On condition numbers in $hp$-FEM with Gauss–Lobatto-based shape functions

J.M. Melenk

Max-Planck Inst. für Mathematik in den Naturwissenschaften, Inselstrasse 22–26, D-04103 Leipzig, Germany

Received 1 August 2000; received in revised form 9 March 2001

Abstract

Sharp bounds on the condition number of stiffness matrices arising in $hp$/spectral discretizations for two-dimensional problems elliptic problems are given. Two types of shape functions that are based on Lagrange interpolation polynomials in the Gauss–Lobatto points are considered. These shape functions result in condition numbers $O(p)$ and $O(p \ln p)$ for the condensed stiffness matrices, where $p$ is the polynomial degree employed. Locally refined meshes are analyzed. For the discretization of Dirichlet problems on meshes that are refined geometrically toward singularities, the conditioning of the stiffness matrix is shown to be independent of the number of layers of geometric refinement. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Condition number; Spectral method; Schur complement; Finite element method; Preconditioning

1. Introduction and model problem

Model problem and FEM formulation: Let $\Omega \subset \mathbb{R}^2$ be a bounded Lipschitz domain with piecewise smooth boundary. We are interested in the discretization by the $hp$-version of the finite element method (FEM) of elliptic problems of which a typical representative is the following model problem:

$$-\nabla \cdot (A(x)\nabla u) + c(x)u = f \quad \text{on } \Omega, \quad u = 0 \text{ on } \partial \Omega. \quad (1.1)$$

The symmetric matrix $A$ satisfies $0 < \alpha \leq A(x) \leq \|A\|_{L^\infty(\Omega)} < \infty$ and $c$ satisfies $0 \leq c \leq \|c\|_{L^\infty(\Omega)} < \infty$ on $\Omega$.

Given $V \subset H^1_0(\Omega)$ the FEM reads: Find $u \in V$ such that

$$B(u,v) := \int_{\Omega} A \nabla u \cdot \nabla v + cuv \, dx \, dy = F(v) := \int_{\Omega} f v \, dx \quad \forall v \in V. \quad (1.2)$$
In $h$-, $p$-, and $hp$-FEM, the approximation spaces $V$ are spaces of piecewise (mapped) polynomials—a more precise definition of these spaces follows below. The purpose of the present paper is to analyze the influence of the polynomial degree $p$ and the mesh size variation on the condition number of the stiffness matrix.

Locally refined meshes in the $h$-FEM and the $hp$-FEM: In practice, the input data $A$, $c$, $f$, and $\partial \Omega$ in (1.1) are (piecewise) smooth or even (piecewise) analytic. The solution $u$ of (1.1) is then also smooth up to boundary with the exception of the vertices of the curvilinear polygon $\Omega$ where it has singularities. In order to preserve the optimal rate of convergence, meshes that are suitably refined near the vertices are employed. In the context of the $h$-version FEM (i.e., the polynomial degree $p$ is fixed), the optimally refined meshes are the so-called radical meshes $T_{\text{rad}}$, [18, 5]. For example, in the case $p = 1$, i.e., if piecewise linear/bilinear shape functions are employed, the optimal mesh is quasi-uniform with mesh size $h_0$ in the interior of $\Omega$ and the refinement in a neighborhood of a vertex $A$ is performed such that the element size $h_k$ of each element $K \in T_{\text{rad}}$ with $A \notin \tilde{K}$ satisfies

$$h_k \sim h_0 \cdot (\text{dist}(A, x_K))^\beta$$

(1.3)

for some $\beta \in (0, 1)$. Here, $x_K$ is some suitable point in the “center” of the element (e.g., $x_K$ is the image of the barycenter of the reference element $\hat{K}$ under the element map $F_K$). The elements abutting on $A$ have to be of size $h_0^{1/(1-\beta)}$. We note that the smallest element size $h_{\text{min}} \approx h_0^{1/(1-\beta)}$ may be much smaller than $h_0$. One can show (cf. [18, 5]) that on such radical meshes the $h$-version FEM satisfies

$$\|u - u_{\text{FE}}\|_{H^1(\Omega)} \leq C h_0, \quad h_0^{-2} \sim |T_{\text{rad}}| = \text{number of elements in } T_{\text{rad}}.$$

Radical meshes for the approximation with piecewise polynomials of (fixed) degree $p \geq 1$ can also be designed and lead, given sufficient smoothness of $A$, $c$, $f$, to the following optimal performance of the $h$-FEM:

$$\|u - u_{\text{FE}}\|_{H^1(\Omega)} \leq C h_0^p, \quad h_0^{-2} \sim |T_{\text{rad}}|.$$

In the context of the $hp$-version of the FEM, the optimal meshes are geometric meshes $T_{\text{geo}}$ (see, e.g., [20, 23] for a precise definition). The key feature of a geometric mesh is that a fixed quasi-uniform mesh with elements of size $O(1)$ is used in the interior of $\Omega$, and the refinement in a neighborhood of a vertex $A$ is performed such that the element size $h_k$ of each element with $A \notin \tilde{K}$ satisfies

$$h_k \sim \text{dist}(A, x_K).$$

In a geometric mesh the size $h_{\text{min}}$ of the small elements abutting on the vertices $A$ is a measure for the number of elements in the mesh: there holds

$$|\log h_{\text{min}}| \sim |T_{\text{geo}}|.$$  

(1.4)

Put differently, $h_{\text{min}}$ is exponentially small in the number of elements of the mesh: $h_{\text{min}} \sim e^{-b|T_{\text{geo}}|}$ (See also Fig. 3 for an example of a geometric mesh). Thus, the variation in element size is very large in geometric meshes. If the input data $A$, $c$, $f$, $\partial \Omega$ are piecewise analytic, the following error bound can be proved for the $hp$-version of the FEM on geometric meshes, [4]:

$$\|u - u_{\text{FE}}\|_{H^1(\Omega)} \leq C [e^{-b|T_{\text{geo}}|} + e^{-b|T_{\text{geo}}|}].$$
Condition number estimates: dependence on the mesh: The locally refined meshes introduced above contain elements of greatly varying sizes. One of the purposes of the present paper is to show that this need not necessarily adversely affect the condition number of the corresponding stiffness matrix.

The standard technique to estimate the condition number of the stiffness matrix consists in splitting the bilinear form into element contributions, using polynomial inverse estimates on the reference element, and then combining these element contributions. This is the procedure of [6] and also our approach in Section 2.1. A consequence of this approach is that the condition number can be bounded in the form

$$\kappa \leq C_{\text{data}}C_pC_{\mathcal{F}}. \quad (1.5)$$

The factor $C_{\text{data}}$ depends only on the given data $A, c, \Omega$. The factor $C_p$ depends only on the polynomial degree and the choice of polynomial basis on the reference element, and the factor $C_{\mathcal{F}}$ reflects the dependence on the mesh. The bound (1.5) shows that the influence of the polynomial degree and the mesh can be considered separately.

Condition number estimates in the $h$-version FEM are only concerned with estimating $C_{\mathcal{F}}$. On general, shape regular meshes, $C_{\mathcal{F}}$ was found in [6] to be bounded by

$$C_{\mathcal{F}} \leq C|\mathcal{F}|(1 + \ln(|\mathcal{F}|h_{\min})), \quad |\mathcal{F}| = \text{number of elements in } \mathcal{F}. \quad (1.6)$$

For $p = 1$ and radical meshes as discussed above, we have $|\mathcal{T}_{\text{rad}}| \sim h_0^{-2}$ and $h_{\min} \sim h_0^{1/(1-\beta)}$ for some $\beta \in (0, 1)$. Thus, the condition number $\kappa$ is bounded by

$$\kappa \leq CC_{\mathcal{T}_{\text{rad}}} \leq C|\mathcal{T}_{\text{rad}}|(1 + \ln|\mathcal{T}_{\text{rad}}|) \quad (1.7)$$

in terms of the number of elements in the triangulation. A similar result holds for radical meshes for any fixed polynomial degree $p$. A by-product of the analysis presented in this paper is that, for radical meshes and the model problem (1.1) with Dirichlet boundary conditions, estimate (1.7) can be sharpened to $\kappa \leq C|\mathcal{T}_{\text{rad}}|$.

A similar situation is given for the $hp$-version of the FEM. The general estimate (1.6) implies

$$\kappa \leq C_{\text{data}}C_pC_{\mathcal{T}_{\text{geo}}} \leq CC_p|\mathcal{T}_{\text{geo}}|\{1 + \ln(|\mathcal{T}_{\text{geo}}|h_{\min})\} \leq CC_p|\mathcal{T}_{\text{geo}}|^2, \quad (1.8)$$

where the constant $C_p$ depends on the chosen basis of the polynomial space on the reference element (we will elaborate on this below). Again, this estimate does not make use of the fact that a Dirichlet problem is considered. Exploiting the homogeneous Dirichlet conditions gives bounds that are independent of $|\mathcal{T}_{\text{geo}}|$, i.e., $\kappa \leq C_p$ independent of the number of elements in the geometric mesh $\mathcal{T}_{\text{geo}}$.

Remark 1.1. The results of [6] are fairly general: the mesh need only be shape regular and boundary conditions are not explicitly exploited. Hence, the results of [6] are also applicable to meshes that are locally refined in the interior of the domain or to Neumann problems. Bank and Scott [6] shows that the bound (1.6) is attained if the differential equation (1.1) is considered with Neumann boundary conditions and a geometric mesh refinement toward vertices is employed. It is the presence of Dirichlet boundary conditions that allows us to sharpen some of their results. Examples 3.2 and 3.3 illustrate this phenomenon numerically.
Condition number estimates: dependence on $p$. We now turn to the $p$-dependence of the condition numbers, i.e., to estimates on $C_p$ in (1.5). On meshes consisting of quadrilaterals in 2D and cubes in 3D, results for the $p$-dependence of the condition number are available [7,17,15,13,9] for different choices of tensor product bases. For example, for the tensor product Babuška–Szabó polynomials (3.1), the constants $C_p$ can be bounded by $C_p \leq C p^4$, [15,13]. In the present paper, we will be concerned with two types of shape functions, called type GL and type SP. For tensor products of shape functions of type GL and SP (see Remark 1.4 for the precise definition of these tensor product shape functions), we obtain bounds $C_p \leq C p^4(1+\ln p)$ and $C_p \leq C p^3$, respectively (cf. Propositions 2.10 and 2.8).

In practice, many $hp$-FEM implementations perform static condensation as part of the element stiffness matrix generation and merely assemble the condensed element stiffness matrices. Bound on the condition number of the condensed stiffness matrix are therefore relevant. In fact, this local static condensation is a very good preconditioner in 2D. A further advantage of our analysis of condensed stiffness matrices is that it applies to meshes that may contain both quadrilaterals and triangles. For shape functions of type SP and type GL we obtain bounds $O(p)$ and $O(p \ln p)$, respectively, for the condition number of the condensed stiffness matrix.

Our results concerning the $p$-dependence of the condition number of condensed stiffness matrices are closely related to those of [9]. For the condensed stiffness matrix in the spectral method (corresponding to the shape functions of type SP in the notation of the present paper) [9] obtains the bound $O(p \ln p)$ and conjectures, based on numerical evidence, a bound $O(p)$. The present paper rigorously establishes this conjecture.

Outline of the paper: The paper is organized as follows: We start with the requisite notation in Section 1.1. In Section 2, we present the main theoretical results of this paper. In Section 2.1 we present the condition number estimates for the condensed stiffness matrix on locally refined meshes. In Section 2.2, we give bounds for the condition number on quadrilateral elements. We illustrate our theoretical results with several numerical examples in Section 3. Finally, Section 4 is devoted to the proof of several technical results.

1.1. Notation

Let $I=\{−1,1\} \subset \mathbb{R}$, $S=I \times I \subset \mathbb{R}^2$, $T=\{ (x, y) \mid −1 < x < 1, 0 < y < \sqrt{3}(1−|x|) \}$ be the reference interval, square, and triangle, respectively. For $p \in \mathbb{N}$, $\mathcal{P}_p(I) = \text{span}\{x^i \mid i = 0, \ldots, p\}$ denotes the space of all polynomials of degree $p$ on $I$, the tensor product space $\mathcal{Q}_p(S) = \text{span}\{x^i y^j \mid 0 \leq i, j \leq p\}$ is the space of polynomials of degree at most $p$ in each variable on $S$, and the space $\mathcal{Q}_p(T) = \text{span}\{x^i y^j \mid 0 \leq i + j \leq p\}$ is defined as the space of polynomials of total degree $p$ on $T$. We introduce the notation

$$V_p(\hat{K}) = \begin{cases} \mathcal{Q}_p(S) & \text{if } \hat{K} = S, \\ \mathcal{Q}_p(T) & \text{if } \hat{K} = T. \end{cases}$$

(1.9)

We will denote by $P_p(x)$ the $p$-th Legendre polynomial normalized such that $P_p(1) = 1$. For each $p \in \mathbb{N}$, the Gauss–Lobatto points $x_i$, $i = 0, \ldots, p$ are the zeros of the polynomial $(1−x^2)P'_p(x)$. As is well-known [7], they are all distinct and lie in the interval $[-1,1]$. Interpolation in the Gauss–Lobatto points $(x_i)_{i=0}^p$ is most conveniently formulated in terms of the cardinal polynomials
\( l_i \in \mathcal{P}_p(I) \) defined by
\[
l_i(x) := \prod_{j=0}^{p} \frac{x - x_j}{x_i - x_j}, \quad i = 0, \ldots, p.
\] (1.10)

We next introduce the following one-dimensional shape functions:
\[
\mathcal{B}_{\text{int}} := \{ l_i(x) \mid i = 1, \ldots, p - 1 \},
\]
\[
\mathcal{B}_{\text{SP,ext}} := \{ l_0(x), l_p(x) \}, \quad \mathcal{B}_{\text{GL,ext}} := \left\{ \frac{1}{2} (1 - x), \frac{1}{2} (1 + x) \right\}.
\]

In the following, we will analyze two types of bases for the spaces \( \mathcal{Q}_p(S) \) and \( \mathcal{P}_p(T) \). Bases of the first type will called SP (reminiscent of “spectral”) and the second GL (reminiscent of Gauss–Lobatto); they are required to satisfy the following conditions:

Definition 1.2. Let \( \hat{K} = S \) or \( \hat{K} = T \). Let \( v_j, j = 1, \ldots, n \) be the vertices of \( \hat{K} \). Set \( v_{n+1} = v_1 \). Let \( \Gamma_j \) be the edge of \( \hat{K} \) connecting \( v_j \) and \( v_{j+1} \), and set \( \Gamma_0 := \Gamma_n \).

For \( j \in \{0, \ldots, n\} \) let \( \gamma_j : I \rightarrow \Gamma_j \) be the parametrization by arc length of the edge \( \Gamma_j \) oriented such that \( \gamma_j(+1) = v_{j+1} \). A basis \( \mathcal{B} = \mathcal{N} \cup \mathcal{S} \cup \mathcal{I} \) of the space \( \mathcal{V}_p(\hat{K}) \) (cf. (1.9)) is said to be of type SP or GL if the following holds:

(C1) The internal shape functions \( \mathcal{I} \) vanish on \( \partial \hat{K} \); their span is denoted by \( \tilde{\mathcal{I}}(\hat{K}) \).

(C2) The side shape functions \( \mathcal{S} = \{ s_{i,j} \mid i = 1, \ldots, n, j = 1, \ldots, p - 1 \} \) satisfy:
\[
\text{(a)} \quad s_{i,j} |_{\Gamma_k} = 0 \quad \text{for} \quad k \neq i,
\]
\[
\text{(b)} \quad s_{i,j} |_{\Gamma_i} \circ \gamma_i \in \mathcal{B}_{\text{int}}.
\]

The span of the side shape functions is denoted by \( \tilde{\mathcal{S}}(\hat{K}) \).

(C3) The vertex shape functions \( \mathcal{N} = \{ n_i \mid i = 1, \ldots, n \} \) satisfy \( n_i(v_j) = \delta_{ij} \) and \( n_i |_{\Gamma_j} = 0 \) for \( j \notin \{i - 1, i\} \). Furthermore, only one of the following two cases can occur:
\[
\text{(a)} \quad n_i |_{\Gamma_j} \circ \gamma_j \in \mathcal{B}_{\text{SP,ext}} \quad \text{holds for all} \quad i \in \{1, \ldots, n\} \quad \text{and} \quad j \in \{i - 1, i\}.
\]
\[
\text{(b)} \quad n_i |_{\Gamma_j} \circ \gamma_j \in \mathcal{B}_{\text{GL,ext}} \quad \text{holds for all} \quad i \in \{1, \ldots, n\} \quad \text{and} \quad j \in \{i - 1, i\}.
\]

The span of the vertex shape functions is denoted by \( \tilde{\mathcal{N}}(\hat{K}) \).

Bases \( \mathcal{B} \) satisfying conditions (C1), (C2), (C3a) are said to be of type SP, those satisfying (C1), (C2), (C3b) of type GL. Vertex and side shape functions form the external shape functions:
\[
\mathcal{E} = \mathcal{N} \cup \mathcal{S} := \{ e_i \mid i = 1, \ldots, np \}.
\]

Vertex shape functions \( \mathcal{N} \) or external shape functions \( \mathcal{E} \) are said to be type SP or GL if they can be completed to a basis \( \mathcal{B} \) of \( \mathcal{V}_p(\hat{K}) \) that is of type SP or GL.

Remark 1.3. The distinction between internal shape functions, side shape functions, and vertex shape functions is standard in \( h p \)-FEM. Bases of type SP and GL have the same side shape functions but differ in the vertex shape functions (and possibly in the internal shape functions). In both cases, the side shape functions are, on each edge, Lagrange interpolation polynomials in the Gauss–Lobatto points. The vertex shape functions of type GL are, on each edge, linear while the vertex shape functions of type SP are also Lagrange interpolation polynomials in the Gauss–Lobatto points. For
\( \hat{K} = S \), a specific example of a basis of type SP is the tensor product \( B_{SP} \otimes B_{SP} \) with \( B_{SP} = B_{SP, \text{ext}} \cup B_{\text{int}} \).

An example of a basis of type GL for \( \hat{K} = S \) is \( B_{GL} \otimes B_{GL} \) with \( B_{GL} = B_{GL, \text{ext}} \cup B_{\text{int}} \). We recognize \( B_{SP} \otimes B_{SP} \) as customary choice in spectral methods, which motives the label “SP”.

**Remark 1.4.** Definition 1.2 prescribes the behavior of the vertex and side shape functions on \( \partial \hat{K} \) only; they are thus defined up to additive elements from \( \tilde{I} \).

We introduce finite element spaces \( V \) as spaces of piecewise mapped polynomials in the standard way. Let \( \Omega \subset \mathbb{R}^2 \) be the computational domain. A triangulation \( \mathcal{T} = \{ K \} \) of \( \Omega \) is a partitioning of \( \Omega \) into elements \( K \); with each element \( K \), a bijective element map \( F_K : \hat{K} \rightarrow \tilde{K} \) is associated where the reference element \( \hat{K} \) is either the reference square \( S \) or the reference triangle \( T \). The elements and the element maps satisfy the following conditions:

1. (M1) The elements \( K \) are mutually disjoint and \( \bigcup K = \tilde{K} \); for two elements \( K, K' \), the intersection \( \tilde{K} \cap \tilde{K}' \) is either empty, a vertex, or a whole edge (vertices and edges are the images of the vertices and edges of the reference element \( \hat{K} \) under the element maps).

2. (M2) The element maps \( F_K : \hat{K} \rightarrow \tilde{K} \) are \( C^1 \) diffeomorphisms. There is a constant \( C_M > 0 \) and constants \( h_K > 0 \) (the “element sizes”) such that
   \[
   \| F_K' \|_{L_\infty(\hat{K})} \leq C_M h_K, \quad C_M^{-1} h_K^2 \leq \det F_K' \leq C_M h_K^2 \quad \text{on} \ \hat{K}.
   \]

3. (M3) The number of elements sharing a vertex is bounded by \( C_M \).

With the spaces \( V_p(\hat{K}) \) of (1.9) we can define the finite element space \( V \) as
\[
V := \{ u \in H^1_0(\Omega) \mid u|_K \circ F_K \in V_p(\hat{K}) \}.
\]

For given bases of \( V_p(\hat{K}) \) problem (1.2) can then be recast as a linear system of equations. Setting up this linear system is part of the so-called *assembly*, for which we refer to [20,23]. For the purpose of this paper, we assume that the space \( V \) is either of type SP or GL in the sense that the chosen basis for \( V_p(\hat{K}) \) is of type SP for all elements \( K \in \mathcal{T} \) or that the basis for \( V_p(\hat{K}) \) is of type GL for all elements \( K \in \mathcal{T} \).

Finally, we use standard notation for the Sobolev spaces \( H^k \), [1]. The fractional order spaces \( H^{1/2}(I) \), \( H^{1/2}_{00}(I) \) on the interval \( I = (-1,1) \) are characterized by the norms
\[
\| u \|^2_{H^{1/2}(I)} := \int_I \int_I \frac{|u(x) - u(y)|^2}{|x - y|^2} \, dx \, dy,
\]
\[
\| u \|^2_{H^{1/2}_{00}(I)} := \int_I \left( \frac{1}{\text{dist}(x, \partial I)} \right) u^2(x) \, dx.
\]

If \( \hat{K} \) is the reference square or the reference triangle with edges \( \Gamma_i, i = 1, \ldots, n \) (\( n = 3 \) for \( \hat{K} = T \) and \( n = 4 \) for \( \hat{K} = S \)), then the trace on the boundary of functions of \( H^1(\hat{K}) \) can be characterized by (see [11,
Theorem 1.5.2.3:}
\[ \|u\|_{H^1(\hat{K})}^2 := \inf \{ \|v\|_{H^1(\hat{K})}^2 \mid v|_{\partial \hat{K}} = u \} \] (1.11a)
\[ \sim \sum_{i=1}^n \|u\|_{H^1(T_i)}^2 + \sum_{i=1}^n \int_0^1 \frac{1}{t} |u(x_i(t)) - u(x_i(-t))|^2 \, dt, \] (1.11b)
where \( x_i: (-2, 2) \to \mathbb{R}^2 \) is the parametrization by arclength of two edges of \( \partial \hat{K} \) that meet in the vertex \( A_i \) such that \( x_i(0) = A_i \).

Finally, we denote the \( l^2 \)-norm for vectors and the associated matrix norm by \( \| \cdot \|_2 \).

2. Condition number estimates

2.1. Condition numbers with static condensation

We consider continuous, symmetric, and coercive bilinear forms \( B \) on the finite element space \( V \subset H^1_0(\Omega) \) satisfying for some \( c_1, c_2 > 0 \)
\[ c_1 \|u\|_{H^1(\Omega)}^2 \leq B(u, u) \leq c_2 \|u\|_{H^1(\Omega)}^2 \quad \forall u \in V. \] (2.1)
The bilinear form \( B \) is assumed to be of the form
\[ B(u, v) := \sum_{K \in \mathcal{T}} \hat{B}_K(\hat{u}, \hat{v}), \quad \hat{u} = u|_K \circ F_K, \quad \hat{v} = v|_K \circ F_K, \quad u, v \in V, \]
where the elemental bilinear forms \( \hat{B}_K \) are symmetric and satisfy
\[ 0 < \hat{B}_K(u, u) \quad \forall 0 \neq u \in \mathcal{E}(\hat{K}), \quad 0 \leq \hat{B}_K(u, u) \leq C_B \|u\|_{H^1(\hat{K})}^2 \quad \forall u \in V_p(\hat{K}). \] (2.2)

Remark 2.1. The bilinear form \( B \) analyzed in this section is slightly more general than that obtained as the \( hp \)-Galerkin discretization of (1.1). The main motivation for working with the assumptions (2.1) and (2.2) is that it allows us to analyze the numerical treatment of (1.1) by fully discrete schemes such as Galerkin scheme with quadrature.

Using the element shape functions \( \mathcal{N}(\hat{K}), \mathcal{I}(\hat{K}), \mathcal{J}(\hat{K}) \) (cf. Definition 1.2) and the elemental bilinear forms \( \hat{B}_K \), the element stiffness matrices \( A_K \) are defined in the standard way: \( A_K \) has block structure
\[ A_K = \begin{pmatrix} A_{EE} & A_{EI} \\ A_{IE} & A_{II} \end{pmatrix}, \]
where the entries of the submatrices \( A_{EE}, A_{EI}, A_{IE} = A_{EI}^T, \) and \( A_{II} \) are given by
\[ (A_{EE})_{ij} = \hat{B}_K(e_i, e_j), \quad e_i, e_j \in \mathcal{E}(\hat{K}), \]
\[ (A_{EI})_{ij} = \hat{B}_K(e_i, b_j), \quad e_i \in \mathcal{E}(\hat{K}), \quad b_j \in \mathcal{J}(\hat{K}), \]
\[ (A_{II})_{ij} = \hat{B}_K(b_i, b_j), \quad b_i, b_j \in \mathcal{J}(\hat{K}). \]
The standard finite element assembly operator \( \mathcal{A} \) (see, e.g., [14]) generates a basis \( \mathcal{B} \) of \( V \) from the elemental basis functions and the global stiffness matrix \( A \) from the elemental stiffness matrices \( A_K \):

\[
\mathcal{B} = \mathcal{A} \left( \mathcal{N}(\hat{K}), \mathcal{S}(\hat{K}), \mathcal{I}(\hat{K}) \right), \quad A = \mathcal{A} A_K.
\]

The assumptions on the elemental bilinear forms allow forming the Schur complements

\[
A^c_K := A_{EE} - A_E A_{II}^{-1} A_{IE},
\]

and the global condensed system is then obtained by assembling these condensed matrices into

\[
A^c := \mathcal{A} A^c_K.
\]

For our purposes, it is essential to note that the condensed matrix \( A^c \) can be obtained as the stiffness matrix corresponding to the bilinear form \( \mathcal{B}^c \) acting on the basis functions \( \mathcal{B}^c \) of a space \( \hat{V} \subset V \), where \( \mathcal{B}^c \) is obtained by assembling modified external shape functions:

\[
\mathcal{B}^c = \mathcal{A} \delta^c(\hat{K}), \quad \delta^c(\hat{K}) = \{ e_i + h_i | i = 1, \ldots, np \}.
\]

The functions \( h_i \) occurring in the modified external shape functions \( \delta^c(\hat{K}) \) are (the unique) solutions to the problems:

\[
\text{find } h_i \in \hat{\mathcal{I}}(\hat{K}) \text{ s.t. } \hat{B}_K(e_i + h_i, v) = 0 \quad \forall v \in \hat{\mathcal{I}}(\hat{K}),
\]

i.e., the functions \( e_i + h_i \) are discretely harmonic. We are now ready to formulate the main result of this section:

**Theorem 2.2.** Let \( r(x) := \text{dist}(x, \partial \Omega) \). Assume that there are \( c_{\text{geo}}, h_0 > 0 \) such that the elements \( K \) of a triangulation \( \mathcal{T} \) satisfy \( h_K \geq c_{\text{geo}} h_0 r(x_K) \), where \( x_K \) is the image of the barycenter of \( \hat{K} \) under \( F_K \). Let the space \( V \subset H_0^1(\Omega) \) be of type SP or GL. Assume that the elemental bilinear forms \( \hat{B}_K \) and the bilinear form \( B \) satisfy (2.1) and (2.2). Then, for some \( C > 0 \) depending only on \( C_M \) of (M2), \( c_{\text{geo}} \), and the constants of (2.1) and (2.2), the condensed stiffness matrix \( A^c \) in (2.4) satisfies

\[
\|A^c\|_2 \leq C, \quad (A^c)^{-1} \|_2 \leq C h_0^{-2} p(1 + \ln p) \quad \text{if } V \text{ is of type GL},
\]

\[
\|A^c\|_2 \leq C, \quad (A^c)^{-1} \|_2 \leq C h_0^{-2} p \quad \text{if } V \text{ is of type SP}.
\]

**Proof.** Define \( \tilde{C}_p := p \) if \( V \) is of type SP and \( \tilde{C}_p := p(1 + \ln p) \) if \( V \) is of type GL. Lemmas 2.5 and 2.6 give the existence of \( C > 0 \) such that for all elements \( K \) there holds for all functions \( \hat{u} = \sum_{i=1}^{np} u_i (e_i + h_i) \)

\[
\sum_{i=1}^{np} u_i^2 \leq C \tilde{C}_p \| \hat{u} \|_{H^1(\hat{K})}^2, \quad \hat{B}_K(u, u) \leq C C_B \sum_{i=1}^{np} u_i^2.
\]

Next, if \( \mathcal{B}^c = \{ b_i^c \mid i = 1, \ldots, N \} \) is the basis obtained from assembling the modified shape functions \( \delta^c(\hat{K}) \), (2.5) and the assembly process imply together that for all \( u \in \text{span} \mathcal{B}^c \) of the form...
\[ u = \sum_{i=1}^{N} u_i b_i^e \] 

there holds

\[ \sum_{i=1}^{N} u_i^2 \leq C \mathcal{C}_p \sum_{K \in \mathcal{F}} \| \nabla u \|^2_{L^2(K)} + h^{-2}_\mathring{K} \| u \|^2_{L^2(K)}, \quad B(u, u) \leq C C_B \sum_{i=1}^{N} u_i^2. \]

Lemma 2.7 allows us to estimate further

\[ \sum_{K \in \mathcal{F}} \| \nabla u \|^2_{L^2(K)} + h^{-2}_\mathring{K} \| u \|^2_{L^2(K)} \leq C h_0^{-2} B(u, u). \]

Thus

\[ \sum_{i=1}^{N} u_i^2 \leq C \mathcal{C}_p h_0^{-2} B(u, u), \quad B(u, u) \leq C \sum_{i=1}^{N} u_i^2. \]

The claim of the Theorem now follows easily from these last two bounds. \( \Box \)

Several comments concerning Theorem 2.2 are in order:

**Remark 2.3.**

1. The condition \( V \subset H^1_0(\Omega) \) expresses the fact that we consider the discretization of a Dirichlet problem. The boundary conditions are used in an essential way (Lemma 2.7).
2. We mentioned in Section 1.1 that the optimal meshes for the \( hp \)-version of the FEM are meshes that are geometrically refined toward the vertices. Theorem 2.2 can be adapted to this case and it can be shown that the condition number of the condensed stiffness matrix does not depend on the number of layers of the geometric refinement. Additionally, Theorem 2.2 permits (shape regular) refinement toward \( \partial \mathcal{G} \) and not merely toward the vertices.
3. Theorem 2.2 is formulated for condensed stiffness matrices due to their importance in practice. The mesh-independence of the condition number is due to Lemma 2.7 and thus valid for many other choices of polynomial bases. In particular, the mesh independence holds for uncondensed stiffness matrices also.
4. For \( h \)-FEMs the optimal meshes are radical meshes characterized by (1.3). Such meshes are of the type considered in Theorem 2.2. Hence, the condition number of the stiffness matrix of the \( h \)-FEM on radical meshes is \( O(h_0^{-2}) \).
5. For our model problem (1.2), the elemental bilinear form \( \mathcal{B} \) is \( \mathcal{B}_K(u, v) := \int_K \nabla u \hat{A}(x) \nabla v + c u v \) (the coefficients \( \hat{A}, \hat{c} \) are obtained from the change of variables to the reference element \( \hat{K} \)), and it satisfies the assumptions of Theorem 2.2. The results of Theorem 2.2 therefore apply if the integration over \( \hat{K} \) is performed exactly. From an implementational point of view it is interesting to allow for numerical quadrature, in effect leading then to spectral methods (see also [16]). For certain types of quadrature (e.g., Gauss–Lobatto quadrature with \((p+1)\times(p+1)\) points if \( \hat{K} = \hat{S} \)), hypotheses (2.1), (2.2) can be ascertained with constants independent of \( p \), and Theorem 2.2 is again applicable.

The remainder of this subsection is devoted to the proof of auxiliary results that were used in the proof of Theorem 2.2.

**Lemma 2.4.** Let \( \hat{K} = \hat{S} \) or \( \hat{K} = \hat{T} \). Let \( \mathcal{N} = \{ n_i \mid i = 1, \ldots, n \} \) be vertex shape functions of type \( SP \) or \( GL \). Then there exists \( C > 0 \) independent of \( p \) such that

\[ \| n_i \|_{H^{1/2}(\hat{K})} \leq C, \quad i \in \{ 1, \ldots, n \}. \]
Proof. Vertex shape functions \( n_i \) of type GL are piecewise linear on \( \partial \hat{K} \) and independent of \( p \); the assertion therefore follows trivially. For vertex shape functions of type SP, we first note that by (1.11)
\[
\|u\|^2_{H^{1/2}(\partial \hat{K})} \sim \sum_{i=1}^{n} \|u\|^2_{H^{1/2}(\Gamma_i)} + \sum_{i=1}^{n} \int_0^1 \|u(x_i(t)) - u(x_i(-t))\|^2 dt,
\]
where the functions \( x_i : (-2, 2) \to \partial \hat{K} \) are parametrizations by arclength of the two edges meeting at the vertices \( v_i \) with \( x_i(0) = v_i \). Hence,
\[
\|n_i\|^2_{H^{1/2}(\partial \hat{K})} \leq C \left\| l_0 \right\|^2_{H^{1/2}(\Gamma)} + \int_I \frac{1}{1 - x} l_0^2(x) dx.
\]
Finally, \( \|l_0\|^2_{H^{1/2}(\Gamma)} \leq C \) by Theorem 4.1(iii) and
\[
\int_{-1}^{1} \frac{1}{1 - x} l_0(x) dx = \frac{1}{2} \frac{1}{p(p + 1)} \leq 1
\]
by a reasoning similar to that in (4.22). This completes the proof. \( \square \)

Lemma 2.5. Let \( \hat{K} = S \) or \( \hat{K} = T \). Then there is \( C > 0 \) such that the following holds. Let \( \mathcal{E} = \{e_i | i = 1, \ldots, np\} \) be a set of external shape functions of either type SP or GL. Then for all \( u \) of the form \( u = \sum_{i=1}^{np} u_i e_i \):
\[
C^{-1} \|u\|^2_{H^{1/2}(\partial \hat{K})} \leq \sum_{i=1}^{np} u_i^2 \leq \begin{cases} \quad C p \|u\|^2_{H^{1}(\hat{K})} & \text{if } \mathcal{E} \text{ is of type SP}, \\ C p (1 + \ln p) \|u\|^2_{H^{1}(\hat{K})} & \text{if } \mathcal{E} \text{ is of type GL}. \end{cases}
\] (2.6)

Proof. We start with the upper bound in (2.6). We observe that the assumptions on the behavior of the functions \( e_i \) on \( \partial \hat{K} \) imply that \( |u_i| \leq 2 \|u\|_{L^\infty(\partial \hat{K})} \). Hence, for \( u = \sum_{i=1}^{np} u_i e_i \) we can bound
\[
\sum_{i=1}^{np} u_i^2 \leq 4n p \|u\|^2_{L^\infty(\partial \hat{K})} \leq C p (1 + \ln p) \|u\|^2_{H^{1}(\hat{K})}
\]
by [3, Corollary 6.3]. This shows the upper bound in (2.6) for external shape functions of type GL. For functions of type SP, we employ Theorem 4.1(vi) to obtain the sharper bound
\[
\sum_{i=1}^{np} u_i^2 \leq 4 p \sum_{i=1}^{n} \|u\|^2_{H^{1/2}(\Gamma_i)} \leq C p \|u\|^2_{H^{1}(\hat{K})}.
\]
We now turn to the lower bound in (2.6). Upon writing \( u = \sum_{i=1}^{np} u_i e_i = \sum_{i=1}^{n} u_{i,0} e_i + \sum_{j=1}^{n} \sum_{i=1}^{p} u_{i,j} s_{i,j} \) and exploiting the support conditions on the side shape functions \( s_{i,j} \), we obtain
\[
\|u\|_{H^{1/2}(\partial \hat{K})} \leq \sum_{i=1}^{n} \|u_{i,0}\|_{H^{1/2}(\partial \hat{K})} + \sum_{j=1}^{p} \| \sum_{i=1}^{n} u_{i,j} s_{i,j} \|_{H^{1/2}(\Gamma_j)}.
\]
By Lemma 2.4 \( \|n_i\|^2_{H^{1/2}(\partial \hat{K})} \leq C \) for \( i = 1, \ldots, n \). From Theorem 4.1, we furthermore get the existence of \( C > 0 \) independent of \( p \) such that
\[
\left\| \sum_{j=1}^{p} u_{i,j} s_{i,j} \right\|_{H^{1/2}(\Gamma_j)} \leq C \left\{ \sum_{j=1}^{p} |u_{i,j}|^2 \right\}^{1/2}.
\]
We conclude that
\[
\|u\|_{H^{1/2}(\partial \hat{K})}^2 \leq C \left[ \sum_{i=1}^{n} |u_{i,0}|^2 + \sum_{i=1}^{n} \sum_{j=1}^{p-1} |u_{i,j}|^2 \right] \leq C \sum_{i=1}^{np} |u_i|^2. \]

Lemma 2.6. Let \( \hat{K} = S \) or \( \hat{K} = T \). Let \( \hat{B} \) be a symmetric bilinear form on \( V_p(\hat{K}) \) satisfying
\[
0 \leq \hat{B}(u,u) \leq C_B \|u\|_{H^1(\hat{K})}^2 \quad \forall u \in V_p(\hat{K}).
\]
Then there exists \( C > 0 \) independent of \( p \) such that the following is true. Let \( \mathcal{E} = \{e_i | i=1,\ldots,np\} \) be a set of external shape functions (either of type SP or GL) that are discrete harmonic, i.e.,
\[
\hat{B}(e_i,v) = 0 \quad \forall v \in \tilde{\mathcal{I}}, \ i = 1,\ldots,np. \tag{2.7}
\]
Then for all \( u = \sum_{i=1}^{np} u_ie_i \) there holds
\[
|\hat{B}(u,u)| \leq CC_B \sum_{i=1}^{np} u_i^2.
\]

Proof. Babuška et al. [3, Theorems 7.4 and 7.5] give the existence of \( C_E > 0 \) independent of \( p \) such that for every polynomial \( u \) there exists an \( h \in \mathcal{I} \) with
\[
\|u + h\|_{H^1(\hat{K})} \leq C_E \|u\|_{H^{1/2}(\partial \hat{K})}.
\]
Next, we can estimate using (2.7)
\[
0 \leq \hat{B}(u,u) = \hat{B}(u,u) + 2\hat{B}(u,h) \leq \hat{B}(u,u) + 2\hat{B}(u,h) + \hat{B}(h,h) = \hat{B}(u + h,u + h)
\leq C_B \|u + h\|_{H^1(\hat{K})}^2 \leq C_B C_E^2 \|u\|_{H^{1/2}(\partial \hat{K})}^2.
\]
Appealing to Lemma 2.5 concludes the argument. \( \square \)

Lemma 2.7. Let \( r(x) := \text{dist}(x, \partial \Omega) \). Assume that there is \( c_{\text{geo}} > 0 \) such that the elements \( K \) of a triangulation \( \mathcal{T} \) satisfy \( h_K \geq c_{\text{geo}} \eta_0 r(x_K) \) where \( x_K \) is the image of the barycenter of \( \hat{K} \) under \( F_K \). Then there is \( C > 0 \) depending only on \( C_M \) of (M2), \( c_{\text{geo}} \), and the constants of (2.1) such that
\[
\sum_{K \in \mathcal{T}} \| \nabla u \|_{L^2(K)}^2 + h_K^{-2} \| u \|_{L^2(K)}^2 \leq Ch_0^{-2} B(u,u) \quad \forall u \in H^1_0(\Omega).
\]

Proof. The result follows readily from the observation
\[
\sum_{K \in \mathcal{T}} h_K^{-2} \| u \|_{L^2(K)}^2 \leq \sum_{K \in \mathcal{T}} c^{-2} h_0^{-2} \| r^{-1} u \|_{L^2(K)}^2 \leq Ch_0^{-2} \| \nabla u \|_{L^2(\Omega)}^2 \quad \forall u \in H^1_0(\Omega)
\]
by a standard embedding result in weighted Sobolev spaces [11, Theorem 1.4.4.3]. \( \square \)
2.2. Condition number estimates without condensation

For the case of rectangular elements, i.e., $\hat{K} = S$, we defined in Remark 1.4 the sets $\mathcal{B}_{\text{SP}} \otimes \mathcal{B}_{\text{SP}}$ and $\mathcal{B}_{\text{GL}} \otimes \mathcal{B}_{\text{GL}}$ as examples of bases for $V_p(\hat{K})$ that are of type SP and GL. In the present section we derive bounds on the condition number for these two choices.

Proposition 2.8. Let $l_i, i = 0, \ldots, p$, be the Lagrangian interpolation polynomials in the Gauss–Lobatto points given by (1.10). There exists $C > 0$ independent of $p$ such that for all polynomials $u$ of the form $u(x, y) = \sum_{i,j=0}^p u_{ij} l_i(x) l_j(y)$

$$C^{-1} p^{-2} \sum_{i,j=0}^p u_{ij}^2 \leq \|u\|_{H^1(S)}^2 \leq C p \sum_{i,j=0}^p u_{ij}^2.$$

Proof. By Theorem 4.1(vi) there exists $C > 0$ such that for all $p$ and all $j \in \{0, \ldots, p\}$

$$\sum_{i=0}^p u_{ij}^2 \leq Cp \|u(\cdot, x_j)\|_{H^1/2(I)}^2. \quad (2.8)$$

It is easy to see that there exists $C > 0$ independent of $j$ such that

$$\|u(\cdot, x_j)\|_{H^1/2(I)} \leq C \|u\|_{H^1(S)}. \quad (2.9)$$

Inserting this bound in (2.8) and summing on $j$ gives the desired lower bound. For the upper bound, we employ Lemma 4.3 and Theorem 4.1(iv) to get

$$\|u\|_{L^2(S)}^2 \leq 9 \sum_{i,j=0}^p \rho_j \rho_i u_{ij}^2 \leq C p^{-2} \sum_{i,j=0}^p u_{ij}^2,$$

$$\|u_x\|_{L^2(S)}^2 \leq 9 \sum_{j=0}^p \rho_j \|u_x(\cdot, x_j)\|_{L^2(I)}^2 \leq C \sum_{j=0}^p \rho_j p^2 \sum_{i=0}^p u_{ij}^2 \leq C p \sum_{i,j=1}^p u_{ij}^2$$

as, by Lemma 4.6 and the definition of $\rho_j$, we have $\rho_j \leq C p^{-1}$. \qed

Remark 2.9. Proposition 2.8 shows that the condition number of the spectral method is $O(p^3)$. Proposition 2.8 is a generalization of existing results: The condition number bound $O(p^3)$ for the case of homogeneous Dirichlet condition, i.e., the restriction of Proposition 2.8 to polynomials of the form $u(x, y) = \sum_{i,j=0}^{p-1} u_{ij} l_i(x) l_j(y)$, is known in the spectral element literature [7,2,15].

Proposition 2.10. Let $l_i, i = 0, \ldots, p$, be the Lagrangian interpolation polynomials in the Gauss–Lobatto points given by (1.10). Let $\phi_i(x) = l_i(x)$ for $i = 1, \ldots, p = 1$, $\phi_0(x) = \frac{1}{2}(1 - x)$, $\phi_p(x) = \frac{1}{2}(1 + x)$. Then there exists $C > 0$ independent of $p$ such that for all polynomials $u$ of the form $u(x, y) = \sum_{i,j=0}^p u_{ij} \phi_i(x) \phi_j(y)$ there holds

$$C^{-1} \frac{1}{p^2(1 + \ln p)} \sum_{i,j=0}^p u_{ij}^2 \leq \|u\|_{H^1(S)}^2 \leq C p^2 \sum_{i,j=0}^p u_{ij}^2.$$
Proof. We start with the upper bound. We will only demonstrate the arguments for the contribution \( \|u_\chi\|_{L^2(S)}^2 \), the other terms being handled analogously. Using Theorem 4.1 (iv), we estimate

\[
\|u_\chi(\cdot, x_j)\|_{L^2(I)}^2 \leq 2 \sum_{k=1}^{p} \sum_{l=0}^{p} \sum_{l'=0}^{p} \|u_{kl'}(\cdot)\phi_l(x_j)\|_{L^2(I)}^2 + 2 \sum_{k=0}^{p} \sum_{l=0}^{p} \|u_{kl'}(\cdot)\phi_l(x_j)\|_{L^2(I)}^2
\]

where we exploited the definition of the functions \( \phi_k \). Using Lemma 4.3 and the bound \( \rho_j \leq Cp^{-1} \), which follows from Lemma 4.6 and (4.3), we get

\[
\|u_\chi\|_{L^2(S)}^2 \leq C \sum_{j=0}^{p} \rho_j \|u_\chi(\cdot, x_j)\|_{L^2(I)}^2 \leq C p^2 \sum_{k,l=0}^{p} u_{kl'}^2.
\]

For the lower bound, we start by using [3, Corollary 6.3] to bound for each fixed \( j \):

\[
\sum_{i=0}^{p} u_{ij}^2 \leq (p + 1)\|u(\cdot, x_j)\|_{L^\infty(I)}^2 \leq C p(1 + \ln p)\|u(\cdot, x_j)\|_{H^1(I)}^2 \leq C p(1 + \ln p)\|u\|_{H^1(S)}^2,
\]

where we employed (2.9). Summing on \( j \) then gives the desired upper bound. □

3. Numerical examples

In the present section, we present numerical examples that illustrate our theoretical results of Sections 2.1 and 2.2. We start by illustrating Propositions 2.8 and 2.10.

Example 3.1. We consider the conditioning of the stiffness matrix corresponding to the bilinear form

\[
\hat{B}(u, v) = \int_S \nabla u \cdot \nabla v \, dx \, dy.
\]

Taking as a basis of \( \mathcal{P}_p(S) \) the set \( \mathcal{P}_{SP} \otimes \mathcal{P}_{SP} \) yields the stiffness matrix \( A^{SP} \); the choice \( \mathcal{P}_{GL} \otimes \mathcal{P}_{GL} \) leads to the matrix \( A^{GL} \). By \( A_{II}^{SP} \), \( A_{II}^{GL} \) we denote the submatrices of \( A^{SP} \), \( A^{GL} \) that correspond to the internal degrees of freedom, and we write \( A_{c}^{SP} \), \( A_{c}^{GL} \) for the condensed stiffness matrices defined by (2.3). From Proposition 2.8 we infer

\[
\kappa(A^{SP}_{II}) \leq Cp^3, \quad \kappa(A^{GL}_{II}) \leq Cp^3.
\]

Since the null space of the matrices \( A^{SP} \), \( A^{GL} \), \( A_{c}^{SP} \), \( A_{c}^{GL} \) is spanned by the coefficient vector corresponding to the function \( u \equiv 1 \), we define \( \kappa(A^{SP}) \), \( \kappa(A^{GL}) \), \( \kappa(A_{c}^{SP}) \), and \( \kappa(A_{c}^{GL}) \) as the quotient of the largest and the smallest non-vanishing eigenvalue. With this understanding Theorem 2.2 and
Propositions 2.8, 2.10 imply

\[ \kappa(A^{\text{SP}}) \leq Cp^3, \quad \kappa(A^{\text{GL}}) \leq Cp^4(1 + \ln p), \]

\[ \kappa(A^{\text{SP}})_c \leq Cp, \quad \kappa(A^{\text{GL}})_c \leq Cp(1 + \ln p). \]

The left panel in Fig. 1 shows \( \kappa(A^{\text{GL}}), \kappa(A^{\text{GL}})_II, \kappa(A^{\text{GL}})_c \) as functions of the polynomial degree \( p \), and the right panel in Fig. 1 depicts \( \kappa(A^{\text{SP}}), \kappa(A^{\text{SP}})_II, \kappa(A^{\text{SP}})_c \). We clearly note that \( \kappa(A^{\text{SP}}) \) and \( \kappa(A^{\text{SP}})_II \) grow at the same rate as predicted; the growth is indeed \( O(p^3) \).

Our estimates for \( \kappa(A^{\text{SP}})_c \) and \( \kappa(A^{\text{GL}})_c \) differ by a factor \( (1 + \ln p) \). The presence of this logarithmic factor is shown in Fig. 2 by graphing \( p \mapsto \exp(\kappa(A^{\text{SP}})_c/p), \ p \mapsto \exp(\kappa(A^{\text{GL}})_c/p). \)

We now show that for meshes refined geometrically toward a vertex, the condition number of the global stiffness matrix for the Dirichlet problem is independent of the number of layers of geometric refinement as ascertained in Theorem 2.2.

**Example 3.2.** We consider the stiffness matrix corresponding to the problem

\[-\Delta u = f \quad \text{on } \Omega = (0,1)^2, \quad u = 0 \quad \text{on } \partial \Omega.\]

The meshes \( \mathcal{T}_L \) employed consist of quadrilaterals and are geometrically refined toward one of the vertices with grading factor \( \sigma = 0.15 \). The left panel of Fig. 3 shows the mesh \( \mathcal{T}_L \) for \( L = 2 \). For higher values of \( L \), further self-similar refinement is performed in the small square of size \( \sigma^L \times \sigma^L \) in Fig. 3 until the element abutting on the vertex is a square of size \( \sigma^L \times \sigma^L \). The number of elements in \( \mathcal{T}_L \) is \( 2L + 1 \), and the ratio of smallest element size to largest element size is approximately \( \sigma^L = 0.15^L \). The shape functions employed are tensor products of the following one-dimensional
Fig. 2. \( \exp(\kappa(A_{c}^{GL})/p) \) and \( \exp(\kappa(A_{c}^{SP})/p) \) vs. \( p \).

Fig. 3. Geometric mesh \( \mathcal{F}_L \) for \( L=2 \) (left) and condition number dependence on \( p \) and \( L \) for Dirichlet boundary conditions (right).

shape functions (the “Babuška–Szabó” polynomials—cf. [23]):

\[
\varphi_0(x) = \frac{1}{2}(1 - x), \quad \varphi_1(x) = \frac{1}{2}(1 + x), \quad (3.1a)
\]

\[
\varphi_i(x) = \frac{1}{\|P_{i-1}\|_{L^2(I)}} \int_{I_{i-1}}^x P_{i-1}(t) \, dt, \quad i = 2, \ldots, p. \quad (3.1b)
\]

From Remark 2.3(3) we expect the condition number of the stiffness matrix to be independent of the number of layers \( L \) of geometric refinement. The condition number does, however, depend on the polynomial degree \( p \)—by [13] the condition number is \( O(p^4) \). The right panel of Fig. 3 lists the condition numbers of the global stiffness matrix in dependence on the polynomial degree \( p \) and the number of layers \( L \). We clearly see that the condition number is independent of the number of levels of geometric refinement as predicted by our analysis.
Our last example illustrates that the condition number of the stiffness matrix may be mesh dependent if boundary conditions other than those of Dirichlet type are considered. To that end, we consider meshes that are geometrically refined toward a “Neumann–Neumann” vertex, i.e., a vertex where two edges meet on which Neumann boundary conditions are prescribed.

**Example 3.3.** We consider the discretization of

$$-\Delta u = f \quad \text{on } \Omega = (0,1)^2, \quad u = 0 \quad \text{on } \Gamma_D, \quad \partial_{\nu} u = 0 \quad \text{on } \Gamma_N,$$

where \(\Gamma_N = \{(x,y) \in \partial\Omega \mid y = 0 \text{ or } x = 1\}\) and \(\Gamma_D = \partial\Omega \setminus \Gamma_N\). The shape functions employed are the Babuška–Szabó shape functions of (3.1), and the meshes are again the geometric meshes \(T_L\) of Example 3.2. The general bound (1.8) of [6] implies

$$\kappa \leq C_p|T_L| \leq C_pL^2.$$  

Fig. 4 shows indeed that the condition number depends on \(L\) and even suggests the quadratic dependence on \(L\).

### 4. Discrete norms

The purpose of this section is proving the following theorem, which was repeatedly used in the preceding analysis.

**Theorem 4.1.** Let \(I = (-1,1)\). For \(p \in \mathbb{N}\) let \((x_i)_{i=0}^p\) be the Gauss–Lobatto points. Then there exists \(C > 0\) independent of \(p\) such that for all \(u \in \mathcal{P}_p(I)\) there holds upon writing \(\|u\|_2 = \left\{\sum_{i=0}^p |u(x_i)|^2\right\}^{1/2} :\)

(i) \(\|u\|_{L^\infty(I)} \leq \|u\|_2,\)

(ii) \(\|u\|_{L^2(I)} \leq Cp^{-1/2}\|u\|_2,\)

(iii) \(\|u\|_{H^{1/2}(I)} \leq C\|u\|_2,\)

(iv) \(\|u\|_{H^1(I)} \leq Cp\|u\|_2,\)

(v) \(\|u\|_{H^{1/2}_0(I)} \leq C\|u\|_2\) if additionally \(u(\pm 1) = 0,\)

(vi) \(\|u\|_2^2 \leq Cp \int_{-1}^{1} (1 - x^2)^{-1/2} u^2(x) \, dx \leq Cp\|u\|_{H^{1/2}(I)}^2.\)

Furthermore, the bounds are sharp with respect to the spectral order \(p\).
The remainder of this section is devoted to the proof of Theorem 4.1. At the heart of the ensuing analysis is the demonstration (Proposition 4.9) that the matrix \( H \in \mathbb{R}^{(p+1) \times (p+1)} \) given by
\[
H_{ij} = \frac{1}{p(p+1)} \frac{1}{P_p^2(x_i)} l_i'(x_j), \quad 0 \leq i, j \leq p
\]
(4.1)
satisfies \( \|H\|_2 = O(1) \) uniformly in \( p \).

4.1. Analysis of the matrix \( H \)

Let \( x_i, i = 0, \ldots, p \) be the Gauss–Lobatto points. As they lie in the interval \([-1, 1]\), they may also be written in the form
\[
x_i = \cos \phi_i.
\]
(4.2)
We start by recalling the following lemma, due to Sündermann [21,22]:

**Lemma 4.2.** For \( p \in \mathbb{N} \) the Gauss–Lobatto points \( x_i, i = 0, \ldots, p \) are of form (4.2) with
\[
\frac{i}{p+1/2} \pi \leq \phi_i \leq \frac{i+1/2}{p+1/2} \pi, \quad i = 0, \ldots, p.
\]
Together with the weights
\[
\rho_i = \frac{1}{p(p+1)P_p^2(x_i)}, \quad i = 0, \ldots, p,
\]
(4.3)
the Gauss–Lobatto nodes \((x_i)_{i=0}^p\) generate a quadrature rule of the form \( \int_{-1}^{1} f(x) \, dx \approx \sum_{i=0}^{p} \rho_i f(x_i) \). It is known that all weights \( \rho_i \) are finite (implying that \( P_p(x_i) \neq 0 \)) and that the quadrature rule is exact for polynomial of degree \( 2p-1 \):
\[
\int_{-1}^{1} f(x) \, dx = \sum_{i=0}^{p} \rho_i f(x_i) \quad \forall f \in \mathcal{P}_{2p-1}(-1,1).
\]
Additionally, it generates a norm that is equivalent to the \( L^2 \) norm on the spaces \( \mathcal{P}_p(I) \), \( \mathcal{Q}_p(S) \) (see, e.g., [7]):

**Lemma 4.3.** There holds for \( I = (-1,1) \)
\[
\sum_{i=0}^{p} \rho_i |u(x_i)|^2 \leq \|u\|_{L^2(I)}^2 \leq 3 \sum_{i=0}^{p} \rho_i |u(x_i)|^2 \quad \forall u \in \mathcal{P}_p(I),
\]
\[
\sum_{i=0}^{p} \sum_{j=0}^{p} \rho_i \rho_j |u(x_i,x_j)|^2 \leq \|u\|_{L^2(I \times I)}^2 \leq 9 \sum_{i=0}^{p} \sum_{j=0}^{p} \rho_i |u(x_i,x_j)|^2 \quad \forall u \in \mathcal{Q}_p(I \times I).
\]
From well-known identities for Jacobi polynomials (cf., e.g., [8, eq. (2.3.25)]), we have the following representation for the values of \( l_i'(x_j) \):
Lemma 4.4. The polynomials $l_i$ defined in (1.10) satisfy
\[
l'_i(x_j) = \begin{cases} 
\frac{1}{4} p(p + 1) & \text{if } i = j = 0, \\
-\frac{1}{4} p(p + 1) & \text{if } i = j = p, \\
0 & \text{if } i = j \in \{1, \ldots, p - 1\}, \\
P_p(x_j) & \text{else.}
\end{cases}
\]

We observe that there holds
\[
x_i - x_j = \cos \varphi_i - \cos \varphi_j = -2 \sin \left( \frac{\varphi_i + \varphi_j}{2} \right) \sin \left( \frac{\varphi_i - \varphi_j}{2} \right), \quad 0 \leq i, j \leq p. \tag{4.4}
\]

This observation motivates the following shorthand notation, which we will use extensively in the remainder of this section: For $i, j \in \{0, \ldots, p\}$ we set
\[
s(i,j) := (p + 1/2) \sin \left( \frac{\varphi_i + \varphi_j}{2} \right), \quad s(i, -j) := (p + 1/2) \sin \left( \frac{\varphi_i - \varphi_j}{2} \right).
\]

Lemma 4.5. There is $C > 0$ independent of $p$ such that
\[
1 < \frac{3}{4} \sqrt{3} \leq s(i,i), \quad i \in \{1, \ldots, p - 1\}, \tag{4.5}
\]
\[
C^{-1} \leq s(i,0), \quad i \in \{1, \ldots, p - 1\}, \tag{4.6}
\]
\[
C^{-1} \min \{i, p - i\} \leq s(i,i) \leq C \min \{i, p - i\}, \quad i \in \{0, \ldots, p\}, \tag{4.7}
\]
\[
C^{-1} |i - j| \leq |s(i, -j)| \leq C |i - j|, \quad i, j \in \{0, \ldots, p\}, \tag{4.8}
\]
\[
C^{-1} \min \{i + j, 2p - (i + j)\} \leq s(i,j) \leq C \min \{i + j, 2p - (i + j)\}, \quad i, j \in \{0, \ldots, p\}, \tag{4.9}
\]
\[
|s(i,i) - s(i,j)| \leq C |s(i, -j)|, \quad i, j \in \{0, \ldots, p\}, \tag{4.10}
\]
\[
s(i,i) - s(j,j) = 2 \cos \left( \frac{\varphi_i + \varphi_j}{2} \right) s(i, -j), \tag{4.11}
\]
\[
s(i,i) + s(j,j) = 2 \cos \left( \frac{\varphi_i - \varphi_j}{2} \right) s(i,j). \tag{4.12}
\]

Proof. The assertions of the lemma follows from Lemma 4.2 and trigonometric identities. \qed

Lemma 4.6. There are constants $c_1, c_2$ independent of $p$ such that
\[
\frac{c_1}{1 + s(i,i)} \leq P^2_p(x_i) \leq \frac{c_2}{1 + s(i,i)}, \quad i \in \{0, \ldots, p\}.
\]
**Proof.** By symmetry properties of the Gauss–Lobatto points and the functions $P_p$, $s(i,i)$, it suffices to see the estimate for $i \in \{0, \ldots, \lfloor p/2 \rfloor + 1\}$. Since there exists $\varepsilon > 0$ such that $\varphi_i \leq \pi - \varepsilon$ for $i \leq \lfloor p/2 \rfloor + 1$, we base our analysis on the following asymptotic expansion, [24, Theorem 8.21.6]:

$$P_p^2(\cos \varphi) = \frac{\varphi}{\sin \varphi} J_0^2((p + 1/2)\varphi) + O(p^{-3/2});$$

(4.13)

where the implied constant in the remainder $O(p^{-3/2})$ is bounded uniformly in $\varphi \in (0, \pi - \varepsilon]$. Key to estimating $P_p^2(\cos \varphi)$ is the following result about the behavior of the Bessel function $J_0$:

**Assertion:** Set $X := \mathbb{N}_0 + [0,1/2] := \{x \geq 0 \mid \text{there exists } m \in \mathbb{N}_0 \text{ s.t. } x - m \in [0,1/2]\}$. Then there exist $c_1, c_2 > 0$ such that

$$\frac{c_1}{1 + x\pi} \leq J_0^2(x\pi) \leq \frac{c_2}{1 + x\pi} \quad \forall x \in X.$$  

(4.14)

In order to show (4.14) for large values of $x$, we employ an asymptotic expansion of the Bessel function $J_0$ for large arguments that is due to Poisson, [26, Chapter 7.1]:

$$J_0(z) = \sqrt{\frac{2}{\pi z}} \left[ \cos(z - \pi/4) + O(z^{-1}) \right], \quad z \to \infty.$$  

This asymptotic expansion leads to

$$J_0^2(x\pi) = \frac{2}{\pi^2 x} \cos^2((x - 1/4)\pi) + O(x^{-2}), \quad x \to \infty.$$  

(4.15)

Since $x \in X$ implies $1/2 \leq \cos^2[(x - 1/4)\pi] \leq 1$, we get the existence of $x_0, c_1, c_2 > 0$ such that

$$\frac{c_1}{1 + x\pi} \leq J_0^2(x\pi) \leq \frac{c_2}{1 + x\pi} \quad \forall x \in X \cap (x_0, \infty).$$  

(4.16)

It remains to see (4.14) for $x \in X \cap [0,x_0]$. By continuity of $J_0$ on the compact set $X \cap [0,x_0]$, it suffices to show that $J_0(x\pi) \neq 0$ for $x \in X \cap [0,x_0]$. This is a result due to Schafheitlin, [19], a proof of which can be found in [26, Chapter 15.32]. The proof of the Assertion (4.14) is complete.

Since Lemma 4.2 implies

$$(p + 1/2)\varphi_i \in [i\pi, (i + 1/2)\pi],$$

the Assertion (4.14) gives

$$\frac{c_1}{i + 1} \leq J_0^2((p + 1/2)\varphi_i) \leq \frac{c_2}{i + 1} \quad \forall i \in \{0, \ldots, \lfloor p/2 \rfloor + 1\}$$

for some constants $c_1, c_2$ independent of $p$ and $i$. For $p \geq p_0$ with $p_0$ sufficiently large the claim of the lemma now follows from this by combining (4.7) of Lemma 4.5 (implying $i \sim s(i,i)$) and the expansion (4.13). For the remaining finitely many cases $p \in \{1, \ldots, p_0\}$, we use the fact that the quadrature weights $\rho_i$ are known to be finite for all $i$ and $p$, which guarantees the existence of $\delta > 0$ with $0 < \delta \leq P_p^2(x_i) \leq 1$ for $p \in \{1, \ldots, p_0\}$.
Lemma 4.7. Let \( A = (A_{ij})_{i,j=0}^{p} \in \mathbb{R}^{(p+1) \times (p+1)} \), \( B = (B_{ij})_{i,j=0}^{p} \in \mathbb{R}^{(p+1) \times (p+1)} \), be given by

\[
A_{ij} = \begin{cases} 
\frac{1}{s(i,j)} & \text{if } i \neq j, \\
0 & \text{if } i = j,
\end{cases} \quad B_{ij} = \begin{cases} 
1 & \text{if } i = j \in \{0, p\}, \\
\frac{1}{s(i,j)} & \text{else.}
\end{cases}
\]

Then there exists \( C > 0 \) independent of \( p \) such that \( \|A\|_2 \leq C \) and \( \|B\|_2 \leq C \).

Proof. The assertion \( \|A\|_2 \) follows from a perturbation argument. To that end, set \( \tilde{\phi}_i = i\pi/(p + 1/2), i = 0, \ldots, p \) and define the matrix \( \tilde{A} \) by

\[
\tilde{A}_{ij} = \frac{1}{p + 1/2} \frac{1}{(\tilde{\phi}_i - \tilde{\phi}_j)/2} \quad \text{for } i \neq j \quad \text{and} \quad \tilde{A}_{ii} = 0.
\]

By [12, Theorem 294]

\[
\|\tilde{A}\|_2 \leq 2.
\]

The result then follows from the triangle inequality if we can show that \( \|A - \tilde{A}\|_2 \) is bounded uniformly in \( p \). For that, it suffices to show by [10, Cor. 2.3.2] the existence of \( C > 0 \) independent of \( p \) such that

\[
\max_{j \in \{0, \ldots, p\}} \sum_{i=0}^{p} |A_{ij} - \tilde{A}_{ij}| \leq C \quad \text{and} \quad \max_{j \in \{0, \ldots, p\}} \sum_{i=0}^{p} |A_{ij} - \tilde{A}_{ij}| \leq C. \tag{4.17}
\]

For \( i \neq j \), we calculate

\[
|A_{ij} - \tilde{A}_{ij}|
= \left| \frac{1}{p + 1/2} \left( \frac{(\tilde{\phi}_i - \tilde{\phi}_j)/2 - \sin((\varphi_i - \varphi_j)/2)}{(\tilde{\phi}_i - \tilde{\phi}_j)/2 \sin((\varphi_i - \varphi_j)/2)} \right) \right|
= \left| \frac{1}{p + 1/2} \left[ (\tilde{\phi}_i - \tilde{\phi}_j)/2 - (\varphi_i - \varphi_j)/2 \right] + (\varphi_i - \varphi_j)/2 - \sin((\varphi_i - \varphi_j)/2) \right|.
\]

By (4.8) of Lemma 4.5 there is \( c_1 > 0 \) such that the denominator satisfies

\[
\left| \frac{1}{p + 1/2} (\tilde{\phi}_i - \tilde{\phi}_j)/2 \sin \left( \frac{\varphi_i - \varphi_j}{2} \right) \right| \geq c_1 (i - j)^2 \geq c_1' (i - j)^2.
\]

Expanding the sine function in the numerator in a Taylor series and exploiting that \( \varphi_i = \tilde{\phi}_i + O(p^{-1}) \) and employing again Lemma 4.5 we get

\[
|A_{ij} - \tilde{A}_{ij}| \leq \frac{C (i - j)^3}{c_1(i - j)^2/p} \leq C(i - j)^2 + p^{-1},
\]

where \( C \) is independent of \( i, j, \) and \( p \). From this, we can readily infer the estimates of (4.17) and thus complete the proof that \( \|A\|_2 \) is bounded uniformly in \( p \).
We now turn to the proof that $\|B\|_2$ is bounded uniformly in $p$. From (4.9) of Lemma 4.5 we get the existence of $C > 0$ independent of $p$ such that
\[ 0 \leq B_{ij} \leq C \frac{1}{1 + \min\{i + j, 2p - (i + j)\}} \leq C \left[ \frac{1}{1 + i + j} + \frac{1}{1 + (p - i) + (p - j)} \right]. \]

Hardy et al. [12, Theorem 294] assert that the matrices $D, E$ with
\[ D_{ij} = \frac{1}{1 + i + j}, \quad E_{ij} = \frac{1}{1 + (p - i) + (p - j)} \]
satisfy $\|D\|_2 = \|E\|_2 \leq \pi$. Hence, $\|B\|_2$ is bounded uniformly in $p$.  

Lemma 4.8. Let $A, B, C \in \mathbb{R}^{(p+1) \times (p+1)}$ satisfy
\[ |A_{ij}| |C_{ij} - C_{jj}| \leq B_{ij} \quad \forall i, j. \]
Then the matrix $D \in \mathbb{R}^{(p+1) \times (p+1)}$ given by $D_{ij} = A_{ij}C_{ij}$ satisfies
\[ \|D\|_2 \leq \|B\|_2 + \max_j |C_{jj}| \|A\|_2. \]

Proof. We split $D = D_1 + D_2$ by writing $D_{ij} = C_{jj}A_{ij} + (C_{ij} - C_{jj})A_{ij}$. We now readily observe that $\|D_1\|_2 \leq \max_j |C_{jj}| \|A\|_2$ and that by the hypothesis $\|D_2\|_2 \leq \|B\|_2$.  

Proposition 4.9. Let the matrix $H \in \mathbb{R}^{(p+1) \times (p+1)}$ be given by (4.1). Then it has the form
\[ H_{ij} = \frac{1}{p(p+1)} \begin{cases} \frac{1}{4} p(p+1) & \text{if } i = j = 0, \\ -\frac{1}{4} p(p+1) & \text{if } i = j = p, \\ 0 & \text{if } i = j \text{ and } i \notin \{0, p\}, \\ \frac{1}{P_p(x_i)P_p(x_j)} & \text{else} \end{cases} \]
and $\|H\|_2 \leq C$ for some $C > 0$ independent of $p$.  

Proof. Lemma 4.4 implies immediately that the matrix $H$ of (4.1) has the form given in the statement of Proposition 4.9. In order to show that $\|H\|_2$ is bounded uniformly in $p$, we start by showing the following.

Assertion: There exists $C > 0$ independent of $p$ such that $\|\tilde{H}\|_2 \leq C$, where $\tilde{H}$ is given by
\[ \tilde{H}_{ij} = \frac{\sqrt{1 + s(i, i)\sqrt{1 + s(j, j)}}}{s(i, -j)s(i, j)} \quad \text{if } i \neq j, \]
\[ 0 \quad \text{else}. \]

Proof of the Assertion: Define the matrix
\[ C_{ij} = \begin{cases} 0 & \text{if } i = j \in \{0, p\}, \\ \frac{\sqrt{1 + s(i, i)\sqrt{1 + s(j, j)}}}{s(i, j)} & \text{else}. \end{cases} \]
Let $A$, $B$ be the matrices of Lemma 4.7. We first show the existence of $C' > 0$ independent of $p$ with

$$|A_{ij}(C_{ij} - C_{jj})| \leq C'B_{ij} \quad \forall i, j. \quad (4.18)$$

Clearly, it suffices to show this for $i \neq j$. A direct calculation shows

$$|C_{p0} - C_{00}| = |C_{p0}| = \frac{1}{s(p, 0)} = \frac{1}{p + 1/2} \leq 1.$$

Using (4.5), (4.6) of Lemma 4.5, we estimate for $i \in \{0, p\}$:

$$|C_{i0} - C_{00}| = \frac{\sqrt{1 + s(i, i)}}{s(i, 0)} \leq 2\sqrt{s(i, i)} \leq 2\sqrt{\frac{s(i, 0)}{s(i, 0)}} \leq 2\sqrt{\frac{1}{p + 1/2}} \leq C'$$

for some $C' > 0$ independent of $p$. Similarly, we have

$$|C_{op} - C_{pp}| = |C_{op}| = \frac{1}{s(0, p)} = \frac{1}{p + 1/2} \leq 1,$$

and for $i \in \{0, p\}$ we estimate

$$|C_{ip} - C_{pp}| = \frac{\sqrt{1 + s(i, i)}}{s(i, p)} \leq 2\sqrt{s(i, i)} \leq 2\sqrt{\frac{s(i, i)}{s(i, p)}} \leq 2\sqrt{\frac{1}{p + 1/2}} \leq C'.$$

As $A_{i0} = B_{i0}$ for $i \neq 0$ and $A_{ip} = -B_{ip}$ for $i \neq p$, these last two estimates readily imply that (4.18) holds for $j \in \{0, p\}$. Let us now see that (4.18) holds for $j \in \{1, \ldots, p - 1\}$ and $i \neq j$. We have

$$C_{ij} - C_{jj} = \frac{\sqrt{1 + s(j, j)}}{s(i, j)} (\sqrt{1 + s(i, i)} - \sqrt{1 + s(j, j)}) + \frac{\sqrt{1 + s(j, j)}}{s(i, j)s(j, j)} (s(j, j) - s(i, j)) =: C_1 + C_2.$$

We estimate further

$$|C_1| \leq \frac{\sqrt{1 + s(j, j)}}{s(i, j)} \frac{|s(i, i) - s(j, j)|}{\sqrt{1 + s(i, i) + \sqrt{1 + s(j, j)}}} \leq 2\frac{|s(i, -j)|}{s(i, j)}$$

by (4.11) of Lemma 4.5. Additionally, (4.10) of Lemma 4.5 gives the bound $|s(j, j) - s(i, j)| \leq C'|s(i, -j)|$ for some $C' > 0$ independent of $p$ thus leading to

$$|C_2| \leq \frac{\sqrt{1 + s(j, j)}}{s(i, j)s(j, j)} \sqrt{1 + s(j, j)} C'|s(i, -j)| \leq 2C' \frac{|s(i, -j)|}{s(i, j)}.$$
These two bounds on $C_1, C_2$ allows us to conclude that (4.18) holds for $j \in \{1, \ldots, p - 1\}$ as well. Finally, we note that (4.5) of Lemma 4.5 implies

$$|C_j| \leq 2.$$  

Thus, Lemmata 4.7 and 4.8 together imply the existence of $C > 0$ independent of $p$ such that $\|\hat{H}\|_2 \leq C$. This concludes the proof of the assertion.

Let us now introduce the diagonal matrices $D$ and $H'$ by

$$D_{ij} = \frac{1}{\sqrt{2}P_p(x_i)} \frac{1}{\sqrt{1 + s(i, i)}} \frac{p + 1/2}{\sqrt{p(p + 1)}}, \quad H'_{ij} = \frac{1}{4}[\delta_{i0}\delta_{j0} - \delta_{i,p}\delta_{j,p}].$$

From (4.4) we have $(p + 1/2)s^2(x_i - x_j) = 2s(i, -j)s(i, j)$ and therefore

$$H = D\hat{H}D + H'.$$

Lemma 4.6 gives the existence of $C¿_0$ independent of $p$ such that $\|D\|_2 \leq C$, and the claim of the proposition now follows from $\|H\|_2 \leq \|D\|_2 \|\hat{H}\|_2 \|D\|_2 + \|H'\|_2$. \hfill $\square$

A consequence of Proposition 4.9 is

**Corollary 4.10.** There exists $C > 0$ independent of $p$ such that

$$C^{-1}p^4 \leq \sum_{j \neq 0} \frac{1}{(1 - x)^2} \frac{1}{P_p^2(x_j)} \leq \max_{i \in \{0, \ldots, p\}} \sum_{j \neq i} \frac{1}{(1 - x_j)^2} \frac{1}{P_p^2(x_i)P_p^2(x_j)} \leq Cp^4.$$  

**Proof.** From the representation of the matrix $H$ in Proposition 4.9, it follows that the upper bound is proved if we can show that

$$\max_{j \in \{0, \ldots, p\}} \sum_{j = 0}^p H_{ij}^2 \leq \|H\|_2^2. \quad (4.19)$$

In order to see (4.19), fix $i \in \{0, \ldots, p\}$ and define the vector $u$ by $u_j := H_{ij}$. We get

$$\left(\sum_{j = 0}^p H_{ij}^2\right)^2 = \left(\sum_{j = 0}^p H_{ij}u_j\right)^2 \leq \|H\|_2^2\|u\|^2 = \|H\|_2^2\sum_{j = 0}^p H_{ij}^2,$$

implying (4.19). In order to see that the lower bound in the statement of Corollary 4.10 is also true, we compute

$$\sum_{j = 1}^p H_{ij}^2 = \frac{1}{(p(p + 1))^2} \sum_{j = 1}^p \frac{1}{P_p^2(x_j)} \frac{1}{(1 - x_j)^2} \geq \frac{1}{(p(p + 1/2))^2} \frac{1}{P_p^2(x_1)} \frac{1}{(1 - x_1)^2} \geq \frac{1}{4P_p^2(x_1)} \frac{1}{s^2(1, 0)s^2(1, 0)} \geq C$$

for some $C > 0$ independent of $p$. \hfill $\square$
4.2. Pointwise bounds in Gauss–Lobatto points

We start with a lemma:

**Lemma 4.11.** Let $0 = x_0 < x_1 < \cdots < x_N = \pi$ define a quasi-uniform mesh on $(0, \pi)$, i.e., there exists $c_1 > 0$ such that $c_1^{-1}N^{-1} \leq x_{i+1} - x_i \leq c_1N^{-1}$ for $i \in \{0, \ldots, N-1\}$. Assume $p \in \mathbb{N}$ satisfies $c_2^{-1}p \leq N \leq c_2p$ for some $c_2 > 0$. Then there exists $C > 0$ depending only on $c_1$, $c_2$ such that for all trigonometric polynomials of degree $p$ of the form $u(\theta) = \sum_{k=0}^{p} u_i \cos(k\theta)$

$$\frac{1}{p} \sum_{k=0}^{N} |u(x_k)|^2 \leq C\|u\|_{L^2(0, \pi)}^2.$$  

**Proof.** Let $Iu$ be the piecewise linear interpolant of $u$ on the mesh given by the points $x_i$. By the quasi-uniformity of the mesh, we have

$$\|Iu\|_{L^2(0, \pi)} - \|u\|_{L^2(0, \pi)} \leq \|u - Iu\|_{L^2(0, \pi)} \leq C \frac{1}{p} \|u'\|_{L^2(0, \pi)},$$

where $C$ depends only on $c_1$ and $c_2$. Noting that $\|Iu\|_{L^2(0, \pi)}^2 \sim \frac{1}{p} \sum_{k=0}^{N} |u(x_k)|^2$, where the implied constants in the $\sim$-notation again depend only on $c_1$, $c_2$, we conclude

$$\frac{1}{p} \sum_{k=0}^{N} |u(x_k)|^2 \leq C \left[ \|u\|_{L^2(0, \pi)}^2 + \frac{1}{p^2} \|u'\|_{L^2(0, \pi)}^2 \right].$$

The function $u$, being a trigonometric polynomial of degree $p$, satisfies the inverse estimate $\|u'\|_{L^2(0, \pi)} \leq p \|u\|_{L^2(0, \pi)}$. Thus, we arrive at

$$\frac{1}{p} \sum_{k=0}^{N} |u(x_k)|^2 \leq C\|u\|_{L^2(0, \pi)}^2. \quad \square$$

**Proposition 4.12.** Let $c_1 > 0$ be given and let $p$, $N$ satisfy $c_1^{-1}p \leq N \leq c_1p$. Let $x_0, \ldots, x_N$ be the $N+1$ Gauss–Lobatto points. Then there exists $C > 0$ depending only on $c_1$ such that

$$\sum_{i=0}^{N} |u(x_i)|^2 \leq Cp \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} u^2(x) \, dx \quad \forall u \in \mathcal{P}_p(-1, 1).$$

**Proof.** Let $u \in \mathcal{P}_p(-1, 1)$ and set $\theta_i = \arccos(x_i)$. Lemma 4.2 implies that the points $\theta_0, \ldots, \theta_N$ form a quasi-uniform mesh with meshsize $O(1/p)$ on $(0, \pi)$. We apply Lemma 4.11 to the trigonometric polynomial $\hat{u}(\theta) = u(\cos \theta)$ to obtain

$$\sum_{i=0}^{N} |u(x_i)|^2 = \sum_{i=0}^{N} |\hat{u}(\theta_i)|^2 \leq C p \|\hat{u}\|_{L^2(0, \pi)}^2 = C p \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} |u(x)|^2 \, dx. \quad \square$$
Remark 4.13. It is not essential to take the points $x_i$ as the Gauss–Lobatto points in Proposition 4.12. It is only required that a) the number of points be essentially proportional to $p$ and that b) the points $\theta_i = \arccos(x_i)$ be distributed quasi-uniformly on $(0, \pi)$. Hence, Proposition 4.12 also holds if the points $x_i$ are the Chebyshev points or the Gauss–Legendre points.

4.3. Proof of Theorem 4.1

With the polynomials $l_i$ defined in (1.10), we observe that every $u \in \mathcal{P}_p(I)$ can be written as

$$u(x) = \sum_{i=0}^{p} u_i l_i(x), \quad u_i := u(x_i).$$

For the remainder of this proof, $u$ will always be an element of $\mathcal{P}_p(I)$ and the corresponding $u_i$ are collected in the vector $u$.

Proof of Theorem 4.1(i): From [25] we have $\sum_{i=0}^{p} l_i^2(x) \leq 1$ on $I$ and the desired bound follows. Taking $u = l_0$ shows the sharpness of the bound.

Proof of Theorem 4.1(ii): From Lemma 4.3, we have

$$\|u\|^2_{L^2(I)} \leq 3 \sum_{i=0}^{p} \rho_i u_i^2 \leq 3 \max_i \|u\|^2.$$

Combining (4.3) and Lemma 4.6, we see that there is $C > 0$ such that $\rho_i \leq Cp^{-1}$, and the desired bound follows. Choosing $q = [p/2]$, we furthermore see that for some $C > 0$ independent of $p$ there holds

$$Cp^{-1} \leq \rho_q = \sum_{i=0}^{p} \rho_i l_q^2(x_i) \leq \|l_q\|^2_{L^2(I)},$$

showing that the result is sharp for the choice $u = l_q$.

Proof of Theorem 4.1(iii): It suffices to prove

$$\int_I \int_I \frac{|u(x) - u(y)|^2}{|x - y|^2} \, dx \, dy \leq C\|u\|_{L^2}^2.$$

Define

$$v(x, y) := \begin{cases} \frac{u(x) - u(y)}{x - y} & \text{if } y \neq x, \\ u'(x) & \text{if } y = x. \end{cases}$$

As $v \in \mathcal{P}_p(I \times I)$, Lemma 4.3 implies

$$\int_I \int_I |v(x, y)|^2 \, dx \, dy \leq 9 \sum_{i,j=0}^{p} \rho_i \rho_j |v(x_i, x_j)|^2$$

$$= 9 \sum_{i=0}^{p} \rho_i^2 |u'(x_i)|^2 + 9 \sum_{i=0}^{p} \sum_{j \neq i} \rho_i \rho_j \frac{|u(x_i) - u(x_j)|^2}{|x_i - x_j|^2}.$$
In order to estimate the first term, we write \( u'(x_i) = \sum_{j=0}^{p} u_j l'_j(x_i) \) and therefore get with the matrix \( H \) of (4.1)

\[
\sum_{i=0}^{p} \rho_i^2 |u'(x_i)|^2 = \sum_{i=0}^{p} \left| \sum_{j=0}^{p} \rho_i l'_j(x_i) u_j \right|^2 = \|Hu\|_2^2 \leq \|H\|_2^2 \|u\|_2^2.
\]

Proposition 4.9 thus asserts that for some \( C > 0 \) independent of \( p \) there holds

\[
\sum_{i=0}^{p} \rho_i^2 |u'(x_i)|^2 \leq C \|u\|_2^2.
\] (4.20)

Next, we estimate

\[
\sum_{i=0}^{p} \sum_{j \neq i} \rho_i \rho_j \frac{|u(x_i) - u(x_j)|^2}{|x_i - x_j|^2} \leq 4 \sum_{i=0}^{p} \sum_{j \neq i} \rho_i \rho_j \frac{1}{(x_i - x_j)^2} u_i^2
\]

\[
\leq 4 \frac{1}{p(p + 1)^2} \sum_{i=0}^{p} \sum_{j \neq i} \frac{1}{p^2 p^2(x_i)(x_j)(x_i - x_j)} u_i^2 \leq C ||u||_2^2
\]

by Corollary 4.10. This proves Theorem 4.1(iii). In order to see that this estimate is sharp, we consider the Chebyshev polynomial

\[
T_p(x) := \cos(p \arccos x).
\] (4.21)

Choosing \( u(x) = T_p(x) \), we have on the one hand

\[
\|u\|_2^2 = \sum_{i=0}^{p} |T_p(x_i)|^2 \leq (p + 1)
\]

and on the other hand

\[
\|T_p\|_{H^{1/2}(I)} \geq Cp
\]

for some \( C > 0 \) independent of \( p \) by [3, Lemma 6.1]. The assertion concerning the sharpness of Theorem 4.1(iii), now follows.

**Proof of Theorem 4.1(iv):** As \( u' \in \mathcal{P}_{p-1}(I) \), we have

\[
\|u'\|_{L^2(I)}^2 = \sum_{i=0}^{p} \rho_i |u'(x_i)|^2 \leq p(p + 1) \sum_{i=0}^{p} \rho_i^2 |u'(x_i)|^2 \leq Cp^2 \|u\|_2^2,
\]

where we used (4.20) in the last step. Again, in order to see that the result is sharp, we consider the polynomial \( u(x) = l_0(x) \). Then by the lower bound in Corollary 4.10:

\[
\|u'\|_{L^2(I)}^2 = \sum_{i=0}^{p} \rho_i |l'_0(x_i)|^2 \geq \frac{1}{p(p + 1)} \sum_{i=1}^{p-1} \rho_i^2 \frac{1}{p^2 p^2(x_j)(1 - x_j)} u_i^2 \geq Cp^2 \|u||_2^2.
\]

**Proof of Theorem 4.1(v):** It suffices to bound

\[
\int_I \frac{1}{1 - x^2} u_i^2(x) \, dx.
\]
As \( u(\pm 1) = 0 \), the expression \( u^2(x)/(1-x^2) \) is a polynomial of degree \( 2p - 2 \) and vanishes at the endpoints \( x = \pm 1 \). Thus, by the exactness of the Gauss–Lobatto quadrature rule

\[
\int_{-1}^{1} \frac{1}{1-x^2} u^2(x) \, dx = \sum_{i=1}^{p-1} \frac{\rho_i}{1-x_i^2} = \frac{(p+1/2)^2}{p(p+1)} \sum_{i=1}^{p-1} \frac{1}{P^2_p(x_i) s^2(i,i)} u_i.
\]

(4.22)

Combining Lemma 4.6 and (4.5) of Lemma 4.5, we conclude that there is \( C > 0 \) independent of \( p \) such that for all polynomials \( u \) with \( u(\pm 1) = 0 \)

\[
C^{-1} p^{-1} \| u \|_2^2 \leq \int_{-1}^{1} \frac{1}{1-x^2} u^2(x) \, dx \leq C \| u \|_2^2.
\]

(4.23)

The result of Theorem 4.1(v) follows. In order to see that the estimate of Theorem 4.1(v) is sharp, it suffices to consider the polynomial \( u(x) = T_p(x) - l(x) \) where \( T_p \) is the Chebyshev polynomial of \( (4.21) \) and \( l \) is a linear function such that \( u(\pm 1) = 0 \). Exploiting that \( |T_p(\pm 1)| = 1 \), it is easy to proceed in the same fashion as in the proof of Theorem 4.1(iii).

Proof of Theorem 4.1(vi): The first estimate follows from Proposition 4.12 with \( N = p \). The second bound follows easily from Sobolev’s embedding theorem \( H^{1/2}(I) \subset L^q(I) \), \( q \in [1, \infty) \). The choice \( u \equiv 1 \) shows that the bound is sharp.

References


