Numerical Analysis (Leture Note)

Edited by Shin, Byeong Chun

2001. 5.

A. Quarteroni and A. Valli : Numerical approximation of PDE

C. Johnson : Numerical solution of PDE by FEM

S.C. Brenner and L.R. Scott : The mathematical theory of FEMs

P.G. Ciarlet : The Finite element method for elliptic problems

V. Girault and P.-A. Raviart : Finite element methods for Navier-Stokes equations

W. Hackbusch : Iterative solution of large sparse systems of equations

Contents

0 Introduction to Finite element methods

Let Ω be a bounded open domain in $\mathbb{R}^d$ and $\partial\Omega = \Gamma_D \cap \Gamma_N$ its Lipschitz continuous boundary. Consider the simple model problem:

$$\text{(D)} \qquad \begin{cases} -\Delta u + \mathbf{b}\cdot\nabla u + c\,u &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \Gamma_D, \\ \frac{\partial u}{\partial n} &= g \quad \text{on } \Gamma_N \end{cases}$$

where $\Gamma_D \neq \emptyset$ and n is the unit normal vector to G_N.

Let $V = H^1_D(\Omega)$ where

$$H^1_D(\Omega) = \{ v \in H^1(\Omega) \,:\, v|_{G_D} = 0 \}.$$

Variational Formulation : Find $u \in V = H^1_D(\Omega)$ such that

$$\text{(V)} \qquad a(u,v) = \mathcal{F}(v) := (f,v) + \langle g, v\rangle_{G_N}, \quad \forall\, v \in V$$

where the bilinear form $a(\cdot,\cdot)$ is given by

$$a(u,v) = (\nabla u, \nabla v).$$

Let V_h be a finite dimensional subspace of $V = H^1_D(\Omega)$.

Galerkin Approximation : Find $u_h \in V_h$ such that

$$\text{(V}_h\text{)} \qquad a_h(u_h, v_h) = \mathcal{F}_h(v_h), \quad \forall\, v_h \in V_h,$$

where $a_h(\cdot,\cdot)$ and $\mathcal{F}_h(\cdot)$ are appropriate approximates of $a(\cdot,\cdot)$ and $\mathcal{F}(\cdot)$, respectively.

The existence and uniqueness of the solution to (V) or (V$_h$) can be verified by Lax-Milgram Lemma under conditions of the ellipticity and continuity of $a(\cdot,\cdot)$ or $a_h(\cdot,\cdot)$, and the continuity of $\mathcal{F}(\cdot)$ or $\mathcal{F}_h(\cdot)$.

Let $\{\phi_j\}_{j=1}^{N_h}$ be a basis of V_h. Then, for any $u_h \in V_h$, it can be represented by $u_h = \sum_{j=1}^{N_h} u_j \phi_j$. By substituting u_h and replacing v by ϕ_i in (V$_h$), we have the following linear system

$$AU = F,$$

where $A(i,j) = a_h(\phi_j, \phi_i)$, $F(i) = \mathcal{F}_h(\phi_i)$, $U(i) = u_i$.

Using direct methods or various iteration methods, we solve $AU = F$. (Jacobi, Gauss-Seidel, CGM, Multi-Grid, GmRes)

According to the finite dimensional space V_h, we have the error estimates $\|u - u_h\|_0$ and $\|u - u_h\|_1$. (Interpolation errors, Projection errors, Strang Lemma)

1 Preliminaries

1.1 Hilbert and Banach spaces

Let V be a (real) linear vector space.

Definition (Inner product)**.** A *scalar product* (or an inner product) on V is a linear map $(\cdot,\cdot) : V \times V \longrightarrow \mathbb{R}$ such that

(a) $(w, v) = (v, w) \quad \forall\, w, v \in V$,

(b) $(v, v) \geq 0 \quad \forall\, v \in V$,

(c) $(v, v) = 0$ iff $v = 0$.

Definition (Norm)**.** A *norm* is a map $\|\cdot\| : V \longrightarrow \mathbb{R}$ such that

(a) $\|v\| \geq 0 \quad \forall\, v \in V$,

(b) $\|cv\| = |c|\,\|v\| \quad \forall\, c \in \mathbb{R}, \quad v \in V$,

(c) $\|v + w\| \leq \|v\| + \|w\| \quad \forall\, v, w \in V$,

(d) $\|v\| = 0$ iff $v = 0$.

Definition (Norm equivalence)**.** Two norms $\|\cdot\|$ and $|||\cdot|||$ on V are *equivalent* if there exists $C_1, C_2 > 0$ such that

$$C1\|v\| \leq |||v||| \leq C_2\|v\| \quad \forall\, v \in V.$$

Definition (Banach, Hilbert Spaces)**.** A liner space V equipped with a scalar product (or a norm) is called *pre-Hilbert* (or *normed*) space.
If any Cauchy sequence in a pre-Hilbert (or normed) space V is convergent, then V is called a *Hilbert* (or *Banach*) space.

(Schwarz inequality)**.** In any Hilbert space, the *Schwarz inequality* holds:

$$|(w, v)| \leq \|w\|\,\|v\| \quad \forall\, w, v \in V.$$

Definition (Dual Spaces)**.** Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed spaces. Denoted by $\mathcal{L}(V; W)$ the set of linear continuous functional from V into W.

For $L \in \mathcal{L}(V; W)$, define the norm:

$$\|L\|_{\mathcal{L}(V;W)} = \sup_{\substack{v \in V \\ v \neq 0}} \frac{\|Lv\|_W}{\|v\|_V}.$$

Then, $\mathcal{L}(V; W)$ is a normed space.

- If W is a Banace space, then $\mathcal{L}(V;W)$ is also a Banach space.
- If $W = \mathbb{R}$, then the space $\mathcal{L}(V;\mathbb{R})$ is called the *dual space* of V and denoted by V'.
- The bilinear form $\langle \cdot, \cdot \rangle$ from $V' \times V$ into $\mathbb{R}$ defined by $\langle L, v \rangle := L(v)$ is called the *duality pairing* between V' and V.

(Weak convergence).

A sequence $\{v_n\}$ in V converges to v weakly if $\langle L, v_n \rangle \to \langle L, v \rangle \quad \text{as} \quad n \to \infty \quad \text{for all } L \in V'$.

(Weak* convergence).

A sequence $\{L_n\}$ in V' converges to L weakly* if $\langle L_n, v \rangle \to \langle L, v \rangle \quad \text{as} \quad n \to \infty \quad \text{for all } v \in V$.

1.2 $L^p(\Omega)$ Spaces

Let Ω be an open set in $\mathbb{R}^d$ and let $1 \le p \le \infty$.

Define

$$\|v\|_{L^p(\Omega)} := \left(\int_\Omega |v(x)|^p \, dx \right)^{\frac{1}{p}}, \quad 1 \le p < \infty,$$

$$\|v\|_{L^\infty(\Omega)} := \sup\{|v(x)| \,:\, x \in \Omega\}.$$

Denote by

$$L^p(\Omega) = \{v \,:\, \|v\|_{L^p(\Omega)} < \infty\}.$$

The space $L^2(\Omega)$ is a Hilbert space, endowed with the scalar product

$$(w, v)_{L^2(\Omega)} := \int_\Omega w(x)v(x) \, dx.$$

Denote by

$$\| \cdot \| := \| \cdot \|_0 = \| \cdot \|_{0,\Omega} = \| \cdot \|_{L^2(\Omega)}, \quad (\cdot, \cdot) := (\cdot, \cdot)_0 = (\cdot, \cdot)_{0,\Omega} = (\cdot, \cdot)_{L^2(\Omega)}.$$

(Hölder Inequality). For $w \in L^p(\Omega)$, $v \in L^q(\Omega)$ with $\frac{1}{p} + \frac{1}{q} = 1$, $1 \le p < \infty$,

$$\left| \int_\Omega w(x)v(x) \, dx \right| \le \|w\|_{L^p(\Omega)} \, \|v\|_{L^q(\Omega)}.$$

1.3 Distribution

Let $C_0^\infty := \mathcal{D}(\Omega)$ be the space of infinitely differentiable function having compact support. Denote by, for $\alpha = (\alpha_1, \cdots, \alpha_d)$ $\alpha_i \geq 0$,

$$D^\alpha v := \frac{\partial^{|\alpha|} v}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}, \quad v \in \mathcal{D}(\Omega)$$

where $|\alpha| = \alpha_1 + \cdots + \alpha_d$.

Definition (Distribution)**.** A sequence $v_n \in \mathcal{D}(\Omega)$ *converges* to $v \in \mathcal{D}(\Omega)$ if there exists a compact subset $K \subset \Omega$ such that v_n vanishes outside K for each n and for every α, $D^\alpha v_n$ converges to $D^\alpha v$ uniformly in Ω.
Denoted by $\mathcal{D}'(\Omega)$ the dual space of $\mathcal{D}(\Omega)$ and its elements are called *distribution* which is continuous in the above sense.

- Each function $w \in L^p(\Omega)$ is a distribution:

$$v \longrightarrow \int_\Omega w(x)v(x)\,dx, \quad \forall v \in \mathcal{D}(\Omega).$$

- The Dirac functional δ does not belong $L^p(\Omega)$ but it is also a distribution such that

$$v \longrightarrow \delta(v) = \int_{-\infty}^{\infty} \delta(t)v(t)\,dt = v(0), \quad \forall v \in \mathcal{D}(\Omega)$$

where the Dirac delta function is defined as follows. Define

$$\delta_\tau(t) = \begin{cases} \frac{1}{2\tau} & (-\tau < t < \tau) \\ 0 & (t \leq -\tau, \quad t \geq \tau). \end{cases}$$

Then

$$\int_{-\infty}^{\infty} \delta_\tau(t)\,dt = 1.$$

Let

$$\delta(t) = \lim_{\tau \to 0} \delta_\tau(t).$$

Then

$$\int_{-\infty}^{\infty} \delta(t)\,dt = 1 \quad \text{and} \quad \delta(t) = 0 \quad (t \neq 0).$$

c.f. Using the mean value theorem yields

$$\int_{-\infty}^{\infty} \delta(t)v(t)\,dt = \lim_{\tau \to 0} \int_{-\infty}^{\infty} \delta_\tau(t)v(t)\,dt = \lim_{\tau \to 0} \frac{1}{2\tau} \int_{-\tau}^{\tau} v(t)\,dt = \lim_{\tau \to 0} v(t^*) = v(0).$$

Definition (Derivative of a distribution). Let α be a non-negative multi-index and $L \in \mathcal{D}'(\Omega)$.

Then $D^\alpha L$ is the distribution defined as

$$\langle D^\alpha L, v\rangle := (-1)^{|\alpha|}\langle L, D^\alpha v\rangle, \quad \forall v \in \mathcal{D}(\Omega).$$

- A distribution is infinitely differentiable.
- When L is a smooth funtion, the derivatives in the sense of distribution coincides with the usual derivatives.
- Define the Heaviside funtion H as

$$H(x) := \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

  Then $H' = \delta$ in the distribution sense,

$$\langle H', v\rangle = -\int_{\mathbb{R}} H(t)v'(t)\,dt = \int_0^\infty v'(t)\,dt = v(0) = \langle \delta, v\rangle, \quad \forall v \in \mathcal{D}(\Omega).$$

1.4 Sobolev Spaces

Define

$$W^{k,p}(\Omega) := \{v \in L^p(\Omega) : D^\alpha v \in L^p(\Omega),\ |\alpha| \leq k\}$$

and the corresponding norms

$$\|v\|_{k,p,\Omega} = \left(\sum_{|\alpha|\leq k} \|D^\alpha v\|^p_{L^p(\Omega)}\right)^{\frac{1}{p}}, \quad 1 \leq p < \infty,$$

$$\|v\|_{k,\infty,\Omega} = \max_{|\alpha|\leq k} \|D^\alpha v\|^p_{L^\infty(\Omega)}$$

and the corresponding semi-norms

$$|v|_{k,p,\Omega} = \left(\sum_{|\alpha|=k} \|D^\alpha v\|^p_{L^p(\Omega)}\right)^{\frac{1}{p}}, \quad 1 \leq p < \infty,$$

$$|v|_{k,\infty,\Omega} = \max_{|\alpha|=k} \|D^\alpha v\|^p_{L^\infty(\Omega)}.$$

Then, the space $W^{k,p}(\Omega)$ is a Banach space.
When $p = 2$, denote by $H^k(\Omega) := W^{k,2}(\Omega)$.
The space $H^k(\Omega)$ is a Hilbert space with respect to

$$(w, v)_{k,\Omega} := \sum_{|\alpha|\leq k} (D^\alpha w, D^\alpha v)_{0,\Omega}.$$

Denote by

$$
\begin{aligned}
W_0^{k,p}(\Omega) &= \overline{C_0^\infty(\Omega)} \quad \text{w.r.t. the norm } \|\cdot\|_{k,p,\Omega},\\
H_0^k(\Omega) &= W_0^{k,2}(\Omega),\\
W^{-k,p'}(\Omega) &= \text{the dual space of } W_0^{k,p}(\Omega),\\
H^{-k}(\Omega) &= W^{-k,2}(\Omega) = \text{the dual space of } H_0^k(\Omega),
\end{aligned}
$$

where

$$
\|f\|_{W^{-k,p'}(\Omega)} = \sup_{\substack{v \in W_0^{k,p}(\Omega)\\ v\neq 0}} \frac{\langle f, v\rangle}{\|v\|_{W_0^{k,p}(\Omega)}}.
$$

Define

$$
H(\mathrm{div};\Omega) = \{\mathbf{v} \in L^2(\Omega)^d \,:\, \nabla\cdot\mathbf{v} \in L^2(\Omega)\},
$$

equipped with the norm

$$
\|\mathbf{v}\|_{H(\mathrm{div};\Omega)} := \left(\|\mathbf{v}\|_{0,\Omega}^2 + \|\nabla\cdot\mathbf{v}\|_{0,\Omega}^2\right)^{\frac{1}{2}}.
$$

($W^{s,p}(\Gamma) \,:\, \Gamma = \partial\Omega$).

$$
\begin{aligned}
\|v\|_{0,p,\Gamma} &= \left(\int_\Gamma |v(x)|^p\, ds(x)\right)^{\frac{1}{p}} \quad \text{in } W^{0,p}(\Gamma) := L^p(\Gamma),\\
\|f\|_{s-\frac{1}{p},p,\Gamma} &= \inf_{\substack{v\in W^{s,p}(\Omega)\\ v|_\Gamma = f}} \|v\|_{s,p,\Omega} \quad \text{in } W^{s-\frac{1}{p},p}(\Gamma).
\end{aligned}
$$

The space $H^{-\frac{1}{2}}(\Gamma)$ is the dual space of $H^{\frac{1}{2}}(\Gamma) := W^{\frac{1}{2},2}(\Gamma)$ and

$$
\|f\|_{-\frac{1}{2},\Gamma} = \sup_{\substack{v\in H^{\frac{1}{2}}(\Gamma)\\ v\neq 0}} \frac{\langle f, v\rangle}{\|v\|_{\frac{1}{2},\Gamma}}.
$$

Denote by $C^0(\Omega)$ the space of all continuous functions in Ω and

$$
\begin{aligned}
C^m(\Omega) &= \{u \in C^0(\Omega) : \partial^\alpha u \in C^0(\Omega) \quad \forall |a| \le m\},\\
C^m(\bar\Omega) &= \{u \in C^m(\Omega) : \partial^\alpha u \,\text{are bounded and uniformly continuous on}\, \Omega\, \forall |a| \le m\},\\
C^{m,1}(\bar\Omega) &= \{u \in C^m(\bar\Omega) : \partial^\alpha u \,\text{are Lipschitz continuous in}\, \bar\Omega \quad \forall |a| \le m\},
\end{aligned}
$$

equipped with the norms

$$
\|u\|_{C^m(\bar\Omega)} = \max_{0\le|\alpha|\le m} \sup_{x\in\bar\Omega} |\partial^\alpha u(x)|,
$$

$$
\|u\|_{C^{m,1}(\bar\Omega)} = \|u\|_{C^m(\bar\Omega)} + \max_{0\le|\alpha|\le m} \sup_{\substack{x,y\in\Omega\\ x\neq y}} \frac{|\partial^\alpha u(x) - \partial^\alpha u(y)|}{\|x-y\|}.
$$

(Dense of Sobolev spaces).

1. $\mathcal{D}(\Omega) = C_0^\infty(\Omega)$ is dense in $W_0^{k,p}(\Omega)$, $(1 \le p < \infty,\ k \ge 0)$.
2. $\mathcal{D}(\Omega) = C_0^\infty(\Omega)$ is dense in $L^p(\Omega)$, $(1 \le p < \infty)$.
3. $C^\infty(\bar{\Omega})$ is dense in $W^{k,p}(\Omega)$, $(1 \le p < \infty,\ k \ge 0)$ if Ω is Lipschitz domain.
4. $C^k(\bar{\Omega})$ is dense in $W^{k,p}(\Omega)$, $(1 \le p < \infty,\ k \ge 0)$ if Ω is Lipschitz domain.

1.5 Some Results

Definition (Lipschitz continuous). A function f is *Lipschitz continuous* on D if there exists $L > 0$ such that

$$|f(x) - f(y)| \le L\,|x - y| \quad \forall x, y \in D.$$

Definition (Lipschitz Domain). A domain Ω in $\mathbb{R}$ is called a *Lipschitz domain* if there are bounded open sets $G_1, \cdots, G_k$ such that

i) $\partial\Omega \subset \cup_{j=1}^k G_j$,

ii) for every j, $G_j \cap \partial\Omega$ is the graph of a Lipschitz continuous function φ_j satisfying $G_j \cap \mathrm{epi}\varphi_j \subset \Omega$.

For examples, triangles, parallelograms, discs, annuli, parallelepipeds, balls, polygons and polytopes are all Lipschitz domains.

(Trace Theorem). Let Ω be a bounded open set of $\mathbb{R}^d$ with Lipschitz continuous boundary $\partial\Omega$ and let $s > \frac{1}{2}$.

(a) There exists a unique linear continuous map $\gamma_0 : H^s(\Omega) \longrightarrow H^{s-1/2}(\partial\Omega)$ such that $\gamma_0 v = v|_{\partial\Omega}$ for each $v \in H^s(\Omega) \cap C^0(\bar{\Omega})$.

(b) There exists a unique linear continuous map $\mathcal{R}_0 : H^{s-1/2}(\partial\Omega) \longrightarrow H^s(\Omega)$ such that $\gamma_0 \mathcal{R}_0 \varphi = \varphi$ for each $\varphi \in H^{s-1/2}(\partial\Omega)$.

c.f. If Σ is a Lipschitz continuous subset of $\partial\Omega$, then γ_Σ has the analogous results. And

$$\|v\|_{L^p(\partial\Omega)} \le C\,\|v\|_{L^p(\Omega)}^{1-1/p}\,\|v\|_{W^{1,p}(\Omega)}^{1/p}, \quad 1 \le p \le \infty.$$

(Normal Trace Theorem). Let Ω be a bounded open set of $\mathbb{R}^d$ with Lipschitz continuous boundary $\partial\Omega$ and let $s > \frac{1}{2}$.

(a) There exists a unique linear continuous map $\gamma^* : H(\mathrm{div};\Omega) \longrightarrow H^{-1/2}(\partial\Omega)$ such that $\gamma^*\mathbf{v} = (\mathbf{v} \cdot \mathbf{n})|_{\partial\Omega}$ for each $v \in H(\mathrm{div};\Omega) \cap C^0(\bar{\Omega})^d$.

(b) There exists a unique linear continuous map $\mathcal{R}^* : H^{-1/2}(\partial\Omega) \longrightarrow H(\mathrm{div};\Omega)$ such that $\gamma^*\mathcal{R}^*\varphi = \varphi$ for each $\varphi \in H^{-1/2}(\partial\Omega)$.

Note that

$$H_0(\mathrm{div}\,;\Omega) := \overline{C_0^\infty(\Omega)^d}^{H(\mathrm{div};\Omega)}.$$

If $\partial\Omega$ is Lipschitz continuous,

$$\begin{aligned} H_0^1(\Omega) &= \{v \in H^1(\Omega) : \gamma_0 v = 0\}, \\ H_0(\mathrm{div}\,;\Omega) &= \{v \in H(\mathrm{div};\Omega) : \gamma^* v = 0\}, \\ H_\Sigma^1(\Omega) &= \{v \in H^1(\Omega) : \gamma_\Sigma v = 0\}. \end{aligned}$$

(Poincaré Inequlity). Let Ω be a bounded connected open set of $\mathbb{R}^d$ and let Σ be a (non-empty) Lipschitz continuous subset of $\partial\Omega$. Then there exists a constant $C_\Omega > 0$ such that

$$\int_\Omega |v(x)|^2 dx \le C_\Omega \int_\Omega |\nabla v(x)|^2 dx, \quad \forall v \in H_\Sigma^1(\Omega).$$

(Green Formula and Divergence Theorem).

For all $w, v \in H^1(\Omega)$,

$$\int_\Omega \frac{\partial w}{\partial x_j}\, v\, dx = -\int_\Omega w\, \frac{\partial v}{\partial x_j}\, dx + \int_{\partial\Omega} w\, v\, \mathbf{n}_j\, ds.$$

If $\mathbf{w} \in H(\mathrm{div};\Omega)$, $v \in H^1(\Omega)$,

$$\int_\Omega (\mathrm{div}\, \mathbf{w})\, v\, dx = -\int_\Omega \mathbf{w} \cdot \nabla v\, dx + \int_{\partial\Omega} (\mathbf{w} \cdot \mathbf{n}) v\, ds.$$

(Sobolev Embedding Theorem).

Let Ω be an open set of $\mathbb{R}^d$ with Lipschitz continuous boundary $\partial\Omega$, and $s \ge 0$, $1 \le p \le \infty$. Then, the continuous embeddings hold:

(a) If $\quad 0 \le sp < d, \quad W^{s,p}(\Omega) \subset L^{p^*}(\Omega), \quad p^* = \frac{dp}{d-sp}$.

$$\begin{aligned} d = 2, p = 2 \quad &: \quad H^{1/2}(\Omega) \subset L^4(\Omega) \\ d = 3, p = 2 \quad &: \quad H^1(\Omega) \subset L^6(\Omega) \end{aligned}$$

(b) If $\quad sp = d, \quad W^{s,p}(\Omega) \subset L^q(\Omega), \quad p \le q < \infty$.

$$\begin{aligned} d = 1, p = 2 \quad &: \quad H^{1/2}(\Omega) \subset L^q(\Omega), \quad q \ge 2, \\ d = 2, p = 2 \quad &: \quad H^1(\Omega) \subset L^q(\Omega), \quad q \ge 2, \\ d = 3, p = 2 \quad &: \quad H^{3/2}(\Omega) \subset L^q(\Omega), \quad q \ge 2, \end{aligned}$$

(c) If $\quad sp > d, \quad W^{s,p}(\Omega) \subset C^0(\bar{\Omega})$.

$$\begin{aligned} d = 1, p = 2 \quad &: \quad H^s(\Omega) \subset C^0(\bar{\Omega}), \quad s > 1/2, \\ d = 2, p = 2 \quad &: \quad H^s(\Omega) \subset C^0(\bar{\Omega}), \quad s > 1, \\ d = 3, p = 2 \quad &: \quad H^s(\Omega) \subset C^0(\bar{\Omega}), \quad s > 3/2. \end{aligned}$$

Definition (Compact Operator)**.** Let X and Y be Banach spaces.

An operator $T : X \longrightarrow Y$ is *compact* if

i) T is continuous,

ii) for given any bounded sequence x_n in X, there exists a subsequence x_{n_k} such that $T(x_{n_k})$ is convergent in Y.

Note. Let V is compactly embedding to W. Then,

i) if u_n is bounded in V, then there exists a subsequence u_{n_k} which is strongly convergent to a function u in W.
ii) u_n converges to u weakly in $V \Longrightarrow u_n$ is bounded in V and then there exists a subsequence u_{n_k} which converges to u in W.

(Compact Sobolev Embedding Theorem)**.**

Let Ω be an open set of $\mathbb{R}^d$ with Lipschitz continuous boundary $\partial\Omega$, and $s \geq 0$, $1 \leq p \leq \infty$. Then, the following embeddings are compact:

(a) If $0 \leq sp < d, \quad W^{s,p}(\Omega) \subset L^q(\Omega), \quad 1 \leq q < \frac{dp}{d-sp}$.

$$d = 2, p = 2 \quad : \quad H^{1/2}(\Omega) \subset L^q(\Omega), \quad 1 \leq q < 4,$$
$$d = 3, p = 2 \quad : \quad H^{1}(\Omega) \subset L^q(\Omega), \quad 1 \leq q < 6,$$

(b) If $sp = d, \quad W^{s,p}(\Omega) \subset L^q(\Omega), \quad p \leq q < \infty$.

$$d = 1, p = 2 \quad : \quad H^{1/2}(\Omega) \subset L^q(\Omega), \quad 1 \leq q < \infty,$$
$$d = 2, p = 2 \quad : \quad H^{1}(\Omega) \subset L^q(\Omega), \quad 1 \leq q < \infty,$$
$$d = 3, p = 2 \quad : \quad H^{3/2}(\Omega) \subset L^q(\Omega), \quad 1 \leq q < \infty,$$

(c) If $sp > d, \quad W^{s,p}(\Omega) \subset C^0(\bar{\Omega})$.

$$d = 1, p = 2 \quad : \quad H^{s}(\Omega) \subset C^0(\bar{\Omega}), \quad s > 1/2,$$
$$d = 2, p = 2 \quad : \quad H^{s}(\Omega) \subset C^0(\bar{\Omega}), \quad s > 1,$$
$$d = 3, p = 2 \quad : \quad H^{s}(\Omega) \subset C^0(\bar{\Omega}), \quad s > 3/2.$$

(d) If $p > \frac{2d}{d+2}, \quad L^p(\Omega) \subset H^{-1}(\Omega)$.

(e) $H^k(\Omega) \subset H^{k-1}(\Omega)$, $k \geq 0$.

(Gagliardo-Nirenberg Interpolation Inequlity)**.**

$$\max_{a \leq x \leq b} |v(x)| \leq \left(\frac{1}{b-a} + 2 \right)^{\frac{1}{2}} \|v\|^{\frac{1}{2}} \|v\|_1^{\frac{1}{2}}, \quad \forall v \in H^1((a,b)).$$

(Interpolation Theorem).

Let Ω be an open set of $\mathbb{R}^d$ with Lipschitz continuous boundary $\partial\Omega$. Let $s_1 < s_2$ be real numbers and $r = (1-\theta)s_1 + \theta s_2$, $(0 \le \theta \le 1)$. Then, there exists a constant $C > 0$ such that

$$\|v\|_r \le C \, \|v\|_{s_1}^{1-\theta} \, \|v\|_{s_2}^{\theta}, \quad \forall v \in H^{s_2}(\Omega).$$

(Gronwall Lemma for IBVP). Let $f \in L^1(t_0, T)$ be a non-negative function, g and φ be continuous functions on $[t_0, T]$.

If φ satisfies

$$\varphi(t) \le g(t) + \int_{t_0}^{t} f(\tau)\varphi(\tau)\, d\tau, \quad \forall t \in [t_0, T],$$

then

$$\varphi(t) \le g(t) + \int_{t_0}^{t} f(s)g(s) \exp\left(\int_s^t f(\tau)\, d\tau \right) ds, \quad \forall t \in [t_0, T].$$

In addition, if g is non-decreasing, then

$$\varphi(t) \le g(t) \exp\left(\int_s^t f(\tau)\, d\tau \right), \quad \forall t \in [t_0, T].$$

c.f. We often use a form of

$$g(t) = \varphi(0) + \int_0^t \psi(s)\, ds, \quad \psi(s) \ge 0.$$

2 Finite Element Approximation

2.1 Triangulation

Let $\Omega \subset \mathbb{R}^d, d = 2, 3$ be a polygonal domain, i.e., Ω is an open bounded connected subset such that $\bar{\Omega}$ is the union of a finite number of polyhedron.

Conside a finite decomposition

$$\bar{\Omega} = \cup_{K \in \mathcal{T}_h} K,$$

where

1. K is a polyhedron with the interior of K, Int(K), is non-empty;

2. Int$(K_1)\cap$Int$(K_2)= \emptyset$ for each distinct $K_1, K_2 \in \mathcal{T}_h$;

3. if $F = K_1 \cap K_2 \neq \emptyset (K_1 \neq K_2)$, then F is a common face, side, or vertex of K_1, K_2;

4. diam$(K) \leq h$ for each $K \in \mathcal{T}_h$.

Here, $\mathcal{T}_h$ is called a *triangulation* of $\bar{\Omega}$.

From now on, we assume that for each $K \in \mathcal{T}_h$,

$$K = T_K(\hat{K}) \quad \text{with} \quad T_K(\hat{\mathbf{x}}) = B_K \hat{\mathbf{x}} + b_K$$

where $\hat{K}$ is a reference polyhedron and T_K is a suitable invertible affine map with a non-singular matrix B_K.

(Triangle Finite Elements).

The reference polyhedron $\hat{K}$ is the unit d-simplex, i.e., the triangle of vertices (0,0), (1,0), (0,1) when $d = 2$, or the tetrahedron of vertices (0,0,0), (1,0,0), (0,1,0), (0,0,1) when $d = 3$.

As a consequence, each $K = T_K(\hat{K})$ is a triangle or tetrahedron.

(Parallelepipedal Finite Elements).

The reference polyhedron $\hat{K}$ is the unit d-cube $[0, 1]^d$.

As a consequence, each $K = T_K(\hat{K})$ is a parallelogram when $d = 2$ or a parallelepiped when $d = 3$.

If for each $K \in \mathcal{T}_h$ the matrix B_K defining the affine transformation T_K is diagonal, the triangulation is made by d-rectangles (**Rectangular Finite Elements**).

(Quadrilateral Finite Elements).

Dealing with general quadrilaterals or hexahedrons would require admitting that each conponent of the invertible transformation T_K is no longer an affine map but a linear polynomial with respect to each single variable $x_1, \cdots, x_d$. See Ciarlet.

2.2 Piecewise Polynomial Subspaces

Denote by $\mathbb{P}_k$ the space of polynomials of degree less than or equal to k and $\mathbb{Q}_k$ the space of polynomials of degree less than or equal to k with respect to each variable $x_1, \cdots, x_d$.

$$\dim \mathbb{P}_k = \begin{pmatrix} d+k \\ k \end{pmatrix}, \qquad \dim \mathbb{Q}_k = (k+1)^d, \qquad \mathbb{P}_k \subset \mathbb{Q}_k \subset \mathbb{P}_{dk}.$$

Define a space of vector polynomials

$$\mathbb{D}_k = (\mathbb{P}_{k-1})^d \oplus \mathbf{x}\,\mathbb{P}_{k-1}, \quad (k \geq 1),$$

where $\mathbf{x} \in \mathbb{R}^d$ is the independent variable.

$$\dim \mathbb{D}_k = (d+k)\frac{(d+k-2)!}{(d-1)!\,(k-1)!}, \qquad (\mathbb{P}_{k-1})^d \subset \mathbb{D}_k \subset (\mathbb{P}_k)^d.$$

Define the space of *triangular finite elements*:

$$X_h = X_h^k := \{v_h \in C^0(\bar{\Omega}) : v_h|_K \in \mathbb{P}_k,\, \forall K \in \mathcal{T}_h\}, \quad k \geq 1,$$

or the space of *parallelepipedal finite elements*:

$$X_h = X_h^k := \{v_h \in C^0(\bar{\Omega}) : v_h|_K \circ T_K \in \mathbb{Q}_k,\, \forall K \in \mathcal{T}_h\}, \quad k \geq 1.$$

Note that $X_h^k \subset H^1(\Omega)$, $\forall k \geq 1$.

Proposition 2.1. *A function v belongs to $H^1(\Omega)$ if and only if*

(a) $v|_K \in H^1(K)$ for each $K \in \mathcal{T}_h$,

(b) the trace of $v|_{K_1}$ is equal to the trace of $v|_{K_2}$ on F for each common face $F = K_1 \cap K_2$.

Proof. Using (a), define the function $w_j \in L^2(\Omega)$ such that

$$w_j|_K := D_j(v|_K), \quad K \in \mathcal{T}_h,\, j = 1, \cdots, d.$$

By the Green formula, for each $\varphi \in \mathcal{D}(\Omega)$,

$$\int_\Omega w_j\varphi = \sum_K \int_K w_j\varphi = -\sum_K \int_K (v|_K) D_j\varphi + \sum_K \int_{\partial K} v|_K \varphi\, \mathbf{n}_{K,j},$$

where $\mathbf{n}_K$ is the unit normal vector on ∂K. Since φ is vanishing on $\partial\Omega$ and $\mathbf{n}_{K_1} = -\mathbf{n}_{K_2} := \mathbf{n}$ on a common face $F = K_1 \cap K_2$, by (b) we have

$$\int_\Omega w_j\varphi = -\int_\Omega vD_j\varphi + \sum_F \int_F (v|_{K_1} - v|_{K_2})\varphi\, n_j \quad \Big(= -\int_\Omega vD_j\varphi\Big). \tag{2.1}$$

Hence, $w_j = D_j v \in L^2(\Omega)$ and $v \in H^1(\Omega)$.

If $v \in H^1(\Omega)$, then we have

$$\int_K |D_j(v|_K)|^2 \le \int_\Omega |D_j v|^2 < \infty$$

which implies (a). With $w_j = D_j v$, from (2.1)

$$\sum_F \int_F (v|_{K_1} - v|_{K_2})\varphi\, n_j = 0 \quad \forall\, \varphi \in \mathcal{D}(\Omega),\ j = 1, \cdots, d.$$

This completes the proof (b). □

Define the space for vector functions:

$$W_h^k := \{\mathbf{v}_h \in H(\mathrm{div}\,;\Omega) : \mathbf{v}_h|_K \in \mathbb{D}_k \quad \forall\, K \in \mathcal{T}_h\},\ \ k \ge 1.$$

Proposition 2.2. *Let $\mathbf{v} : \Omega \to \mathbb{R}^d$ be such that $\mathbf{v}|_K \in H^1(K)^d$ for each $K \in \mathcal{T}_h$. Then, for $K_1, K_2 \in \mathcal{T}_h$*

$$(a)\ \mathbf{v} \in H(\mathrm{div}\,;\Omega) \quad \textit{if and only if} \quad (b)\ \mathbf{n}\cdot\mathbf{v}|_{K_1} = \mathbf{n}\cdot\mathbf{v}|_{K_2}\ \textit{on}\ F = K_1 \cap K_2.$$

i.e., The traces of the normal components are the same on each common face $F = K_1 \cap K_2$ for $K_1, K_2 \in \mathcal{T}_h$.

Proof. Define $w \in L^2(\Omega)$ such that

$$w|_K := \nabla\cdot(\mathbf{v}|_K) \quad \forall\, K \in \mathcal{T}_h.$$

By the Green formula, if (b) holds, then for each $\varphi \in \mathcal{D}(\Omega)$,

$$\begin{aligned}\langle \nabla\cdot\mathbf{v}, \varphi\rangle &= -\int_\Omega \mathbf{v}\cdot\nabla\varphi = -\sum_K \int_K (\mathbf{v}|_K)\cdot\nabla\varphi \\ &= \sum_K \int_K \nabla\cdot(\mathbf{v}|_K)\varphi - \sum_F \int_F (\mathbf{n}\cdot\mathbf{v}|_{K_1} - \mathbf{n}\cdot\mathbf{v}|_{K_2})\varphi = \int_\Omega w\varphi.\end{aligned} \tag{2.2}$$

Thus, $\nabla\cdot\mathbf{v} = w \in L^2(\Omega)$ and $\mathbf{v} \in H(\mathrm{div};\Omega)$.

If $\mathbf{v} \in H(\mathrm{div};\Omega)$, $w := \nabla\cdot\mathbf{v} \in L^2(\Omega)$. Since $\mathbf{v}|_K \in H^1(K)^d$, by Trace theorem the trace on F is well defined, and using (2.2) we obtain

$$\sum_F \int_F (\mathbf{n}\cdot\mathbf{v}|_{K_1} - \mathbf{n}\cdot\mathbf{v}|_{K_2})\varphi = 0 \quad \forall\, \varphi \in \mathcal{D}(\Omega).$$

Hence, (b) holds. □

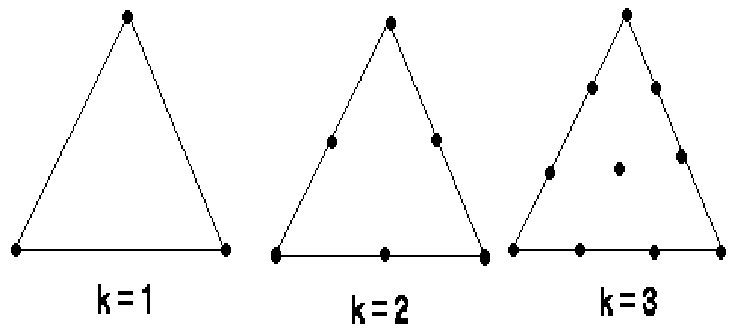

Figure 1: Degrees of freedom for triangular elements in 2D

2.3 Degrees of Freedom and Shape functions

In the constructing a basis for the space X_h^k, an important point is concerned with the choice of a set of *degrees of freedom* on each element K (i.e., the parameters which permit to uniquely idetify a function in $\mathbb{P}_k$, $\mathbb{Q}_k$ or $\mathbb{D}_k$.

2.3.1 The Scalar Case: Triangular finite elements

In two dimensional space, to identify $v_h|_K$ in X_h^k, when $k = 1$ we have to choose three degrees of freedom on each element K, with the additional constraint that $v_h \in C^0(\bar{\Omega})$. The simplest choice is that of the values at the vertices of each K.

Otherwise, if we consider

$$Y_h^1 := \{v_h \in L^2(\Omega) \, : \, v_h|_K \in \mathbb{P}_1, \forall K \in \mathcal{T}_h\},$$

we are free to choose the degrees of freedom on K as the values at three arbitrary points(not necessarily coincident with the vertices).

(Discontinuous FEM). One can take as nodes three internal points, or else the midpoints of each side without continuity at the midpoint.

(Nonconforming FEM). One can take as nodes at the midpoints of each side with continuity at the midpoints.

When $k = 2$, we assume that the element degrees of freedom in X_h^k are given by the value at the vertices and in the middle point of each side.

Denote the vertices of the triangle K by a^i, $i = 1, 2, 3$, and the midpoints by a^{ij}, $i < j$, $i, j = 1, 2, 3$.

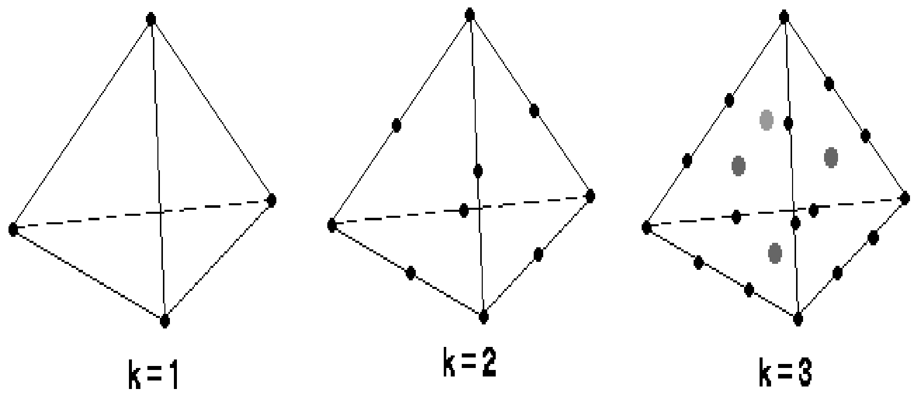

Figure 2: Degrees of freedom for triangular elements in 3D

Proposition 2.3. *A function $p \in \mathbb{P}_2$ is uniquely determined by the six values $p(a^i)$, $1 \le i \le 3$, and $p(a^{ij})$, $1 \le i < j \le 3$.*

Proof. Since the number of the degrees of freedom is equal to the dimension of $\mathbb{P}_2 (= 6)$, we have only to prove that $p(a^i) = p(a^{ij}) = 0$ then $p \equiv 0$.

Note that the restriction of p over each side is a quadratic function of one variable vanishing in three distinct points, hence p is vanishing over each side. Thus we can write

$$p(\mathbf{x}) = cp_1(\mathbf{x})p_2(\mathbf{x})p_3(\mathbf{x}),$$

where $p_i(\mathbf{x})$ are linear functions, each one vanishing on one side of K. Since $p \in \mathbb{P}_2$, it follows $c = 0$. □

This choice of degrees of freedom guarantees that $v_h \in C^0(\bar{\Omega})$, since the degrees of freedom on each side uniquely identify the restriction of v_h on that side.

(Cubic elements $k = 3$)**.** In similar way, one can prove that the degrees of freedom for a cubic triangle elements are given by ten values at the following nodes:

a. the three vertices

b. two other nodes on each side, dividing it into three subintervals of equal length

c. the center of gravity

When $d = 3$, it is not difficult to see that the degrees of freedom are the values at the nodes indicated in Figure 3.

A basis for X_h^k is now easily constructed. For the global set of nodes $\{a_j\}_{j=1}^{N_h}$ in $\bar{\Omega}$, if $\phi_i \in X_h^k$ satisfies

$$\phi_i(a_j) = \delta_{ij}, \quad \forall\, i, j = 1, \cdots, N_h,$$

then ϕ_i is called a *shape function* or *nodal basis function* corresponding to the node a_i.

Let $\hat{K}$ be the reference triangle with three vertices $a_1 = (1, 0)$, $a_2 = (0, 1)$ and $a_3 = (0, 0)$. Denote by a_4, a_5 and a_6 the midpoints of a_1 through a_3.
In P_1 elements, i.e., continuous piecewise linear functions, the *baricentric coordinate* corresponding to the reference triangle $\hat{K}$ is given by

$$\lambda_1 = x, \quad \lambda_2 = y, \quad \lambda_3 = 1 - x - y, \quad \text{and } \sum_{i=1}^{3} \lambda_i = 1.$$

In P_2 elements, i.e., continuous piecewise quadratic functions, the *baricentric coordinate* corresponding to the reference triangle $\hat{K}$ is given by

$$\begin{array}{lll} \phi_1 = \lambda_1(2\lambda_1 - 1), & \phi_2 = \lambda_2(2\lambda_2 - 1), & \phi_3 = \lambda_3(2\lambda_3 - 1), \\ \phi_4 = 4\lambda_1\lambda_2, & \phi_5 = 4\lambda_2\lambda_3, & \phi_6 = 4\lambda_3\lambda_1, \end{array} \quad \text{and } \sum_{i=1}^{6} \phi_i = 1.$$

2.3.2 The Scalar Case: parallelepipedal finite elements

The reference square $\hat{K} = [0, 1]^d$.

Let us prove that a function in $\mathbb{Q}_k$ is uniquely determined by its values at the nodes given in Figure 3.

Proposition 2.4. *If $q \in \mathbb{Q}_k$ ($k = 1, 2, 3$) vanishes at the nodes, then $q \equiv 0$.*

Proof. For the case of $k = 1$, the restriction of q to each side is a linear polynomial of one variable. Hence q vanishes over each side and therefore it can be written as

$$q(\mathbf{x}) = c_1 x_1 (1 - x_1) x_2 (1 - x_2),$$

which implies $c_1 = 0$.

A similar argument applied to the cases $k = 2$ and $k = 3$ implies that q has the form

$$q(\mathbf{x}) = c_2 x_1 (\frac{1}{2} - x_1)(1 - x_1) x_2 (1 - x_2), \quad k = 2,$$

or

$$q(\mathbf{x}) = c_3 x_1 (\frac{1}{3} - x_1)(\frac{2}{3} - x_1)(1 - x_1) x_2 (1 - x_2), \quad k = 3.$$

Since $x_1^3 x_2^2 \notin \mathbb{Q}_2$ and $x_1^4 x_2^2 \notin \mathbb{Q}_3$, it follows that $c_2 = c_3 = 0$. $\square$

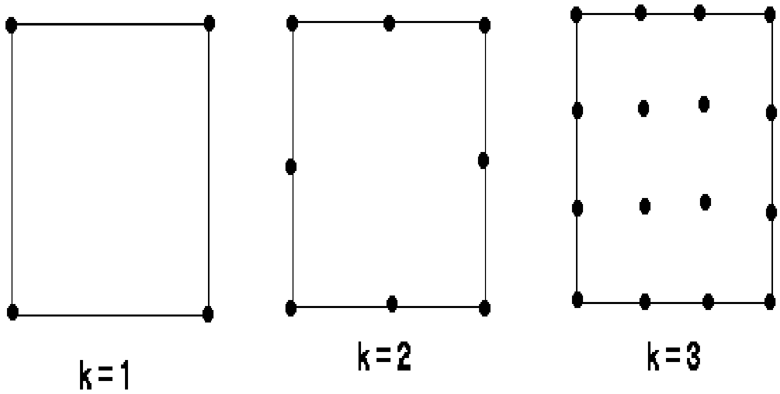

Figure 3: Degrees of freedom for parallelepipedal elements in 2D

2.3.3 The vector case

Let $d = 2$. Recall the space of vector functions:

$$W_h^k := \{\mathbf{v}_h \in H(\mathrm{div}\,;\Omega) : \mathbf{v}_h|_K \in \mathbb{D}_k \quad \forall\, K \in \mathcal{T}_h\},\ k \geq 1.$$

On each, the dimension of $\mathbb{D}_k$ is $(k+2)k$, e.g., $\begin{pmatrix} k & : & 1 & 2 & 3 \\ \mathrm{dim} & : & 3 & 8 & 15 \end{pmatrix}$.

Also, it must hold that $\mathbf{v}_h \in H(\mathrm{div}\,;\Omega)$. Hence it is necessary and sufficient that

$$\mathbf{n} \cdot \mathbf{v}_h|_{K_1} = \mathbf{n} \cdot \mathbf{v}_h|_{K_2} \quad \text{on } F = K_1 \cap K_2.$$

(Ex 1). Prove that $(\mathbf{n} \cdot \mathbf{q})|_F \in \mathbb{P}_{k-1}$ and $\nabla \cdot \mathbf{q} \in \mathbb{P}_{k-1}$ for each $\mathbf{q} \in \mathbb{D}_k$.

The (Ex 1) suggests that k degrees of freedom can be given by the values of $\mathbf{n} \cdot \mathbf{q}$ at k distinct points of each side. This is sufficient for the case $k = 1$.

Proposition 2.5. *Let $k = 1, 2, 3$. Assume that $\mathbf{q} \in \mathbb{D}_k$ is such that $\mathbf{n} \cdot \mathbf{q}$ vanishes at k distinct points on each side of K. Assume moreover that*

$$\int_K q_1 = \int_K q_2 = 0 \quad (\textit{if } k \geq 2) \tag{2.3}$$

and

$$\int_K x_1 q_1 = \int_K x_2 q_1 = \int_K x_1 q_2 = \int_K x_2 q_2 = 0 \quad (\textit{only if } k = 3). \tag{2.4}$$

Then $\mathbf{q} \equiv 0$.

Proof. Since $(\mathbf{n}\cdot\mathbf{q})|_F \in \mathbb{P}_{k-1}$, it vanishes on each side F of K. By the Green formula and (2.3), (2.4), for each $\Psi \in \mathbb{P}_{k-1}$ we have

$$\int_K \Psi \nabla\cdot\mathbf{q} = -\int_K \nabla\Psi\cdot\mathbf{q} + \int_{\partial K} \Psi\mathbf{n}\cdot\mathbf{q} = 0,$$

since $\nabla\Psi \in (\mathbb{P}_{k-2})^2$ and $k \le 3$. As $\nabla\cdot\mathbf{q} \in \mathbb{P}_{k-1}$, it follows that $\nabla\cdot\mathbf{q} = 0$ in K by substituting $\Psi = \nabla\cdot\mathbf{q}$.

For $\mathbf{q} \in \mathbb{D}_k$, it can be written by $\mathbf{q} = \mathbf{p}_{k-1} + \mathbf{x}p^*_{k-1}$, where $\mathbf{p}_{k-1} \in \mathbb{P}^2_{k-1}$ and p^*_{k-1} is a homogeneous function of degree $k-1$ (function consisting of only terms of the highest degree $k-1$). Then, we have

$$\begin{aligned} 0 = \nabla\cdot\mathbf{q} &= \nabla\cdot\mathbf{p}_{k-1} + 2p^*_{k-1} + \mathbf{x}\cdot\nabla p^*_{k-1} \\ &= \nabla\cdot\mathbf{p}_{k-1} + (2+k-1)p^*_{k-1}. \end{aligned}$$

Thus, $p^*_{k-1} \in \mathbb{P}_{k-2}$ and so $p^*_{k-1} = 0$, and consequently $\mathbf{q} \in \mathbb{P}^2_{k-1}$.

Since $\nabla\cdot\mathbf{q} = 0$, we can find a polynomial $w \in \mathbb{P}_k$ (unique up to an additive constant) such that (see Helmholz decomposition below)

$$\mathbf{q} = (D_2 w,\ -D_1 w).$$

Moreover, since $(\mathbf{n}\cdot\mathbf{q})|_F = 0$, we can assume that w is vanishing on each side F, and consequently

$$w(\mathbf{x}) = c_0 p_1(\mathbf{x}) p_2(\mathbf{x}) p_3(\mathbf{x}),$$

where $p_i(\mathbf{x})$ are linear functions, each one vanishing on one side of K. If $k = 1$ or 2, then by $w \in \mathbb{P}_k$ we have $c_0 = 0$ and this completes the proof for $k = 1$ or 2. When $k = 3$, using (2.3) and (2.4) we obtain for each $\mathbf{r} \in \mathbb{P}^2_1$

$$0 = \int_K \mathbf{q}\cdot\mathbf{r} = \int_K [(D_2 w) r_1 - (D_1 w) r_2] = \int_K w(D_2 r_1 - D_1 r_2).$$

Choosing $\mathbf{r}$ such that $D_2 r_1 - D_1 r_2 = c_0$, it follows

$$c_0^2 \int_K p_1 p_2 p_3 = 0.$$

Thus, $c_0 = 0$ and $\mathbf{q} \equiv 0$. □

(Helmholz decomposition). ⌈ For a give $\mathbf{q} \in L^2(\Omega)^2$, it can be decomposed as

$$\mathbf{q} = \nabla u + \nabla\times w \in H^\perp + H \quad \text{where} \quad \nabla\times w = (D_2 w,\ -D_1 w)^t$$

where

$$H = \{\mathbf{v} \in H(\mathrm{div};\Omega) \,:\, \nabla\cdot\mathbf{v} = 0,\ (\mathbf{n}\cdot\mathbf{v})|_\Gamma = 0\}$$

and $u \in H^1(\Omega)/\mathbb{R}$ is the only solution of the Neumann's problem

$$(\nabla u, \nabla \mu) = (\mathbf{q}, \nabla \mu) \quad \forall \mu \in H^1(\Omega),$$

and $w \in H_0^1(\Omega)$ is the only solution of

$$(\nabla \times w, \nabla \times \chi) = (\mathbf{q} - \nabla u, \nabla \times \chi) \quad \forall \chi \in H_0^1(\Omega). \rfloor$$

Note that

$$(\nabla \times w, \mathbf{f})_K = (w, \nabla^{\perp} \mathbf{r})_K \qquad \text{if } w = 0 \text{ on } \partial K$$

where $\nabla^{\perp}\mathbf{r} = d_2 r_1 - D_1 r_2$ denotes the formal adjoint of $\nabla\times$.
The construction of a basis of W_h^k is some how less evident than for X_h^k.
Let $\{a_j\}$ be the set of all nodes $\bar{\Omega}$. Let us denote by

$$\begin{cases} m_j(\mathbf{v}),\ j = 1, \cdots, N_{1,h} & : \text{the values } (\mathbf{n} \cdot \mathbf{v})(a_j) \\ m_\ell(\mathbf{v}),\ \ell = N_{1,h}+1, \cdots, N_h & : \text{the set of all } K\text{-moments of the function } \mathbf{v}. \end{cases}$$

Now, a basis of W_h^k is constructed by requiring that

$$m_s(\boldsymbol{\phi}) = \delta_{is}, \quad i, s = 1, 2, \cdots, N_h.$$

2.4 The Interpolation Operator

Denote by

a_i : the global nodes on $\bar{\Omega}$,

$a_{i,K}$: the local nodes in K,

ϕ_i : the corresponding shape function to a_i in X_h^k.

Define a *local interpolation operator* π_K^k:

$$\pi_K^k(v) := \sum v(a_{i,K}) \phi_i|_K \quad \forall v \in C^0(K),$$

and define an *interpolation operator* $\pi_h^k : C^0(\bar{\Omega}) \to X_h^k$ as

$$\pi_h^k(v)|_K = \pi_K^k(v|_K) \quad \forall K \in \mathcal{T}_h,\ v \in C^0(\bar{\Omega}). \quad \text{i.e.,} \quad \pi_h^k(v) := \sum_{i=1}^{N_h} v(a_i)\phi_i.$$

Denote $\hat{v} = v \circ T_K$ for any $v \in H^m(K)$, where $T_K(\hat{\mathbf{x}}) = B_K(\hat{\mathbf{x}}) + b_K$ for each $\hat{\mathbf{x}} \in \hat{K}$.

Proposition 2.6. *For any $v \in H^m(K)$, $m \geq 0$, we have $\hat{v} = v \circ T_K \in H^m(\hat{K})$, and there exists a constant $C = C(m,d)$ such that*

(a) $|\hat{v}|_{m,\hat{K}} \leq C \|B_K\|^m \,|\det B_K|^{-\frac{1}{2}}\, |v|_{m,K} \quad \forall v \in H^m(K),$

(b) $|v|_{m,K} \le C\|B_K^{-1}\|^m \,|\det B_K|^{\frac{1}{2}}\, |\hat{v}|_{m,\hat{K}} \quad \forall\, \hat{v} \in H^m(\hat{K}),$

where $\|\cdot\|$ *is the matrix norm associated to the euclidean norm in* $\mathbb{R}^d$*.*

Proof. Claim (1)holds for smooth v.

Using the chain rule, with $|\alpha| = m$

$$\begin{aligned}
\|D^\alpha \hat{v}\|^2_{0,\hat{K}} &= \int_{\hat{K}} |D^\alpha \hat{v}|^2 d\hat{\mathrm{x}} \le C\|B_K\|^{2m} \int_{\hat{K}} |(D^\alpha v)\circ T_K|^2 d\hat{\mathrm{x}} \\
&= C\|B_K\|^{2m} \int_{\hat{K}} |D^\alpha v|^2 |detB_K|^{-1} d\mathrm{x} \\
&= C\|B_K\|^{2m} |detB_K|^{-1} \|D^\alpha v\|^2_{0,K} d\mathrm{x}
\end{aligned}$$

Summation gives the conclusion (1).

(2) is similary followed by using $|\hat{v}|^2_{m,\hat{K}} = \Sigma_{|\alpha|=m} \|D^\alpha \hat{v}\|^2_{0,\hat{K}}$. □

Define

$$h_K = \mathrm{diam}(K), \quad \rho_K = \sup\{\mathrm{diam}(S) \,:\, S \text{ is a ball contained in } K\}.$$

The same quantities will be denoted by $\hat{h}$ and $\hat{\rho}$ when they are refered to the reference domain $\hat{K}$.

Proposition 2.7. *The following estimates hold*

$$\|B_K\| \le \frac{h_K}{\hat{\rho}} \quad \textit{and} \quad \|B_K^{-1}\| \le \frac{\hat{h}}{\rho_K}.$$

Proof. We can write

$$\|B_K\| = sup_{|\xi|=1} |B_K \xi| = \frac{1}{\hat{\rho}} sup_{|\xi|=\hat{\rho}} |B_K \xi|.$$

For each ξ satisfying $|\xi| = \hat{\rho}$, we find two points $\hat{\mathrm{x}}, \hat{\mathrm{y}} \in \hat{K}$ such that $\hat{\mathrm{x}} - \hat{\mathrm{y}} = \xi$.

Since $B_K \xi = T_K \hat{\mathrm{x}} - T_K \hat{\mathrm{y}}$, we deduce $|B_K \xi| \le h_K$.

Hence $\|B_K\| \le \frac{h_K}{\hat{\rho}}$.

Similary, $\|B_K^{-1}\| \le \frac{\hat{h}}{\rho_K}$. □

Denote by

$$[\pi^k_K(v)]^\wedge = \pi^k_K(v) \circ T_K.$$

Using $\hat{\phi}_i = \phi_i \circ T_K$,

$$[\pi^k_K(v)]^\wedge = \sum_{i=1}^{M_K} v(a_{i,K})(\phi_i \circ T_K) = \sum_{i=1}^{M_K} v(T_K(\hat{a}_i))\hat{\phi}_i = \pi^k_{\hat{K}}(\hat{v}).$$

Hence, in order to estimate for the seminorm $[v - \pi^k_K(v)] \circ T_K$ in $H^m(\hat{K})$, we have to estimate $\hat{v} - \pi^k_{\hat{K}}(\hat{v})$ in $H^m(\hat{K})$.

Proposition 2.8 (Bramble-Hilbert Lemma).

Let $\hat{\ell} : H^{k+1}(\hat{K}) \to H^m(\hat{K})$, $m \geq 0$, $k \geq 0$ *be a linear continuous mapping such that*

$$\hat{\ell}(\hat{p}) = 0 \quad \forall \hat{p} \in \mathbb{P}_k.$$

Then, for each $\hat{v} \in H^{k+1}(\hat{K})$

$$|\hat{\ell}(\hat{v})|_{m,\hat{K}} \leq \|\hat{\ell}\|_{\mathcal{L}(H^{k+1}(\hat{K});H^m(\hat{K}))} \inf_{\hat{p}\in\mathbb{P}_k} \|\hat{v} + \hat{p}\|_{k+1,\hat{K}}.$$

Proof. Let $\hat{v} \in H^{k+1}(\hat{K})$. For any $\hat{p} \in \mathbb{P}_k$ we have $\hat{\ell}(\hat{p}) = 0$.

$$|\hat{\ell}(\hat{v})|_m, \hat{K} = |\hat{\ell}(\hat{v} + \hat{p})|_m, \hat{K} \leq \|\hat{\ell}\|_{\mathcal{L}(H^{k+1}(\hat{K});H^m(\hat{K}))}\|\hat{v} + \hat{p}\|_{k+1,\hat{K}} \quad \forall \hat{p} \in \mathbb{P}_k.$$

□

Proposition 2.9 (Deny-Lions Lemma).

For each $k \geq 0$, *there exists a constant* $C = C(k, \hat{K})$ *such that*

$$\inf_{\hat{p}\in\mathbb{P}_k} \|\hat{v} + \hat{p}\|_{k+1,\hat{K}} \leq C|\hat{v}|_{k+1,\hat{K}} \quad \forall \hat{v} \in H^{k+1}(\hat{K}).$$

Proof. (Compactness argument)

Step1) Let us prove that there exists a constant $C = C(\hat{K})$ such that

(1) $\|\hat{v}\|_{k+1,\hat{K}} \leq C\{|\hat{v}|^2_{k+1,\hat{K}} + \sum_{|\alpha|\leq k}(\int_{\hat{K}} D^\alpha \hat{v})^2\}^{\frac{1}{2}}$

for each $\hat{v} \in H^{k+1}(\hat{K})$. We proceed by contradiction. If (1) doesn't hold, thenwe could find a sequence $\hat{v}_s \in H^{k+1}(\hat{K})$ such that

(2) $\|\hat{v}_s\|_{k+1,\hat{K}} = 1$

and

(3) $|\hat{v}_s|^2_{k+1,\hat{K}} + \sum_{|\alpha|\leq k}(\int_{\hat{K}} D^\alpha \hat{v}_s)^2 < \frac{1}{s^2}$.

Since the immersion $H^{k+1}(\hat{K}) \hookrightarrow H^k(\hat{K})$ is compact, we can select a subsequence, still denoted by $\hat{v}_s$, strongly convergent in $H^k(\hat{K})$.

As a consequence of (3) $\hat{v}_s$ is indeed a Cauchy sequence in $H^{k+1}(\hat{K})$, therefore a function $\hat{\omega}$ exists such that $\hat{v}_s$ converge to $\hat{\omega}$ in $H^{k+1}(\hat{K})$ and $\|\hat{\omega}\|_{k+1,\hat{K}} = 1$.

Moreover, by (3) $\int_{\hat{K}} D^\alpha \hat{\omega} = 0$ for $|\alpha| \leq k$ and $D^\alpha \hat{\omega} = 0$ for $|\alpha| = k + 1$. This last relation implies that $\hat{\omega} \in \mathbb{P}_k$ and then $\hat{\omega} = 0$. this is contradiction to $\|\hat{\omega}\|_{k+1,\hat{K}} = 1$.

Step2) For each $\hat{v} \in H^{k+1}(\hat{K})$ we can construct a unique $\hat{q} \in \mathbb{P}_k$ such that

$$\int_{\hat{K}} D^\alpha \hat{q} = -\int_{\hat{K}} D^\alpha \hat{v}, \quad \forall |\alpha| \leq k.$$

Hence from (1) applied to $\hat{v}+\hat{q}$ that $D^\alpha(hatv+hatq)=0$, we obtain

$$\inf_{\hat{p}\in\mathbb{P}_k}\|\hat{v}+\hat{p}\|_{k+1,\hat{K}}\le\|\hat{v}+\hat{q}\|_{k+1,\hat{K}}\le C|\hat{v}+\hat{q}|_{k+1,\hat{K}}=C|\hat{v}|_{k+1,\hat{K}}.$$

□

Theorem 2.10 (Local interpolation error)**.** *If $0\le m\le l+1$, $1\le l\le k$, then there exists a constant $C=C(\hat{K},\pi^k_{\hat{K}},l,m,d)$ such that*

$$|v-\pi^k_K(v)|_{m,K}\le C\,\frac{h_K^{l+1}}{\rho_K^m}\,|v|_{l+1,K}\quad\forall\,v\in H^{l+1}(K).$$

Note that high order interpolation on v do not give, in principle, better error estimates if v is not regular enough.

Proof. First of all, let us remark that the Sovolev embedding theorem yields $H^{l+1}(K)\subset C^0(K)$ for $k\ge1$. Hence the interpolation operator π^k_K is well-defined in $H^{l+1}(K)$. By previous proposition,

$$\begin{aligned}|v-\pi^k_K(v)|_{m,K}&\le C\,\|B_K^{-1}\|^m\,|detB_K|^{\frac12}\,|\hat{v}-\pi^k_{\hat{K}}(\hat{v})|_{m,\hat{K}}\\&\le C\,\frac{\hat{h}^m}{{\rho_K}^m}\,|detB_K|^{\frac12}\,|\hat{\ell}(\hat{v})|_{m,\hat{K}}\quad\hat{\ell}=I-\pi^k_K.\end{aligned}$$

Since

$$\begin{aligned}|\hat{\ell}(\hat{v})|_{m,\hat{K}}&\le\|\hat{\ell}\|\inf_{\hat{p}\in\mathbb{P}_k}\|\hat{v}+\hat{p}\|_{k+1,\hat{K}}\le C\,|\hat{v}|_{k+1,\hat{K}}\\&\le C\,\|B_K\|^m\,|detB_K|^{-\frac12}\,|v|_{k+1,K}\\&\le C\,\frac{h_K^{k+1}}{\hat{\rho}^{k+1}}\,|detB_K|^{-\frac12}\,|v|_{k+1,K}.\end{aligned}$$

Hence $|v-\pi^k_K(v)|_{m,K}\le C\,\frac{h_K^{k+1}}{\rho_K^m}\,|v|_{k+1,K}$. □

Remark (L^∞-interpolation error)**.**

(a) The similar results hold for interpolation in the Sobolev space $W^{k+1,p}(\Omega)$ $p\in[1,\infty]$. (see Ciarlet)

(b) For $1\le\ell\le k$, $0\le m\le\ell+1-\frac{d}{2}$, $d=2,3$,

$$|v-\pi^k_K(v)|_{m,\infty,K}\le C\,[\text{meas}(K)]^{-\frac12}\,\frac{h_K^{\ell+1}}{\rho_K^m}\,|v|_{\ell+1,K}\quad\forall\,v\in H^{\ell+1}(K).$$

(To prove) Using Theorem 2.10, $H^{l+1}(\hat{K})\subset W^{m,\infty}(\hat{K})$, $0\le m<l+1-d/2$ and that

$$\|D^\alpha v\|_{\infty,K}\le C\|B_K^{-1}\|^m\,\|D^\alpha\hat{v}\|_{\infty,\hat{K}},\quad|\alpha|=m,$$

$$|v-\pi^k_K(v)|_{m,\infty,K}\le C\|B_K^{-1}\|^m|\hat{v}-\pi^k_{\hat{K}}(\hat{v})|_{m,\infty,\hat{K}}\le\frac{C}{\rho_K^m}\,|\hat{v}|_{l+1,\hat{K}}\le C\text{meas}(K)^{-\frac12}\,\frac{h_K^{l+1}}{\rho_K^m}\,|v|_{l+1,K}.$$

(c) For $1 \le \ell \le k$, $0 \le m \le \ell+1$,

$$|v - \pi_K^k(v)|_{m,\infty,K} \le C\, \frac{h_K^{\ell+1}}{\rho_K^m}\, |v|_{\ell+1,\infty,K} \quad \forall\, v \in W^{\ell+1,\infty}(K).$$

(To prove) Use Bramble-Hilbert and Deny-Lions lemmas and replace $H^{k+1}(\hat{K})$ and $H^m(\hat{K})$ by $W^{k+1,\infty}(\hat{K})$ and$W^{m,\infty}(\hat{K})$

Definition (Regular triangulation)**.**

A family of triangulation $\mathcal{T}_h$ $(h > 0)$ is called *regular* if there exists $\sigma \ge 1$ such that

$$\max_{K \in \mathcal{T}_h} \frac{h_K}{\rho_K} \le \sigma \quad \forall\, h > 0.$$

Theorem 2.11 (Interpolation error)**.**

Let $\mathcal{T}_h$ be a regular family of triangulations and assume that $m = 0, 1$, $k \ge 1$. Then there exists C, independent of h, such that

$$|v - \pi_h^k(v)|_{m,\Omega} \le C\, h^{\ell+1-m}\, |v|_{\ell+1,\Omega} \quad \forall\, v \in H^{\ell+1}(\Omega), \quad 1 \le \ell \le k.$$

Proof.

$$\begin{aligned} |v - \pi_h^k(v)|_{m,\Omega}^2 &= \sum_K |v - \pi_h^k(v)|_{m,K}^2 \\ &\le C \sum_K \Big(\frac{h_K^{l+1}}{\rho_K^m}\Big)^2 |v|_{l+1,K}^2 \\ &\le C \sum_K (h_K^{l+1-m})^2 |v|_{l+1,K}^2, \quad h_k \le \rho_K\, \sigma \\ &\le C\, h_K^{l+1-m}\, |v|_{l+1,\Omega}^2. \end{aligned}$$

□

Note that the restriction on the index m is due to the fact that the inclusion $X_h^k \subset H^m(\Omega)$ holds only if $m \le 1$.

The construction of a finite dimensional space contained in $H^2(\Omega)$ would require higher order continuity across the interelement boundaries.

2.4.1 Interpolation Error: the vector case

Recall the space of vector functions:

$$W_h^k := \{\mathbf{v}_h \in H(\mathrm{div}\,; \Omega) : \mathbf{v}_h|_K \in \mathbb{D}_k \quad \forall\, K \in \mathcal{T}_h\}, \ k \ge 1.$$

To define the interpolation operator we must give a meaning to the point value $\mathbf{n} \cdot \mathbf{v}$ at all nodes $a_i \in \bar{\Omega}$ and to all the K-moments $m_l(\mathbf{v})$.

If $\mathbf{v} \in C^0(\bar{\Omega})^d$, this is easily doable.

But it will be useful to define the interpolation operator even in spaces of functions that are not necessarily continuous.

Instead of the point values of $\mathbf{n} \cdot \mathbf{v}$ on a face F_K of K, consider the following degrees of freedom

$$\int_{F_K} \mathbf{n} \cdot \mathbf{q}\, \psi, \quad \psi \in \mathbb{P}_{k-1},$$

which are called F_K-moments.

Denote the global set of these F_K-moments relative to a function $\mathbf{v}$ by $m_l(\mathbf{v})$, $l = 1, \cdots, N_{1,h}$ and denote the set of K-moments by $m_l(\mathbf{v})$, $l = N_{1,h}+1, \cdots, N_h$. Let $\boldsymbol{\varphi}_i$ be the shape functions such that

$$m_s(\boldsymbol{\varphi}_i) = \delta_{is}, \quad i, s = 1, \cdots, N_h.$$

Define the interpolation operator $\boldsymbol{\omega}_h^k : H^1(\Omega)^d \to W_h^k$ by

$$\boldsymbol{\omega}_h^k(\mathbf{v}) := \sum_{i=1}^{N_h} m_i(\mathbf{v}) \boldsymbol{\varphi}_i.$$

Then, $\boldsymbol{\omega}_h^k(\mathbf{v})$ is the only one function in W_h^k satisfying

$$m_i(\boldsymbol{\omega}_h^k(\mathbf{v})) = m_i(\mathbf{v}), \quad i = 1, \cdots, N_h.$$

Denote by $m_{i,K}(\mathbf{v})$, $i = 1, \cdots, M_K$ the set of K-moments and F_K-moments relative to K. Define a local interpolation operator

$$\boldsymbol{\omega}_K^k(\mathbf{v}) := \sum_{i=1}^{M_K} m_{i,K}(\mathbf{v}) \boldsymbol{\varphi}_i|_K, \quad \mathbf{v} \in H^1(\Omega)^d.$$

We have

$$m_{i,K}(\mathbf{v}) = m_{i,K}(\boldsymbol{\omega}_K^k(\mathbf{v})), \quad i = 1, \cdots, M_K$$

and

$$\boldsymbol{\omega}_h^k(\mathbf{v})|_K = \boldsymbol{\omega}_K^k(\mathbf{v}|_K), \quad \forall\, K \in \mathcal{T}_h, \forall\, \mathbf{v} \in H^1(\Omega)^d.$$

Let P_K^{k-1} be the orthogonal projection in $L^2(K)$ onto $\mathbb{P}_{k-1}$.

Then, we have an important property:

$$\operatorname{div}(\boldsymbol{\omega}_K^k(\mathbf{v})) = P_K^{k-1}(\operatorname{div} \mathbf{v}), \quad \forall\, \mathbf{v} \in H^1(K)^d.$$

(Proof) Since $\operatorname{div}(\boldsymbol{\omega}_K^k(\mathbf{v})) \in \mathbb{P}_{k-1}$, for each $\psi \in \mathbb{P}_{k-1}$,

$$\begin{aligned}
\int_K \psi \operatorname{div}(\boldsymbol{\omega}_K^k(\mathbf{v})) &= -\int_K \nabla\psi \cdot \boldsymbol{\omega}_K^k(\mathbf{v}) + \int_{\partial K} \psi\, \mathbf{n} \cdot \boldsymbol{\omega}_K^k(\mathbf{v}) \\
&= -\int_K \nabla\psi \cdot \mathbf{v} + \int_{\partial K} \psi\, \mathbf{n} \cdot \mathbf{v} = \int_K \psi \cdot \operatorname{div} \mathbf{v},
\end{aligned}$$

owing to the fact that the moments of $\mathbf{v}$ and $\boldsymbol{\omega}_K^k(\mathbf{v})$ are the same, i.e,

$$m_{i,K}(\mathbf{v}) = m_{i,K}(\boldsymbol{\omega}_K^k(\mathbf{v})), \quad i = 1, \cdots, M_K. \qquad \square$$

Similarly we have

$$\operatorname{div}(\boldsymbol{\omega}_h^k(\mathbf{v})) = p_h^{k-1}(\operatorname{div}\mathbf{v}), \quad \forall\, \mathbf{v} \in H^1(\Omega)^d,$$

where p_h^{k-1} is the $L^2(\Omega)$-orthogonal projection onto

$$Y_h^{k-1} := \{v \in L^2(\Omega) \,:\, v|_K \in \mathbb{P}_{k-1} \quad \forall\, K \in \mathcal{T}_h\}.$$

This commutativity property is very important for the approximation theory in $H(\operatorname{div};\Omega)$, and it turns out to be useful also when considering optimal error estimates for boundary value problems. The introduction of the polynomial spaces $\mathbb{D}_k$ is in fact motivated by such a matter.

$$\begin{array}{ccc} H^1(\Omega)^d & \xrightarrow{\operatorname{div}} & L^2(\Omega) \\ \boldsymbol{\omega}_h^k \downarrow & & \downarrow p_h^{k-1} \\ W_h^k & \xrightarrow{\operatorname{div}} & Y_h^{k-1} \end{array}$$

Define

$$\hat{\mathbf{v}} := |\det B_K|\, B_K^{-1}\, \mathbf{v} \circ T_K.$$

Note that

$$\hat{\mathbf{v}} \in H^1(\hat{K})^d \quad \text{if and only if} \quad \mathbf{v} \in H^1(K)^d,$$

$$\hat{\mathbf{v}} \in \mathbb{D}_k \quad \text{if and only if} \quad \mathbf{v} \in \mathbb{D}_k.$$

For $\mathbf{v} \in H^1(K)^d$, $\psi \in H^1(K)$,

$$\begin{aligned} \int_{\hat{K}} \hat{\mathbf{v}} \cdot \nabla\hat{\psi} d\hat{x} &= \int_{\hat{K}} |\det B_K| (B_K^{-1}\mathbf{v} \circ T_K)(B_K^t \nabla\psi \circ T_K) d\hat{x} \\ &= \int_K (B_K^{-1}\mathbf{v})(B_K^t \nabla\psi) dx = \int_K \mathbf{v} \cdot \nabla\psi, \end{aligned}$$

where $\hat{\psi} = \psi \circ T_K$, and analogously

$$\int_{\hat{K}} \hat{\psi} \operatorname{div}\hat{\mathbf{v}} = \int_K \psi \operatorname{div}\mathbf{v} \quad \text{and then} \quad \int_{\partial\hat{K}} \hat{\psi}\, \hat{\mathbf{v}} \cdot \hat{\mathbf{n}} = \int_{\partial K} \psi\, \mathbf{v} \cdot \mathbf{n}.$$

For each $\mathbf{v} \in H^1(K)^d$,

$$\boldsymbol{\omega}_{\hat{K}}^k(\hat{\mathbf{v}}) = [\boldsymbol{\omega}_K^k(\mathbf{v})]^\wedge \Big(:= |\det B_K|\, B_K^{-1}\, \boldsymbol{\omega}_K^k(\mathbf{v}) \circ T_K \Big).$$

(Proof) We have to show that all $\hat{K}$-moments and $F_{\hat{K}}$-moments of $\hat{\mathbf{v}}$ and $[\boldsymbol{\omega}_K^k(\mathbf{v})]^\wedge$ coincide. In fact we have that any $\mathbf{w} \in (\mathbb{P}_{k-2})^d$,

$$\begin{aligned}\int_{\hat{K}} [\boldsymbol{\omega}_K^k(\mathbf{v})]^\wedge \cdot \hat{\mathbf{w}} &= \int_{\hat{K}} |\det B_K| \left(B_K^{-1}\boldsymbol{\omega}_K^k(\mathbf{v}) \circ T_K\right) \cdot \hat{\mathbf{w}} = \int_K \left(B_K^{-1}\boldsymbol{\omega}_K^k(\mathbf{v})\right) \cdot \mathbf{w} \\ &= \int_K \boldsymbol{\omega}_K^k(\mathbf{v}) \cdot (B_K^{-T}\mathbf{w}) = \int_K \mathbf{v} \cdot (B_K^{-T}\mathbf{w}) = \int_K \left(B_K^{-1}\mathbf{v}\right) \cdot \mathbf{w} \\ &= \int_{\hat{K}} |\det B_K| \left(B_K^{-1}\mathbf{v} \circ T_K\right) \cdot \hat{\mathbf{w}} = \int_{\hat{K}} \hat{\mathbf{v}} \cdot \hat{\mathbf{w}},\end{aligned}$$

and for any $\hat{\psi} \in \mathbb{P}_{k-1}$,

$$\int_{\partial\hat{K}} \hat{\psi}\, \hat{\mathbf{v}} \cdot \hat{\mathbf{n}} = \int_{\partial K} \psi\, \mathbf{v} \cdot \mathbf{n} = \int_{\partial K} \psi\, \boldsymbol{\omega}_K^k(\mathbf{v}) \cdot \mathbf{n} = \int_{\partial\hat{K}} \hat{\psi}\, [\boldsymbol{\omega}_K^k(\mathbf{v})]^\wedge \cdot \mathbf{n}.$$

Proposition 2.12. *For any $\mathbf{v} \in H^m(K)^d$, $m \geq 0$, we have $\hat{\mathbf{v}} \in H^m(\hat{K})^d$, and there exists a constant $C = C(m,d)$ such that*

(a) $|\hat{\mathbf{v}}|_{m,\hat{K}} \leq C\|B_K^{-1}\|\, \|B_K\|^m\, |\det B_K|^{\frac{1}{2}}\, |\mathbf{v}|_{m,K} \quad \forall\, \mathbf{v} \in H^m(K)^d,$

(b) $|\mathbf{v}|_{m,K} \leq C\|B_K\|\, \|B_K^{-1}\|^m\, |\det B_K|^{-\frac{1}{2}}\, |\hat{\mathbf{v}}|_{m,\hat{K}} \quad \forall\, \hat{\mathbf{v}} \in H^m(\hat{K})^d.$

Theorem 2.13. *If $1 \leq l \leq k$ and $0 \leq m \leq l$, then there exists a constant $C = C(\hat{K}, \boldsymbol{\omega}_{\hat{K}}^k, l, m, d)$ such that*

$$|\mathbf{v} - \boldsymbol{\omega}_K^k(\mathbf{v})|_{m,K} \leq C\, \frac{h_K^{l+1}}{\rho_K^{m+1}}\, |\mathbf{v}|_{l,K} \quad \forall\, \mathbf{v} \in H^l(K)^d$$

and for each $\mathbf{v} \in H^1(K)^d$ with $\operatorname{div}\mathbf{v} \in H^l(K)$

$$|\operatorname{div}\mathbf{v} - \operatorname{div}\boldsymbol{\omega}_K^k(\mathbf{v})|_{m,K} \leq C\, \frac{h_K^{l}}{\rho_K^{m}}\, |\operatorname{div}\mathbf{v}|_{l,K} \quad \forall\, \mathbf{v} \in H^l(K)^d.$$

Theorem 2.14 (Global Interpolation error)**.**
Let $\mathcal{T}_h$ be a regular family of triangulations and assume that $k \geq 1$. Then there exists C, independent of h, such that

$$\|\mathbf{v} - \boldsymbol{\omega}_K^k(\mathbf{v})\|_{H(\mathrm{div};\Omega)} \leq C\, h^l \left(|\mathbf{v}|_{l,\Omega} + |\operatorname{div}\mathbf{v}|_{l,\Omega}\right)$$

for each $\mathbf{v} \in H^l(\Omega)^d$ with $\operatorname{div}\mathbf{v} \in H^l(K)$, $1 \leq l \leq k$.

2.5 Projection Operators

The interpolation operator gives optimal error estimates in Sobolev norms whenever the function to be interpolated enjoys the minimal requirement to be continuous. In view of finite element analysis, it is useful to introduce other approximation operators, remarkably the $L^2(\Omega)$- and

$H^1(\Omega)$-orthogonal projection operators, which make sense on functions which need not to be continuous.

Let H be a Hilber space and S a closed subspace of H.
Define the *orthogonal projection operator* P_S in H over S such that

$$P_S(v) \in S \,:\, \big(P_S(v), \varphi\big)_H = \big(v, \varphi\big)_H \quad \forall\, \varphi \in S.$$

It is characterized by the property

$$\|v - P_S(v)\|_H = \min_{\varphi \in S} \|v - \varphi\|_H.$$

Note that

$$P_S^2 = P_S \quad \text{and} \quad \|P_S(v)\|_H \le \|v\|_H \quad \forall\, v \in H.$$

We are interested in the following projection operators

$$P_h^k : L^2(\Omega) \longrightarrow X_h^k, \qquad P_{1,h}^k : H^1(\Omega) \longrightarrow X_h^k,$$

$$p_h^k : L^2(\Omega) \longrightarrow Y_h^k, \qquad Q_h^k : H(\mathrm{div}\,;\Omega) \longrightarrow W_h^k,$$

where

$$Y_h^k := \{v_h \in L^2(\Omega) \,:\, v_h|_K \in \mathbb{P}_k \quad \forall\, K \in \mathcal{T}_h\}, \quad k \ge 0.$$

Proposition 2.15. *If $\mathcal{T}_h$ is a regular family of triangulations, then*

(a) $\|v - P_h^k(v)\| \le C\, h^{\ell+1}\, |v|_{\ell+1}, \quad 1 \le \ell \le k, \quad \forall\, v \in H^{\ell+1}(\Omega),$

(b) $\|v - P_{1,h}^k(v)\|_1 \le C\, h^{\ell}\, |v|_{\ell+1}, \quad 1 \le \ell \le k, \quad \forall\, v \in H^{\ell+1}(\Omega),$

(c) $\|\mathbf{v} - Q_h^k(\mathbf{v})\|_{H(\mathrm{div}\,;\Omega)} \le C\, h^{\ell}\, \big(|\mathbf{v}|_\ell + |\mathrm{div}\, \mathbf{v}|_\ell\big),$
$\qquad 1 \le \ell \le k, \quad \forall\, \mathbf{v} \in H^\ell(\Omega)^d,\ \mathrm{div}\, \mathbf{v} \in H^\ell(\Omega)^d.$

(d) $\|v - P_{1,h}^k(v)\|_1 \le C\, |v|_1$ *and* $\|\mathbf{v} - Q_h^k(\mathbf{v})\|_{H(\mathrm{div}\,;\Omega)} \le \|\mathbf{v}\|_{H(\mathrm{div}\,;\Omega)}.$
Under H^2 regularity assumption, we have the following $L^2(\Omega)$ error estimates.

(e) $\|v - P_{1,h}^k(v)\| \le C\, h^{\ell+1}\, |v|_{\ell+1}, \quad 0 \le \ell \le k, \quad \forall\, v \in H^{\ell+1}(\Omega),$

(f) $\|v - P_h^k(v)\| \le C\, h\, |v|_1, \quad \forall\, v \in H^1(\Omega),$

(g) $\|v - P_h^k(v)\|_1 \le C\, \|v - P_{1,h}^k(v)\|_1 \quad \forall\, v \in H^1(\Omega),$

(h) $\|v - p_h^k(v)\| \le C\, h^{\ell+1}\, |v|_{\ell+1}, \quad 0 \le \ell \le k, \quad \forall\, v \in H^{\ell+1}(\Omega).$

Proof. (a), (b), (c) are immediate consequences of the orthogonal projection property and interpolation error estimate.

The second result of (d) is also an immediate consequences of the orthogonal projection property.

For the first result of (d), from the orthogonal projection property we have

$$\|v - P_{1,h}^k(v)\|_1 = \min_{\phi \in X_h^k} \|v - \phi\|_1 \le \|v\|_1,$$

and then using the standard compactness argument yields the conclusion.

Now we will prove (e) using the duality argument.

(Duality Argument).

For a given $r \in L^2(\Omega)$, by Riesz representation theorem there exists $\phi(r) \in H^1(\Omega)$ such that

$$(\phi(r), \psi)_{H^1(\Omega)} = (r, \psi)_{L^2(\Omega)} \quad \forall\, \psi \in H^1(\Omega).$$

Assume that $\phi(r) \in H^2(\Omega)$, i.e., H^2-regularity.
By closed graph theorem, there exists a constant $C = C(\Omega)$ such that

$$|\phi(r)|_2 \le C\,\|r\| \quad \forall\, r \in L^2(\Omega).$$

($L^2(\Omega)$ error estimate).
Set $e = v - v_h \in L^2(\Omega)$ with $v_h = P_{1,h}^k(v)$.

$$\begin{aligned}
\|e\|^2 &= (e, e) = (\phi(e), e)_{H^1(\Omega)} \quad \text{by duality argument} \\
&= (e, \phi(e) - w_h)_{H^1(\Omega)} \quad \forall\, w_h \in X_h^k(\Omega), \quad \text{by } (e, w_h) = 0 \\
&\le \|e\|_1\, \|\phi(e) - w_h\|_1 \quad \forall\, w_h \in X_h^k(\Omega).
\end{aligned}$$

Since $\phi(e) \in H^2(\Omega) \subset C^0(\bar{\Omega})$ by Sobolev embedding theorem, we can take $w_h = \pi_h^k(\phi(e))$ so that

$$\begin{aligned}
\|e\|^2 &\le \|e\|_1\, \|\phi(e) - \pi_h^k(\phi(e))\|_1 \\
&\le C\,h\,\|e\|_1\, |\phi(e)|_2 \quad \text{by interpolation error} \\
&\le C\,h\,\|e\|_1\, \|e\| \quad \text{by } H^2\text{-regularity}
\end{aligned}$$

Hence, by the last inequality and (b) of Proposition 2.15,

$$\|v - P_{1,h}^k(v)\| \le C\,h\,\|v - P_{1,h}^k(v)\|_1 \le C\,h^{\ell+1}\,|v|_{\ell+1}, \quad 1 \le \ell \le k.$$

Combining the last inequality with (d) yields (e).

Using (e) and the orthogonal projection property, we have

$$\|v - P_h^k(v)\| \le \min_{\phi \in X_h^k} \|v - \phi\| \le \|v - P_{1,h}^k(v)\| \le Ch\,|v|_1, \quad v \in H^1(\Omega)$$

which completes (f).

Since $P^k_{1,h}(v) \in X^k_h$,

$$\begin{aligned}\|v - P^k_h(v)\|_1 &\le \|v - P^k_{1,h}(v)\|_1 + \|P^k_h(v) - P^k_{1,h}(v)\|_1 \\ &= \|v - P^k_{1,h}(v)\|_1 + \|P^k_h[v - P^k_{1,h}(v)]\|_1.\end{aligned}$$

Using the inverse inequality and the fact that

$$\|Q(v)\|_H \le \|v\|_H, \quad \text{if } Q \text{ is } H\text{-orthogonal projection}$$

we have

$$\begin{aligned}\|P^k_h[v - P^k_{1,h}(v)]\|_1 &\le \sqrt{1 + Ch^{-2}}\|P^k_h[v - P^k_{1,h}(v)]\| \le \sqrt{1 + Ch^{-2}}\|v - P^k_{1,h}(v)\| \\ &\le \sqrt{1 + Ch^{-2}}\, Ch\, \|v - P^k_{1,h}(v)\|_1 \le C\, \|v - P^k_{1,h}(v)\|_1.\end{aligned}$$

This completes (g).

Let P^k_K be the $L^2(K)$-orthogonal projection onto $\mathbb{P}_k$. Then, we have

$$p^k_h(v)|_K = P^k_K(v|_K) \quad \forall\, v \in L^2(\Omega).$$

Using the similar arguments of Theorem 2.11 we have the conclusion (h). □

Definition (Quasi-uniform triangulation)**.**

A family of triangulation $\mathcal{T}_h$ $(h > 0)$ is called *quasi-uniform* if it is regular and there exists $\tau > 0$ such that

$$\min_{K \in \mathcal{T}_h} h_K \ge \tau\, h \quad \forall\, h > 0.$$

This yields the so-called *inverse-inequality*: there exists a constant C such that

$$\|\nabla v_h\| \le C\, \frac{1}{h}\, \|v_h\| \quad \forall\, v_h \in X^k_h(\Omega).$$

Theorem 2.16 (Approximation Properties)**.**

Let $\mathcal{T}_h$ be a family of quasi-uniform triangulations. Then we have

(a) $\|v - P^k_h(v)\| + h\, \|v - P^k_h(v)\|_1 \le C\, h^{\ell+1}\, |v|_{\ell+1}, \quad 0 \le \ell \le k, \quad \forall\, v \in H^{\ell+1}(\Omega),$

(b) $\|v - P^k_{1,h}(v)\| + h\, \|v - P^k_{1,h}(v)\|_1 \le C\, h^{\ell+1}\, |v|_{\ell+1}, \quad 0 \le \ell \le k, \quad \forall\, v \in H^{\ell+1}(\Omega),$

if we assume the H^2 regularity assumption for $\ell = 0$.

3 Variational Formulation

3.1 Variational Formulation

Let Ω be a bounded domain of $\mathbb{R}^d$, $d = 2, 3$ and let $\partial\Omega$ be its boundary. Consider a boundary value problem of the form

$$\begin{cases} Lu &= f \qquad \text{in } \Omega, \\ Bu &= 0 \qquad \text{on } \partial\Omega^*, \end{cases} \tag{3.1}$$

where f is a given function, u is the unknown, L is a linear differential operator and B is an affine boundary operator. Also, $\partial\Omega^*$ is a subset of $\partial\Omega$ possibly the whole boundary.

The problem (3.1) can generally be reformulated in a weak (or variational) form. This approach allows the search of weak solutions, which don't necessarily satisfy the equations (3.1) in a pointwise manner.

Formally speaking, the weak formulation can be derived after multiplication of the differential equation by a suitable set of test functions and performing an integration upon the domain.

As a result, we obtain a problem that reads

$$\text{find} \quad u \in W \; : \; \mathcal{A}(u, v) = \mathcal{F}(v) \quad \forall\, v \in V, \tag{3.2}$$

where W is the space of admissible solutions and V is the space of test functions. Both W and V can be assumed to be a Hilbert spaces.

$\mathcal{F}$ is a linear functional on V that accounts for the right hand side f as well as for possible non-homogeneous boundary terms.

$\mathcal{A}(\cdot, \cdot)$ is a bilinear form corresponding to the differential operator L.

The boundary conditions on u can be enforced directly in the definition of W (*essential boundary conditions*), or they can be achieved indirectly through a suitable choice of the bilinear form $\mathcal{A}$ as well as the functional $\mathcal{F}$ (*natural boundary conditions*).

In most case, $W = V$. Denote by $(\cdot, \cdot)$ the $L^2(\Omega)$ inner product.

Example (Poisson problem)**.**

$$\begin{cases} -\Delta u &= f \qquad \text{in } \Omega, \\ \quad u &= 0 \qquad \text{on } \partial\Omega \end{cases} \quad \Longrightarrow \quad \begin{cases} W = V = H_0^1(\Omega) \\ \mathcal{A}(u, v) = (\nabla u, \nabla v) \\ \mathcal{F}(v) = (f, v) \end{cases}$$

Example (Stokes problem)**.**

$$\begin{cases} -\nu\Delta\mathbf{u} + \nabla p &= \mathbf{f} \qquad \text{in } \Omega, \\ \text{div } \mathbf{u} &= 0 \qquad \text{in } \Omega, \\ \mathbf{u} &= 0 \qquad \text{on } \partial\Omega \end{cases} \quad \Longrightarrow \quad \begin{cases} W = V = H_0^1(\Omega)^2 \times L_0^2(\Omega) \\ \mathcal{A}(u, v) = (\nu\nabla\mathbf{u}, \nabla\mathbf{v}) - (p, \text{div } \mathbf{v}) + (q, \text{div } \mathbf{u}) \\ \mathcal{F}(v) = (\mathbf{f}, \mathbf{v}) \end{cases}$$

where $u = (\mathbf{u}, p)$, $v = (\mathbf{v}, q) \in V$.

3.2 Some results of functional analysis

In this section we present basic functional theorems about existence and uniqueness of the solution of the variational problem.

Proposition 3.1 (Projection Theorem). *Given a closed subspace M of H and $v \in H$, there exists a unique decomposition*

$$v = P_M v + P_{M^\perp} v,$$

where $P_M : H \to M$ and $P_{M^\perp} : H \to M^\perp$ are orthogonal projections, respectively. In other words,

$$H = M \oplus M^\perp.$$

Theorem 3.2 (Riesz Representation Theorem).
Any continuous linear functional L on a Hilbert space H can be represented uniquely as

$$L(v) = (u, v)_H, \quad \textit{for some } u \in H.$$

Furthermore, we have

$$\|L\|_{H'} = \|u\|_H.$$

Proof. Uniqueness is given by

$$\begin{aligned} 0 &= L(u_1 - u_2) - L(u_1 - u_2) = (u_1, u_1 - u_2)_H - (u_2, u_1 - u_2)_H \\ &= (u_1 - u_2, u_1 - u_2)_H = \|u_1 - u_2\|_H^2. \end{aligned}$$

(Existence) Let $M = \{v \in H : L(v) = 0\} = \mathrm{Ker}(L)$. Then M is a subspace of H and $H = M \oplus M^\perp$.

Case (1): $M^\perp = \{0\}$.

In this case, $M = H$ so that $L \equiv 0$. So take $u = 0$.

Case (2): $M^\perp \neq \{0\}$.

Pick $z \in M^\perp$, $z \neq 0$. Then $L(z) \neq 0$. For $v \in H$ and $\beta = L(v)/L(z)$ we have

$$L(v - \beta z) = L(v) - \beta L(z) = 0 \quad \text{or} \quad v - \beta z \in M.$$

Thus, $v - \beta z = P_M v$ and $\beta z = P_{M^\perp} v$. In particular, if $v \in M^\perp$, then $v = \beta z$ which proves that $M^\perp$ is one-dimensional.

Choose

$$u := \frac{L(z)}{\|z\|_H^2} z.$$

Note that $u \in M^{\perp}$. We have

$$\begin{aligned}(u,v)_H &= (u,(v-\beta z)+\beta z)_H = (u,v-\beta z)_H + (u,\beta z)_H \\ &= (u,\beta z)_H = \beta \frac{L(z)}{\|z\|_H^2}(z,z)_H = \beta L(z) = L(v).\end{aligned}$$

Thus, this u is the desired element of H.

It remains to prove that $\|L\|_{H'} = \|u\|_H$. Using the dual norm,

$$\|L\|_{H'} = \sup_{0\neq v\in H} \frac{L(v)}{\|v\|_H} = \sup_{0\neq v\in H} \frac{|(u,v)_H|}{\|v\|_H} \leq \|u\|_H = \frac{|L(z)|}{\|z\|_H} \leq \|L\|_{H'}.$$

Therefore, $\|u\|_H = \|L\|_{H'}$. □

Remark. According to the Riesz Representation Theorem, there is a natural isometry between H and H' ($u \in H \longleftrightarrow L_u \in H'$). For this reason, H and H' are often identified. for example, $[L^2(\Omega)]' = L^2(\Omega)$.

Let us consider the case $W = V$ in (3.2)

$$\text{find } \ u \in V \ : \mathcal{A}(u,v) = \mathcal{F}(v) \quad \forall\, v \in V. \tag{3.3}$$

Theorem 3.3 (Lax-Milgram lemma).

Let V be a (real) Hilbert space, endowed with the norm $\|\cdot\|_V$, $\mathcal{A}(\cdot,\cdot)$ a bilinear form on $V\times V$ and $\mathcal{F}(\cdot)$ a linear continuous functional on V, i.e., $\mathcal{F}\in V'$.

Assume that $\mathcal{A}(\cdot,\cdot)$ is continuous:

$$\exists\,\gamma > 0 \quad s.t. \quad |\mathcal{A}(v,w)| \leq \gamma\,\|v\|_V\,\|w\|_V \quad \forall\, v,w \in V,$$

and coercive:

$$\exists\,\alpha > 0 \quad s.t. \quad \mathcal{A}(v,v) \geq \alpha\,\|v\|_V^2 \quad \forall\, v \in V.$$

Then, there exists a unique solution $u \in V$ solution to (3.3) and

$$\|u\|_V \leq \frac{1}{\alpha}\,\|\mathcal{F}\|_{V'}.$$

Proof. For convenience, denote by $\|\cdot\| = \|\cdot\|_V$ and $(\cdot,\cdot)$ the inner product in V.

By the Riesz representation theorem, we can write

$$\mathcal{F}(v) = (R\mathcal{F}, v) \quad \forall\, v \in V$$

and for each fixed $w \in V$

$$\mathcal{A}(w,v) = (Aw, v) \quad \forall\, v \in V$$

where the isometric operator $R : V' \to V$ and $A : V \to V$ are linear continuous operators since

$$\|R\mathcal{F}\| = \sup_{0\neq v\in V} \frac{(R\mathcal{F}, v)}{\|v\|} = \sup_{0\neq v\in V} \frac{\mathcal{F}(v)}{\|v\|} = \|\mathcal{F}\|_{V'}$$

and

$$\|Aw\| = \sup_{0\neq v\in V} \frac{(Aw, v)}{\|v\|} = \sup_{0\neq v\in V} \frac{\mathcal{A}(w, v)}{\|v\|} \le \gamma\,\|w\|.$$

Problem (3.3) is thus equivalent to the following one: for each $\mathcal{F} \in V'$, find a unique $u \in V$ such that

$$Au = R\mathcal{F}.$$

It is enough to show that A is bijection.

(Injective) From the fact that

$$\|v\|^2 \le \frac{1}{\alpha}\,(Av, v) \le \frac{1}{\alpha}\|Av\|\,\|v\|,$$

we have $\|v\| \le \frac{1}{\alpha}\|Av\|$. Thus, the uniqueness is proven.

(Surjective : the range $\mathcal{R}(A)$ of $A = V$) It is enough to show that $\mathcal{R}(A)$ is closed and $\mathcal{R}(A)^\perp = \{0\}$ because $V = \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$ if $\mathcal{R}(A)$ is a closed subspace.

Suppose that $Av_n \to w$ in V. From the fact that

$$\|v_n - v_m\| \le \frac{1}{\alpha}\|Av_n - Av_m\|,$$

v_n is a Cauchy sequence in the Hilbert space V. Set $v = \lim v_n$. Since A is continuous, $Av = w \in \mathcal{R}(A)$ and hence $\mathcal{R}(A)$ is closed.

Let $z \in \mathcal{R}(A)^\perp$. Since

$$0 = (Az, z) = \mathcal{A}(z, z) \ge \alpha\|z\|^2,$$

we have that $z = 0$.

Finally, we have

$$\|u\|^2 \le \frac{1}{\alpha}\mathcal{A}(u, u) = \frac{1}{\alpha}\mathcal{F}(u) \le \frac{1}{\alpha}\|\mathcal{F}\|_{V'}\|u\|$$

which completes the theorem. □

Remark (Symmetric Case)**.** If the bilinear form is symmetric, then $\mathcal{A}(\cdot,\cdot)$ defines a scalar product on V and hence the Riesz representation theorem suffices to infer existence and uniqueness for the solution (3.3).

In this case, the solution can be regarded as the unique solution to the *minimization problem*

$$\text{find } u \in V \text{ such that } \; J(u) \le J(v) \quad \forall\, v \in V,$$

where $J(v)$ is a quadratic functional given by

$$J(v) := \frac{1}{2}\,\mathcal{A}(v,v) - \mathcal{F}(v).$$

$\square$

Remark (Complex Case)**.** Let V be a complex Hilbert space, endowed with the norm $\|\cdot\|_V$ and $\mathcal{A}(\cdot,\cdot)$ a sesquilinear form on $V \times V$:

$$\mathcal{A}(w, c_1 v_1 + c_2 v_2) = \bar{c_1}\mathcal{A}(w,v_1) + \bar{c_2}\mathcal{A}(w,v_2), \quad w, v_1, v_2 \in V,\ c_1, c_2 \in \mathbb{C}$$

and $\mathcal{F}(\cdot)$ a linear continuous functional on V.

Assume that $\mathcal{A}(\cdot,\cdot)$ is continuous:

$$\exists\,\gamma > 0 \quad \text{s.t.} \quad |\mathcal{A}(v,w)| \le \gamma\,\|v\|_V\,\|w\|_V \quad \forall\,v,w \in V,$$

and coercive:

$$\exists\,\alpha > 0 \quad \text{s.t.} \quad |\mathcal{A}(v,v)| \ge \alpha\,\|v\|_V^2 \quad \forall\,v \in V.$$

Then, there exists a unique solution $u \in V$ solution to (3.3) and

$$\|u\|_V \le \frac{1}{\alpha}\,\|\mathcal{F}\|_{V'}.$$

$\square$

Now, let us consider the general case

$$\text{find}\ \ u \in W\ :\ \mathcal{A}(u,v) = \mathcal{F}(v) \quad \forall\,v \in V. \tag{3.4}$$

Theorem 3.4 (Extension of the Lax-Milgram Lemma)**.**
Let W and V be two (real) Hilbert spaces, endowed with the norms $\|\cdot\|_W$ and $\|\cdot\|_V$, respectively, and let $\mathcal{A}(\cdot,\cdot)$ be a bilinear form on $W \times V$ and $\mathcal{F}(\cdot)$ a linear continuous functional on V. i.e., $\mathcal{F} \in V'$.

Assume that

$$\begin{aligned}
&\exists\,\gamma > 0 \quad s.t. \quad |\mathcal{A}(w,v)| \le \gamma\,\|w\|_W\,\|v\|_V \quad \forall\,w \in W,\ v \in V,\\
&\exists\,\alpha > 0 \quad s.t. \quad \sup_{0\ne v\in V} \frac{\mathcal{A}(w,v)}{\|v\|_V} \ge \alpha\,\|w\|_W \quad \forall\,w \in W,\\
&\qquad\qquad\qquad\quad \sup_{w\in W} \mathcal{A}(w,v) > 0 \quad \forall\,0 \ne v \in V.
\end{aligned}$$

Then, there exists a unique solution $u \in W$ solution to (3.4) and

$$\|u\|_W \le \frac{1}{\alpha}\,\|\mathcal{F}\|_{V'}.$$

Proof. By the Riesz representation theorem, we can construct a linear continuous operator $A : W \to V$ such that for each $w \in V$,

$$\mathcal{A}(w,v) = (Aw,v)_V \quad \forall\, v \in V$$

and

$$\|Aw\|_V \le \sup_{v\in V} \frac{\mathcal{A}(w,v)}{\|v\|_V} \le \gamma \|w\|_W \quad \forall\, w \in W.$$

With the isometric operator $R : V' \to V$ constructed in Lax-Milgram lemma, we can reduce the problem to find a unique $u \in W$ such that

$$Au = R\mathcal{F}.$$

If $Aw = 0$, then $\mathcal{A}(w,v) = 0$ for any $v \in V$ so that $w = 0$ by the second hypothesis. Hence, A is injective.

Moreover the range $\mathcal{R}(A)$ of A is closed. In fact, if $Aw_n \to v$ in V, we have

$$\|w_n - w_m\|_W \le \frac{1}{\alpha} \sup_{0\neq v\in V} \frac{\big(A(w_n - w_m), v\big)_V}{\|v\|_V} \le \frac{1}{\alpha}\|A(w_n - w_m)\|_V,$$

hence $w_n \to w$ for some $w \in W$ and $Aw_n \to Aw$ in V.

Also, if $z \in \mathcal{R}(A)^{\perp}$, i.e.,

$$(Aw,z)_V = \mathcal{A}(w,z) = 0 \quad \forall\, w \in W,$$

it follows $z = 0$ by the third hypothesis, hence A is surjective. Finally, the stability is easily given:

$$\|u\|_W \le \frac{1}{\alpha} \sup_{0\neq v\in V} \frac{\mathcal{A}(u,v)}{\|v\|_V} \le \frac{1}{\alpha}\|Au\|_W \le \frac{1}{\alpha}\|R\mathcal{F}\|_W \le \frac{1}{\alpha}\|\mathcal{F}\|_{V'}.$$

□

□

3.3 Galerkin Method

In this section we assume that $W = V$. We consider the following variational problem:

$$\text{find } \ u \in V \ : \ \mathcal{A}(u,v) = \mathcal{F}(v) \quad \forall\, v \in V. \tag{3.5}$$

Let $\mathcal{T}_h$ be a family of regular triangulations of Ω with the mesh size h and let $\{V_h\}$ denote a family of finite dimensional subspaces of V, for example, $V_h = X_h^k$.

Assume that

$$\text{for all } v \in V, \qquad \inf_{v_h \in V_h} \|v - v_h\| \to 0 \quad \text{as } h \to 0. \tag{3.6}$$

The Galerkin approximation to (3.5) reads:

$$\text{find } \; u_h \in V_h \; : \; \mathcal{A}(u_h, v_h) = \mathcal{F}(v_h) \quad \forall\, v_h \in V_h. \tag{3.7}$$

From the algebraic point of view, let $\{\phi_j \; : \; j = 1, \cdots, N_h\}$ be a basis for V_h, so that we can set

$$u_h(\mathbf{x}) = \sum_{j=1}^{N_h} \xi_j \, \phi_j(\mathbf{x}).$$

Then, from (3.7) we deduce the following linear system of dimension N_h:

$$A\boldsymbol{\xi} = F,$$

where $\boldsymbol{\xi} = \big(\xi_j\big)$, $F := \mathcal{F}(\phi_i)$, $A_{ij} = \mathcal{A}(\phi_j, \phi_i)$ for $i, j = 1, \cdots, N_h$.

The matrix A is called the *stiffness matrix.*

By subtracting the equation (3.7) from (3.5) we have the *fundamental orthogonality*

$$\mathcal{A}(u - u_h, v_h) = 0 \quad \forall\, v_h \in V_h.$$

Theorem 3.5 (Céa Lemma)**.** *Under the assumption of Lax-Milgram lemma, there exists a unique solution u_h to (3.7) such that*

$$\|u_h\|_V \leq \frac{1}{\alpha} \|\mathcal{F}\|_{V'}.$$

If u is the solution to (3.5), then it follows

$$\|u - u_h\|_V \leq \frac{\gamma}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|_V.$$

Proof. Since V_h is a subspace of V, applying the Lax-Milgram lemma yields the existence and uniqueness of u_h and the stability.

Using the fundamental orthogonality, we have that for any $w_h \in V_h$

$$\begin{aligned} \alpha \|u - u_h\|_V^2 &\leq \mathcal{A}(u - u_h, u - u_h) = \mathcal{A}(u - u_h, u - u_h) + \mathcal{A}(u - u_h, u_h + w_h) \\ &= \mathcal{A}(u - u_h, u - w_h) \leq \gamma \|u - u_h\|_V \, \|u - w_h\|_V. \end{aligned}$$

This completes the theorem. $\square$

Remark (Ritz Galerkin Method). When $\mathcal{A}(\cdot,\cdot)$ is symmetric, Galerkin method is referred to as the Ritz Galerkin method, in this case existence and uniqueness follows from the Riesz representation theorem. Also, u_h turns out to be the orthogonal projection of u upon V_h with respect to the the scalar product $\mathcal{A}(\cdot,\cdot)$.

Remark (Stiffness Matrix A). The stiffness matrix A is positive definite, i.e.,

$$(A\boldsymbol{\eta},\boldsymbol{\eta}) = \sum_{i,j=1}^{N_h} \eta_i \mathcal{A}(\phi_j,\phi_i)\eta_j = \mathcal{A}(\eta_h,\eta_h) > 0, \quad \forall\, \mathbf{0} \neq \boldsymbol{\eta} \in \mathbb{R}^{N_h}$$

since $(A\boldsymbol{\eta},\boldsymbol{\eta}) = \mathcal{A}(\eta_h,\eta_h)$ where $\eta_h(\mathbf{x}) = \sum_{j=1}^{N_h} \eta_j\phi_j(\mathbf{x})$ and $\boldsymbol{\eta} = \big(\eta_j\big)$.

In particular, any eigenvalue of A has positive real part:

Let $A(\mathbf{x}_1 + i\mathbf{x}_2) = (\lambda_1 + i\lambda_2)(\mathbf{x}_1 + i\mathbf{x}_2)$.

$$(A\mathbf{x}_1,\mathbf{x}_1) = \lambda_1(\mathbf{x}_1,\mathbf{x}_1) - \lambda_2(\mathbf{x}_2,\mathbf{x}_1), \quad (A\mathbf{x}_2,\mathbf{x}_2) = \lambda_1(\mathbf{x}_2,\mathbf{x}_2) + \lambda_2(\mathbf{x}_1,\mathbf{x}_2)$$

by summing

$$(A\mathbf{x}_1,\mathbf{x}_1) + (A\mathbf{x}_2,\mathbf{x}_2) = \lambda_1\big[(\mathbf{x}_1,\mathbf{x}_1) + (\mathbf{x}_2,\mathbf{x}_2)\big].$$

Since $(A\mathbf{x}_1,\mathbf{x}_1)+(A\mathbf{x}_2,\mathbf{x}_2) > 0$ and $(\mathbf{x}_1,\mathbf{x}_1)+(\mathbf{x}_2,\mathbf{x}_2) > 0$, we are led to $\lambda_1 > 0$.

When the bilinear form $\mathcal{A}$ is symmetric, it follows immediately that A is also symmetric and any eigenvalue of A is positive real value.

Remark (Example). If $V = H^1(\Omega)$ and $V_h = X_h^k$, then by Céa Lemma and approximation property we have the following error estimate in $H^1(\Omega)$-norm:

$$\|u - u_h\|_1 \leq \frac{\gamma}{\alpha} \inf_{v_h\in V_h} \|u - v_h\|_1 \leq \frac{\gamma\, C}{\alpha}\, h^\ell\, |u|_{\ell+1}$$

provided $u \in H^{\ell+1}(\Omega)$, $0 \leq \ell \leq k$.

3.4 Petrov-Galerkin Method

In this section we consider the following variational problem:

$$\text{(3.8)} \qquad \text{find } \; u_h \in W_h \,:\, \mathcal{A}_h(u_h,v_h) = \mathcal{F}_h(v_h) \quad \forall\, v_h \in V_h$$

where W_h and V_h are two relative finite dimensional subspaces of W and V such that $W_h \neq V_h$ but $\dim W_h = \dim V_h = N_h$ for all $h > 0$, and $\mathcal{A}_h$ and $\mathcal{F}_h$ are convenient approximations to $\mathcal{A}$ and $\mathcal{F}$, respectively. Spaces W and V need not be necessarily different.

Due to Babuška and Aziz we have the following theorem.

Theorem 3.6. *Under the assumptions of Theorem 3.4, suppose further that $\mathcal{F}_h$ is a linear map and that $\mathcal{A}_h$ is a bilinear form satisfying the same properties of $\mathcal{A}$, (2) and (3) in Theorem 3.4, with the constant α_h. Then, there exists a unique solution u_h to (3.8) such that*

$$\|u_h\|_W \leq \frac{1}{\alpha_h} \sup_{0 \neq v_h \in V_h} \frac{\mathcal{F}_h(v_h)}{\|v_h\|_V}.$$

Moreover, if u is the solution of (3.4), it follows

$$\|u - u_h\|_W \leq \inf_{w_h \in W_h} \left[\left(1 + \frac{\gamma}{\alpha_h}\right) \|u - w_h\|_W + \frac{1}{\alpha_h} \sup_{0 \neq v_h \in V_h} \frac{|\mathcal{A}(w_h, v_h) - \mathcal{A}_h(w_h, v_h)|}{\|v_h\|_V} \right] + \frac{1}{\alpha_h} \sup_{0 \neq v_h \in V_h} \frac{|\mathcal{F}(v_h) - \mathcal{F}_h(v_h)|}{\|v_h\|_V}.$$

Proof. For any fixed h, existence, uniqueness and stability follow from the extension of Lax-Milgram lemma.

For all $w_h \in W_h$ and $v_h \in V_h$, we have

$$\mathcal{A}_h(u_h - w_h, v_h) = \mathcal{A}(u - w_h, v_h) + \mathcal{A}(w_h, v_h) - \mathcal{A}_h(w_h, v_h) + \mathcal{F}_h(v_h) - \mathcal{F}(v_h)$$

so that

$$\alpha_h \|u_h - w_h\|_W \leq \gamma \|u - w_h\|_W + \sup_{0 \neq v_h \in V_h} \left[\frac{|\mathcal{A}(w_h, v_h) - \mathcal{A}_h(w_h, v_h)|}{\|v_h\|_V} + \frac{|\mathcal{F}(v_h) - \mathcal{F}_h(v_h)|}{\|v_h\|_V} \right].$$

Finally, using the triangle inequality

$$\|u - u_h\|_W \leq \|u - w_h\|_W + \|u_h - w_h\|_W$$

yields the theorem. □

Examples of Petrov-Galerkin approximations are furnished by the so-called τ-method (the trial functions do not individually satisfy the boundary conditions; thus, some equations are needed to ensure that the global expansion satisfies the boundary conditions).

3.5 Generalized Galerkin Method

In this section we consider the following variational problem:

$$\text{find } \; u_h \in V_h \, : \, \mathcal{A}_h(u_h, v_h) = \mathcal{F}_h(v_h) \quad \forall\, v_h \in V_h \tag{3.9}$$

where V_h is a family of finite dimensional subspaces of V, and $\mathcal{A}_h$ and $\mathcal{F}_h$ are convenient approximations to $\mathcal{A}$ and $\mathcal{F}$, respectively. This is a special subcase of Petrov-Galerkin Method including the collocation method in its weak form. $\mathcal{F}_h(\cdot)$ is a linear form defined on V_h and $\mathcal{A}_h(\cdot, \cdot)$ is a bilinear form defined over $V_h \times V_h$, and they do not necessarily make sense when applied to elements of V.

Theorem 3.7 (The first Strang Lemma).

Under the assumptions of Theorem 3.3, suppose further that $\mathcal{F}_h$ is a linear map and that $\mathcal{A}_h$ is a bilinear form which is uniformly coercive (independent of h) over $V_h \times V_h$:

$$\mathcal{A}_h(v_h, v_h) \geq \alpha^* \|v_h\|_V^2 \quad \forall\, v_h \in V_h.$$

Then, there exists a unique solution u_h to (3.9) such that

$$\|u_h\|_V \leq \frac{1}{\alpha^*} \sup_{0 \neq v_h \in V_h} \frac{\mathcal{F}_h(v_h)}{\|v_h\|_V}.$$

Moreover, if u is the solution of (3.5), it follows

$$\begin{aligned}\|u - u_h\|_V \leq \inf_{w_h \in V_h} &\left[\left(1 + \frac{\gamma}{\alpha^*}\right) \|u - w_h\|_V + \frac{1}{\alpha^*} \sup_{0 \neq v_h \in V_h} \frac{|\mathcal{A}(w_h, v_h) - \mathcal{A}_h(w_h, v_h)|}{\|v_h\|_V} \right] \\ &+ \frac{1}{\alpha^*} \sup_{0 \neq v_h \in V_h} \frac{|\mathcal{F}(v_h) - \mathcal{F}_h(v_h)|}{\|v_h\|_V}.\end{aligned}$$

Proof. The existence, uniqueness and stability follow from the Lax-Milgram lemma.

Let $w_h \in V_h$. Setting $\sigma_h = u_h - w_h \neq 0$, we obtain

$$\begin{aligned}\alpha^* \|\sigma_h\|_V^2 &\leq \mathcal{A}_h(\sigma_h, \sigma_h) \\ &= \mathcal{A}(u - w_h, \sigma_h) + \mathcal{A}(w_h, \sigma_h) - \mathcal{A}_h(w_h, \sigma_h) + \mathcal{F}_h(\sigma_h) - \mathcal{F}(\sigma_h).\end{aligned}$$

We have

$$\begin{aligned}\alpha^* \|\sigma_h\|_V &\leq \gamma \|u - w_h\|_V + \frac{|\mathcal{A}(w_h, \sigma_h) - \mathcal{A}_h(w_h, \sigma_h)|}{\|\sigma_h\|_V} + \frac{|\mathcal{F}_h(\sigma_h) - \mathcal{F}(\sigma_h)|}{\|\sigma_h\|_V} \\ &\leq \gamma \|u - w_h\|_V + \sup_{0 \neq v_h \in V_h} \frac{|\mathcal{A}(w_h, v_h) - \mathcal{A}_h(w_h, v_h)|}{\|v_h\|_V} + \sup_{0 \neq v_h \in V_h} \frac{|\mathcal{F}_h(v_h) - \mathcal{F}(v_h)|}{\|v_h\|_V}.\end{aligned}$$

The above inequality is true also when $\sigma_h = 0$. Using the triangle inequality

$$\|u - u_h\|_W \leq \|u - w_h\|_W + \|u_h - w_h\|_W$$

yields the theorem. □

Proposition 3.8. *Under the same assumptions of the previous theorem, suppose further that the bilinear form $\mathcal{A}_h(\cdot, \cdot)$ is defined at (u, v_h), where u is the solution to (3.5) and $v_h \in V_h$, and satisfies for a suitable $\gamma^* > 0$*

$$|\mathcal{A}_h(u - w_h, v_h)| \leq \gamma^* \|u - w_h\|_V \|v_h\|_V \quad \forall\, w_h, v_h \in V_h,$$

uniformly with respect to $h > 0$. Then the following convergence estimates holds

$$\|u - u_h\|_V \leq \left(1 + \frac{\gamma^*}{\alpha^*}\right) \inf_{w_h \in V_h} \|u - w_h\|_V + \frac{1}{\alpha^*} \sup_{0 \neq v_h \in V_h} \frac{|\mathcal{A}_h(u, v_h) - \mathcal{F}_h(v_h)|}{\|v_h\|_V}.$$

Proof. For all $w_h \in V_h$, we have

$$\mathcal{A}_h(u_h - w_h, u_h - w_h) = \mathcal{A}_h(u - w_h, u_h - w_h) + \mathcal{F}_h(u_h - w_h) - \mathcal{A}_h(u, u_h - w_h)$$

so that

$$\alpha^* \|u_h - w_h\|_V \leq \gamma^* \|u - w_h\|_V + \sup_{0 \neq v_h \in V_h} \frac{|\mathcal{F}_h(v_h) - \mathcal{A}_h(u, v_h)|}{\|v_h\|_V}.$$

The triangle inequality yields the theorem. □

Remark (Non-conforming approximation). When $V_h \nsubseteq V$, the bilinear form $\mathcal{A}$ is thus not necessarily defined on $V_h \times V_h$. Assume that a norm $\|\cdot\|_h$ and the approximate bilinear form $\mathcal{A}_h$ are defined in $(V + V_h)$, and that the approximate linear functional $\mathcal{F}_h$ is defined on V_h. We require that there exist constants $\alpha^* > 0$ and $\gamma^* > 0$ such that for each $h > 0$

$$\mathcal{A}_h(v_h, v_h) \geq \alpha^* \|v_h\|_h^2 \quad \forall\, v_h \in V_h,$$
$$|\mathcal{A}_h(w, v_h)| \leq \gamma^* \|w\|_h \|v_h\|_h \quad \forall\, w \in (V + v_h),\ v_h \in V_h.$$

Then by the so-called *second Strang lemma* we have

$$\|u - u_h\|_h \leq \left(1 + \frac{\gamma^*}{\alpha^*}\right) \inf_{w_h \in V_h} \|u - w_h\|_h + \frac{1}{\alpha^*} \sup_{0 \neq v_h \in V_h} \frac{|\mathcal{A}_h(u, v_h) - \mathcal{F}_h(v_h)|}{\|v_h\|_h}.$$

The proof is quite similar to that of the previous proposition.

For example, consider P_1 non-conforming triangular finite element method for Poisson problem:

$$\mathcal{A}_h(v_h, w_h) := \sum_{K \in \mathcal{T}_h} (\nabla v_h, \nabla w_h)_K \quad \text{and} \quad \|v_h\| := \mathcal{A}_h(v_h, v_h)^{\frac{1}{2}}.$$

4 Galerkin Approximation of Elliptic Problems

4.1 Problem Formulation

Let Ω be a bounded domain in $\mathbb{R}^d$ with a Lipschitz continuous boundary $\partial\Omega$.

Consider the second order linear operator L defined by

$$Lu = -\nabla \cdot A\nabla u + \mathbf{b} \cdot \nabla u + \nabla \cdot (\tilde{\mathbf{b}}\, u) + c\, u, \tag{4.1}$$

where a matrix function $A = (a_{ij})$, vector functions $\mathbf{b}$ and $\hat{\mathbf{b}}$, and a scalar function c are given coefficients.

If the coefficients $\mathbf{b}$ and $\tilde{\mathbf{b}}$ are regular enough, we can omit either $\mathbf{b} \cdot \nabla u$ or $\nabla \cdot (\tilde{\mathbf{b}}\, u)$ without loosing generality.

Definition (Elliptic oprator)**.** The differential operator L is said to be *elliptic* in Ω if there exists a constant $\alpha_0 > 0$ such that

$$\alpha_0\, |\boldsymbol{\xi}|^2 \le \boldsymbol{\xi}^T A\, \boldsymbol{\xi} \quad \text{for each } \boldsymbol{\xi} \in \mathbb{R}^d \text{ and a.e. } x \in \Omega.$$

The bilinear form associating to L is

$$a(u,v) = (A\nabla u, \nabla v) + (\mathbf{b} \cdot \nabla u, v) - (\tilde{\mathbf{b}}\, u, \nabla v) + (cu, v).$$

Assume that the coefficients hold

$$a_{ij},\, b_i,\, \tilde{b}_i,\, c \in L^\infty(\Omega).$$

Let V be a closed subspace of $H^1(\Omega)$ satisfying

$$H_0^1(\Omega) \subset V \subset H^1(\Omega).$$

The variational problem we are interested is as follows: For a given $\mathcal{F} \in V'$,

$$\text{find } u \in V \quad : \quad \mathcal{A}(u,v) = \mathcal{F}(v) \quad \forall\, v \in V, \tag{4.2}$$

where the bilinear form $\mathcal{A}(\cdot,\cdot)$ coincides with $a(\cdot,\cdot)$ up to the sum of possible boundary terms.

With the equation

$$Lu = f \quad \text{in } \Omega$$

(Examples of Boundary Conditions).

a. **The Dirichlet Problem**

$$u = 0 \quad \text{on } \partial\Omega.$$

$$\Longrightarrow \quad \mathcal{A}(u,v) = a(u,v), \quad \mathcal{F}(v) = (f,v), \quad V = H_0^1(\Omega).$$

b. **The Neumann Problem**

$$\frac{\partial u}{\partial \mathbf{n}_L} = g \quad \text{on } \partial\Omega$$

where the conormal derivative of u is given by

$$\frac{\partial u}{\partial n_L} := \mathbf{n} \cdot A\nabla u - (\tilde{\mathbf{b}} \cdot \mathbf{n})u$$

If $A = I$ and $\tilde{\mathbf{b}} = \mathbf{0}$, then $\dfrac{\partial u}{\partial \mathbf{n}_L} = \dfrac{\partial u}{\partial \mathbf{n}} = \mathbf{n} \cdot \nabla u$.

$$\Longrightarrow \quad \mathcal{A}(u,v) = a(u,v), \quad \mathcal{F}(v) = (f,v) + (g,v)_{\partial\Omega}, \quad V = H^1(\Omega).$$

c. **The Mixed Problem**

$$\begin{aligned} u &= 0 \qquad \text{on } \Gamma_D, \\ \frac{\partial u}{\partial \mathbf{n}_L} &= g \qquad \text{on } \Gamma_N. \end{aligned}$$

$$\Longrightarrow \quad \mathcal{A}(u,v) = a(u,v), \quad \mathcal{F}(v) = (f,v) + (g,v)_{\Gamma_N}, \quad V = H_{\Gamma_D}^1(\Omega).$$

d. **The Robin Problem**

$$\frac{\partial u}{\partial \mathbf{n}_L} + \kappa\, u = g \quad \text{on } \partial\Omega$$

where κ is a given function.

$$\Longrightarrow \quad \mathcal{A}(u,v) = a(u,v) + (\kappa u, v)_{\partial\Omega}, \quad \mathcal{F}(v) = (f,v) + (g,v)_{\partial\Omega}, \quad V = H^1(\Omega).$$

4.2 Existence and Uniqueness

The basic ingredient for proving the existence of a solution is the Lax-Milgram lemma.

For any $f \in L^2(\Omega)$, $v \to (f, v)$ on $H_0^1(\Omega)$ is a continuous linear functional.

The continuity of bilinear form $a(\cdot, \cdot)$ can be easily verified by using the $L^\infty(\Omega)$ coefficients.

Hence, we need only to check the coercivity of $a(\cdot, \cdot)$ under suitable assumptions on the data.

The ellipticity assumption yields

$$\alpha_0 \|\nabla v\|^2 = \alpha_0 (\nabla v, \nabla v) \le (A \nabla v, \nabla v) \quad \forall\, v \in H^1(\Omega).$$

For a convenience, denote the remaining term in $a(v, v)$ by

$$R := (\mathbf{b} \cdot \nabla v, v) - (\tilde{\mathbf{b}}\, v, \nabla v) + (cv, v).$$

Note that

$$(\mathbf{b} \cdot \nabla v, v) = \frac{1}{2} \left(\mathbf{b}, \nabla(v^2)\right) \quad \text{and} \quad - (\tilde{\mathbf{b}}\, v, \nabla v) = -\frac{1}{2} \left(\tilde{\mathbf{b}}, \nabla(v^2)\right).$$

Hence, we have

$$R := \frac{1}{2} \left(\mathbf{b} - \tilde{\mathbf{b}}, \nabla(v^2)\right) + (cv, v) = \left(-\frac{1}{2} \operatorname{div} (\mathbf{b} - \tilde{\mathbf{b}}) + c, v^2\right) + \frac{1}{2} \left(\mathbf{n} \cdot (\mathbf{b} - \tilde{\mathbf{b}}), v^2\right)_{\partial\Omega}.$$

Let C_Ω be the Poincaré constant satisfying

$$\|v\|^2 \le C_\Omega \|\nabla v\|^2 \quad \forall\, v \in H_0^1(\Omega).$$

Assume that $\operatorname{div} (\tilde{\mathbf{b}} - \mathbf{b}) \in L^\infty(\Omega)$.

a. Dirichlet Problem :

If there exists a constant η such that

$$-\eta \le -\frac{1}{2} \operatorname{div} (\mathbf{b} - \tilde{\mathbf{b}}) + c, \quad \text{a.e. } x \in \Omega \quad \text{with} \quad -\infty < \eta < \frac{\alpha_0}{C_\Omega},$$

then $a(\cdot, \cdot)$ is coercive since $\alpha_0 - \eta C_\Omega > 0$ and

$$a(v, v) = (A \nabla v, \nabla v) + R \ge \alpha_0 \|\nabla v\|^2 - \eta \|v\|^2 \ge \begin{cases} (\alpha_0 - \eta C_\Omega) \|\nabla v\|^2 & \text{if } \eta \ge 0, \\ \alpha_0 \|\nabla v\|^2 & \text{if } \eta < 0, \end{cases}$$

$$\ge \begin{cases} \frac{\alpha_0 - \eta C_\Omega}{1 + C_\Omega} \, \|v\|_1^2 & \text{if } \eta \ge 0, \\ \frac{\alpha_0}{1 + C_\Omega} \, \|v\|_1^2 & \text{if } \eta < 0. \end{cases}$$

b. Neumann Problem :

If there exists a constant $\mu_0 > 0$ such that

$$0 < \mu_0 \le -\frac{1}{2} \operatorname{div} (\mathbf{b} - \tilde{\mathbf{b}}) + c \quad \text{a.e. } x \in \Omega \quad \text{and} \quad \mathbf{n} \cdot (\mathbf{b} - \tilde{\mathbf{b}}) \ge 0 \quad \text{a.e. } x \in \partial\Omega, \tag{4.3}$$

then

$$R = \left(-\frac{1}{2}\operatorname{div}(\mathbf{b}-\tilde{\mathbf{b}}) + c, v^2\right) + \frac{1}{2}\left(\mathbf{n}\cdot(\mathbf{b}-\tilde{\mathbf{b}}), v^2\right)_{\partial\Omega} \geq \mu_0\|v\|^2$$

so that $a(\cdot,\cdot)$ is coercive:

$$a(v,v) \geq \alpha_0\|\nabla v\|^2 + \mu_0\|v\|^2 \geq \min\{\alpha_0,\mu_0\}\ \|v\|_1^2.$$

The second condition of (4.3) can be easily replaced by

$$\|\mathbf{b}-\tilde{\mathbf{b}}\|_{L^\infty(\partial\Omega)} \leq \epsilon_0, \quad 0 \leq \epsilon_0 < \frac{2\min\{\alpha_0,\mu_0\}}{C^*}$$

where C^* is the constant of trace inequality:

$$\int_{\partial\Omega} v^2 \leq C^* \int_\Omega (v^2 + |\nabla v|^2) \quad \forall\, v \in H^1(\Omega).$$

If $g \in L^2(\partial\Omega)$, $v \to (g,v)_{\partial\Omega}$ is a continuous linear form and hence $\mathcal{F}(v) = (f,v) + (g,v)_{\partial\Omega}$ is also a continuous linear form.

When the coefficients $\mathbf{b}$, $\tilde{\mathbf{b}}$ and c are all zeros, and

$$\int_\Omega f dx + \int_{\partial\Omega} g ds = 0,$$

taking the space $V = H^1(\Omega)\cap L_0^2(\Omega)$ guarantees the Poincaré inequality and we can easily show the coerciveness of $a(\cdot,\cdot)$.

c. Mixed Problem :

The Poincaré inequality is still valid in $H^1_{\Gamma_D}(\Omega)$.

Thus, if there exists a constant η such that

$$-\eta \leq -\frac{1}{2}\operatorname{div}(\mathbf{b}-\tilde{\mathbf{b}}) + c, \quad \text{a.e. } x\in\Omega \quad \text{with} \quad -\infty < \eta < \frac{\alpha_0}{C_\Omega},$$

and either

$$\mathbf{n}\cdot(\mathbf{b}-\tilde{\mathbf{b}}) \geq 0 \quad \text{a.e. } x \in \Gamma_N$$

or

$$\|\mathbf{b}-\tilde{\mathbf{b}}\|_{L^\infty(\Gamma_N)} \leq \epsilon_0, \quad 0 \leq \epsilon_0 < \frac{2\min\{\alpha_0,\mu_0\}}{C^*},$$

then $a(\cdot,\cdot)$ is clearly coercive.

If $g \in L^2(\Gamma_N)$, $v \to (g,v)_{\Gamma_N}$ is a continuous linear form and hence $\mathcal{F}(v) = (f,v) + (g,v)_{\Gamma_N}$ is also a continuous linear form.

d. Robin Problem :

If there exists a constant $\mu_0 > 0$ such that

$$0 < \mu_0 \leq -\frac{1}{2}\operatorname{div}(\mathbf{b}-\tilde{\mathbf{b}}) + c \quad \text{a.e. } x\in\Omega$$

and either

$$\kappa + \frac{1}{2}\,\mathbf{n}\cdot(\mathbf{b}-\tilde{\mathbf{b}}) \geq 0 \quad \text{a.e. } x \in \partial\Omega$$

or

$$\|\kappa + \frac{1}{2}(\mathbf{b}-\tilde{\mathbf{b}})\cdot\mathbf{n}\|_{L^\infty(\partial\Omega)} \quad \text{is small enough,}$$

then $\mathcal{A}(u,v) = a(u,v) + (\kappa u, v)_{\partial\Omega}$ is coercive:

$$\begin{aligned}\mathcal{A}(u,v) &= (A\nabla v, \nabla v) + \left(-\frac{1}{2}\operatorname{div}(\mathbf{b}-\tilde{\mathbf{b}}) + c, v^2\right) + \frac{1}{2}\left(\kappa + \mathbf{n}\cdot(\mathbf{b}-\tilde{\mathbf{b}}), v^2\right)_{\partial\Omega}\\ &\geq \alpha_0\|\nabla v\|^2 + \mu_0\|v\|^2 \geq \min\{\alpha_0,\mu_0\}\ \|v\|_1^2.\end{aligned}$$

Also, if $\kappa \in L^\infty(\partial\Omega)$, then $\mathcal{A}(u,v) = a(u,v) + (\kappa u, v)_{\partial\Omega}$ is continuous. Finally, if $g \in L^2(\partial\Omega)$, $v \to (g,v)_{\partial\Omega}$ is a continuous linear form and hence $\mathcal{F}(v) = (f,v) + (g,v)_{\partial\Omega}$ is also a continuous linear form.

(A priori estimate).

The coerciveness of $\mathcal{A}$ and continuity of $\mathcal{F}$ yields

$$\alpha\|u\|_1^2 \leq \mathcal{A}(u,u) = \mathcal{F}(u) \leq \|F\|_{V'}\|u\|_1$$

and then

$$\alpha\|u\|_1 \leq \|F\|_{V'} \leq \begin{cases} C\,\|f\|_{-1} & \text{for Dirichlet problem,}\\ C\,(\|f\| + \|g\|_{-\frac{1}{2},\partial\Omega}) & \text{for Neumann problem.}\end{cases}$$

4.3 Non-homogeneous Dirichlet Problem

Consider the following non-homogeneous Dirichlet problem:

$$\text{(D)} \qquad \begin{cases} Lu = f & \text{in } \Omega,\\ u = \varphi & \text{on } \partial\Omega.\end{cases}$$

Let $\tilde{\varphi}$ be the extension of φ in the whole Ω such that

$$\|\tilde{\varphi}\|_1 \leq C\|\varphi\|_{\frac{1}{2},\partial\Omega} \quad \text{if } \varphi \in H^{\frac{1}{2}}(\partial\Omega).$$

By the change of variable $\tilde{u} = u - \tilde{\varphi}$, the problem (D) is equivalent to

$$\text{(D}^*\text{)} \qquad \begin{cases} L\tilde{u} = f - L\tilde{\varphi} & \text{in } \Omega,\\ \tilde{u} = 0 & \text{on } \partial\Omega.\end{cases}$$

The variational problem is to

$$\text{find} \quad \tilde{u} \in H_0^1(\Omega) \quad : \quad a(\tilde{u}, v) = (f,v) - a(\tilde{\varphi}, v) \quad \forall\, v \in H_0^1(\Omega).$$

Since

$$|a(\tilde{\varphi}, v)| \le C\|\tilde{\varphi}\|_1 \|v\|_1 \le C\|\varphi\|_{\frac{1}{2},\partial\Omega}\|v\|_1 \quad \forall\, v \in H^1(\Omega),$$

$\mathcal{F}(v) := (f, v) - a(\tilde{\varphi}, v)$ is a continuous linear form on $H_0^1(\Omega)$. Thus, the a priori estimate is given by

$$\|u\|_1 \le \|\tilde{u}\|_1 + \|\tilde{\varphi}\|_1 \le C(\|f\|_{-1} + \|\varphi\|_{\frac{1}{2},\partial\Omega}).$$

4.4 Regularity of Solutions

Assuming additional regularity on the data it is possible to prove that the weak solution is indeed more regular, i.e., it belongs to $H^s(\Omega)$ for some $s > 1$ (see Grisvard(1985)).

It is worthwhile mentioning that the smoothness degree of the solution of a boundary value problem does affect the order of convergence of a numerical approximation.

(Regularity of Solution)**.**

Assume that for some $k \ge 0$, $\partial\Omega$ is a C^{k+2} manifold and the coefficients hold

$$a_{ij}, \tilde{b}_i \in C^{k+1}(\bar{\Omega}), \quad b_i, c \in C^k(\bar{\Omega}) \quad \text{and} \quad f \in H^k(\Omega).$$

Assume further that

i) $\varphi \in H^{k+\frac{3}{2}}(\partial\Omega)$ for the non-homogeneous Dirichlet problem (D),

ii) $g \in H^{k+\frac{1}{2}}(\partial\Omega)$ for the Neumann problem,

iii) $\kappa \in C^{k+1}(\partial\Omega)$ for the Robin problem.

Then the respective solution u belongs to $H^{k+2}(\Omega)$.

In particular, if all data are C^∞, then u is C^∞.

(Polygonal Domain)**.**

On the plane convex polygonal domain, the homogeneous Dirichlet problem for the Laplace operator:

$$-\Delta u = f \;\text{ in } \Omega, \quad u = 0 \;\text{ on } \partial\Omega$$

has the solution $u \in H^2(\Omega)$ if $f \in L^2(\Omega)$, the homogeneous Neumann problem for the Laplace operator:

$$-\Delta u = f \;\text{ in } \Omega, \quad \frac{\partial u}{\partial \mathbf{n}} = 0 \;\text{ on } \partial\Omega$$

has the solution $u \in H^2(\Omega)$ if $f \in L^2(\Omega)$ and the compatibility condition $\int_\Omega f = 0$ holds.

If Ω is not convex, and $\omega > \pi$ is the angle of a concave corner of $\partial\Omega$, then it turns out that $(u - s)$ is an H^2 function locally near that corner, but $u \in H^{\frac{3}{2}}(\Omega)$, where s is a suitable singular function.

On the contrary, the solution of the mixed problem in general is not regular. There exist examples in which the data and the boundary are smooth, while the solution belongs to $H^s(\Omega)$ for any $s < \frac{3}{2}$, but not to $H^{\frac{3}{2}}(\Omega)$.

(Example having corner singularity).

Let Ω be an open, bounded polygonal domain in $\mathbb{R}^2$ with one re-entrant angle. Extension to the domain with the finite number of re-entrant angles is straightforward.

Let ω be the internal angle of Ω satisfying $\pi < \omega < 2\pi$. Without the loss of generality, assume that the corresponding vertex is at the origin. Define the *singular function* s and the *dual singular function* by

$$s = r^{\frac{\pi}{\omega}} \sin \frac{\pi\theta}{\omega} \quad \text{and} \quad s_- = r^{-\frac{\pi}{\omega}} \sin \frac{\pi\theta}{\omega}$$

in the polar coordinate (r, θ) which is chosen at the origin so that the internal angle ω is spanned by the two half-lines $\theta = 0$ and $\theta = \omega$. Consider a family of cut-off functions of r, $\eta_\rho(r)$, defined as follows:

$$\eta_\rho(r) = \begin{cases} 1, & 0 < r \le \frac{\rho R}{2}, \\ \frac{15}{16}\left\{\frac{8}{15} - \left(\frac{4r}{\rho R} - 3\right) + \frac{2}{3}\left(\frac{4r}{\rho R} - 3\right)^3 - \frac{1}{5}\left(\frac{4r}{\rho R} - 3\right)^5\right\}, & \frac{\rho R}{2} < r \le \rho R, \\ 0, & r > \rho R \end{cases}$$

where ρ is a parameter in $(0, 2]$ and $R \in \mathbb{R}$ is a fixed number. It is well known that the solution has the representation of the type

$$u = w + \lambda \eta_\rho s,$$

where $w \in H^2(\Omega) \cap H_0^1(\Omega)$ is the regular part of the solution satisfying

$$-\Delta w - \lambda \Delta(\eta_\rho s) = f, \quad \text{in } \Omega$$

and $\lambda \in \mathbb{R}$ is the so-called *stress intensity factor*. Moreover, the following regularity estimate holds:

$$\|w\|_2 + |\lambda| \le C_R \|f\|$$

where C_R is a positive constant depending on the domain and the diameter of the support of η_ρ. Especially, C_R increases if the diameter of η_ρ is chosen smaller.

4.5 Galerkin Method : Finite Element Approximation

Consider the following Galerkin approximation:

$$\text{find} \quad u_h \in V_h \quad : \quad \mathcal{A}(u_h, v_h) = \mathcal{F}(v_h) \quad \forall\, v_h \in V_h \tag{4.4}$$

where V_h is a suitable finite dimensional subspace of V.

Here, the bilinear form $\mathcal{A}(\cdot,\cdot)$ is continuous and coercive, and the linear functional $\mathcal{F}(\cdot)$ is continuous.

The main point toward probing the convergence of u_h to u is to verify that

$$\lim_{h\to 0} \inf_{v_h \in V_h} \|v - v_h\|_V = 0 \quad \forall\, v \in V.$$

Proposition 4.1. *Assume there exists a subset $\mathcal{V}$ dense in V such that*

$$\inf_{v_h \in V_h} \|v - v_h\|_V \longrightarrow 0 \quad \textit{as} \quad h \longrightarrow 0 \quad \forall\, v \in \mathcal{V}. \tag{4.5}$$

Then, the Galerkin method is convergent, i.e., the solution u_h of (4.4) converges in V to the solution u of (4.2) with respect to the norm $\|\cdot\|_V$.

Proof. Since $\mathcal{V}$ is dense in V, for each $\epsilon > 0$ we can find $v \in \mathcal{V}$ such that

$$\|u - v\|_V < \epsilon.$$

Due to (4.5) there exist $h_0(\epsilon) > 0$ and, for any positive $h < h_0(\epsilon)$, $v_h \in V_h$ such that

$$\|v - v_h\| < \epsilon.$$

Hence, using the error estimate we have

$$\|u - u_h\|_V \le \frac{\gamma}{\alpha}\|u - v_h\|_V \le \frac{\gamma}{\alpha}\left(\|u - v\| + \|v - v_h\|_V\right)$$

which completes the theorem. □

Assume that $\Omega \subset \mathbb{R}^d$, $d = 2, 3$ is a polygonal domain with Lipschitz boundary and $\mathcal{T}_h$ is a regular family of triangulations of Ω.

The finite dimensional subspace V_h is one of the followings:

i) $V_h = X_h^k \cap H_0^1(\Omega)$ $(k \ge 1)$ for the Dirichlet problem

ii) $V_h = X_h^k$ $(k \ge 1)$ for the Neumann problem

iii) $V_h = X_h^k \cap H^1_{\Gamma_D}(\Omega)$ $(k \ge 1)$ for the Mixed problem

iv) $V_h = X_h^k$ $(k \ge 1)$ for the Robin problem.

Theorem 4.2 (H^1 error estimate)**.**

If the exact solution $u \in H^s(\Omega)$ for some $s \geq 2$, the following error estimate holds

$$\|u - u_h\|_1 \leq C\, h^\ell \, \|u\|_{\ell+1} \quad \textit{where } \ell = \min(k, s-1). \tag{4.6}$$

Proof. Since $C^\infty(\bar{\Omega})$ is dense in $H^1(\Omega)$, we can choose $\mathcal{V} = C^\infty(\bar{\Omega})$ for both Neumann and Robin problem, $\mathcal{V} = C^\infty(\bar{\Omega}) \cap H^1_0(\Omega)$ for the Dirichlet problem and $\mathcal{V} = C^\infty(\bar{\Omega}) \cap H^1_{G_D}(\Omega)$ for the mixed problem. Furthermore, for each $v \in \mathcal{V}$

$$\inf_{v_h \in V_h} \|v - v_h\|_1 \leq \|v - \pi_h^k(v)\|_1 \leq C h^\ell |v|_{\ell+1},$$

hence it converges to zero.

Since $u \in H^s(\Omega)$, $s \geq 2$, $u \in C^0(\bar{\Omega})$ and hence $\pi_h^k(u) \in V_h$ holds the respective boundary conditions. Then, using the interpolation error

$$\|u - \pi_h^k(u)\|_1 \leq C h^\ell \, \|u\|_{\ell+1}$$

and the Céa lemma

$$\|u - u_h\|_1 \leq \frac{\gamma}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|_1,$$

we have the conclusion. □

The convergence result (4.6) is *optimal* in the $H^1(\Omega)$-norm, i.e., it provides the highest possible rate of convergence in the $H^1(\Omega)$-norm allowed by the polynomial degree k.

(Adjoint Problem)**.** Consider the following *adjoint problem*: given $r \in L^2(\Omega)$,

$$\text{find} \quad \phi(r) \in V \quad : \quad \mathcal{A}(v, \phi(r)) = (r, v) \quad \forall\, v \in V. \tag{4.7}$$

The solution to (4.7) enjoys the same regularity property than the one of the original problem (4.2).

In particular, if Ω is a polygonal domain the solution $\phi(r)$ belongs to $H^2(\Omega)$ and holds

$$\|\phi(r)\|_2 \leq C \|r\|_0 \quad \forall\, r \in L^2(\Omega), \tag{4.8}$$

provided that Ω is convex, $a_{ij} \in C^1(\bar{\Omega})$, and $\kappa \in C^1(\partial\Omega)$ (see Grisvard).

This is true for all but the mixed boundary value problem: the solution of the mixed problem belongs to $H^2(\Omega)$ for any $s < 3/2$ but in general not to $H^{3/2}(\Omega)$, even for smooth data.

Theorem 4.3 (L^2-error estimate)**.** *Assume that it holds (4.8).*

If the exact solution $u \in H^s(\Omega)$ for some $s \geq 2$, then the following estimate holds

$$\|u - u_h\| \leq C\, h^{\ell+1}\, \|u\|_{\ell+1} \quad \textit{where } \ell = \min(k, s-1). \tag{4.9}$$

Proof. Let $r = u - u_h$. For any $w_h \in V_h$ we have

$$\begin{aligned}\|u - u_h\|^2 &= (r, u - u_h) = \mathcal{A}(u - u_h, \phi(r)) = \mathcal{A}(u - u_h, \phi(r) - w_h) \\ &\leq \gamma \|u - u_h\|_1\, \|\phi(r) - w_h\|_1.\end{aligned}$$

Taking $w_h = \pi_h^k(\phi(r))$, we have from the interpolation error and (4.8) that

$$\begin{aligned}\|u - u_h\|^2 &\leq \gamma \|u - u_h\|_1\, \|\phi(r) - \pi_h^k(\phi(r))\|_1 \\ &\leq C\gamma \|u - u_h\|_1\, h\, \|\phi(r)\|_2 \leq C\, h\, \|u - u_h\|_1\, \|r\| \\ &= C\, h\, \|u - u_h\|_1\, \|u - u_h\|.\end{aligned}$$

Thus, we have the conclusion. □

(L^∞-error estimate)**.** We have the following error estimate in $L^\infty(\Omega)$:

$$\|u - u_h\|_\infty \leq C\, h^{\ell+1-\frac{d}{2}}\, |u|_{\ell+1} \quad \forall\, u \in H^{\ell+1}(\Omega).$$

(The non-homogeneous Dirichlet problem)**.**

The variational problem is to

$$\text{find} \quad \tilde{u} \in H_0^1(\Omega) \quad : \quad a(\tilde{u}, v) = (f, v) - a(\tilde{\varphi}, v) \quad \forall\, v \in H_0^1(\Omega).$$

Let $V_h = X_h^k \cap H_0^1(\Omega)$. Then, the finite element approximation is to

$$\text{find} \quad \tilde{u}_h \in V_h \quad : \quad a(\tilde{u}_h, v_h) = (f, v_h) - a(\tilde{\varphi}, v_h) \quad \forall\, v_h \in V_h. \tag{4.10}$$

Here, the construction of the extension operator $\varphi \to \tilde{\varphi}$ is not easily performed.

Assuming that $\varphi \in H^{1/2}(\partial\Omega) \cap C^0(\partial\Omega)$, we can get an alternative approach.

Denote by $\{x_s \,:\, s = 1, \cdots, M_h\}$ the nodes on $\partial\Omega$ and $\{a_i \,:\, i = 1, \cdots, N_h\}$ the internal nodes. Set

$$V_h^* := \{v_h \in X_h^k \,:\, v_h(x_s) = \varphi(x_s) \quad s = 1, \cdots, M_h\}.$$

The approximate problem reads:

$$\text{find} \quad u_h \in V_h^* \quad : \quad a(u_h, v_h) = (f, v_h) \quad \forall\, v_h \in V_h. \tag{4.11}$$

Any $u_h \in V_h^*$ can be written by

$$u_h = \sum_{i=1}^{N_h} u_h(a_i)\varphi_i + \sum_{s=1}^{M_h} \varphi(x_s)\tilde{\varphi}_s := z_h + \tilde{\varphi}_h,$$

where φ_i and $\tilde{\varphi}_i$ are the basis functions of X_h^k relative to the internal and boundary nodes, respectively.

Then we have the new discrete problem for non-homogeneous Dirichlet problem:

$$\text{find} \quad z_h \in V_h \quad : \quad a(z_h, v_h) = (f, v_h) - a(\tilde{\varphi}_h, v_h) \quad \forall\, v_h \in V_h.$$

Using the orthogonality

$$a(u - u_h, v_h) = 0 \quad \forall\, v_h \in V_h$$

and the Céa lemma, we have the error estimate that if $u \in H^s(\Omega)$ $s \geq 2$,

$$\|u - u_h\|_1 = O(\,h^{\ell}\,), \quad \ell = \min(k, s-1).$$

4.6 Non-coercive Variational Problem

Consider the following elliptic problem:

$$Lu = -\nabla \cdot A\nabla u + \mathbf{b} \cdot \nabla u + c\,u. \tag{4.12}$$

Define

$$a(u,v) = (A\nabla u, \nabla v) + (\mathbf{b} \cdot \nabla u, v) + (cu, v).$$

When the bilinear form $a(\cdot,\cdot)$ is not coercive, we need the following theorem.

Theorem 4.4 (Gårding Inequality)**.**

Suppose that the coefficient matrix A is uniformly bounded, i.e.,

$$\alpha\,|\boldsymbol{\xi}|^2 \leq \boldsymbol{\xi}^T A\,\boldsymbol{\xi} \quad \forall\,\boldsymbol{\xi} \in \mathbb{R}^d, \quad a.e. \quad x \in \Omega,$$

and the coefficient $\mathbf{b} \in L^\infty(\Omega)^d$. Then there is a constant $K < \infty$ such that

$$\frac{\alpha}{2}\,\|v\|_1^2 \leq a(v,v) + K\|v\|^2 \quad \forall\, v \in H^1(\Omega). \tag{4.13}$$

Proof. By Hölder inequality

$$\begin{aligned}
|(\mathbf{b}\cdot\nabla v, v)| &= \left| \int_\Omega \sum_{k=1}^d b_k(x)\partial_{x_k} v(x)\ v(x)\,dx \right| \\
&\leq \int_\Omega \sum_{k=1}^d |b_k(x)|\ |\partial_{x_k} v(x)|\ \ |v(x)|\,dx \leq \sum_{k=1}^d \|b_k\|_\infty \int_\Omega |\partial_{x_k} v(x)|\ \ |v(x)|\,dx \\
&\leq \sum_{k=1}^d \|b_k\|_\infty\,\|\partial_{x_k} v\|\,\|v\| \leq B\,|v|_1\,\|v\|,
\end{aligned}$$

where

$$B^2 := \sum_{k=1}^{d} \|b_k\|_\infty^2.$$

Now we have

$$\begin{aligned} a(v,v) + K\|v\|^2 &\geq \alpha|v|_1^2 + (\mathbf{b}\cdot\nabla v, v) + (c+K, v^2) \\ &\geq \alpha|v|_1^2 - B\,|v|_1\,\|v\| + (\beta+K)\|v\|^2, \end{aligned}$$

where

$$\beta := \operatorname{ess\,inf}\{c(x) \,:\, x\in\Omega\}.$$

From the arithmetic-geometric mean inequality, we have

$$a(v,v) + K\|v\|^2 \geq \frac{\alpha}{2}\left(|v|_1^2 + \|v\|_1^2\right),$$

provided

$$K \geq \frac{\alpha}{2} + \frac{B^2}{2\alpha} - \beta.$$

Note that K need not be positive, if $\beta > 0$. □

Assume that

(a) $a(\cdot,\cdot)$ is continuous on $H^1(\Omega)$:

$$|a(u,v)| \leq C_1\|u\|_1\,\|v\|_1 \quad \forall\, u,v\in H^1(\Omega)$$

(b) there exists a constant $K\in\mathbb{R}$ satisfying the Gårding inequality

$$a(v,v) + K\|v\|^2 \geq \alpha\,\|v\|_1^2 \quad \forall\, v\in H^1(\Omega)$$

(c) there is some $V\subset H^1(\Omega)$ such that there is a unique solution u, to the variational problem

$$a(u,v) = (f,v) \quad \forall\, v\in V$$

as well as to the adjoint variational problem

$$a(v,u) = (f,v) \quad \forall\, v\in V$$

(d) in both cases, the regularity estimate holds: for all $f\in L^2(\Omega)$

$$|u|_2 \leq C_R\,\|f\|_2.$$

Let V_h be a finite element subspace of V which satisfies

(e) $$\inf_{v\in V_h}\|u-v\|_1 \leq C_A\,h\,|u|_2 \quad \forall\, u\in H^2(\Omega).$$

Consider the following variational problem:

(4.14) $$\text{find}\quad u_h\in V_h \quad : \quad a(u_h,v) = (f,v) \quad \forall\, v\in V_h.$$

Theorem 4.5. *Under the conditions (a), (b), (c), (d) and (e), there are constants $h_0 > 0$ and $C > 0$ such that for all $h \le h_0$, there is a unique solution to (4.14) satisfying*

$$\|u - u_h\|_1 \le C \inf_{v \in V_h} \|u - v\|_1 \le C\, h\, |u|_2, \tag{4.15}$$

and

$$\|u - u_h\| \le C\, h\, \|u - u_h\|_1 \le C\, h^2\, |u|_2. \tag{4.16}$$

In particular, we may take

$$h_0 = \left(\frac{\alpha}{2K}\right)^{\frac{1}{2}} \frac{1}{C_1 C_A C_R}.$$

Proof. We begin by deriving an estimate for any solution to (4.14) that may exists.

Using (a), (b) and the orthogonality

$$a(u - u_h, v) = 0 \quad \forall\, v \in V_h,$$

we have that for any $v \in V_h$,

$$\begin{aligned}
\alpha \|u - u_h\|_1^2 &\le a(u - u_h, u - u_h) + K(u - u_h, u - u_h) \\
&= a(u - u_h, u - v) + K\|u - u_h\|_2^2 \\
&\le C_1 \|u - u_h\|_1\, \|u - v\|_1 + K\|u - u_h\|^2.
\end{aligned}$$

We apply standard duality techniques to bound $\|u - u_h\|$. Let w be the solution to the adjoint problem satisfying (d):

$$a(v, w) = (u - u_h, v) \quad \forall\, v \in V.$$

Then, for any $w_h \in V_h$,

$$\begin{aligned}
(u - u_h, u - u_h) &= a(u - u_h, w) = a(u - u_h, w - w_h) \\
&\le C_1 \|u - u_h\|_1\, \|w - w_h\|_1.
\end{aligned}$$

Using (d) and (e) yields

$$(u - u_h, u - u_h) \le C_1 C_A\, h\, \|u - u_h\|_1\, |w|_2 \le C_1 C_A C_R\, h\, \|u - u_h\|_1\, \|u - u_h\|_2.$$

Therefore

$$\|u - u_h\| \le C_1 C_A C_R\, h\, \|u - u_h\|_1.$$

Now, we have

$$\alpha \|u - u_h\|_1^2 \le C_1 \|u - u_h\|_1\, \|u - v\|_1 + K(C_1 C_A C_R)^2\, h^2\, \|u - u_h\|_1^2.$$

Thus, for $h \le h_0$ where

$$h_0 = \left(\frac{\alpha}{2K}\right)^{\frac{1}{2}} \frac{1}{C_1 C_A C_R},$$

we obtain

$$\alpha \|u - u_h\|_1^2 \le 2C_1 \, \|u - v\|_1 \quad \forall\, v \in V_h. \tag{4.17}$$

This yields (4.15) and (4.16) follows (4.15).

So far, we have been operating under the assumption of the existence of a solution u_h. Since (4.14) is a finite dimensional system having the same number of unknowns as equations, uniqueness implies existence.

Set $f \equiv 0$. Then, $u \equiv 0$ from (d). Also, (4.17) implies that $u_h \equiv 0$ as well, provided h is sufficiently small. Hence, the problem (4.14) has unique solution for h sufficiently small, since $f \equiv 0$ implies $u_h \equiv 0$. □

4.7 Generalized Galerkin Method

Consider the generalized Galerkin method

$$\text{find} \quad u_h \in V_h \quad : \quad \mathcal{A}_h(u_h, v_h) = \mathcal{F}_h(v_h) \quad \forall\, v_h \in V_h \tag{4.18}$$

associating to the homogeneous Dirichlet problem:

$$Lu = -\nabla \cdot A\nabla u = f \quad \text{in} \quad \Omega.$$

Let $V = H_0^1(\Omega)$ and $V_h = X_h^k \cap H_0^1(\Omega)$, $k = 1, 2, 3$.

Consider the numerical integration:

$$\int_\Omega \varphi \, dx \simeq \sum_{K \in \mathcal{T}_h} \sum_{j=1}^{M} \int_K w_{j,K} \, \varphi(b_{j,k}) \, dx =: Q_h(\varphi),$$

where the weights $w_{j,K}$ and the nodes $b_{j,K}$ are derived from a quadrature formula defined on the reference element $\hat{K}$. More precisely, these weights and nodes are defined as

$$w_{j,K} = |\det B_K| \hat{w}_j, \quad b_{j,K} = T_K(\hat{b}_j),$$

where $\hat{w}_j$ and $\hat{b}_j$ are the weights and nodes of the quadrature formula chosen on $\hat{K}$, and $T_K(\hat{\mathbf{x}}) = B_K \hat{x} + b_K$ is the affine map from $\hat{K}$ onto K.

Note that, with $\varphi(x) = \hat{\varphi}(\hat{x})$ for all $x = T_K(\hat{x})$, $\hat{x} \in \hat{K}$,

$$\int_K \varphi(x) \, dx = |\det(B_K)| \int_{\hat{K}} \hat{\varphi}(\hat{x}) \, d\hat{x}.$$

Define

$$\mathcal{F}_h(v_h) = Q_h(fv_h), \quad \mathcal{A}_h(u_h, v_h) = Q_h(A\nabla u_h \cdot \nabla v_h).$$

Due to the presence of pointvalues, both $\mathcal{A}_h$ and $\mathcal{F}_h$ should now be defined on V_h and not on V. We need to assume that the coefficients of both the operator L and the right hand side f are continuous functions on $\bar{\Omega}$.

Let $\hat{K}$ be an d-simplex. Denote by

$$\begin{aligned}
\hat{b}_i \quad &: \text{the vertices for } i = 1, \cdot, d+1, \\
\hat{b}_{ij} \quad &: \text{the midpoints of each side for } i, j = 1, \cdots, d+1, \\
\hat{b}_0 \quad &: \text{the center of gravity of } \hat{K}, \textit{ i.e.}, \hat{b}_0 := \tfrac{1}{n+1} \textstyle\sum_{j=1}^{d+1} \hat{b}_j.
\end{aligned}$$

The following numerical quadrature formulae are exact on $\mathbb{P}_k$:

$$\begin{aligned}
&\int_{\hat{K}} \hat{\varphi}\, dx \simeq \operatorname{meas}(\hat{K})\, \hat{\varphi}(\hat{b}_0), \quad k = 1 \\
&\int_{\hat{K}} \hat{\varphi}\, dx \simeq \frac{1}{3} \operatorname{meas}(\hat{K}) \sum_{1 \le i < j \le 3} \hat{\varphi}(\hat{b}_{ij}), \quad k = 2, \quad d = 2, \\
&\int_{\hat{K}} \hat{\varphi}\, dx \simeq \frac{1}{60} \operatorname{meas}(\hat{K}) \left[3 \sum_{i=1}^{3} \hat{\varphi}(\hat{b}_i) + 8 \sum_{1 \le i < j \le 3} \hat{\varphi}(\hat{b}_{ij}) + 27 \varphi(\hat{b}_0) \right], \quad k = 3 \quad d = 2.
\end{aligned}$$

To check these, let $\hat{\lambda}_i(x)$, $1 \le i \le d+1$, denote the barycentric coordinates of a point x with respect to the vertices of the n-simplex $\hat{K}$.

Then, for any integers $\alpha_i \ge 0$ $(1 \le i \le d+1)$, one has (show exercise 4.1.1 in Ciarlet)

$$\int_{\hat{K}} \hat{\lambda}_1(\hat{x})^{\alpha_1} \cdots \hat{\lambda}_{d+1}(\hat{x})^{\alpha_{d+1}}\, d\hat{x} = \frac{\alpha_1! \cdots \alpha_{d+1}!\, d!}{(\alpha_1 + \cdots + \alpha_{d+1} + d)!} \operatorname{meas}(\hat{K}). \tag{4.19}$$

To show the first formula, let

$$\hat{\varphi} = \sum_{j=1}^{d+1} \hat{\varphi}(\hat{b}_j) \hat{\lambda}_j$$

be any polynomial of degree ≤ 1. Then, we have

$$\int_{\hat{K}} \hat{\varphi}(\hat{x})\, d\hat{x} = \frac{\operatorname{meas}(\hat{K})}{d+1} \sum_{j=1}^{d+1} \hat{\varphi}(\hat{b}_j) = \operatorname{meas}(\hat{K})\, \varphi(\hat{b}_0).$$

The uniform coerciveness of $\mathcal{A}_h$ using the above three quadrature formulae is given in Cíarlet, pp. 193–196. Under the assumption that $\mathcal{A}_h$ is uniformly coercive in $V_h \times V_h$, the error estimate is given in Theorem 3.7.

To check the consistency errors define the quadrature error

$$E_K(\varphi) := \int_K \varphi\, dx - \sum_{j=1}^{L} w_{j,K}\, \varphi(b_{j,K}).$$

The following theorem is given in Theorem 4.1.4 and 4.1.5 of Ciarlet:

Proposition 4.6. *Assume that the quadrature formula on $\hat{K}$ is exact on $\mathbb{P}_{2k-2}$. Then there exists a constant C, independent of h, such that for all $K \in \mathcal{T}_h$*

$$|E_K(aD_ipD_jq)| \le C\, h_K^k \, \|a\|_{W^{k,\infty}(K)} \, \|p\|_{k,K} \, |q|_{1,K}$$

for any $a \in W^{k,\infty}(K)$, $p, q \in \mathbb{P}_k$, $i, j = 1, \cdots, d$, and

$$|E_K(fp)| \le C\, h_K^k \, [meas(K)]^{\frac{1}{2}} \, \|f\|_{W^{k,\infty}(K)} \, \|p\|_{1,K}$$

for any $f \in W^{k,\infty}(K)$, $p \in \mathbb{P}_k$.

Theorem 4.7. *Let $\mathcal{T}_h$ be a regular family of triangulations.*
Assume that the quadrature formula on $\hat{K}$ is exact on $\mathbb{P}_{2k-2}$ and that its weights $\hat{w}_j$ are positive. If the solution $u \in H^{k+1}(\Omega)$, the coefficients $a_{ij} \in W^{k,\infty}(\Omega)$ and the datum $f \in W^{k,\infty}(\Omega)$, then there exists a constant C independent of h such that

$$\|u - u_h\|_1 \le C\, h^k \left[|u|_{k+1} + \|u\|_{k+1} \sum_{i,j=1}^{d} \|a_{ij}\|_{W^{k,\infty}(\Omega)} + \|f\|_{W^{k,\infty}(\Omega)} \right].$$

Proof. Recall the first Strang Lemma:

$$\|u - u_h\|_V \le \inf_{w_h \in V_h} \left[\left(1 + \frac{\gamma}{\alpha^*}\right) \|u - w_h\|_V + \frac{1}{\alpha^*} \sup_{0 \ne v_h \in V_h} \frac{|\mathcal{A}(w_h, v_h) - \mathcal{A}_h(w_h, v_h)|}{\|v_h\|_V} \right] + \frac{1}{\alpha^*} \sup_{0 \ne v_h \in V_h} \frac{|\mathcal{F}(v_h) - \mathcal{F}_h(v_h)|}{\|v_h\|_V}.$$

Note that the approximation error:

$$\inf_{w_h \in V_h} \|u - u_h\|_1 \le C h^k |u|_{k+1}.$$

Using the previous proposition, we have

$$\begin{aligned}
|\mathcal{A}(\pi_h^k(u), v_h) - \mathcal{A}_h(\pi_h^k(u), v_h)| &\le \sum_{K \in \mathcal{T}_h} \sum_{i,j=1}^{d} \left| E_K \left(a_{ij} D_j \pi_h^k(u) D_i v_h \right) \right| \\
&\le C \sum_{K \in \mathcal{T}_h} \sum_{i,j=1}^{d} \left(h_K^k \|a_{ij}\|_{W^{k,\infty}(K)} \|\pi_h^k(u)\|_{k,K} |v_h|_{1,K} \right) \\
&\le C h^k \left(\sum_{i,j=1}^{d} \|a_{ij}\|_{W^{k,\infty}(\Omega)} \right) \left(\sum_{K \in \mathcal{T}_h} \|\pi_h^k(u)\|_{k,K} |v_h|_{1,K} \right) \\
&\le C h^k \left(\sum_{i,j=1}^{d} \|a_{ij}\|_{W^{k,\infty}(\Omega)} \right) \left(\sum_{K \in \mathcal{T}_h} \|\pi_h^k(u)\|_{k,K}^2 \right)^{\frac{1}{2}} |v_h|_{1,\Omega}.
\end{aligned}$$

In addition, writing $\pi_h^k(u)$ as $\pi_h^k(u) - u + u$, we obtain

$$\begin{aligned}\sum_{K\in\mathcal{T}_h} \|\pi_h^k(u)\|_{k,K}^2 &\leq 2\sum_{k\in\mathcal{T}_h}\left(\|u\|_{k,K}^2 + \|u-\pi_h^k(u)\|_{k,K}^2\right)\\ &\leq 2\left(\|u\|_{k,\Omega}^2 + \sum_{k\in\mathcal{T}_h}\|u-\pi_h^k(u)\|_{k,K}^2\right)\\ &\leq C\left(\|u\|_k^2 + h^2\,|u|_{k+1,\Omega}^2\right) \leq C\|u\|_{k+1}^2.\end{aligned}$$

Also we have

$$|\mathcal{F}(v_h) - \mathcal{F}_h(v_h)| \leq \sum_{k\in\mathcal{T}_h} |E_K(fv_h)| \leq C\,h^k\,[\mathrm{meas}(\Omega)]^{\frac{1}{2}}\,\|f\|_{W^{k,\infty}(\Omega)}\,\|v_h\|_1.$$

Thus, by the first Strang Lemma 3.7 we obtain the conclusion. □

(Example on P_1 finite element space, i.e., $k = 1$).

Let $\mathcal{T}_h$ be a regular family of triangulations.

The three points quadrature formula on $\hat{K}$ is exact on $\mathbb{P}_2$. If the solution $u \in H^2(\Omega)$, the coefficients $a_{ij} \in W^{1,\infty}(\Omega)$ and the datum $f \in W^{1,\infty}(\Omega)$, then there exists a constant C independent of h such that

$$\|u-u_h\|_1 \leq C\,h\left[|u|_2 + \|u\|_2\sum_{i,j=1}^{d}\|a_{ij}\|_{W^{1,\infty}(\Omega)} + \|f\|_{W^{1,\infty}(\Omega)}\right].$$

4.8 Condition number of Stiffness matrix and Inverse inequality

Let A_{fe} be the stiffness matrix given by

$$A_{fe} \quad : \quad A_{fe}(i,j) = \mathcal{A}(\varphi_j, \varphi_i)$$

where φ_j are basis functions of $V_h \subset X_h^k$.

Recall that the condition number $\chi(B)$ of a non-singular matrix B is given by

$$\chi(B) := \|B\|\,\|B^{-1}\|,$$

where $\|\cdot\|$ is a suitable matrix-norm.

When B is symmetric, we have $\|B\|_2 = \rho(B)$ where $\rho(B)$ denotes the spectral radius of B. In addition, if B is positive definite, then

$$\chi_2(B) = \chi_{sp}(B) := \frac{\lambda_{\max}(B)}{\lambda_{\min}(B)}.$$

In this subsection, we will show that

$$\chi_{sp}(A_{fe}) = O(h^{-2}).$$

Proposition 4.8 ($\boldsymbol{\eta}^T M \boldsymbol{\eta} \sim h^d \boldsymbol{\eta}^T \boldsymbol{\eta}$)**.**

Let $\mathcal{T}_h$ be a quasi-uniform family of triangulations of $\bar{\Omega}$. Then there exist positive constants C_1 and C_2 such that for each $v_h \in V_h$,

$$C_1\, h^d\, |\boldsymbol{\eta}|^2 \le \|v_h\|^2 \le C_2\, h^d\, |\boldsymbol{\eta}|^2, \quad v_h = \sum_{j=1}^{N_h} \eta_j \varphi_j.$$

Proof. Since $\mathcal{T}_h$ is quasi-uniform, it is enough to show that for any element K

$$C_1^*\, h_K^d \sum_{j=1}^{M} \eta_j^2 \le \int_K v_h^2 dx \le C_2^* h_K^d \sum_{j=1}^{M} \eta_j^2,$$

where M is the number of degrees of freedom associated with K.

For the reference element $\hat{K}$, set $\hat{v} = v_h \circ T_K$. Then, $\hat{v} = \sum_j \eta_j \varphi_j$. Note from (4.19) that

$$\begin{aligned}\int_{\hat{K}} \hat{v}^2 &= \sum_j \eta_j^2 \int_{\hat{K}} \hat{\varphi}_j^2 + 2 \sum_{i<j} \eta_i \eta_j \int_{\hat{K}} \hat{\varphi}_i \hat{\varphi}_j \\ &= \frac{d!\,\mathrm{meas}(\hat{K})}{(2+d)!} \left(2 \sum_j \eta_j^2 + 2 \sum_{i<j} \eta_i \eta_j \right) = \frac{d!\,\mathrm{meas}(\hat{K})}{(2+d)!} \left(\sum_j \eta_j^2 + \left(\sum_j \eta_j \right)^2 \right).\end{aligned}$$

Thus we have

$$C_1^{**} \sum_{j=1}^{M} \eta_j^2 \le \int_{\hat{K}} \hat{v}^2 d\hat{x} \le C_2^{**} \sum_{j=1}^{M} \eta_j^2.$$

Using the last inequality together with the fact that

$$c\, h_K^d \le |\det B_K| = \frac{\mathrm{meas}(K)}{\mathrm{meas}(\hat{K})} \le C\, h_K^d$$

and

$$\int_K v_h^2 = \int_K (\hat{v} \circ T_K^{-1})^2 = |\det B_K| \int_{\hat{K}} \hat{v}^2,$$

we have the conclusion. □

Proposition 4.9 (Inverse Inequality for piecewise polynomials)**.**

Let $\mathcal{T}_h$ be a quasi-uniform family of triangulations of $\bar{\Omega}$. Then there exist positive constants C_3 such that for each $v_h \in V_h$,

$$\|\nabla v_h\|^2 \le C_3\, h^{-2}\, \|v_h\|^2.$$

Proof. Since $\mathbb{P}_k(\hat{K})$ is finite dimensional, we have by the equivalence of norms that

$$\|\nabla \hat{v}\|_{0,\hat{K}}^2 \le \|\hat{v}\|_{1,\hat{K}}^2 \le C\, \|\hat{v}\|_{0,\hat{K}}^2 \quad \forall\, \hat{v} \in \mathbb{P}_k(\hat{K}).$$

With $\hat{v} = v_h \circ T_K$, we have from Proposition 2.6 and 2.7 that

$$\begin{aligned} \|\nabla v_h\|_{0,K}^2 &\le C\|B_K^{-1}\|^2 \cdot |\det B_K| \cdot \|\nabla \hat{v}\|_{0,\hat{K}}^2 \le C\|B_K^{-1}\|^2 \cdot |\det B_K| \cdot \|\hat{v}\|_{0,\hat{K}}^2 \\ &\le C\|B_K^{-1}\|^2 \cdot \|v_h\|_{0,K}^2 \le \frac{C}{(\rho_K)^2}\,\|v_h\|_{0,K}^2. \end{aligned}$$

Since $\mathcal{T}_h$ is a quasi-uniform family of triangulations, we have the conclusion. □

Writing $v_h = \sum_j \eta_j \varphi_j$, we have

$$\frac{(A_{fe}\boldsymbol{\eta}, \boldsymbol{\eta})}{|\boldsymbol{\eta}|^2} = \frac{\mathcal{A}(v_h, v_h)}{|\boldsymbol{\eta}|^2}.$$

Since $\mathcal{A}(\cdot,\cdot)$ is continuous and coercive,

$$\alpha\, C_1\, h^d \le \frac{(A_{fe}\boldsymbol{\eta}, \boldsymbol{\eta})}{|\boldsymbol{\eta}|^2} \le \gamma\, C_2\, h^d\,(1 + C_3\, h^{-2}).$$

Assume that A_{fe} is symmetric and positive definite. Then, we have

$$\alpha\, C_1\, h^d \le \lambda \le \gamma\, C_2\, h^d\,(1 + C_3\, h^{-2}), \quad \text{for any eigenvalue } \lambda \text{ of } A_{fe}.$$

Hence, the condition number of A_{fe} has the following bounds:

$$\chi_{sp}(A_{fe}) = \frac{\lambda_{\max}(A_{fe})}{\lambda_{\min}(A_{fe})} \le \frac{\gamma C_2}{\alpha C_1}\,(1 + C_3 h^{-2}) = O(h^{-2}).$$

The spectrum of $\mathcal{A}(\cdot,\cdot)$ is defined as the set of $\mu \in \mathbb{R}$ such that there exists an eigenfunction $\omega_h \in V_h$, $\omega_h \neq 0$, satisfying

$$\mathcal{A}(\omega_h, v_h) = \mu\,(\omega_h, v_h) \quad \forall\, v_h \in V_h.$$

Thus, if (μ, ω_h) is an eigenpair, then we have

$$\alpha \le \frac{\mathcal{A}(\omega_h, \omega_h)}{\|\omega_h\|^2} = \mu \le \gamma\,\frac{\|\omega_h\|_1^2}{\|\omega\|^2} \le \gamma\,(1 + C_3\, h^{-2}).$$

The convergence properties of Conjugate Gradient iteration is reflected by the estimate

$$|e^k|_A \le 2\left(\frac{\sqrt{\chi_{sp}(A)} - 1}{\sqrt{\chi_{sp}(A)} + 1}\right)^k |e^0|_A,$$

where e^k denotes the error of k-iterations and $|e|_A^2 = (Ae,\, e)$ denotes the vector norm.

4.9 Finite Elements with Interpolated Boundary Conditions

Let $\Omega \subset \mathbb{R}^2$ be a bounded domain with smooth boundary, and $\mathcal{T}_h$ be a triangulation of Ω, where each triangle at the boundary has at most one curved side.

Assume that there exists $\rho > 0$ such that for each triangle $K \in \mathcal{T}_h$ we can find two concentric circular discs D_1 and D_2 such that

$$D_1 \subset K \subset D_2 \quad \text{and} \quad \frac{\text{diam}D_2}{\text{diam}D_1} \leq \rho.$$

Using the proof of Lemma 4.5.3 in [Brenner and Scott] that

$$\|\phi\|_{W_\infty^{k-1}(D_2)} \leq C_{k,\rho} \, (\text{diam}D_2)^{1-k} \, \|\phi\|_{H^1(D_1)} \quad \forall \, \phi \in \mathbb{P}_{k-1}. \tag{4.20}$$

We consider the Lobatto quadrature formula. Let the polynomial $L_k(\xi)$ of degree k be defined by

$$L_k(x) = \left(\frac{d}{dx}\right)^{k-2} \big(x(1-x)\big)^{k-1}.$$

$L_k(\xi)$ has k distinct roots $0 = \xi_0 < \xi_1 < \cdots < \xi_{k-1} = 1$.

For each j $(0 \leq j \leq k-1)$, let P_j be the Lagrange interpolating polynomial of degree $k-1$ such that $P_j(\xi_i) = \delta_{ij}$, and let

$$w_j = \int_0^1 P_j(x)dx.$$

Lemma 4.10. *We have*

$$\int_0^1 P(x)dx = \sum_{j=0}^{k-1} w_j \, P(\xi_j) \quad \forall \, P \in \mathbb{P}_{2k-3}.$$

Corollary 4.11. *We have*

$$\left| \int_0^h f(x)dx - h \sum_{j=0}^{k-1} w_j \, f(h\xi_j) \right| \leq C_k \, h^{2k-1} \, \|f^{(2k-2)}(x)\|_{L^\infty(0,h)} \quad \forall \, f \in C^{2k-2}([0,h]).$$

For each boundary edge

$$e = \{x(s) \, : \, s \in [s_e, s_e + h_e], \, s \text{ is arc elngth}\},$$

let the boundary nodes be $x(s_e + h_e\xi_j)$, $j = 0, \cdots, k-1$.

Define the finite element space V_h:

$$V_h = \{v \in C^0(\bar{\Omega}) \, : \, v|_K \in \mathbb{P}_{k-1} \, \text{and } v \text{ vanishes at the boundary nodes}\}.$$

Let

$$a(u,v) = (\nabla u, \nabla v) \quad \text{for} \quad u, v \in V := H_0^1(\Omega).$$

We consider the variational problem

$$\text{find} \quad u \in V \quad : \quad a(u,v) = (f,v) \quad \forall\, v \in V,$$

and the discrete variational problem

$$\text{find} \quad u_h \in V_h \quad : \quad a(u_h,v) = (f,v) \quad \forall\, v \in V_h.$$

By Green's Theorem, we have that for any $w \in V_h$,

$$\begin{aligned} a(u-u_h,w) = a(u,w) - (f,w) &= (-\Delta u, w) + \int_{\partial\Omega} (\mathbf{n}\cdot\nabla u)\, w\, ds - (f,w) \\ &= \int_{\partial\Omega} (\mathbf{n}\cdot\nabla u)\, w\, ds. \end{aligned}$$

Lemma 4.12. *Let K be a triangle with a curved edge e. Assume that $u \in W_\infty^{2k-1}(K)$, and $w \in \mathbb{P}_{k-1}$ vanishes at the Lobatto nodes along e. Then we have*

$$\Big| \int_e (\mathbf{n}\cdot\nabla u)\, w\, ds \Big| \le C_{k,\rho}\, h_e^{2k-1}\, (diam D_2)^{1-k}\, \|u\|_{W_\infty^{2k-1}(K)} \cdot \|w\|_{H^1(K)},$$

where h_e = length of e.

Proof. Let s denote arc length. Using parameterization $x(s)$, $0 \le s \le h_e$, and (4.20) yields

$$\begin{aligned} \left| \int_e (\mathbf{n}\cdot\nabla u)\, w\, ds \right| &= \left| \int_0^{h_e} \big(\mathbf{n}\cdot\nabla u\big)(x(s))\, w(x(s))\, ds \right| \\ &\le C_k\, h_e^{2k-1}\, \|\mathbf{n}\cdot\nabla u\|_{W_\infty^{2k-2}(K)}\, \|w\|_{W_\infty^{k-1}(K)} \\ &\le C_{k,\rho}\, h_e^{2k-1}\, (\mathrm{diam} D_2)^{1-k}\, \|u\|_{W_\infty^{2k-1}(K)}\, \|w\|_{H^1(K)}. \end{aligned}$$

□

Lemma 4.13. *Assume that $u \in W_\infty^{2k-1}(\Omega)$. For small h and fixed k, we have*

$$\sup_{w \in V_h \setminus \{0\}} \frac{|a(u-u_h,w)|}{\|w\|_1} \le C_\rho\, h^{k-\frac{1}{2}}\, \|u\|_{W_\infty^{2k-1}(\Omega)}.$$

Lemma 4.14. *We have*

$$\beta\, \|v\|_1 \le |v|_1 + \left| \int_{\partial\Omega} v \right| \qquad \forall\, v \in H^1(\Omega).$$

Lemma 4.15. *For h small enough, we have*

$$a(v,v) \ge \gamma\, \|v\|_1^2 \qquad \forall\, v \in V_h.$$

Theorem 4.16. *Assume that $u \in W_\infty^{2k-1}(\Omega)$ and (4.20) holds for a $\rho > 0$ independent of h. Then we have the following error estimate*

$$\|u-u_h\|_1 \le C_\rho\, h^{k-1}\, \|u\|_{W_\infty^{2k-1}(\Omega)}.$$

The last estimate also holds for $u \in H^k(\Omega)$.

4.10 Isoparametric Polynomial Approximation

Let Ω be a smooth domain in $\mathbb{R}^d$ and let Ω_h be a base polyhedral domain which is close to Ω. Let $\tilde{V}_h$ be a base finite element space defined on Ω_h. (e.g., P_{k-1} finite element space on Ω_h.) We construct a one-to-one continuous mapping

$$F_h : \Omega_h \to \mathbb{R}^n \qquad \text{where each component } F_{h,i} \in \tilde{V}_h.$$

The resulting space

$$V_h := \left\{ v\left(F_h^{-1}(x)\right) \,:\, x \in F_h(\Omega_h),\, v \in \tilde{V}_h \right\}$$

is called an isoparametric-equivalent finite element space.

Let $\mathcal{T}_h$ denote corresponding triangulations consisting of simplices of size at most h. Then, it is possible to construct piecewise polynomial mappings, F_h, of degree $k-1$ which

a) equal the identity map away from the boundary of Ω_h,

b) have the property that the distance from any point on $\partial\Omega$ to the closet point on $\partial F_h(\Omega_h)$ is at most $C\,h^k$,

c) $\|J_{F_h}\|_{W^k_\infty(\Omega_h)} \le C$ and $\|J^{-1}_{F_h}\|_{W^k_\infty(\Omega_h)} \le C$, independent of h.

Note that Ω is only approximated by $F_h(\Omega_h)$, not equal.

We assume that there is an auxilliary mapping $F : \Omega_h \to \Omega$ satisfying the above three conditions a), b), c), and that $F_{h,i} = \mathcal{I}_h F_i$ for each component of the mapping.

Define

$$a_h(v,w) = \int_{F_h(\Omega_h)} \nabla v \cdot \nabla w \, dx$$

and define $\Phi_h : \Omega \to F_h(\Omega_h)$ by $\Phi_h(x) = F_h\left(F^{-1}(x)\right)$. Then, by using chain rule we can write

$$a_h(v,w) = \int_\Omega \Big(J_{\Phi_h}(x)^{-T} \nabla \hat{v}(x) \Big) \cdot \Big(J_{\Phi_h}(x)^{-T} \nabla \hat{w}(x) \Big)\, |\det J_{\Phi_h}(x)|\, dx,$$

where $\hat{v}(x) := v\left(\Phi_h(x)\right)$ for any function v defined on $F_h(\Omega_h)$.

The variational problem is to find $u_h \in V_h$ such that

$$a_h(u_h, v_h) = (f, v)_{F_h(\Omega_h)} \qquad \forall\, v_h \in V_h.$$

Theorem 4.17. *For h sufficiently small, we have the following error estimate:*

$$\|u - \hat{u}_h\|_{1,\Omega} \le C\,h^{k-1} \Big(\|u\|_{k,\Omega} + \|f\|_{W^1_\infty(\Omega)} \Big),$$

where $\hat{u}_h(x) := u_h\left(\Phi_h(x)\right)$.

This can be improved w.r.t. the norm on f:

$$\|u - \hat{u}_h\|_{1,\Omega} \le C\,h^{k-1} \Big(\|u\|_{k,\Omega} + \|f\|_{k-2,\Omega} \Big),$$

5 Mixed Method

5.1 Abstract Formulation

Let X and M be two Hilbert spaces, with norms $\|\cdot\|_X$ and $\|\cdot\|_M$, respectively. Let X' and M' be their dual spaces, and introduce two bilinear forms

$$a(\cdot,\cdot)\,:\,X\times X\to\mathbb{R},\qquad b(\cdot,\cdot)\,:\,X\times M\to\mathbb{R}$$

which are continuous:

$$|a(\mathbf{w},\mathbf{v})|\le\gamma\,\|\mathbf{w}\|_X\,\|\mathbf{v}\|_X,\qquad |b(\mathbf{w},q)|\le\delta\,\|\mathbf{w}\|_X\,\|q\|_M$$

for each $\mathbf{w},\mathbf{v}\in X$ and $q\in M$.

Consider the following constrained problem : find $(\mathbf{u},p)\in X\times M$ such that

$$\begin{aligned} a(\mathbf{u},\mathbf{v})+b(\mathbf{v},p)&=\langle\mathbf{f},\mathbf{v}\rangle &&\forall\,\mathbf{v}\in X,\\ b(\mathbf{u},q)&=\langle g,q\rangle &&\forall\,q\in M,\end{aligned}\tag{5.1}$$

where $\mathbf{f}\in X'$ and $g\in M'$, and $\langle\cdot,\cdot\rangle$ denotes the duality pairing between X' and X or M' and M.

We associate to $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$ the operators $A\in\mathcal{L}(X;X')$ and $B\in\mathcal{L}(X;M')$ defined by

$$\begin{aligned}\langle A\mathbf{w},\mathbf{v}\rangle&=a(\mathbf{w},\mathbf{v}) &&\forall\,\mathbf{w},\mathbf{v}\in V,\\ \langle B\mathbf{v},q\rangle&=b(\mathbf{v},q) &&\forall\,\mathbf{v}\in V,\ q\in M.\end{aligned}$$

Denote by $B'\in\mathcal{L}(M;X')$ the adjoint operator of B:

$$\langle B'q,\mathbf{v}\rangle=\langle B\mathbf{v},q\rangle=b(\mathbf{v},q)\qquad\forall\,\mathbf{v}\in V,\ q\in M.$$

Thus we can write (5.1) as : find $(\mathbf{u},p)\in X\times M$ such that

$$\begin{aligned} A\mathbf{u}+B'p&=\mathbf{f} &&\text{in}\quad X',\\ B\mathbf{u}&=g &&\text{in}\quad M'.\end{aligned}\tag{5.2}$$

Define the affine manifold

$$X^g:=\{\mathbf{v}\in X\,:\,b(\mathbf{v},q)=\langle g,q\rangle\quad\forall\,q\in M\}.$$

Clearly $X^0=\ker B$ is a closed subspace of X. We can now associate to problem (5.1) the following problem:

$$\text{find}\quad\mathbf{u}\in X^g\quad:\quad a(\mathbf{u},\mathbf{v})=\langle\mathbf{f},\mathbf{v}\rangle\qquad\forall\,\mathbf{v}\in X^0.\tag{5.3}$$

That is, if $(\mathbf{u}, p)$ is a solution to (5.1), then $\mathbf{u}$ is a solution to (5.3).
We will introduce suitable conditions ensuring that the converse is also true, and that the solution to (5.3) does exist and is unique, thus construction a solution to (5.1).

Denote by X_p^0 the polar set of X^0,

$$X_p^0 := \{\mu \in X' \,:\, \langle \mu, \mathbf{v}\rangle = 0 \quad \forall\, \mathbf{v} \in X^0\}.$$

Since $X^0 = \ker\, B$, $X_p^0 = (\ker\, B)_p$. Let us decompose X as follows:

$$X = X^0 \oplus (X^0)^\perp.$$

B is not an isomorphism form X onto M', as in general ker B$= X^0 \neq \{0\}$.

We are going to introduce a condition which is equivalent to the fact that B is indeed an isomorphism from $(X^0)^\perp$ onto M' (and moreover B' is an isomorphism form M onto X_p^0).

Proposition 5.1 (Compatibility Condition). *The following statements are equivalent:*

a) there exists a constant $\beta^* > 0$ *such that*

$$\forall\, q \in M \quad \exists \mathbf{v} \in X,\, \mathbf{v} \neq \mathbf{0} \quad : \quad b(\mathbf{v}, q) \geq \beta^* \, \|\mathbf{v}\|_X \, \|q\|_M; \tag{5.4}$$

b) B' *is an isomorphism from* M *onto* X_p^0 *and Inf-Sup Condition holds*

$$\|B'q\|_{X'} := \sup_{0\neq \mathbf{v}\in X} \frac{\langle B'q, \mathbf{v}\rangle}{\|\mathbf{v}\|_X} \geq \beta^* \, \|q\|_M \quad \forall\, q \in M; \tag{5.5}$$

c) B *is an isomorphism from* $(X^0)^\perp$ *onto* M' *and*

$$\|B\mathbf{v}\|_{M'} := \sup_{0\neq q\in M} \frac{\langle B\mathbf{v}, q\rangle}{\|q\|_M} \geq \beta^* \, \|\mathbf{v}\|_X \quad \forall\, \mathbf{v} \in (X^0)^\perp. \tag{5.6}$$

Proof. Clearly (5.4) and (5.5) are equivalent. We have only to prove that B' is an isomorphism from M onto X_p^0. Clearly (5.5) shows that B' is an one-to-one operator from M onto its range $\mathcal{R}(B')$, with a continuous inverse. Thus $\mathcal{R}(B')$ is a closed subspace of X'. It remains to be proven that $\mathcal{R}(B') = X_p^0$. Applying the Closed Range theorem gives

$$\mathcal{R}(B') = (\ker\, B)_p = X_p^0.$$

Hence, a) and b) are equivalent.

$$\boxed{\text{c.f.) } \mathcal{N}(B') = \mathcal{R}(B)_p,\, \mathcal{N}(B) = \mathcal{R}(B')_p,\, \mathcal{N}(B')_p = \mathcal{R}(B),\, \mathcal{N}(B)_p = \mathcal{R}(B').}$$

For each $\mu \in \big((X^0)^\perp\big)'$ we associate $\hat{\mu} \in X'$ satisfying

$$\langle \hat{\mu}, \mathbf{v}\rangle = \langle \mu, P^\perp \mathbf{v}\rangle \quad \forall\, \mathbf{v} \in X,$$

where $P^\perp$ is the orthogonal projection onto $(X^0)^\perp$. Since $P^\perp \mathbf{v} = 0$ for $\mathbf{v} \in X^0$, $\hat{\mu} \in X_p^0$. Using the fact that

$$\|\hat{\mu}\|_{X'} = \sup_{\mathbf{v}\in X} \frac{\langle \hat{\mu}, \mathbf{v}\rangle}{\|\mathbf{v}\|_X} = \sup_{\mathbf{v}\in X} \frac{\langle \mu, P^\perp \mathbf{v}\rangle}{\|\mathbf{v}\|_X} \leq \sup_{\mathbf{v}\in X} \frac{\langle \mu, P^\perp \mathbf{v}\rangle}{\|P^\perp \mathbf{v}\|_X} = \sup_{\mathbf{v}\in (X^0)^\perp} \frac{\langle \mu, \mathbf{v}\rangle}{\|\mathbf{v}\|_X} = \|\mu\|_{\left((X^0)^\perp\right)'}$$

and

$$\|\mu\|_{\left((X^0)^\perp\right)'} = \sup_{\mathbf{v}\in (X^0)^\perp} \frac{\langle \mu, \mathbf{v}\rangle}{\|\mathbf{v}\|_X} = \sup_{\mathbf{v}\in (X^0)^\perp} \frac{\langle \mu, P^\perp \mathbf{v}\rangle}{\|\mathbf{v}\|_X} = \sup_{\mathbf{v}\in (X^0)^\perp} \frac{\langle \hat{\mu}, \mathbf{v}\rangle}{\|\mathbf{v}\|_X} \leq \sup_{\mathbf{v}\in X} \frac{\langle \hat{\mu}, \mathbf{v}\rangle}{\|\mathbf{v}\|_X} = \|\hat{\mu}\|_{X'}$$

yields that $\mu \to \hat{\mu}$ is an isometric bijection from $\left((X^0)^\perp\right)'$ onto X_p^0. Hence, X_p^0 can be identified with the dual of $(X^0)^\perp$. As consequence, B' is an isomorphism from M onto $\left((X^0)^\perp\right)'$ satisfying

$$\|(B')^{-1}\|_{\mathcal{L}(X_p^0;M)} \leq \frac{1}{\beta^*}$$

if and only if B is an isomorphism from $(X^0)^\perp$ onto M' satisfying

$$\|B^{-1}\|_{\mathcal{L}(M';(X^0)^\perp)} \leq \frac{1}{\beta^*}.$$

This proof is now complete. □

Theorem 5.2. *Assume that the bilinear form $a(\cdot,\cdot)$ is continuous and coercive on X^0:*

$$a(\mathbf{v},\mathbf{v}) \geq \alpha \|\mathbf{v}\|_X^2 \qquad \forall \mathbf{v} \in X^0.$$

Assume further that the bilinear form $b(\cdot,\cdot)$ is continuous and the compatibility condition (5.4) holds.

Then, for each $(\mathbf{f}, g) \in X' \times M'$ there exists a unique solution $\mathbf{u}$ to (5.3), and a unique $p \in M$ such that $(\mathbf{u}, p)$ is the unique solution to (5.1).

Furthermore, the map $(\mathbf{f}, g) \to (\mathbf{u}, p)$ is an isomorphism from $X' \times M'$ onto $X \times M$, and

$$\|\mathbf{u}\|_X \leq \frac{1}{\alpha}\left(\|\mathbf{f}\|_{X'} + \frac{\alpha+\gamma}{\beta^*}\|g\|_{M'}\right)$$
$$\|p\|_M \leq \frac{1}{\beta^*}\left(\frac{\alpha+\gamma}{\alpha}\|\mathbf{f}\|_{X'} + \frac{\gamma(\alpha+\gamma)}{\alpha\beta^*}\|g\|_{M'}\right).$$

Proof. Uniqueness of the solution to (5.3) is a straightforward consequence of the coerciveness.

From (5.6), there exists a unique $\mathbf{u}^0 \in (X^0)^\perp$ such that $B\mathbf{u}^0 = g$ and

$$\|\mathbf{u}^0\|_X \leq \frac{1}{\beta^*}\|g\|_{M'}.$$

Thus, we rewrite (5.3) as:

$$\text{(5.7)} \qquad \text{find} \quad \tilde{\mathbf{u}} \in X^0 \quad : \quad a(\tilde{\mathbf{u}}, \mathbf{v}) = \langle \mathbf{f}, \mathbf{v}\rangle - a(\mathbf{u}^0, \mathbf{v}) \qquad \forall\, \mathbf{v} \in X^0,$$

and define the solution $\mathbf{u} \in X^g$ to (5.3) as $\mathbf{u} = \tilde{\mathbf{u}} + \mathbf{u}^0$. The existence of a unique solution to (5.7) is assured by Lax-Milgram lemma, and moreover

$$\|\tilde{\mathbf{u}}\|_X \leq \frac{1}{\alpha}\left(\|\mathbf{f}\|_{X'} + \gamma\,\|\mathbf{u}^0\|_X\right).$$

Let us now consider problem (5.1). As (5.7) can be written in the form

$$\langle A\mathbf{u} - \mathbf{f}, \mathbf{v}\rangle = 0 \qquad \forall\, \mathbf{v} \in X^0,$$

we have $(A\mathbf{u} - \mathbf{f}) \in X^0_p$. Moreover, from (5.5) we can find a unique $p \in M$ such that $A\mathbf{u} - \mathbf{f} = -B'p$, i.e., $(\mathbf{u}, p)$ is a solution to (5.1) and satisfies

$$\|p\|_M \leq \frac{1}{\beta^*}\,\|A\mathbf{u} - \mathbf{f}\|_{X'}.$$

Each solution $(\mathbf{u}, p)$ to (5.1) gives a solution $\mathbf{u}$ to (5.3), also for problem (5.3) uniqueness thus holds. Finally, summing up inequalities the proof is completed. □

The approximation of the abstract constrained problem (5.1) is as follows. Let X_h and M_h be finite dimensional subspace of X and M, respectively. The discrete constrained problem is to: find $(\mathbf{u}_h, p_h) \in X_h \times M_h$ such that

$$\begin{aligned} a(\mathbf{u}_h, \mathbf{v}) + b(\mathbf{v}, p_h) &= \langle \mathbf{f}, \mathbf{v}\rangle \qquad \forall\, \mathbf{v} \in X_h, \\ b(\mathbf{u}_h, q) &= \langle g, q\rangle \qquad \forall\, q \in M_h. \end{aligned} \tag{5.8}$$

Define the space

$$X^g_h := \{\mathbf{v}_h \in X_h \,:\, b(\mathbf{v}_h, q) = \langle g, q\rangle \quad \forall q \in M_h\}.$$

Since M_h is in general a proper subspace of M, $X^g_h \not\subseteq X^g$.

The finite dimensional problem corresponding to (5.3) is to

$$\text{find} \quad \mathbf{u}_h \in X^g_h \quad : \quad a(\mathbf{u}_h, \mathbf{v}) = \langle \mathbf{f}, \mathbf{v}\rangle \qquad \forall\, \mathbf{v} \in X^g_h. \tag{5.9}$$

5.2 Analysis of Stability and Convergence

Theorem 5.3 (Stability). *Assume that the bilinear form $a(\cdot,\cdot)$ is continuous and coercive on X^0_h:*

$$a(\mathbf{v}_h, \mathbf{v}_h) \geq \alpha_h\,\|\mathbf{v}_h\|^2_X \qquad \forall \mathbf{v}_h \in X^0_h.$$

Assume further that the bilinear form $b(\cdot,\cdot)$ is continuous and the following compatibility condition holds: there exists a constant $\beta_h > 0$ such that

$$\forall\, q_h \in M_h \quad \exists \mathbf{v}_h \in X_h,\, \mathbf{v}_h \neq \mathbf{0} \quad : \quad b(\mathbf{v}_h, q_h) \geq \beta_h\,\|\mathbf{v}_h\|_X\,\|q_h\|_M. \tag{5.10}$$

Then, for each $(\mathbf{f}, g) \in X' \times M'$ *there exists a unique solution* $(\mathbf{u}_h, p_h)$ *to (5.8) which satisfies*

$$\|\mathbf{u}_h\|_X \leq \frac{1}{\alpha_h}\left(\|\mathbf{f}\|_{X'} + \frac{\alpha_h + \gamma}{\beta_h}\|g\|_{M'}\right)$$
$$\|p_h\|_M \leq \frac{1}{\beta_h}\left(\frac{\alpha_h + \gamma}{\alpha_h}\|\mathbf{f}\|_{X'} + \frac{\gamma(\alpha_h + \gamma)}{\alpha_h\beta_h}\|g\|_{M'}\right)$$

where both α_h *and* β_h *are independent of* h*, this is a stability result for* $(\mathbf{u}_h, p_h)$*.*

The proof is similar to the preceding Theorem. But, note that $X_h^0 \not\subseteq X^0$ and that the compatibility condition (5.4) does not imply (5.10) since X_h is a proper subspace of X.

Theorem 5.4 (Convergence)**.** *Let the assumptions of two preceding Theorems be satisfied. Then, we have the error estimates*

$$\|\mathbf{u} - \mathbf{u}_h\|_X \leq \left(1 + \frac{\gamma}{\alpha_h}\right) \inf_{\mathbf{v}_h^* \in X_h^g} \|\mathbf{u} - \mathbf{v}_h^*\|_X + \frac{\delta}{\alpha_h} \inf_{q_h \in M_h} \|p - q_h\|_M,$$
$$\|p - p_h\|_M \leq \frac{\gamma}{\beta_h}\left(1 + \frac{\gamma}{\alpha_h}\right) \inf_{\mathbf{v}_h^* \in X_h^g} \|\mathbf{u} - \mathbf{v}_h^*\|_X + \left(1 + \frac{\delta}{\beta_h} + \frac{\gamma\delta}{\alpha_h\beta_h}\right) \inf_{q_h \in M_h} \|p - q_h\|_M.$$

Moreover, the following estimate holds

$$\inf_{\mathbf{v}_h^* \in X_h^g} \|\mathbf{u} - \mathbf{v}_h^*\|_X \leq \left(1 + \frac{\delta}{\beta_h}\right) \inf_{\mathbf{v}_h \in X_h} \|\mathbf{u} - \mathbf{v}_h\|_X.$$

The convergence is optimal if both α_h *and* β_h *are independent of* h*.*

Proof. Take $\mathbf{v}_h \in X_h$, $\mathbf{v}_h^* \in X_h^g$ and $q_h \in M_h$. By subtracting $(5.8)_1$ from $(5.1)_1$ it follows

$$a(\mathbf{u}_h - \mathbf{v}_h^*, \mathbf{v}_h) + b(\mathbf{v}_h, p_h - q_h) = a(\mathbf{u} - \mathbf{v}_h^*, \mathbf{v}_h) + b(\mathbf{v}_h, p - q_h).$$

Choosing $\mathbf{v}_h = (\mathbf{u}_h - \mathbf{v}_h^*) \in X_h^0$, we have

$$\|\mathbf{u}_h - \mathbf{v}_h^*\|_X \leq \frac{1}{\alpha_h}\left(\gamma\|\mathbf{u} - \mathbf{v}_h^*\|_X + \delta\|p - q_h\|_M\right),$$

and consequently the first estimate of the theorem holds true.

For each $q_h \in M_h$, we find

$$\|p_h - q_h\|_M \leq \frac{1}{\beta_h} \sup_{0 \neq \mathbf{v}_h \in X_h} \frac{b(\mathbf{v}_h, p_h - q_h)}{\|\mathbf{v}_h\|_X}.$$

By subtracting $(5.8)_2$ from $(5.1)_2$ it follows

$$b(\mathbf{v}_h, p_h - q_h) = a(\mathbf{u} - \mathbf{u}_h, \mathbf{v}_h) + b(\mathbf{v}_h, p - q_h).$$

Combining this with the last inequality and using the continuities, we obtain

$$\|p_h - q_h\|_M \leq \frac{1}{\beta_h}\Big(\gamma\|\mathbf{u} - \mathbf{u}_h\|_X + \delta\|p - q_h\|_M\Big)$$

which yields the second estimate of the theorem.

For each $\mathbf{v}_h \in X_h$, from (5.10) and Proposition 5.1 (we similarly show that there exists an isomorphism B_h from $(X_h^0)^\perp$ to M_h' which has the similar properties in Proposition 5.1) there exists a unique $\mathbf{w}_h \in (X_h^0)^\perp$ such that

$$\langle B_h \mathbf{w}_h, q_h\rangle = \langle B_h(\mathbf{u} - \mathbf{v}_h), q_h\rangle \quad \text{or} \quad b(\mathbf{w}_h, q_h) = b(\mathbf{u} - \mathbf{v}_h, q_h) \qquad \forall\, q_h \in M_h$$

and

$$\|\mathbf{w}_h\|_X \leq \frac{1}{\beta_h} \sup_{0 \neq q_h \in M_h} \frac{\langle B_h \mathbf{w}_h, q_h\rangle}{\|q_h\|_M} \leq \frac{\delta}{\beta_h}\|\mathbf{u} - \mathbf{v}_h\|_X.$$

Setting $\mathbf{v}_h^* := \mathbf{w}_h + \mathbf{v}_h$, we have $\mathbf{v}_h^* \in X_h^g$ as $b(\mathbf{v}_h^*, q_h) = b(\mathbf{u}, q_h) = \langle g, q_h\rangle$ for all $q_h \in M_h$. Furthermore,

$$\|\mathbf{u} - \mathbf{v}_h^*\|_X \leq \|\mathbf{u} - \mathbf{v}_h\|_X + \|\mathbf{w}_h\|_X \leq \left(1 + \frac{\delta}{\beta_h}\right)\|\mathbf{u} - \mathbf{v}_h\|_X.$$

Since $\mathbf{v}_h$ is arbitrary, this completes the proof. □

(Spurious Modes)**.** The compatibility condition (5.10) is necessary to achieve uniqueness of p_h. Actually it can be written as:

$$\text{if } q_h \in M_h \text{ and } b(\mathbf{v}_h, q_h) = 0 \text{ for each } \mathbf{v}_h \in X_h, \text{ then } q_h = 0.$$

Thus, if (5.10) is not satisfied, there exists $q_h^* \in M_h$, $q_h^* \neq 0$, such that

$$b(\mathbf{v}_h, q_h^*) = 0 \qquad \forall\, \mathbf{v}_h \in X_h.$$

As a consequence, if $(\mathbf{u}_h, p_h)$ solves (5.8), also $(\mathbf{u}_h, p_h + \tau q_h^*)$, $\tau \in \mathbb{R}$ is a solution to the same problem.

Any such element q_h^* is called *spurious* (or *parasitic*) *mode*, as it cannot be detected by the numerical problem (5.8).

5.3 How to verify the uniform compatibility condtion

Lemma 5.5 (Fortin)**.** *Assume that the compatibility condition (5.4) is satisfied, and, moreover, that there exists an operator $r_h : X \to X_h$ such that*

$$(i) \quad b\big(\mathbf{v} - r_h(\mathbf{v}), q_h\big) = 0 \quad \forall\, \mathbf{v} \in X, \quad \forall\, q_h \in M_h$$
$$(ii) \quad \|r_h(\mathbf{v})\|_X \le C_* \, \|\mathbf{v}\|_X \quad \forall\, \mathbf{v} \in X,$$

where $C_* > 0$ doesn't depend on h.
Then, the compatibility condition (5.10) is satisfied with $\beta = \beta^*/C_*$.

For the connected domain Ω, let $X = H_0^1(\Omega)^d$ and $M = L_0^2(\Omega)$, and let

$$V = \{\mathbf{v} \in X \, : \, \nabla \cdot \mathbf{v} = 0\}.$$

Then, V is a closed subspace of $H_0^1(\Omega)^d$ and we have the decomposition:

$$H_0^1(\Omega)^d = V \,\oplus\, V^\perp, \qquad V^\perp \text{ is orthogonal of } V \text{ in } H_0^1(\Omega)^d.$$

From the arguments in p. 24 of [GR], we have
a) the operator $\nabla \; L_0^2(\Omega) \to V_p$ is an isomorphism;
b) the operator $\nabla \cdot \, : \, V^\perp \to L_0^2(\Omega)$ is an isomorphism
where $V_p = \{\mathbf{f} \in H^{-1}(\Omega)^d \, : \, \langle \mathbf{f}, \mathbf{v} \rangle \quad \forall\, \mathbf{v} \in V\}$ denotes the polar set of V.
For the bilinear forms

$$a(\mathbf{w}, \mathbf{v}) = \nu\,(\nabla \mathbf{w}, \nabla \mathbf{v}) \quad \text{and} \quad b(\mathbf{v}, q) = -(q, \nabla \cdot \mathbf{v}), \qquad \nu > 0,$$

we have the following result:

Lemma 5.6 (Verfürth)**.** *Let $\mathcal{T}_h$ be a quasi-uniform family of triangulations of Ω. Assume that $M_h \subset H^1(\Omega) \cap L_0^2(\Omega)$ and there exists $\hat{\beta} > 0$ such that*

$$(5.11) \qquad \forall\, q_h \in M_h \quad \exists\, \mathbf{v}_h \in X_h, \quad \mathbf{v}_h \neq 0 \quad : \quad b(\mathbf{v}_h, q_h) \ge \hat{\beta}\, \|\mathbf{v}_h\|_0 \, \|q_h\|_1.$$

Assume, moreover, that there exists an operator $R_h : X \to X_h$ and a constant $K > 0$ such that

$$(5.12) \qquad \|\mathbf{v} - R_h(\mathbf{v})\|_0 + h\, \|\mathbf{v} - R_h(\mathbf{v})\|_1 \le K\, h\, |\mathbf{v}|_1 \qquad \forall \mathbf{v} \in X.$$

Then, the compatibility condition (5.10) is satisfied.

Proof. Recall the inverse inequality:

$$\|\mathbf{v}_h\|_1 \le C\,h^{-1}\,\|\mathbf{v}_h\|_0 \qquad \forall\,\mathbf{v}_h \in X_h.$$

The condition (5.11) implies that

$$\text{(5.13)} \qquad \forall\,q_h \in M_h \quad \exists\,\mathbf{v}_h \in X_h, \quad \mathbf{v}_h \neq 0 \quad : \quad b(\mathbf{v}_h, q_h) \ge K_1\,h\,\|\mathbf{v}_h\|_1\,\|q_h\|_1.$$

Note that for each $q_h \in M_h \subset L_0^2(\Omega)$ there exists $\mathbf{w} \in V^\perp \subset X = H_0^1(\Omega)^d$ such that

$$\nabla\cdot\mathbf{w} = -q_h \quad \text{and} \quad \|\mathbf{w}\|_1 \le K_2\,\|q_h\|_0.$$

Thus from (5.12) we find

$$\|R_h(\mathbf{w})\|_1 \le \|R_h(\mathbf{w}) - \mathbf{w}\|_1 + \|\mathbf{w}\|_1 \le (1+K)\|\mathbf{w}\|_1 \le K_2(1+K)\|q_h\|_0.$$

If q_h is such that

$$\|q_h\|_0 \le KK_2\,h\,\|q_h\|_1,$$

then (5.13) yields

$$b(\mathbf{v}_h, q_h) \ge \frac{K_1}{KK_2}\,\|\mathbf{v}_h\|_1\,\|q_h\|_0.$$

On the contrary, if

$$\|q_h\|_0 > KK_2\,h\,\|q_h\|_1,$$

$$\begin{aligned}
b(R_h(\mathbf{w}), q_h) &= b(\mathbf{w}, q_h) + b(R_h(w) - \mathbf{w}, q_h) = (q_h, q_h) + b(R_h(w) - \mathbf{w}, q_h)\\
&\ge \|q_h\|_0^2 - \|R_h(\mathbf{w}) - \mathbf{w}\|_0\,\|q_h\|_1 \ge \|q_h\|_0^2 - K\,h\,|\mathbf{w}|_1\,\|q_h\|_1\\
&\ge \|q_h\|_0^2 - KK_2\,h\,\|q_h\|_0\,\|q_h\|_1\\
&\ge \frac{1}{K_2(1+K)}\Big(\|q_h\|_0 - KK_2\,h\,\|q_h\|_1\Big)\|R_h(\mathbf{w})\|_1.
\end{aligned}$$

Combining this with (5.13), we are led to that for each $q_h \in M_h$ satisfying

$$\|q_h\|_0 \le KK_2\,h\,\|q_h\|_1,$$

there exists $\mathbf{z}_h \in X_h$, $\mathbf{z}_h \neq 0$, such that

$$b(\mathbf{z}_h, q_h) \ge Q(h\,\|q_h\|_1)\|\mathbf{z}_h\|_1,$$

where

$$Q(\xi) := \max\left\{K_1\,\xi, \quad \frac{1}{K_2(1+K)}\Big(\|q_h\|_0 - KK_2\,\xi\Big)\right\}, \quad \xi > 0.$$

We have the minimum of $Q(\xi)$ over $\{\xi > 0\}$ at $\xi = \|q_h\|_0 / \big(K_2[K_1(1+K)+K]\big)$ and hence (5.10) holds with $\beta = K_1\big(K_2[K_1(1+K)+K]\big)^{-1}$. $\square$

The condition (5.11) has been proven to hold to some finite element spaces frequently used in the approximation of the Stokes problem (see for instance the Taylor-Hood or Bercovier-Pironneau elements).