On the asymptotic spectrum of Finite Element matrix sequences

by

Bernhard Beckermann\textsuperscript{1 2} & Stefano Serra-Capizzano\textsuperscript{3 4}

Abstract

We derive a new formula for the asymptotic eigenvalue distribution of stiffness matrices obtained by applying $P_1$ Finite Elements with standard mesh refinement to the semi elliptic PDE of second order in divergence form $-\nabla(K\nabla^Tu) = f$ on $\Omega$, $u = g$ on $\partial\Omega$. Here $\Omega \subset \mathbb{R}^2$ and $K$ is supposed to be piecewise continuous and pointwise symmetric semi positive definite. The symbol describing this asymptotic eigenvalue distribution depends on the PDE, but also both on the numerical scheme for approaching the underlying bilinear form and on the geometry of triangulation of the domain. Our work is motivated by recent results on the superlinear convergence behavior of the Conjugate Gradient method, which requires the knowledge of such asymptotic eigenvalue distributions for sequences of matrices depending on a discretization parameter $h$ when $h \to 0$.

We compare our findings with similar results for the Finite Difference method which were published over the last years. In particular we observe that also our sequence of stiffness matrices is part of the class of Generalized Locally Toeplitz sequences for which many theoretical tools are available. This enables us to derive some results on the conditioning and preconditioning of such stiffness matrices.

Key words: Finite Element methods, matrix sequence, joint eigenvalue distribution

AMS Classification (2000): 65F10, 65N22, 15A18, 15A12, 47B65

1 Introduction and preliminary discussion

Consider the semi elliptic PDE of second order in divergence form

$$
-\nabla(K\nabla^Tu) = f \quad \text{on } \Omega, \quad u = g \quad \text{on } \partial\Omega,
$$

where $\Omega \subset \mathbb{R}^2$ is a bounded open "smooth" set (say, with piecewise $C^1$ boundary), and $K : \Omega \mapsto \mathbb{R}^{2 \times 2}$ is piecewise continuous in $\Omega$, and symmetric semi positive definite at each point of $\Omega$. In this paper we are interested in describing the asymptotic distribution of eigenvalues of the matrix of coefficients obtained by discretizing the above elliptic PDE by $P_1$ Finite Elements in the case where the position of the vertices can be described by some mapping.

The task of finding the asymptotic eigenvalue distribution is motivated by some recent results on the (superlinear) convergence behavior for method of conjugate gradients (CG) \cite{3, 4, 5}: a discretization of (1) for some sequence of stepsizes $h$ tending to zero leads to a sequence of systems of linear equations $A_n x_n = b_n$ with $A_n$ some symmetric positive definite matrix of order $n$, where of course $n$ depends on $h$, and tends to $\infty$ for $h \to 0$. The method of conjugate

\textsuperscript{1}Laboratoire de Mathématiques Paul Painlevé, UMR CNRS 8524, Université des Sciences et Technologies de Lille, F-59655 Villeneuve d’Ascq, France, e-mail: Bernhard.Beckermann@univ-lille1.fr.

\textsuperscript{2}Supported in part by INTAS network NeCCA 03-51-6637, and in part by the Ministry of Science and Technology (MCYT) of Spain and the European Regional Development Fund (ERDF) through the grant BFM2001-3878-C02-02.

\textsuperscript{3}Dipartimento di Fisica e Matematica, Università dell’Insubria, Via Valleggio 11, 22100 Como – Italy; e-mail: stefano.serrac@uninsubria.it, serra@mail.dm.unipi.it

\textsuperscript{4}Supported in part by MIUR grant no. 2002014121.
gradients is a popular method for solving such systems, and its convergence properties have been analyzed by many authors (see e.g. [2, 35]). For instance, one has a simple upper bound for the CG error in energy norm in terms of the spectral condition number of $A_n$, that is, the ratio of the largest divided by the smallest eigenvalue of $A_n$, see, e.g., [19, Eqn. (6.106)]. Both for Finite Difference and Finite Element discretizations, asymptotics for the smallest eigenvalue of $A_n$ in terms of $h$ and the smallest eigenvalue of the differential operator of (1) are known, see for instance [15]. By elementary means one also obtains upper bounds for the largest eigenvalue, and hence upper bounds for the CG error.

However, the (linear) upper bound based on the condition number is usually quite rough, especially in the range of superlinear convergence of CG. This superlinear convergence behavior is observed numerically to be quite pronounced in the context of discretized elliptic problems in $\mathbb{R}^2$ dimensions, in particular for small stepsizes $h$. Here CG convergence is known to be governed by the distribution of the spectrum $\Lambda(A_n)$ of $A_n$, which at least for very simple model problems can be computed explicitly. Roughly speaking, superlinear CG convergence occurs if the eigenvalue distribution of $A_n$ is far from being a worst case eigenvalue distribution. Recently [3, 4, 5], the authors give asymptotic error estimates for CG in terms of the joint asymptotic eigenvalue distribution of $(A_n)_{n \geq 0}$ which provide a theoretical background for the above heuristic observations. The findings in [3, 4, 5] motivated our present work on explicit formulas for the joint asymptotic eigenvalue distribution of sequences of matrices obtained by discretizing (1).

We recall that for a sequence of matrices $(A_n)_{n \geq 0}$, $A_n$ Hermitian of order $n$, with spectrum $\Lambda(A_n) \subset \mathbb{R}$, the measure $\sigma$ is said to describe a joint asymptotic spectrum if for all functions $f \in C_0(\mathbb{R})$, that is, the subset of functions from $C(\mathbb{R})$ having bounded support, there holds

$$\lim_{n \to \infty} \frac{1}{n} \sum_{\lambda \in \Lambda(A_n)} f(\lambda) = \int f(\lambda) \, d\sigma(\lambda), \quad (2)$$

where each eigenvalue is counted according to its multiplicity (and hence $\sigma$ is a probability measure supported on the extended real line $\mathbb{R} = \mathbb{R} \cup \{\pm \infty\}$). It should be observed that often (2) is written in a more informative way as

$$\lim_{n \to \infty} \frac{1}{n} \sum_{\lambda \in \Lambda(A_n)} f(\lambda) = \int_D f(\omega(t)) \frac{dt}{m(D)}, \quad (3)$$

where $D$ is a domain having finite Lebesgue measure in $\mathbb{R}^d$, $d \geq 1$, and where $m(\cdot)$ indicates the standard Lebesgue measure. In that case $\omega$ is also called the symbol of $(A_n)_n$. For instance, in the case of Hermitian Toeplitz matrices, the symbol (see e.g. [34, 31]) coincides with the generating function whose Fourier coefficients determine the entries of any matrix of the sequence while in the case of (symmetric) Finite Difference discretization of PDEs, the symbol includes information on the coefficients and the domain of the PDE and information on the discretization schemes for the derivatives including the used meshes, see Subsection 1.1. We will prove similar results for the Finite Element case.

Notice that, in reference to formulation (2), by compactifying the extended real axis, from any subsequence of $(A_n)_n$ we may extract a subsequence having a joint asymptotic spectrum, but in general there is no joint asymptotic spectrum for the whole sequence $(A_n)_n$. In the present paper, however, the matrices $A_n$ will result from the same discretization process when using different (decreasing) stepsizes. Indeed, when using a Finite Difference discretization for differential operators, explicit formulas for an joint asymptotic spectrum have been given in [18, 32, 27, 21] (one-dimensional setting) and [26, 25, 23, 29] (two-dimensional and multi-dimensional setting). To our knowledge, results for Finite Element discretizations are still lacking (except for brief philosophical observations in [21, 26]).
Before stating our results on stiffness matrices for Finite Elements in Subsection 1.2, we first recall in Subsection 1.1 some known examples of joint asymptotic spectra in the Finite Difference case.

1.1 The case of Finite Difference discretizations

Consider the discretization of the one-dimensional boundary value problem

\[
\begin{cases}
-\frac{d}{dx} \left( k(x) \frac{d}{dx} u(x) \right) = f(x), & x \in (0, 1), \\
u(0), u(1) & \text{given numbers}
\end{cases}
\]

on a uniformly spaced grid using centered Finite Differences of precision order 2 and minimal bandwidth. The resulting linear systems are of tridiagonal type with coefficient matrices \((A_n)_n\) having entries which are weighted samples of \(k\):

\[
A_n = \begin{bmatrix}
  k_{\frac{1}{2}} + k_{\frac{3}{2}} & -k_{\frac{1}{2}} & 0 & \cdots & 0 \\
-k_{\frac{3}{2}} & k_{\frac{1}{2}} + k_{\frac{3}{2}} & -k_{\frac{1}{2}} & \cdots & 0 \\
0 & -k_{\frac{1}{2}} & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & -k_{\frac{2n-1}{2}} & k_{\frac{2n-1}{2}} + k_{\frac{2n+1}{2}}
\end{bmatrix}, 
\]

with \(k_t = k(t \cdot h), h = (n + 1)^{-1}\). When \(k(x) \equiv 1\), the matrix \(A_n\) reduces to the Toeplitz matrix \(T_n(a)\) of size \(n\)

\[
T_n(a) = \begin{bmatrix}
2 & -1 & 0 & \cdots & 0 \\
-1 & 2 & -1 & \cdots & 0 \\
0 & -1 & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & -1 & 2
\end{bmatrix}
\]

generated by \(a(s) = 2 - 2 \cos(s)\): note that the numbers \(-1, 2, -1\) are the (nonzero) Fourier coefficients \(c_1, c_0, c_{-1}\) of \(a\) and represent also the stencil of the Finite Difference formula. This latter statement is not a coincidence: if we change the stencil (for instance in order to obtain more precise discretization schemes), then we obtain Toeplitz matrices generated by a new function \(a\) having Fourier coefficients given by the entries of this new stencil [27]. A well-known fact from the theory of Toeplitz matrices is that \((T_n(a))_n\) has a joint asymptotic spectrum given by \(\omega(s) = a(s)\) with \(D = [-\pi, \pi]\), see for instance the seminal work by Szegö [13]. In the more general case of variable coefficients, it follows from the Locally Toeplitz analysis of [32] that the matrices \(A_n\) of (4) have a joint asymptotic eigenvalue distribution given by the symbol

\[
\omega(x, s) = k(x) a(s)
\]

with \(D = [-\pi, \pi] \times (0, 1)\) (see also [18]). We observe that the result is some sense natural since the samplings of \(k\) move along the diagonals of \(A_n\) smoothly (if \(k\) is smooth) and therefore also the algebraic structure of \(A_n\) looks like a Toeplitz if we restrict the attention to a local portion of the matrix: this nice algebraic behavior has a natural counterpart in the global spectral behavior. As in the constant coefficient case, the change of the discretization scheme, i.e. of the stencil, will change only the function \(a\) in the symbol (compare [27] and [32]). Finally, we observe that the matrices \((A_n)_n\) are essentially of the same type as those which one encounters
when dealing with sequences of orthogonal polynomials with varying coefficients. Here again Locally Toeplitz tools have been used for finding the distribution of the zeros of the considered orthogonal polynomials under very weak assumptions (only measurability) on the regularity of the coefficients [17].

A further variation which could considered in the discretization of the above one-dimensional boundary value problem is the use of non-equispaced grids. Indeed, if the new grid of size \( n \) is obtained as the image under a map \( \phi : [0, 1] \to [0, 1] \) of a uniform grid of the same size \( n \), then the corresponding matrix sequence \( (A_n)_n \) has an asymptotic distribution described by the symbol

\[
\omega(x, s) = \frac{k(\phi(x))}{|\phi'(x)|^2} a(s) \quad \text{with} \quad D = [-\pi, \pi] \times (0, 1).
\]

(6)

For these results, motivated by the use of collocation techniques (see e.g. [16]) for approximating the solution of one-dimensional and multi-dimensional boundary value problems, see [29].

In the case of a two-dimensional problem as (1), the analysis is also quite complete concerning Finite Difference approximations. For instance, when \( \Omega = (0, 1)^2 \), and \( K = I_2 \), using the classical 5 point stencil or the 7 point stencil (in this case there is no difference since \( K_{1,2} = K_{2,1} = 0 \)) we obtain the two-level Toeplitz matrix

\[
T_N(b) = T_{n_1}(a) \otimes I_{n_2} + I_{n_1} \otimes T_{n_2}(a)
\]

(7)

where \( N = (n_1, n_2) \) (\( n_1 \) is the number of internal points in the \( x_1 \) direction and \( n_2 \) is the number of internal points in the \( x_2 \) direction, \( n = n_1n_2 \) is the size, \( b(s_1, s_2) = a(s_1) + a(s_2) \) with \( a(s) = 2 - 2 \cos(s) \). Also in this case the bi-variate stencil represents the non-zero Fourier coefficients of the bi-variate generating function \( b \), and this property remains valid for other stencils. Moreover, according to relation (3), the joint spectrum of \( (T_N(b))_N \) is described by the symbol \( \omega(s_1, s_2) = b(s_1, s_2) \) with \( D = [-\pi, \pi]^2 \) (see e.g. [33]). We observe that the same matrix, with \( n_1 = n_2 = \nu - 1 \), is obtained when employing the \( P_1 \) Finite Element approximation with triangles having the vertices

\[
\left( \frac{(j, k)}{\nu}, \frac{(j + \epsilon, k)}{\nu}, \frac{(j, k + \epsilon)}{\nu} \right), \quad \epsilon = \pm 1.
\]

(8)

More generally, as a consequence of the theory of Generalized Locally Toeplitz sequences presented in [26], asymptotic spectra can be given for Finite Difference approximations of (1) for general functions \( K \) and a domain \( \Omega \) which guarantees the symmetry of the resulting matrix (e.g. a pluri-rectangle that is a connected finite union of rectangles with edges parallel to the main axes, see [30]). For instance, for a seven point stencil (see the proof of Corollary 1.3(b) below) we know that the resulting matrix sequence has a joint asymptotic spectrum with symbol

\[
\omega(s, x) = \left[ \begin{array}{cc} 1 - e^{is_1} & 1 - e^{is_2} \\ 1 - e^{is_2} & 1 - e^{is_2} \end{array} \right] \cdot K(x) \cdot \left[ \begin{array}{cc} 1 - e^{is_1} & 1 - e^{is_1} \\ 1 - e^{is_2} & 1 - e^{is_2} \end{array} \right],
\]

(9)

with \( D = [-\pi, \pi]^2 \times \Omega \). Notice that if \( \Omega = (0, 1)^2 \) and \( K(x) = I_2 \) then the above symbol reduces to the one of (7) since

\[
\left[ \begin{array}{cc} 1 - e^{is_1} & 1 - e^{is_1} \\ 1 - e^{is_2} & 1 - e^{is_2} \end{array} \right] \cdot \left[ \begin{array}{cc} 1 - e^{is_1} & 1 - e^{is_2} \\ 1 - e^{is_2} & 1 - e^{is_2} \end{array} \right] = |1 - e^{is_1}|^2 + |1 - e^{is_2}|^2 = a(s_1) + a(s_2) = b(s).
\]

Furthermore, for non-equispaced tensor grids obtained as the image under a bijective map \( \phi(x) = (\phi_1(x_1), \phi_2(x_2))^T \) of an equispaced tensor grid, the general structure of the symbol
(see [29, 26]) is the natural generalization of (6): denoting by $\nabla \phi$ the (diagonal) Jacobian of $\phi(x) = (\phi_1(x_1), \phi_2(x_2))^T$, we have

$$\omega(s, x) = \begin{bmatrix} 1 - e^{is_1} \\ 1 - e^{is_2} \end{bmatrix} \cdot \tilde{K}(x) \cdot \begin{bmatrix} 1 - e^{is_1} \\ 1 - e^{is_2} \end{bmatrix} \cdot \tilde{K}(x) = \nabla \phi(x)^{-1} K(\phi(x)) \nabla \phi(x)^{-T}$$

(10)

over $D = [-\pi, \pi] \times \tilde{\Omega}$, $\tilde{\Omega} := \phi^{-1}(\Omega)$. We notice that (10) is the natural two-dimensional generalization of (6) and that the symbol in (10) reduces to the one in (9) if $\phi_1(x_1) = x_1$ and $\phi_2(x_2) = x_2$, i.e. in the case where the grids are uniform.

**Remark 1.1.** It should be noted that the Generalized Locally Toeplitz approach allows to derive much more general results. For instance, the coefficient $K_{i,j}$ can be chosen only Riemann integrable and the set $\Omega$ only Peano Jordan measurable (see [26, 14]). The reason for that very weak assumptions can be condensed in the fact that we need only that our domain is approximated in measure by a finite union of rectangles and that our coefficients $K_{i,j}$ can be approximated by a linear combination of characteristic functions of rectangles (the essence of the Peano Jordan measurability and of the Riemann integrability). In this way, our matrix is approximated by a linear combination of matrices which are zero except for a block which is of Toeplitz type and this is the very basic idea in any Locally Toeplitz analysis.

### 1.2 Finite Element examples

Taking into account the results of the previous subsection, the natural question arises whether similar results on the asymptotic spectrum hold for stiffness matrices obtained by applying Finite Elements to (1). We mentioned already before the well-known fact that for the special case $K = I_2$, $\Omega = (0, 1)^2$ and a uniform triangulation on the square such as (8), the stiffness matrix for $P_1$ elements is identical to the one obtained by Finite Differences using a 5 point stencil. However, this connection is no longer true in the general case, and is not sufficient to fully understand the asymptotic properties of stiffness matrices, since for Finite Elements, for instance, a triangulation does not need to be of tensor form.

In order to avoid heavy notation, we will discuss in this paper only the example of a discretization of (1) using $P_1$ Finite Elements. We think however that our techniques carry over to more general settings, along the same lines as in [26] for Finite Differences. In our analysis, we include non-rectangular domains (e.g. L shaped), non-constant coefficients, and non-trivial geometries in the triangulation like grading functions of the form $t \mapsto t^\beta$ for inner angles of size $\pi \beta$, $\beta > 1$ (as the L-shape with $\beta = 3/2$). More specifically, in the following we suppose that we have some $\nu \geq 1$, some open bounded set $\overline{\Omega}$, and a triangulation of $\text{Clos}(\Omega)$ with vertices described by a bijective mapping $\phi : \text{Clos}(\overline{\Omega}) \mapsto \text{Clos}(\Omega)$ of the form

$$(j/\nu, k/\nu)^T \in \text{Clos}(\overline{\Omega}) : \quad P_{j,k} = \phi((j/\nu, k/\nu))$$

(11)

and triangles

$$(P_{j,k}, P_{j+\epsilon,k}, P_{j,k+\epsilon}), \quad \epsilon = \pm 1.$$  

The usual procedure for solving (the variational form of) (1) via $P_1$ Finite Elements (see e.g. [8, 11]) is to consider for $P_{j,k} \in \Omega$ the hat function $\psi_{j,k}$ being linear on each of the triangles, taking the value 1 on the vertex $P_{j,k}$ and 0 on any other vertex (and thus having a support given by the set of the six triangles which share the vertex $P_{j,k}$, see Figure 1), and to solve the system of linear equations

$$A_n x_n = b_n, \quad A_n = \left( \int_{\Omega} \nabla \psi_{j,k}(x) K(x) \nabla \psi_{j',k'}(x)^T dx \right)_{P_{j,k}, P_{j',k'} \in \Omega}$$

(12)
with a suitable right hand side $b_n$ depending on $f$ and $g$. The matrix $A_n$ is usually referred to as the stiffness matrix. Notice that the same matrix of coefficients but a different right hand side is obtained if the Dirichlet boundary conditions are partly replaced by Neumann boundary conditions. In what follows, the letter $n$ will always denote the size of the matrix $A_n$, i.e., the number of vertices in $\Omega$ (which is proportional to $n^2$, compare with (17) below).

**Theorem 1.2.** Consider the above triangulation of $\text{Clos}(\Omega)$ with bijective $\phi : \text{Clos}(\overline{\Omega}) \mapsto \text{Clos}(\Omega)$, where we suppose that $m(\overline{\Omega}) > 0$, and that there exists an “exceptional” compact set $\Gamma \subset \text{Clos}(\Omega)$ with $\partial \Omega \subset \Gamma$ and with Lebesgue measure $m(\Gamma) = 0$ such that $K \circ \phi$ is continuous in $\overline{\Omega} \setminus \Gamma$, and $\phi$ is of class $C^1$ in $\Omega \setminus \Gamma$, with nonsingular Jacobian $\nabla \phi$. Then a joint asymptotic spectrum of the stiffness matrices $A_n$ of (12) for $n \to \infty$ exists, and is given by the formula

$$
\int f \, ds = \frac{1}{(2\pi)^2} \frac{1}{m(\Omega)} \int_{[\pi,\pi]^2} ds \int_{\overline{\Omega}} dx f(\omega(s, x)),
$$

where

$$
\omega(s, x) = \left[ \begin{array}{cc}
1 - e^{is_1} \\
1 - e^{is_2}
\end{array} \right]^{\ast} \cdot \tilde{K}(x) \cdot \left[ \begin{array}{cc}
1 - e^{is_1} \\
1 - e^{is_2}
\end{array} \right], \quad \tilde{K}(x) = |\det \nabla \phi(x)||\nabla \phi(x)^{-1}K(\phi(x))\nabla \phi(x)|^{-T}.
$$

Moreover, this formula for the joint asymptotic spectrum remains valid if one uses numerical integration for evaluating the entries of $A_n$, as long as the quadrature formula has positive weights and integrates constants exactly.

Some consequences of Theorem 1.2 are summarized in the following result.

**Corollary 1.3.** With the notations and assumptions of Theorem 1.2 there holds:

**a)** The matrix of coefficients $A_n$ has the same joint asymptotic spectrum as the one obtained by applying $P_1$ elements on the uniform triangulation (8) to the PDE

$$
-\nabla(K \nabla u) = \tilde{f} \quad \text{on } \Omega, \quad u = \tilde{g} \quad \text{on } \partial \Omega.
$$

Moreover, the bilinear form in the weak formulation of problems (1) and (13) are equivalent via variable transformation.

**b)** One obtains for $(A_n)_n$ the same asymptotic spectrum as the one for matrices obtained by applying Finite Differences based on a seven-point stencil (see Figure 1) to (13). Moreover, $(A_n)_n$ is a Generalized Locally Toeplitz sequence with the symbol $\omega(s, x)$ of Theorem 1.2.
It is quite instructive to compare the results of Theorem 1.2 and Corollary 1.3 with those of Subsection 1.1 for Finite Difference discretizations. We observe that the symbol in formula (10) and the expression of $\omega$ in Theorem 1.2 have a similar structure, in particular we have the same dependency on the domain $\Omega$ and on the matrix-valued coefficient function $K$. Also, the trigonometric polynomials in $s_1, s_2$ occurring in Theorem 1.2 are the same as those in (10). These polynomials translate the dependency of the joint asymptotic spectrum on the discretization scheme (five/seven point stencil or $P_1$ finite elements). The main difference between the two symbols is the dependency on the triangulation described by our function $\phi$: in case of Finite Elements there is an additional factor $|\det \nabla \phi|$, leading to a smoother symbol in neighborhoods of points $x \in \Gamma$ with $|\det \nabla \phi(x)| \neq 0$ (corresponding, e.g., to non-convex edges of $\Omega$, compare with Example 1.5 below), and implying that the FE matrix of coefficients has less eigenvalues of “large” magnitude than the corresponding FD matrix of coefficients. We expect that, as in the Finite Difference setting (see [26]), the results of Theorem 1.2 can be generalized to the cases of higher dimensions, higher order differential operators, Peano Jordan measurable domains, Riemann integrable PDEs coefficients using the same techniques of proof (see also Remark 1.1).

Also, again in analogy with the Finite Difference case, we think that there are similar results for other finite elements, by adapting the choice of the trigonometric polynomials in $s$.

**Remark 1.4.** A typical case covered by Theorem 1.2 is that $\Omega, \Omega$ are polygons, $\Gamma$ consists of a finite number of smooth arcs, and $\phi$ has a $C^1$ extension on the closure of each of the connected components of $\Omega \setminus \Gamma$, up to a finite number of points $a_j \in \Gamma$ being vertices, with $|\nabla \phi|$ remaining bounded while approaching these $a_j$, see for instance Example 1.5 or Example 1.6 below. Under these assumptions, one easily shows that the lengths of all edges of the triangulation of $\Omega$ with parameter $\nu$ are bounded by $\sqrt{2} \sup_{x \in \Omega} |\nabla \phi(x)|/\nu$, in other words, the finess parameter or the largest of the diameters of the triangles of this triangulation is $O(1/\nu)$.

As we will see in Example 1.6 below, our assumptions on $\phi$ do not imply that our family of triangulations of $\Omega$ for varying $\nu$ is quasi-uniform [1, 15]. We recall that a family of triangulations $T_{\nu}$ is called quasi-uniform if the length of the shortest edge in $T_{\nu}$ divided by the finess parameter of $T_{\nu}$ is bounded below by some constant uniformly in $\nu$. We also recall that the family of triangulations $T_{\nu}$ is called shape-regular [7, Definition II.5.1] if, the ratio of the diameter divided by the radius of the largest inscribed disk is bounded uniformly for of each triangle $T \in T_{\nu}$ and all $\nu$ (or, equivalently, if all angles are bounded away from zero uniformly in $\nu$).

We finally should notice that in the proof of Theorem 1.2 we do not need any properties of the finite elements sharing a vertex with $\Gamma$ (or having such a neighbor). Thus Theorem 1.2 remains valid if one uses for instance curved elements in order to fit more complicated boundaries.

The assumption on the Jacobian in Theorem 1.2 implies that $\nabla \phi$ has a norm bounded uniformly in a compact subset $C$ of $\Omega \setminus \Gamma$, and the same is true for $1/\det(\nabla \phi)$. One may show that this implies that the restriction of our triangulation of $\Omega$ to triangles being a subset of $\phi(C)$ will be both quasi-uniform and shape-regular. However, while approaching $\Gamma$, these conditions need no longer hold.

**Example 1.5.** Consider a Finite Element discretization on some non-convex polygon $\Omega$. Suppose that the non-convex edges are given by $a_j, j = 1, \ldots, p$, with inner angles $\beta_j, \pi \in (\pi, 2\pi)$. If $d > 0$ is such that the distance of any two (non-convex) edges is bounded below by $2d$, then a typical grading function verifying the mesh refinement condition $||\phi(x) - a_j|| \approx ||x - a_j||^{\beta_j}$ for $x \to a_j$ could be

$$
\phi(x) = \begin{cases} 
   a_j + (x - a_j) \cdot \left( \frac{||x - a_j||}{d} \right)^{\beta_j - 1} & \text{for } ||x - a_j|| < d, \\
   x & \text{else}.
\end{cases}
$$
Figure 2: Triangulation of an L-shape for $\nu = 12$. On the left we find the uniform triangulation, and on the right its image under the map $\phi(x) = x \cdot \min\{1, \sqrt{||x||}\}$ leading to some grid refinement around the origin.

Its Jacobian for $||x - a_j|| < d$ is given by

$$\nabla \phi(x) = \frac{||x - a_j||^{\beta_j - 1}}{d^{\beta_j - 1}} \left[ I_2 + (\beta_j - 1) \frac{(x - a_j)(x - a_j)^T}{||x - a_j||^2} \right],$$

and $|\det \nabla \phi(x)| = \beta_j (||x - a_j||/d)^{2\beta_j - 2}$ tends to 0 for $x \to a_j$. For the inverse of the normalized Jacobian we find

$$\sqrt{|\det(\nabla \phi(x))|} \nabla \phi(x)^{-1} = \sqrt{\beta_j} \left[ I_2 - \left( \frac{1}{\beta_j} \right) \frac{(x - a_j)(x - a_j)^T}{||x - a_j||^2} \right].$$

Thus clearly $\phi : \text{Clos}(\Omega) \mapsto \text{Clos}(\Omega)$ is a bijective continuous function which is of class $C^1$ in $\{z \in \Omega : ||z - a_j|| \neq d \text{ for } j = 1, 2, \ldots, p\}$.

Example 1.6. A typical example covered by Example 1.5 is a triangulation of an L-shape with vertices $(0, 0), (-1, 0), (-1, 1), (1, 1), (1, -1), (0, -1)$, the only non-convex edge being at the origin $a_1 = 0$, with $\beta_1 : = \beta = 3/2$. Here we can choose $d = 1$ in Example 1.5, leading to the function $\phi(x) = x \cdot \min\{1, ||x||^{3/2 - 1}\}$, with the inverse of the normalized Jacobian given by

$$\sqrt{|\det(\nabla \phi(x))|} \nabla \phi(x)^{-1} = \sqrt{\beta} I_2 - \left( \frac{1}{\sqrt{\beta}} \right) \frac{x x^T}{||x||^2}, \quad ||x|| < 1.$$

In Figure 2 we have drawn both the uniform triangulation and its image under under $\phi$, leading to some graduation around the origin. Since $\nu ||\phi((0, 0)) - \phi((1/\nu, 0))|| = 1/\nu^{3-1}$ is not bounded away from zero for $\nu \to \infty$, we obtain for varying $\nu$ a family of triangulations of $\Omega$ which is not quasi-uniform.

Notice that the triangulation depends of the choice of the Hölder norm: there is for instance a nice geometric interpretation when choosing the 1-norm (and the corresponding family of refined triangulations are known to be shape-regular). However, here we stick to the Euclidean norm leading to a smoother function $\phi$.

The remaining of the paper is organized as follows: in Section 2 we give the proof of Theorem 1.2 and Corollary 1.3 and in Section 3 we discuss relations between the uniform and the non-uniform case, the role of preconditioning strategies, and we draw conclusions.
In what follows we write \( \lambda_1(A_n) \leq \lambda_2(A_n) \leq \cdots \leq \lambda_n(A_n) \) for denoting the eigenvalues of some symmetric matrix \( A_n \) of order \( n \), and \( \mu(A_n) = \frac{1}{n} \sum_{j=1}^{n} \delta_{\lambda_j(A_n)} \) for the corresponding counting measure.

For proving the above result we make use of the following result on Generalized Locally Toeplitz matrix sequences (see [26]) which we will not cite in its greatest generality: we will focus instead on a subclass of matrix sequences that are Generalized Locally Toeplitz and also banded and symmetric. Let \( (M_n) \) be a sequence of matrices of size \( n \) and of level \( \gamma \in \mathbb{N} \) defined according to the following multi-index rule

\[
M_n = (M_{a,a'})_{a,a' \in \nu D \cap \mathbb{Z}^\gamma}, \quad M_{a,a'} = \frac{1}{(2\pi)^\gamma} \int_{[-\pi,\pi]^\gamma} dse^{-s'\omega(s)} \omega(s, a + a'),
\]

and corresponding to some open \( D \subset \mathbb{R}^\gamma \), some integer \( \nu \geq 1 \), and some symbol \( \omega : [-\pi, \pi]^\gamma \times D \in \mathbb{R} \) with \( \omega(s, x) = \omega(-s, x) \) being a polynomial in \( e^{is}, e^{-is} \) with coefficients continuous in \( x \). We observe that a matrix \( M_n \) of such a type and level 1 is just an ordinary banded matrix, where succeeding elements on any diagonal vary only slightly (for large \( n \) and therefore a fortiori for large \( n \)) since they are values of some continuous function at arguments differing only by \( 1/\nu \) (which tends to zero as \( n = n(\nu) \) tends to infinity). Also, a matrix \( M_n \) of level \( \gamma \) is block banded with blocks being themselves of the same structure as in (14) of level \( \gamma - 1 \). Finally, if the symbol \( \omega(s, x) \) does not depend on \( x \) and \( D = \bigotimes_{j=1}^{\gamma} (0, \alpha_j) \), we obtain the classical Toeplitz matrices of level \( \gamma \) and order \( \prod_{j=1}^{\gamma} [\nu \cdot \alpha_j - 1] \). A basic result on such symmetric banded Generalized Locally Toeplitz matrix sequences is that their joint asymptotic spectrum is known [26] and is given by the following formula

\[
\lim_{n \to \infty} \mu(M_n) = \mu, \quad \int f \, dm = \frac{1}{(2\pi)^\gamma} \frac{1}{m(D)} \int_{[-\pi,\pi]^\gamma} ds \int_D dx f(\omega(s, x)).
\]

We will also apply the following statement on the behavior of a joint asymptotic spectrum under perturbations: the idea relies upon the use of some kind of (matrix) approximation theory for reducing the computation of the symbol of a complex matrix sequence to the computation of the symbol of simpler matrix sequences (see [24, 26]).

**Lemma 2.1.** Let \( A_n \in \mathbb{C}^{n \times n} \) being symmetric, and suppose that there exist probability measures \( \sigma, \sigma' \) such that, for each \( \epsilon > 0 \), we may write \( A_n = A'_n + A''_n + A'''_n \) with symmetric matrix sequences \( A'_n := A'_n(\epsilon), A''_n := A''_n(\epsilon), A'''_n := A'''_n(\epsilon) \), where

\[
\limsup_{n \to \infty} ||A'_n|| < \epsilon, \quad \limsup_{n \to \infty} \frac{\text{rank}(A'''_n)}{n} < \epsilon,
\]

and \( (A'_n)_n \) having a joint asymptotic spectrum \( \mu \leq \epsilon \sigma' + \sigma \). Then \( (A_n)_n \) has the joint asymptotic spectrum \( \sigma \).

**Proof.** Suppose that the assertion of the Lemma is not true. Then by Helley’s Theorem [20, Theorem 0.1.3] there exists an infinite set of natural numbers \( \mathcal{N} \) such that \( (\mu(A_n))_{n \in \mathcal{N}} \) tends to some probability measure \( \nu \) different from the probability measure \( \sigma \). By possibly replacing \( A_n \) by \( -A_n \) we may conclude that there exists a \( b \in \mathbb{R} \) with

\[
\nu([-\infty, b)) > \sigma([-\infty, b)) = \sigma([-\infty, b]).
\]
Write $r_n = \text{rank}(A_n'')$. Any $V \subset \mathbb{C}^n$ can be written as direct sum $V' \oplus V''$, $V'$ being a subset of the kernel of $A_n''$, $V''$ being therefore a subset of the image of $(A_n'')^* = A_n''$, implying that $\dim(V') \geq \dim(V) - r_n$. Consequently, using the Courant min-max principle we obtain for any $1 \leq j \leq n - r_n$ \[
abla_j(A_n') = \max_{V \subset \mathbb{C}^n, \dim(V) = n+1-j} \min_{y \in V} \frac{y^*A_n'y}{y^*y} \leq \max_{V' \subset \text{Ker}(A_n''), \dim(V') \geq n+1-j-r_n} \min_{y \in V'} \frac{y^*(A_n' + A_n'')y}{y^*y} + ||A_n''|| \leq \max_{V' \subset \mathbb{C}^n, \dim(V') \geq n+1-j-r_n} \min_{y \in V'} \frac{y^*A_n'y}{y^*y} + ||A_n''|| = \lambda_j + r_n(A_n) + ||A_n''||. \]

Taking into account [20, Theorem 0.1.4], we conclude that \[
\nu([-\infty, b]) \leq \limsup_{n \to \infty} \mu(A_n)((-\infty, b]) = \limsup_{n \to \infty} \frac{\#\{j : \lambda_j(A_n) \leq b\}}{n} \leq \limsup_{n \to \infty} r_n + \frac{\#\{j > r_n : \lambda_j(A_n) \leq b + ||A_n''||\}}{n} \leq \epsilon + \limsup_{n \to \infty} \mu(A_n)((-\infty, b + 2\epsilon]) \leq \epsilon + \sigma([-\infty, b + 2\epsilon]). \]

For $\epsilon \to 0$, we are left with $\nu([-\infty, b]) \leq \sigma([-\infty, b])$, in contradiction with (16). Hence the Lemma is shown. \qed

The above lemma is essentially contained in original work by Tilli on (one-level) Locally Toeplitz sequences [32] and can be considered an evolution of the low-rank, low-norm splittings used by Tyrtysnikov [33]. A form which is closer to the present approach can be found in [26] where the main role is played by the symbols of the involved matrix sequences. However, in the present version the language and the tools of Lemma 2.1 are a bit different since the results are expressed in terms of measures (recall formulation (2)) rather than symbols (recall formulation (3)).

Proof of Theorem 1.2: We start by establishing the formula
\[
\lim_{\nu \to \infty} \frac{n(\nu)}{\nu^2} = m(\Omega), \quad \text{where} \quad n(n(\nu)) = \# \left\{ \frac{(j,k)}{\nu} \in \tilde{\Omega} \right\}
\]

is the size of the stiffness matrix (12) for the triangulation with parameter $\nu$. For $d > 0$, denote by $\Gamma_d := \{ y \in \mathbb{R}^2 : \text{dist}(y, \Gamma) \leq d \}$ the closed $d$–neighborhood of $\Gamma$, where we recall that $\partial \tilde{\Omega} \subset \Gamma$ by assumption on $\Gamma$. For any $\frac{(j,k)}{\nu} \in \tilde{\Omega}$ we find an open square of Lebesgue measure $1/\nu^2$ being a subset of the $(2/\nu)$-neighborhood of $\tilde{\Omega}$, any two of such squares having an empty intersection, and thus $n(\nu)/\nu^2 \leq m(\Omega \cup \Gamma_{2/\nu})$. On the other hand, the set $\Omega \setminus \Gamma_{2/\nu}$ is a subset of the union of closed squares of Lebesgue measure $1/\nu^2$ centered at $\frac{(j,k)}{\nu} \in \tilde{\Omega}$, implying that $n(\nu)/\nu^2 \geq m(\Omega \setminus \Gamma_{2/\nu})$. Taking into account that $m(\Gamma_d) \to m(\Gamma) = 0$ for $d \to 0$ by assumption of Theorem 1.2, we arrive at relation (17).

Let $\epsilon > 0$. We now choose suitable subsets of $\tilde{\Omega}$. Let $d > 0$ with $m(\tilde{\Omega} \setminus \Gamma_{3d}) > (1 - \frac{\epsilon}{2}) m(\tilde{\Omega})$. By compactness of $\Gamma$, we may cover $\Gamma$ with a finite number of open $\infty$–neighborhoods $U_d(x_j) = \{ y \in \mathbb{R}^2 : ||y - x_j||_\infty < d \}$, $j = 1, \ldots, p$, with $x_j \in \Gamma$. Defining the pluri-rectangles\[
\tilde{\Omega} := \tilde{\Omega} \setminus \bigcup_{j=1}^p \text{Clos} \left( U_{2d}(x_j) \right), \quad \tilde{\Omega}'' := \tilde{\Omega} \setminus \bigcup_{j=1}^p U_d(x_j)
\]
we find that $\Omega' \cap 3d \subset \Omega' \subset \Omega' \cap \Gamma$, with $\Omega''$ being compact, $\Omega'$ being open, and
\[
\lim_{\nu \to \infty} \frac{n'(\nu)}{\nu^2} = m(\Omega') \geq \left( 1 - \frac{\epsilon}{3} \right) m(\Omega), \quad \text{where} \quad n' = n'(\nu) = \# \left\{ (j, k) \in \Omega' \right\}.
\] (18)

Thus, for sufficiently large $\nu$,
\[
\frac{n'(\nu)}{n(\nu)} > 1 - \frac{\epsilon}{2}.
\] (19)

We are now prepared to introduce a suitable splitting of the stiffness matrix $A_n$ of (12): we first apply a suitable simultaneous permutation of row and columns such that the first $n'(\nu)$ rows and columns of $A_n$ correspond to indices with $(j, k) / \nu \in \Omega'$. Then the matrix $A_n^{\prime\prime}$ defined by
\[
A_n - A_n^{\prime\prime} = \begin{bmatrix}
\tilde{A}_n & 0 \\
0 & 0
\end{bmatrix}, \quad \tilde{A}_n = \left( \int_{\Omega} \nabla \psi_{j,k}(x) \nabla \psi_{j',k'}(x)^T \, dx \right)_{(j,k) / \nu, (j',k') / \nu \in \Omega'}
\]
is symmetric, and has a rank bounded above by twice the difference of the order $n = n(\nu)$ of $A_n$ minus the order $n' = n'(\nu)$ of $A_n$. A combination with (19) leads to the relations
\[
(A_n^{\prime\prime})^* = A_n^{\prime\prime}, \quad \text{rank}(A_n^{\prime\prime}) \leq \epsilon n.
\] (20)

We want to apply Lemma 2.1 via a splitting $\tilde{A}_n = \tilde{A}_n' + \tilde{A}_n^{\prime\prime}$, and
\[
A_n = A_n' + A_n' + A_n^{\prime\prime}, \quad A_n' = \begin{bmatrix}
\tilde{A}_n' & 0 \\
0 & 0
\end{bmatrix}, \quad A_n^{\prime\prime} = \begin{bmatrix}
\tilde{A}_n^{\prime\prime} & 0 \\
0 & 0
\end{bmatrix},
\] (21)
where $\tilde{A}_n^{\prime\prime}$ will be a symmetric matrix of small norm, and $\tilde{A}_n'$ symmetric and banded. Moreover, $(\tilde{A}_n')$ will be Generalized Locally Toeplitz of level 2 in the sense of (14), and thus we know the existence and the explicit form of the joint asymptotic spectrum of $\tilde{A}_n'$ for $\nu \to \infty$.

We make use of the classical assembling procedure of a $P_1$ Finite Element matrix $A_n$: starting from the zero matrix, the stiffness matrix $A_n$ is obtained after applying for all triangles $T$ of the form $(P_{j,k}, P_{j+\eta,k}, P_{j,k+\eta})$, $\eta = \pm 1$, the updating formula
\[
A_n(\begin{array}{c}
(j, k), (j + \eta, k), (j, k + \eta) \\
(j, k), (j + \eta, k), (j, k + \eta)
\end{array}) \leftarrow A_n(\begin{array}{c}
(j, k), (j + \eta, k), (j, k + \eta) \\
(j, k), (j + \eta, k), (j, k + \eta)
\end{array}) + \frac{1}{2|\det(C^{-1})|} B^T C^{-1} \int_T K(x) \, dx \, C^{-T} B,
\] (22)
where the linear application $x \mapsto P_{j,k} + Cx$ maps the points $(0, 0), (1, 0), (0, 1)$ to $P_{j,k}, P_{j+\eta,k}, P_{j,k+\eta}$, respectively, and
\[
B = \begin{bmatrix}
-1 & 1 & 0 \\
-1 & 0 & 1
\end{bmatrix}.
\]

An important observation in our proof is that the updating term in (22) behaves like $\frac{1}{2} B^T \tilde{K}(\zeta) B$ for some $\zeta \in \phi^{-1}(T)$ for “most” triangles $T$. In order to make this claim more precise in (23) below, we notice that, by construction, $\Omega''$ is a compact subset of $\Omega' \cap \Gamma$, and hence the Jacobian $\nabla \phi$ of $\phi$, its inverse $\nabla \phi(x)^{-1}$ and the function $K \circ \phi$ are uniformly continuous in $\Omega''$. Let
\[
M := \sup_{x \in \Omega''} \max \left\{ ||\nabla \phi(x)||, ||\nabla \phi(x)^{-1}||, ||K(\phi(x))||, \sqrt{2}\epsilon \right\} \geq 1,
\]
and choose $\nu$ sufficiently large such that a triangle having at least one vertex in $\Omega' \cap \Omega''$ is a subset of $\Omega''$, and that any of the above functions varies at most by $\epsilon / (4M^2)$ by choosing two arguments
Table 1: The six adjacent vertices of \((j,k) \in \tilde{\Omega}^\prime\) and the corresponding off-diagonal entries of \(\tilde{A}^\prime_n\): in the first column we find the index \((j',k')\) of an adjacent vertex, in the second and third column the index of the third vertex of the two triangles giving a non-trivial contribution to the entry in row \((j,k)\) and column \((j',k')\) of \(A_n\), and in the last column the entry of \(\tilde{A}^\prime_n\) at the same position.

<table>
<thead>
<tr>
<th>((j',k'))</th>
<th>((j'',k''))</th>
<th>((j'',k'''))</th>
<th>Corresponding entry of (\tilde{A}^\prime_n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>((j-1,k))</td>
<td>((j,k-1))</td>
<td>((j-1,k+1))</td>
<td>(B_1^T \tilde{K} \left( \frac{(j'+i,k'+k')}{{2}} \right) B_2)</td>
</tr>
<tr>
<td>((j,k-1))</td>
<td>((j+1,k-1))</td>
<td>((j-1,k))</td>
<td>(B_1^T \tilde{K} \left( \frac{(j+j',k+k')}{{2}} \right) B_3)</td>
</tr>
<tr>
<td>((j+1,k-1))</td>
<td>((j+1,k))</td>
<td>((j,k-1))</td>
<td>(B_2^T \tilde{K} \left( \frac{(j+j',k+k')}{{2}} \right) B_3)</td>
</tr>
<tr>
<td>((j+1,k))</td>
<td>((j,k+1))</td>
<td>((j+1,k-1))</td>
<td>(B_1^T \tilde{K} \left( \frac{(j+j',k+k')}{{2}} \right) B_2)</td>
</tr>
<tr>
<td>((j,k+1))</td>
<td>((j-1,k+1))</td>
<td>((j+1,k))</td>
<td>(B_2^T \tilde{K} \left( \frac{(j+j',k+k')}{{2}} \right) B_3)</td>
</tr>
<tr>
<td>((j-1,k+1))</td>
<td>((j,k-1))</td>
<td>((j,k+1))</td>
<td>(B_2^T \tilde{K} \left( \frac{(j+j',k+k')}{{2}} \right) B_3)</td>
</tr>
</tbody>
</table>

in any triangle being a subset of \(\tilde{\Omega}^\prime\). For the matrix \(\tilde{A}_n\) we only need to consider triangles \(T\) having at least one vertex with pre-image in \(\tilde{\Omega}^\prime\). Denoting by \(\tilde{T} \subset \tilde{\Omega}^\prime\) the corresponding triangle with vertices \(\frac{(j,k)}{{\nu}}, \frac{(j+n,k)}{{\nu}}, \frac{(j+k)}{{\nu}}\), we may conclude with help of the mean value theorem that, for any \(\zeta \in \tilde{T}\),

\[
\left| \frac{1}{\int_{\tilde{T}} K(x) \, dx} K(\phi(\zeta)) \right| \leq \frac{\epsilon}{4M^5} \leq M, \quad \left| \frac{\nu}{\eta} C - \nabla \phi(\zeta) \right| \leq \frac{\epsilon}{M^5} \leq \frac{1}{2\|\nabla \phi(\zeta)\|^2}.
\]

and hence

\[
\left| \left( \frac{\nu}{\eta} C \right)^{-1} - \nabla \phi(\zeta)^{-1} \right| \leq \frac{2\epsilon}{M^3} \leq M, \quad \left| \det \left( \frac{\nu}{\eta} C \right) - \det(\nabla \phi(\zeta)) \right| \leq \frac{4\epsilon}{M^2} \leq M.
\]

Applying several times the triangular inequality, we obtain after some elementary computations the (quite rough) estimate

\[\max_{\zeta \in \tilde{T}} \left| \frac{1}{\det(C^{-1})} C^{-1} \int_{\tilde{T}} K(x) \, dx - \tilde{K}(\zeta) \right| \leq 80\epsilon, \quad (23)\]

with \(\tilde{K}\) as in the statement of Theorem 1.2. We remark that the same conclusion holds if instead of exact integration one uses a quadrature formula with positive weights for the entries of the stiffness matrix, provided that this quadrature formula integrates constants exactly.

Notice that, in the updating procedure (22), an offdiagonal entry of \(A_n\) is updated twice since a fixed edge of the triangulation is adjacent to two triangles, and a diagonal entry is updated six times since there are 6 triangles adjacent to a vertex, compare with Figure 1. More precisely, in row labeled \((j,k)\), the matrix \(\tilde{A}_n\) has non-zero offdiagonal entries in columns labeled

\[
(j',k') \in \{(j-1,k+1), (j,k+1), (j-1,k), (j+1,k), (j,k-1), (j+1,k-1)\},
\]

i.e., the indices of adjacent vertices. For instance, for the entry in column \((j',k') = (j-1,k)\) we have to consider the two triangles \(T\) with third vertex labeled \((j'',k'') = (j,k-1)\) and \((j'',k''') = (j-1,k+1)\), respectively, and the corresponding updating quantities can be found at position \((1,2)\) and \((2,1)\), respectively, of the symmetric \(3 \times 3\) updating matrix on the right of (22). Thus, defining the corresponding offdiagonal entry of \(\tilde{A}_n\) by

\[
\tilde{A}_n \left( \frac{(j',k')}{(j,k)} \right) = B_1^T \tilde{K} \left( \frac{1}{2} \left( \frac{(j,k)}{{\nu}} + \frac{(j',k')}{\nu} \right) \right) B_2 = (-1, -1) \tilde{K} \left( \frac{(j+j',k+k')}{2\nu} \right) (1,0)^T,
\]

12
$B_\ell$ denoting the $\ell$th column of $B$, we find according to (23) that

$$\left| \tilde{A}_n'(\frac{(j', k')}{(j, k)}) - \tilde{A}_n(\frac{(j', k')}{(j, k)}) \right| \leq 80\epsilon \|B\|^2 = 240\epsilon.$$ 

The off-diagonal entries of $\tilde{A}_n'$ for the other five adjacent vertices $(j', k')$ of $(j, k)$ are given in Table 1, and each time we obtain the same estimate for the off-diagonal entries of $\tilde{A}_n' - \tilde{A}_n$. We define the diagonal entries of $\tilde{A}_n'$ by

$$\tilde{A}_n'(\frac{(j, k)}{(j, k)}) = \text{trace}(B^T \tilde{K} \left( \frac{(j, k)}{\nu} \right) B) = -2 \left( B_1^T \tilde{K} \left( \frac{(j, k)}{\nu} \right) B_2 + B_1^T \tilde{K} \left( \frac{(j, k)}{\nu} \right) B_3 + B_2^T \tilde{K} \left( \frac{(j, k)}{\nu} \right) B_3 \right),$$

and find according to (23) that

$$\left| \tilde{A}_n'(\frac{(j, k)}{(j, k)}) - \tilde{A}_n(\frac{(j, k)}{(j, k)}) \right| \leq 240\epsilon \|B\|^2 = 720\epsilon,$$

and thus, by (21),

$$\|\tilde{A}_n'\| = \|\tilde{A}_n - \tilde{A}_n'\| \leq \sqrt{\|\tilde{A}_n - \tilde{A}_n'\|_1 \|\tilde{A}_n - \tilde{A}_n'\|_\infty} \leq (6 \cdot 240 + 720)\epsilon = 2160\epsilon.$$

It remains to analyze $\tilde{A}_n'$. Comparing the definition (14) with the last column of Table 1 and with (24), we see that $(\tilde{A}_n')$ is a banded and symmetric Generalized Locally Toeplitz matrix sequence of level 2 corresponding to the domain $\Omega'$ and the symbol

$$\omega(s, x) = \begin{cases} 
(2 \cos(s_1) - 2) B_1^T \tilde{K}(x)B_2 + (2 \cos(s_2) - 2) B_1^T \tilde{K}(x)B_3 \\
+ (2 \cos(s_2 - s_1) - 2) B_2^T \tilde{K}(x)B_3 \\
4 \sin^2 \left( \frac{s_1}{2} \right) \tilde{K}_{1,1}(x) + 4 \sin^2 \left( \frac{s_2}{2} \right) \tilde{K}_{2,2}(x) \\
+ 4 \left[ \sin^2 \left( \frac{s_1}{2} \right) + \sin^2 \left( \frac{s_2}{2} \right) - \sin^2 \left( \frac{s_2 - s_1}{2} \right) \right] \tilde{K}_{1,2}(x),
\end{cases}$$

that is, the same symbol (but a different domain) as in the statement of Theorem 1.2. Using (15), we may conclude that $(\mu(\tilde{A}_n'))$ has the limit $\tilde{\sigma}$, with

$$\int f \, d\tilde{\sigma} = \frac{1}{(2\pi)^2 m(\Omega')} \int_{[-\pi, \pi]^2} ds \int_{\Omega'} dx \, f(\omega(s, x)).$$

According to (21), for the corresponding counting measures for $\nu \to \infty$, we get using (17), (18),

$$\mu(\tilde{A}_n') = \frac{n(\nu') - n'(\nu)}{n(\nu)} \cdot \delta_0 + \frac{n'(\nu)}{n(\nu)} \mu(\tilde{A}_n') \to \frac{m(\tilde{\Omega}) - m(\tilde{\Omega}')}{m(\Omega)} \cdot \delta_0 + \frac{m(\tilde{\Omega}')}{m(\Omega)} \tilde{\sigma},$$

and

$$\frac{m(\tilde{\Omega}) - m(\tilde{\Omega}')}{m(\Omega)} \cdot \delta_0 + \frac{m(\tilde{\Omega}')}{m(\Omega)} \tilde{\sigma} \leq \epsilon \cdot \delta_0 + \frac{m(\tilde{\Omega}')}{m(\Omega)} \tilde{\sigma} \leq \epsilon \cdot \delta_0 + \sigma,$$

since $\tilde{\sigma}$ differs from $\sigma$ by using a different normalization and a smaller set of integration $\tilde{\Omega}' \subset \tilde{\Omega}$. Thus we may apply Lemma 2.1, giving the joint asymptotic spectrum for $(A_n)$ as claimed in Theorem 1.2. \(\Box\)
Proof of Corollary 1.3: The first sentence of part (a) follows immediately by applying twice the formulas of Theorem 1.2. With respect to the second one, consider the variable transformation \( x = \phi(x) \) in (1); with \( \tilde{f}(\tilde{x}) = f(\phi(x)) \), we have \( \nabla \tilde{f}(\tilde{x}) = (\nabla f)(\phi(x)) \nabla \phi(\tilde{x}) \), and hence
\[
\int_{\Omega} (\nabla u)(x) K(x) (\nabla v)(x) \, dx = \int_{\Omega} (\nabla \tilde{u}(\tilde{x})) \nabla \phi(\tilde{x})^{-1} K(\phi(x)) \nabla \phi(\tilde{x})^{-T} (\nabla \tilde{v})(\tilde{x}) T \det \nabla \phi(\tilde{x}) |d\tilde{x}|
\]
\[
= \int_{\Omega} (\nabla \tilde{u}(\tilde{x})) \tilde{K}(\tilde{x})(\nabla \tilde{v})(\tilde{x})^{T} \, d\tilde{x}.
\]
For a proof of part (b), we consider
\[
y_{\nu} = (u_{j,k})_{(j,k)\in\tilde{\Omega}} \quad \tilde{u}_{j,k} \approx u \left( \frac{(j,k)}{\nu} \right)
\]
and the second order central difference operators using the seven point stencil of Figure 1
\[
\Delta_1 u_{j,k} = u_{j+1/2,k} - u_{j-1/2,k} \approx \frac{1}{\nu} \frac{\partial}{\partial \tilde{x}_1} u \left( \frac{(j,k)}{\nu} \right),
\]
\[
\Delta_2 u_{j,k} = u_{j,k+1/2} - u_{j,k-1/2} \approx \frac{1}{\nu} \frac{\partial}{\partial \tilde{x}_2} u \left( \frac{(j,k)}{\nu} \right),
\]
\[
\Delta_3 u_{j,k} = u_{j+1/2,k-1/2} - u_{j-1/2,k+1/2} \approx \frac{1}{\nu} \left( \frac{\partial}{\partial \tilde{x}_1} - \frac{\partial}{\partial \tilde{x}_2} \right) u \left( \frac{(j,k)}{\nu} \right).
\]
Let \( \tilde{\Omega} \) and \( \tilde{A}_n \) be as in the preceding proof, and let \( C_n \) be obtained from the matrix \( \tilde{A}_n \) by replacing the diagonal entries (24) by
\[
C_n \left( \frac{(j,k)}{\nu} \right) = -\tilde{B}_1 \left( \tilde{K} \left( \frac{2j - 1, 2k}{2\nu} \right) \right) + \tilde{K} \left( \frac{2j + 1, 2k}{2\nu} \right) \left( \frac{(j,k)}{\nu} \right),
\]
and hence \( ||\tilde{A}_n - C_n|| \) is of order \( \epsilon \), compare with (23). For a grid point \( \frac{(j,k)}{\nu} \in \tilde{\Omega} \) having all its adjacent vertices in \( \Omega' \), the component of \( C_n y_{\nu} \) with index \( (j,k) \) can be written as
\[
\left[ \tilde{K}_{1,1} + \tilde{K}_{1,2} \right] \left( \frac{2j - 1, 2k}{2\nu} \right) (u_{j,k} - u_{j-1,k}) + \left[ \tilde{K}_{1,1} + \tilde{K}_{1,2} \right] \left( \frac{2j + 1, 2k}{2\nu} \right) (u_{j,k} - u_{j+1,k})
\]
\[
+ \left[ \tilde{K}_{2,2} + \tilde{K}_{1,2} \right] \left( \frac{2j + 2k - 1}{2\nu} \right) (u_{j,k} - u_{j,k-1}) + \left[ \tilde{K}_{2,2} + \tilde{K}_{1,1} \right] \left( \frac{2j + 2k + 1}{2\nu} \right) (u_{j,k} - u_{j,k+1})
\]
\[
+ \tilde{K}_{1,2} \left( \frac{2j + 1, 2k - 1}{2\nu} \right) (u_{j+1,k-1} - u_{j,k}) + \tilde{K}_{1,2} \left( \frac{2j - 1, 2k + 1}{2\nu} \right) (u_{j-1,k+1} - u_{j,k})
\]
\[
= -\Delta_1 \left[ \tilde{K}_{1,1} + \tilde{K}_{1,2} \right] \Delta_1 u_{j,k} - \Delta_2 \left[ \tilde{K}_{2,2} + \tilde{K}_{1,2} \right] \Delta_2 u_{j,k} + \Delta_3 \tilde{K}_{1,2} \Delta_3 u_{j,k}.
\]
If some of the vertices \( \frac{(j',k')}{\nu} \) adjacent to \( \frac{(j,k)}{\nu} \) lie outside of \( \Omega' \), we get a similar expression, where the corresponding values \( u_{j',k'} \) have to be dropped. Therefore the matrix \( C_n \) describes a Finite Difference discretization in \( \tilde{\Omega} \) based on the seven point stencil of Figure 1 for the differential expression
\[
-\frac{\partial}{\partial \tilde{x}_1} \left( \frac{\tilde{K}_{1,1} + \tilde{K}_{1,2}}{\partial \tilde{x}_1} \frac{\partial u}{\partial \tilde{x}_1} \right) - \frac{\partial}{\partial \tilde{x}_2} \left( \frac{\tilde{K}_{2,2} + \tilde{K}_{1,2}}{\partial \tilde{x}_2} \frac{\partial u}{\partial \tilde{x}_2} \right) + \left( \frac{\partial}{\partial \tilde{x}_1} - \frac{\partial}{\partial \tilde{x}_2} \right) \left( \frac{\partial u}{\partial \tilde{x}_1} - \frac{\partial u}{\partial \tilde{x}_2} \right)
\]
3 Uniform versus non-uniform triangulations and preconditioning

In the previous sections we have considered a triangulation \( \mathcal{T}_\nu \) of \( \Omega \) being the image under a bijective map \( \phi \) of a uniform triangulation \( \overline{\mathcal{T}_\nu} \) of \( \overline{\Omega} \) with stepsize \( 1/\nu \). Denote by \( A_n(K, \mathcal{T}_\nu) \) the corresponding stiffness matrix (12). Since in general the two triangulations \( \mathcal{T}_\nu \) and \( \overline{\mathcal{T}_\nu} \) lead to stiffness matrices of the same size, we want to discuss in this section in more detail some spectral properties of the matrix \( A_n(I_2, \overline{\mathcal{T}_\nu})^{-1} A_n(K, \mathcal{T}_\nu) \) and other related matrices, being motivated by the task of finding efficient preconditioning strategies for the method of conjugate gradients applied to the stiffness matrix \( A_n(K, \mathcal{T}_\nu) \). Notice that our uniform triangulation \( (\overline{\mathcal{T}_\nu})_\nu \) is trivially both quasi-uniform and shape-regular (see Remark 1.4), but not necessarily \( (\mathcal{T}_\nu)_\nu \). We will be in particular interested in the case where \( (\mathcal{T}_\nu)_\nu \) is only shape-regular.

The main results of this Section are given in Subsection 3.2: in Theorem 3.2 we first relate two stiffness matrices with respect to the partial ordering of Hermitian matrices (\( M_1 \preceq M_2 \) if \( M_1, M_2 \) are Hermitian and \( M_2 - M_1 \) is semi positive definite). Subsequently, in Corollary 3.4 we deduce bounds for the smallest and the largest eigenvalue of such preconditioned stiffness matrices, and, in Theorem 3.5, we give results on the joint asymptotic spectrum for such matrices. But first we provide in Subsection 3.1 a basic proposition (based on the local analysis of Finite Element matrices) which is the keystone for proving the results of Subsection 3.2.

3.1 Local domain analysis of the Finite Element matrices

In order to understand better the local properties of a stiffness matrix, let us go back to the classical assembling procedure of a \( P_1 \) Finite Element matrix \( A_n \) mentioned already in the proof of Theorem 1.2. Starting from the zero matrix we have the following updating formulas: any two stiffness matrices with respect to the partial ordering of Hermitian matrices (\( T \triangleq \) classical assembling procedure of a Finite Element matrices, and, in Theorem 3.5, we give results on the joint asymptotic spectrum for such matrices. It follows. The second assertion now is a simple consequence of the above relationship between \( A_n \) and the 7 point stencil Finite Difference matrix and of the fact that every Finite Difference discretization of second order PDEs leads to Generalized Locally Toeplitz sequences (see [26]).

\[
\mathbf{3.1 \ Local \ domain \ analysis \ of \ the \ Finite \ Element \ matrices}
\]

In order to understand better the local properties of the elements of a finite element stiffness matrix, let us go back to the classical assembling procedure of a \( P_1 \) finite element matrix \( A_n \) mentioned already in the proof of Theorem 1.2. Starting from the zero matrix we have the following updating formulas: any triangle \( T \) of the form \( (P_{j,k}, P_{j+k,n}, P_{j+k+n}) \), \( \eta \in \{\pm 1\} \), gives the contribution

\[
\begin{align*}
A_n((j, k), (j + \eta, k), (j + k, \eta) &\sim (j, k), (j + \eta, k), (j, k + \eta)) \\
&\sim \frac{1}{2|\det(C^{-1})|} B^T C^{-1} \int_T K(x) dx B C^{-T} B,
\end{align*}
\]

where the linear application \( x \mapsto P_{j,k} + Cx \) maps the points \((0, 0), (1, 0), (0, 1)\) to \( P_{j,k}, P_{j+k,n}, P_{j+k+n} \), respectively, and

\[
B = \begin{bmatrix}
-1 & 1 & 0 \\
-1 & 0 & 1
\end{bmatrix}.
\]

Suppose that the three points \( (P_{j,k}, P_{j+k,n}, P_{j+k+n}) \) have positive orientation, and define by \( \alpha, \beta, \gamma \), respectively, the angles of the triangle \( T \) at these vertices. In addition, define \( \Pi \) to be a rotation matrix mapping the half line \((0, P_{j+k,n} - P_{j,k})\) to the half line \((0, 0), (1, 0))\), then

\[
\Pi C = \frac{||P_{j+k,n} - P_{j,k}||}{\sin(\alpha)} \begin{bmatrix}
\sin(\gamma) & \sin(\beta) \cos(\alpha) \\
0 & \sin(\beta) \sin(\alpha)
\end{bmatrix},
\]

15
and in addition

\[
\frac{C^{-1}}{\sqrt{|\det(C^{-1})|}} = \frac{1}{\sqrt{\sin(\alpha) \sin(\beta) \sin(\gamma)}} \begin{bmatrix}
\sin(\alpha) \sin(\beta) & -\cos(\alpha) \sin(\beta) \\
0 & \sin(\gamma)
\end{bmatrix}.
\]

Observe also that \(C^{-1}/\sqrt{|\det(C^{-1})|}\) has has the singular values \(\sqrt{\delta_T}\) and \(1/\sqrt{\delta_T}\) and thus a spectral condition number \(\delta_T\) which can be computed explicitly in terms of the angles of \(T\)

\[
\delta_T := \text{cond} \left( \frac{C^{-1}}{\sqrt{|\det(C^{-1})|}} \right) = y_T + \sqrt{y_T^2 - 1}, \quad y_T = \frac{\sin^2(\beta) + \sin^2(\gamma)}{2\sin(\alpha) \sin(\beta) \sin(\gamma)}.
\]

Therefore

\[
\frac{1}{\delta_T} I_2 \leq \frac{1}{|\det(C^{-1})|} C^{-T} C - T \leq \delta_T I_2.
\]

If the three points \((P_{j,k}, P_{j+\eta,k}, P_{j,k+\eta})\) have negative orientation, then we switch axes, that is, we exchange the role of \(\beta\) and \(\gamma\), but the conclusions in (26) and (27) are the same. For instance, for a triangle \(T \in T_{\nu}\) of a uniform triangulation we get \(\alpha = \pi/2\) and \(\beta = \gamma = \pi/4\), leading to \(\delta_T = 1\), but in general \(\delta_T \geq 1\).

The relation (27) enables us to compare the updating matrices in (25) for different meshes and \(K = I_2\), and, by a similar argument, for different (pointwise symmetric positive definite) coefficient functions \(K\).

**Proposition 3.1.** With

\[
\kappa_{\min} = \text{essinf}_{x \in T} \lambda_{\min}(K(x)) \geq 0, \quad \kappa_{\max} = \text{esssup}_{x \in T} \lambda_{\max}(K(x)),
\]

and \(B, C\) as in (25) we have that

\[
\kappa_{\min} B^T C^{-1} C^{-T} B \leq \frac{1}{2|\det(C^{-1})|} B^T C^{-1} \int_T K(x) \frac{dx}{C^{-T}} B \leq \kappa_{\max} \frac{B^T C^{-1} C^{-T} B}{2|\det(C^{-1})|},
\]

and, with \(\delta_T \geq 1\) as in (26),

\[
\frac{1}{\delta_T} \frac{B^T B}{2} \leq \frac{B^T C^{-1} C^{-T} B}{2|\det(C^{-1})|} \leq \delta_T \frac{B^T B}{2}.
\]

There are many ways of writing the constant \(\delta_T\) of (26). For instance, if \(\beta, \gamma \in (0, \pi/2)\), we find using the relation \(\alpha + \beta + \gamma = \pi\) that

\[
y_T = \frac{\sin^2(\beta) + \sin^2(\gamma)}{\sin^2(\beta) \sin(2\gamma) + \sin^2(\gamma) \sin(2\beta)} \leq \frac{1}{\min(\sin(2\beta), \sin(2\gamma))},
\]

which is quite precise if \(\beta\) or \(\gamma\) is small compared to the other two angles. We also have that \(\delta_T\) is uniformly bounded for \(T \in T_{\nu}\) for all \(\nu\) if and only if all angles occurring in \(T_{\nu}\) are bounded away from zero uniformly in \(\nu\), i.e., \((T_{\nu})_\nu\) is shape-regular. Moreover, there holds

\[
\delta_T \leq 2y_T = \frac{b^2 + c^2}{2m(T)} \leq \frac{a + b + c}{2m(T)} \max\{a, b, c\}
\]

the expression on the right being bounded above by the ratio of the diameter of the triangle \(T\) divided by the radius of the largest disk contained in \(T\).
For our triangulation $T_\nu$ obtained as the image of the uniform triangulation, we also know from the proof of Theorem 1.2 that
\[
\frac{C}{\sqrt{|\det(C)|}} \approx \eta \frac{\nabla \phi(\zeta)}{\sqrt{|\det(\nabla \phi(\zeta))|}}, \quad \zeta \in \phi^{-1}(T),
\]
and hence
\[
\delta := \sup_{\nu} \max_{T \in T_\nu} \delta_T = \sup_{\nu} \max_{T \in T_\nu} \text{cond} \left( \frac{C}{\sqrt{|\det(C)|}} \right) \approx \sup_{\zeta \in \Omega \setminus \Gamma} \text{cond} \left( \frac{\nabla \phi(\zeta)}{\sqrt{|\det(\nabla \phi(\zeta))|}} \right).
\]
This latter quantity turns out to be very simple for the refined triangulations discussed in Example 1.5 and Example 1.6, namely $\delta \approx \beta$, with $\beta \pi \in (\pi, 2\pi)$ being the largest inner angle of $\Omega$. We should notice that these last arguments are not completely rigorous, since in general relation (28) can only shown to be true for triangles $T$ with $\phi^{-1}(T)$ having a certain distance to $\Gamma$. However, there exist similar mesh refinements where the resulting family $(T_\nu)_\nu$ is shape-regular, and where explicit lower bounds for the angles are known.

### 3.2 Extremal eigenvalues, condition numbers, and preconditioning

The four statements in this section will have a short proof since they are related to previously known results. For our first statement we have been strongly inspired by similar results for so-called matrix-valued Linear and Positive Operators (LPOs) (see [22, 28]). Here we give a short direct proof.

**Theorem 3.2.** Assume that the matrix $K$ is uniformly elliptic and bounded, i.e., there exist positive constant $\kappa_{\min}$ and $\kappa_{\max}$ such that $\kappa_{\min} I_2 \leq K(x) \leq \kappa_{\max} I_2$ almost everywhere with respect to $x$ (for instance $\kappa_{\min} = \text{essinf}_x \lambda_{\min}(K(x))$, $\kappa_{\max} = \text{esssup}_x \lambda_{\max}(K(x))$). Then
\[
(A_n(K, T_\nu))_\nu \text{ and } (A_n(I_2, T_\nu))_\nu \text{ are uniformly equivalent}
\]
and more precisely $\kappa_{\min} A_n(I_2, T_\nu) \leq A_n(K, T_\nu) \leq \kappa_{\max} A_n(I_2, T_\nu),
\]
and the same result is true if one replaces $T_\nu$ in (29) by $\tilde{T}_\nu$.

Assume that the family of triangulations $(T_\nu)_\nu$ is shape-regular, and define
\[
\delta := \sup_{\nu} \max_{T \in T_\nu} \delta_T < \infty
\]
with $\delta_T$ as in (26). Then
\[
(A_n(I_2, T_\nu))_\nu \text{ and } (A_n(I_2, \tilde{T}_\nu))_\nu \text{ are uniformly equivalent}
\]
and more precisely $\frac{1}{\delta} A_n(I_2, \tilde{T}_\nu) \leq A_n(I_2, T_\nu) \leq \delta A_n(I_2, \tilde{T}_\nu).
\]

**Proof.** The main work for proving statements (29) and (30) has been done already in Subsection 3.1: according to (25), the claimed inequalities in (29) are obtained by summing over all triangles $T \in T_\nu$ the first inequality of Proposition 3.1. Similarly, relating the triangulations $T_\nu$ and $\tilde{T}_\nu$ for $K = I_2$ means that we have to study how the stiffness matrix changes if $C$ in (25) is replaced by $I_2$: the answer is obtained by summing the last inequality of Proposition 3.1 for all triangles (after replacing $\delta_T$ by $\delta$).

The preceding result enables us to give more precise bounds for the smallest and largest eigenvalue of the different stiffness matrices occurring in Theorem 3.2.
Corollary 3.3. Assume that the matrix $K$ is uniformly elliptic and bounded, and that $(T_{\nu})_{\nu}$ is shape-regular. Then the largest eigenvalue of $A_n(K, T_{\nu})$ is uniformly bounded in $\nu$ and the smallest behaves like $1/\nu^2$ for $\nu \to \infty$. 

In particular, the spectral condition number of $A_n(K, T_{\nu})$ behaves like $n$, the number of vertices of $T_{\nu}$.

Proof. Since $\Omega$ is bounded, it is contained in a square with sides of size $d_o$, and contains a square of size $d_i$. Then $A_n(I_2, \tilde{T}_{\nu})$ contains as submatrix the Toeplitz matrix generated by $4 - 2\cos(s_1) - 2\cos(s_2)$ of order $(d_i(\nu - 1))^2$, and in addition $A_n(I_2, \tilde{T}_{\nu})$ is a submatrix of a Toeplitz matrix generated by $4 - 2\cos(s_1) - 2\cos(s_2)$ of order $(d_o^2\nu)^2$ (see [30]). Since the eigenvalues of Toeplitz matrices generated by linear cosine polynomials are explicitly known, it follows that the smallest eigenvalue of $A_n(I_2, \tilde{T}_{\nu})$ is of order $1/\nu^2 \sim n^{-1}$ and its maximal eigenvalue is uniformly bounded by 8, which is also its limit for $\nu \to \infty$. Using for instance the well-known representation of extremal eigenvalues of Hermitian matrices in terms of Rayleigh quotients, it follows from Theorem 3.2 by combining (29) and (30) that all three matrices $A_n(K, T_{\nu})$, $A_n(I_2, T_{\nu})$, and $A_n(K, \tilde{T}_{\nu})$ have a smallest eigenvalue of order $1/\nu^2 \sim n^{-1}$ and a maximal eigenvalue being bounded uniformly in $\nu$. \hfill $\Box$

Corollary 3.3 has been proved in [1, relations (5.102c), page 235, and pp. 236-238], [7, Lemma V.2.6] and [15, page 61 and Lemma 2.6, page 233] under the additional assumption that $(T_{\nu})_{\nu}$ is also quasi-uniform. It seems that, for their technique of proof (comparing suitable Sobolev norms), this latter condition cannot be dropped.

Notice that $T_{\nu}$ as an image under $\phi$ of a uniform triangulation is always a 6-triangulation, that is, inner vertices are shared by exactly six triangles, and vertices on the boundary by at most 5 triangles. We believe that an assertion similar to Corollary 3.3 is true for any shape-regular 6-triangulation, even if it is not generated by some function $\phi$, as long as the diameter of the underlying graph is asymptotically related to the length of the shortest path of a fixed inner vertex to the boundary.

Let us finally turn to the problem of designing a preconditioner for the CG method applied to the system $A_n(K, T_{\nu})x_n = b_n$. We recall that the matrix $A_n(I_2, \tilde{T}_{\nu})$ corresponding to the uniform triangulation $T_{\nu}$ coincides with the one obtained by applying the classical FD 5 point stencil to the Poisson problem $-\Delta u = f$. Thus solving the system $A_n(I_2, \tilde{T}_{\nu})y_n = c_n$ can be performed in $\mathcal{O}(n)$ operations using, e.g., the method of cyclic reductions [9, 10, 12] and thus such a matrix would be a practical preconditioner. Define also the matrix

$$D_n = \text{diag} \left( \left\| \tilde{K} \left( \frac{(j, k)}{\nu} \right) \right\|_{\frac{1}{2} \leq \frac{j, k}{\nu} \in \tilde{\Omega}_h} \right)$$

which again is a practical preconditioner. Then, under the assumptions of Proposition 3.1, the condition number of $A_n(I_2, \tilde{T}_{\nu})^{-1}A_n(K, T_{\nu})$ and of $A_n(I_2, \tilde{T}_{\nu})^{-1}D_n^{-1/2}A_n(K, T_{\nu})D_n^{-1/2}$ can be bounded independently of the stepsize $1/\nu$ in terms of the smallest angle used in the triangulation of $\Omega$, plus possibly the norm and the ellipticity constant of $K$. This means that the associated preconditioned CG (PCG) will achieve a fixed precision in $\mathcal{O}(n)$ operations also in the non-constant coefficient case with a non-uniform triangulation.

In the following two results we give a complete picture (localization and distribution) of the spectral behavior of preconditioned matrix sequences arising from the use of the above mentioned preconditioners.
the eigenvalues of $A_n(I_2, \mathcal{T}_\nu)^{-1} A_n(K, \mathcal{T}_\nu)$ belong to $[\kappa_{\min}, \kappa_{\max}]$, \hspace{1cm} (31)

and the same result is true if one replaces $\mathcal{T}_\nu$ in (31) by $\mathcal{T}_\nu$.

Assume also that the family of triangulations $(\mathcal{T}_\nu)_\nu$ is shape-regular such that $\delta := \sup_{\nu} \max_{x \in \mathcal{T}_\nu} \delta_T < \infty$ with $\delta_T$ as in (26). Then

\[ \text{the eigenvalues of } A_n(I_2, \mathcal{T}_\nu)^{-1} A_n(I_2, \mathcal{T}_\nu) \text{ belong to } [1/\delta, \delta]; \hspace{1cm} (32) \]

\[ \text{the eigenvalues of } A_n(I_2, \mathcal{T}_\nu)^{-1} A_n(K, \mathcal{T}_\nu) \text{ belong to } [\kappa_{\min}/\delta, \kappa_{\max}\delta]. \hspace{1cm} (33) \]

**Proof.** Statements (31) and (32) follow from the corresponding statements (29) and (30) in Theorem 3.2 and the fact that, for Hermitian positive definite $X, Y$, we have for the spectrum $\Lambda(Y^{-1}X)$ the localization

\[ \Lambda(Y^{-1}X) \subset \left\{ \frac{u^* X u}{u^* Y u} : u \neq 0 \right\}. \]

The claim (33) follows from (29) and (30) by rewriting the Rayleigh quotient as

\[ \frac{u^* A_n(K, \mathcal{T}_\nu) u}{u^* A_n(I_2, \mathcal{T}_\nu) u} = \frac{u^* A_n(I_2, \mathcal{T}_\nu) u u^* A_n(I_2, \mathcal{T}_\nu) u}{u^* A_n(I_2, \mathcal{T}_\nu) u u^* A_n(I_2, \mathcal{T}_\nu) u}. \]

\[ \blacksquare \]

**Theorem 3.5.** Assume that the matrix $K$ is uniformly elliptic in the sense of Corollary 3.4. Consider the preconditioned sequences $(Y_n^{-1}X_n)$ with

\[ [Y_n, X_n] \subset \{ [A_n(I_2, \mathcal{T}_\nu), A_n(K, \mathcal{T}_\nu)], [A_n(I_2, \mathcal{T}_\nu), A_n(K, \mathcal{T}_\nu)], \]

\[ [A_n(I_2, \mathcal{T}_\nu), A_n(I_2, \mathcal{T}_\nu)], [A_n(I_2, \mathcal{T}_\nu), A_n(K, \mathcal{T}_\nu)] \} \]

Then, calling $\omega_X$ the symbol of $(X_n)$ and calling $\omega_Y$ the symbol of $(Y_n)$, we have that the joint asymptotic distribution of $(Y_n^{-1}X_n)$ is given by $\omega_X/\omega_Y$.

**Proof.** It is enough to observe that all the involved matrix sequences are such that both $X_n$ and $Y_n$ come from the same matrix valued linear positive operator for which the distribution is known (see Theorem 1.2) and is sparsely vanishing (i.e. the symbol vanishes in a set of zero Lebesgue measure). The conclusion follows from the general theory of linear positive operators as in Theorem 2.9 [23] (compare also Theorem 4.6 in [27] and Theorem 3.7 in [21]). \[ \blacksquare \]

With the notation of the above theorem, we remark that the same result could be proved for the matrices $[Y_n, X_n] = [D_n^{1/2} A_n(I_2, \mathcal{T}_\nu) D_n^{1/2}, A_n(K, \mathcal{T}_\nu)]$. Indeed $D_n^{1/2}, A_n(I_2, \mathcal{T}_\nu)$, and $A_n(K, \mathcal{T}_\nu)$ are all Generalized Locally Toeplitz sequences with sparsely vanishing symbols (i.e. zero on at most a set of zero Lebesgue measure): for $D_n$ the statement is trivial since the matrix is diagonal while for the remaining two matrix sequences this has been proven in Corollary 1.3. Then our claims follows from the fact that, if the symbols are all sparsely vanishing and sparsely unbounded (the inverse of a sparsely vanishing), then the operation $X_n \otimes Y_n$ gives also a sequence in the Generalized Locally Toeplitz class, with asymptotic joint spectrum described by the symbol $\omega_X \otimes \omega_Y$: this has been shown in [26, Theorem 5.8] for $\otimes$ being multiplication, in the same paper for $\otimes$ being addition or subtraction, and is known to be true also for inversion, that is, for the sequence $(Y_n^{-1}X_n)$.
In order to illustrate Theorem 3.5 and its link with Theorem 1.2, we mention more explicitly the example that the sequence of matrices \( (A_n(I_2, T)^{-1}A_n(K, T_\nu)) \) for \( \nu \to \infty \) has an joint asymptotic spectrum described by the measure \( \sigma \), with

\[
\int f \, d\sigma = \frac{1}{(2\pi)^2} \frac{1}{m(\Omega)} \int_{[-\pi,\pi]^2} ds \int_{\Omega} dx \, f(\omega(x, s)),
\]

\[
\tilde{K}(x) = |\det \nabla \phi(x)| \nabla \phi(x)^{-1} K(\phi(x)) \nabla \phi(x)^{-T}
\]
as before,

\[
\omega(x, s) = \frac{\omega_X(x, s)}{\omega_Y(x, s)} = \begin{bmatrix} 1 - e^{is_1} & 1 - e^{is_2} \\ 1 - e^{is_2} & 1 - e^{is_1} \end{bmatrix} \cdot \tilde{K}(x) \cdot \begin{bmatrix} 1 - e^{is_1} & 1 - e^{is_2} \\ 1 - e^{is_2} & 1 - e^{is_1} \end{bmatrix},
\]

and with \( \omega_X(x, s), \omega_Y(x, s) \) according to the notations of Theorem 3.5.

In particular (compare with (33)), the most important part of its eigenvalues lies in the interval

\[
[\kappa_{\min}, \kappa_{\max}] = \left[ \text{essinf}_{x \in \Omega} \lambda_{\min}(\tilde{K}(x)), \text{esssup}_{x \in \Omega} \lambda_{\max}(\tilde{K}(x)) \right].
\]

4 Concluding remarks

We have analyzed the spectral behavior of some Finite Element matrix sequences and related preconditioned sequences in terms of localization, extremal and, especially, distributional spectral results. The analysis could be used for deducing more precise bounds on the (P)CG convergence in view of the results in [3, 4, 5]: the related specific study and the related numerical experiments will be part of a subsequent work. We expect that, as in the Finite Difference setting (see [26]), the results of Theorem 1.2 can be generalized to the cases of higher dimensions, higher order differential operators, Peano Jordan measurable domains, Riemann integrable PDEs coefficients using the same techniques of proof (see also Remark 1.1). Also, again in analogy with the Finite Difference case, we think that there are similar results for other finite elements, by adapting the choice of the trigonometric polynomials in \( s \).

Beside the Locally Toeplitz idea, we have used in Section 3.1 another purely linear algebra tool, namely the local domain analysis: it consists in decomposing complicate matrix structures in linear combinations of nonnegative definite dyads or low-rank matrices for which the (spectral) analysis is very simple, and then to combine these results for deducing properties of the original matrix (for Finite Differences compare with [27, Section 3.5], [6, Theorem 3.7], and for general matrices see [21]). We mention that this simple tool is especially useful for preconditioning analysis and for the analysis of extremal eigenvalues asymptotics. Indeed, in Corollary 3.3 we have shown that the condition number is exactly of order \( h^{-2} \) under the classical angle condition of shape-regularity, but without requiring the additional condition of quasi-uniformity as imposed in [1, 7, 15]. As a byproduct we have deduced in Corollary 3.4 that the Finite Element matrix sequence with uniform triangulation and the non-uniform one (not necessarily verifying the quasi-uniformity) are spectrally equivalent. Thus the simpler two-level Toeplitz structure associated with the uniform triangulation can be employed as preconditioner requiring a constant number of iterations independently of the size of the problem.
References


