3.4 and 3.6].

Theorem 5.1. *The following properties hold.*

1. $c_{\boldsymbol{p},\boldsymbol{\nu}} \sum_{k=1}^{d}(2-2\cos\theta_k) \le f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}(\boldsymbol{\theta}) \le C_{\boldsymbol{\nu}} \sum_{k=1}^{d}(2-2\cos\theta_k)$, *where*

$$c_{\boldsymbol{p},\boldsymbol{\nu}} := \left(\frac{4}{\pi^2}\right)^{\sum_{i=1}^{d} p_i + d - 1} \frac{\min_{i=1,\dots,d} \nu_i^2}{N(\boldsymbol{\nu})}, \quad C_{\boldsymbol{\nu}} := \frac{\max_{i=1,\dots,d} \nu_i^2}{N(\boldsymbol{\nu})}.$$

2. *Let* $M_{f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}} := \max_{\boldsymbol{\theta}\in[-\pi,\pi]^d} f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}(\boldsymbol{\theta})$. *Then*

$$\frac{f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}(\theta_1,\dots,\theta_{j-1},\pi,\theta_{j+1},\dots,\theta_d)}{M_{f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}}} \le \frac{f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}(\theta_1,\dots,\theta_{j-1},\pi,\theta_{j+1},\dots,\theta_d)}{f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}(\theta_1,\dots,\theta_{j-1},\frac{\pi}{2},\theta_{j+1},\dots,\theta_d)} \le 2^{2-p_j},$$

for all $j = 1,\dots,d$.

In particular, $f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}$ *has a unique zero of order two at* $\boldsymbol{\theta} = \mathbf{0}$, *like the function* $\sum_{k=1}^{d}(2-2\cos\theta_k)$, *and the value* $f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}(\theta_1,\dots,\theta_{j-1},\pi,\theta_{j+1},\dots,\theta_d)/M_{f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}}$ *converges to* 0 *exponentially as* $p_j \to \infty$.

The zero of the symbol $f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}$ at $\boldsymbol{\theta} = \mathbf{0}$ is expected, because it is a canonical feature of the symbol associated with the discretization matrices of differential problems like (1.1). Indeed, this was already observed in the FD and FE settings. Such a zero is interpreted as a source of ill-conditioning for the corresponding IgA Galerkin matrices $n^{d-2}A^{[\boldsymbol{p}]}_{\boldsymbol{n}}$ in the low frequencies; see [7, Section 3.2.2] for the terminology of frequencies. On the other hand, when the degrees $\boldsymbol{p}$ are large, the normalized symbol $f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}/M_{f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}}$ takes very small values at the so-called π-edge points

$$\{\boldsymbol{\theta} \in [0,\pi]^d : \ \theta_i = \pi \text{ for some } i\}. \tag{5.1}$$

These 'numerical zeros' are interpreted as an ill-conditioning of $n^{d-2}A^{[\boldsymbol{p}]}_{\boldsymbol{n}}$ in the high frequencies corresponding to such points. This second (non-canonical) source of ill-conditioning is responsible for the convergence deterioration of standard multigrid methods when the p_i increase. A way to overcome this problem consists of adopting a multi-iterative strategy, as suggested in [6, 7] (see also [8]).

The next result concerns the matrix $H_{\boldsymbol{p}}$. We use the abbreviation SPSD for Symmetric Positive Semi-Definite.

Theorem 5.2. *For $d = 1, 2, 3$, the matrix $H_{\boldsymbol{p}}(\boldsymbol{\theta})$ in (4.1) is SPSD for all $\boldsymbol{\theta} \in [-\pi, \pi]^d$ and SPD for all $\boldsymbol{\theta} \in [-\pi, \pi]^d$ such that $\theta_1 \cdots \theta_d \neq 0$.*

Proof. The result for $d = 1, 2$ follows immediately from [9, Lemma 3.6 and Theorem 5.2], taking into account that $H_{\boldsymbol{p}} = f_p$ for $d = 1$. In the remainder, we focus on the case $d = 3$.

The symmetry of $H_{\boldsymbol{p}} = H_{p_1,p_2,p_3}$ can be directly seen from its definition, i.e.,

$$H_{p_1,p_2,p_3} := \begin{bmatrix} f_{p_1} \otimes h_{p_2} \otimes h_{p_3} & g_{p_1} \otimes g_{p_2} \otimes h_{p_3} & g_{p_1} \otimes h_{p_2} \otimes g_{p_3} \\ g_{p_1} \otimes g_{p_2} \otimes h_{p_3} & h_{p_1} \otimes f_{p_2} \otimes h_{p_3} & h_{p_1} \otimes g_{p_2} \otimes g_{p_3} \\ g_{p_1} \otimes h_{p_2} \otimes g_{p_3} & h_{p_1} \otimes g_{p_2} \otimes g_{p_3} & h_{p_1} \otimes h_{p_2} \otimes f_{p_3} \end{bmatrix}.$$

For any $p \geq 1$, let $e_p := f_p h_p - (g_p)^2$. Then,

$$\begin{aligned} \det(H_{p_1,p_2,p_3}) &= (f_{p_1} \otimes h_{p_2} \otimes h_{p_3})(h_{p_1} \otimes f_{p_2} \otimes h_{p_3})(h_{p_1} \otimes h_{p_2} \otimes f_{p_3}) \\ &\quad + 2\,(h_{p_1} \otimes g_{p_2} \otimes g_{p_3})(g_{p_1} \otimes h_{p_2} \otimes g_{p_3})(g_{p_1} \otimes g_{p_2} \otimes h_{p_3}) \\ &\quad - (f_{p_1} \otimes h_{p_2} \otimes h_{p_3})(h_{p_1} \otimes g_{p_2} \otimes g_{p_3})(h_{p_1} \otimes g_{p_2} \otimes g_{p_3}) \\ &\quad - (g_{p_1} \otimes h_{p_2} \otimes g_{p_3})(h_{p_1} \otimes f_{p_2} \otimes h_{p_3})(g_{p_1} \otimes h_{p_2} \otimes g_{p_3}) \\ &\quad - (g_{p_1} \otimes g_{p_2} \otimes h_{p_3})(g_{p_1} \otimes g_{p_2} \otimes h_{p_3})(h_{p_1} \otimes h_{p_2} \otimes f_{p_3}) \\ &= (h_{p_1} \otimes h_{p_2} \otimes h_{p_3}) \big[(e_{p_1} + (g_{p_1})^2) \otimes (e_{p_2} + (g_{p_2})^2) \otimes (e_{p_3} + (g_{p_3})^2) \\ &\qquad\qquad - e_{p_1} \otimes (g_{p_2})^2 \otimes (g_{p_3})^2 - (g_{p_1})^2 \otimes e_{p_2} \otimes (g_{p_3})^2 \\ &\qquad\qquad - (g_{p_1})^2 \otimes (g_{p_2})^2 \otimes (e_{p_3} + (g_{p_3})^2)\big] \\ &= (h_{p_1} \otimes h_{p_2} \otimes h_{p_3}) \big[e_{p_1} \otimes e_{p_2} \otimes e_{p_3} + (g_{p_1})^2 \otimes e_{p_2} \otimes e_{p_3} \\ &\qquad\qquad + e_{p_1} \otimes (g_{p_2})^2 \otimes e_{p_3} + e_{p_1} \otimes e_{p_2} \otimes (g_{p_3})^2\big]. \end{aligned}$$

We recall from [9, Lemmas 3.4–3.7] that

$$\begin{aligned} &h_p(\theta) > 0, \quad \theta \in [-\pi, \pi], \\ &f_p(\theta) > 0, \quad \theta \in [-\pi, \pi] \setminus \{0\}, \\ &e_p(\theta) > 0, \quad \theta \in [-\pi, \pi] \setminus \{0\}, \end{aligned}$$

and

$$f_p(0) = 0, \quad g_p(0) = 0, \quad e_p(0) = 0.$$

Hence, we deduce that $\det(H_{p_1,p_2,p_3}) \geq 0$ over $[-\pi, \pi]^3$. Moreover, if $\theta_1\theta_2\theta_3 \neq 0$, then $\det(H_{p_1,p_2,p_3}) > 0$. In addition, from [9, Lemma 3.6 and Theorem 5.2] we infer that $(H_{p_1,\ldots,p_d})_{1,1} > 0$ if $\theta_1 \neq 0$ and $\det\big(\big[(H_{p_1,p_2,p_3})_{i,j}\big]_{i,j=1}^2\big) > 0$ if $\theta_1\theta_2 \neq 0$. Therefore, Sylvester's criterion implies that H_{p_1,p_2,p_3} is SPD if $\theta_1\theta_2\theta_3 \neq 0$. Finally, H_{p_1,p_2,p_3} is SPSD over $[-\pi, \pi]^3$ by a continuity argument. ☐

Remark 5.3. We conjecture that Theorem 5.2 holds for every $d \geq 1$. It is clear from its definition (4.1) that $H_{\boldsymbol{p}}$ is symmetric for any d. To prove the positive definiteness of $H_{\boldsymbol{p}}$, it suffices to prove that the determinant of $H_{\boldsymbol{p}}$ is positive for any d. Indeed, from (4.1) we see that the upper-left $k \times k$ submatrix of $H_{\boldsymbol{p}} = H_{p_1,\ldots,p_d}$ can be written as

$$\big[(H_{p_1,\ldots,p_d})_{i,j}\big]_{i,j=1}^k = \big[(H_{p_1,\ldots,p_k})_{i,j} \otimes \big(\textstyle\bigotimes_{r=k+1}^d h_{p_r}\big)\big]_{i,j=1}^k,$$

and so

$$\det\big(\big[(H_{p_1,\dots,p_d})_{i,j}\big]_{i,j=1}^{k}\big) = \det(H_{p_1,\dots,p_k}) \otimes \big(\bigotimes_{r=k+1}^{d}(h_{p_r})^k\big).$$

Since $h_p > 0$ over $[-\pi,\pi]$, the k-th leading principal minor of $H_{\boldsymbol{p}}$ is positive if $\det(H_{p_1,\dots,p_k}) > 0$, and $H_{\boldsymbol{p}}$ is positive definite if

$$\det(H_{p_1,\dots,p_k}) > 0, \quad k = 1,\dots,d.$$

The fact that $H_{\boldsymbol{p}}$ is SPD follows by induction from Sylvester's criterion.

In the following, we assume that $H_{\boldsymbol{p}}(\boldsymbol{\theta})$ is SPSD for all $\boldsymbol{\theta} \in [-\pi,\pi]^d$. Under this assumption, we derive some interesting properties of the symbol $f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}$, which are certainly true for $d = 1,2,3$ by Theorem 5.2.

Theorem 5.4. *Assume that $H_{\boldsymbol{p}}(\boldsymbol{\theta})$ is SPSD for all $\boldsymbol{\theta} \in [-\pi,\pi]^d$. Then, the following properties hold.*

1. *The symbol $f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}$ is non-negative a.e. in $[0,1]^d \times [-\pi,\pi]^d$.*
2. *For every $(\hat{\mathbf{x}},\boldsymbol{\theta}) \in [0,1]^d \times [-\pi,\pi]^d$, we have*

$$\begin{aligned} f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}(\hat{\mathbf{x}},\boldsymbol{\theta}) &\geq \lambda_{\min}(K_{\mathbf{G}}(\hat{\mathbf{x}}))|\det(J_{\mathbf{G}}(\hat{\mathbf{x}}))|f_{\boldsymbol{p}}^{(\boldsymbol{\nu})}(\boldsymbol{\theta}), \\ f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}(\hat{\mathbf{x}},\boldsymbol{\theta}) &\leq \lambda_{\max}(K_{\mathbf{G}}(\hat{\mathbf{x}}))|\det(J_{\mathbf{G}}(\hat{\mathbf{x}}))|f_{\boldsymbol{p}}^{(\boldsymbol{\nu})}(\boldsymbol{\theta}). \end{aligned} \tag{5.2}$$

As a consequence, if

$$cI \leq K_{\mathbf{G}}|\det(J_{\mathbf{G}})| \leq CI \quad \text{a.e. in } [0,1]^d, \quad \text{for some } c,C > 0, \tag{5.3}$$

then

$$cf_{\boldsymbol{p}}^{(\boldsymbol{\nu})}(\boldsymbol{\theta}) \leq f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}(\hat{\mathbf{x}},\boldsymbol{\theta}) \leq Cf_{\boldsymbol{p}}^{(\boldsymbol{\nu})}(\boldsymbol{\theta}),$$

for all $\boldsymbol{\theta} \in [-\pi,\pi]^d$ and for almost every $\hat{\mathbf{x}} \in [0,1]^d$.

Proof. The first statement can be proved by following the argument used in the proof of [9, Theorem 5.4] or [10, Theorem 5.5], also taking into account that K is SPD over Ω by hypothesis.

Let us prove the second statement. It is clear that

$$\lambda_{\min}(K_{\mathbf{G}}(\hat{\mathbf{x}}))I \leq K_{\mathbf{G}}(\hat{\mathbf{x}}) \leq \lambda_{\max}(K_{\mathbf{G}}(\hat{\mathbf{x}}))I.$$

Therefore, by using the assumption that $H_{\boldsymbol{p}}(\boldsymbol{\theta})$ is SPSD and the definitions of $f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}$ and $f_{\boldsymbol{p}}^{(\boldsymbol{\nu})}$, from (2.5) we get (5.2). □

Remark 5.5. Alternatively, we could prove the first statement in Theorem 5.4 even without the assumption that $H_{\boldsymbol{p}}(\boldsymbol{\theta})$ is SPSD for all $\boldsymbol{\theta} \in [-\pi,\pi]^d$, at least in the case where $\mathbf{G}$ is regular and the components of K are continuous over $\overline{\Omega}$. Indeed, we know from Theorem 4.1 (applied with $\boldsymbol{\alpha} = \mathbf{0}$ and $\gamma = 0$) that $\{n^{d-2}K_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}\}_n \sim_\lambda f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}$. Therefore, by [16, Theorem 2.4], each point of the essential range of $f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}$ strongly attracts the spectrum of $n^{d-2}K_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}$ with infinite order (see [16, Definition 2.3] for the concept of spectral attraction). Roughly speaking, this means that, for each point in the essential range of $f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}$, the number of eigenvalues of $n^{d-2}K_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}$ that collapse on this point tends to ∞ with n. Since every matrix $n^{d-2}K_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}$ is positive definite, it is clear that no eigenvalue of $n^{d-2}K_{\mathbf{G},\boldsymbol{n}}^{[\boldsymbol{p}]}$ can collapse on a negative real number. Hence, the essential range of $f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}$ is contained in $[0,\infty)$ and, consequently, $f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})} \geq 0$ (a.e.) over its domain $[0,1]^d \times [-\pi,\pi]^d$.

We remark that the condition (5.3) is usually satisfied in practice. For instance, it is satisfied if

- $K \geq c_K I$ a.e. in Ω, for some $c_K > 0$;
- $\mathbf{G}$ is regular, as in Theorem 4.1.

Note that we only require the inequality $K \geq c_K I$ and not the opposite $K \leq C_K I$, because the latter is satisfied as the components of K are assumed to be in $L^\infty(\Omega)$.

Assuming that (5.3) is met and $H_{\boldsymbol{p}}$ is SPSD over $[-\pi,\pi]^d$, the second statement of Theorem 5.4 implies that $f^{(\boldsymbol{\nu})}_{\mathbf{G},\boldsymbol{p}}$ has the same behavior of $f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}$. In particular, for (almost) every $\hat{\mathbf{x}} \in [0,1]^d$, the function $f^{(\boldsymbol{\nu})}_{\mathbf{G},\boldsymbol{p}}(\hat{\mathbf{x}},\cdot) : [-\pi,\pi]^d \to \mathbb{R}$ has a unique zero of order two at $\boldsymbol{\theta} = \mathbf{0}$, like the functions $f^{(\boldsymbol{\nu})}_{\boldsymbol{p}}$ and $\sum_{k=1}^d (2 - 2\cos\theta_k)$. Furthermore, if one of the $\boldsymbol{p}$ parameters, say p_i, is large, then $f^{(\boldsymbol{\nu})}_{\mathbf{G},\boldsymbol{p}}(\hat{\mathbf{x}},\cdot)$ also has infinitely many numerical zeros located at the points $\{\boldsymbol{\theta} \in [0,\pi]^d : \ \theta_i = \pi\}$; and if all the $\boldsymbol{p}$ parameters are large, then all the π-edge points (5.1) are numerical zeros for $f^{(\boldsymbol{\nu})}_{\mathbf{G},\boldsymbol{p}}(\hat{\mathbf{x}},\cdot)$.

6. Strong and weak points of the analysis

In this section we discuss the strong and weak aspects of the presented GLT analysis, and we give a few illustrative numerical examples.

6.1. **Strong points of the analysis.**

There are mainly three strong points: the mildness of the assumptions on the coefficients of the differential problem and on the geometry map, the insight into the subspaces where the ill-conditioning arises, and the applicability to any dimensionality.

Mildness of the assumptions. The assumptions on K and $\mathbf{G}$ in Theorem 4.1 are quite mild. Actually, they can be relaxed further, even without changing the technique of the proof, but at the price of making its presentation much heavier. Let us detail such relaxations here.

The components of K can be allowed to be piecewise continuous over $\overline{\Omega}$ instead of continuous over $\overline{\Omega}$. Under this milder assumption, for $d = 1$, an additional term of rank bounded by Cp will show up in the analysis of the stiffness matrix, because the number of basis functions affected by the jumps will be proportional to p. For $d = 2$, if the loci of the jumps form a finite set of 'nice curves', then the number of basis functions affected by the jumps will be proportional to

$$\|\boldsymbol{p}\|_\infty (n_1 + n_2),$$

and therefore the rank of the new term in the analysis of the stiffness matrix is bounded by $C\|\boldsymbol{p}\|_\infty(n_1+n_2)$. For a general $d \geq 3$, a similar argument shows that the rank of this new term is bounded by

$$C\|\boldsymbol{p}\|_\infty N(\boldsymbol{n})(n_1^{-1} + \cdots + n_d^{-1}),$$

with C depending linearly on the number of jumps and on their analytic features. Since $C\|\boldsymbol{p}\|_\infty N(\boldsymbol{n})(n_1^{-1} + \cdots + n_d^{-1})$ is $o(N(\boldsymbol{n}))$, it follows that the sequence of stiffness matrices will have the same distribution as indicated in (4.5). Indeed, the new additional terms form a GLT sequence with zero symbol (i.e., a zero-distributed sequence) which does not change the global distribution function. Note that the great flexibility of the theory of GLT sequences admits a further weakening of the assumptions on K, as indicated in Remark 4.3.

The geometry map $\mathbf{G}$ can be allowed to be piecewise regular instead of regular (i.e., continuously differentiable and with invertible Jacobian); see Remark 4.5. For instance, a square can be mapped into an L-shaped domain or, analogously, a cube can be mapped into a cube where a smaller cube has been eliminated (Fichera corner). In such a situation, the symbol remains formally unchanged, but it will become unbounded; see Section 4.6 and [25, Section 5.2]. By looking at (4.5) and taking into account (3.5), we see that the inverse of the Jacobian of $\mathbf{G}$ is unbounded, and so the symbol is unbounded. This reflects the large eigenvalues, going to infinity as the mesh size $h = \frac{1}{n}$ becomes small.

The above phenomena are illustrated numerically in Section 6.3.

Insight into the subspaces producing ill-conditioning. The presented analysis gives precise indications on the nature of the subspaces where the ill-conditioning arises. In this sense, the IgA setting is really self-explanatory. If one of the components of $\boldsymbol{p}$, say p_i, becomes large, then numerical zeros appear in the domain where $\theta_i = \pi$ (see Section 5). This makes us aware that an ill-conditioned subspace is present in high frequencies and the latter information is crucial for designing a proper iterative solver. For instance, a standard black-box AMG (Algebraic MultiGrid) does not encapsulate this information and thus it is not robust with respect to $\boldsymbol{p}$. On the contrary, by exploiting such information from the symbol, a proper multigrid method endowed with a preconditioned Krylov smoother can be designed, resulting in an optimal and robust method with respect to $\boldsymbol{p}$; see [6–8] for more details.

Applicability to any dimensionality. The formal structure of the symbol in (4.5) and its properties analyzed in Section 5 are valid and remain the same in any dimensionality d. The multigrid method mentioned in the previous point (described in detail in [6–8]) exploits the information from the symbol regardless of the dimensionality d and is basically robust also with respect to d, overcoming in this way the 'curse of dimensionality'.

6.2. **Weak points of the analysis.** There are two weak points: boundary conditions are basically neglected, and in some cases the spectral information from the symbol is difficult to use in practice. The weak points can be summarized in the popular slogan 'there is no free lunch'.

Negligence of boundary conditions. The analysis does not capture the information related to the boundary conditions. Indeed, a change in the boundary conditions is reflected in a 'small rank' change in the matrix, i.e., in an additional GLT term with zero symbol. Consequently, the asymptotic distribution does not change. On the contrary, in the matrix the change is present in specific outliers whose behavior may be important in a structural analysis and for precisely understanding the behavior of (preconditioned) Krylov methods.

Difficult use of some spectral information. The spectral information delivered by the symbol can sometimes be difficult to use in practice, such as in the case of singularly perturbed problems and strong anisotropy.

In singularly perturbed problems, the elliptic part is multiplied by a small coefficient ϵ. If we make a coefficient ϵ dependent on the mesh size h (which is unnatural from a theoretical viewpoint, but it may happen in practice), then the picture of the symbol can be wildly affected. If $h = o(\epsilon(h))$, then the symbol is the same as in the case where ϵ is just a constant and the spectrum is not really affected. On the

contrary, if $\epsilon(h) \sim h$, then the first-order terms (usually neglected in the analysis, according to the notion of principal symbol by Hörmander [17]) become crucial and the symbol becomes complex-valued and wildly dependent on the proportionality constants hidden in the relation $\epsilon(h) \sim h$. In other words, our symbol analysis gives spectral information also in this case, but this spectral information is very difficult to use in practice for designing efficient iterative solvers.

The same picture shows up in the case of strong anisotropy: our symbol analysis delivers again the correct spectral information, but this spectral information may be difficult to exploit for designing efficient iterative solvers.

6.3. **Numerical examples.** In this section we present some numerical examples to illustrate the effectiveness of the symbol in Theorem 4.1 in describing the spectral properties of the IgA stiffness matrices.

In all the examples we consider our model problem (1.1) with $d = 1$, $\alpha = \gamma = 0$ and $\Omega = (0, 1)$. The diffusion coefficient κ and the geometry map G are specified in each example. Only non-constant diffusion coefficients and non-trivial geometry maps are chosen. Since $d = 1$, the symbol in Theorem 4.1 reads as

$$f^{(1)}_{G,p}(\hat{x}, \theta) := \frac{\kappa(G(\hat{x}))}{|G'(\hat{x})|} f_p(\theta), \tag{6.1}$$

where f_p is defined in (4.4). We only focus on the eigenvalue distribution because we are dealing with symmetric matrices. As explained in Section 2.3, the spectrum of a normalized stiffness matrix $\frac{1}{n} A^{[p]}_{G,n}$ behaves approximately like a uniform sampling of this symbol. According to Remark 2.2, we have $q = 2$ and $\hat{x} \in [0, 1]$, $\theta \in [-\pi, \pi]$. Suppose for the sake of simplicity that the stiffness matrices are of size m^2. Then, for each variable, we take a set consisting of m uniformly distributed points, namely

$$\left\{ \frac{i}{m},\ i = 1, \dots, m \right\}, \qquad \left\{ \frac{i\pi}{m},\ i = 1, \dots, m \right\}. \tag{6.2}$$

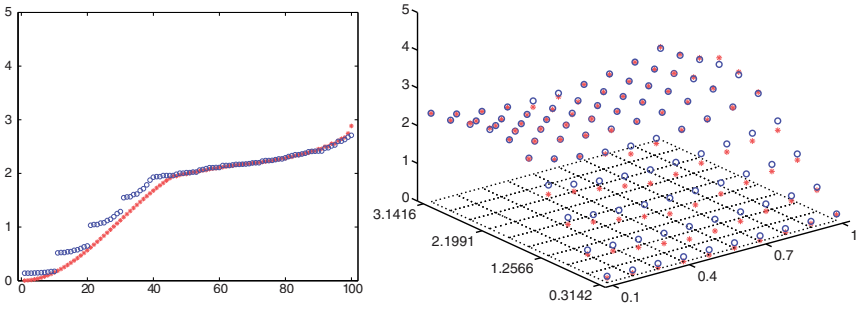
Note that, concerning the Fourier variables, it suffices to sample in the interval $[0, \pi]$ because of the symmetry of the function f_p.

In all the examples we compare the eigenvalues of a normalized stiffness matrix $\frac{1}{n} A^{[p]}_{G,n}$ and the sampled values of the symbol $f^{(1)}_{G,p}$ in (6.1) over the tensor-product grid constructed by the point sets in (6.2) where $n = m^2 - p + 2$. To make a fair comparison, both sets of values are sorted. The plots confirm that the eigenvalues of the normalized stiffness matrices follow the symbol, also when the diffusion coefficient is not continuous or the geometry map is not regular.

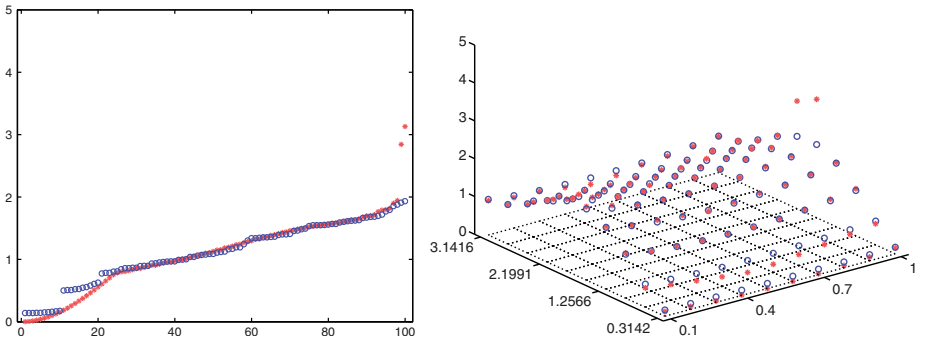
Example 1. Consider the continuous coefficient and regular geometry map

$$\kappa(x) = \mathrm{e}^x, \qquad G(\hat{x}) = \frac{1}{2} \hat{x}(\hat{x} + 1).$$

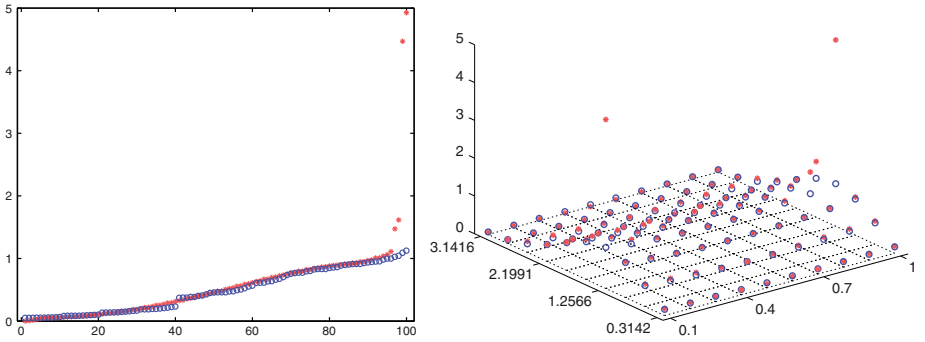
The eigenvalues of $\frac{1}{n} A^{[p]}_{G,n}$ and the sampled values of $f^{(1)}_{G,p}$ are depicted in Figure 1 for the values $p = 2, 3, 6$ and $m = 10$ ($n = 102 - p$). In addition to a plot of both sets of sorted values, we also show a bivariate plot taking into account the positional information of the symbol samples. No outliers are observed for $p = 2$, two outliers for $p = 3$ and four outliers for $p = 6$. The number of outliers depends on the degree p, but not on the parameter m.



(a) $p = 2$, $m = 10$

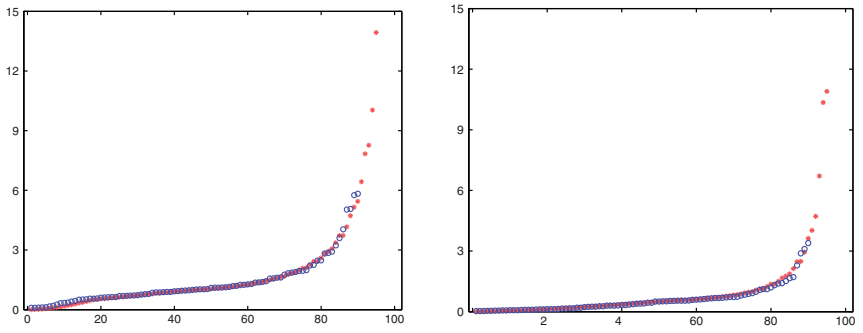


(b) $p = 3$, $m = 10$



(c) $p = 6$, $m = 10$

FIGURE 1. Example 1: comparison between equispaced samples of $f_{G,p}^{(1)}$ (blue circles) and the eigenvalues of $\frac{1}{n}A_{G,n}^{[p]}$ (red stars) with $n = m^2 - p + 2$.



(a) $p = 3$, $m = 10$ (b) $p = 6$, $m = 10$

FIGURE 2. Example 2: comparison between equispaced samples of $f_{G,p}^{(1)}$ (blue circles) and the eigenvalues of $\frac{1}{n}A_{G,n}^{[p]}$ (red stars) with $n = m^2 - p + 2$.

Example 2. Consider the continuous coefficient and non-regular geometry map

$$\kappa(x) = \mathrm{e}^x, \qquad G(\hat{x}) = \frac{1}{2}(1 - \cos(\pi\hat{x})).$$

Note that $G'(\hat{x})$ is not strictly positive for $\hat{x} \in [0, 1]$, so the symbol $f_{G,p}^{(1)}$ is unbounded. This is in agreement with the blow-up of the large eigenvalues of $\frac{1}{n}A_{G,n}^{[p]}$ when the mesh size $h = \frac{1}{n}$ becomes small. The eigenvalues of $\frac{1}{n}A_{G,n}^{[p]}$ and the sampled values of $f_{G,p}^{(1)}$ are depicted in Figure 2 for the values $p = 3, 6$ and $m = 10$ ($n = 102 - p$). The largest values of both sets are not visualized in the figure in order to improve the quality of the plot for the majority of the spectrum.

Example 3. Consider the discontinuous coefficient and regular geometry map

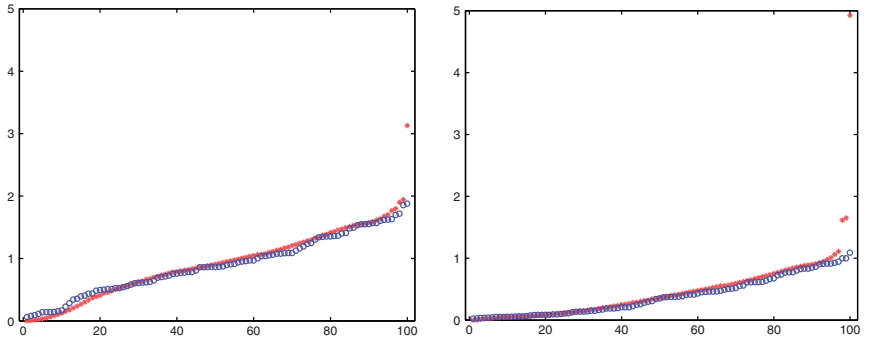
$$\kappa(x) = \begin{cases} \mathrm{e}^x, & x \in [0, 1/2], \\ 2 - x, & \text{elsewhere}, \end{cases} \qquad G(\hat{x}) = \frac{1}{2}\hat{x}(\hat{x} + 1).$$

The eigenvalues of $\frac{1}{n}A_{G,n}^{[p]}$ and the sampled values of $f_{G,p}^{(1)}$ are depicted in Figure 3 for the values $p = 3, 6$ and $m = 10$ ($n = 102 - p$).

Example 4. Consider the discontinuous coefficient and non-regular geometry map

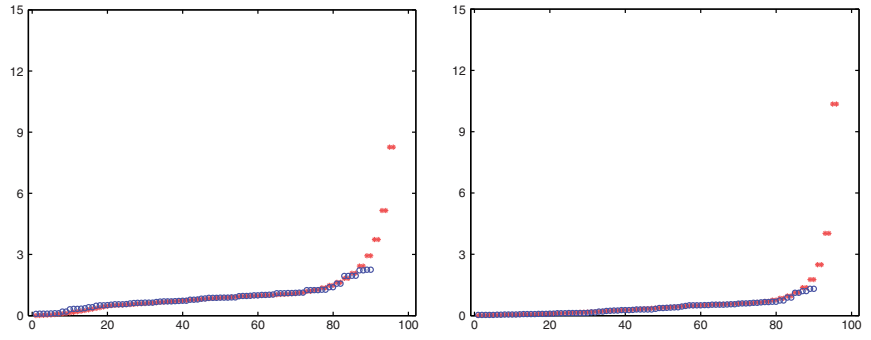
$$\kappa(x) = \begin{cases} \mathrm{e}^x, & x \in [0, 1/2], \\ 2 - x, & \text{elsewhere}, \end{cases} \qquad G(\hat{x}) = \frac{1}{2}(1 - \cos(\pi\hat{x})).$$

The eigenvalues of $\frac{1}{n}A_{G,n}^{[p]}$ and the sampled values of $f_{G,p}^{(1)}$ are depicted in Figure 4 for the values $p = 3, 6$ and $m = 10$ ($n = 102 - p$). As in Example 2, the largest values of both sets are not visualized in the figure in order to improve the quality of the plot for the majority of the spectrum.



(a) $p = 3$, $m = 10$ (b) $p = 6$, $m = 10$

FIGURE 3. Example 3: comparison between equispaced samples of $f^{(1)}_{G,p}$ (blue circles) and the eigenvalues of $\frac{1}{n}A^{[p]}_{G,n}$ (red stars) with $n = m^2 - p + 2$.



(a) $p = 3$, $m = 10$ (b) $p = 6$, $m = 10$

FIGURE 4. Example 4: comparison between equispaced samples of $f^{(1)}_{G,p}$ (blue circles) and the eigenvalues of $\frac{1}{n}A^{[p]}_{G,n}$ (red stars) with $n = m^2 - p + 2$.

7. Conclusions

We presented an asymptotic spectral analysis of the discretization matrices associated with the Galerkin B-spline IgA approximation of full elliptic PDEs. In particular, we computed the spectral and singular value symbol $f^{(\boldsymbol{\nu})}_{\mathbf{G},\boldsymbol{p}}$, in the sense of Definition 2.1, of the normalized sequence of matrices $\{n^{d-2}A^{[\boldsymbol{p}]}_{\mathbf{G},\boldsymbol{n}}\}_n$, with $\boldsymbol{n} = \boldsymbol{\nu}n$. We also collected some properties of the symbol $f^{(\boldsymbol{\nu})}_{\mathbf{G},\boldsymbol{p}}$, which – in agreement with previous results in the FD/FE/collocation contexts [2, 9, 23, 24] – has the canonical structure described in item b) of the introduction. The symbol is a non-negative function with a unique zero of order two at $\boldsymbol{\theta} = \mathbf{0}$. However, when p_i is large, there also appear infinitely many numerical zeros at the points $(\hat{\mathbf{x}}, \boldsymbol{\theta})$ where $\theta_i = \pi$. While the zero at $\boldsymbol{\theta} = \mathbf{0}$ is expected, because it is common to any approximation method

(see analogous features in the FD/FE/collocation symbols [2, 9, 23, 24]), the second type of zero leads to the uncommon fact that, for large $\|\boldsymbol{p}\|_\infty$, there is a subspace of high frequencies where the (normalized) Galerkin B-spline IgA stiffness matrices are ill-conditioned. This non-canonical feature is responsible for the slowdown, with respect to $\boldsymbol{p}$, of standard iterative methods. On the other hand, its knowledge and the knowledge of other properties of the (simplified) symbol $f_{\boldsymbol{p}}^{(\boldsymbol{\nu})}$ allowed us to construct a (multi-iterative) multigrid solver involving the PCG/PGMRES method as a smoother at the finest level, for which optimal convergence properties are numerically observed, with a remarkable robustness with respect to all the relevant parameters; see [6, 7].

The technique presented herein for computing the symbol $f_{\mathbf{G},\boldsymbol{p}}^{(\boldsymbol{\nu})}$ is based on the theory of GLT sequences [12, 23, 24], as well as on the results in [13, 16] for the non-symmetric case. This technique, called GLT analysis, is quite general and offers an abstract framework for the computation of the singular value and eigenvalue distribution of matrix-sequences coming from a PDE discretization. Besides Galerkin B-spline IgA, it has already been applied to FDs [23, Section 6], FEs [2, 14], and isogeometric collocation methods [9]. Future lines of research may include:

a) the application of the GLT analysis to the computation of the symbol of the discretization matrices resulting from the approximation of (1.1) by means of other numerical methods, such as, for instance, the isogeometric Galerkin and collocation methods based on NURBS instead of B-splines;
b) the generalization of the given results, especially Theorem 4.1, to the case where the components of K belong to $L^\infty(\Omega)$, as stated in Remark 4.4.

References

[1] R. A. Adams and J. J. F. Fournier, *Sobolev Spaces*, 2nd ed., Pure and Applied Mathematics (Amsterdam), vol. 140, Elsevier/Academic Press, Amsterdam, 2003. MR2424078 (2009e:46025)

[2] B. Beckermann and S. Serra-Capizzano, *On the asymptotic spectrum of finite element matrix sequences*, SIAM J. Numer. Anal. **45** (2007), no. 2, 746–769 (electronic), DOI 10.1137/05063533X. MR2300295 (2007m:65103)

[3] C. de Boor, *A Practical Guide to Splines*, Revised edition, Applied Mathematical Sciences, vol. 27, Springer-Verlag, New York, 2001. MR1900298 (2003f:41001)

[4] A. Böttcher and B. Silbermann, *Introduction to Large Truncated Toeplitz Matrices*, Universitext, Springer-Verlag, New York, 1999. MR1724795 (2001b:47043)

[5] J. A. Cottrell, T. J. R. Hughes, and Y. Bazilevs, *Isogeometric Analysis: Toward Integration of CAD and FEA*, John Wiley & Sons, 2009.

[6] M. Donatelli, C. Garoni, C. Manni, S. Serra-Capizzano, and H. Speleers, *Symbol-based multigrid methods for Galerkin B-spline isogeometric analysis*, Tech. Report TW650 (2014), Dept. Computer Science, KU Leuven. Available online at https://lirias.kuleuven.be/handle/123456789/461954.

[7] M. Donatelli, C. Garoni, C. Manni, S. Serra-Capizzano, and H. Speleers, *Robust and optimal multi-iterative techniques for IgA Galerkin linear systems*, Comput. Methods Appl. Mech. Engrg. **284** (2015), 230–264, DOI 10.1016/j.cma.2014.06.001. MR3310285

[8] M. Donatelli, C. Garoni, C. Manni, S. Serra-Capizzano, and H. Speleers, *Robust and optimal multi-iterative techniques for IgA collocation linear systems*, Comput. Methods Appl. Mech. Engrg. **284** (2015), 1120–1146, DOI 10.1016/j.cma.2014.11.036. MR3310320

[9] M. Donatelli, C. Garoni, C. Manni, S. Serra-Capizzano, and H. Speleers, *Spectral analysis and spectral symbol of matrices in isogeometric collocation methods*, Math. Comp. **85** (2016), no. 300, 1639–1680, DOI 10.1090/mcom/3027. MR3471103

[10] C. Garoni, *Structured matrices coming from PDE approximation theory: spectral analysis, spectral symbol and design of fast iterative solvers*, Ph.D. Thesis in Mathematics of Computation, University of Insubria, Como, Italy (2015). Available online at http://hdl.handle.net/10277/568.

[11] C. Garoni, C. Manni, F. Pelosi, S. Serra-Capizzano, and H. Speleers, *On the spectrum of stiffness matrices arising from isogeometric analysis*, Numer. Math. **127** (2014), no. 4, 751–799, DOI 10.1007/s00211-013-0600-2. MR3229992

[12] C. Garoni and S. Serra-Capizzano, *The theory of Generalized Locally Toeplitz sequences: A review, an extension, and a few representative applications*, Tech. Report 2015-023 (2015), Dept. Information Technology, Uppsala University. Available online at http://www.it.uu.se/research/publications/reports/2015-023/. This is the preliminary version of: C. Garoni and S. Serra-Capizzano, *Generalized Locally Toeplitz Sequences: Theory and Applications*, Springer Monograph (in preparation).

[13] C. Garoni, S. Serra-Capizzano, and D. Sesana, *Tools for determining the asymptotic spectral distribution of non-Hermitian perturbations of Hermitian matrix-sequences and applications*, Integral Equations Operator Theory **81** (2015), no. 2, 213–225, DOI 10.1007/s00020-014-2157-6. MR3299836

[14] C. Garoni, S. Serra-Capizzano, and D. Sesana, *Spectral analysis and spectral symbol of d-variate $\mathbb{Q}_p$ Lagrangian FEM stiffness matrices*, SIAM J. Matrix Anal. Appl. **36** (2015), no. 3, 1100–1128, DOI 10.1137/140976480. MR3376130

[15] C. Garoni, S. Serra-Capizzano, and P. Vassalos, *A general tool for determining the asymptotic spectral distribution of Hermitian matrix-sequences*, Oper. Matrices **9** (2015), no. 3, 549–561. MR3399336

[16] L. Golinskii and S. Serra-Capizzano, *The asymptotic properties of the spectrum of nonsymmetrically perturbed Jacobi matrix sequences*, J. Approx. Theory **144** (2007), no. 1, 84–102, DOI 10.1016/j.jat.2006.05.002. MR2287378 (2007k:47047)

[17] L. Hörmander, *Pseudo-differential operators and non-elliptic boundary problems*, Ann. of Math. (2) **83** (1966), 129–209. MR0233064 (38 #1387)

[18] A. Quarteroni, *Numerical Models for Differential Problems*, MS&A. Modeling, Simulation and Applications, vol. 2, Springer-Verlag Italia, Milan, 2009. Translated from the 4th (2008) Italian edition by Silvia Quarteroni. MR2522375 (2010h:65004)

[19] W. Rudin, *Real and Complex Analysis*, 3rd ed., McGraw-Hill Book Co., New York, 1987. MR924157 (88k:00002)

[20] S. Serra Capizzano, *Distribution results on the algebra generated by Toeplitz sequences: a finite-dimensional approach*, Linear Algebra Appl. **328** (2001), no. 1-3, 121–130, DOI 10.1016/S0024-3795(00)00311-6. MR1823514 (2002b:15016)

[21] S. Serra Capizzano, *Spectral behavior of matrix sequences and discretized boundary value problems*, Linear Algebra Appl. **337** (2001), 37–78, DOI 10.1016/S0024-3795(01)00335-4. MR1856851 (2002j:15005)

[22] S. Serra-Capizzano, *A note on antireflective boundary conditions and fast deblurring models*, SIAM J. Sci. Comput. **25** (2003/04), no. 4, 1307–1325, DOI 10.1137/S1064827502410244. MR2045058 (2005e:94013)

[23] S. Serra Capizzano, *Generalized locally Toeplitz sequences: spectral analysis and applications to discretized partial differential equations*, Linear Algebra Appl. **366** (2003), 371–402, DOI 10.1016/S0024-3795(02)00504-9. Special issue on structured matrices: analysis, algorithms and applications (Cortona, 2000). MR1987730 (2004d:47062)

[24] S. Serra-Capizzano, *The GLT class as a generalized Fourier analysis and applications*, Linear Algebra Appl. **419** (2006), no. 1, 180–233, DOI 10.1016/j.laa.2006.04.012. MR2263117 (2008a:15023)

[25] S. Serra Capizzano and C. Tablino Possio, *Analysis of preconditioning strategies for collocation linear systems*, Linear Algebra Appl. **369** (2003), 41–75, DOI 10.1016/S0024-3795(02)00719-X. MR1988478 (2004c:65035)

[26] P. Tilli, *Locally Toeplitz sequences: spectral properties and applications*, Linear Algebra Appl. **278** (1998), no. 1-3, 91–120, DOI 10.1016/S0024-3795(97)10079-9. MR1637331 (99g:47057)

[27] P. Tilli, *A note on the spectral distribution of Toeplitz matrices*, Linear and Multilinear Algebra **45** (1998), no. 2-3, 147–159, DOI 10.1080/03081089808818584. MR1671591 (99j:65063)

[28] E. E. Tyrtyshnikov, *A unifying approach to some old and new theorems on distribution and clustering*, Linear Algebra Appl. **232** (1996), 1–43, DOI 10.1016/0024-3795(94)00025-5. MR1366576 (96m:15018)

[29] E. E. Tyrtyshnikov and N. L. Zamarashkin, *Spectra of multilevel Toeplitz matrices: advanced theory via simple matrix relationships*, Linear Algebra Appl. **270** (1998), 15–27. MR1484073 (98i:65030)

[30] H. K. Versteeg and W. Malalasekera, *An Introduction to Computational Fluid Dynamics: The Finite Volume Method*, Second Edition, Pearson Education Limited, 2007.

Department of Mathematics, University of Rome 'Tor Vergata', Via della Ricerca Scientifica 1, 00133 Rome, Italy – and – Department of Science and High Technology, University of Insubria, Via Valleggio 11, 22100 Como, Italy

E-mail address: garoni@mat.uniroma2.it

E-mail address: carlo.garoni@uninsubria.it

Department of Mathematics, University of Rome 'Tor Vergata', Via della Ricerca Scientifica 1, 00133 Rome, Italy

E-mail address: manni@mat.uniroma2.it

Department of Science and High Technology, University of Insubria, Via Valleggio 11, 22100 Como, Italy – and – Department of Information Technology, Division of Scientific Computing, Uppsala University, Box 337, SE-751 05 Uppsala, Sweden

E-mail address: stefano.serrac@uninsubria.it

E-mail address: stefano.serra@it.uu.se

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

E-mail address: debora.sesana@uninsubria.it

Department of Mathematics, University of Rome 'Tor Vergata', Via della Ricerca Scientifica 1, 00133 Rome, Italy

E-mail address: speleers@mat.uniroma2.it