ESAIM: M2AN
Vol. 40, N° 1, 2006, pp. 29–48
DOI: 10.1051/m2an:2006006

ESAIM: Mathematical Modelling and Numerical Analysis
www.edpsciences.org/m2an

EVALUATION OF THE CONDITION NUMBER IN LINEAR SYSTEMS ARISING IN FINITE ELEMENT APPROXIMATIONS

Alexandre Ern[1] and Jean-Luc Guermond[2]

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 finite element approximations of elliptic PDE's in variational and in mixed form, and to first-order PDE's approximated using the Galerkin–Least Squares technique or by means of a non-standard Galerkin technique in $L^1(\Omega)$. Numerical simulations are presented to illustrate the theoretical results.

Mathematics Subject Classification. 65F35, 65N30.

Received: March 7, 2005. Revised: July 6, 2005.

1. Introduction

The finite element method provides an extremely powerful tool to approximate partial differential equations arising in engineering sciences. Since the linear systems obtained with this technique are generally very large and sparse, the most practical way to solve them is to resort to iterative methods. Estimates for the convergence rate of iterative methods usually depend on the condition number of the system matrix (see, *e.g.*, [11, 14]). Although in general it is the distribution of eigenvalues rather than the condition number that controls the convergence rate of iterative methods, a study of the condition number *per se* is still of interest. In particular, it is important to assess this quantity as a function of the meshsize used in the finite element method.

It is well-known that second-order elliptic equations, *e.g.*, a Laplacian, yield stiffness matrices whose Euclidean condition number explodes as the reciprocal of the square of the meshsize; see, *e.g.*, [3]. 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

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

Keywords and phrases. Finite elements, condition number, partial differential equations, linear algebra.

[1] CERMICS, École nationale des ponts et chaussées, Champs sur Marne, 77455 Marne la Vallée Cedex 2, France. ern@cermics.enpc.fr

[2] Dept. Math, Texas A&M, College Station, TX 77843-3368, USA and LIMSI (CNRS-UPR 3152), BP 133, 91403, Orsay, France. guermond@math.tamu.edu

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

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

Recent work on the conditioning of finite element matrices has focused on upper bounds for the Euclidean condition number in the case of locally refined meshes; see, *e.g.*, [1,3]. The objective of the present paper is to give upper and lower bounds on $\kappa_p(\mathcal{A})$ for $p \in [1, +\infty]$ when $\mathcal{A}$ is the stiffness matrix associated with the finite element approximation of a linear, abstract model problem posed in Banach spaces. The analysis is restricted to finite element bases that are localized in space, *i.e.*, nodal bases. The case of hierarchical and modal bases is not discussed. Technical aspects related to locally refined meshes are not addressed either.

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. For most of the examples (with the exception of the last one where the case $p = 1$ is considered), the analysis in Section 4 focuses on the case $p = 2$. 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

$$A : W \longrightarrow V. \tag{3}$$

Recall that as a consequence of the Open Mapping Theorem and the Closed Range Theorem [15], the following holds:

Lemma 2.1. *The following statements are equivalent:*

(i) *A is bijective.*

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

$$\forall w \in W, \quad \|Aw\|_V \ge \alpha \|w\|_W, \tag{4}$$

$$\forall v' \in V', \quad (A^T v' = 0) \implies (v' = 0). \tag{5}$$

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

$$\forall (w, v') \in W \times V', \qquad a(w, v') = \langle v', Aw\rangle_{V',V}, \tag{6}$$

where $\langle\cdot,\cdot\rangle_{V',V}$ denotes the duality paring. 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 2.1, henceforth referred to as the BNB Theorem [2, 8, 13], is the following:

Theorem 2.2 (Banach–Nečas–Babuška)**.** *The following statements are equivalent:*

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

$$\begin{cases} \text{Seek } u \in W \text{ such that} \\ a(u, v') = \langle v', f\rangle_{V',V}, \quad \forall v' \in V', \end{cases} \tag{7}$$

is well-posed.

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

$$\inf_{w\in W}\sup_{v'\in V'}\frac{a(w,v')}{\|w\|_W\|v'\|_{V'}}\geq\alpha, \tag{8}$$

$$\forall v'\in V',\quad (\forall w\in W,\ a(w,v')=0)\implies (v'=0). \tag{9}$$

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 $A:W\ni u\longmapsto u+u_x\in V$.

Remark 2.3. When supremun and/or infimum over sets of functions are considered, it is always implicitly understood that the zero function is excluded from the set in question whenever it makes sense. This convention is meant to alleviate the notation.

2.2. **The finite element setting**

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

To construct an approximate solution to (7), 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 $h_K=\operatorname{diam}(K)$. 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; henceforth, W_h is referred to as the solution space and V_h as the test space. 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 $W_h\subset[L^p(\Omega)]^n$ and $V_h\subset[L^{p'}(\Omega)]^n$. The spaces W_h and V_h are equipped with some norms, say $\|\cdot\|_{W_h}$ and $\|\cdot\|_{V_h}$, respectively.

Let $A:W\to V$ be an isomorphism. Problem (7) is approximated by replacing the spaces W and V' by their finite-dimensional counterparts, yielding the approximate problem:

$$\begin{cases}\text{Seek } u_h\in W_h \text{ such that}\\ a_h(u_h,v_h)=\langle v_h,f\rangle_h,\quad\forall v_h\in V_h.\end{cases} \tag{10}$$

Problem (10) involves a consistent approximation a_h of the bilinear form a and a consistent approximation of the linear form in the right-hand side. The way $\langle v_h,f\rangle_h$ is defined is not important for the present investigation. Henceforth, we assume

$$\inf_{w_h\in W_h}\sup_{v_h\in V_h}\frac{a_h(w_h,v_h)}{\|w_h\|_{W_h}\|v_h\|_{V_h}}>0. \tag{11}$$

This, together with the fact that $\dim(W_h)=\dim(V_h)$, implies that the discrete problem (10) 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 $C_{V_h}:V_h\longrightarrow\mathbb{R}^N$ 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 $C_{W_h}:W_h\longrightarrow\mathbb{R}^N$ 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 $\big(a_h(\psi_j,\varphi_i)\big)_{1\le i,j\le N}$. This definition is such that for all $(w_h,v_h)\in W_h\times V_h$, $(C_{V_h}v_h,\mathcal{A}C_{W_h}w_h)_N=a_h(w_h,v_h)$. The discrete problem (10) yields the linear system:

$$\left\{\begin{array}{l}\text{Seek } \mathcal{U}\in\mathbb{R}^N \text{ such that}\\ \mathcal{A}\mathcal{U}=\mathcal{F},\end{array}\right. \tag{12}$$

where the entries of $\mathcal{F}$ are $\mathcal{F}_i=\langle\varphi_i,f\rangle_h$ for $1\le i\le N$. The solution u_h to (10) 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 following positive constants

$$m_{s,p,h}=\inf_{w_h\in W_h}\frac{\|w_h\|_{L^p(\Omega)}}{\|C_{W_h}w_h\|_p},\qquad M_{s,p,h}=\sup_{w_h\in W_h}\frac{\|w_h\|_{L^p(\Omega)}}{\|C_{W_h}w_h\|_p}, \tag{13}$$

$$m_{t,p,h}=\inf_{v_h\in V_h}\frac{\|v_h\|_{L^{p'}(\Omega)}}{\|C_{V_h}v_h\|_{p'}},\qquad M_{t,p,h}=\sup_{v_h\in V_h}\frac{\|v_h\|_{L^{p'}(\Omega)}}{\|C_{V_h}v_h\|_{p'}}. \tag{14}$$

Here, the subscripts s and t refer to the solution space and the test space, respectively. The constants introduced in (13)–(14) are such that

$$\forall w_h\in W_h,\quad m_{s,p,h}\|\mathcal{W}\|_p\le\|w_h\|_{L^p(\Omega)}\le M_{s,p,h}\|\mathcal{W}\|_p, \tag{15}$$

$$\forall v_h\in V_h,\quad m_{t,p,h}\|\mathcal{V}\|_{p'}\le\|v_h\|_{L^{p'}(\Omega)}\le M_{t,p,h}\|\mathcal{V}\|_{p'}, \tag{16}$$

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

$$\kappa_{s,p,h}=\frac{M_{s,p,h}}{m_{s,p,h}},\qquad \kappa_{t,p,h}=\frac{M_{t,p,h}}{m_{t,p,h}},\qquad \kappa_{p,h}=\sqrt{\kappa_{s,p,h}\kappa_{t,p,h}}. \tag{17}$$

It is possible to estimate $m_{s,p,h}$ and $M_{s,p,h}$ (resp. $m_{t,p,h}$ and $M_{t,p,h}$) 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, $\kappa_{s,p,h}$ and $\kappa_{t,p,h}$ 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:

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

$$\omega_{p,h}=\sup_{w_h\in W_h}\sup_{v_h\in V_h}\frac{a_h(w_h,v_h)}{\|w_h\|_{L^p(\Omega)}\|v_h\|_{L^{p'}(\Omega)}}. \tag{19}$$

A first result is the following:

Theorem 3.1. *Under the above assumptions,*

$$\forall h, \quad \kappa_{p,h}^{-2}\,\frac{\omega_{p,h}}{\alpha_{p,h}} \le \kappa_p(\mathcal{A}) \le \kappa_{p,h}^{2}\,\frac{\omega_{p,h}}{\alpha_{p,h}}. \tag{20}$$

Proof. (1) Upper bound on $\|\mathcal{A}\|_p$. Consider $\mathcal{W} \in \mathbb{R}^N$. Then, owing to definition (19) and using the notation $C_{V_h} v_h = \mathcal{V}$ and $C_{W_h} w_h = \mathcal{W}$,

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

Using inequalities (15)–(16) yields $\|\mathcal{A}\mathcal{W}\|_p \le \omega_{p,h}\, M_{s,p,h}\, M_{t,p,h}\, \|\mathcal{W}\|_p$. That is to say,

$$\|\mathcal{A}\|_p \le \omega_{p,h}\, M_{s,p,h}\, M_{t,p,h}.$$

(2) Upper bound on $\|\mathcal{A}^{-1}\|_p$. Using again (15)–(16) together with definition (18) yields

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

Hence, setting $\mathcal{Z} = \mathcal{A}\mathcal{W}$, we infer $\alpha_{p,h}\, m_{s,p,h}\, \|\mathcal{A}^{-1}\mathcal{Z}\|_p \le m_{t,p,h}^{-1}\, \|\mathcal{Z}\|_p$. Since $\mathcal{Z}$ is arbitrary, this means

$$\|\mathcal{A}^{-1}\|_p \le \frac{1}{\alpha_{p,h}}\, m_{s,p,h}^{-1}\, m_{t,p,h}^{-1}.$$

The upper bound in (20) 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

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

Hence,

$$\frac{1}{\alpha_{p,h}}\, M_{t,p,h}^{-1}\, M_{s,p,h}^{-1} \le \|\mathcal{A}^{-1}\|_p.$$

(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

$$m_{s,p,h}\|\mathcal{W}\|_p \le \|w_h\|_{L^p(\Omega)} = \frac{1}{\omega_{p,h}} \sup_{v_h\in V_h} \frac{a_h(w_h,v_h)}{\|v_h\|_{L^{p'}(\Omega)}}$$
$$= \frac{1}{\omega_{p,h}} \sup_{v_h\in V_h} \frac{(\mathcal{A}\mathcal{W},\mathcal{V})_N}{\|\mathcal{V}\|_{p'}} \frac{\|\mathcal{V}\|_{p'}}{\|v_h\|_{L^{p'}(\Omega)}} \le \frac{1}{\omega_{p,h}} m_{t,p,h}^{-1}\|\mathcal{A}\mathcal{W}\|_p \le \frac{1}{\omega_{p,h}} m_{t,p,h}^{-1}\|\mathcal{A}\|_p\|\mathcal{W}\|_p.$$

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

$$\omega_{p,h}\, m_{s,p,h}\, m_{t,p,h} \le \|\mathcal{A}\|_p.$$

The lower bound in (20) 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:

$$\exists\gamma,\quad \begin{cases} 0 < c_{\rm inf}^{\alpha} = \liminf_{h\to 0} \alpha_{p,h}\, h^{-\gamma} < +\infty, \\ 0 < c_{\rm sup}^{\alpha} = \limsup_{h\to 0} \alpha_{p,h}\, h^{-\gamma} < +\infty, \end{cases} \tag{21}$$

$$\exists\delta,\quad \begin{cases} 0 < c_{\rm inf}^{\omega} = \liminf_{h\to 0} \omega_{p,h}\, h^{\delta} < +\infty, \\ 0 < c_{\rm sup}^{\omega} = \limsup_{h\to 0} \omega_{p,h}\, h^{\delta} < +\infty. \end{cases} \tag{22}$$

As a consequence of Theorem 3.1, we deduce the following:

Theorem 3.2. *Under the assumptions* (21)–(22)*, the following holds true: for all* $\epsilon\in\,]0,1[$*, there is* h_ϵ *such that for all* $h\le h_\epsilon$*,*

$$(1-\epsilon)\,\frac{c_{\rm inf}^{\omega}}{c_{\rm sup}^{\alpha}}\,\kappa_{p,h}^{-2}\,h^{-\gamma-\delta} \le \kappa_p(\mathcal{A}) \le (1+\epsilon)\,\frac{c_{\rm sup}^{\omega}}{c_{\rm inf}^{\alpha}}\,\kappa_{p,h}^{2}\,h^{-\gamma-\delta}. \tag{23}$$

Proof. Let $\epsilon\in\,]0,1[$.
(1) There is h_ϵ such that for all $h\le h_\epsilon$, $(1-\frac{\epsilon}{3})c_{\rm inf}^{\alpha}h^{\gamma} \le \alpha_{p,h}$ and $\omega_{p,h} \le \left(1+\frac{\epsilon}{3}\right) c_{\rm sup}^{\omega}h^{-\delta}$. Then, apply Theorem 3.1 to deduce the upper bound.
(2) Owing to the definition of $c_{\rm sup}^{\alpha}$, there is h_ϵ such that for all $0<h\le h_\epsilon$, there is $w_h\in W_h$ satisfying

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

Then, proceed as in step (3) of the proof of Theorem 3.1 to derive the lower bound

$$\|\mathcal{A}^{-1}\|_p \ge (1+\frac{\epsilon}{2})^{-1}\,(c_{\rm sup}^{\alpha})^{-1}\,M_{s,p,h}^{-1}\,M_{t,p,h}^{-1}\,h^{-\gamma}.$$

(3) The definition of $c_{\rm inf}^{\omega}$ implies the existence of h_ϵ such that for all $0<h\le h_\epsilon$, there is $w_h\in W_h$ satisfying

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

Then, proceed as in step (4) of the proof of Theorem 3.1 to infer $\|\mathcal{A}\|_p \ge (1-\frac{\epsilon}{2})\,c_{\rm inf}^{\omega}\,m_{s,p,h}\,m_{t,p,h}\,h^{-\delta}$. The lower bound on $\kappa_p(\mathcal{A})$ then results from the above estimates. □

Remark 3.3. Observe that in (20) and (23) the lower bound is multiplied by $\kappa_{p,h}^{-2}$ and the upper bound is multiplied by $\kappa_{p,h}^2$; hence, these estimates are sharp only if $\kappa_{p,h} = \sqrt{\kappa_{s,p,h}\kappa_{t,p,h}}$ is uniformly bounded with respect to h. It is shown in the appendix that this holds true when the global shape functions $\{\varphi_1, \dots \varphi_N\}$ and $\{\psi_1, \dots \psi_N\}$ spanning V_h and W_h respectively have localized supports. The definition of $\alpha_{p,h}$ and $\omega_{p,h}$ must be modified if the bases of V_h and W_h are hierarchical.

3.2. **Estimates based on natural stability norms**

Introduce the quantities

$$\alpha_h = \inf_{w_h \in W_h} \sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|w_h\|_{W_h}\|v_h\|_{V_h}}, \tag{24}$$

$$\omega_h = \sup_{w_h \in W_h} \sup_{v_h \in V_h} \frac{a_h(w_h, v_h)}{\|w_h\|_{W_h}\|v_h\|_{V_h}}. \tag{25}$$

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.

For this purpose, we make the following technical assumptions:

$$\exists c_{sP} > 0, \quad \forall w_h \in W_h, \quad c_{sP}\|w_h\|_{L^p(\Omega)} \le \|w_h\|_{W_h}, \tag{26}$$

$$\exists c_{tP} > 0, \quad \forall v_h \in V_h, \quad c_{tP}\|v_h\|_{L^{p'}(\Omega)} \le \|v_h\|_{V_h}, \tag{27}$$

$$\exists s > 0, \exists c_{sI}, \quad \forall w_h \in W_h, \quad \|w_h\|_{W_h} \le c_{sI} h^{-s}\|w_h\|_{L^p(\Omega)}, \tag{28}$$

$$\exists t > 0, \exists c_{tI}, \quad \forall v_h \in V_h, \quad \|v_h\|_{V_h} \le c_{tI} h^{-t}\|v_h\|_{L^{p'}(\Omega)}. \tag{29}$$

Estimates (26) and (27) 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 $W_h \subset [L^p(\Omega)]^n$ and $V_h \subset [L^{p'}(\Omega)]^n$ are uniformly continuous. Furthermore, (28) and (29) 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 3.1, we deduce the following:

Corollary 3.4. *Under the assumptions* (26)–(29)*, the following bound holds:*

$$\forall h, \quad \kappa_p(\mathcal{A}) \le \kappa_{s,p,h}\frac{c_{sI}}{c_{sP}}\,\kappa_{t,p,h}\frac{c_{tI}}{c_{tP}}\,\frac{\omega_h}{\alpha_h}\,h^{-s-t}. \tag{30}$$

Proof. Let us estimate $\alpha_{p,h}$ and $\omega_{p,h}$.
(1) It is clear that

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

Hence $\alpha_{p,h} \ge c_{sP}\, c_{tP}\, \alpha_h$.
(2) Moreover,

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

Hence, $\omega_{p,h} \le c_{sI}\, c_{tI}\, \omega_h\, h^{-s-t}$.
(3) Conclude using Theorem 3.1. □

Remark 3.5. It may happen that (30) is not sharp; see Section 4.3 and (68).

In addition to (26)–(29), we assume the following:

$$\exists \mu, \quad \begin{cases} 0 < d_{\inf}^{\alpha} = \liminf_{h\to 0} \dfrac{\alpha_{p,h}}{\alpha_h} h^{\mu} < +\infty, \\ 0 < d_{\sup}^{\alpha} = \limsup_{h\to 0} \dfrac{\alpha_{p,h}}{\alpha_h} h^{\mu} < +\infty, \end{cases} \tag{31}$$

$$\exists \nu, \quad \begin{cases} 0 < d_{\inf}^{\omega} = \liminf_{h\to 0} \dfrac{\omega_{p,h}}{\omega_h} h^{s+t-\nu} < +\infty, \\ 0 < d_{\sup}^{\omega} = \limsup_{h\to 0} \dfrac{\omega_{p,h}}{\omega_h} h^{s+t-\nu} < +\infty. \end{cases} \tag{32}$$

The constants μ and ν are meant to measure the possible default to optimality of Corollary 3.4. Proceeding as in the proof of Theorem 3.2, it is clear that the following holds true:

Corollary 3.6. *Under the assumptions* (26)–(29) *and* (31)–(32)*, the following holds true: for all* $\epsilon \in]0,1[$*, there is* h_ϵ *such that for all* $h \le h_\epsilon$*,*

$$(1-\epsilon)\,\frac{d_{\inf}^{\omega}}{d_{\sup}^{\alpha}}\,\kappa_{p,h}^{-2}\,\frac{\omega_h}{\alpha_h}\,h^{-s-t+\mu+\nu} \le \kappa_p(\mathcal{A}) \le (1+\epsilon)\,\frac{d_{\sup}^{\omega}}{d_{\inf}^{\alpha}}\,\kappa_{p,h}^{2}\,\frac{\omega_h}{\alpha_h}\,h^{-s-t+\mu+\nu}. \tag{33}$$

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 $A : W \ni w \longmapsto -\Delta w \in V$. Clearly $A : W \to V$ 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-conforming. We assume that W_h is such that there is c, independent of h, such that the following global inverse inequality holds:

$$\forall w_h \in W_h, \quad \|\nabla w_h\|_{L^2(\Omega)} \le c\, h^{-1}\, \|w_h\|_{L^2(\Omega)}. \tag{34}$$

This hypothesis holds whenever W_h is a finite element space constructed using a quasi-uniform mesh family; see, *e.g.*, [4, 5, 8, 10].

Consider the approximate problem:

$$\begin{cases} \text{Seek } u_h \in W_h \text{ such that} \\ a(u_h, v_h) = (f, v_h)_{L^2(\Omega)}, \quad \forall v_h \in W_h, \end{cases} \tag{35}$$

for some data $f \in L^2(\Omega)$. Let $\mathcal{A}$ be the stiffness matrix associated with (35). The main result concerning the Euclidean condition number of $\mathcal{A}$ is the following:

Theorem 4.1. *If the mesh family* $\{\mathcal{T}_h\}_{h>0}$ *is quasi-uniform and* (34) *holds, there are* $0 < c_1 \le c_2$*, both independent of* h*, such that*

$$c_1 \kappa_{2,h}^{-2}\, h^{-2} \le \kappa_2(\mathcal{A}) \le c_2 \kappa_{2,h}^{2}\, h^{-2}. \tag{36}$$

Proof. (1) For $w_h \in W_h \setminus \{0\}$, define $R(w_h) = \frac{\|\nabla w_h\|^2_{L^2(\Omega)}}{\|w_h\|^2_{L^2(\Omega)}}$. Then

$$\alpha_{2,h} = \inf_{w_h \in W_h} R(w_h), \qquad \omega_{2,h} = \sup_{w_h \in W_h} R(w_h). \tag{37}$$

(2) Let $\tilde{z}_h$ be given by Lemma A.5 with $Z = H^1_0(\Omega)$, $Z_h = W_h$ equipped with the H^1-seminorm, and $L = L^2(\Omega)$. Since $R(\tilde{z}_h) \le c$, we infer $\alpha_{2,h} \le R(\tilde{z}_h) \le c$ uniformly in h. Moreover, the Poincaré inequality in $H^1_0(\Omega)$ implies that $\alpha_{2,h}$ is uniformly bounded from below away from zero.
(3) Letting w_h in (37) be one of the global shape functions in W_h, it is clear that $\omega_{2,h} \ge ch^{-2}$. Moreover, owing to the inverse inequality (34), $\omega_{2,h} \le c'h^{-2}$.
(4) To conclude, use Theorem 3.1 (or Thm. 3.2 with $\gamma = 0$ and $\delta = 2$). □

Remark 4.2. Observe that owing to the quasi-uniformity of $\{\mathcal{T}_h\}_{h>0}$ and Lemma A.1, the constant $\kappa_{2,h} = \sqrt{\kappa_{s,2,h}\kappa_{t,2,h}}$ is bounded from below and from above uniformly with respect to h when the global shape functions have localized supports.

Remark 4.3. The Euclidean condition number $\kappa_2(\mathcal{A})$ can also be estimated using Corollary 3.4. One readily verifies that α_h can be bounded from below uniformly with respect to h, ω_h can be bounded from above uniformly with respect to h, and that $s = t = 1$. Hence, (30) yields $\kappa_2(\mathcal{A}) \le ch^{-2}$, *i.e.*, the estimate is sharp. One also verifies that $\mu = \nu = 0$ in (31)–(32), confirming the optimality of Corollary 3.4.

Remark 4.4. Estimate (36) 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 the following non-standard Galerkin technique introduced in [6] to approximate the Laplacian in mixed form. Let $H(\mathrm{div};\Omega) = \{v \in [L^2(\Omega)]^d;\, \nabla\cdot v \in L^2(\Omega)\}$, $W = H(\mathrm{div};\Omega) \times H^1_0(\Omega)$, and $V = [L^2(\Omega)]^d \times L^2(\Omega)$. Introduce the operator

$$A : W \ni (u,p) \longmapsto (u + \nabla p, \nabla\cdot u) \in V. \tag{38}$$

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

$$a((u,p),(v,q)) = (u,v)_{L^2(\Omega)} + (\nabla p, v)_{L^2(\Omega)} + (\nabla\cdot u, q)_{L^2(\Omega)}. \tag{39}$$

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

The non-standard Galerkin approximation 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 $\mathcal{F}_h$, $\mathcal{F}_h^\partial$, and $\mathcal{F}_h^\mathrm{i}$ the set of faces, boundary faces, and interior faces of the mesh, respectively. Define

$$X_h = \{u_h;\, \forall K \in \mathcal{T}_h,\, u_{h|K} \in \mathbb{RT}_0;\, \forall F \in \mathcal{F}_h^\mathrm{i},\, \textstyle\int_F [\![u_h\cdot n]\!] = 0\}, \tag{40}$$

$$Y_h = \{p_h;\, \forall K \in \mathcal{T}_h,\, p_{h|K} \in \mathbb{P}_1;\, \forall F \in \mathcal{F}_h,\, \textstyle\int_F [\![p_h]\!] = 0\}, \tag{41}$$

where $\mathbb{RT}_0 = [\mathbb{P}_0]^d \oplus x\mathbb{P}_0$, $[\![u_h\cdot n]\!]$ denotes the jump of the normal component of u_h across interfaces, and $[\![p_h]\!]$ 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

$$V_h = \{(v_h,q_h);\, \forall K \in \mathcal{T}_h,\, v_{h|K} \in [\mathbb{P}_0]^d \text{ and } q_{h|K} \in \mathbb{P}_0\}, \tag{42}$$

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

$$a_h((u_h,p_h),(v_h,q_h)) = (u_h,v_h)_{L^2(\Omega)} + (\nabla\cdot u_h, q_h)_{L^2(\Omega)} + \sum_{K\in\mathcal{T}_h} (\nabla p_h, v_h)_{L^2(K)}, \tag{43}$$

the discrete problem is formulated as follows:

$$\begin{cases} \text{Seek } (u_h, p_h) \in W_h \text{ such that} \\ a_h((u_h, p_h), (v_h, q_h)) = (f, q_h)_{L^2(\Omega)}, \quad \forall (v_h, q_h) \in V_h, \end{cases} \tag{44}$$

for some data $f \in L^2(\Omega)$. Note that the approximation setting is conforming on the velocity and non-conforming on the pressure. Moreover, it is readily checked that the total number of unknowns in (44) 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

$$\|(u_h, p_h)\|_{W_h}^2 = \|u_h\|_{L^2(\Omega)}^2 + \|\nabla{\cdot}u_h\|_{L^2(\Omega)}^2 + \|p_h\|_{L^2(\Omega)}^2 + \sum_{K \in \mathcal{T}_h} \|\nabla p_h\|_{L^2(K)}^2, \tag{45}$$

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 (44) is the following:

Lemma 4.5. *There are $c > 0$ and h_0 such that for all $h \le h_0$,*

$$\inf_{(u_h, p_h) \in W_h} \sup_{(v_h, q_h) \in V_h} \frac{a_h((u_h, p_h), (v_h, q_h))}{\|(u_h, p_h)\|_{W_h} \|(v_h, q_h)\|_{V_h}} \ge c. \tag{46}$$

Proof. Since this is a non-classical result, the proof is briefly sketched; see [6] and [8] for further details.
(1) Let $(u_h, p_h) \in W_h$. Denote by $\overline{u}_h$ the function whose restriction to each element $K \in \mathcal{T}_h$ is the mean value of u_h. Use a similar notation for $\overline{p}_h$. Denote by $\nabla_h p_h$ the function whose restriction to each element $K \in \mathcal{T}_h$ is $\nabla p_{h|K}$. Set $v_h = \overline{u}_h + \nabla_h p_h$ and $q_h = 2\overline{p}_h + \nabla{\cdot}u_h$. Note that the pair (v_h, q_h) is in V_h since the gradient and the divergence terms are piecewise constant owing to the present choice of discrete spaces. Hence,

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

since $(\nabla{\cdot}u_h, \overline{p}_h)_{L^2(\Omega)} + \sum_{K \in \mathcal{T}_h} (u_h, \nabla p_h)_{L^2(K)} = 0$.
(2) For $u_h \in X_h$, one readily verifies that $\forall K \in \mathcal{T}_h$, $\forall x \in K$, $u_h(x) = \overline{u}_h + \frac{1}{d}(x - g_K)\nabla{\cdot}u_h$ where g_K is the barycenter of K. This implies that there is c, independent of h, such that

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

Hence,

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

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.*, [7, 8] for a proof)

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

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 (44). Our main result is the following:

Theorem 4.6. *If the mesh family $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform, there are $0 < c_1 \leq c_2$, both independent of h, such that*

$$c_1\kappa_{2,h}^{-2}\,h^{-1} \leq \kappa_2(\mathcal{A}) \leq c_2\kappa_{2,h}^2\,h^{-1}. \tag{47}$$

Proof. (1) Owing to (46),

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

Hence, $\alpha_{2,h} \geq c$.
(2) Take $u_h = 0$ and $p_h = \tilde{z}_h$ given by Lemma A.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 (34) 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 3.1 (or Thm. 3.2 with $\gamma = 0$ and $\delta = 1$). □

Remark 4.7. As for elliptic PDE's in variational form, $\kappa_2(\mathcal{A})$ can also be estimated using Corollary 3.4. 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, (30) yields $\kappa_2(\mathcal{A}) \leq ch^{-1}$, *i.e.*, the estimate is sharp. One also verifies that $\mu = \nu = 0$ in (31)–(32), confirming the optimality of Corollary 3.4.

Remark 4.8. It is also possible to consider a standard Galerkin approximation to the Laplacian in mixed form. In this case, the solution 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 (40) and Z_h denotes the space of piecewise constant functions. The discrete problem is (44) with the bilinear form

$$a_h((u_h,p_h),(v_h,q_h)) = (u_h,v_h)_{L^2(\Omega)} - (\nabla\cdot v_h,p_h)_{L^2(\Omega)} + (\nabla\cdot u_h,q_h)_{L^2(\Omega)}. \tag{48}$$

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 4.9. 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

$$\partial\Omega^- = \{x \in \partial\Omega;\, \beta(x)\cdot n(x) < 0\}, \qquad \partial\Omega^+ = \{x \in \partial\Omega;\, \beta(x)\cdot n(x) > 0\}, \tag{49}$$

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

$$\begin{cases} \rho u + \beta\cdot\nabla u = f, \\ u_{|\partial\Omega^-} = 0. \end{cases} \tag{50}$$

To give a mathematical meaning to (50), introduce the so-called graph space

$$G = \{w \in L^2(\Omega);\, \beta\cdot\nabla w \in L^2(\Omega)\}. \tag{51}$$

Equipped with the graph norm $\|w\|_G = \|w\|_{L^2(\Omega)} + \|\beta\cdot\nabla w\|_{L^2(\Omega)}$, G is a Hilbert space. Assume that

$$\mathcal{C}^1(\overline{\Omega}) \text{ is dense in } G, \tag{52}$$

$$\partial\Omega^- \text{ and } \partial\Omega^+ \text{ are well-separated, } \textit{i.e.}, \text{ dist}(\partial\Omega^-, \partial\Omega^+) > 0. \tag{53}$$

Hypothesis (52) amounts to a regularity assumption on Ω; for instance, it holds whenever Ω is Lipschitz. Owing to (52)–(53), it can be shown (see, *e.g.*, [9]) that the trace operator $\tau : \mathcal{C}^1(\overline{\Omega}) \ni v \longmapsto v_{|\partial\Omega} \in L^2(\partial\Omega; |\beta\cdot n|)$ extends uniquely to a continuous operator in G. Consider the spaces

$$W = \{w \in G;\, \tau(w) = 0\}, \qquad V = L^2(\Omega), \tag{54}$$

and define the differential operator

$$A : W \ni w \longmapsto \rho w + \beta\cdot\nabla w \in V. \tag{55}$$

It is clear that A is continuous. Moreover, assuming that there is $\rho_0 > 0$ such that almost everywhere in Ω,

$$\rho(x) - \tfrac{1}{2}\nabla\cdot\beta(x) \geq \rho_0 > 0, \tag{56}$$

it can be shown that $A : W \to V$ is an isomorphism; see, *e.g.*, [9].

We want to illustrate Theorem 3.1 by analyzing the Euclidean condition number of the stiffness matrix associated with the GaLS approximation of (50). To this end 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-conforming, and assume that (34) holds. 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

$$a_h(w, v) = a(w, v) + \sum_{K\in\mathcal{T}_h} \delta(h_K)(Aw, Av)_{L^2(K)}, \tag{57}$$

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 [12]:

$$\begin{cases} \text{Seek } u_h \in W_h \text{ such that} \\ a_h(u_h, v_h) = (f, v_h)_{L^2(\Omega)} + \sum_{K\in\mathcal{T}_h} \delta(h_K)(f, Av_h)_{L^2(K)}, \quad \forall v_h \in W_h. \end{cases} \tag{58}$$

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

$$\forall (w_1, w_2) \in W \times W, \quad a_s(w_1, w_2) = \tfrac{1}{2}\big((Aw_1, w_2)_{L^2(\Omega)} + (w_1, Aw_2)_{L^2(\Omega)}\big). \tag{59}$$

It is clear that a_s is positive definite since

$$\forall w \in W, \quad a_s(w,w) = a(w,w) \geq \rho_0 \|w\|^2_{L^2(\Omega)}. \tag{60}$$

The main result of this section is the following:

Theorem 4.10. *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 $\mathcal{C}^{0,1}(\overline{\Omega}_0)$. Assume that the mesh family $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform and (34) holds. Then, there are $0 < c_1 \leq c_2$ and h_0 such that for all $h \leq h_0$,*

$$c_1 \kappa_{2,h}^{-2}\, h^{-1} \leq \kappa_2(\mathcal{A}) \leq c_2 \kappa_{2,h}^2\, h^{-1}. \tag{61}$$

Proof. (1) Owing to (60),

$$\rho_0 \leq \inf_{w_h \in W_h} \frac{a_h(w_h,w_h)}{\|w_h\|^2_{L^2(\Omega)}} \leq \inf_{w_h \in W_h} \sup_{v_h \in W_h} \frac{a_h(w_h,v_h)}{\|w_h\|_{L^2(\Omega)} \|v_h\|_{L^2(\Omega)}} = \alpha_{2,h},$$

i.e., $\rho_0 \leq \alpha_{2,h}$.
(2) To derive a bound on $\alpha_{2,h}$, use Lemma A.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 A.5 implies that there exists $\tilde{c} > 0$ and $\tilde{h}$ such that for all $h \leq \tilde{h}$, there is $\tilde{z}_h \in W_h \backslash \{0\}$ satisfying $\|A\tilde{z}_h\|_{L^2(\Omega)} \leq \tilde{c} \|\tilde{z}_h\|_{L^2(\Omega)}$. Then using this bound together with

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

a direct computation using (34) shows that $\alpha_{2,h}$ is bounded uniformly with respect to h.
(3) Using again (34) it is clear that there is c, independent of h, such that $\omega_{2,h} \leq ch^{-1}$.
(4) A simple computation yields

$$a_h(w_h,w_h) \geq \int_\Omega (\rho_0 - h\|\rho\|^2_{L^\infty(\Omega)}) w_h^2 + \sum_{K \in \mathcal{T}_h} \frac{h_K}{2} \int_K |\beta \cdot \nabla w_h|^2.$$

Assume that h is small enough so that $\rho_0 - h\|\rho\|^2_{L^\infty(\Omega)} \geq 0$ and there is a mesh cell $K_0 \subset \Omega_0$. Then for all $w_h \in W_h$,

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

where $\overline{\beta}$ 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

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

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

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

(5) To conclude, use Theorem 3.1 (or Thm. 3.2 with $\gamma = 0$ and $\delta = 1$). □

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

$$\forall w \in W, \quad \begin{cases} \|w\|_{h,A}^2 = a_s(w,w) + \displaystyle\sum_{K\in\mathcal{T}_h} \delta(h_K)\|Aw\|_{L^2(K)}^2, \\ \|w\|_{h,\frac{1}{2}}^2 = \|w\|_{h,A}^2 + \displaystyle\sum_{K\in\mathcal{T}_h} h_K^{-1}\|w\|_{L^2(K)}^2. \end{cases} \tag{62}$$

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

$$\forall w \in W, \quad a_h(w,w) \geq \|w\|_{h,A}^2, \tag{63}$$

$$\forall w \in W,\ \forall w_h \in W_h, \quad a_h(w,w_h) \leq c\,\|w\|_{h,\frac{1}{2}}\|w_h\|_{h,A}, \tag{64}$$

from which the convergence analysis of the GaLS approximation directly follows; see [8] for more details.

Proposition 4.11. *Equip W_h and V_h with the norm $\|\cdot\|_{h,A}$ to define α_h and ω_h in (24)–(25). Then if the mesh family $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform, there is c, independent of h, such that*

$$\alpha_h \geq 1, \tag{65}$$

$$\omega_h \leq c\,h^{-\frac{1}{2}}, \tag{66}$$

$$s = t = \frac{1}{2}\cdot \tag{67}$$

Proof. (1) Estimate (65) is a direct consequence of (63).
(2) Owing to the quasi-uniformity hypothesis and (60),

$$\|w\|_{h,\frac{1}{2}}^2 = \|w\|_{h,A}^2 + ch^{-1}\|w\|_{L^2(\Omega)}^2 \leq \|w\|_{h,A}^2 + \frac{c}{\rho_0}h^{-1}a_s(w,w) \leq (1+\frac{c}{\rho_0}h^{-1})\|w\|_{h,A}^2.$$

The bound (66) follows readily from (64).
(3) Statement (67) is an easy consequence of (34). □

Remark 4.12. If we apply Corollary 3.4, we obtain

$$\kappa_2(\mathcal{A}) \leq c\kappa_{2,h}h^{-\frac{3}{2}}. \tag{68}$$

This result shows that Corollary 3.4 may not be optimal; in fact, one readily verifies that $\mu = 0$ and $\nu = \frac{1}{2}$ in Corollary 3.6.

4.4. First-order PDE's in L^1

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

$$\begin{cases} \rho u + u_x = f, \\ u(0) = 0, \end{cases} \tag{69}$$

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

$$W = \{w \in W^{1,1}(\Omega);\, w(0) = 0\}, \qquad V = L^1(\Omega). \tag{70}$$

Define the operator

$$A : W \ni w \longmapsto \rho w + w_x \in V. \tag{71}$$

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

$$\exists \alpha > 0,\ \forall w \in W, \quad \|Aw\|_{L^1(\Omega)} \geq \alpha \|w\|_{W^{1,1}(\Omega)}. \tag{72}$$

Define the finite element spaces

$$W_h = \{w_h \in \mathcal{C}^0(\overline{\Omega});\, \forall K \in \mathcal{T}_h,\, w_{h|K} \in \mathbb{P}_1;\, w_h(0) = 0\}, \tag{73}$$

$$V_h = \{v_h \in L^1(\Omega);\, \forall K \in \mathcal{T}_h,\, v_{h|K} \in \mathbb{P}_0\}. \tag{74}$$

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

$$\forall (w,v) \in W \times V', \quad a(w,v) = \int_0^1 (\rho w + w_x)v. \tag{75}$$

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

$$\left\{ \begin{array}{l} \text{Seek } u_h \in W_h \text{ such that} \\ a(u_h, v_h) = \int_0^1 f v_h, \quad \forall v_h \in V_h. \end{array} \right. \tag{76}$$

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 (76) is the following:

Lemma 4.13. *There is $\gamma > 0$ and h_0 such that for all $h \leq h_0$,*

$$\inf_{w_h \in W_h} \sup_{v_h \in V_h} \frac{a(w_h, v_h)}{\|w_h\|_{W^{1,1}(\Omega)} \|v_h\|_{L^\infty(\Omega)}} \geq \gamma. \tag{77}$$

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

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

The conclusion follows readily. □

Let $\{\psi_1, \ldots, \psi_N\}$ be the standard $\mathbb{P}_1$ nodal shape functions of W_h. Let $\{\varphi_1, \ldots, \varphi_N\}$ 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

$\big(a(\psi_j,\varphi_i)\big)_{1\le i,j\le N}$. The main result of this section is the following:

Theorem 4.14. *If the mesh family $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform, there are $0<c_1\le c_2$ and h_0 such that for all $h\le h_0$,*

$$c_1\kappa_{1,h}^{-2}\,h^{-1}\le\kappa_1(\mathcal{A})\le c_2\kappa_{1,h}^{2}\,h^{-1}. \tag{78}$$

Proof. (1) From Lemma 4.13, it is clear that $\alpha_{1,h}\ge\gamma$.
(2) To derive a bound on $\alpha_{1,h}$, we use Lemma A.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 A.5 implies that there exists $\tilde c>0$ and $\tilde h$ such that for all $h\le\tilde h$, there is $\tilde z_h\in Z_h\backslash\{0\}$ satisfying $\|\tilde z_h\|_{W^{1,1}(\Omega)}\le\tilde c\|\tilde z_h\|_{L^1(\Omega)}$. Since

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

one readily infers that $\alpha_{1,h}$ is bounded uniformly with respect to h.
(3) Using standard inverse inequalities yields $\omega_{1,h}\le ch^{-1}$.
(4) Let ψ_i be a shape function in W_h and set $v_h=\mathrm{sg}(\psi_{i,x})$. Then, $v_h\in V_h$, $\|v_h\|_{L^\infty(\Omega)}=1$ and

$$\begin{aligned}a(\psi_i,v_h)&\ge-\rho\|\psi_i\|_{L^1(\Omega)}+\|\psi_{i,x}\|_{L^1(\Omega)}\ge-\rho\|\psi_i\|_{L^1(\Omega)}+\tfrac{c}{h}\|\psi_i\|_{L^1(\Omega)}\\&\ge(\tfrac{c}{h}-\rho)\|\psi_i\|_{L^1(\Omega)}\|v_h\|_{L^\infty(\Omega)}.\end{aligned}$$

This implies $\omega_{1h}\ge ch^{-1}$.
(5) Apply Theorem 3.1 to conclude. □

Remark 4.15. The above result can be easily adapted to the situation where ρ 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. Results are collected in Table 1 for 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. The Euclidean condition number $\kappa_2(\mathcal{A})$ behaves like h^{-1} in agreement with Theorem 4.6. Furthermore, we observe that the condition numbers $\kappa_1(\mathcal{A})$ and $\kappa_\infty(\mathcal{A})$ behave like h^{-1} also and that both numbers approximately take the same value; this value is slightly larger than that of $\kappa_2(\mathcal{A})$.
- Case 2 (GaLS): the first-order PDE (50) posed in the unit square of $\mathbb{R}^2$ with $\rho=1$ 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 meshsize $h=0.1$, 0.05, 0.025, and 0.0125 at the boundary of Ω. The Euclidean condition number $\kappa_2(\mathcal{A})$ behaves like h^{-1} in agreement with Theorem 4.10. Furthermore, the condition number $\kappa_\infty(\mathcal{A})$ appears to behave like h^{-1} also, while the condition number $\kappa_1(\mathcal{A})$ explodes more slowly than h^{-1}. It is also observed that the Euclidean condition number takes larger values than those of the two other condition numbers, as opposed to the results obtained for the Laplacian in mixed form.
- Case 3 (NGL1): the first-order PDE (69) with $\rho=1$ 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. The condition number $\kappa_1(\mathcal{A})$ behaves like h^{-1} in agreement with Theorem 4.14. The Euclidean condition number $\kappa_2(\mathcal{A})$ appears to behave like h^{-1} also and $\kappa_2(\mathcal{A})<\kappa_1(\mathcal{A})$. Finally, owing to the particular structure of the stiffness matrix, the condition number $\kappa_\infty(\mathcal{A})$ is equal to $\kappa_1(\mathcal{A})$.

TABLE 1. Condition numbers of the stiffness matrix as a function of meshsize for the three test cases.

h^{-1}	LapMix			h^{-1}	GaLS			h^{-1}	NGL1	
	$\kappa_1(\mathcal{A})$	$\kappa_2(\mathcal{A})$	$\kappa_\infty(\mathcal{A})$		$\kappa_1(\mathcal{A})$	$\kappa_2(\mathcal{A})$	$\kappa_\infty(\mathcal{A})$		$\kappa_1(\mathcal{A})$	$\kappa_2(\mathcal{A})$
4	14.6	8.5	13.5	10	20.9	68.5	41.5	4	50.7	36.6
8	26.6	16.7	25.5	20	36.1	139.0	81.9	8	101.2	72.7
16	50.6	32.9	49.5	40	60.6	282.8	157.9	16	202.3	143.9
32	99.5	64.9	97.5	80	123.5	430.3	320.7	32	404.6	285.7
64	194.5	129.0	193.5	—				64	809.1	568.8

To sum up, we observe that in the three test cases, the numerical predictions match the theoretical results whenever available.

A. TECHNICAL RESULTS

A.1. Estimates of $\kappa_{s,p,h}$ and $\kappa_{t,p,h}$

Let $\{\mathcal{T}_h\}_{h>0}$ be a shape-regular family of meshes of Ω. Recall that the 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 $h \le c \min_{K\in\mathcal{T}_h}(h_K)$. This section collects the main estimates of $\kappa_{s,p,h}$ and $\kappa_{t,p,h}$ under the assumption that the family $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform. The proof is well-known for $p = 2$ and can be easily adapted to handle the case $p \in [1, +\infty]$. For completeness, the proof is briefly presented in the general case.

Let $\{\widehat{K}, \widehat{P}, \widehat{\Sigma}\}$ be the reference finite element on which W_h is constructed. For each cell K, denote by $T_K : \widehat{K} \longrightarrow K$ 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:

$$W_h \subset \{w_h \in [L^1(\Omega)]^n; \forall K \in \mathcal{T}_h,\ (w_h \circ T_K^{-1})_{|K} \in \widehat{P}\}. \tag{79}$$

See [4, 5, 8, 10] for more details on the construction of finite element spaces.

Lemma A.1. *If $\{\mathcal{T}_h\}_{h>0}$ is quasi-uniform, there exist $0 < c_1 \le c_2$ such that*

$$\forall h,\ \forall w_h \in W_h, \quad c_1 h^{\frac{d}{p}} \|C_{W_h} w_h\|_p \le \|w_h\|_{L^p(\Omega)} \le c_2 h^{\frac{d}{p}} \|C_{W_h} w_h\|_p. \tag{80}$$

As a result,

$$\forall h, \quad \frac{c_1}{c_2} \le \kappa_{s,p,h} \le \frac{c_2}{c_1}. \tag{81}$$

Proof. Assume $1 \le p < +\infty$. The case $p = +\infty$ can be treated similarly.
(1) Let $\{\widehat{\theta}_1, \dots, \widehat{\theta}_{n_{\rm sh}}\}$ be the local shape functions for the reference finite element. Denote by $\mathcal{S}^{n_{\rm sh}}$ the unit sphere in $\mathbb{R}^{n_{\rm sh}}$ for the $\|\cdot\|_p$-norm and define the operator

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

The operator ψ is clearly continuous. Moreover, since $\mathcal{S}^{n_{\rm sh}}$ 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 $\eta \in \mathcal{S}^{n_{\rm sh}}$ such that $\psi(\eta) = 0$, yielding $\sum_{k=1}^{n_{\rm sh}} \eta_k \widehat{\theta}_k = 0$. Since $\{\widehat{\theta}_1, \dots, \widehat{\theta}_{n_{\rm sh}}\}$ is a basis, this implies $\eta_1 = \dots = \eta_{n_{\rm sh}} = 0$, contradicting the fact that $\eta \in \mathcal{S}^{n_{\rm sh}}$. Therefore, $\widehat{c}_1 > 0$. Consider now $\widehat{\mathcal{U}} \in \mathbb{R}^{n_{\rm sh}}$ with $\widehat{\mathcal{U}} \neq 0$. Let $\widehat{u} = \sum_{i=1}^{n_{\rm sh}} \widehat{\mathcal{U}}_i \widehat{\theta}_i$ and $\eta_i(\widehat{u}) = \widehat{\mathcal{U}}_i / \|\widehat{\mathcal{U}}\|_p$

for $1 \le i \le n_{\rm sh}$. Clearly, $\eta(\widehat{u}) = (\eta_i(\widehat{u}))_{1\le i\le n_{\rm sh}}$ is in $\mathcal{S}^{n_{\rm sh}}$. Since $\psi(\eta(\widehat{u})) = \|\widehat{u}\|_{L^p(\widehat{K})}/\|\widehat{\mathcal{U}}\|_p$, the following inequalities hold:

$$\forall \widehat{\mathcal{U}} \in \mathbb{R}^{n_{\rm sh}}, \quad \widehat{c}_1 \, \|\widehat{\mathcal{U}}\|_p \le \|\widehat{u}\|_{L^p(\widehat{K})} \le \widehat{c}_2 \, \|\widehat{\mathcal{U}}\|_p. \tag{82}$$

(2) Consider now an arbitrary element K in the mesh. Denote by $T_K : \widehat{K} \to K$ the corresponding transformation and by $\{\theta_1, \ldots, \theta_{n_{\rm sh}}\}$ the local shape functions. For $\mathcal{U} \in \mathbb{R}^{n_{\rm sh}}$, set $u = \sum_{i=1}^{n_{\rm sh}} \mathcal{U}_i \theta_i$ and $\widehat{u} = u \circ T_K$. Observing that $\widehat{\mathcal{U}} = \mathcal{U}$ and changing variables in the integral in (82) yields

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

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

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

(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

$$\forall h, \ \forall K \in \mathcal{T}_h, \quad c_1 h^d \sum_{i\in\Upsilon_K} |\mathcal{W}_i|^p \le \|w_h\|^p_{L^p(K)} \le c_2 h^d \sum_{i\in\Upsilon_K} |\mathcal{W}_i|^p,$$

where Υ_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

$$c_1 h^d \sum_{K\in\mathcal{T}_h} \sum_{i\in\Upsilon_K} |\mathcal{W}_i|^p \le \|w_h\|^p_{L^p(\Omega)} \le c_2 h^d \sum_{K\in\mathcal{T}_h} \sum_{i\in\Upsilon_K} |\mathcal{W}_i|^p.$$

Since $\{\mathcal{T}_h\}_{h>0}$ is shape-regular, it is clear that the cardinal of Υ_K is bounded uniformly in h; hence, (80) holds. (4) Estimate (81) is a direct consequence of (80). □

Remark A.2. 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 A.3. If $\{\mathcal{T}_h\}_{h>0}$ is not quasi-uniform, the lower bound in (80) holds with $h_{\rm min}^{\frac{d}{p}}$ and the upper bound holds with $h_{\rm max}^{\frac{d}{p}}$, where $h_{\rm max}$ and $h_{\rm min}$ are the largest and smallest cell diameters in the mesh, respectively; see, *e.g.*, [1].

Remark A.4. When $p = 2$, it is possible to interpret $m_{s,p,h}$, $M_{s,p,h}$, $m_{t,p,h}$, and $M_{t,p,h}$ in terms of eigenvalues. Define the mass matrix $\mathcal{M}_s = (\int_\Omega \psi_i \psi_j)_{1\le i,j\le N}$. Observe that $\mathcal{M}_s$ is symmetric positive definite. Let λ_s and Λ_s 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.*, $\mathcal{M}_t = (\int_\Omega \varphi_i \varphi_j)_{1\le i,j\le N}$. The smallest and largest eigenvalue of $\mathcal{M}_t$ are denoted by λ_t and Λ_t, respectively. Definitions (13) and (14) imply

$$m_{s,2,h} = \lambda_s^{\frac{1}{2}}, \qquad M_{s,2,h} = \Lambda_s^{\frac{1}{2}}, \tag{83}$$

$$m_{t,2,h} = \lambda_t^{\frac{1}{2}}, \qquad M_{t,2,h} = \Lambda_t^{\frac{1}{2}}. \tag{84}$$

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.*,

$$\frac{1}{c_P} = \sup_{z\in Z} \frac{\|z\|_L}{\|z\|_Z}. \tag{85}$$

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 $c_{\rm inj}$ the uniform embedding constant, *i.e.*, $\|z\|_{Z(h)} \le c_{\rm inj}\|z\|_Z$ for all $z \in Z$. Assume moreover that the family $\{Z_h\}_{h>0}$ has the following approximability property:

$$\forall z \in Z, \quad \lim_{h\to 0} \inf_{z_h\in Z_h} \|z - z_h\|_L + \|z - z_h\|_{Z(h)} = 0. \tag{86}$$

Lemma A.5. *Under the above assumptions, there is h_0 such that for all $h \le h_0$, there is $\tilde z_h \in Z_h\backslash\{0\}$ such that*

$$\|\tilde z_h\|_{Z_h} \le 2c_P c_{\rm inj}\|\tilde z_h\|_L. \tag{87}$$

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

$$\|\tilde z - \tilde z_h\|_L \le \epsilon\|\tilde z\|_L, \qquad \|\tilde z - \tilde z_h\|_{Z(h)} \le \epsilon c_P c_{\rm inj}\|\tilde z\|_L.$$

Then,

$$\|\tilde z_h\|_{Z_h} \le \|\tilde z - \tilde z_h\|_{Z(h)} + \|\tilde z\|_{Z(h)} \le \epsilon c_P c_{\rm inj}\|\tilde z\|_L + c_{\rm inj}\|\tilde z\|_Z \le c_P c_{\rm inj}(\epsilon + \tfrac{3}{2})\|\tilde z\|_L.$$

Moreover,

$$\|\tilde z_h\|_L \ge \|\tilde z\|_L - \|\tilde z - \tilde z_h\|_L \ge (1-\epsilon)\|\tilde z\|_L.$$

Then,

$$\frac{\|\tilde z_h\|_{Z_h}}{\|\tilde z_h\|_L} \le c_P c_{\rm inj}\frac{\frac{3}{2}+\epsilon}{1-\epsilon}.$$

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

Remark A.6. 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 A.5 is termed a large-scale discrete function.

References

[1] M. Ainsworth, W. McLean and T. Tran, The conditioning of boundary element equations on locally refined meshes and preconditioning by diagonal scaling. *SIAM J. Numer. Anal.* **36** (1999) 1901–1932.

[2] 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)*. Academic Press, New York (1972) 1–359.

[3] R.E. Bank and L.R. Scott, On the conditioning of finite element equations with highly refined meshes. *SIAM J. Numer. Anal.* **26** (1989) 1383–1384.

[4] S.C. Brenner and R.L. Scott, *The Mathematical Theory of Finite Element Methods*. Springer, New York, *Texts Appl. Math.* **15** (1994).

[5] P.G. Ciarlet, *The Finite Element Method for Elliptic Problems*. North Holland, Amsterdam (1978).

[6] J.-P. Croisille, Finite volume box schemes and mixed methods. *ESAIM: M2AN* **34** (2000) 1087–1106.

[7] J.-P. Croisille and I. Greff, Some nonconforming mixed box schemes for elliptic problems. *Numer. Methods Partial Differential Equations* **18** (2002) 355–373.

[8] A. Ern and J.-L. Guermond, *Theory and Practice of Finite Elements*, Springer-Verlag, New York. *Appl. Math. Ser.* **159** (2004)
[9] A. Ern and J.-L. Guermond, Discontinuous Galerkin methods for Friedrichs' systems. I. General theory. *SIAM J. Numer. Anal.* (2005) (in press).
[10] 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).
[11] G.H. Golub and C.F. van Loan, *Matrix Computations*. John Hopkins University Press, Baltimore, second edition (1989).
[12] C. Johnson, U. Nävert and J. Pitkäranta, Finite element methods for linear hyperbolic equations. *Comput. Methods Appl. Mech. Engrg.* **45** (1984) 285–312.
[13] 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** (1962) 305–326.
[14] Y. Saad, *Iterative Methods for Sparse Linear Systems*. PWS Publishing Company, Boston (1996).
[15] K. Yosida, *Functional Analysis*, Classics in Mathematics. Springer-Verlag, Berlin (1995). Reprint of the sixth edition (1980).