

Evaluation of the condition number in linear systems arising in finite element approximations

A. Ern¹, J.-L. Guermond^{1,2}

¹ CERMICS, Ecole nationale des ponts et chaussées, Champs sur Marne, 77455 Marne la Vallée Cedex 2, France e-mail: ern@cermics.enpc.fr

² LIMSI (CNRS-UPR 3152), BP 133, 91403, Orsay, France e-mail: guermond@limsi.fr

Draft version: 24th September 2004

Abstract This paper derives upper and lower bounds for the ℓ^p -condition number of the stiffness matrix resulting from the finite element approximation of a linear, abstract model problem. Sharp estimates in terms of the meshsize h are obtained. The theoretical results are applied to various finite element approximations of PDE's on quasi-uniform mesh families. For elliptic PDE's in variational form, the Euclidean condition number of the stiffness matrix scales as h^{-2} , while it scales as h^{-1} if the PDE is approximated in mixed form. When first-order PDE's are approximated using the Galerkin–Least Squares technique, the Euclidean condition number scales as h^{-1} . The same result is obtained for the ℓ^1 -condition number if the first-order PDE is approximated by means of a non-standard Galerkin technique in $L^1(\Omega)$. Numerical simulations are presented to illustrate the theoretical results.

Key words Finite Elements – Condition Number – Partial Differential Equations – Linear Algebra

Mathematics Subject Classification (1991): 65N30, 65F35

1 Introduction

The finite element method provides an extremely powerful tool to approximate partial differential equations arising in engineering sciences. The linear systems obtained using this technique are generally

very large and sparse; the most practical way to solve them is to resort to an iterative method. Since the convergence rate of such methods is strongly affected by the condition number of the system matrix (see, e.g., [8, 11]), it is important to assess this quantity as a function of the meshsize. For instance, second-order elliptic equations, e.g., a Laplacian, in variational form yield a stiffness matrix whose Euclidean condition number explodes as the reciprocal of the square of the meshsize. More generally, let $p \in [1, +\infty]$ and denote by $\|\cdot\|_p$ the ℓ^p -norm in \mathbb{R}^N , i.e., for all $\mathcal{W} \in \mathbb{R}^N$, set

$$\|\mathcal{W}\|_p = \left(\sum_{i=1}^N |\mathcal{W}_i|^p \right)^{\frac{1}{p}}, \quad (1)$$

if $1 \leq p < +\infty$ and $\|\mathcal{W}\|_\infty = \max_{1 \leq i \leq N} |\mathcal{W}_i|$. Use a similar notation for the associated matrix norm over $\mathbb{R}^{N,N}$. Then, upon defining the ℓ^p -condition number of a matrix $\mathcal{A} \in \mathbb{R}^{N,N}$ by

$$\kappa_p(\mathcal{A}) = \|\mathcal{A}\|_p \|\mathcal{A}^{-1}\|_p, \quad (2)$$

the objective of this paper is to give upper and lower bounds on $\kappa_p(\mathcal{A})$ when \mathcal{A} is the stiffness matrix associated with the finite element approximation of a linear, abstract model problem posed in Banach spaces.

This paper is organized as follows. Section 2 collects preliminary results. Necessary and sufficient conditions for wellposedness of an abstract model problem are stated, and the finite element setting for the approximation of this problem is introduced. Section 3 contains the main results of the paper. Section 4 presents various applications to finite element approximations of PDE's. Elliptic PDE's either in variational or in mixed form are first considered. Then, first-order PDE's approximated using either the Galerkin–Least Squares (GaLS) technique or a non-standard Galerkin technique in $L^1(\Omega)$ are analyzed. Numerical illustrations are reported in Section 5. Finally, Appendix A collects technical results concerning norm equivalence constants and the existence of large-scale discrete functions in finite element spaces.

2 Preliminaries

2.1 Wellposedness

Let W and V be two real Banach spaces equipped with some norms, say $\|\cdot\|_W$ and $\|\cdot\|_V$, respectively. Consider a linear bounded operator

$$A : W \longrightarrow V.$$

Recall that as a consequence of the Open Mapping Theorem and the Closed Range Theorem [12], the following holds:

Lemma 1 *The following statements are equivalent:*

- (i) *A is bijective.*
- (ii) *There exists a constant $\alpha > 0$ such that*

$$\forall w \in W, \quad \|Aw\|_V \geq \alpha \|w\|_W, \quad (3)$$

$$\forall v' \in V', \quad (A^T v' = 0) \implies (v' = 0). \quad (4)$$

Another way of interpreting A consists of introducing the bilinear form $a \in \mathcal{L}(W \times V'; \mathbb{R})$ such that

$$\forall (w, v') \in W \times V', \quad a(w, v') = \langle v', Aw \rangle_{V', V}, \quad (5)$$

where $\langle \cdot, \cdot \rangle_{V', V}$ denotes the duality pairing. Owing to a standard corollary of the Hahn–Banach Theorem, for all $f \in V$ and for all $w \in W$, $Aw = f$ if and only if $a(w, v') = \langle v', f \rangle_{V', V}$ for all $v' \in V'$. Then, a reformulation of Lemma 1, henceforth referred to as the BNB Theorem [1, 10, 6], is the following:

Theorem 1 (Banach–Nečas–Babuška) *The following statements are equivalent:*

- (i) *For all $f \in V$, the problem*

$$\begin{cases} \text{Seek } u \in W \text{ such that} \\ a(u, v') = \langle v', f \rangle_{V', V}, \quad \forall v' \in V', \end{cases} \quad (6)$$

is well-posed;

- (ii) *There exists a constant $\alpha > 0$ such that*

$$\inf_{w \in W} \sup_{v' \in V'} \frac{a(w, v')}{\|w\|_W \|v'\|_{V'}} \geq \alpha, \quad (7)$$

$$\forall v' \in V', \quad (\forall w \in W, a(w, v') = 0) \implies (v' = 0). \quad (8)$$

If V is reflexive, the above setting is unchanged if V is substituted by V' and V' by V . As an illustration of a nonreflexive situation, the reader may think of $W = W^{1,1}(\Omega)$, $V = L^1(\Omega)$, $V' = L^\infty(\Omega)$, and $A : W \ni u \mapsto u + u_x \in V$.

2.2 The finite element setting

Let Ω be an open domain in \mathbb{R}^d . Let m be a positive integer. In the sequel, we assume that W and V are Banach spaces of \mathbb{R}^m -valued functions on Ω . For $p \in [1, +\infty]$, equip $[L^p(\Omega)]^m$ with the norm $\|w\|_{L^p(\Omega)} = (\int_{\Omega} \sum_{i=1}^m |w_i|^p)^{\frac{1}{p}}$ if $p \neq \infty$ and for $p = \infty$, set $\|w\|_{L^\infty(\Omega)} = \max_{1 \leq i \leq m} \text{ess sup}_{\Omega} |w_i|$. Let $(w, v)_{L^2(\Omega)} = \int_{\Omega} \sum_{i=1}^m w_i v_i$ denote the $[L^2(\Omega)]^m$ -inner product. Likewise, for all measurable subset $K \subset \Omega$, we denote $(w, v)_{L^2(K)} = \int_K \sum_{i=1}^m w_i v_i$.

To construct an approximate solution to (6), we introduce a family of meshes of Ω that we denote by $\{\mathcal{T}_h\}_{h>0}$. The parameter h refers to the maximum meshsize, i.e., $h = \max_{K \in \mathcal{T}_h} h_K$ where $h_K = \text{diam}(K)$. Let W_h and V_h be finite-dimensional approximation spaces based on the mesh \mathcal{T}_h . These spaces are meant to approximate W and V' respectively. Let $p \in [1, +\infty]$ and denote by p' its conjugate, i.e., $\frac{1}{p} + \frac{1}{p'} = 1$ with the convention that $p' = 1$ if $p = +\infty$ and $p' = +\infty$ if $p = 1$. We assume hereafter that $\dim(W_h) = \dim(V_h)$ and that there is $p \in [1, +\infty]$ such that $W_h \subset [L^p(\Omega)]^m$ and $V_h \subset [L^{p'}(\Omega)]^m$. The spaces W_h and V_h are equipped with some norms, say $\|\cdot\|_{W_h}$ and $\|\cdot\|_{V_h}$, respectively.

Let $A : W \rightarrow V$ be an isomorphism. Problem (6) is approximated by replacing the spaces W and V' by their finite-dimensional counterparts, yielding the approximate problem:

$$\begin{cases} \text{Seek } u_h \in W_h \text{ such that} \\ a_h(u_h, v_h) = \langle v_h, f_h \rangle_{V', V}, \quad \forall v_h \in V_h. \end{cases} \quad (9)$$

Problem (9) involves an approximation a_h to the bilinear form a and an approximation f_h to the data f . Henceforth, we assume

$$\inf_{w_h \in W_h} \sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|w_h\|_{W_h} \|v_h\|_{V_h}} > 0.$$

This, together with the fact that $\dim(W_h) = \dim(V_h)$, implies that the discrete problem (9) has a unique solution.

Let $N = \dim(W_h) = \dim(V_h)$. Assume we are given a basis for V_h , say $\{\varphi_1, \dots, \varphi_N\}$. The elements in this basis are hereafter referred to as the global shape functions of V_h . Likewise let $\{\psi_1, \dots, \psi_N\}$ be the global shape functions in W_h . For a function $v_h \in V_h$, denote by $\mathcal{V} \in \mathbb{R}^N$ the coordinate vector of v_h relative to the basis $\{\varphi_1, \dots, \varphi_N\}$, i.e., $v_h = \sum_{i=1}^N \mathcal{V}_i \varphi_i \in V_h$. Denote by $C_{V_h} : V_h \rightarrow \mathbb{R}^N$ the linear operator that maps vectors in V_h to their coordinate vectors in \mathbb{R}^N , i.e., $C_{V_h} v_h = \mathcal{V}$. Similarly, denote by $C_{W_h} : W_h \rightarrow \mathbb{R}^N$ the operator

that maps vectors in W_h to their coordinate vectors in \mathbb{R}^N . It is clear that both C_{V_h} and C_{W_h} are isomorphisms. Denote by $(\cdot, \cdot)_N$ the Euclidean scalar product in \mathbb{R}^N .

Define the so-called stiffness matrix \mathcal{A} with entries $(a_h(\psi_j, \varphi_i))_{1 \leq i, j \leq N}$. This definition is such that $(C_{V_h} v_h, \mathcal{A} C_{W_h} w_h)_N = a_h(w_h, v_h)$ for all $(w_h, v_h) \in W_h \times V_h$. The discrete problem (9) yields the linear system:

$$\begin{cases} \text{Seek } \mathcal{U} \in \mathbb{R}^N \text{ such that} \\ \mathcal{A}\mathcal{U} = \mathcal{F}, \end{cases} \quad (10)$$

where the entries of \mathcal{F} are $\mathcal{F}_i = \langle \varphi_i, f_h \rangle_{V', V}$ for $1 \leq i \leq N$. The solution u_h to (9) is then $u_h = C_{W_h}^{-1} \mathcal{U}$.

2.3 Norm equivalence constants

Since W_h and V_h are finite-dimensional and since C_{W_h} and C_{V_h} are isomorphisms, it is legitimate to introduce the positive constants (depending on h)

$$\mu_{s, \min} = \inf_{w_h \in W_h} \frac{\|w_h\|_{L^p(\Omega)}}{\|C_{W_h} w_h\|_p}, \quad \mu_{s, \max} = \sup_{w_h \in W_h} \frac{\|w_h\|_{L^p(\Omega)}}{\|C_{W_h} w_h\|_p}, \quad (11)$$

$$\mu_{t, \min} = \inf_{v_h \in V_h} \frac{\|v_h\|_{L^{p'}(\Omega)}}{\|C_{V_h} v_h\|_{p'}}, \quad \mu_{t, \max} = \sup_{v_h \in V_h} \frac{\|v_h\|_{L^{p'}(\Omega)}}{\|C_{V_h} v_h\|_{p'}}. \quad (12)$$

These constants are such that

$$\forall w_h \in W_h, \quad \mu_{s, \min} \|\mathcal{W}\|_p \leq \|w_h\|_{L^p(\Omega)} \leq \mu_{s, \max} \|\mathcal{W}\|_p, \quad (13)$$

$$\forall v_h \in V_h, \quad \mu_{t, \min} \|\mathcal{V}\|_{p'} \leq \|v_h\|_{L^{p'}(\Omega)} \leq \mu_{t, \max} \|\mathcal{V}\|_{p'}, \quad (14)$$

with $\mathcal{W} = C_{W_h} w_h$ and $\mathcal{V} = C_{V_h} v_h$. Henceforth, we denote

$$\kappa_{s, p} = \frac{\mu_{s, \max}}{\mu_{s, \min}}, \quad (15)$$

$$\kappa_{t, p'} = \frac{\mu_{t, \max}}{\mu_{t, \min}}. \quad (16)$$

It is possible to estimate $\mu_{s, \min}$ and $\mu_{s, \max}$ (resp. $\mu_{t, \min}$ and $\mu_{t, \max}$) when W_h (resp. V_h) is a finite element space and the global shape functions are such that their support is restricted to a number of mesh cells that is uniformly bounded with respect to the meshsize. For instance, if the mesh family $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform, $\kappa_{s, p}$ and $\kappa_{t, p'}$ are uniformly bounded with respect to h ; see Appendix A.

3 Bounds on $\kappa_p(\mathcal{A})$

The goal of this section is to derive upper and lower bounds for the ℓ^p -condition number of the stiffness matrix \mathcal{A} .

3.1 Main results

Introduce the following notation:

$$\alpha_{p,h} = \inf_{w_h \in W_h} \sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|w_h\|_{L^p(\Omega)} \|v_h\|_{L^{p'}(\Omega)}}, \quad (17)$$

$$\omega_{p,h} = \sup_{w_h \in W_h} \sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|w_h\|_{L^p(\Omega)} \|v_h\|_{L^{p'}(\Omega)}}. \quad (18)$$

A first result is the following:

Theorem 2 *Under the above assumptions,*

$$\forall h, \quad \kappa_{s,p}^{-1} \kappa_{t,p'}^{-1} \frac{\omega_{p,h}}{\alpha_{p,h}} \leq \kappa_p(\mathcal{A}) \leq \kappa_{s,p} \kappa_{t,p'} \frac{\omega_{p,h}}{\alpha_{p,h}}. \quad (19)$$

Proof (1) Upper bound on $\|\mathcal{A}\|_p$. Consider $\mathcal{W} \in \mathbb{R}^N$. Then, owing to definition (18) and using the notation $C_{V_h} v_h = \mathcal{V}$ and $C_{W_h} w_h = \mathcal{W}$,

$$\begin{aligned} \|\mathcal{A}\mathcal{W}\|_p &= \sup_{\mathcal{V} \in \mathbb{R}^N} \frac{(\mathcal{A}\mathcal{W}, \mathcal{V})_N}{\|\mathcal{V}\|_{p'}} \\ &= \sup_{\mathcal{V} \in \mathbb{R}^N} \frac{a_h(w_h, v_h)}{\|w_h\|_{L^p(\Omega)} \|v_h\|_{L^{p'}(\Omega)}} \frac{\|w_h\|_{L^p(\Omega)} \|v_h\|_{L^{p'}(\Omega)}}{\|\mathcal{W}\|_p \|\mathcal{V}\|_{p'}} \|\mathcal{W}\|_p \\ &\leq \omega_{p,h} \frac{\|w_h\|_{L^p(\Omega)}}{\|\mathcal{W}\|_p} \sup_{\mathcal{V} \in \mathbb{R}^N} \frac{\|v_h\|_{L^{p'}(\Omega)}}{\|\mathcal{V}\|_{p'}} \|\mathcal{W}\|_p. \end{aligned}$$

Using inequalities (13)–(14) yields

$$\|\mathcal{A}\mathcal{W}\|_p \leq \omega_{p,h} \mu_{s,\max} \mu_{t,\max} \|\mathcal{W}\|_p.$$

That is to say,

$$\|\mathcal{A}\|_p \leq \omega_{p,h} \mu_{s,\max} \mu_{t,\max}.$$

(2) Upper bound on $\|\mathcal{A}^{-1}\|_p$. Using again (13)–(14) together with definition (17) yields

$$\begin{aligned} \alpha_{p,h} \mu_{s,\min} \|\mathcal{W}\|_p &\leq \alpha_{p,h} \|w_h\|_{L^p(\Omega)} \leq \sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|v_h\|_{L^{p'}(\Omega)}} \\ &= \sup_{\mathcal{V} \in \mathbb{R}^N} \frac{(\mathcal{A}\mathcal{W}, \mathcal{V})_N}{\|v_h\|_{L^{p'}(\Omega)}} \leq \|\mathcal{A}\mathcal{W}\|_p \sup_{\mathcal{V} \in \mathbb{R}^N} \frac{\|\mathcal{V}\|_{p'}}{\|v_h\|_{L^{p'}(\Omega)}} \leq \mu_{t,\min}^{-1} \|\mathcal{A}\mathcal{W}\|_p. \end{aligned}$$

Hence, setting $\mathcal{Z} = \mathcal{A}\mathcal{W}$, we infer

$$\alpha_{p,h} \mu_{s,\min} \|\mathcal{A}^{-1}\mathcal{Z}\|_p \leq \mu_{t,\min}^{-1} \|\mathcal{Z}\|_p.$$

Since \mathcal{Z} is arbitrary, this means

$$\|\mathcal{A}^{-1}\|_p \leq \frac{1}{\alpha_{p,h}} \mu_{s,\min}^{-1} \mu_{t,\min}^{-1}.$$

The upper bound in (19) is a direct consequence of the above estimates.

(3) Lower bound on $\|\mathcal{A}^{-1}\|_p$. Since W_h is finite-dimensional, there is $w_h \neq 0$ in W_h such that

$$\alpha_{p,h} = \sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|w_h\|_{L^p(\Omega)} \|v_h\|_{L^{p'}(\Omega)}}.$$

As a result, setting $\mathcal{W} = C_{W_h} w_h$ and $\mathcal{V} = C_{V_h} v_h$ yields

$$\begin{aligned} \|\mathcal{A}\mathcal{W}\|_p &= \sup_{\mathcal{V} \in \mathbb{R}^N} \frac{(\mathcal{A}\mathcal{W}, \mathcal{V})_N}{\|\mathcal{V}\|_{p'}} \\ &= \sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|w_h\|_{L^p(\Omega)} \|v_h\|_{L^{p'}(\Omega)}} \frac{\|v_h\|_{L^{p'}(\Omega)}}{\|\mathcal{V}\|_{p'}} \frac{\|w_h\|_{L^p(\Omega)}}{\|\mathcal{W}\|_p} \|\mathcal{W}\|_p \\ &\leq \alpha_{p,h} \mu_{t,\max} \mu_{s,\max} \|\mathcal{W}\|_p. \end{aligned}$$

Hence,

$$\frac{1}{\alpha_{p,h}} \mu_{t,\max}^{-1} \mu_{s,\max}^{-1} \leq \|\mathcal{A}^{-1}\|_p.$$

(4) Lower bound on $\|\mathcal{A}\|_p$. Since W_h is finite-dimensional, there is $w_h \neq 0$ in W_h such that

$$\omega_{p,h} = \sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|w_h\|_{L^p(\Omega)} \|v_h\|_{L^{p'}(\Omega)}}.$$

This implies

$$\begin{aligned} \mu_{s,\min} \|\mathcal{W}\|_p &\leq \|w_h\|_{L^p(\Omega)} = \frac{1}{\omega_{p,h}} \sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|v_h\|_{L^{p'}(\Omega)}} \\ &= \frac{1}{\omega_{p,h}} \sup_{v_h \in V_h} \frac{(\mathcal{A}\mathcal{W}, \mathcal{V})_N}{\|\mathcal{V}\|_{p'}} \frac{\|\mathcal{V}\|_{p'}}{\|v_h\|_{L^{p'}(\Omega)}} \\ &\leq \frac{1}{\omega_{p,h}} \mu_{t,\min}^{-1} \|\mathcal{A}\mathcal{W}\|_p \leq \frac{1}{\omega_{p,h}} \mu_{t,\min}^{-1} \|\mathcal{A}\|_p \|\mathcal{W}\|_p. \end{aligned}$$

Since $\mathcal{W} \neq 0$ this yields

$$\omega_{p,h} \mu_{s,\min} \mu_{t,\min} \leq \|\mathcal{A}\|_p.$$

The lower bound in (19) easily follows from the above estimates.

To account for a possible polynomial dependence of $\alpha_{p,h}$ and $\omega_{p,h}$ on h , we make the following additional technical hypotheses:

$$\exists \gamma, \quad \begin{cases} 0 < c_{\inf}^{\alpha} = \liminf_{h \rightarrow 0} \alpha_{p,h} h^{-\gamma} < +\infty, \\ 0 < c_{\sup}^{\alpha} = \limsup_{h \rightarrow 0} \alpha_{p,h} h^{-\gamma} < +\infty, \end{cases} \quad (20)$$

$$\exists \delta, \quad \begin{cases} 0 < c_{\inf}^{\omega} = \liminf_{h \rightarrow 0} \omega_{p,h} h^{\delta} < +\infty, \\ 0 < c_{\sup}^{\omega} = \limsup_{h \rightarrow 0} \omega_{p,h} h^{\delta} < +\infty. \end{cases} \quad (21)$$

As a consequence of Theorem 2, we deduce the following:

Theorem 3 *Under the assumptions (20)–(21), the following holds true: For all $\epsilon \in]0, 1[$, there is h_{ϵ} such that for all $h \leq h_{\epsilon}$,*

$$(1 - \epsilon) \frac{c_{\inf}^{\omega}}{c_{\sup}^{\alpha}} \kappa_{s,p}^{-1} \kappa_{t,p'}^{-1} h^{-\gamma-\delta} \leq \kappa_p(\mathcal{A}) \leq (1 + \epsilon) \frac{c_{\sup}^{\omega}}{c_{\inf}^{\alpha}} \kappa_{s,p} \kappa_{t,p'} h^{-\gamma-\delta}.$$

Proof Let $\epsilon \in]0, 1[$.

(1) There is h_{ϵ} such that for all $h \leq h_{\epsilon}$, $(1 - \frac{\epsilon}{3})c_{\inf}^{\alpha}h^{\gamma} \leq \alpha_{p,h}$ and $\omega_{p,h} \leq (1 + \frac{\epsilon}{3})c_{\sup}^{\omega}h^{-\delta}$. Then, apply Theorem 2 to deduce the upper bound.

(2) Owing to the definition of c_{\sup}^{α} , there is h_{ϵ} such that for all $0 < h \leq h_{\epsilon}$ there is $w_h \in W_h$ satisfying

$$\sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|w_h\|_{L^p(\Omega)} \|v_h\|_{L^{p'}(\Omega)}} \leq (1 + \frac{\epsilon}{2}) c_{\sup}^{\alpha} h^{\gamma}.$$

Then, proceed as in step (3) of the proof of Theorem 2 to derive the lower bound $\|\mathcal{A}^{-1}\|_p \geq (1 + \frac{\epsilon}{2})^{-1} (c_{\sup}^{\alpha})^{-1} \mu_{s,\max}^{-1} \mu_{t,\max}^{-1} h^{-\gamma}$.

(3) The definition of c_{\inf}^{ω} implies the existence of h_{ϵ} such that for all $0 < h \leq h_{\epsilon}$ there is $w_h \in W_h$ satisfying

$$(1 - \frac{\epsilon}{2}) c_{\inf}^{\omega} h^{-\delta} \leq \sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|w_h\|_{L^p(\Omega)} \|v_h\|_{L^{p'}(\Omega)}}.$$

Then, proceed as in step (4) of the proof of Theorem 2 to derive the lower bound $\|\mathcal{A}\|_p \geq (1 - \frac{\epsilon}{2}) c_{\inf}^{\omega} \mu_{s,\min} \mu_{t,\min} h^{-\delta}$. The lower bound on $\kappa_p(\mathcal{A})$ then results from the above estimates.

3.2 Estimates based on natural stability norms

Introduce the quantities

$$\alpha_h = \inf_{w_h \in W_h} \sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|w_h\|_{W_h} \|v_h\|_{V_h}}, \quad (22)$$

$$\omega_h = \sup_{w_h \in W_h} \sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|w_h\|_{W_h} \|v_h\|_{V_h}}. \quad (23)$$

In general, one may expect that the norms of W_h and V_h are selected so that α_h is uniformly bounded from below away from zero and ω_h is uniformly bounded. Hence, bounding $\kappa_p(\mathcal{A})$ in terms of α_h and ω_h may yield valuable information.

To this purpose, we make the following technical assumptions:

$$\exists c_{sP} > 0, \quad \forall w_h \in W_h, \quad c_{sP} \|w_h\|_{L^p(\Omega)} \leq \|w_h\|_{W_h}, \quad (24)$$

$$\exists c_{tP} > 0, \quad \forall v_h \in V_h, \quad c_{tP} \|v_h\|_{L^{p'}(\Omega)} \leq \|v_h\|_{V_h}, \quad (25)$$

$$\exists s > 0, \exists c_{sI}, \quad \forall w_h \in W_h, \quad \|w_h\|_{W_h} \leq c_{sI} h^{-s} \|w_h\|_{L^p(\Omega)}, \quad (26)$$

$$\exists t > 0, \exists c_{tI}, \quad \forall v_h \in V_h, \quad \|v_h\|_{V_h} \leq c_{tI} h^{-t} \|v_h\|_{L^{p'}(\Omega)}. \quad (27)$$

Estimates (24) and (25) are Poincaré-like inequalities expressing the fact that the norms equipping W_h and V_h control the L^p -norm and the $L^{p'}$ -norm, respectively. In other words, the injections $W_h \subset [L^p(\Omega)]^m$ and $V_h \subset [L^{p'}(\Omega)]^m$ are uniformly continuous. Furthermore, (26) and (27) are inverse inequalities. When the mesh family $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform, the constants s and t can be interpreted as the order of the differential operator used to define the norms in W_h and V_h , respectively.

As a consequence of Theorem 2, we deduce the following:

Corollary 1 *Under the assumptions (24)–(27), the following bound holds:*

$$\forall h, \quad \kappa_p(\mathcal{A}) \leq \kappa_{s,p} \kappa_{t,p'} \frac{c_{sI} c_{tI}}{c_{sP} c_{tP}} \frac{\omega_h}{\alpha_h} h^{-s-t}. \quad (28)$$

Proof Let us estimate $\alpha_{p,h}$ and $\omega_{p,h}$.

(1) It is clear that

$$\begin{aligned} \alpha_h &= \inf_{w_h \in W_h} \sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|w_h\|_{W_h} \|v_h\|_{V_h}} \\ &\leq \frac{1}{c_{sP} c_{tP}} \inf_{w_h \in W_h} \sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|w_h\|_{L^p(\Omega)} \|v_h\|_{L^{p'}(\Omega)}}. \end{aligned}$$

Hence $\alpha_{p,h} \geq c_{sP} c_{tP} \alpha_h$.

(2) Moreover,

$$\begin{aligned} \omega_{p,h} &= \sup_{w_h \in W_h} \sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|w_h\|_{W_h} \|v_h\|_{V_h}} \\ &\leq \omega_h \sup_{w_h \in W_h} \frac{\|w_h\|_{W_h}}{\|w_h\|_{L^p(\Omega)}} \sup_{v_h \in V_h} \frac{\|v_h\|_{V_h}}{\|v_h\|_{L^{p'}(\Omega)}} \leq \omega_h c_{sI} c_{tI} h^{-s-t}. \end{aligned}$$

Hence, $\omega_{p,h} \leq c_{sI} c_{tI} \omega_h h^{-s-t}$.

(3) Conclude using Theorem 2.

Remark 1 It may happen that (28) is not sharp; see §4.3 and (61).

In addition to (24)–(27), we assume the following:

$$\exists \mu, \quad \begin{cases} 0 < d_{\inf}^\alpha = \liminf_{h \rightarrow 0} \frac{\alpha_{p,h}}{\alpha_h} h^\mu < +\infty, \\ 0 < d_{\sup}^\alpha = \limsup_{h \rightarrow 0} \frac{\alpha_{p,h}}{\alpha_h} h^\mu < +\infty, \end{cases} \quad (29)$$

$$\exists \nu, \quad \begin{cases} 0 < d_{\inf}^\omega = \liminf_{h \rightarrow 0} \frac{\omega_{p,h}}{\omega_h} h^{s+t-\nu} < +\infty, \\ 0 < d_{\sup}^\omega = \limsup_{h \rightarrow 0} \frac{\omega_{p,h}}{\omega_h} h^{s+t-\nu} < +\infty. \end{cases} \quad (30)$$

The constants μ and ν are meant to measure the possible default to optimality of Corollary 1. Proceeding as in the proof of Theorem 3, it is clear that the following holds true:

Corollary 2 *Under the assumptions (24)–(27) and (29)–(30), the following holds true: For all $\epsilon \in]0, 1[$, there is h_ϵ such that for all $h \leq h_\epsilon$,*

$$\begin{aligned} (1 - \epsilon) \frac{d_{\inf}^\omega}{d_{\sup}^\alpha} \kappa_{s,p}^{-1} \kappa_{t,p'}^{-1} \frac{\omega_h}{\alpha_h} h^{-s-t+\mu+\nu} &\leq \kappa_p(A) \\ &\leq (1 + \epsilon) \frac{d_{\sup}^\omega}{d_{\inf}^\alpha} \kappa_{s,p} \kappa_{t,p'} \frac{\omega_h}{\alpha_h} h^{-s-t+\mu+\nu}. \end{aligned} \quad (31)$$

4 Applications

This section presents various applications of the theoretical results derived in Section 3 to finite element approximations of PDE's posed on a bounded domain Ω in \mathbb{R}^d . For the sake of simplicity, we assume that Ω is a polyhedron. Let $\{\mathcal{T}_h\}_{h>0}$ be a shape-regular family of meshes of Ω .

4.1 Elliptic PDE's in variational form

Consider the Laplacian with homogeneous Dirichlet boundary conditions. Set $W = H_0^1(\Omega)$, $V = H^{-1}(\Omega)$, and $A : W \ni w \mapsto -\Delta w \in V$. Clearly $A : W \rightarrow V$ is an isomorphism. Introduce the bilinear form $a(w_1, w_2) = \int_{\Omega} \nabla w_1 \cdot \nabla w_2$, $\forall (w_1, w_2) \in W \times W$.

Let W_h be a finite-dimensional space based on the mesh \mathcal{T}_h . We assume that $W_h \subset W$, i.e., the approximation is H^1 -conformal. We assume that W_h is such that there is c independent of h such that the following global inverse inequality holds:

$$\forall w_h \in W_h, \quad \|\nabla w_h\|_{L^2(\Omega)} \leq c h^{-1} \|w_h\|_{L^2(\Omega)}. \quad (32)$$

This hypothesis holds whenever W_h is a finite element space constructed using a quasi-uniform mesh family; see, e.g., [2, 3, 6, 7].

Consider the approximate problem:

$$\begin{cases} \text{Seek } u_h \in W_h \text{ such that} \\ a(u_h, v_h) = (f, v_h)_{L^2(\Omega)}, \quad \forall v_h \in W_h, \end{cases} \quad (33)$$

for some data $f \in L^2(\Omega)$. Let \mathcal{A} be the stiffness matrix associated with (33). The main result concerning the Euclidean condition number of \mathcal{A} is the following:

Theorem 4 *If the mesh family $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform, there are $0 < c_1 \leq c_2$ independent of h such that*

$$c_1 h^{-2} \leq \kappa_2(\mathcal{A}) \leq c_2 h^{-2}. \quad (34)$$

Proof (1) For $w_h \in W_h$, define $R(w_h) = \frac{\|\nabla w_h\|_{L^2(\Omega)}^2}{\|w_h\|_{L^2(\Omega)}^2}$. Then

$$\alpha_{2,h} = \inf_{w_h \in W_h} R(w_h), \quad \omega_{2,h} = \sup_{w_h \in W_h} R(w_h). \quad (35)$$

(2) Let \tilde{z}_h be given by Lemma 5 with $Z = H_0^1(\Omega)$, $Z_h = W_h$ equipped with the H^1 -seminorm, and $L = L^2(\Omega)$. Since $R(\tilde{z}_h) \leq c$, we infer $\alpha_{2,h} \leq R(\tilde{z}_h) \leq c$ uniformly in h . Moreover, the Poincaré inequality in $H_0^1(\Omega)$ implies that $\alpha_{2,h}$ is uniformly bounded from below away from zero.

(3) Letting w_h in (35) be one of the global shape functions in W_h , it is clear that $\omega_{2,h} \geq c h^{-2}$. Moreover, owing to the inverse inequality (32), $\omega_{2,h} \leq c' h^{-2}$.

(4) To conclude, use Theorem 2 (or Theorem 3 with $\gamma = 0$ and $\delta = 2$) and observe that owing to the quasi-uniformity of $\{\mathcal{T}_h\}_{h>0}$ and Lemma 4, the constants $\kappa_{s,p}$ and $\kappa_{t,p'}$ are independent of h .

Remark 2 The Euclidean condition number $\kappa_2(\mathcal{A})$ can also be estimated using Corollary 1. One readily verifies that α_h and ω_h are independent of h and that $s = t = 1$. Hence, (28) yields $\kappa_2(\mathcal{A}) \leq ch^{-2}$, i.e., the estimate is sharp. One also verifies that $\mu = \nu = 0$ in (29)–(30), confirming the optimality of Corollary 1.

Remark 3 Estimate (34) extends to more general second-order elliptic operators, e.g., advection–diffusion–reaction equations.

4.2 Elliptic PDE's in mixed form

In this section we investigate a non-standard Galerkin technique to approximate the Laplacian in mixed form.

Let $H(\operatorname{div}; \Omega) = \{v \in [L^2(\Omega)]^d; \nabla \cdot v \in L^2(\Omega)\}$, $W = H(\operatorname{div}; \Omega) \times H_0^1(\Omega)$, and $V = [L^2(\Omega)]^d \times L^2(\Omega)$. Introduce the operator

$$A : W \ni (u, p) \longmapsto (u + \nabla p, \nabla \cdot u) \in V. \quad (36)$$

One readily verifies that $A : W \rightarrow V$ is an isomorphism. For $(w, v) \in W \times V$, define the bilinear form

$$a((u, p), (v, q)) = (u, v)_{L^2(\Omega)} + (\nabla p, v)_{L^2(\Omega)} + (\nabla \cdot u, q)_{L^2(\Omega)}. \quad (37)$$

By analogy with Darcy's equations, u is termed the velocity and p the pressure.

The non-standard Galerkin approximation introduced in [4] consists of seeking the discrete velocity in the Raviart–Thomas finite element space of lowest order and the discrete pressure in the Crouzeix–Raviart finite element space. Denote by \mathcal{F}_h , \mathcal{F}_h^∂ , and \mathcal{F}_h^i the set of faces, boundary faces, and interior faces of the mesh, respectively. Define

$$X_h = \{u_h; \forall K \in \mathcal{T}_h, u_h|_K \in \mathbb{RT}_0; \forall F \in \mathcal{F}_h^i, \int_F \llbracket u_h \cdot n \rrbracket = 0\}, \quad (38)$$

$$Y_h = \{p_h; \forall K \in \mathcal{T}_h, p_h|_K \in \mathbb{P}_1; \forall F \in \mathcal{F}_h, \int_F \llbracket p_h \rrbracket = 0\}, \quad (39)$$

where $\mathbb{RT}_0 = [\mathbb{P}_0]^d \oplus x\mathbb{P}_0$, $\llbracket u_h \cdot n \rrbracket$ denotes the jump of the normal component of u_h across interfaces, and $\llbracket p_h \rrbracket$ the jump of p_h across interfaces (with the convention that a zero outer value is taken whenever $F \in \mathcal{F}_h^\partial$). Test functions for both the velocity and the pressure are taken to be piecewise constants. Introducing the spaces $W_h = X_h \times Y_h$ and

$$V_h = \{(v_h, q_h); \forall K \in \mathcal{T}_h, v_h|_K \in [\mathbb{P}_0]^d \text{ and } q_h|_K \in \mathbb{P}_0\}, \quad (40)$$

and defining the bilinear form $a_h \in \mathcal{L}(W_h \times V_h; \mathbb{R})$ such that

$$\begin{aligned} a_h((u_h, p_h), (v_h, q_h)) &= (u_h, v_h)_{L^2(\Omega)} \\ &+ (\nabla \cdot u_h, q_h)_{L^2(\Omega)} + \sum_{K \in \mathcal{T}_h} (\nabla p_h, v_h)_{L^2(K)}, \end{aligned} \quad (41)$$

the discrete problem is formulated as follows:

$$\begin{cases} \text{Seek } (u_h, p_h) \in W_h \text{ such that} \\ a_h((u_h, p_h), (v_h, q_h)) = (f, q_h)_{L^2(\Omega)}, \quad \forall (v_h, q_h) \in V_h, \end{cases} \quad (42)$$

for some data $f \in L^2(\Omega)$. Note that the approximation setting is conformal on the velocity and non-conformal on the pressure. Moreover, it is readily checked that the total number of unknowns in (42) equals the total number of equations. Indeed, the former is the number of faces plus the number of interior faces, the latter is equal to $(d+1)$ times the number of elements, and these two quantities are equal owing to the Euler relations.

Equip W_h with the norm

$$\|(u_h, p_h)\|_{W_h}^2 = \|u_h\|_{L^2(\Omega)}^2 + \|\nabla \cdot u_h\|_{L^2(\Omega)}^2 + \|p_h\|_{L^2(\Omega)}^2 + \sum_{K \in \mathcal{T}_h} \|\nabla p_h\|_{L^2(K)}^2, \quad (43)$$

and equip V_h with the norm $\|(v_h, q_h)\|_{V_h}^2 = \|v_h\|_{L^2(\Omega)}^2 + \|q_h\|_{L^2(\Omega)}^2$. In the framework of the BNB Theorem, the main stability result for (42) is the following:

Lemma 2 *There are $c > 0$ and h_0 such that for all $h \leq h_0$,*

$$\inf_{(u_h, p_h) \in W_h} \sup_{(v_h, q_h) \in V_h} \frac{a_h((u_h, p_h), (v_h, q_h))}{\|(u_h, p_h)\|_{W_h} \|(v_h, q_h)\|_{V_h}} \geq c. \quad (44)$$

Proof Since this is a non-classical result, the proof is briefly sketched; see [4] and [6] for further details.

(1) Let $(u_h, p_h) \in W_h$. Denote by \bar{u}_h the function whose restriction to each element $K \in \mathcal{T}_h$ is the mean value of u_h . Denote by $\nabla_h p_h$ the function whose restriction to each element $K \in \mathcal{T}_h$ is $\nabla p_h|_K$. Set $v_h = \bar{u}_h + \nabla_h p_h$ and $q_h = 2p_h + \nabla \cdot u_h$. Note that (v_h, q_h) is in V_h . Hence,

$$\begin{aligned} a_h((u_h, p_h), (v_h, q_h)) &= (u_h, \bar{u}_h)_{L^2(\Omega)} + \|\nabla \cdot u_h\|_{L^2(\Omega)}^2 + \sum_{K \in \mathcal{T}_h} \|\nabla p_h\|_{L^2(K)}^2 \\ &+ 2(\nabla \cdot u_h, p_h)_{L^2(\Omega)} + \sum_{K \in \mathcal{T}_h} (u_h, \nabla p_h)_{L^2(K)} + (\nabla p_h, \bar{u}_h)_{L^2(K)} \\ &= \|\bar{u}_h\|_{L^2(\Omega)}^2 + \|\nabla \cdot u_h\|_{L^2(\Omega)}^2 + \sum_{K \in \mathcal{T}_h} \|\nabla p_h\|_{L^2(K)}^2, \end{aligned}$$

since $(\nabla \cdot u_h, p_h)_{L^2(\Omega)} + \sum_{K \in \mathcal{T}_h} (u_h, \nabla p_h)_{L^2(K)} = 0$.

(2) For $u_h \in X_h$, one readily verifies that $\forall K \in \mathcal{T}_h, \forall x \in K, u_h(x) = \bar{u}_h + \frac{1}{d}(x - g_K) \nabla \cdot u_h$ where g_K is the barycenter of K . This implies that there is c independent of h such that

$$\forall u_h \in X_h, \quad \|u_h\|_{L^2(K)} \leq \|\bar{u}_h\|_{L^2(K)} + c h_K \|\nabla \cdot u_h\|_{L^2(K)}.$$

Hence,

$$\begin{aligned} a_h((u_h, p_h), (v_h, q_h)) &\geq c \|u_h\|_{L^2(\Omega)}^2 + (1 - c'h^2) \|\nabla \cdot u_h\|_{L^2(\Omega)}^2 \\ &\quad + \sum_{K \in \mathcal{T}_h} \|\nabla p_h\|_{L^2(K)}^2. \end{aligned}$$

If h is small enough, $(1 - c'h^2)$ is bounded from below by $\frac{1}{2}$.

(3) Use the extended Poincaré inequality (see, e.g., [5, 6] for a proof)

$$\forall p_h \in Y_h, \quad \sum_{K \in \mathcal{T}_h} \|\nabla p_h\|_{L^2(K)}^2 \geq c \|p_h\|_{L^2(\Omega)}^2,$$

and the above estimates to conclude that $a_h((u_h, p_h), (v_h, q_h)) \geq c \|(u_h, p_h)\|_{W_h}^2 \geq c' \|(u_h, p_h)\|_{W_h} \|(v_h, q_h)\|_{V_h}$.

We now estimate the Euclidean condition number of the stiffness matrix \mathcal{A} resulting from (42). Our main result is the following:

Theorem 5 *If the mesh family $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform, there are $0 < c_1 \leq c_2$ independent of h such that*

$$c_1 h^{-1} \leq \kappa_2(\mathcal{A}) \leq c_2 h^{-1}. \quad (45)$$

Proof (1) Owing to (44),

$$\sup_{(v_h, q_h) \in V_h} \frac{a_h((u_h, p_h), (v_h, q_h))}{\|(v_h, q_h)\|_{V_h}} \geq c \|(u_h, p_h)\|_{W_h} \geq c \|(u_h, p_h)\|_{L^2(\Omega)}.$$

Hence, $\alpha_{2,h} \geq c$.

(2) Take $u_h = 0$ and $p_h = \tilde{z}_h$ given by Lemma 5 with $Z = H_0^1(\Omega)$, $Z_h = Y_h \cap H_0^1(\Omega)$ equipped with the H^1 -seminorm, and $L = L^2(\Omega)$. Then,

$$\alpha_{2,h} \leq \sup_{(v_h, q_h) \in V_h} \frac{a_h((0, \tilde{z}_h), (v_h, q_h))}{\|(0, \tilde{z}_h)\|_{L^2(\Omega)} \|(v_h, q_h)\|_{L^2(\Omega)}} \leq c'.$$

(3) Since the mesh family $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform, it is clear that an inverse inequality of the form (32) holds in W_h . This implies that

$\omega_{2,h} \leq ch^{-1}$. Moreover, setting $u_h = 0$ and letting p_h be one the global shape functions in Y_h , say ψ_i , yields

$$\omega_{2,h} \geq \sup_{(v_h, q_h) \in V_h} \frac{a_h((0, \psi_i), (v_h, q_h))}{\|(0, \psi_i)\|_{L^2(\Omega)} \|(v_h, q_h)\|_{L^2(\Omega)}} \geq c'h^{-1}.$$

(4) To conclude, use Theorem 2 (or Theorem 3 with $\gamma = 0$ and $\delta = 1$) and Lemma 4.

Remark 4 As for elliptic PDE's in variational form, $\kappa_2(\mathcal{A})$ can also be estimated using Corollary 1. One readily verifies that α_h and ω_h can be uniformly bounded from below and above, and that $s = 1$ and $t = 0$. Hence, (28) yields $\kappa_2(\mathcal{A}) \leq ch^{-1}$, i.e., the estimate is sharp. One also verifies that $\mu = \nu = 0$ in (29)–(30), confirming the optimality of Corollary 1.

Remark 5 It is also possible to consider a standard Galerkin approximation to the Laplacian in mixed form. In this case, the trial space and the test space are identical and given by $W_h = V_h = X_h \times Z_h$ where X_h is defined by (38) and Z_h denotes the space of piecewise constant functions. The discrete problem is (42) with the bilinear form

$$a_h((u_h, p_h), (v_h, q_h)) = (u_h, v_h)_{L^2(\Omega)} - (\nabla \cdot v_h, p_h)_{L^2(\Omega)} + (\nabla \cdot u_h, q_h)_{L^2(\Omega)}. \quad (46)$$

One readily verifies that the Euclidean condition number of the resulting stiffness matrix scales as h^{-1} , i.e., the same asymptotic behavior as for the non-standard Galerkin approximation is obtained. This result is essentially due to the fact that the mixed form only involves first-order PDE's.

Remark 6 Although the Euclidean condition number of the stiffness matrix associated with the mixed form is one order smaller in h than that associated with the variational form, the matrix in the first case is larger than that in the second case so that it is not a priori clear to decide which linear system is the easiest to solve by an iterative method.

4.3 First-order PDE's and GaLS

Let β be a vector field in \mathbb{R}^d , assume $\beta \in [L^\infty(\Omega)]^d$, $\nabla \cdot \beta \in L^\infty(\Omega)$, and define the inflow and outflow boundaries

$$\partial\Omega^- = \{x \in \partial\Omega; \beta(x) \cdot n(x) < 0\}, \quad \partial\Omega^+ = \{x \in \partial\Omega; \beta(x) \cdot n(x) > 0\},$$

where n is the outward normal to Ω . Let μ be a function in $L^\infty(\Omega)$ and consider the advection–reaction equation

$$\begin{cases} \mu u + \beta \cdot \nabla u = f, \\ u|_{\partial\Omega^-} = 0. \end{cases} \quad (47)$$

To give a mathematical meaning to (47), consider the spaces

$$W = \{w \in L^2(\Omega); \beta \cdot \nabla w \in L^2(\Omega); w|_{\partial\Omega^-} = 0\}, \quad (48)$$

$$V = L^2(\Omega). \quad (49)$$

Equipped with the norm $\|w\|_W = \|w\|_{L^2(\Omega)} + \|\beta \cdot \nabla w\|_{L^2(\Omega)}$, W is a Hilbert space. Now, define the differential operator

$$A : W \ni w \longmapsto \mu w + \beta \cdot \nabla w \in V.$$

It is clear that A is continuous. Moreover, assuming that there is $\mu_0 > 0$ such that

$$\mu(x) - \frac{1}{2} \nabla \cdot \beta(x) \geq \mu_0 > 0 \quad \text{a.e. in } \Omega, \quad (50)$$

$A : W \rightarrow V$ is an isomorphism.

We want to illustrate Theorem 2 by analyzing the Euclidean condition number of the stiffness matrix associated with the GaLS approximation of (47). To this purpose introduce a finite-dimensional approximation space W_h based on the mesh \mathcal{T}_h . Assume that $W_h \subset H^1(\Omega) \cap W$, i.e., the approximation is H^1 -conformal. Introduce the bilinear form $a \in \mathcal{L}(W \times V; \mathbb{R})$ such that $a(w, v) = (Aw, v)_{L^2(\Omega)}$ and set

$$a_h(w, v) = a(w, v) + \sum_{K \in \mathcal{T}_h} \delta(h_K) (Aw, Av)_{L^2(K)}, \quad (51)$$

where $\delta(h_K) = c_{\text{GaLS}} h_K$ and c_{GaLS} is a (user-defined) mesh-independent constant. Assume $f \in L^2(\Omega)$. The GaLS approximate problem consists of the following [9]:

$$\begin{cases} \text{Seek } u_h \in W_h \text{ such that} \\ a_h(u_h, v_h) = (f, v_h)_{L^2(\Omega)} + \sum_{K \in \mathcal{T}_h} \delta(h_K) (f, Av_h)_{L^2(K)}, \quad \forall v_h \in W_h. \end{cases} \quad (52)$$

Note that the solution space and the test space are identical here, i.e., $V_h = W_h$. Define the symmetric bilinear form $a_s \in \mathcal{L}(W \times W; \mathbb{R})$ such that

$$\forall (w_1, w_2) \in W \times W, \quad a_s(w_1, w_2) = \frac{1}{2} ((Aw_1, w_2)_{L^2(\Omega)} + (w_1, Aw_2)_{L^2(\Omega)}).$$

It is clear that a_s is positive definite since

$$\forall w \in W, \quad a_s(w, w) = a(w, w) \geq \mu_0 \|w\|_{L^2(\Omega)}^2. \quad (53)$$

The main result of this section is the following:

Theorem 6 *Assume that there is a nonempty open subset of $\overline{\Omega}$, say Ω_0 , in which $\inf_{\Omega_0} \|\beta\| > 0$ and β is in $C^{0,1}(\overline{\Omega}_0)$. Assume that the mesh family $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform. Then, there are $0 < c_1 \leq c_2$ and h_0 such that for all $h \leq h_0$,*

$$c_1 h^{-1} \leq \kappa_2(\mathcal{A}) \leq c_2 h^{-1}. \quad (54)$$

Proof (1) Owing to (53),

$$\mu_0 \leq \inf_{w_h \in W_h} \frac{a_h(w_h, w_h)}{\|w_h\|_{L^2(\Omega)}^2} \leq \inf_{w_h \in W_h} \sup_{v_h \in W_h} \frac{a_h(w_h, v_h)}{\|w_h\|_{L^2(\Omega)} \|v_h\|_{L^2(\Omega)}} = \alpha_{2,h},$$

i.e., $\mu_0 \leq \alpha_{2,h}$.

(2) To derive a bound on $\alpha_{2,h}$, use Lemma 5. Set $Z = W$, $Z_h = W_h$, $L = L^2(\Omega)$, and equip Z_h with the norm $\|z_h\|_{Z_h} = \|Az_h\|_{L^2(\Omega)}$.

Lemma 5 implies that there exists $\tilde{c} > 0$ and \tilde{h} such that for all $h \leq \tilde{h}$, there is $\tilde{z}_h \in W_h \setminus \{0\}$ satisfying $\|A\tilde{z}_h\|_{L^2(\Omega)} \leq \tilde{c} \|\tilde{z}_h\|_{L^2(\Omega)}$. Then

$$\alpha_{2,h} \leq \sup_{v_h \in W_h} \frac{a_h(\tilde{z}_h, v_h)}{\|\tilde{z}_h\|_{L^2(\Omega)} \|v_h\|_{L^2(\Omega)}}.$$

A direct computation using (32) shows that $\alpha_{2,h}$ is bounded uniformly with respect to h .

(3) Using again (32) it is clear that there is c independent of h such that $\omega_{2,h} \leq ch^{-1}$.

(4) A simple computation yields

$$a_h(w_h, w_h) \geq \int_{\Omega} (\mu_0 - h \|\mu\|_{L^\infty(\Omega)}^2) w_h^2 + \sum_{K \in \mathcal{T}_h} \frac{h_K}{2} \int_K |\beta \cdot \nabla w_h|^2.$$

Assume that h is small enough so that $\mu_0 - h \|\mu\|_{L^\infty(\Omega)}^2 \geq 0$ and there is a mesh cell $K_0 \subset \Omega_0$. Then for all $w_h \in W_h$,

$$a_h(w_h, w_h) \geq \frac{h_{K_0}}{2} \int_{K_0} \frac{1}{2} |\bar{\beta} \cdot \nabla w_h|^2 - |(\beta - \bar{\beta}) \cdot \nabla w_h|^2,$$

where $\bar{\beta}$ is the value of β at the barycenter of K_0 . Then it is always possible to find a global shape function φ_i that is nonzero on K_0 and such that

$$\|\bar{\beta} \cdot \nabla \varphi_i\|_{L^2(K_0)} \geq ch^{-1} \|\varphi_i\|_{L^2(K_0)} \geq c'h^{-1} \|\varphi_i\|_{L^2(\Omega)},$$

where c' is positive and independent of h . Hence, if h is small enough

$$\omega_{2,h} \geq \frac{a_h(\varphi_i, \varphi_i)}{\|\varphi_i\|_{L^2(\Omega)}^2} \geq c'h^{-1}.$$

(5) To conclude, use Theorem 2 (or Theorem 3 with $\gamma = 0$ and $\delta = 1$) and Lemma 4.

We now estimate the Euclidean condition number $\kappa_2(\mathcal{A})$ using the natural stability norms. For the GaLS technique, such norms are

$$\forall w \in W, \quad \begin{cases} \|w\|_{h,A}^2 = a_s(w, w) + \sum_{K \in \mathcal{T}_h} \delta(h_K) \|Aw\|_{L^2(K)}^2, \\ \|w\|_{h, \frac{1}{2}}^2 = \|w\|_{h,A}^2 + \sum_{K \in \mathcal{T}_h} h_K^{-1} \|w\|_{L^2(K)}^2. \end{cases} \quad (55)$$

The introduction of these norms is motivated by the following stability and boundedness results:

$$\forall w \in W, \quad a_h(w, w) \geq \|w\|_{h,A}^2, \quad (56)$$

$$\forall w \in W, \quad \forall w_h \in W_h, \quad a_h(w, w_h) \leq c \|w\|_{h, \frac{1}{2}} \|w_h\|_{h,A}. \quad (57)$$

from which the convergence analysis of the GaLS approximation directly follows; see [6] for more details.

Proposition 1 *Equip W_h and V_h with the norm $\|\cdot\|_{h,A}$ to define α_h and ω_h in (22)–(23). Then if the mesh family $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform, there is c independent of h such that*

$$\alpha_h \geq 1, \quad (58)$$

$$\omega_h \leq c h^{-\frac{1}{2}}, \quad (59)$$

$$s = t = \frac{1}{2}. \quad (60)$$

Proof (1) Estimate (58) is a direct consequence of (56).

(2) Owing to the quasi-uniformity hypothesis and (53),

$$\begin{aligned} \|w\|_{h, \frac{1}{2}}^2 &= \|w\|_{h,A}^2 + ch^{-1} \|w\|_{L^2(\Omega)}^2 \\ &\leq \|w\|_{h,A}^2 + \frac{c}{\mu_0} h^{-1} a_s(w, w) \leq \left(1 + \frac{c}{\mu_0} h^{-1}\right) \|w\|_{h,A}^2. \end{aligned}$$

The bound (59) follows readily from (57).

(3) Statement (60) is an easy consequence of (32).

Remark 7 If we apply Corollary 1, we obtain

$$\kappa_2(\mathcal{A}) \leq ch^{-\frac{3}{2}}. \quad (61)$$

This result shows that Corollary 1 may not be optimal; in fact, one readily verifies that $\mu = 0$ and $\nu = \frac{1}{2}$ in Corollary 2.

4.4 First-order PDE's in L^1

Let $\Omega =]0, 1[$, $f \in L^1(\Omega)$, and consider the following problem:

$$\begin{cases} \mu u + u_x = f, \\ u(0) = 0, \end{cases} \quad (62)$$

where μ is a nonnegative constant. This problem has a unique solution in the framework

$$W = \{w \in W^{1,1}(\Omega); w(0) = 0\}, \quad (63)$$

$$V = L^1(\Omega). \quad (64)$$

Define the operator

$$A : W \ni w \mapsto \mu w + w_x \in V. \quad (65)$$

$A \in \mathcal{L}(W; V)$ is an isomorphism, implying that

$$\exists \alpha > 0, \forall w \in W, \quad \|Aw\|_{L^1(\Omega)} \geq \alpha \|w\|_{W^{1,1}(\Omega)}. \quad (66)$$

Define the finite element spaces

$$W_h = \{w_h \in \mathcal{C}^0(\overline{\Omega}); \forall K \in \mathcal{T}_h, w_h|_K \in \mathbb{P}_1; w_h(0) = 0\}, \quad (67)$$

$$V_h = \{v_h \in L^1(\Omega); \forall K \in \mathcal{T}_h, v_h|_K \in \mathbb{P}_0\}. \quad (68)$$

The discrete solution space W_h consists of continuous piecewise affine functions while the discrete test space V_h consists of piecewise constant functions. Introduce the bilinear form

$$\forall (w, v) \in W \times V', \quad a(w, v) = \int_0^1 (\mu w + w_x)v.$$

Clearly $a \in \mathcal{L}(W \times V'; \mathbb{R})$ where $V' = L^\infty(\Omega)$. The discrete problem is the following:

$$\begin{cases} \text{Seek } u_h \in W_h \text{ such that} \\ a(u_h, v_h) = \int_0^1 f v_h, \quad \forall v_h \in V_h. \end{cases} \quad (69)$$

Obviously W_h and V_h have the same dimension, say N , the number of mesh cells. In the framework of the BNB Theorem, the main stability result for (69) is the following:

Lemma 3 *There is $\gamma > 0$ and h_0 such that for all $h \leq h_0$,*

$$\inf_{w_h \in W_h} \sup_{v_h \in V_h} \frac{a(w_h, v_h)}{\|w_h\|_{W^{1,1}(\Omega)} \|v_h\|_{L^\infty(\Omega)}} \geq \gamma. \quad (70)$$

Proof Let $w_h \in W_h \setminus \{0\}$. Denote by sg the sign function, i.e., $\text{sg}(x) = \frac{x}{|x|}$ if x is not zero and $\text{sg}(0) = 0$. For $w_h \in W_h$, let $\bar{w}_h \in V_h$ be the function such that the restriction of \bar{w}_h to a mesh cell K is the mean value of w_h over this mesh cell. Set $z_h = \text{sg}(\mu\bar{w}_h + w_{h,x})$. Clearly $\mu\bar{w}_h + w_{h,x} \neq 0$, otherwise w_h would be zero; hence, $\|z_h\|_{L^\infty(\Omega)} = 1$. Observing that $z_h \in V_h$, we infer

$$\begin{aligned} \sup_{v_h \in V_h} \frac{a(w_h, v_h)}{\|v_h\|_{L^\infty(\Omega)}} &\geq \frac{a(w_h, z_h)}{\|z_h\|_{L^\infty(\Omega)}} = \sum_{K \in \mathcal{T}_h} \mu z_h \int_K w_h + \int_0^1 w_{h,x} z_h \\ &= \sum_{K \in \mathcal{T}_h} \mu z_h \int_K \bar{w}_h + \int_0^1 w_{h,x} z_h = \int_0^1 (\mu\bar{w}_h + w_{h,x}) z_h \\ &= \|\mu\bar{w}_h + w_{h,x}\|_{L^1} \geq \|\mu w_h + w_{h,x}\|_{L^1} - \|\mu(w_h - \bar{w}_h)\|_{L^1} \\ &\geq \alpha \|w_h\|_{W^{1,1}(\Omega)} - ch \|w_h\|_{W^{1,1}(\Omega)}. \end{aligned}$$

The conclusion follows readily.

Let $\{\psi_1, \dots, \psi_N\}$ be the standard \mathbb{P}_1 shape functions of W_h . Let $\{\varphi_1, \dots, \varphi_N\}$ be the standard \mathbb{P}_0 shape functions of V_h , i.e., the characteristic functions of mesh cells. Let \mathcal{A} be the stiffness matrix with entries $(a(\psi_j, \varphi_i))_{1 \leq i, j \leq N}$. The main result of this section is the following:

Theorem 7 *If the mesh family $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform, there are $0 < c_1 \leq c_2$ and h_0 such that for all $h \leq h_0$,*

$$c_1 h^{-1} \leq \kappa_1(\mathcal{A}) \leq c_2 h^{-1}. \quad (71)$$

Proof (1) From Lemma 3, it is clear that $\alpha_{1,h} \geq \gamma$.

(2) To derive a bound on $\alpha_{1,h}$, we use Lemma 5. Set $Z = W$, $Z_h = W_h$, $L = L^1(\Omega)$, and equip Z_h with the norm $\|\cdot\|_{W^{1,1}(\Omega)}$. Lemma 5 implies that there exists $\tilde{c} > 0$ and \tilde{h} such that for all $h \leq \tilde{h}$, there is $\tilde{z}_h \in Z_h \setminus \{0\}$ satisfying $\|\tilde{z}_h\|_{W^{1,1}(\Omega)} \leq \tilde{c} \|\tilde{z}_h\|_{L^1(\Omega)}$. Since

$$\alpha_{1,h} \leq \sup_{v_h \in V_h} \frac{a(\tilde{z}_h, v_h)}{\|\tilde{z}_h\|_{L^1(\Omega)} \|v_h\|_{L^1(\Omega)}},$$

one readily infers that $\alpha_{1,h}$ is bounded uniformly with respect to h .

(3) Using standard inverse inequalities yields $\omega_{1,h} \leq ch^{-1}$.

(4) Let ψ_i be a shape function in W_h and set $v_h = \text{sg}(\psi_{i,x})$. Then, $v_h \in V_h$, $\|v_h\|_{L^\infty(\Omega)} = 1$ and

$$\begin{aligned} a(\psi_i, v_h) &\geq -\mu \|\psi_i\|_{L^1(\Omega)} + \|\psi_{i,x}\|_{L^1(\Omega)} \geq -\mu \|\psi_i\|_{L^1(\Omega)} + \frac{\epsilon}{h} \|\psi_i\|_{L^1(\Omega)} \\ &\geq \left(\frac{\epsilon}{h} - \mu\right) \|\psi_i\|_{L^1(\Omega)} \|v_h\|_{L^\infty(\Omega)}. \end{aligned}$$

LapMix		GaLS		NGL1	
h^{-1}	$\kappa_2(\mathcal{A})$	h^{-1}	$\kappa_2(\mathcal{A})$	h^{-1}	$\kappa_1(\mathcal{A})$
4	8.5	10	68.5	4	50.7
8	16.7	20	139.0	8	101.2
16	32.9	40	282.8	16	202.3
32	64.9	50	345.2	32	404.6
64	129.0			64	809.1

Table 1. Condition number of the stiffness matrix as a function of meshsize for the three test cases.

This implies $\omega_{1h} \geq ch^{-1}$.

(5) Apply Theorem 2 to conclude.

Remark 8 The above result can be easily adapted to the situation where μ is a nonconstant function in $L^\infty(\Omega)$.

5 Numerical illustrations

The purpose of this section is to numerically illustrate the theoretical results derived in the previous sections. Consider the following test cases:

- Case 1 (LapMix): the Laplacian in mixed form is approximated by the non-standard Galerkin technique described in Section 4.2; the domain is $\Omega =]0, 1[$ and a family of uniform meshes with stepsize $h = 2^{-i}$, $i \in \{2, \dots, 6\}$, is employed.
- Case 2 (GaLS): the first-order PDE (47) posed in the unit square of \mathbb{R}^2 with $\mu = 1$ and $\beta = (1, 0)^T$ is approximated by the GaLS technique with parameter c_{GaLS} set to 1; the meshes are quasi-Delaunay triangulations constructed using a frontal method by imposing a uniform mesh of stepsize $h = 0.1, 0.05, 0.025$, and 0.02 on the boundary of Ω .
- Case 3 (NGL1): the first-order PDE (62) with $\mu = 1$ is approximated by the non-standard Galerkin technique based on the L^1 -setting described in Section 4.4; a family of uniform meshes with stepsize $h = 2^{-i}$, $i \in \{2, \dots, 6\}$, is employed.

Results are collected in Table 1. In the three cases we observe that the numerical predictions match the theoretical results.

A Technical results

A.1 Estimates of $\kappa_{s,p}$ and $\kappa_{t,p'}$

Let $\{\mathcal{T}_h\}_{h>0}$ be a shape-regular family of meshes of Ω . Recall that the shape-regular mesh family $\{\mathcal{T}_h\}_{h>0}$ is said to be quasi-uniform if there is c independent of $h = \max_{K \in \mathcal{T}_h} (h_K)$ such that $h \leq c \min_{K \in \mathcal{T}_h} (h_K)$.

Let $\{\widehat{K}, \widehat{P}, \widehat{\Sigma}\}$ be the reference finite element on which W_h is constructed. For each cell K , denote by $T_K : \widehat{K} \rightarrow K$ the transformation that maps the reference cell \widehat{K} to K . For the sake of simplicity, assume that T_K is affine, i.e., Ω is a polyhedron. Moreover, assume the following:

$$W_h \subset \{w_h \in [L^1(\Omega)]^m; \forall K \in \mathcal{T}_h, (w_h \circ T_K^{-1})|_K \in \widehat{P}\}. \quad (72)$$

See [2,3,6,7] for more details on the construction of finite element spaces.

Lemma 4 *If $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform, there exist $0 < c_1 \leq c_2$ such that*

$$\forall h, \forall w_h \in W_h, \quad c_1 h^{\frac{d}{p}} \|C_{W_h} w_h\|_p \leq \|w_h\|_{L^p(\Omega)} \leq c_2 h^{\frac{d}{p}} \|C_{W_h} w_h\|_p. \quad (73)$$

As a result,

$$\forall h, \quad \frac{c_1}{c_2} \leq \kappa_{s,p} \leq \frac{c_2}{c_1}. \quad (74)$$

Proof Assume $1 \leq p < +\infty$. The case $p = +\infty$ can be treated similarly.

(1) Let $\{\widehat{\theta}_1, \dots, \widehat{\theta}_{n_{\text{sh}}}\}$ be the local shape functions for the reference finite element. Denote by $\mathcal{S}^{n_{\text{sh}}}$ the unit sphere in $\mathbb{R}^{n_{\text{sh}}}$ for the $\|\cdot\|_p$ -norm and define the operator

$$\psi : \mathcal{S}^{n_{\text{sh}}} \ni \eta \longmapsto \left\| \sum_{k=1}^{n_{\text{sh}}} \eta_k \widehat{\theta}_k \right\|_{L^p(\widehat{K})} \in \mathbb{R}.$$

The operator ψ is clearly continuous. Moreover, since $\mathcal{S}^{n_{\text{sh}}}$ is compact, ψ reaches its minimum and its maximum, say \widehat{c}_1 and \widehat{c}_2 , respectively. Assume that $\widehat{c}_1 = 0$. Then, there exists $\eta \in \mathcal{S}^{n_{\text{sh}}}$ such that $\psi(\eta) = 0$, yielding $\sum_{k=1}^{n_{\text{sh}}} \eta_k \widehat{\theta}_k = 0$. Since $\{\widehat{\theta}_1, \dots, \widehat{\theta}_{n_{\text{sh}}}\}$ is a basis, this implies $\eta_1 = \dots = \eta_{n_{\text{sh}}} = 0$, contradicting the fact that $\eta \in \mathcal{S}^{n_{\text{sh}}}$. Therefore, $\widehat{c}_1 > 0$. Consider now $\widehat{\mathcal{U}} \in \mathbb{R}^{n_{\text{sh}}}$ with $\widehat{\mathcal{U}} \neq 0$. Let $\widehat{u} = \sum_{i=1}^{n_{\text{sh}}} \widehat{\mathcal{U}}_i \widehat{\theta}_i$ and $\eta_i(\widehat{u}) = \widehat{\mathcal{U}}_i / \|\widehat{\mathcal{U}}\|_p$ for $1 \leq i \leq n_{\text{sh}}$. Clearly, $\eta(\widehat{u}) = (\eta_i(\widehat{u}))_{1 \leq i \leq n_{\text{sh}}}$ is in $\mathcal{S}^{n_{\text{sh}}}$. Since $\psi(\eta(\widehat{u})) = \|\widehat{u}\|_{L^p(\widehat{K})} / \|\widehat{\mathcal{U}}\|_p$, the following inequalities hold:

$$\forall \widehat{\mathcal{U}} \in \mathbb{R}^{n_{\text{sh}}}, \quad \widehat{c}_1 \|\widehat{\mathcal{U}}\|_p \leq \|\widehat{u}\|_{L^p(\widehat{K})} \leq \widehat{c}_2 \|\widehat{\mathcal{U}}\|_p. \quad (75)$$

(2) Consider now an arbitrary element K in the mesh. Denote by $T_K : \widehat{K} \rightarrow K$ the corresponding transformation and by $\{\theta_1, \dots, \theta_{n_{\text{sh}}}\}$ the local shape functions. For $\mathcal{U} \in \mathbb{R}^{n_{\text{sh}}}$, set $u = \sum_{i=1}^{n_{\text{sh}}} \mathcal{U}_i \theta_i$ and $\widehat{u} = u \circ T_K$. Observing that $\widehat{\mathcal{U}} = \mathcal{U}$ and changing variables in the integral in (75) yields

$$\forall \mathcal{U} \in \mathbb{R}^{n_{\text{sh}}}, \quad \left(\frac{\text{meas}(K)}{\text{meas}(\widehat{K})} \right)^{\frac{1}{p}} \widehat{c}_1 \|\mathcal{U}\|_p \leq \|u\|_{L^p(K)} \leq \left(\frac{\text{meas}(K)}{\text{meas}(\widehat{K})} \right)^{\frac{1}{p}} \widehat{c}_2 \|\mathcal{U}\|_p.$$

Clearly, $\frac{\text{meas}(K)}{\text{meas}(\widehat{K})} \leq ch_K^d \leq ch^d$. Furthermore, the quasi-uniformity of the mesh family implies $c'h^d \leq \frac{\text{meas}(K)}{\text{meas}(\widehat{K})}$. As a result, there are $0 < c_1 \leq c_2$ such that

$$\forall h, \forall K \in \mathcal{T}_h, \forall \mathcal{U} \in \mathbb{R}^{n_{\text{sh}}}, \quad c_1 h^{\frac{d}{p}} \|\mathcal{U}\|_p \leq \|u\|_{L^p(K)} \leq c_2 h^{\frac{d}{p}} \|\mathcal{U}\|_p.$$

(3) Let $w_h \in W_h$ and set $\mathcal{W} = C_{W_h} w_h$, i.e., $w_h = \sum_{i=1}^N \mathcal{W}_i \psi_i$. Step 2 shows that

$$\forall h, \forall K \in \mathcal{T}_h, \quad c_1 h^d \sum_{i \in \mathcal{Y}_K} |\mathcal{W}_i|^p \leq \|w_h\|_{L^p(K)}^p \leq c_2 h^d \sum_{i \in \mathcal{Y}_K} |\mathcal{W}_i|^p,$$

where \mathcal{Y}_K is the set of indices i such that the intersection of K with the support of the global shape function ψ_i has non-zero measure. Summing over the elements yields

$$c_1 h^d \sum_{K \in \mathcal{T}_h} \sum_{i \in \mathcal{Y}_K} |\mathcal{W}_i|^p \leq \|w_h\|_{L^p(\Omega)}^p \leq c_2 h^d \sum_{K \in \mathcal{T}_h} \sum_{i \in \mathcal{Y}_K} |\mathcal{W}_i|^p.$$

Since $\{\mathcal{T}_h\}_{h>0}$ is shape-regular, it is clear that the cardinal of \mathcal{Y}_K is bounded uniformly in h ; hence, (73) holds.

(4) Estimate (74) is a direct consequence of (73).

Remark 9 The above proof can be easily adapted if the finite elements are not locally defined by the change of variable $(w_h \circ T_K^{-1})|_K \in \widehat{P}$ but by some other scaling like for Raviart–Thomas-like elements or Nédélec-like elements.

Remark 10 If $\{\mathcal{T}_h\}_{h>0}$ is not quasi-uniform, the lower bound in (73) holds with $h_{\min}^{\frac{d}{p}}$ and the upper bound holds with $h_{\max}^{\frac{d}{p}}$, where h_{\max} and h_{\min} are the largest and smallest cell diameters in the mesh, respectively.

Remark 11 When $p = 2$, it is possible to interpret $\mu_{s,\min}$, $\mu_{s,\max}$, $\mu_{t,\min}$, and $\mu_{t,\max}$ in terms of eigenvalues. Define the mass matrix $\mathcal{M}_s = (\int_{\Omega} \psi_i \psi_j)_{1 \leq i, j \leq N}$. Observe that \mathcal{M}_s is symmetric positive definite. Let $\lambda_{s,\min}$ and $\lambda_{s,\max}$ be the smallest and largest eigenvalue of \mathcal{M}_s , respectively. Likewise define the mass matrix associated with the global shape functions in V_h , i.e., $\mathcal{M}_t = (\int_{\Omega} \varphi_i \varphi_j)_{1 \leq i, j \leq N}$. The smallest and largest eigenvalue of \mathcal{M}_t are denoted by $\lambda_{t,\min}$ and $\lambda_{t,\max}$, respectively. Definitions (11) and (12) imply

$$\mu_{s,\min} = \lambda_{s,\min}^{\frac{1}{2}}, \quad \mu_{s,\max} = \lambda_{s,\max}^{\frac{1}{2}}, \quad (76)$$

$$\mu_{t,\min} = \lambda_{t,\min}^{\frac{1}{2}}, \quad \mu_{t,\max} = \lambda_{t,\max}^{\frac{1}{2}}. \quad (77)$$

A.2 Existence of large-scale discrete functions

Let $Z \subset L$ be two Banach spaces with continuous embedding. Denote by $\frac{1}{c_P}$ the norm of the embedding operator, i.e.,

$$\frac{1}{c_P} = \sup_{z \in Z} \frac{\|z\|_L}{\|z\|_Z}. \quad (78)$$

Let $\{Z_h\}_{h>0}$ be a family of finite-dimensional vector spaces equipped with the norm $\|\cdot\|_{Z_h}$. Assume $Z_h \subset L$ for all $h > 0$. Introduce $Z(h) = Z + Z_h$ and equip this space with a norm $\|\cdot\|_{Z(h)}$ such that $\|\cdot\|_{Z(h)} = \|\cdot\|_{Z_h}$ on Z_h and Z is uniformly continuously embedded in $Z(h)$. Denote by c_{inj} the uniform embedding constant, i.e., $\|z\|_{Z(h)} \leq c_{\text{inj}} \|z\|_Z$ for all $z \in Z$. Assume moreover that the family $\{Z_h\}_{h>0}$ has the approximability property, i.e.,

$$\forall z \in Z, \quad \lim_{h \rightarrow 0} \inf_{z_h \in Z_h} \|z - z_h\|_L + \|z - z_h\|_{Z(h)} = 0. \quad (79)$$

Lemma 5 *Under the above assumptions, there is h_0 such that for all $h \leq h_0$, there is $\tilde{z}_h \in Z_h \setminus \{0\}$ such that*

$$\|\tilde{z}_h\|_{Z_h} \leq 2c_P c_{\text{inj}} \|\tilde{z}_h\|_L. \quad (80)$$

Proof The definition of c_P implies that there exists $\tilde{z} \in Z \setminus \{0\}$ such that $\|\tilde{z}\|_Z \leq \frac{3}{2}c_P \|\tilde{z}\|_L$. Let $\epsilon > 0$. The approximability property implies that there is h_ϵ such that for all $h \leq h_\epsilon$, there is $\tilde{z}_h \in Z_h$ satisfying

$$\|\tilde{z} - \tilde{z}_h\|_L \leq \epsilon \|\tilde{z}\|_L, \quad \|\tilde{z} - \tilde{z}_h\|_{Z(h)} \leq \epsilon c_P c_{\text{inj}} \|\tilde{z}\|_L.$$

Then

$$\begin{aligned}\|\tilde{z}_h\|_{Z_h} &\leq \|\tilde{z} - \tilde{z}_h\|_{Z(h)} + \|\tilde{z}\|_{Z(h)} \leq \epsilon c_P c_{\text{inj}} \|\tilde{z}\|_L + c_{\text{inj}} \|\tilde{z}\|_Z \\ &\leq c_P c_{\text{inj}} \left(\epsilon + \frac{3}{2}\right) \|\tilde{z}\|_L.\end{aligned}$$

Moreover,

$$\|\tilde{z}_h\|_L \geq \|\tilde{z}\|_L - \|\tilde{z} - \tilde{z}_h\|_L \geq (1 - \epsilon) \|\tilde{z}\|_L.$$

Then

$$\frac{\|\tilde{z}_h\|_{Z_h}}{\|\tilde{z}_h\|_L} \leq c_P c_{\text{inj}} \frac{\frac{3}{2} + \epsilon}{1 - \epsilon}.$$

Conclude using $\epsilon = \frac{1}{6}$.

Remark 12 If $Z = H_0^1(\Omega)$, $L = L^2(\Omega)$, and $\|z\|_Z^2 = \int_\Omega \nabla z \cdot \nabla z$, then c_P is the square root of the first eigenvalue of the Laplace operator supplemented with homogeneous Dirichlet boundary conditions. This motivates the fact that the function \tilde{z}_h provided by Lemma 5 is termed a large-scale discrete function.

References

1. I. Babuška and A.K. Aziz. Survey lectures on the mathematical foundations of the finite element method. In *The mathematical foundations of the finite element method with applications to partial differential equations (Proc. Sympos., Univ. Maryland, Baltimore, Md., 1972)*, pages 1–359. Academic Press, New York, 1972.
2. S.C. Brenner and R.L. Scott. *The Mathematical Theory of Finite Element Methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, 1994.
3. P.G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North Holland, Amsterdam, 1978.
4. J.-P. Croisille. Finite volume box schemes and mixed methods. *M2AN Math. Model. Numer. Anal.*, 34(2):1087–1106, 2000.
5. J.-P. Croisille and I. Greff. Some nonconforming mixed box schemes for elliptic problems. *Numer. Methods Partial Differential Equations*, 18(3):355–373, 2002.
6. A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements*, volume 159 of *Applied Mathematical Series*. Springer-Verlag, New York, 2004.
7. V. Girault and P.-A. Raviart. *Finite Element Methods for Navier-Stokes Equations. Theory and Algorithms*. Springer Series in Computational Mathematics. Springer-Verlag, Berlin, 1986.
8. G.H. Golub and C.F. van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore, second edition, 1989.
9. C. Johnson, U. Nävert, and J. Pitkäranta. Finite element methods for linear hyperbolic equations. *Comput. Methods Appl. Mech. Engrg.*, 45:285–312, 1984.

10. J. Nečas. Sur une méthode pour résoudre les équations aux dérivées partielles de type elliptique, voisine de la variationnelle. *Ann. Scuola Norm. Sup. Pisa*, 16:305–326, 1962.
11. Y. Saad. *Iterative Methods for Sparse Linear Systems*. PWS Publishing Company, Boston, 1996.
12. K. Yosida. *Functional Analysis*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the sixth (1980) edition.