
Introduction to the Finite Element method

Gustavo C. Buscaglia

ICMC-USP, São Carlos, Brasil
gustavo.buscaglia@gmail.com

Motivation

- For **elliptic** and **parabolic** problems, the most popular approximation method is the FEM.
- It is **general**, not restricted to linear problems, or to isotropic problems, or to any subclass of mathematical problems.
- It is **geometrically flexible**, complex domains are quite easily treated, not requiring adaptations of the method itself.
- It is **easy to code**, and the coding is quite problem-independent. Boundary conditions are much easier to deal with than in other methods.
- It is **robust**, because in most cases the mathematical problem has an underlying variational structure (energy minimization, for example).

Overview

- **Galerkin approximations:** Differential, variational and extremal formulations of a simple 1D boundary value problem. Well-posedness of variational formulations. Functional setting. Strong and weak coercivity. Lax-Milgram lemma. Banach's open mapping theorem. Céa's best-approximation property. Convergence under weak coercivity. (2 lectures)
- **The spaces of FEM and their implementation:** (3 lectures)
- **Interpolation error and convergence:** (2 lectures)
- **Application to convection-diffusion-reaction problems:** (2 lectures)
- **Application to linear elasticity:** (2 lectures)
- **Mixed problems:** (2 lectures)
- **FEM for parabolic problems:** (2 lectures)

1 Galerkin approximations

1.1 Variational formulation of a simple 1D example

Let u be the solution of

$$\begin{cases} -u'' + u = f & \text{in } (0, 1) \\ u(0) = u(1) = 0 \end{cases} \quad (1.1)$$

The **differential formulation** (DF) of the problem requires $-u'' + u$ to be exactly equal to f in **all** points $x \in (0, 1)$.

Multiplying the equation by any function v and integrating by parts (recall that

$$\int_0^1 w' z \, dx = w(1)z(1) - w(0)z(0) - \int_0^1 w z' \, dx \quad (1.2)$$

holds for all w and z that are *regular enough*) one obtains that u satisfies

$$\int_0^1 (u' v' + u v) \, dx - u'(1)v(1) + u'(0)v(0) = \int_0^1 f v \, dx \quad \forall v. \quad (1.3)$$

- The requirement “for all x ” of the DF has become “for all functions v ”.
- Does equation (1.3) fully determine u ?
- What happened with the boundary conditions?

Consider the following problem in **variational formulation** (VF): “Determine $u \in W$, such that $u(0) = u(1) = 0$ and that

$$\int_0^1 (u' v' + u v) dx = \int_0^1 f v dx \quad (1.4)$$

holds for all $v \in W$ satisfying $v(0) = v(1) = 0$.”

Prop. 1.1 *The solution u of the DF (eq. 1.1) is also a solution of the VF if W consists of continuous functions of sufficient regularity. As a consequence, problem VF admits at least one solution whenever DF does.*

Proof. Following the steps that lead to the VF, it becomes clear that the only requirement for u to satisfy (1.4) is that the integration by parts formula (1.2) be valid. \square

Exo. 1.1 *Show that the solution of*

$$\begin{cases} -u'' + u = f & \text{in } (0, 1) \\ u(0) = 0, \quad u'(1) = g \in \mathbb{R} \end{cases} \quad (1.5)$$

is a solution to: “Find $u \in W$ such that $u(0) = 0$ and that

$$\int_0^1 (u' v' + u v) dx = \int_0^1 f v dx + g v(1) \quad (1.6)$$

holds for all $v \in W$ satisfying $v(0) = 0$.”

Consider the following problem in **extremal formulation** (EF): “Determine $u \in W$ such that it minimizes the function

$$J(w) = \int_0^1 \left(\frac{1}{2}w'(x)^2 + \frac{1}{2}w(x)^2 - f w \right) dx \quad (1.7)$$

over the functions $w \in W$ that satisfy $w(0) = w(1) = 0$.”

Prop. 1.2 *The unique solution u of (1.1) is also a solution to EF. As a consequence, EF admits at least one solution.*

Proof. We need to show that $J(w) \geq J(u)$ for all $w \in W_0$, where

$$W_0 = \{w \in W, w(0) = w(1) = 0\}$$

Writing $w = u + \alpha v$ and replacing in (1.7) one obtains

$$J(u + \alpha v) = J(u) + \alpha \left[\int_0^1 (u' v' + u v - f v) dx \right] + \alpha^2 \int_0^1 \left(\frac{1}{2}v'(x)^2 + \frac{1}{2}v(x)^2 \right) dx$$

The last term is not negative and the second one is zero. \square

Exo. 1.2 *Identify the EF of the previous exercise.*

Prop. 1.3 Let u be the solution of

$$\begin{cases} -u'' + u = f & \text{in } (0, 1) \\ u(0) = 1, \quad u'(1) = g \in \mathbb{R} \end{cases} \quad (1.8)$$

then u is also a solution of “Determine $u \in W$ such that $u(0) = 1$ and that

$$\int_0^1 (u' v' + u v) dx = \int_0^1 f v dx + g v(1) \quad (1.9)$$

holds for all $v \in W$ satisfying $v(0) = 0$.”

Further, defining for any $a \in \mathbb{R}$

$$W_a = \{w \in W, w(0) = a\},$$

u minimizes over W_1 the function

$$J(w) = \int_0^1 \left(\frac{1}{2} w'(x)^2 + \frac{1}{2} w(x)^2 - f w \right) dx - g w(1). \quad (1.10)$$

Exo. 1.3 Prove the last proposition.

Let us define the bilinear and linear forms corresponding to problem (1.1):

$$a(v, w) = \int_0^1 (v'w' + vw) \, dx \qquad \ell(v) = \int_0^1 f v \, dx \qquad (1.11)$$

and the function $J(v) = \frac{1}{2}a(v, v) - \ell(v)$. Remember that W is a space of functions with some (yet unspecified) regularity and let $W_0 = \{w \in W, w(0) = w(1) = 0\}$.

The three formulations that we have presented up to now are, thus:

DF: Find a function u such that

$$-u''(x) + u(x) = f(x) \qquad \forall x \in (0, 1), \qquad u(0) = u(1) = 0$$

VF: Find a function $u \in W_0$ such that

$$a(u, v) = \ell(v) \quad \forall v \in W_0$$

EF: Find a function $u \in W_0$ such that

$$J(u) \leq J(w) \quad \forall w \in W_0$$

and we know that the exact solution of DF is also a solution of VF and of EF.

The logic of the construction is justified by the following

Theorem 1.4 *If W is taken as*

$$W = \{w : (0, 1) \rightarrow \mathbb{R}, \int_0^1 w(x)^2 dx < +\infty, \int_0^1 w'(x)^2 dx < +\infty\} \stackrel{\text{def}}{=} H^1(0, 1)$$

and if f is such that there exists $C \in \mathbb{R}$ for which

$$\int_0^1 f(x) w(x) dx \leq C \sqrt{\int_0^1 w'(x)^2 dx} \quad \forall w \in W_0 \quad (1.12)$$

then problems (VF) and (EF) have one and only one solution, and their solutions coincide.

The proof will be given later, now let us consider its consequences:

- The differential equation has at most one solution in W .
- If the solution u to (VF)-(EF) is regular enough to be considered a solution to (DF), then u is the solution to (DF).
- If the solution u to (VF)-(EF) is not regular enough to be considered a solution to (DF), then (DF) has no solution.

\Rightarrow (VF) is a generalization of (DF).

Exo. 1.4 Show that $W_0 \subset C^0(0, 1)$. Further, compute $C \in \mathbb{R}$ such that

$$\max_{x \in [0, 1]} |w(x)| \leq C \sqrt{\int_0^1 w'(x)^2 dx} \quad \forall w \in W_0$$

Hint: You may assume that $\int_0^1 f(x)g(x) dx \leq \sqrt{\int_0^1 f(x)^2 dx} \sqrt{\int_0^1 g(x)^2 dx}$ for any f and g (Cauchy-Schwarz).

Exo. 1.5 Consider $f(x) = |x - 1/2|^\gamma$. For which exponents γ is $\int_0^1 f(x)w(x) dx < +\infty$ for all $w \in W_0$?

Exo. 1.6 Consider as f the “Dirac delta function” at $x = 1/2$, that we will denote by $\delta_{1/2}$. It can be considered as a “generalized” function defined by

$$\int_0^1 \delta_{1/2}(x) w(x) dx = w(1/2) \quad \forall w \in C^0(0, 1)$$

Prove that $\delta_{1/2}$ satisfies (1.12) and determine the analytical solution to (VF).

Exo. 1.7 Determine the DF and the EF corresponding to the following VF: “Find $u \in W = H^1(0, 1)$, $u(0) = 1$, such that

$$\int_0^1 (u'w' + uw) dx = w(1/2) \quad \forall w \in W_0 \tag{1.13}$$

where $W_0 = \{w \in W, w(0) = 0\}$.”

1.2 Variational formulations in general

Let V be a Hilbert space with norm $\|\cdot\|_V$. Let $a(\cdot, \cdot)$ and $\ell(\cdot)$ be bilinear and linear forms on V satisfying (continuity), for all $v, w \in V$,

$$a(v, w) \leq N_a \|v\|_V \|w\|_V, \quad \ell(v) \leq N_\ell \|v\|_V \quad (1.14)$$

This last inequality means that $\ell \in V'$, the (topological) dual of V . The minimum N_ℓ that satisfies this inequality is called the norm of ℓ in V' , i.e.

$$\|\ell\|_{V'} \stackrel{\text{def}}{=} \sup_{0 \neq v \in V} \frac{\ell(v)}{\|v\|_V} \quad (1.15)$$

The abstract VF we consider here is:

$$\text{“Find } u \in V \text{ such that } \quad a(u, v) = \ell(v) \quad \forall v \in V\text{”} \quad (1.16)$$

Exo. 1.8 Assume that V is finite dimensional, of dimension n , and let $\{\phi^1, \phi^2, \dots, \phi^n\}$ be a basis. Show that (1.16) is then equivalent to

$$\underline{V}^T \underline{A} \underline{U} = \underline{V}^T \underline{L} \quad \forall \underline{V} \in \mathbb{R}^n, \quad (1.17)$$

which in turn is equivalent to the linear system

$$\underline{A} \underline{U} = \underline{L}; \quad (1.18)$$

where

$$A_{ij} \stackrel{\text{def}}{=} a(\phi^j, \phi^i), \quad L_i \stackrel{\text{def}}{=} \ell(\phi^i) \quad (1.19)$$

and \underline{U} is the coefficient column vector of the expansion of u , i.e.,

$$u = \sum_{i=1}^n U_i \phi^i \quad (1.20)$$

Def. 1.5 The bilinear form $a(\cdot, \cdot)$ is said to be **strongly coercive** if there exists $\alpha > 0$ such that

$$a(v, v) \geq \alpha \|v\|_V^2 \quad \forall v \in V \quad (1.21)$$

Def. 1.6 The bilinear form $a(\cdot, \cdot)$ is said to be **weakly coercive** (or to satisfy an **inf-sup** condition) if there exists $\beta > 0$ such that

$$\sup_{0 \neq w \in V} \frac{a(v, w)}{\|w\|_V} \geq \beta \|v\|_V \quad \forall v \in V \quad (1.22)$$

and

$$\sup_{0 \neq v \in V} \frac{a(v, w)}{\|v\|_V} \geq \beta \|w\|_V \quad \forall w \in V \quad (1.23)$$

Exo. 1.9 Prove that strong coercivity implies weak coercivity.

Exo. 1.10 Prove that, if V is finite dimensional, then **(i)** $a(\cdot, \cdot)$ is strongly coercive iff $\underline{\underline{A}}$ is positive definite ($\underline{\underline{X}}^T \underline{\underline{A}} \underline{\underline{X}} > 0 \forall \underline{\underline{X}} \in \mathbb{R}^n$), and **(ii)** $a(\cdot, \cdot)$ is weakly coercive iff $\underline{\underline{A}}$ is invertible.

Exo. 1.11 Prove that, if $a(\cdot, \cdot)$ is weakly coercive, then the solution u of (1.16) depends continuously on the forcing $\ell(\cdot)$. Specifically, prove that

$$\|u\|_V \leq \frac{1}{\beta} \|\ell\|_{V'} \quad (1.24)$$

Theorem 1.7 Assuming V to be a Hilbert space, problem (1.16) is well posed for any $\ell \in V'$ if and only if **(i)** $a(\cdot, \cdot)$ is continuous, and **(ii)** $a(\cdot, \cdot)$ is weakly coercive.

A simpler version of this result is known as **Lax-Milgram lemma**:

Theorem 1.8 Assuming V to be a Hilbert space, if $a(\cdot, \cdot)$ is continuous and strongly coercive then problem (1.16) is well posed for any $\ell \in V'$.

Proof. This proof uses the so-called ‘‘Galerkin method’’, which will be useful to introduce... the Galerkin method!

Let $\{\phi^i\}$ be a basis of V . Denoting $V_N = \text{span}(\phi^1, \dots, \phi^N)$ we can define $u_N \in V_N$ as the unique solution of $a(u_N, v) = \ell(v)$ for all $v \in V_N$. This generates a sequence $\{u_N\}_{N=1,2,\dots}$ in V . Further, this sequence is bounded, because

$$\|u_N\|_V^2 \leq \frac{1}{\alpha} a(u_N, u_N) = \frac{1}{\alpha} \ell(u_N) \leq \frac{\|\ell\|_{V'}}{\alpha} \|u_N\|_V \quad \Rightarrow \quad \|u_N\|_V \leq \frac{\|\ell\|_{V'}}{\alpha}, \quad \forall N$$

Recalling the weak compactness of bounded sets in Hilbert spaces, there exists $u \in V$ such that a subsequence of $\{u_N\}$ (still denoted by $\{u_N\}$ for simplicity) converges to u weakly. It remains to prove that $a(u, v) = \ell(v)$ for all $v \in V$. To see this, notice that

$$a(u, \phi^i) = a(\lim_N u_N, \phi^i) = \lim_N a(u_N, \phi^i) = \ell(\phi^i)$$

where the last equality holds because $a(u_N, \phi^i) = \ell(\phi^i)$ whenever $N \geq i$. Uniqueness is left as an exercise. \square

Exo. 1.12 Prove uniqueness in the previous theorem (bounded sequences may have several accumulation points).

Remark 1.9 The space $L^2(a, b)$ (also denoted by $H^0(a, b)$) is the Hilbert space of functions $f : (a, b) \rightarrow \mathbb{R}$ such that $\int_a^b f^2(x) dx < +\infty$.

The scalar product is

$$(f, g)_{L^2(a,b)} = \int_a^b f(x)g(x) dx \quad (1.25)$$

and accordingly

$$\|f\|_{L^2(a,b)} = (f, f)_{L^2(a,b)}^{1/2} = \sqrt{\int_a^b f^2(x) dx} . \quad (1.26)$$

Also of frequent use are the Hilbert spaces $H^1(a, b)$ and $H^2(a, b)$:

$$H^1(a, b) = \{f \in L^2(a, b) \mid f' \in L^2(a, b)\} \quad (1.27)$$

$$|f|_{H^1(a,b)} = \|f'\|_{L^2(a,b)} \quad (1.28)$$

$$\|f\|_{H^1(a,b)} = \|f\|_{L^2(a,b)} + |f|_{H^1(a,b)} \quad (1.29)$$

$$H^2(a, b) = \{f \in H^1(a, b) \mid f'' \in L^2(a, b)\} \quad (1.30)$$

$$|f|_{H^2(a,b)} = \|f''\|_{L^2(a,b)} \quad (1.31)$$

$$\|f\|_{H^2(a,b)} = \|f\|_{H^1(a,b)} + |f|_{H^2(a,b)} \quad (1.32)$$

Exo. 1.13 Other equivalent norms can be defined in $H^1(a, b)$, e.g.,

1. $\| \|f\| \|_{H^1(a,b)} = \left(\|f\|_{L^2(a,b)}^2 + |f|_{H^1(a,b)}^2 \right)^{1/2}$

2. $\| \|f\| \|_{H^1(a,b)} = \max \left(\|f\|_{L^2(a,b)}, |f|_{H^1(a,b)} \right)$

3. $\| \|f\| \|_{H^1(a,b)} = \|f\|_{L^2(a,b)} + \|\ell f'\|_{L^2(a,b)}$, where $\ell : (a, b) \rightarrow \mathbb{R}$ satisfies $0 < \ell_{\min} \leq \ell(x) \leq \ell_{\max}$ for all $x \in (a, b)$. Notice that if $\ell(x)$ has dimensions of length then this norm is unit-consistent.

Find the constants c and C such that $c\|f\| \leq \|f\| \leq C\|f\|$.

Remark 1.10 For the spaces $H^1(a, b)$ and $H^2(a, b)$ to be complete, one needs a weaker definition of the derivative. For this purpose, one first introduces the space

$$\mathcal{D}(a, b) = C_0^\infty(a, b) = \{\varphi \in C^\infty(a, b) \mid \varphi \text{ has compact support in } (a, b)\} . \quad (1.33)$$

Given a function $f : (a, b) \rightarrow \mathbb{R}$, if there exists $g : (a, b) \rightarrow \mathbb{R}$ such that

$$\int_a^b g(x) \varphi(x) dx = - \int_a^b f(x) \varphi'(x) dx , \quad \forall \varphi \in \mathcal{D}(a, b) , \quad (1.34)$$

then we say that f' exists **in a weak sense**, and that $f' = g$.

Exo. 1.14 Show that the function

$$\phi(x) = \begin{cases} \exp(1/(|x|^2 - 1)) & \text{if } |x| < 1 \\ 0 & \text{if } |x| \geq 1 \end{cases} \quad (1.35)$$

belongs to $\mathcal{D}(\mathbb{R})$. By suitably shifting and scaling the argument of ϕ show that $\mathcal{D}(a, b)$ has infinite dimension for all $a < b$. (Hint: See Brenner-Scott, p. 27)

Exo. 1.15 Consider $f(x) = 1 - |x|$ in the domain $(-1, 1)$. Prove that its weak derivative is given by

$$f'(x) = \begin{cases} 1 & \text{if } x < 0 \\ -1 & \text{if } x > 0 \end{cases} . \quad (1.36)$$

Prove also that f'' does not exist. (Hint: See Brenner-Scott, p. 28)

Exo. 1.16 Let $f \in L^2(a, b)$, and let $V = H^1(a, b)$. Show that $\ell(v) = \int_a^b f(x) v(x) dx$ belongs to V' and that $\|\ell\|_{V'} \leq \|f\|_{L^2(a, b)}$.

1.3 Galerkin approximations

The previous proof suggests a numerical method, the Galerkin method, to approximate the solution of a variational problem and thus of an elliptic PDE. The idea is simply to restrict the variational problem to a subspace of V that we will denote by V_h .

Discrete variational problem (Galerkin): Find $u_h \in V_h$ such that

$$a(u_h, v_h) = \ell(v_h) \quad \forall v_h \in V_h \quad (1.37)$$

When the bilinear form $a(\cdot, \cdot)$ is symmetric and strongly coercive, this discrete problem is equivalent to

Discrete extremal problem (Galerkin): Find $u_h \in V_h$ which minimizes over V_h the function

$$J(w) = \frac{1}{2} a(w, w) - \ell(w) \quad (1.38)$$

Exo. 1.17 *Prove this last assertion.*

The natural questions that arise are:

- Does u_h exist? Is it unique?
- Does u_h approximate u (the exact solution)?
- How difficult is it to compute u_h ?

Does u_h exist? Is it unique?

Case 1) Strong coercivity of the form $a(\cdot, \cdot)$ over V

If $a(\cdot, \cdot)$ is strongly coercive over V , then

$$\inf_{0 \neq w \in V} \frac{a(w, w)}{\|w\|_V^2} = \alpha > 0.$$

If $V_h \subset V$, then $a(\cdot, \cdot)$ is strongly coercive over V_h (because the infimum is taken over a smaller set). Then u_h exists and is unique as a consequence of Exo. 1.10.

Case 2) Weak coercivity of the form $a(\cdot, \cdot)$ over V

If $a(\cdot, \cdot)$ is just weakly coercive over V , then it may or may not be weakly coercive over V_h . Compare the two following conditions

$$(A) \inf_{w \in V} \sup_{v \in V} \frac{a(w, v)}{\|w\|_V \|v\|_V} = \beta > 0, \quad (B) \inf_{w \in V_h} \sup_{v \in V_h} \frac{a(w, v)}{\|w\|_V \|v\|_V} = \beta_h > 0.$$

It is not true that (A) \Rightarrow (B) because the sup in (B) is taken over a smaller set. In this case the weak coercivity of the discrete problem must be proven independently, it is not inherited from the weak coercivity over the whole space V .

Does u_h approximate u ?

Case 1) Strong coercivity of the form $a(\cdot, \cdot)$ over V

Lemma 1.11 (J. Céa) *If $a(\cdot, \cdot)$ and $\ell(\cdot)$ are continuous in V and $a(\cdot, \cdot)$ is strongly coercive, then*

$$\|u - u_h\|_V \leq \frac{N_a}{\alpha} \|u - v_h\|_V \quad \forall v_h \in V_h \quad (1.39)$$

Proof. Notice the so-called **Galerkin orthogonality**:

$$a(u - u_h, v_h) = 0 \quad \forall v_h \in V_h \quad (1.40)$$

which implies that $a(u - u_h, u - u_h) = a(u - u_h, u - v_h)$ for all $v_h \in V_h$. Using this,

$$\|u - u_h\|_V^2 \leq \frac{1}{\alpha} a(u - u_h, u - u_h) = \frac{1}{\alpha} a(u - u_h, u - v_h) \leq \frac{N_a}{\alpha} \|u - u_h\|_V \|u - v_h\|_V \quad \forall v_h \in V_h$$

In other words, $\|u - u_h\|_V \leq C \inf_{v_h \in V_h} \|u - v_h\|_V$. \square

Let h be a real parameter, typically a “mesh size”. We say that a family $\{V_h\}_{h>0} \subset V$ satisfies the **approximability property** if:

$$\lim_{h \rightarrow 0} \text{dist}(u, V_h) = \lim_{h \rightarrow 0} \inf_{v \in V_h} \|u - v\|_V = 0 \quad (1.41)$$

Corollary 1.12 *If $a(\cdot, \cdot)$ and $\ell(\cdot)$ are continuous in V , $a(\cdot, \cdot)$ is strongly coercive, and the family $\{V_h\}_{h>0} \subset V$ satisfies (1.41), then*

$$\lim_{h \rightarrow 0} u_h = u$$

in the sense of the norm $\|\cdot\|_V$.

If the strongly coercive bilinear form is **symmetric**, then $a(\cdot, \cdot)$ is a **scalar product** over V . In this case, Galerkin orthogonality corresponds to: **The Galerkin solution u_h is the orthogonal projection of u onto V_h .**

Further, the **energy norm** can be defined

$$\|v\|_a = \sqrt{a(v, v)}, \quad (1.42)$$

which satisfies the equivalence

$$\alpha^{\frac{1}{2}} \|v\|_V \leq \|v\|_a \leq N_a^{\frac{1}{2}} \|v\|_V. \quad (1.43)$$

Exo. 1.18 Show that the Galerkin approximation is **optimal** in the energy norm,

$$\|u - u_h\|_a \leq \|u - v_h\|_a, \quad \forall v_h \in V_h, \quad (1.44)$$

without the constants that appear in Céa's lemma. Further show that the following sharper estimate holds:

$$\|u - u_h\|_V \leq \left(\frac{N_a}{\alpha}\right)^{\frac{1}{2}} \|u - v_h\|_V, \quad \forall v_h \in V_h. \quad (1.45)$$

Case 2) Weak coercivity of the form $a(\cdot, \cdot)$ over V_h

Assume now that the weak coercivity constant β_h is positive for all $h > 0$, so that u_h exists and is unique. Notice that Galerkin orthogonality still holds.

Lemma 1.13 *If $a(\cdot, \cdot)$ and $\ell(\cdot)$ are continuous in V , and $a(\cdot, \cdot)$ is weakly coercive in V_h with constant $\beta_h > 0$, then*

$$\|u - u_h\|_V \leq \left(1 + \frac{N_a}{\beta_h}\right) \|u - v_h\|_V \quad \forall v_h \in V_h \quad (1.46)$$

Proof. One begins by decomposing the error as follows (we omit the subindex V in the norm)

$$\|u - u_h\| \leq \|u - v_h\| + \|u_h - v_h\| \quad \forall v_h \in V_h \quad (1.47)$$

and then using the weak coercivity

$$\|u_h - v_h\| \leq \frac{1}{\beta_h} \sup_{w_h \in V_h} \frac{a(u_h - v_h, w_h)}{\|w_h\|} = \frac{1}{\beta_h} \sup_{w_h \in V_h} \frac{a(u - v_h, w_h)}{\|w_h\|} \leq \frac{N_a}{\beta_h} \|u - v_h\|$$

Substituting this into (1.47) one proves the claim. \square

Corollary 1.14 *Under the hypotheses of Lemma 1.13, if there exists $\beta_0 > 0$ such that $\beta_h > \beta_0$ for all h and the family $\{V_h\}_{h>0} \subset V$ satisfies (1.41), then*

$$\lim_{h \rightarrow 0} u_h = u$$

in the sense of the norm $\|\cdot\|_V$.

How difficult is it to compute u_h ?

Let us go back to our problem $-u'' + u = f$ in $(0, 1)$ with $u(0) = u(1) = 0$, which in VF requires to compute $u \in H^1(0, 1)$ satisfying the boundary conditions and such that

$$\int_0^1 [u'(x)v'(x) + u(x)v(x)] dx = \int_0^1 f(x)v(x) dx \quad (1.48)$$

Suitable spaces for the Galerkin approximation are, for example,

- \mathcal{P}_k : The polynomials of degree up to k .
- \mathcal{F}_k : The space generated by the functions $\phi^m(x) = \sin(m\pi x)$, $m = 1, 2, \dots, k$.

Exo. 1.19 Show that $a(\cdot, \cdot)$ is continuous and strongly coercive over $V = H^1(0, 1)$ with the norm

$$\|w\|_V \stackrel{\text{def}}{=} \left[\int_0^1 [w'(x)^2 + w(x)^2] dx \right]^{\frac{1}{2}}$$

Exo. 1.20 Build a small program in Matlab or Octave (or something else) that solves the Galerkin approximation of problem (1.48) considering $f = \delta_{1/4}$ and the spaces \mathcal{P}_k and/or \mathcal{F}_k , for some values of k . Compare the results to the analytical solution building plots of u and u_h . Also, build graphs of $\|u - u_h\|$ vs k .

In general, however, the construction of spaces of global basis functions, as the ones above, is not practical because it leads to dense matrices. In the next chapter we will introduce the spaces of the FEM, which are characterized by having bases with small support and thus lead to sparse matrices.

Exercises

Reading assignment: Read Chapter 1 of Duran's notes (all of it).

Exo. 1.21 Carry out the “easy computation” that shows that $\underline{\underline{A}}$ is the tridiagonal matrix such that the diagonal elements are $2/h + 2h/3$ and the extra-diagonal elements are $-1/h + h/6$ (Durán, page 3).

Exo. 1.22 Can a symmetric bilinear form be weakly coercive but not strongly coercive?

Exo. 1.23 To what variational formulation and what differential formulation corresponds the following extremal formulation?

Find $u \in V$, V consisting of functions that are smooth in $(0, 1/2)$ and $(1/2, 1)$ but can exhibit a (bounded) discontinuity at $x = 1/2$, that minimizes the function

$$J(w) = \int_0^1 [w'(x)^2 + 2w(x)^2] dx + 4 [w(1/2+) - w(1/2-)]^2 - \int_0^{1/2} 7 w(x) dx - 9w(0) \quad (1.49)$$

where $w(1/2\pm)$ represent the values on each side of the discontinuity. Notice that the space V (is it a vector space really?) has no boundary condition imposed. What are the boundary conditions of the DF at $x = 0$ and $x = 1$?

Exo. 1.24 Consider the bilinear form

$$a(u, v) = \int_0^1 u'(x) v'(x) dx.$$

Prove that this form is not strongly coercive in $H^1(0, 1)$ considering the norm

$$\|w\|_{H^1} \stackrel{\text{def}}{=} \left\{ \int_0^1 [u'(x)^2 + u(x)^2] dx \right\}^{\frac{1}{2}}$$

and that it is, with the same norm, in

$$H_0^1(0, 1) \stackrel{\text{def}}{=} \{w \in H^1(0, 1), w(0) = w(1) = 0\}.$$

1.4 Variational formulations in 2D and 3D

The ideas are similar, but we need another integration by parts formula:

Lemma 1.15 *Let $f : \Omega \rightarrow \mathbb{R}$ be an integrable function, with Ω a Lipschitz bounded open set in \mathbb{R}^d and $\partial_i f$ integrable over Ω , then*

$$\int_{\Omega} \partial_i f \, d\Omega = \int_{\partial\Omega} f n_i \, d\Gamma \quad (1.50)$$

Notice that this implies that

$$\int_{\Omega} \nabla \cdot \mathbf{v} \, d\Omega = \int_{\partial\Omega} \mathbf{v} \cdot \check{\mathbf{n}} \, d\Gamma \quad (1.51)$$

and that

$$\int_{\Omega} v \nabla^2 u \, d\Omega = \int_{\partial\Omega} v \nabla u \cdot \check{\mathbf{n}} \, d\Gamma - \int_{\Omega} \nabla v \cdot \nabla u \, d\Omega \quad (1.52)$$

We will also introduce the notation

Def. 1.16 *The Lebesgue space $L^p(\Omega)$, where $p \geq 1$, is the set of all functions such that their $L^p(\Omega)$ -norm is finite,*

$$\|w\|_{L^p(\Omega)} \stackrel{\text{def}}{=} \left[\int_{\Omega} |w(x)|^p \, dx \right]^{\frac{1}{p}} \quad (1.53)$$

Exa. 1.17 (Poisson equation) Consider the DF

$$-\nabla^2 u = f \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega \quad (1.54)$$

where ∇ is the gradient operator and $\nabla^2 u = \sum_{i=1}^d \partial_{ii}^2 u$.

A suitable variational formulation is: Find $u \in V$ such that

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

where

$$a(u, v) = \int_{\Omega} \nabla u \cdot \nabla v \, d\Omega, \quad \ell(v) = \int_{\Omega} f v \, d\Omega \quad \text{and} \quad (1.55)$$

$$V = H_0^1(\Omega) = \{w \in L^2(\Omega), \partial_i w \in L^2(\Omega) \forall i = 1, \dots, d, w = 0 \text{ on } \partial\Omega\}$$

which is a Hilbert space with the norm

$$\|w\|_{H^1} = (\|w\|_{L^2}^2 + \|\nabla w\|_{L^2}^2)^{\frac{1}{2}} \quad (1.56)$$

Exo. 1.25 Prove that if u is a solution of the DF, then it solves the VF.

Exo. 1.26 Prove that $a(\cdot, \cdot)$ is continuous in V . Prove that $\ell(\cdot)$ is continuous in V if $f \in L^2(\Omega)$. Is this last condition necessary?

Exo. 1.27 Determine the EF of the Poisson problem.

Exo. 1.28 Is $a(\cdot, \cdot)$ strongly coercive?

Exo. 1.29 Let Ω be the unit circle. Determine for which exponents γ is the function r^γ in $H^1(\Omega)$.

Exo. 1.30 Assume that the domain Ω is divided into subdomains Ω_1 and Ω_2 by a smooth internal boundary Γ . Let V consist of functions such that their restrictions to Ω_i belong to $H^1(\Omega_i)$ and that are continuous across Γ . Determine the VF corresponding to the following EF: Find $u \in V$ that minimizes

$$J(w) = \int_{\Omega_1} \frac{w^2 + \|\nabla w\|^2}{2} d\Omega + \int_{\Omega_2} \frac{3\|\nabla w\|^2}{2} d\Omega + \int_{\Gamma} (5w^2 - w) d\Gamma$$

over V .

Exo. 1.31 Determine the DF that corresponds to the previous exercise.

2 Finite element spaces and interpolation

The basic reference for what follows is Ciarlet [5]. Basically, the idea is to define finite element spaces that are locally polynomial and that contain complete polynomials of degree k in the space variables. With a judicious choice of the nodes (degrees of freedom), these piecewise polynomial functions can be made continuous by construction (if needed).

In the previous chapter it was shown that if there exists $\beta > 0$ such that, for all $w_h \in V_h$ and all $h > 0$,

$$\sup_{v_h \in V_h} \frac{a(w_h, v_h)}{\|v_h\|_V} \geq \beta \|w_h\|_V \quad (2.1)$$

then there exists $C > 0$ such that

$$\|u - u_h\|_V \leq C \inf_{v_h \in V_h} \|u - v_h\|_V \quad (2.2)$$

Notice that (2.1) is automatically satisfied if the bilinear form $a(\cdot, \cdot)$ is strongly coercive.

Denoting by $\mathcal{I}_h u$ the element-wise Lagrange interpolant of $u \in V \cap C^0(\overline{\Omega})$, it is obvious from (2.2) that

$$\|u - u_h\|_V \leq C \|u - \mathcal{I}_h u\|_V \quad (2.3)$$

The goal of this section is to introduce estimates of the interpolation error $\|u - \mathcal{I}_h u\|_V$ for some spaces V that appear in the applications.

2.1 Basic definitions

Def. 2.1 A finite element in \mathbb{R}^n is a triplet (K, P_K, Σ_K) where

- (i) K is a closed (bounded) subset of \mathbb{R}^n with a nonempty interior and Lipschitz boundary;
- (ii) P_K is a finite-dimensional space of functions defined in K , of dimension m ;

(iii) Σ_K is a set of m linear forms $\{\sigma_i\}_{i=1,\dots,m}$ which is P_K -unisolvent; i.e., if $p \in P_K$ then

$$\sigma(p) = 0 \quad \forall \sigma \in \Sigma_K \quad \Rightarrow \quad p = 0$$

It is implicitly assumed that the finite element is viewed with a larger function space $V(K)$ associated to it, in general a Sobolev space. Each $\sigma_i \in \Sigma_K$ is then assumed to be extended as an element of $V(K)'$.

Prop. 2.2 *There exists a basis $\{\mathcal{N}_i\}$ such that $\sigma_i(\mathcal{N}_j) = \delta_{ij}$.*

Whenever needed, we will write $\sigma_{K,i}$ instead of σ_i and $\mathcal{N}_{K,i}$ instead of \mathcal{N}_i to make explicit the element K being considered.

Def. 2.3 *If the degrees of freedom correspond to nodal values of the functions in $V(K)$ the element is called a Lagrange finite element. In this case, there exist X^1, \dots, X^m in K such that $\sigma_i(v) = v(X^i)$ for all $i = 1, \dots, m$.*

Exa. 2.4 *P_k elements.*

Finite elements are usually built by mapping a unique *master element* $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ onto (K, P_K, Σ_K) in a clever way. We denote by $\{\widehat{\sigma}_i\}$ the degrees of freedom of the master, and $\{\widehat{\mathcal{N}}_i\}$ the corresponding basis functions.

One begins by defining a linear bijective transformation

$$T_K : V(K) \rightarrow V(\widehat{K}) , \tag{2.4}$$

mapping functions defined on K onto functions defined on \widehat{K} . Its inverse, T_K^{-1} allows one to build P_K , i.e.,

$$P_K = \{T_K^{-1}\widehat{p}, \widehat{p} \in \widehat{P}\} . \tag{2.5}$$

Exo. 2.1 Two elements \widehat{K} , K , are said to be affine equivalent if there exists a bijective mapping $F_K : \widehat{K} \rightarrow K$ of the form

$$F_K(\hat{x}) = A_K \hat{x} + b . \quad (2.6)$$

Show that if $\widehat{P} = \mathbb{P}_k$ and T_K is defined by

$$(T_K v)(\hat{x}) = v(F_K(\hat{x})) \quad (2.7)$$

then $P_K = \mathbb{P}_k$.

This preservation of polynomial spaces makes the analysis of affine-equivalent elements much easier, but if F_K is not affine one still uses (2.7) for the definition of P_K in Lagrange finite elements (P_K will not consist of polynomials).

Prop. 2.5 If $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ is a master finite element and $T_K : V(K) \rightarrow V(\widehat{K})$ is linear and bijective, then the triplet (K, P_K, Σ_K) given by

$$K = F_K(\widehat{K}) \quad (2.8)$$

$$P_K = T_K^{-1} \widehat{P} \quad (2.9)$$

$$\Sigma_K = \{ \sigma_i \mid \sigma_i(p) = \alpha_i \widehat{\sigma}_i(T_K p), \forall p \in P_K \} \quad (2.10)$$

(where all α_i are non-zero) is a finite element. Further, the basis functions on K are given by

$$\mathcal{N}_i = \frac{1}{\alpha_i} T_K^{-1} \widehat{\mathcal{N}}_i . \quad (2.11)$$

In the case of Lagrange finite elements one takes $\alpha_i = 1$ and obtains

$$X^i = F_K(\widehat{X}^i) , \quad \sigma_i(p) = p(X^i) = p(F_K(\widehat{X}^i)) , \quad \mathcal{N}_i(F_K(\hat{x})) = \widehat{\mathcal{N}}_i(\hat{x}) . \quad (2.12)$$

Exo. 2.2 *Prove the previous proposition.*

Hint: One has to assume that \widehat{K} has Lipschitz boundary and that F_K is regular enough for $F_K(\widehat{K})$ to have Lipschitz boundary too. Because T_K is linear bijective, P_K will be a vector space of the same dimension as \widehat{P} . It remains to show that Σ_K is unisolvent. Let $p \in P_K$ such that $\sigma_i(p) = 0$ for all $i = 1, \dots, m$. Then $\widehat{\sigma}_i(T_K p) = 0$ for all i and thus $T_K p = 0$ because $\widehat{\Sigma}$ is unisolvent. The last assertion follows from $\sigma_i(\mathcal{N}_j) = \delta_{ij}$ and the case of Lagrange finite elements is a particular case of the former.

A case in which the scaling factors α_i in the previous proposition are needed is that of **Hermite finite elements**.

Exo. 2.3 *Build a basis for a cubic 1D Hermite finite element. For this, let K be an interval $[a, b]$, let $V(K) = H^2(K)$, $P_K = \mathbb{P}_3$ (cubic polynomials), and*

$$\Sigma_K = \{\theta_a, \theta_b, \eta_a, \eta_b\} , \quad (2.13)$$

where $\theta_a(v) = v(a)$, $\theta_b(v) = v(b)$, $\eta_a(v) = v'(a)$ and $\eta_b(v) = v'(b)$. Write down the basis functions.

Now consider the master element $\widehat{K} = [-1, 1]$ and the affine mapping

$$F_K(\hat{x}) = a + \frac{b-a}{2}(x+1) . \quad (2.14)$$

Defining T_K as in (2.7), find the factors $\{\alpha_i\}$ so that σ_i relates $\widehat{\sigma}_i$ according to (2.10). Write down the basis functions \widehat{N}_i and verify (2.11).

Raviart-Thomas finite element: The interpolation of vector fields in K with Lagrange finite elements is usually done one component at a time with the tools developed for scalar functions. A notable exception is the Raviart-Thomas element, very popular to approximate velocity fields in porous media. In this case $V(K)$ and P_K consist of **vector fields** and cannot be interpolated one component at a time.

Exo. 2.4 Let K be a simplex (triangle in 2D, tetrahedron in 3D), and let the space P_K be defined as

$$P_K = \mathbb{RT}_0 = (\mathbb{P}_0)^d \oplus x\mathbb{P}_0 , \quad (2.15)$$

which is of dimension $d + 1$. Defining as degrees of freedom the fluxes across each face (edge in 2D),

$$\sigma_i(p) = \int_{F_i} p \cdot \tilde{n} , \quad (2.16)$$

prove that $(K, P_K, \Sigma_K = \{\sigma_i\})$ is a finite element and that

$$\mathcal{N}_i(x) = \frac{1}{d \operatorname{meas}(K)} (x - a_i) , \quad (2.17)$$

where a_i is the vertex opposite to F_i .

The space V_K is in this case $H(\operatorname{div}, K)$ of vector fields in $L^2(K)^d$ with divergence in $L^2(K)$.

Exo. 2.5 To obtain the RT0 element from a master element one needs the mapping T_K to transform vectors normal to ∂K into vectors normal to $\partial \hat{K}$. Prove that this is accomplished by the **Piola transformation**:

$$T_K v(\hat{x}) = \det(A_K) A_K^{-1} v(F_K(\hat{x})) . \quad (2.18)$$

Def. 2.6 *The local interpolation operator $\mathcal{I}_K : V(K) \rightarrow P_K$ is defined as*

$$\mathcal{I}_K v = \sum_{i=1}^m \sigma_i(v) \mathcal{N}_i \quad \forall v \in V(K)$$

Exo. 2.6 *This interpolation is indeed a projection:*

$$\mathcal{I}_K p = p \quad \text{for all } p \in P_K . \quad (2.19)$$

Exo. 2.7 *It is also preserved by composition with the T_K mapping:*

$$\widehat{\mathcal{I}_K v} = T_K \mathcal{I}_K v = \mathcal{I}_{\widehat{K}} T_K v = \mathcal{I}_{\widehat{K}} \widehat{v}, \quad \text{for all } v \in V(K) . \quad (2.20)$$

We now turn to the problem of estimating the interpolation error, i.e., $v - \mathcal{I}_K v$.

2.2 Local $L^\infty(K)$ estimates for P_1 -triangles

We begin by considering the case of P_1 -simplices (triangles in 2D, tetrahedra in 3D). It is a good exercise in which the estimates can be derived explicitly. It is also a good excuse to introduce the multi-point Taylor formula.

Theorem 2.7 *Let K be a P_1 -element, h_K its diameter and ρ_K the radius of the largest ball contained in K . Then, for all $v \in C^2(K)$,*

$$(a) \quad \|v - \mathcal{I}_K v\|_{L^\infty(K)} \leq \frac{d^2 h_K^2}{2} \max_{|\alpha|=2} \|D^\alpha v\|_{L^\infty(K)}$$

$$(b) \quad \max_{|\alpha|=1} \|D^\alpha(v - \mathcal{I}_K v)\|_{L^\infty(K)} \leq \frac{(d+1)d^2 h_K^2}{2\rho_K} \max_{|\alpha|=2} \|D^\alpha v\|_{L^\infty(K)}$$

Proof. Let X^j be the position of the j -th node of the element, then

$$\mathcal{I}_K v(x) = \sum_{j=1}^{d+1} v(X^j) \mathcal{N}^j(x) \quad (2.21)$$

We now perform a Taylor expansion *around* x , and evaluate it at X^j , obtaining

$$v(X^j) = v(x) + \sum_{k=1}^d \frac{\partial v}{\partial x_k}(x) (X_k^j - x_k) + \frac{1}{2} \sum_{k,\ell=1}^d \frac{\partial^2 v}{\partial x_k \partial x_\ell}(\xi) (X_k^j - x_k) (X_\ell^j - x_\ell) \quad (2.22)$$

where $\xi = \eta X^j + (1 - \eta)x$ for some $\eta \in [0, 1]$. Let us denote by $p^j(x)$ the second term in the right-hand side of (2.22), and by $r^j(x)$ the third term. By direct inspection we notice that

$$|r^j(x)| \leq \frac{d^2 h_K^2}{2} \max_{|\alpha|=2} \|D^\alpha v\|_{L^\infty(K)}$$

Let us now insert $v(X^j)$ from (2.22) into (2.21) to get

$$\mathcal{I}_K v(x) = \sum_{j=1}^{d+1} v(x) \mathcal{N}^j(x) + \sum_{j=1}^{d+1} p^j(x) \mathcal{N}^j(x) + \sum_{j=1}^{d+1} r^j(x) \mathcal{N}^j(x)$$

The first term on the right is equal to $v(x)$ because $\sum_j \mathcal{N}^j = 1$. The second term vanishes, since

$$\begin{aligned} \sum_{j=1}^{d+1} \sum_{k=1}^d \frac{\partial v}{\partial x_k}(x) (X_k^j - x_k) \mathcal{N}^j(x) &= \sum_{k=1}^d \frac{\partial v}{\partial x_k}(x) \left\{ \sum_{j=1}^{d+1} X_k^j \mathcal{N}^j(x) - x_k \sum_{j=1}^{d+1} \mathcal{N}^j(x) \right\} = \\ &= \sum_{k=1}^d \frac{\partial v}{\partial x_k}(x) \{x_k - x_k\} = 0 \end{aligned}$$

As a consequence, $v(x) - \mathcal{I}_K v(x) = \sum_{j=1}^{d+1} r^j(x) \mathcal{N}^j(x)$ and thus

$$|v(x) - \mathcal{I}_K v(x)| \leq \max_j |r^j(x)| \sum_j \mathcal{N}^j(x) = \max_j |r^j(x)| \leq \frac{d^2 h_K^2}{2} \max_{|\alpha|=2} \|D^\alpha v\|_{L^\infty(K)}$$

implying assertion (a). Now, by differentiating (2.21) and using (2.22) as before, one obtains

$$\frac{\partial \mathcal{I}_K v}{\partial x_m}(x) = \sum_j v(x) \frac{\partial \mathcal{N}^j}{\partial x_m}(x) + \sum_{j,k} \frac{\partial v}{\partial x_k}(x) (X_k^j - x_k) \frac{\partial \mathcal{N}^j}{\partial x_m}(x) + \sum_{j,k} r^j(x) \frac{\partial \mathcal{N}^j}{\partial x_m}(x)$$

On the right-hand side above, the first term vanishes and the second term happens to be equal to $\frac{\partial v}{\partial x_m}(x)$, since

$$\sum_{j,k} \frac{\partial v}{\partial x_k}(x) (X_k^j - x_k) \frac{\partial \mathcal{N}^j}{\partial x_m}(x) = \sum_k \frac{\partial v}{\partial x_k}(x) \left[\sum_j X_k^j \frac{\partial \mathcal{N}^j}{\partial x_m}(x) - x_m \sum_j \frac{\partial \mathcal{N}^j}{\partial x_m}(x) \right] =$$

$$= \sum_k \frac{\partial v}{\partial x_k}(x) \frac{\partial}{\partial x_m} \sum_j X_k^j \mathcal{N}^j(x) = \sum_k \frac{\partial v}{\partial x_k}(x) \frac{\partial x_k}{\partial x_m} = \frac{\partial v}{\partial x_m}(x)$$

As a consequence

$$\left| \frac{\partial \mathcal{I}_K v}{\partial x_m}(x) - \frac{\partial v}{\partial x_m}(x) \right| = \left| \sum_{j=1}^{d+1} r^j(x) \frac{\partial \mathcal{N}^j}{\partial x_m}(x) \right| \leq \max_j |r^j(x)| \sum_{j=1}^{d+1} \left| \frac{\partial \mathcal{N}^j}{\partial x_m}(x) \right|$$

The reader can convince himself that the norm of the gradient of a P_1 basis function, which equals one at one node and zero on the opposite side/face, can never be greater than $\frac{1}{\rho_K}$, which immediately leads to assertion (b). \square

2.3 Local estimates in Sobolev norms

The previous paragraph provides us with an interpolation estimate in the norm $L^\infty(K)$ for the function and its first derivatives. Most formulations studied so far, however, have $V = H^1(\Omega)$ and we need thus estimates of $u - \mathcal{I}_K u$ in the $H^m(K)$ -norm.

2.3.1 First estimates

A simplistic approach to estimate $\|u - \mathcal{I}_K u\|_{L^2(K)}$ for P_1 elements could be

$$\|u - \mathcal{I}_K u\|_{L^2(K)}^2 = \int_K (u - \mathcal{I}_K u)^2 \leq |K| \|u - \mathcal{I}_K u\|_{L^\infty(K)}^2 \leq 4|K| h_K^4 \max_{|\alpha|=2} \|D^\alpha u\|_{L^\infty(K)}^2$$

so that, with simplified notation,

$$\|u - \mathcal{I}_K u\|_{L^2(K)} \leq 2 \sqrt{|K|} h_K^2 \|D^2 u\|_{L^\infty(K)} \quad (2.23)$$

Proceeding analogously, we obtain a first estimate for $\|\nabla u - \nabla(\mathcal{I}_K u)\|_{L^2(K)}$,

$$\|\nabla u - \nabla(\mathcal{I}_K u)\|_{L^2(K)}^2 = \int_K \sum_{i=1}^d \left[\frac{\partial(u - \mathcal{I}_K u)}{\partial x_i} \right]^2 \leq |K| \sum_{i=1}^d \left\| \frac{\partial(u - \mathcal{I}_K u)}{\partial x_i} \right\|_{L^\infty(K)}^2$$

which from Th. 2.7 implies

$$\|\nabla u - \nabla(\mathcal{I}_K u)\|_{L^2(K)} \leq \sqrt{|K|} \frac{6 d h_K^2}{\rho_K} \|D^2 u\|_{L^\infty(K)} \quad (2.24)$$

Notice that these estimates require $u \in W^{2,\infty}(K)$, which is “too much” regularity.

Exo. 2.8 Consider the function $u(x) = |x|$ and its P_1 interpolant in the 1D simplex $K = (-h/2, h/2)$. Compute $\|u - \mathcal{I}_K u\|_{L^2(K)}$ and $\|u' - (\mathcal{I}_K u)'\|_{L^2(K)}$, compare to the previous estimates, and discuss briefly.

2.3.2 An L^2 -estimate without second derivatives

If the function to be interpolated does not have second derivatives in K , then $\|u - \mathcal{I}_K u\|_{L^2(K)}$ cannot be expected to be of order $\mathcal{O}(\sqrt{|K|} h_K^2)$. The following estimate, proved in *Buscaglia & Agouzal* (IMA J. Numer. Anal. 32, 672-686, 2012), has minimal requirements on both P_K and u . Notice in particular that P_K must contain the constants but not necessarily polynomials of degree 1.

Theorem 2.8 *Assume that the basis functions $\{\mathcal{N}^j\}$ ($j = 1, \dots, d+1$) of an element K satisfy: (H1) $\mathcal{N}^j(X^k) = \delta_{jk}$, (H2) $\sum_j \mathcal{N}^j(x) = 1$, (H3) $0 \leq \mathcal{N}^j(x) \leq 1$ for all j and for all $x \in K$. Then, for all $u \in W^{1,p}(K)$ with $p > d \geq 2$,*

$$\|u - \mathcal{I}_K u\|_{L^2(K)} \leq \frac{p(d+1)}{p-d} |K|^{\frac{1}{2} - \frac{1}{p}} h_K \|\nabla u\|_{L^p(K)} \quad (2.25)$$

If ∇u is bounded we can take $p = +\infty$ to get

$$\|u - \mathcal{I}_K u\|_{L^2(K)} \leq (d+1) \sqrt{|K|} h_K \|\nabla u\|_{L^\infty(K)} \quad (2.26)$$

which is of order $\mathcal{O}(\sqrt{|K|} h_K)$.

2.3.3 Local interpolation estimates for Lagrange finite elements

Lagrange interpolation implies that the function being interpolated is at least in $C^0(K)$, since otherwise its nodal values would not be well defined.

Sobolev's imbedding theorems state that, for bounded convex domain K , $W^{m,p}(K) \subset C^0(K)$ if $mp > d$. Taking $p = 2$ (Hilbert spaces), m needs to be at least 1 in 1D and at least 2 in 2D/3D for $H^m(K)$ to consist of continuous functions.

Theorem 2.9 *Let (K, P_K, Σ_K) be a Lagrange finite element such that (a) P_K contains all polynomials of degree $\leq k$, and (b) it is affine-equivalent to the "master element" $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$. Then, the Lagrange interpolant $\mathcal{I}_K u(x) = \sum_j u(X^j) \mathcal{N}^j(x)$ satisfies*

$$\|u - \mathcal{I}_K u\|_{L^2(K)} \leq C h_K^{\ell+1} \|D^{\ell+1} u\|_{L^2(K)} \quad (2.27)$$

for all $\ell \leq k$, with C depending on ℓ but not on h_K or u .

Similarly,

$$\|u - \mathcal{I}_K u\|_{H^1(K)} = \|\nabla u - \nabla(\mathcal{I}_K u)\|_{L^2(K)} \leq C \frac{h_K^{\ell+1}}{\rho_K} \|D^{\ell+1} u\|_{L^2(K)} \quad (2.28)$$

The proof of this theorem is somewhat involved. The interested reader may refer to Ciarlet [5] or to Ern-Guermond [7].

2.4 Global interpolation error

The obtention of global interpolation estimates is quite straightforward, but needs a few definitions.

2.4.1 Considerations about meshes

A mesh \mathcal{T}_h of a domain Ω in \mathbb{R}^d is a collection of compacts (elements) K_i , $i = 1, \dots, N_e$, such that

$$\bar{\Omega} = \bigcup_{i=1}^{N_e} K_i, \quad K_i \cap K_j = \emptyset \text{ if } i \neq j, \quad \partial\Omega \subset \bigcup_{i=1}^{N_e} \partial K_i \quad (2.29)$$

Def. 2.10 *The global interpolation operator $\mathcal{I}_h : W \rightarrow W_h$, where*

$$W = \{w \in L^1(\Omega), w|_K \in V(K), \forall K \in \mathcal{T}_h\}$$

$$W_h = \{w \in L^1(\Omega), w|_K \in P_K, \forall K \in \mathcal{T}_h\}$$

by

$$\mathcal{I}_h v = \sum_{K \in \mathcal{T}_h} \sum_i \sigma_{K,i}(v|_K) \mathcal{N}_{K,i} \quad (2.30)$$

Notice that, depending on the definition of the degrees of freedom, $\mathcal{I}_h v$ may be multiple-valued at element boundaries. The mesh is said to be conforming when $\mathcal{I}_h v$ belongs to the Sobolev space W in which the variational problem is posed.

The subscript h refers to the mesh size. In fact, in error estimates one has to consider not a single mesh but a family of meshes indexed by h , and study the error as $h \rightarrow 0$. The geometrical properties of the mesh refinement enter thus into consideration. Generally, the mesh-size parameter h is defined as

$$h = \max_{K \in \mathcal{T}_h} h_K \quad (2.31)$$

For global estimates in $H^m(\Omega)$ with $m \geq 1$ the ratio $s_K = \frac{h_K}{\rho_K}$ will appear. This motivates the definition of shape-regular (or, simply, regular) meshes:

Def. 2.11 *A family of meshes \mathcal{T}_h , parameterized by the parameter $h \in H$ (where H is some subset of \mathbb{R}), is said to be **shape-regular** if there exists $S \in \mathbb{R}$ such that*

$$s_K = \frac{h_K}{\rho_K} \leq S \quad \forall K \in \mathcal{T}_h, \quad \forall h \in H \quad (2.32)$$

A shape-regular mesh (rigorously speaking, family of meshes) cannot contain needle-like elements. If the elements are triangles, no angle can tend to zero, the so-called “minimum angle condition”. This condition is known not to be necessary for the convergence of the finite element interpolant in $H^1(\Omega)$, the necessary one being that no angle in the triangulation tend to π (the so-called “maximum angle condition”).

2.4.2 From local to global

The local estimates already obtained can be turned global by simply collecting the contributions from all elements in the mesh.

Consider the estimate of Thm. 2.7(a), to begin with. One can build an $L^\infty(\Omega)$ as follows:

$$\|u - \mathcal{I}_h u\|_{L^\infty(\Omega)} = \max_K \|u - \mathcal{I}_K u\|_{L^\infty(K)} \leq \frac{d^2}{2} \max_K \{h_K^2 \|D^2 u\|_{L^\infty(K)}\} \leq \frac{d^2}{2} h^2 \|D^2 u\|_{L^\infty(\Omega)}$$

which holds without any assumption on the mesh.

Similar estimates based on local to global reasonings are left as exercises.

Exo. 2.9 Starting from Thm. 2.7(b), prove that

$$\|\nabla u - \nabla(\mathcal{I}_h u)\|_{L^\infty(\Omega)} \leq \frac{(d+1)d^2 S}{2} h \|D^2 u\|_{L^\infty(\Omega)}$$

where S is the shape-regularity constant of the mesh. Notice that it is necessary that $\nabla(\mathcal{I}_h u)$ belongs to $L^\infty(\Omega)$, which requires a conforming mesh.

Exo. 2.10 Using (2.26) prove that

$$\|u - \mathcal{I}_h u\|_{L^2(\Omega)} \leq (d+1) \sqrt{|\Omega|} h \|\nabla u\|_{L^\infty(\Omega)} \quad (2.33)$$

Exo. 2.11 Starting from (2.28) prove that, if the family of (conforming) meshes is shape-regular and the function u smooth, then

$$\|u - \mathcal{I}_h u\|_{H^1(\Omega)} \leq C S h^k \|D^{k+1} u\|_{L^2(\Omega)} \quad (2.34)$$

where S is the shape-regularity constant of the mesh.

Exo. 2.12 Assume that there exists a straight line Γ (or planar surface in 3D) in the domain Ω , at which there is a sudden change in material properties. As a consequence, $u \in H^2(\Omega \setminus \Gamma) \cap C^0(\Omega)$, but $u \notin H^2(\Omega)$. Discuss the interpolation estimate for such a function u , showing the advantages of using an “interface-fitting mesh”; i.e., a mesh such that Γ coincides with inter-element boundaries and thus does not cut any element.

2.4.3 Global estimate

Let us state a global estimate more general than the one we have been building up to now.

Theorem 2.12 *Let \mathcal{T}_h , $h > 0$, be a family of shape-regular meshes of a domain $\Omega \subset \mathbb{R}^n$. Let $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ be the (Lagrange) reference element of the mesh, all the mappings $F_K : \widehat{K} \rightarrow K$ being affine. Let \mathcal{I}_h be the global interpolation operator corresponding to \mathcal{T}_h . Assume further that $\mathbb{P}_k \subset \widehat{P}$ (i.e.; that the finite elements are “of degree k ”). Then, for each $1 \leq p < +\infty$, and for each $0 \leq \ell \leq k$, there exists C such that for all h and all $v \in W^{\ell+1,p}(\Omega)$,*

$$\|v - \mathcal{I}_h v\|_{L^p(\Omega)} + \sum_{m=1}^{\ell+1} h^m \left(\sum_{K \in \mathcal{T}_h} |v - \mathcal{I}_h v|_{W^{m,p}(K)}^p \right)^{\frac{1}{p}} \leq C h^{\ell+1} |v|_{W^{\ell+1,p}(\Omega)} \quad (2.35)$$

If $p = +\infty$,

$$\|v - \mathcal{I}_h v\|_{L^\infty(\Omega)} + \sum_{m=1}^{\ell+1} h^m \left(\max_{K \in \mathcal{T}_h} |v - \mathcal{I}_h v|_{W^{m,\infty}(K)}^p \right)^{\frac{1}{p}} \leq C h^{\ell+1} |v|_{W^{\ell+1,\infty}(\Omega)} \quad (2.36)$$

Proof. See Ern-Guermond [7], p. 61. \square

Notice that the previous theorem holds not just for simplicial elements but also for affine-equivalent quadrilaterals, hexahedra, etc.

Exo. 2.13 *Deduce from the theorem that, for P_k and Q_k elements,*

$$\|v - \mathcal{I}_h v\|_{H^1(\Omega)} \leq C h^k, \quad \|v - \mathcal{I}_h v\|_{L^2(\Omega)} \leq C h^{k+1}$$

and explain on what quantities depend the constant C .

The previous theorem establishes, in particular, that the family of spaces $\{W_h\}$ satisfies the **approximability property**.

Prop. 2.13 *For any $v \in L^p(\Omega)$, $p < +\infty$,*

$$\lim_{h \rightarrow 0} \left(\inf_{v_h \in V_h} \|v - v_h\|_{L^p(\Omega)} \right) = 0 \quad (2.37)$$

Exo. 2.14 *Prove the previous proposition.* Hint: One cannot interpolate a generic function in $L^p(\Omega)$ because it is not continuous. Fortunately, **smooth functions are dense in $L^p(\Omega)$ for all $p < +\infty$** , so that for any $\epsilon > 0$ one can find $v^\epsilon \in H^2(\Omega)$ such that $\|v - v^\epsilon\|_{L^p(\Omega)} < \epsilon$. The interpolant $\mathcal{I}_h v^\epsilon$ is well defined and Theorem 2.12 can be applied.

2.5 Inverse inequalities

Inverse inequalities are often useful in the convergence analysis of finite element methods. They provide bounds on operators that are **unbounded** in $H^m(\Omega)$, with $m > 0$, but **bounded** in V_h due to its finite-dimensionality. Intuitively, in a shape-regular mesh for a derivative $\partial u_h / \partial x_i$ to be “very large” the nodal values of the u_h must also be “very large”.

Let $(\hat{K}, \hat{P}, \hat{\Sigma})$ be the “reference” or “master” element. Let K be an element that is affine-equivalent to \hat{K} , as defined before, with $F_K : \hat{K} \rightarrow K$ the corresponding linear mapping:

$$F_K(x) = A_K x + b_K .$$

In this section we only consider finite elements for which

$$T_K v(\hat{x}) = v(F_K(\hat{x})) ,$$

such as Lagrange finite elements. In such a setting, we have

Lemma 2.14

(a)

$$|\det A_K| = \frac{|K|}{|\hat{K}|}, \quad \|A_K\| \leq \frac{h_K}{\rho_{\hat{K}}}, \quad \|A_K^{-1}\| \leq \frac{h_{\hat{K}}}{\rho_K}$$

(b) *There exists C , depending on s and p but independent of K , such that for all $v \in W^{s,p}(K)$,*

$$|\hat{v}|_{W^{s,p}(\hat{K})} \leq C \|A_K\|^s |\det A_K|^{-\frac{1}{p}} |v|_{W^{s,p}(K)} \quad (2.38)$$

$$|v|_{W^{s,p}(K)} \leq C \|A_K^{-1}\|^s |\det A_K|^{\frac{1}{p}} |\hat{v}|_{W^{s,p}(\hat{K})} \quad (2.39)$$

Proof. See, e.g., Ciarlet [5], p. 122. \square

Let us show how to take advantage of this result to prove some simple estimates.

Prop. 2.15 *There exists $C > 0$, independent of K , such that*

$$\|\nabla v_h\|_{L^2(K)} \leq \frac{C}{\rho_K} \|v_h\|_{L^2(K)} \quad (2.40)$$

for any $v_h \in P_K$.

Proof. This proof uses the so-called *scaling* argument. From (2.39) we have, taking $s = 1$ and $p = 2$,

$$\|\nabla v_h\|_{L^2(K)} \leq C \|A_K^{-1}\| |\det A_K|^{\frac{1}{2}} \|\nabla \widehat{v}_h\|_{L^2(\widehat{K})} \quad (2.41)$$

Now let us show that there exists a constant \widehat{C} such that

$$\|\nabla \widehat{v}_h\|_{L^2(\widehat{K})} \leq \widehat{C} \|\widehat{v}_h\|_{L^2(\widehat{K})} \quad (2.42)$$

For this, consider the set $\mathcal{S} = \{w \in P_K \mid \|\widehat{w}\|_{L^2(\widehat{K})} = 1\}$, which is bounded and closed in the finite-dimensional space P_K . Let \widehat{C} be the **maximum** that the **continuous** function $\|\nabla \widehat{w}\|_{L^2(\widehat{K})}$ attains in \mathcal{S} .

Then, denoting by

$$\widehat{z}_h = \frac{1}{\|\widehat{v}_h\|_{L^2(\widehat{K})}} \widehat{v}_h$$

and noticing that $\widehat{z}_h \in \mathcal{S}$, we have that

$$\|\nabla \widehat{z}_h\|_{L^2(\widehat{K})} \leq \widehat{C}$$

and thus (2.42) is proved. Inserting it into (2.41) and using (2.38) one gets

$$\|\nabla v_h\|_{L^2(K)} \leq C \widehat{C} \|A_K^{-1}\| |\det A_K|^{\frac{1}{2}} \|\widehat{v}_h\|_{L^2(\widehat{K})} \leq C^2 \widehat{C} \|A_K^{-1}\| |\det A_K|^{\frac{1}{2}} |\det A_K|^{-\frac{1}{2}} \|v_h\|_{L^2(K)} \leq$$

$$\leq \frac{(C^2 \widehat{C} h_{\widehat{K}})}{\rho_K} \|v_h\|_{L^2(K)}$$

and the proof ends noticing that the product inside the parentheses is a constant independent of K and v_h . \square

Notice that there does **not** exist a constant C that makes

$$\|\nabla v\|_{L^2(K)} \leq \frac{C}{\rho_K} \|v\|_{L^2(K)} \quad (2.43)$$

in the **infinite dimensional case**, i.e., for any v in $H^1(K)$.

Exo. 2.15 *Let K be the unit interval $(0, 1)$ in 1D. Build a sequence $\{\varphi_n\}$ of functions such that $\|\varphi_n\|_{L^2(K)} = 1$ and $\|\nabla \varphi_n\|_{L^2(K)} = n$.*

Argue that the existence of such a sequence is a counterexample to (2.43).

With a scaling argument one can prove the following discrete trace estimate.

Prop. 2.16 *There exists $C > 0$, independent of K , such that*

$$\|v_h\|_{L^2(F)} \leq C h_K^{-\frac{1}{2}} \|v_h\|_{L^2(K)} \quad \forall v_h \in P_K \quad (2.44)$$

where F is an edge (face in 3D) of K .

The proof is left as an optional exercise. Notice that, again, there is no chance of (2.44) holding for all v in an infinite-dimensional space, such as $C^\infty(K)$ for example (build a sequence that shows this!).

Several other inverse inequalities can be extracted as particular cases of the following theorem (see, e.g., [7] p. 75).

Theorem 2.17 *Let \mathcal{T}_h be a shape-regular family of meshes in $\Omega \subset \mathbb{R}^d$. Then, for $0 \leq m \leq \ell$ and $1 \leq p, q \leq \infty$, there exists a constant C such that, for all $h > 0$ and all $K \in \mathcal{T}_h$,*

$$\|v\|_{W^{\ell,p}(K)} \leq C h_K^{m-\ell+d(\frac{1}{p}-\frac{1}{q})} \|v\|_{W^{m,q}(K)} \quad (2.45)$$

for all $v \in P_K$.

This local estimate, to be made global, puts the restriction on the family of meshes that, as $h \rightarrow 0$ the diameter ratio between the largest and smaller h_K in \mathcal{T}_h remain bounded.

Def. 2.18 *A family of meshes $\{\mathcal{T}_h\}_{h>0}$ is said to be **quasi-uniform** if it is shape-regular and there exists c such that*

$$\forall h, \quad \forall K \in \mathcal{T}_h, \quad h_K \geq ch \quad (2.46)$$

Exo. 2.16 *Does the quasi-uniformity of the mesh imply the existence of $C > 0$ such that*

$$\|\nabla v_h\|_{L^2(\Omega)} \leq C h^{-1} \|v_h\|_{L^2(\Omega)} \quad \forall v_h \in V_h ? \quad (2.47)$$

Exo. 2.17 *Does the quasi-uniformity of the mesh imply the existence of $C > 0$ such that*

$$\|v_h\|_{L^2(\partial\Omega)} \leq C h^{-\frac{1}{2}} \|v_h\|_{L^2(\Omega)} \quad \forall v_h \in V_h ? \quad (2.48)$$

Let us now give a try at the easy part of the proof of Theorem 2.27.

The point of departure is the following proposition, in which the carets are inserted to remind that it is to be applied on the master element:

Prop. 2.19 *Let $\mathbb{P}_k \subset \widehat{P}$, $k \geq 1$, then there exists $\widehat{C} > 0$ such that*

$$\|\hat{u} - \mathcal{I}_{\widehat{K}}\hat{u}\|_{L^2(\widehat{K})} + \|\nabla\hat{u} - \nabla\mathcal{I}_{\widehat{K}}\hat{u}\|_{L^2(\widehat{K})} \leq \widehat{C} \|D^{k+1}\hat{u}\|_{L^2(\widehat{K})} \quad \forall u \in H^{k+1}(\widehat{K}) . \quad (2.49)$$

Exo. 2.18 *Taking Proposition 2.19 as established, prove Theorem 2.27. The strategy is a scaling argument analogous to that used in the proof of Prop. 2.15.*

Now, how to prove Prop. 2.19? Leaving the details to be read from the literature, let us just put forward the main conceptual ingredient:

Theorem 2.20 (Bramble-Hilbert lemma) *Let $F : H^{k+1}(\omega) \rightarrow \mathbb{R}$ be a continuous linear functional, satisfying*

$$F(p) = 0 , \quad \forall p \in \mathbb{P}_k . \quad (2.50)$$

Assuming $\omega \subset \mathbb{R}^d$ to be convex and bounded, with Lipschitz boundary, there exists $C(\omega) > 0$ such that

$$|F(v)| \leq C(\omega) \|D^{k+1}v\|_{L^2(\omega)} , \quad \forall v \in H^{k+1}(\omega) . \quad (2.51)$$

2.6 Interpolation in $H(\mathbf{div}, \Omega)$

The Raviart-Thomas element introduced in Exo. 2.4 has **constant normal component** on each face of K , leading to

$$\sigma_i(v_h|_K) = \text{meas}(F_i) (v_h \cdot \tilde{n})(F_i) . \quad (2.52)$$

Because $(v_h \cdot \tilde{n})(F_i)$ is single-valued for the two elements sharing face F_i , the global space W_h generated by RT_0 elements is a subspace of $H(\mathbf{div}, \Omega)$.

The interpolant $\mathcal{I}_h^{RT} : H(\mathbf{div}, \Omega) \rightarrow W_h$ is built using the local-to-global construction as before.

With the same ingredients as before (Bramble-Hilbert lemma, scaling arguments) it is possible to prove the following approximation result:

Theorem 2.21 *Let \mathcal{T}_h be a shape-regular family of triangulations. There exists $c > 0$ such that, for all h and for all $v \in H^1(\Omega)^d$ with $\nabla \cdot v \in H^1(\Omega)$,*

$$\|v - \mathcal{I}_h^{RT} v\|_{L^2(\Omega)} + \|\nabla \cdot (v - \mathcal{I}_h^{RT} v)\|_{L^2(\Omega)} \leq ch \left(\|v - \mathcal{I}_h^{RT} v\|_{H^1(\Omega)} + \|\nabla \cdot (v - \mathcal{I}_h^{RT} v)\|_{H^1(\Omega)} \right) \quad (2.53)$$

2.7 Interpolation of non-smooth functions

As already mentioned, Lagrange interpolation is not defined for arbitrary functions in $L^p(\Omega)$, not even for $H^1(\Omega)$ if $d > 1$. In applications that will be discussed later on, it is important that there exists an interpolation operator $\bar{\mathcal{I}}_h : H^1(\Omega) \rightarrow W_h$ with some useful properties:

- **Stability:** There exists $c > 0$ such that

$$\forall h, \forall v \in L^2(\Omega), \quad \|\bar{\mathcal{I}}_h v\|_{L^2(\Omega)} \leq c \|v\|_{L^2(\Omega)}. \quad (2.54)$$

$$\forall h, \forall v \in H^1(\Omega), \quad \|\bar{\mathcal{I}}_h v\|_{H^1(\Omega)} \leq c \|v\|_{H^1(\Omega)}. \quad (2.55)$$

- **Approximation:** There exists $c > 0$ such that

$$\forall h, \forall K \in \mathcal{T}_h, \forall v \in H^1(\omega_K), \quad \|v - \bar{\mathcal{I}}_h v\|_{L^2(K)} \leq ch \|v\|_{H^1(\omega_K)} \quad (2.56)$$

where ω_K consists of all elements sharing at least a node with K .

- **Preservation of boundary conditions:** If $v|_{\partial\Omega} = 0$, then $(\bar{\mathcal{I}}_h v)|_{\partial\Omega} = 0$.
- **Being a projection:** $\bar{\mathcal{I}}_h v = v$ for all $v \in W_h$.

The **Clément interpolation** operator satisfies the first two properties, while the **Scott-Zhang interpolation** operator satisfies all four.

Exo. 2.19 Read the construction of the Clément and Scott-Zhang interpolation operators, for example in Ern & Guermond, p. 68-71.