MTH651 BOOTCAMP

The topics covered by this book of notes are inspired in the class IMT3130 (2021-1) at PUC Chile, by Dr. Carlos Perez Arancibia. The original set of notes was kindly provided by Martin Guerra (UW-Madison). General references are [Cia13] for Functional Analysis essentials, [Sal09] and [Eva98] for foundational PDE's topics, and [Joh94] for general finite element theory for elliptic problems.

—

This set of notes is not officially reviewed, please do not distribute them.

Date: August 15, 2025.

Summary

1. Measure Theory and Functional Analysis essentials.

1.1. **Spaces of continuous functions.** We start by reviewing some basic notions of functional spaces. For $d \in \mathbb{N}$, we define the multi-index vector

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_d) \in \mathbb{N}^d$$

with absolute value

$$|\boldsymbol{\alpha}| = \alpha_1 + \alpha_2 + ... + \alpha_d$$

and factorial

$$\boldsymbol{\alpha}! = \alpha_1!\alpha_2!...\alpha_d!.$$

For $\boldsymbol{x} \in \mathbb{R}^d$ such that $\boldsymbol{x} = (x_1, x_2, ..., x_d)$, we define

$$\boldsymbol{x}^{\boldsymbol{\alpha}} := x_1^{\alpha_1} x_2^{\alpha_2} ... x_d^{\alpha_d} = \prod_{j=1}^{d} x_j^{\alpha_j},$$

then, if $\Omega \subset \mathbb{R}^d$ and $u : \Omega \mapsto \mathbb{R}$ is a sufficiently smooth function, we have

$$D^{\boldsymbol{\alpha}} u(\boldsymbol{x}) := \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \left(\frac{\partial}{\partial x_2}\right)^{\alpha_2} ... \left(\frac{\partial}{\partial x_d}\right)^{\alpha_d} u(x_1, x_2, ..., x_d).$$

with the convention $D^{\mathbf{0}} u = u$. Let now $\Omega \subset \mathbb{R}^d$ be an arbitrary open and bounded set and $k \in \mathbb{N}_0$. We define $\mathcal{C}^k(\Omega)$ as the space of k-times continuously differentiable functions over Ω. Analogously, we define $\mathcal{C}_b^k(\Omega)$ as the space of k-times continuously differentiable bounded functions over Ω. In other words,

$$\mathcal{C}^k(\Omega) = \left\{ u \in \mathcal{C}(\Omega) \;\middle|\; D^{\boldsymbol{\alpha}} u \in \mathcal{C}(\Omega),\ |\boldsymbol{\alpha}| = i,\ 1 \le i \le k \right\},$$

$$\mathcal{C}_b^k(\Omega) = \left\{ u \in \mathcal{C}(\Omega) \;\middle|\; \sup_{\boldsymbol{x} \in \Omega} |u(\boldsymbol{x})| < \infty \text{ and } D^{\boldsymbol{\alpha}} u \in \mathcal{C}(\Omega),\ |\boldsymbol{\alpha}| = i,\ 1 \le i \le k \right\}.$$

We can endow the space $\mathcal{C}_b^k(\Omega)$ with the norm

$$\|u\|_{\mathcal{C}_b^k(\Omega)} = \sum_{|\boldsymbol{\alpha}| \le k} \sup_{\boldsymbol{x} \in \Omega} |D^{\boldsymbol{\alpha}} u(\boldsymbol{x})| \qquad \forall u \in \mathcal{C}_b^k(\Omega)$$

The spaces

$$\mathcal{C}_b^\infty(\Omega) := \bigcap_{k=0}^{\infty} \mathcal{C}_b^k(\Omega) \quad \text{and} \quad \mathcal{C}^\infty(\Omega) := \bigcap_{k=0}^{\infty} \mathcal{C}^k(\Omega)$$

are then the space of functions which are bounded and infinite continuously differentiable over Ω, and the space of functions which are infinite continuously differentiable over Ω, respectively. Furthermore, for $u : \Omega \mapsto \mathbb{R}$ we denote by,

$$\operatorname{supp}(u) := \overline{\{\boldsymbol{x} \in \Omega \mid u(\boldsymbol{x}) \neq 0\}}$$

the support of u. This motivates the definition

$$\mathcal{C}_0^\infty(\Omega) := \{u \in \mathcal{C}^\infty(\Omega) \mid \operatorname{supp}(u) \subset \Omega\}$$

as the space of functions in $\mathcal{C}^\infty(\Omega)$ with compact support [1]. It is clear from the definitions that

$$\mathcal{C}_0^\infty(\Omega) \subseteq \mathcal{C}_b^\infty(\Omega) \subseteq \mathcal{C}^\infty(\Omega).$$

Now, for $k \in \mathbb{N}_0$ and $\eta \in (0,1]$ we define $\mathcal{C}^{k,\eta}(\Omega)$ as the space of η-Hölder continuous functions, i.e.,

$$\mathcal{C}^{k,\eta}(\Omega) := \{u \in \mathcal{C}_b^k(\Omega) \mid |D^{\boldsymbol{\alpha}}u(\boldsymbol{x}) - D^{\boldsymbol{\alpha}}u(\boldsymbol{y})| \leq L|\boldsymbol{x}-\boldsymbol{y}|^\eta, \ \ \forall \boldsymbol{x},\boldsymbol{y} \in \Omega, \ L > 0\}$$

We endow this space with the norm

$$\|u\|_{\mathcal{C}^{k,\eta}(\Omega)} := \|u\|_{\mathcal{C}_b^k(\Omega)} + \sum_{|\boldsymbol{\alpha}|=k} \sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in\Omega \\ \boldsymbol{x}\neq\boldsymbol{y}}} \frac{|D^{\boldsymbol{\alpha}}u(\boldsymbol{x}) - D^{\boldsymbol{\alpha}}u(\boldsymbol{y})|}{|\boldsymbol{x}-\boldsymbol{y}|^\eta}$$

Observe that in the special case $\eta = 1$ defines the space $\mathcal{C}^{k,1}(\Omega)$ of functions $u \in \mathcal{C}_b^k(\Omega)$ with Lipschitz continuous derivatives $D^{\boldsymbol{\alpha}}u$, $|\boldsymbol{\alpha}| = k$.

1.2. **Lipschitz Domains.** In what follows, we will require the domain of interest $\Omega \subset \mathbb{R}^d$ to be open and bounded such that its boundary, i.e.

$$\Gamma := \partial\Omega = \overline{\Omega} \cap (\mathbb{R}^d \setminus \Omega),$$

can be locally represented as the graph of a smooth function, lets say, at least Lipchitz. This way, let Ω admit the representation

$$\Omega := \{\boldsymbol{x} \in \mathbb{R}^d \mid x_d > \gamma(\tilde{\boldsymbol{x}}), \ \tilde{\boldsymbol{x}} = (x_1, ..., x_{d-1}) \in \mathbb{R}^{d-1}\}, \tag{1}$$

for some continuous function $\gamma : \mathbb{R}^{d-1} \mapsto \mathbb{R}$. So Ω is also characterized a subset of $\mathbb{R} \times \mathbb{R}^{d-1}$.

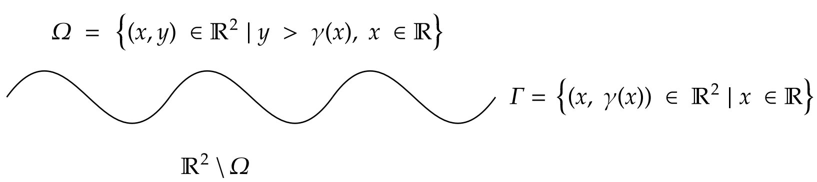


Figure 1. $\mathbb{R}^2$ Lipschitz boundary example.

If the function γ in this characterization is Lipschitz, i.e., there exist $L > 0$ such that

$$|\gamma(\tilde{\boldsymbol{x}}) - \gamma(\tilde{\boldsymbol{y}})| \leq L|\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{y}}| \qquad \forall \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}} \in \mathbb{R}^{d-1},$$

[1] Recall that compact in finite dimensions follows from closedness and boundedness

then we say that Ω is a Lipschitz epigraph with boundary

$$\Gamma = \{\boldsymbol{x} \in \mathbb{R}^d \mid x_d = \gamma(\tilde{\boldsymbol{x}}),\ \tilde{\boldsymbol{x}} \in \mathbb{R}^{d-1}\}$$

If we consider for instance the case $d = 2$, an example of a Lipschitz boundary $\Gamma = \partial\Omega$ would be as in Figure 1. A formal and more general definition of this concept is as follows.

Definition 1.1 ($\mathcal{C}^{k,\eta}$-Class boundary)**.** Let $\Omega \subset \mathbb{R}^d$ be an open bounded set. We say that $\partial\Omega = \Gamma$ is a $\mathcal{C}^{k,\eta}$-Class boundary if for any $\boldsymbol{x}_0 \in \Gamma$ there exists $r > 0$ and a function $\gamma : \mathbb{R}^{d-1} \mapsto \mathbb{R}$ with $\gamma \in \mathcal{C}^{k,\eta}(\mathbb{R}^{d-1})$, such that, up to a coordinates change, together satisfy

$$\Omega \cap B_r(\boldsymbol{x}_0) = \{\boldsymbol{x} \in B_r(\boldsymbol{x}_0) \mid x_d > \gamma(\tilde{\boldsymbol{x}}),\ \tilde{\boldsymbol{x}} \in \mathbb{R}^{d-1}\}.$$

Here, $B_r(\boldsymbol{x}_0) \subset \mathbb{R}^d$ denotes the open ball centered at $\boldsymbol{x}_0$ with radius r. In the particular case $k = 0$ and $\eta = 1$ we say that Γ is a Lipschitz boundary.

Notice that as $\partial\Omega = \Gamma$ is a compact set in $\mathbb{R}^d$, there is a finite sub cover of open balls for Γ, it then follows that there is a collection of positive constants $\{r_i\}_{i=1}^I$, a collection of points $\{\boldsymbol{x}_i\}_{i=1}^I \subset \Gamma$ such that $I < \infty$ and

$$\partial\Omega \subset \bigcup_{i=1}^{I} B_{r_i}(\boldsymbol{x}_i).$$

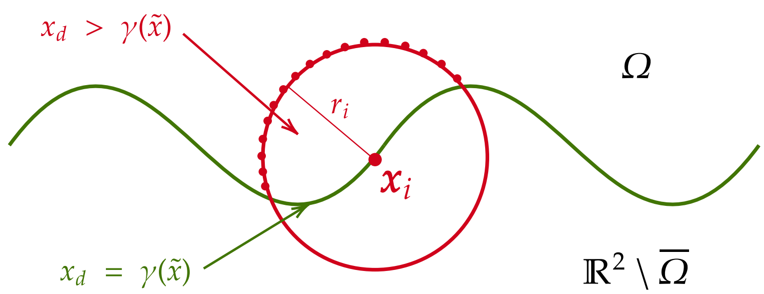


FIGURE 2. Γ open cover example.

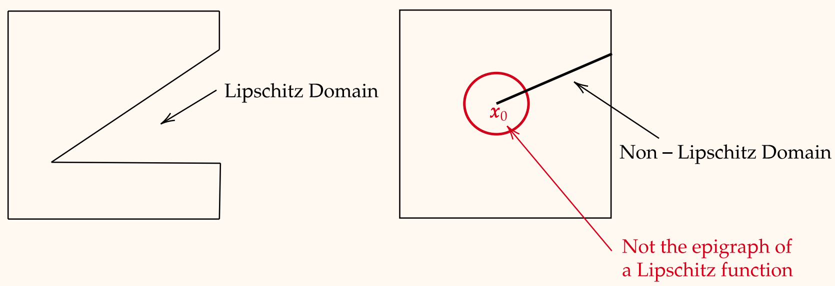
Furthermore, given the particular characterization of Ω provided in (1), we have that, for a collection $\{\gamma_i\}_{i=1}^I$ of Lipschitz continuous functions $\gamma_i : \mathbb{R}^{d-1} \mapsto \mathbb{R}$:

$$\Omega \cap B_{r_i}(\boldsymbol{x}_i) = \{\boldsymbol{x} \in B_{r_i}(\boldsymbol{x}_i) \mid x_d > \gamma_i(\tilde{\boldsymbol{x}}),\ \tilde{\boldsymbol{x}} \in \mathbb{R}^{d-1}\} \tag{2}$$

$$[\mathbb{R}^d \setminus \overline{\Omega}] \cap B_{r_i}(\boldsymbol{x}_i) = \{\boldsymbol{x} \in B_{r_i}(\boldsymbol{x}_i) \mid x_d < \gamma_i(\tilde{\boldsymbol{x}}),\ \tilde{\boldsymbol{x}} \in \mathbb{R}^{d-1}\}, \tag{3}$$

when Ω meets the conditions (2) and (3) we say that it is *locally on the same side of its boundary.* A graphic example of this can be found in Figure 2.

Example 1.1.1. We provide a simple example in $\mathbb{R}^2$ that easily shows the difference between a Lipschitz and a non-Lipschitz domain.



The interest of Lipschitz-continuous boundaries is that, even though they are not too smooth, *surface integrals* can still be defined along them and the *Green's formula* holds, this properties are going to be of particular interest in the following sections. All this considerations motivate the definition of a *domain* in $\mathbb{R}^d$.

Definition 1.2 (Domain)**.** We say that Ω is a domain in $\mathbb{R}^d$ if it is a bounded connected open subset of $\mathbb{R}^d$ with a Lipschitz continuous boundary Γ, the set Ω being locally on the same side of Γ.

1.3. L^p **spaces.** Let μ be the Lebesgue measure, and Ω be an open (thus measurable[2]) subset of $\mathbb{R}^d$. For $1 \le p \le \infty$ we define

$$L^p(\Omega) := \{u \text{ is measurable} \mid \|u\|_{L^p(\Omega)} < \infty\},$$

here, the map $u \in L^p(\Omega) \longmapsto \|u\|_{L^p(\Omega)}$ defined by

$$\|u\|^p_{L^p(\Omega)} = \int_\Omega |u|^p \, d\mu, \quad p \in [1, \infty) \quad \text{and} \quad \|u\|_{L^\infty(\Omega)} = \operatorname*{ess\,sup}_{x \in \Omega} |u(x)|,$$

is the canonical norm for the space $L^p(\Omega)$.

[2]Recall that every open set (with respect to the usual topology) belongs to the Lebesgue σ-algebra in $\mathbb{R}^d$.

Remark 1.2.1. $L^p(\Omega)$ spaces are *complete*.

For simplicity, from now on we just denote $d\mu = d\boldsymbol{x}$, also, we recall the *essential supremum* definition.

$$\operatorname*{ess\,sup}_{\boldsymbol{x}\in\Omega} |u(\boldsymbol{x})| = \inf \{C \geq 0 \mid |u| \leq C \text{ almost everywhere in } \Omega\}.$$

Next, we introduce the *Hölder's and Minkowski's inequalities.*

Theorem 1.3 (Hölder's and Minkowski's inequalities for functions)**.** *Let Ω be an open subset $\mathbb{R}^n$.*

(1) Given real number $p > 1$, let the real number q be defined by

$$\frac{1}{p} + \frac{1}{q} = 1,$$

and let $u : \Omega \mapsto [-\infty, \infty]$ and $v : \Omega \mapsto [-\infty, \infty]$ be two measurable functions that satisfy

$$\int_\Omega |u(\boldsymbol{x})|^p \, d\boldsymbol{x} < \infty \quad \textit{and} \quad \int_\Omega |v(\boldsymbol{x})|^q \, d\boldsymbol{x} < \infty.$$

*Then $uv \in L^1(\Omega)$ and **Hölder's inequality** holds:*

$$\int_\Omega |u(\boldsymbol{x})v(\boldsymbol{x})| \, d\boldsymbol{x} \leq \left(\int_\Omega |u(\boldsymbol{x})|^p \, d\boldsymbol{x}\right)^{1/p} \left(\int_\Omega |v(\boldsymbol{x})|^q \, d\boldsymbol{x}\right)^{1/q}.$$

(2) Given a real number $p \geq 1$, let $u : \Omega \mapsto [-\infty, \infty]$ and $v : \Omega \mapsto [-\infty, \infty]$ be two measurable functions that satisfy

$$\int_\Omega |u(\boldsymbol{x})|^p \, d\boldsymbol{x} < \infty \quad \textit{and} \quad \int_\Omega |v(\boldsymbol{x})|^p \, d\boldsymbol{x} < \infty.$$

*Then $(u + v) \in L^p(\Omega)$ and **Minkowski's inequality** holds:*

$$\left(\int_\Omega |u(\boldsymbol{x}) + v(\boldsymbol{x})|^p \, d\boldsymbol{x}\right)^{1/p} \leq \left(\int_\Omega |u(\boldsymbol{x})|^p \, d\boldsymbol{x}\right)^{1/p} + \left(\int_\Omega |v(\boldsymbol{x})|^p \, d\boldsymbol{x}\right)^{1/p}.$$

Proof. See [Cia13]. ■

Furthermore, for the particular case $p = 2$ the Hölder's inequality is known as the Cauchy-Schwarz inequality, in this case, one can define the L^2 *inner product* as

$$(u, v)_{L^2(\Omega)} := \int_\Omega u(\boldsymbol{x})v(\boldsymbol{x}) \, d\boldsymbol{x},$$

this inner product induces the L^2-norm in the following sense

$$\|u\|^2_{L^2(\Omega)} = (u, u)_{L^2(\Omega)}.$$

In this case we then have that $L^2(\Omega)$ is a *Hilbert space*. L^p-spaces we say that $u \equiv_\mu v \in L^p(\Omega)$ if

$$\mu(\{\boldsymbol{x} \in \Omega \mid u(\boldsymbol{x}) \neq v(\boldsymbol{x})\}) = 0$$

i.e. if they are equal almost everywhere. One can show that $\equiv_\mu$ is actually an equivalence relation. This way the space L^p considered is the set of all equivalence classes induced by $\equiv_\mu$, that also are bounded with respect to the $\|\cdot\|_{L^p(\Omega)}$ norm. So, when u is related to v under $\equiv_\mu$ we handle v, u as they are the same function in the $L^p(\Omega)$ space.

1.4. **Dual Spaces.** We now provide a formal definition of a dual space and identify some of its properties, for this so, first let X, Y be two normed vector spaces and recall the space $\mathcal{L}(X; Y)$ of continuous linear functionals with domain in X and range in Y. Also, consider the *operator norm* defined by

$$\|x'\| = \sup_{\substack{x \in X \\ x \neq 0}} \frac{\|x'(x)\|_Y}{\|x\|_X} \quad \forall x' \in \mathcal{L}(X; Y),$$

here, $\|\cdot\|_X$ and $\|\cdot\|_Y$ are the norms of X and Y respectively. Then we define a dual space as follows.

Definition 1.4 (Dual space)**.** Let X be any normed vector space over some the field $\mathbb{K} \in \{\mathbb{C}, \mathbb{R}\}$. We define

$$X' := \mathcal{L}(X; \mathbb{K}),$$

as the *dual space* of X or simply the *dual* of X, i.e. all the linear functionals $x' : X \mapsto \mathbb{K}$ that are continuous on X. Since the field $\mathbb{K}$ is complete, the space X' equipped with the operator norm, defined in this case by

$$\|x'\| = \sup_{\substack{x \in X \\ x \neq 0}} \frac{|x'(x)|}{\|x\|_X} \quad \forall x' \in X',$$

is always a *Banach space*.

Observe that an immediate consequence of the Hölder's inequality is that $L^q(\Omega)$ is the dual of $L^p(\Omega)$ for $1 \leq p < \infty$. This holds in the following sense: Given $v \in L^q(\Omega)$, we have that,

$$\ell_v(u) := \int_\Omega u(\boldsymbol{x}) v(\boldsymbol{x}) \, d\boldsymbol{x} = (u, v)_{L^2(\Omega)} \qquad \forall u \in L^p(\Omega)$$

defines an element of the dual $(L^p(\Omega))'$ with norm

$$\|\ell_v\|_{(L^p(\Omega))'} = \sup_{\substack{u\in L^p(\Omega)\\ u\neq 0}} \frac{|(u,v)_{L^2(\Omega)}|}{\|u\|_{L^p(\Omega)}} = \|v\|_{L^q(\Omega)}$$

In particular, we have that $(L^1(\Omega))' = L^\infty(\Omega)$, however, $\mathbf{L^1(\Omega)}$ **is not the dual of** $\mathbf{L^\infty(\Omega)}$.

1.5. **Generalized Derivatives.** For this section we start by defining the space

$$L^1_{\text{loc}}(\Omega) := \{u \in L^1(K) \mid \forall K \subset \Omega,\ K \text{ being compact } \}$$

Example 1.4.1. Let $\Omega = (0,1)$ and define $u : (0,1) \mapsto \mathbb{R}$ as $u(x) = \frac{1}{x}$. Observe that $u \notin L^1(\Omega)$, indeed,

$$\int_0^1 u(x)\,dx = \lim_{\varepsilon\to 0^+} \int_\varepsilon^1 \frac{1}{x}\,dx = \lim_{\varepsilon\to 0^+} \ln\left(\frac{1}{\varepsilon}\right) = \infty.$$

However, $u \in L^1_{\text{loc}}(\Omega)$, to see this it suffices to consider an arbitrary compact set $K \subset (0,1)$ of the form $K = [a,b]$ with $0 < a < b < 1$, then observe that

$$\int_K u(x)\,dx = \int_a^b \frac{1}{x}\,dx = \ln\left(\frac{b}{a}\right) < \infty$$

Thus, $u \in L^1_{\text{loc}}(\Omega)$.

The previous example shows that a function $u \in L^1_{\text{loc}}(\Omega)$ can be arbitrarily "ill-behaviored" around the boundary $\partial\Omega$. This, motivates an alternative characterization for $L^1_{\text{loc}}(\Omega)$:

$$\left[\begin{array}{c} v \in L^1_{\text{loc}}(\Omega) \text{ if } \forall \boldsymbol{x} \in \Omega \text{ there exists an open neighborhood } \Omega' \text{ of } \boldsymbol{x} \\ \text{such that } \overline{\Omega'} \subset \Omega \text{ and } v \in L^1(\Omega') \end{array}\right]$$

Now, using integration by parts, we can see that

$$\int_\Omega \frac{\partial}{\partial x_i}\varphi(\boldsymbol{x})\psi(\boldsymbol{x})\,d\boldsymbol{x} = -\int_\Omega \varphi(\boldsymbol{x})\frac{\partial}{\partial x_i}\psi(\boldsymbol{x})\,d\boldsymbol{x}, \qquad \forall \varphi,\psi \in \mathcal{C}_0^\infty(\Omega),$$

as,

$$\int_{\partial\Omega} \hat{\boldsymbol{n}}_i\varphi(\boldsymbol{x})\psi(\boldsymbol{x})\,dS = 0, \qquad \forall \varphi,\psi \in \mathcal{C}_0^\infty(\Omega)$$

This particular property motivates the following generalization of the derivative concept.

Definition 1.5 (Weak derivative)**.** We say that $u \in L^1_{\text{loc}}(\Omega)$ has a weak (or generalized) derivative with respect to x_i, if there exists $v \in L^1_{\text{loc}}(\Omega)$ satisfying

$$\int_\Omega v(\boldsymbol{x})\varphi(\boldsymbol{x})\ d\boldsymbol{x} = -\int_\Omega u(\boldsymbol{x})\frac{\partial \varphi}{\partial x_i}\ d\boldsymbol{x}, \qquad \forall \varphi \in \mathcal{C}_0^\infty(\Omega)$$

In such case we call $v = \frac{\partial u}{\partial x_i}$ the weak derivative of u with respect to x_i. Recursively, we have for $D^{\boldsymbol{\alpha}}u(\boldsymbol{x})$

$$\int_\Omega [D^{\boldsymbol{\alpha}}u(\boldsymbol{x})]\varphi(\boldsymbol{x})\ d\boldsymbol{x} = (-1)^{\boldsymbol{\alpha}}\int_\Omega u(\boldsymbol{x})D^{\boldsymbol{\alpha}}\varphi(\boldsymbol{x})\ d\boldsymbol{x}, \qquad \forall \varphi \in \mathcal{C}_0^\infty(\Omega).$$

We illustrate the advantages of this new derivative notion with the following example

Example 1.5.1. Let $\Omega := [-1, 1]$ and let $u : \Omega \mapsto \mathbb{R}$ be defined as $u(\boldsymbol{x}) = |\boldsymbol{x}|$, then, for every $\varphi \in \mathcal{C}_0^\infty(\Omega)$

$$\begin{aligned}\int_{-1}^1 u(x)\varphi'(x)\ dx &= -\int_{-1}^0 x\varphi'(x)\ dx + \int_0^1 x\varphi'(x)\ dx \\ &= (-x\varphi(x))\Big|_{-1}^0 + \int_{-1}^0 \varphi(x)\ dx + (x\varphi(x))\Big|_0^1 - \int_0^1 \varphi(x)\ dx \\ &= \varphi(-1) + \varphi(1) + \int_{-1}^0 \varphi(x)\ dx - \int_0^1 \varphi(x)\ dx,\end{aligned}$$

and furthermore $\varphi(-1) = \varphi(1) = 0$. It follows that

$$\int_{-1}^1 u(x)\varphi'(x)\ dx = \int_{-1}^0 \varphi(x)\ dx - \int_0^1 \varphi(x)\ dx = -\int_{-1}^1 \text{sign}(x)\varphi(x)\ dx,$$

where,

$$\text{sign}(x) = \begin{cases} -1, & x < 0 \\ 1, & x > 0 \\ a, & x = 0, \quad a \in \mathbb{R} \text{ arbitrary} \end{cases}$$

Given that $u \in L^1_{\text{loc}}(\Omega)$ and $\text{sign} \in L^1_{\text{loc}}(\Omega)$, we have that, the weak derivative u' of u is

$$u'(x) = \text{sign}(x)$$

We try now to calculate the second weak derivative of u, observe that

$$\begin{aligned}\int_{-1}^{1} u'(x)\varphi'(x)\ dx &= \int_{-1}^{1} \operatorname{sign}(x)\varphi'(x)\ dx \\ &= -\int_{-1}^{0} \varphi'(x)\ dx + \int_{0}^{1} \varphi'(x)\ dx \\ &= 2\varphi(0) \qquad \forall \varphi \in \mathcal{C}_0^\infty(\Omega)\end{aligned}$$

However, there is no $w \in L^1_{\text{loc}}(\Omega)$ such that

$$\int_{-1}^{1} w(x)\varphi(x)\ dx = 2\varphi(0), \qquad \forall \varphi \in \mathcal{C}_0^\infty(\Omega).$$

Indeed, suppose the contrary and consider $\{\varphi\}_{n\in\mathbb{N}} \subset \mathcal{C}_0^\infty(\Omega)$ such that

$$\varphi_n(0) = 1, \quad \|\varphi\|_\infty = 1 \quad \text{and} \quad \operatorname{supp}(\varphi_n) = \left[-\frac{1}{n}, \frac{1}{n}\right],$$

then by construction we have that

$$\lim_{n\to\infty} \int_{-1}^{1} w(x)\varphi_n(x)\ dx = 2.$$

Furthermore, notice

$$|w(x)\varphi_n(x)| \leq |w(x)| \|\varphi_n(x)\|_\infty \leq |w(x)| = L^1_{\text{loc}}(\Omega),$$

so in particular $|w(x)\varphi_n(x)| \in L^1\left(\left[-\frac{1}{n}, \frac{1}{n}\right]\right)$ for all $n \in \mathbb{N}$. It follows, by de dominated convergence theorem [Sal09],

$$\lim_{n\to\infty} \int_{-1}^{1} w(x)\varphi_n(x)\ dx = \int_{-1}^{1} w(x) \lim_{n\to\infty} \varphi_n(x)\ dx,$$

but as $\operatorname{supp}(\varphi_n) = \left[-\frac{1}{n}, \frac{1}{n}\right]$, we have $\varphi_n \xrightarrow[n\to\infty]{} 0$ pointwise, and so

$$\lim_{n\to\infty} \int_{-1}^{1} w(x)\varphi_n(x)\ dx = 0 \neq 2,$$

a contradiction. Thus, there is no $w \in L^1_{\text{loc}}(\Omega)$ such that it is the weak derivative of $\operatorname{sign}(x)$.

Now we present a key theorem that will support a lot of results in what follows.

Theorem 1.6 (Approximation by smooth functions). *Let Ω be a domain. Then $\mathcal{C}_0^\infty(\Omega)$ is dense in $L^p(\Omega)$ for all $1 \leq p < \infty$*

Proof. See [Sal09].

Lemma 1.7 (Fundamental lemma of the calculus of variations)**.** *Let Ω be a domain and let $v \in L^1_{\text{loc}}(\Omega)$. If,*

$$\int_\Omega v(\boldsymbol{x})\varphi(\boldsymbol{x})\, d\boldsymbol{x} = 0, \qquad \forall \varphi \in \mathcal{C}_0^\infty(\Omega),$$

then,

$$\mu(\{\boldsymbol{x} \in \Omega \mid v(\boldsymbol{x}) \neq 0\}) = 0,$$

in other words, $v = 0$ almost everywhere in Ω.

Proof. Let $v \in L^1_{\text{loc}}(\Omega)$ and $\boldsymbol{x}_0 \in \Omega$ such that

$$\lim_{\varepsilon \to 0^+} \frac{1}{|B_\varepsilon(\boldsymbol{x}_0)|} \int_{B_\varepsilon(\boldsymbol{x}_0)} |v(\boldsymbol{x}) - v(\boldsymbol{x}_0)|\, d\boldsymbol{x} = 0, \tag{4}$$

by the Lebesgue differentiation theorem [Sal09], this holds true almost everywhere in Ω. Now, let $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^d)$ be such that

$$\operatorname{supp}(\varphi) = \overline{B_1(\boldsymbol{0})}, \qquad \int_{\mathbb{R}^d} \varphi(\boldsymbol{x})\, d\boldsymbol{x} = 1, \quad 0 \leq \varphi(\boldsymbol{x}) \leq 1 \ \ \forall \boldsymbol{x} \in \mathbb{R}^d$$

and notice that the family

$$\varphi_\varepsilon(\boldsymbol{x}) := \frac{1}{\varepsilon^d} \varphi\left(\frac{1}{\varepsilon}(\boldsymbol{x} - \boldsymbol{x}_0)\right),$$

by construction, satisfy

$$\operatorname{supp}(\varphi_\varepsilon) = \overline{B_\varepsilon(\boldsymbol{x}_0)}, \qquad \int_{B_\varepsilon(\boldsymbol{x}_0)} \varphi_\varepsilon(\boldsymbol{x})\, d\boldsymbol{x} = 1, \quad 0 \leq \varphi_\varepsilon \leq \frac{1}{\varepsilon^d},$$

and belongs to $\mathcal{C}_0^\infty(\mathbb{R}^d)$ for every $\varepsilon > 0$. In particular, we also have

$$\int_\Omega v(\boldsymbol{x})\varphi_\varepsilon(\boldsymbol{x})\, d\boldsymbol{x} = 0$$

for all $\varepsilon > 0$, and so,

$$\begin{aligned} |v(\boldsymbol{x}_0)| &= \left| \int_\Omega v(\boldsymbol{x})\varphi_\varepsilon(\boldsymbol{x})\, d\boldsymbol{x} - v(\boldsymbol{x}_0) \right| \\ &= \left| \int_\Omega [v(\boldsymbol{x}) - v(\boldsymbol{x}_0)]\varphi_\varepsilon(\boldsymbol{x})\, d\boldsymbol{x} \right| \end{aligned}$$

$$\leq \frac{1}{\varepsilon^d}\int_\Omega |v(\boldsymbol{x}) - v(\boldsymbol{x}_0)|\, d\boldsymbol{x},$$

but observe that ε^d is proportional to $|B_\varepsilon(\boldsymbol{x}_0)|$, this way, letting $\varepsilon \to 0^+$ we get $v(\boldsymbol{x}_0) = 0$ by means of (4). As (4) holds almost everywhere in Ω, the result follows. ■

Functions with the structure of $\varphi_\varepsilon \in \mathcal{C}_0^\infty$ are known in literature as *mollifiers*, and they enjoy a wide variety of properties. A typical example of a *mollifier* is

$$\varphi_\varepsilon(\boldsymbol{x}) = \begin{cases} ce^{\frac{1}{|x|^2-1}} & |\boldsymbol{x}| < \varepsilon, \quad c > 0, \\ 0 & |\boldsymbol{x}| \geq \varepsilon. \end{cases} \qquad \forall \varepsilon > 0.$$

This functions are of central importance in *Distribution theory*, subject to be addressed in the next section. A more complete but still brief construction around *mollifiers* can be found in [Cia13] as well as in [Sal09].

Theorem 1.8. *Let $u \in L^1_{\text{loc}}(\Omega)$ be such that it admits a weak derivative, then this derivative is uniquely defined, up to sets of measure zero.*

Proof. By contradiction, suppose that u admits two (almost everywhere) different weak derivatives $v_1, v_2 \in L^1_{\text{loc}}(\Omega)$ with respect to x_i. Then,

$$\int_\Omega u(\boldsymbol{x})\frac{\partial \varphi}{\partial x_i}(\boldsymbol{x})\, d\boldsymbol{x} = -\int_\Omega v_1(\boldsymbol{x})\varphi(\boldsymbol{x})\, d\boldsymbol{x}$$

and

$$\int_\Omega u(\boldsymbol{x})\frac{\partial \varphi}{\partial x_i}(\boldsymbol{x})\, d\boldsymbol{x} = -\int_\Omega v_2(\boldsymbol{x})\varphi(\boldsymbol{x})\, d\boldsymbol{x}$$

for all $\varphi \in \mathcal{C}_0^\infty(\Omega)$. It follows that

$$\int_\Omega [v_1(\boldsymbol{x}) - v_2(\boldsymbol{x})]\varphi(\boldsymbol{x})\, d\boldsymbol{x} = 0, \quad \forall \varphi \in \mathcal{C}_0^\infty(\Omega).$$

And by lemma 1.7 we conclude

$$v_1 \equiv_\mu v_2$$

i.e. $v_1 = v_2$ almost everywhere in Ω. ■

Corollary 1.8.1. *If $v \in \mathcal{C}_b^m(\Omega)$, $m \in \mathbb{N}$, then for all $|\boldsymbol{\alpha}| \leq m$, the classical derivative $D^{\boldsymbol{\alpha}} v$ of order $\boldsymbol{\alpha}$ is also the weak derivative of order $\boldsymbol{\alpha}$.*

Proof. Immediate from theorem 1.8. ■

1.6. **Distributions.** We now endow the space $\mathcal{C}_0^\infty(\Omega)$ with a suitable notion of convergence.

Definition 1.9 (Convergence in $\mathcal{C}_0^\infty(\Omega)$)**.** Let $\{\varphi_k\}_{k\in\mathbb{N}} \subset \mathcal{C}_0^\infty(\Omega)$ and $\varphi \in \mathcal{C}_0^\infty(\Omega)$. We say that

$$\varphi_k \xrightarrow[k\to\infty]{} \varphi \quad \text{in } \mathcal{C}_0^\infty(\Omega)$$

if:

- $D^{\boldsymbol{\alpha}}\varphi_k \to D^{\boldsymbol{\alpha}}\varphi$ uniformly in Ω, for all $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_n)$.
- There exists a compact set $K \subset \Omega$ containing the support of every φ_k.

It is possible to show that this limit, defined this particular way, is **unique**. When endowed with this notion of convergence, the space $\mathcal{C}_0^\infty(\Omega)$ is often referred as the space of *test functions* or *test space*, and is denoted by $\mathscr{D}(\Omega)$ in literature.

Observe that $\mathscr{D}(\Omega)$ is not a normed space, its topology is generated over the metric

$$d(\varphi, \psi) := \sum_{k=1}^{\infty} \frac{1}{2^k} \frac{\sup_{\boldsymbol{x}\in\Omega} |\varphi^{(k)}(\boldsymbol{x}) - \psi^{(k)}(\boldsymbol{x})|}{1 + \sup_{\boldsymbol{x}\in\Omega} |\varphi^{(k)}(\boldsymbol{x}) - \psi^{(k)}(\boldsymbol{x})|} \qquad \forall \varphi, \psi \in \mathscr{D}(\Omega).$$

We focus on the linear functionals defined over $\mathscr{D}(\Omega)$, it is common to denote an arbitrary linear functional $T : \mathscr{D}(\Omega) \longmapsto \mathbb{R}$ acting over some test function $\varphi \in \mathscr{D}(\Omega)$ as $T\varphi = \langle T, \varphi\rangle$, this is known as *bracket* (or *pairing*) notation.

Furthermore, we say that T is *continuous* in $\mathscr{D}(\Omega)$ if

$$\langle T, \varphi_k\rangle \xrightarrow[k\to\infty]{} \langle T, \varphi\rangle, \quad \text{whenever} \quad \varphi_k \xrightarrow[k\to\infty]{} \varphi \; \text{ in } \mathscr{D}(\Omega).$$

Definition 1.10 (Distribution)**.** A **distribution** in Ω is a linear continuous function defined over $\mathscr{D}(\Omega)$. The set of all distributions is denoted by $\mathscr{D}'(\Omega)$, in accordance with the previously introduced *Dual space* notation.

For a function $u \in L^1_{\text{loc}}(\Omega)$, the functional

$$Tu(\varphi) = \int_\Omega u(\boldsymbol{x})\varphi(\boldsymbol{x})\, d\boldsymbol{x}, \qquad \forall \varphi \in \mathscr{D}(\Omega)$$

define a distribution over $\mathscr{D}(\Omega)$. This class of distributions are known as *regular distributions*. Observe that $u \in L^1_{\text{loc}}(\Omega)$ can be intuitively identified with $Tu \in \mathscr{D}'(\Omega)$, this way, instead of writing $Tu \in \mathscr{D}'(\Omega)$ we just write $u \in \mathscr{D}'(\Omega)$. Non-regular distributions are called *singular distributions.*

Definition 1.11 (Dirac delta)**.** The d-dimensional *Dirac delta* of the point $\boldsymbol{x}_0 \in \Omega$, denoted by $\delta_{\boldsymbol{x}_0}$, is the distribution that satisfies

$$\delta_{\boldsymbol{x}_0}\varphi = \langle \delta_{\boldsymbol{x}_0}, \varphi \rangle = \varphi(\boldsymbol{x}_0), \qquad \forall \varphi \in \mathscr{D}(\Omega)$$

Furthermore, we will say that $\{T_k\}_{k\in\mathbb{N}}$ converges to T in $\mathscr{D}'(\Omega)$ if,

$$T_k\varphi = \langle T_k, \varphi \rangle \xrightarrow[k\to\infty]{} T\varphi = \langle T, \varphi \rangle, \qquad \forall \varphi \in \mathscr{D}(\Omega).$$

Finally, we introduce the distributional derivatives, inspired in the weak derivative concept introduced before.

Definition 1.12 (Distributional derivative)**.** Let $T \in \mathscr{D}'(\Omega)$. We say that $\partial_{x_i}T$ is the distributional derivative of T if

$$\partial_{x_i}T\varphi = \langle \partial_{x_i}T, \varphi \rangle = -\langle T, \partial_{x_i}\varphi \rangle = T\partial_{x_i}\varphi \qquad \forall \varphi \in \mathscr{D}(\Omega)$$

Then, a distributional derivative of arbitrary order is defined as

$$D^{\boldsymbol{\alpha}}T\varphi = \langle D^{\boldsymbol{\alpha}}T, \varphi \rangle = (-1)^{|\boldsymbol{\alpha}|}\langle T, D^{|\boldsymbol{\alpha}|}\varphi \rangle = (-1)^{|\boldsymbol{\alpha}|}TD^{|\boldsymbol{\alpha}|}\varphi, \qquad \forall \varphi \in \mathscr{D}(\Omega)$$

for an arbitrary multi-index $\boldsymbol{\alpha}$.

We put all the construction together through the following two examples.

Example 1.12.1. Let $h_\varepsilon(x) = 1 - \chi_{[-\varepsilon,\varepsilon]}(x)$ be the characteristic function of $\mathbb{R} \setminus [-\varepsilon, \varepsilon]$. We will define the collection $\{T_\varepsilon\}_{\varepsilon>0}$ as

$$\begin{aligned} T_\varepsilon : \mathscr{D}(\mathbb{R}) &\longmapsto \mathbb{R} \\ \varphi &\longmapsto T_\varepsilon(\varphi) := \int_{\mathbb{R}} \frac{1}{x} h_\varepsilon(x)\varphi(x)\, dx \end{aligned}$$

We want to show that the limit of T_ε when $\varepsilon \to 0$ exists and is well defined in the sense of distributions $\mathscr{D}'(\Omega)$. We proceed by definition, this is, showing that

$$\lim_{\varepsilon\to 0} \langle T_\varepsilon, \varphi \rangle = \lim_{\varepsilon\to 0} \left\langle \frac{1}{x} h_\varepsilon(x), \varphi(x) \right\rangle \in \mathbb{R} \qquad \forall \varphi \in \mathscr{D}(\Omega)$$

For this so, first notice that

$$\int_{\mathbb{R}} \frac{1}{x} h_\varepsilon(x)\, dx = \int_{-\infty}^{-\varepsilon} \frac{1}{x}\, dx + \int_{\varepsilon}^{\infty} \frac{1}{x}\, dx$$
$$= \ln(|-\varepsilon|) - \ln(|\varepsilon|)$$
$$= 0$$

holds for every $\varepsilon > 0$, also, as any $\varphi \in \mathcal{C}_0^\infty$ is a compactly supported function we know that $\mathrm{supp}(\varphi) = [-a, a]$ for some $a > 0$, and so, by the mean value theorem we can formulate a Lipschitz bound as follows

$$\exists c \in [0, x] \ : \ \frac{|\varphi(x) - \varphi(0)|}{|x|} = |\varphi'(c)| \leq \sup_{z \in [-a,a]} |\varphi'(z)| =: L.$$

Observe that this holds for any $x \in [0, a]$ and can be analogously extended to $x \in [-a, 0]$. This way, the two following estimates hold

$$\left| \int_{\varepsilon}^{a} \frac{\varphi(x) - \varphi(0)}{x}\, dx \right| \leq L|a - \varepsilon|, \qquad \left| \int_{-a}^{-\varepsilon} \frac{\varphi(x) - \varphi(0)}{x}\, dx \right| \leq L|a - \varepsilon|.$$

And then we can see that

$$\int_{\mathbb{R}} \frac{h_\varepsilon(x)}{x} \varphi(x)\, dx = \int_{\mathbb{R}} \frac{h_\varepsilon(x)(\varphi(x) - \varphi(0))}{x}\, dx + \varphi(0) \underbrace{\int_{\mathbb{R}} \frac{1}{x} h_\varepsilon(x)\, dx}_{=0}$$
$$= \int_{|x| \geq \varepsilon} \frac{\varphi(x) - \varphi(0)}{x}\, dx$$
$$= \int_{[-a,-\varepsilon] \cup [\varepsilon, a]} \frac{\varphi(x) - \varphi(0)}{x}\, dx$$
$$\leq 2L|a - \varepsilon|$$
$$\xrightarrow[\varepsilon \to 0]{} 2La < \infty$$

Thus, the family $\{T_\varepsilon\}_{\varepsilon > 0} \subset \mathscr{D}'(\mathbb{R})$ converges in the sense of distributions. Its limit, defined by

$$T := \lim_{\varepsilon \to 0} T_\varepsilon : \mathscr{D}(\mathbb{R}) \longmapsto \mathbb{R}$$
$$\varphi \longmapsto \text{p.v.} \left(\frac{1}{x} \right) [\varphi] := \lim_{\varepsilon \to 0} \int_{|x| \geq \varepsilon} \frac{1}{x} \varphi(x)\, dx$$

is known in literature as the *principal value distribution.*

Example 1.12.2. On the other hand, let

$$u(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0, \end{cases}$$

then $u \in L^1_{\text{loc}}(\mathbb{R})$ and so we say $u \in \mathscr{D}'(\mathbb{R})$. We can then calculate, by definition, the distributional derivative u' of u, this is,

$$\langle u', \varphi \rangle = -\langle u, \varphi' \rangle \qquad \forall \varphi \in \mathscr{D}(\mathbb{R}).$$

We notice

$$-\langle u, \varphi' \rangle = \int_{\mathbb{R}} u(x)\varphi'(x)\,dx = \int_0^\infty \varphi'(x)\,dx = \varphi(0) = \delta_0 \varphi$$

as φ is compactly supported, i.e. $\lim_{x\to 0} \varphi(x) = 0$. This way we denote $u' = \delta_0 \in \mathscr{D}'(\mathbb{R})$.

1.7. **Sobolev spaces.** Let $k \in \mathbb{N}_0$, for an arbitrary $u \in \mathcal{C}_0^\infty(\mathbb{R}^d)$ we define the following family of norms

$$\|u\|_{W_p^k(\mathbb{R}^d)} := \begin{cases} \left(\sum_{|\boldsymbol{\alpha}| \leq k} \|D^{\boldsymbol{\alpha}} u\|^p_{L^p(\mathbb{R}^d)} \right)^{1/p}, & 1 \leq p < \infty \\ \max_{|\boldsymbol{\alpha}| \leq k} \|D^{\boldsymbol{\alpha}} u\|_{L^\infty(\mathbb{R}^d)}, & p = \infty. \end{cases} \tag{5}$$

The reader is invited to verify these are indeed norms. With this, we define the *Sobolev space* $W_p^k(\mathbb{R}^d)$.

Definition 1.13. Let $k \in \mathbb{N}$, $p \in [1, \infty]$ and $\Omega \subset \mathbb{R}^d$ be an open set. We define the Sobolev space $W_p^k(\Omega)$ as the set of all functions $v \in L^p(\Omega)$, such that, for all multi-index $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}| \leq k$, the derivative of order $\boldsymbol{\alpha}$ of v, which we denote by $D^{\boldsymbol{\alpha}} v$, not only exists in the weak sense, but also lie in the space $L^p(\Omega)$. In other words,

$$W_p^k(\Omega) := \{ v \in L^p(\Omega) \mid D^{\boldsymbol{\alpha}} v \in L^p(\Omega), \quad 0 \leq |\boldsymbol{\alpha}| \leq k \}.$$

We endow this spaces with the norms defined in (5), and furthermore, we consider the usual topology for $W_p^k(\Omega)$ to be the one induced by those norms.

An alternative, very popular, and perhaps more intuitive definition can be found in literature when $\Omega = \mathbb{R}^d$, in this particular case, **for $\mathbf{1 \leq p < \infty}$**, the Sobolev space $W_p^k(\mathbb{R}^d)$ is defined as the completion of the space $\mathcal{C}_0^\infty(\mathbb{R}^d)$ (i.e.$\mathscr{D}(\mathbb{R}^d)$) with respect to the norm $\|\cdot\|_{W_p^k(\mathbb{R}^d)}$. This is,

$$W_p^k(\mathbb{R}^d) = \overline{\mathcal{C}_0^\infty(\mathbb{R}^d)}^{\|\cdot\|_{W_p^k(\mathbb{R}^d)}}, \qquad p \in [1,\infty). \tag{6}$$

In other words, $W_p^k(\mathbb{R}^d)$ can be characterized as the closure of the space $\mathcal{C}_0^\infty(\mathbb{R}^d)$ with respect to the norm $\|\cdot\|_{W_p^k(\mathbb{R}^d)}$, i.e. if $u \in W_p^k(\mathbb{R}^d)$ then there is always a sequence $\{\varphi_j\}_{j\in\mathbb{N}} \subset \mathcal{C}_0^\infty(\mathbb{R}^d)$ such that

$$\lim_{j\to\infty} \|u - \varphi_j\|_{W_p^k(\mathbb{R}^d)} = 0.$$

Remark 1.13.1. Observe that for the particular case of characterizing the Sobolev space $W_\infty^k(\mathbb{R}^d)$, we only consider the definition 1.13, this is because $\mathcal{C}_0^\infty(\mathbb{R}^d)$ is **not** necessarily dense in $W_\infty^k(\mathbb{R}^d)$.
Furthermore, one can show that the set equality (6) remains true when $\mathbb{R}^d$ is replaced by an open subset Ω such that $\mu(\mathbb{R}^d \setminus \Omega) = 0$, this is, when the Lebesgue measure of $\mathbb{R}^d \setminus \Omega$ is zero.

The previous remark motivates a variant of a Sobolev space $W_p^k(\Omega)$, to be defined through density, and to hold for an arbitrary open set $\Omega \subset \mathbb{R}^d$.

Definition 1.14. Let Ω be any (proper or not) open subset of $\mathbb{R}^d$. For each integer $k \geq 1$ and each real number $1 \leq p < \infty$, we define

$$\mathring{W}_p^k(\Omega) := \overline{\mathcal{C}_0^\infty(\Omega)}^{\|\cdot\|_{W_p^k(\Omega)}}.$$

An immediate and alternative characterization of the space defined in 1.14 is as follows,

$$\mathring{W}_p^k(\Omega) = \left\{ v \in W_p^k(\Omega) \mid D^{\boldsymbol{\alpha}} v\big|_{\partial\Omega} = 0, \quad 0 \leq |\boldsymbol{\alpha}| \leq k-1 \right\}.$$

The reader can check the equivalence of this definitions in [Eva98].

Remark 1.14.1. For a complete and more robust result about density the reader is invited to review the Meyers-Serrin theorem (1964), see [Cia13].

It will be convenient to also introduce the family of *semi-norms*

$$|u|_{W_p^k(\mathbb{R}^d)} := \begin{cases} \left(\displaystyle\sum_{|\boldsymbol{\alpha}|=k} \|D^{\boldsymbol{\alpha}} u\|^p_{L^p(\mathbb{R}^d)} \right)^{1/p}, & 1 \le p < \infty \\ \displaystyle\max_{|\boldsymbol{\alpha}|=k} \|D^{\boldsymbol{\alpha}} u\|_{L^\infty(\mathbb{R}^d)}, & p = \infty, \end{cases} \tag{7}$$

for any $k \in \mathbb{N}_0$, arbitrary $u \in W_p^k(\Omega)$, and open set $\Omega \subset \mathbb{R}^d$. Furthermore, we distinguish the following special cases of Sobolev spaces

$$L^2(\Omega) = W_2^0(\Omega), \qquad H^1(\Omega) := W_2^1(\Omega), \qquad \mathring{H}^1(\Omega) := \mathring{W}_2^1(\Omega).$$

Remark 1.14.2. Notice that, in particular, $|\cdot|_{H^1(\Omega)}$ defines a norm in the space $\mathring{H}^1(\Omega)$.

Now we focus our efforts in showing a key property of Sobolev spaces.

Theorem 1.15 (Completeness of Sobolev spaces)**.** *The Sobolev space $W_p^k(\Omega)$ is a Banach space.*

Proof. Let $\{v_n\}_{n\in\mathbb{N}}$ be a Cauchy sequence in $W_p^k(\Omega)$. Then, for all $\boldsymbol{\alpha}$ such that $|\boldsymbol{\alpha}| \le k$, we have that $\{D^{\boldsymbol{\alpha}} v_n\}_{n\in\mathbb{N}}$ is also a Cauchy sequence in $L^p(\Omega)$, because

$$\|D^{\boldsymbol{\alpha}} v_n\|_{L^p(\Omega)} \le \|v_n\|_{W_p^k(\Omega)}, \qquad 0 \le |\boldsymbol{\alpha}| \le k.$$

And as $L^p(\Omega)$ is complete, then for all $\boldsymbol{\alpha}$ such that $|\boldsymbol{\alpha}| \le k$ there exists an element $v_{\boldsymbol{\alpha}} \in L^p(\Omega)$ such that

$$D^{\boldsymbol{\alpha}} v_n \xrightarrow[n\to\infty]{} v_{\boldsymbol{\alpha}}$$

in $L^p(\Omega)$. This holds in particular for $|\boldsymbol{\alpha}| = 0$, and so, there is $v \in L^p(\Omega)$ such that $v = \lim_{n\to\infty} v_n$, i.e.

$$\|v - v_n\|_{L^p(\Omega)} \xrightarrow[n\to\infty]{} 0.$$

This way, it suffices to show that $v \in W_p^k(\Omega)$ and that $D^{\boldsymbol{\alpha}} v = v_{\boldsymbol{\alpha}}$. For this so, we first notice that

$$\int_\Omega v_n(\boldsymbol{x}) D^{\boldsymbol{\alpha}} \varphi(\boldsymbol{x})\, d\boldsymbol{x} = (-1)^{|\boldsymbol{\alpha}|} \int_\Omega D^{\boldsymbol{\alpha}} v_n(\boldsymbol{x}) \varphi(\boldsymbol{x})\, d\boldsymbol{x} \qquad \forall \varphi \in C_0^\infty(\Omega),\ |\boldsymbol{\alpha}| \le k, \tag{8}$$

for each $v_n \in W_p^k(\Omega)$. Furthermore, $v - v_n \in L^p(\Omega)$ for all $n \in \mathbb{N}$, and so, using the Hölder inequality we can see that

$$\int_\Omega |(v_n(\boldsymbol{x}) - v(\boldsymbol{x}))D^{\boldsymbol{\alpha}}\varphi(\boldsymbol{x})| \; d\boldsymbol{x} \leq \|v - v_n\|_{L^p(\Omega)} \|D^{\boldsymbol{\alpha}}\varphi\|_{L^q(\Omega)} \xrightarrow[n\to\infty]{} 0$$

where $\frac{1}{p} + \frac{1}{q} = 1$, in other words,

$$\lim_{n\to\infty} \int_\Omega v_n(\boldsymbol{x}) D^{\boldsymbol{\alpha}}\varphi(\boldsymbol{x}) \; d\boldsymbol{x} = \int_\Omega v(\boldsymbol{x}) D^{\boldsymbol{\alpha}}\varphi(\boldsymbol{x}) \; d\boldsymbol{x}.$$

Analogously, $D^{\boldsymbol{\alpha}} v_n - v_{\boldsymbol{\alpha}} \in L^p(\Omega)$ for all $n \in \mathbb{N}$, and so,

$$\int_\Omega |(D^{\boldsymbol{\alpha}} v_n(\boldsymbol{x}) - v_{\boldsymbol{\alpha}}(\boldsymbol{x}))\varphi(\boldsymbol{x})| \; d\boldsymbol{x} \leq \|D^{\boldsymbol{\alpha}} v_n - v_{\boldsymbol{\alpha}}\|_{L^p(\Omega)} \|\varphi\|_{L^q(\Omega)} \xrightarrow[n\to\infty]{} 0.$$

Thus,

$$\lim_{n\to\infty} \int_\Omega D^{\boldsymbol{\alpha}} v_n(\boldsymbol{x}) \varphi(\boldsymbol{x}) \; d\boldsymbol{x} = \int_\Omega v_{\boldsymbol{\alpha}}(\boldsymbol{x}) \varphi(\boldsymbol{x}) \; d\boldsymbol{x}.$$

Finally, we let $n \to \infty$ at both sides of (8) to obtain,

$$\int_\Omega v(\boldsymbol{x}) D^{\boldsymbol{\alpha}}\varphi(\boldsymbol{x}) \; d\boldsymbol{x} = (-1)^{|\boldsymbol{\alpha}|} \int_\Omega v_{\boldsymbol{\alpha}}(\boldsymbol{x}) \varphi(\boldsymbol{x}) \; d\boldsymbol{x} \qquad \forall \varphi \in \mathcal{C}_0^\infty(\Omega).$$

This shows that,

$$D^{\boldsymbol{\alpha}} v = \lim_{n\to\infty} D^{\boldsymbol{\alpha}} v_n \quad \text{and} \quad v = \lim_{n\to\infty} v_n,$$

and so, $W_p^k(\Omega)$ is a Banach space. ■

Corollary 1.15.1. *The Sobolev spaces $H^k(\Omega)$ and $\mathring{H}^k(\Omega)$ are Hilbert spaces when endowed with the inner product,*

$$(u, v)_{H^k(\Omega)} = \int_\Omega \sum_{|\boldsymbol{a}| \leq k} D^{\boldsymbol{\alpha}} u(\boldsymbol{x}) D^{\boldsymbol{\alpha}} v(\boldsymbol{x}) \; d\boldsymbol{x}$$

for all $u, v \in H^k(\Omega)$.

Proof. It suffices to notice that $(u, u)_{H^k(\Omega)} = \|u\|^2_{H^k(\Omega)} = \|u\|^2_{W_2^k(\Omega)}$, then the result follows by Theorem 1.15. ■

Example 1.15.1. Let $\Omega \subset \mathbb{R}^d$ be a domain, we divide Ω in $N \in \mathbb{N}$ Lipschitz sub-domains such that

$$\overline{\Omega} = \bigcup_{n=1}^{N} \overline{\Omega}_n, \quad \overline{\Omega}_n \text{ is Lipschitz}$$

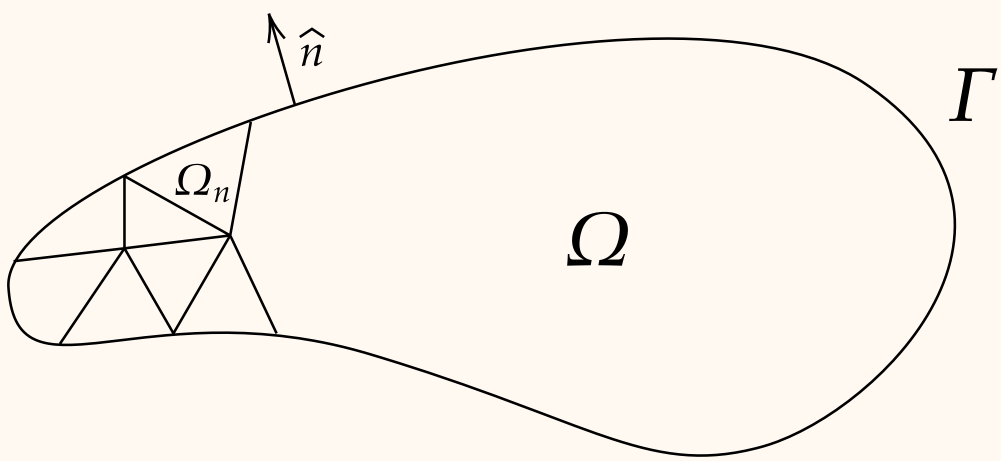


FIGURE 3. Ω subdomains example.

as shown in figure 3. We suppose that, for $p \in [1, \infty]$ we have that $v|_{\Omega_n} \in W_p^1(\Omega_n)$, $1 \leq n \leq N$, $v \in \mathcal{C}_b^k(\overline{\Omega})$. Under this assumptions we will prove that $v \in W_p^1(\Omega)$. For this so, we want to show that $\partial_{x_i} v$ exists in the weak sense along Ω and that $\partial_{x_i} v \in L^p(\Omega)$, for all $i = 1, \ldots, d$. A candidate for $\partial_{x_i} v$ is,

$$w_i(\boldsymbol{x}) = \begin{cases} \partial_{x_i} v|_{\Omega_n}(\boldsymbol{x}), & \text{if } \boldsymbol{x} \in \Omega_n \text{ for some } n \in \{1, \ldots, N\}, \\ a \in \mathbb{R}, & \text{in other case.} \end{cases}$$

Observe that the value $a \in \mathbb{R}$ is completely arbitrary. Furthermore, by assumption $v|_{\Omega_n} \in W_p^1(\Omega_n)$, so $\partial_{x_i} v|_{\Omega_n} \in L^p(\Omega_n)$ for all $n = 1, \ldots, N$. This way, $w_i \in L^p(\Omega)$, as w_i is then L^p-integrable up to a set of measure zero, i.e. $\bigcup_n \partial\Omega_n$. To conclude, it only suffices to verify that

$$\int_\Omega w_i(\boldsymbol{x}) \varphi(\boldsymbol{x})\, d\boldsymbol{x} = -\int_\Omega v(\boldsymbol{x}) \frac{\partial \varphi}{\partial x_i}(\boldsymbol{x})\, d\boldsymbol{x}, \quad \forall \varphi \in \mathcal{C}_0^\infty(\Omega),$$

for all $i = 1, \dots, d$. We notice that each boundary $\partial\Omega_n$ admits a normal vector $\hat{\boldsymbol{n}} = (\hat{n}_1, \dots, \hat{n}_d)^T$, this follows from the lipschitzness of $\partial\Omega_n$, as this implies the existence of a tangent plane, and so $\hat{\boldsymbol{n}}$ is just defined to be the vector orthogonal to the plane under the usual $\mathbb{R}^d$ inner product. Thus,

$$\begin{aligned}\int_\Omega w_i(\boldsymbol{x})\varphi(\boldsymbol{x})\, d\boldsymbol{x} &= \sum_{n=1}^{N} \int_{\Omega_n} \frac{\partial v}{\partial x_i}(\boldsymbol{x})\varphi(\boldsymbol{x})\, d\boldsymbol{x} \\ &= \sum_{n=1}^{N} \left(\int_{\partial\Omega_n} v|_{\Omega_n}(\boldsymbol{x})\varphi(\boldsymbol{x})\hat{n}_i\, dS - \int_{\Omega_n} v(\boldsymbol{x}) \frac{\partial \varphi}{\partial x_i}(\boldsymbol{x})\, d\boldsymbol{x} \right)\end{aligned}$$

But observe that

$$\sum_{n=1}^{N} \int_{\partial\Omega_n} v|_{\Omega_n}(\boldsymbol{x})\varphi(\boldsymbol{x})\hat{n}_i\, dS = 0,$$

as $\varphi = 0$ over $\partial\Omega_n \cap \partial\Omega$ and the integrals over $\partial\Omega_n \setminus \partial\Omega$ appear twice in the summation, one for each subdomain sharing that boundary, but with normal vectors $\hat{\boldsymbol{n}}$ with exact opposite sign as illustrated in figure 4, canceling themselves.

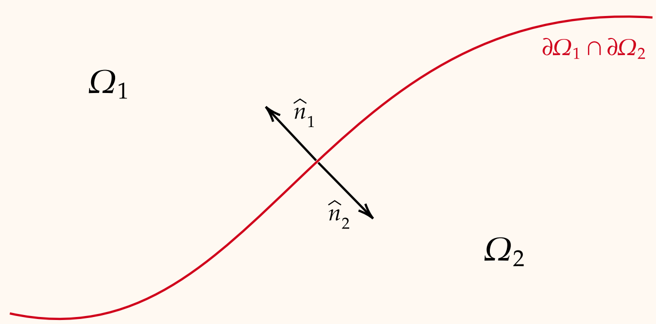


FIGURE 4. Example of normal vectors in interior boundaries.

Thus,

$$\int_\Omega w_i(\boldsymbol{x})\varphi(\boldsymbol{x})\, d\boldsymbol{x} = -\int_\Omega v(\boldsymbol{x}) \frac{\partial \varphi}{\partial x_i}(\boldsymbol{x})\, d\boldsymbol{x}, \quad \forall \varphi \in \mathcal{C}_0^\infty(\Omega),$$

as desired. We then conclude that $v \in W_p^1(\Omega)$.

Now, we can extend the Sobolev space $W_p^k(\Omega)$ with regularity order $k \in \mathbb{N}_0$ to arbitrary order $k \geq 0$.

Definition 1.16 (Sobolev spaces with $k \in \mathbb{R}_0^+$)**.** Let $s = k + \sigma$, $k \in \mathbb{N}_0$, $\sigma \in (0,1)$ and $\Omega \subset \mathbb{R}^d$. We define the Sobolev space,

$$W_p^s(\Omega) := \left\{ v \in W_p^k(\Omega) \;\middle|\; \frac{|D^{\boldsymbol{\alpha}} v(\boldsymbol{x}) - D^{\boldsymbol{\alpha}} v(\boldsymbol{y})|}{|\boldsymbol{x} - \boldsymbol{y}|^{\sigma + d/p}} \in L^p(\Omega \times \Omega), \;\; |\boldsymbol{\alpha}| = k \right\}$$

with norm,

$$\|v\|_{W_p^s(\Omega)} := \left(\|v\|_{W_p^k(\Omega)} + \sum_{|\boldsymbol{\alpha}|=k} \int_\Omega \frac{|D^{\boldsymbol{\alpha}} v(\boldsymbol{x}) - D^{\boldsymbol{\alpha}} v(\boldsymbol{y})|^p}{|\boldsymbol{x} - \boldsymbol{y}|^{\sigma p + d}} \, d\boldsymbol{x} \, d\boldsymbol{y} \right)^{1/p}.$$

In the case $p = 2$, we have that $H^s(\Omega) := W_2^s(\Omega)$ and $\mathring{H}^s(\Omega) := \mathring{W}_2^s(\Omega)$ are Hilbert spaces when endowed with the inner-product,

$$(u, v)_{H^s(\Omega)} = (u, v)_{H^k(\Omega)} + \sum_{|\boldsymbol{\alpha}|=k} \int_{\Omega \times \Omega} \frac{(D^{\boldsymbol{\alpha}} u(\boldsymbol{x}) - D^{\boldsymbol{\alpha}} u(\boldsymbol{y}))(D^{\boldsymbol{\alpha}} v(\boldsymbol{x}) - D^{\boldsymbol{\alpha}} v(\boldsymbol{y}))}{|\boldsymbol{x} - \boldsymbol{y}|^{2\sigma + d}} \, d\boldsymbol{x} \, d\boldsymbol{y}.$$

Moreover, we can extend Sobolev spaces to negative regularity order k.

Definition 1.17 (Sobolev spaces with $k < 0$)**.** Let $s \in \mathbb{R}^+$, $p \in [1, \infty)$ and q such that $\frac{1}{p} + \frac{1}{q} = 1$. We define $W_q^{-s}(\Omega)$ as the dual space of $\mathring{W}_p^s(\Omega)$, in particular, $H^{-s}(\Omega) := W_2^{-s}(\Omega)$. On the other hand, $\mathring{W}_q^{-s}(\Omega)$ is the dual space of $W_p^s(\Omega)$.

An element of a Sobolev space is often guaranteed to also belong to other function spaces, indeed, a simple example of this is to consider the inclusion

$$W_p^k(\Omega) \subset L^p(\Omega), \qquad k \geq 1,$$

which furthermore can be shown to be strict. This notion evidences in effect that a function in $L^p(\Omega)$ acquires some extra regularity or "smoothness" when its weak derivatives up to order $k \geq 1$ also lie in $L^p(\Omega)$. To extend this notion to more sophisticated cases we first introduce the concept of *embedding*[3].

Definition 1.18 (Continuous embeddings)**.** Let V, W be two normed vector spaces such that $V \subset W$. We say that V is **continuously embedded** in W,

[3]Or *imbedding* in some literature.

an we write
$$V \hookrightarrow W,$$
if there exists $c > 0$ such that,
$$\|v\|_W \le c\|v\|_V, \qquad \forall v \in V.$$

Definition 1.19 (Compact embeddings)**.** Let V, W be two normed vector spaces such that $V \subset W$. We say that V is **compactly embedded** in W, and we write
$$V \rightrightarrows W,$$
if $V \hookrightarrow W$ and any bounded sequence in V possesses a convergent subsequence in W.

From the embedding definitions we see that, when $V \hookrightarrow W$, the identity mapping

$$\iota : (V, \|\cdot\|_V) \longmapsto (W, \|\cdot\|_W)$$

is continuous. Furthermore, when $V \rightrightarrows W$, the identity mapping $u \in V \mapsto \iota(u) = u \in W$ is a *compact* linear operator, i.e. if $\{x_i\}_{i=1}^{\infty} \subset V$ is a bounded sequence, then $\{\iota(x_i)\}_{i=1}^{\infty} \subset W$ contains a convergent subsequence.

Theorem 1.20 (Rellich-Kondrachov compact embeddings)**.** *Let Ω be a domain in $\mathbb{R}^d$, let $m \ge 1$ be an integer, let $1 \le p < \infty$, and let p^* be such that $\frac{1}{p^*} = \frac{1}{p} - \frac{k}{d}$. Then the following compact embeddings hold:*
$$\begin{aligned} W_p^k(\Omega) &\rightrightarrows L^q(\Omega) \quad \textit{for all } q \textit{ with } 1 \le q < p^* \textit{ if } k < \frac{d}{p} \\ W_p^k(\Omega) &\rightrightarrows L^q(\Omega) \quad \textit{for all } q \textit{ with } 1 \le q < \infty \textit{ if } k = \frac{d}{p} \\ W_p^k(\Omega) &\rightrightarrows \mathcal{C}_b(\Omega) \qquad\qquad\qquad\qquad\qquad \textit{if } k > \frac{d}{p}. \end{aligned}$$

Proof. See [Rel30]. ■

Corollary 1.20.1. *Let Ω be a domain in $\mathbb{R}^d$, let $k \geq 1$ be an integer and let p^* be such that $\frac{1}{p^*} = \frac{1}{2} - \frac{k}{d}$. Then the following hold true:*

$$\begin{aligned} &\text{if } k < \frac{d}{2} \quad H^k(\Omega) \rightrightarrows L^q(\Omega) \quad \text{for all } q \text{ with } 1 \leq q < p^* \\ &\text{if } k = \frac{d}{2} \quad H^k(\Omega) \rightrightarrows L^q(\Omega) \quad \text{for all } q \text{ with } 1 \leq q < \infty \\ &\text{if } \frac{d}{2} < k \quad H^k(\Omega) \rightrightarrows \mathcal{C}_b(\Omega). \end{aligned}$$

Proof. Straightforward from Theorem 1.20 by letting $p = 2$. ■

Rellich-Kondrachov results are fundamental tools to start talking about norm equivalence in Sobolev spaces.

Theorem 1.21 (Norm equivalence in Sobolev spaces.)**.** *Let $\Omega \subset \mathbb{R}^d$ be a domain, and let $f \in H^{-1}(\Omega)$ be a functional such that*

$$[f(u) = 0] \Rightarrow [u \equiv_\mu 0],$$

for all ***constant*** *function u. Then the map $\|\cdot\|_{H^1(\Omega),f} : H^1(\Omega) \longmapsto \mathbb{R}_0^+$ defined by,*

$$\|v\|_{H^1(\Omega),f} := \Big(|f(v)|^2 + \|\nabla v\|^2_{L^2(\Omega)}\Big)^{1/2} \qquad \forall v \in H^1(\Omega),$$

is an equivalent norm in $H^1(\Omega)$.

Proof. We left as an exercise for the reader to verify that $\|\cdot\|_{H^1(\Omega),f}$ meet the norm axioms. We now show that $\|\cdot\|_{H^1(\Omega),f}$ is an equivalent norm in $H^1(\Omega)$. First, observe that $f : H^1(\Omega) \mapsto \mathbb{R}$ is a continuous linear functional, letting $c_f > 0$ being the continuity constant of f we can see that

$$\begin{aligned} \|v\|^2_{H^1(\Omega),f} &= |f(v)|^2 + \|\nabla v\|^2_{L^2(\Omega)} \\ &\leq c_f^2 \|v\|^2_{H^1(\Omega)} + \|\nabla v\|^2_{L^2(\Omega)} \\ &= c_f^2 (\|v\|^2_{L^2(\Omega)} + \|\nabla v\|^2_{L^2(\Omega)}) + \|\nabla v\|^2_{L^2(\Omega)} \\ &= c_f^2 \|v\|^2_{L^2(\Omega)} + (1 + c_f^2) \|\nabla v\|^2_{L^2(\Omega)} \\ &\leq (1 + c_f^2)(\|v\|^2_{L^2(\Omega)} + \|\nabla v\|^2_{L^2(\Omega)}) \\ &= (1 + c_f^2) \|v\|^2_{H^1(\Omega)}. \end{aligned}$$

This shows that

$$\|v\|_{H^1(\Omega),f} \le \sqrt{(1+c_f^2)}\|v\|_{H^1(\Omega)}.$$

Now we need to show the converse, this is, that there exists a constant $c > 0$ such that,

$$\|v\|_{H^1(\Omega)} \le c\|v\|_{H^1(\Omega),f} \qquad \forall v \in H^1(\Omega).$$

For this so, we will proceed by contradiction, let us then suppose that $\forall c > 0$, there exists $v \in H^1(\Omega)$ such that,

$$c\|v\|_{H^1(\Omega),f} < \|v\|_{H^1(\Omega)}.$$

In this case, it follows that there exists a sequence $\{v_n\}_{n\in\mathbb{N}} \subset H^1(\Omega)$ such that

$$n\|v_n\|_{H^1(\Omega),f} < \|v_n\|_{H^1(\Omega)}.$$

In other words,

$$n < \frac{\|v_n\|_{H^1(\Omega)}}{\|v_n\|_{H^1(\Omega),f}}.$$

We will consider now the normalized sequence $\{\overline{v}_n\}_{n\in\mathbb{N}}$,

$$\overline{v_n} := \frac{v_n}{\|v_n\|_{H^1(\Omega)}},$$

this way $\|\overline{v_n}\|_{H^1(\Omega)} = 1$ and

$$\|\overline{v_n}\|_{H^1(\Omega),f} = \left\|\frac{v_n}{\|v_n\|_{H^1(\Omega)}}\right\|_{H^1(\Omega),f} = \frac{\|v_n\|_{H^1(\Omega),f}}{\|v_n\|_{H^1(\Omega)}} < \frac{1}{n},$$

from which we conclude that

$$\|\overline{v_n}\|_{H^1(\Omega),f} \xrightarrow[n\to\infty]{} 0.$$

Then, from the definition of $\|\cdot\|_{H^1(\Omega),f}$, we obtain

$$|f(\overline{v_n})| \le \|\overline{v_n}\|_{H^1(\Omega),f} \qquad \text{and} \qquad \|\nabla\overline{v_n}\|_{L^2(\Omega)} \le \|\overline{v_n}\|_{H^1(\Omega),f},$$

and so,

$$|f(\overline{v_n})| \xrightarrow[n\to\infty]{} 0 \qquad \text{and} \qquad \|\nabla\overline{v_n}\|_{L^2(\Omega)} \xrightarrow[n\to\infty]{} 0.$$

On the other hand, as $\{\overline{v_n}\}_{n\in\mathbb{N}}$ is a bounded sequence, and by Rellich-Kondrachov Theorem 1.20 we know that $H^1(\Omega) \overset{c}{\hookrightarrow} L^2(\Omega)$, then there exists a subsequence $\{\overline{v_{n'}}\}_{n'\in\mathbb{N}} \subset \{\overline{v_n}\}_{n\in\mathbb{N}}$ that converges in $L^2(\Omega)$. In particular, there exists

$$\overline{v} := \lim_{n'\to\infty} \overline{v_{n'}} \in L^2(\Omega).$$

Observe now that by Cauchy-Schwartz inequality:

$$\int_\Omega |(\overline{v_{n'}} - \overline{v})(\boldsymbol{x})\partial_{x_i}\varphi(\boldsymbol{x})| \; d\boldsymbol{x} \leq \|\overline{v_{n'}} - \overline{v}\|_{L^2(\Omega)} \|\partial_{x_i}\varphi\|_{L^2(\Omega)} \xrightarrow[n\to\infty]{} 0 \qquad \forall \varphi \in \mathcal{C}_0^\infty(\Omega),$$

as $\lim_{n'\to\infty} \|\overline{v} - \overline{v_{n'}}\|_{L^2(\Omega)} = 0$, and

$$\int_\Omega |\partial_{x_i}\overline{v_{n'}}(\boldsymbol{x})\varphi(\boldsymbol{x})| \; d\boldsymbol{x} \leq \|\partial_{x_i}\overline{v_{n'}}\|_{L^2(\Omega)} \|\varphi\|_{L^2(\Omega)} \xrightarrow[n\to\infty]{} 0 \qquad \forall \varphi \in \mathcal{C}_0^\infty(\Omega), \; 1 \leq i \leq d,$$

as $\lim_{n'\to\infty} \|\nabla\overline{v_{n'}}\|_{L^2(\Omega)} = 0$, in other words,

$$\int_\Omega \overline{v}(\boldsymbol{x})\partial_{x_i}\varphi(\boldsymbol{x}) \; d\boldsymbol{x} = \lim_{n'\to\infty} \int_\Omega \overline{v_{n'}}(\boldsymbol{x})\partial_{x_i}\varphi(\boldsymbol{x}) \; d\boldsymbol{x} \qquad \forall \varphi \in \mathcal{C}_0^\infty(\Omega)$$

and,

$$\lim_{n'\to\infty} \int_\Omega \partial_{x_i}\overline{v_{n'}}(\boldsymbol{x})\varphi(\boldsymbol{x}) \; d\boldsymbol{x} = 0 \qquad \forall \varphi \in \mathcal{C}_0^\infty(\Omega).$$

All together, this implies that

$$\begin{aligned}
\int_\Omega \overline{v}(\boldsymbol{x})\partial_{x_i}\varphi(\boldsymbol{x}) \; d\boldsymbol{x} &= \lim_{n'\to\infty} \int_\Omega \overline{v_{n'}}(\boldsymbol{x})\partial_{x_i}\varphi(\boldsymbol{x}) \; d\boldsymbol{x} \\
&= -\lim_{n'\to\infty} \int_\Omega \partial_{x_i}\overline{v_{n'}}(\boldsymbol{x})\varphi(\boldsymbol{x}) \; d\boldsymbol{x} \\
&= 0
\end{aligned}$$

for all $\varphi \in C_0^\infty(\Omega)$ and $1 \leq i \leq d$. This shows, that $\overline{v} \in H^1(\Omega)$ with $\nabla\overline{v} \equiv_\mu 0 \in L^2(\Omega)$. Notice that this conditions force $\overline{v}$ to be almost everywhere constant over Ω, furthermore

$$0 \leq |f(\overline{v})| \leq \lim_{n'\to\infty} |f(\overline{v_{n'}})| = 0,$$

so $f(\overline{v}) = 0$, and by hypothesis it follows that $\overline{v} = 0$. However, this is a contradiction, as

$$\|\overline{v}\|_{H^1(\Omega)} = \lim_{n'\to\infty} \|\overline{v_{n'}}\|_{H^1(\Omega)} = 1.$$

This way, there exists $c > 0$ such that

$$c\|v\|_{H^1(\Omega)} \leq \|v\|_{H^1(\Omega),f} \leq \sqrt{1+c_f^2}\|v\|_{H^1(\Omega)}$$

for all $v \in H^1(\Omega)$. ■

Observe that the condition imposed for $f \in H^{-1}(\Omega)$ in Theorem 1.21 is weaker than injectivity.

Example 1.21.1. Provided Theorem 1.21, the following norms are equivalent in $H^1(\Omega)$:

$$\|v\|_{H^1(\Omega),\Omega} := \left(\left(\int_\Omega v(\boldsymbol{x})\, d\boldsymbol{x}\right)^2 + \|\nabla v\|^2_{L^2(\Omega)}\right)^{1/2} \qquad \forall v \in H^1(\Omega)$$

by defining $f(v) = \int_\Omega v(\boldsymbol{x})\, d\boldsymbol{x}$ and noticing $[f(\text{constant})] = 0 \Leftrightarrow [\text{constant} = 0]$. And,

$$\|v\|_{H^1(\Omega),\partial\Omega} := \left(\left(\int_{\partial\Omega} v(\boldsymbol{x})\, dS_{\boldsymbol{x}}\right)^2 + \|\nabla v\|^2_{L^2(\Omega)}\right)^{1/2} \qquad \forall v \in H^1(\Omega)$$

by defining $f(v) = \int_{\partial\Omega} v(\boldsymbol{x})\, dS_{\boldsymbol{x}}$ and noticing $[f(\text{constant})] = 0 \Leftrightarrow [\text{constant} = 0]$. Furthermore, we can now easily verify the remark 1.14.2. It then follows that there exists $c_{\text{PF}} > 0$ such that

$$\|v\|^2_{L^2(\Omega)} \leq \|v\|^2_{H^1(\Omega)} \leq c_{\text{PF}}\|v\|^2_{H^1(\Omega),\partial\Omega} \qquad \forall v \in H^1(\Omega),$$

and so, we can see that

$$\frac{1}{c_{\text{PF}}}\int_\Omega |v(\boldsymbol{x})|^2\, d\boldsymbol{x} \leq \left(\int_{\partial\Omega} v(\boldsymbol{x})\, dS_{\boldsymbol{x}}\right)^2 + \|\nabla v\|^2_{L^2(\Omega)}.$$

This inequality is widely known in literature as the *Poincaré-Friederichs inequality*. Analogously, there exists $c_{\text{P}} > 0$ such that

$$\frac{1}{c_{\text{P}}}\int_\Omega |v(\boldsymbol{x})|^2\, d\boldsymbol{x} \leq \left(\int_{\Omega} v(\boldsymbol{x})\, d\boldsymbol{x}\right)^2 + \|\nabla v\|^2_{L^2(\Omega)}.$$

This last inequality in known as the *Poincaré inequality*.

The following result is a keystone in the analysis of the finite elements method.

Lemma 1.22 (Bramble-Hilbert lemma). *Let $f \in H^{-(k+1)}(\Omega)$ with $k \in \mathbb{N}_0$ and continuity constant $c_f > 0$, and let $\mathcal{P}^k(\Omega)$ be the space of all polynomials of degree at most k defined over Ω. If f is such that*

$$f(q) = 0 \qquad \forall q \in \mathcal{P}^k(\Omega),$$

then we have that,

$$|f(v)| \leq c(c_{\mathrm{P}}) c_f |v|_{H^{k+1}(\Omega)} \qquad v \in H^{k+1}(\Omega),$$

with $c(c_{\mathrm{P}}) > 0$ depending uniquely on the Poincaré inequality constant $c_{\mathrm{p}} > 0$.

Proof. We will only show the case $k = 1$. In this case, recall that,

$$|v|^2_{H^2(\Omega)} = \sum_{|\boldsymbol{\alpha}|=2} \|D^{\boldsymbol{\alpha}} v\|^2_{L^2(\Omega)} = \sum_{i=1}^{d} \sum_{j=1}^{d} \int_\Omega \left(\frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} v(\boldsymbol{x}) \right)^2 \, d\boldsymbol{x},$$

and that,

$$\mathcal{P}^1(\Omega) = \left\{ a_0 + \sum_{i=1}^{d} a_i x_i \ : \ a_0, ..., a_d \in \mathbb{R} \right\}.$$

Now let $v \in H^2(\Omega)$ and $q \in \mathcal{P}^1$ such that $f(q) = 0$. Then,

$$|f(v)| = |f(v) + f(q)| = |f(v+q)| \leq c_f \|v + q\|_{H^2(\Omega)}.$$

Furhtermore, from the *Poincaré inequality* we have that,

$$\|v+q\|^2_{L^2(\Omega)} \leq c_{\mathrm{P}} \left(\left(\int_\Omega (v(\boldsymbol{x}) + q(\boldsymbol{x})) \, d\boldsymbol{x} \right)^2 + \|\nabla(v+q)\|^2_{L^2(\Omega)} \right).$$

On the other hand, $|q|_{H^2(\Omega)} = 0$ as $D^{\boldsymbol{\alpha}} q = 0$ for all $\boldsymbol{\alpha}$ such that $|\boldsymbol{\alpha}| = 2$. This way,

$$\begin{aligned} \|v+q\|^2_{H^2(\Omega)} &= \|v+q\|^2_{L^2(\Omega)} + \|\nabla(v+q)\|^2_{L^2(\Omega)} + |v|^2_{H^2(\Omega)} \\ &\leq c_{\mathrm{P}} \left(\int_\Omega (v(\boldsymbol{x}) + q(\boldsymbol{x})) \, d\boldsymbol{x} \right)^2 + (1 + c_{\mathrm{P}}) \|\nabla(v+q)\|^2_{L^2(\Omega)} + |v|^2_{H^2(\Omega)} \end{aligned}$$

Observe that we can apply again the *Poincaré inequality* to the term $\|\nabla(v+q)\|_{L^2(\Omega)}$,

$$\|\nabla(v+q)\|^2_{L^2(\Omega)} = \sum_{i=1}^{d} \left\| \frac{\partial}{\partial x_i}(v+q) \right\|^2_{L^2(\Omega)}$$

$$
\begin{aligned}
&\leq \sum_{i=1}^{d} c_{\mathrm{P}} \left(\left(\int_\Omega \frac{\partial (v+q)}{\partial x_i}(\boldsymbol{x}) \, d\boldsymbol{x} \right)^2 + \sum_{j=1}^{d} \left\| \frac{\partial^2 (v+q)}{\partial x_i \partial x_j} \right\|_{L^2(\Omega)}^2 \, d\boldsymbol{x} \right) \\
&= c_{\mathrm{P}} \sum_{i=1}^{d} \left(\left(\int_\Omega \frac{\partial (v+q)}{\partial x_i}(\boldsymbol{x}) \, d\boldsymbol{x} \right)^2 + \sum_{j=1}^{d} \int_\Omega \left(\frac{\partial^2 (v+q)}{\partial x_i \partial x_j}(\boldsymbol{x}) \right)^2 \, d\boldsymbol{x} \right) \\
&= c_{\mathrm{P}} \sum_{i=1}^{d} \left(\int_\Omega \frac{\partial}{\partial x_i}(v(\boldsymbol{x}) + q(\boldsymbol{x})) \, d\boldsymbol{x} \right)^2 + c_{\mathrm{P}} |v+q|_{H^2(\Omega)}^2 \\
&= c_{\mathrm{P}} \sum_{i=1}^{d} \left(\int_\Omega \frac{\partial}{\partial x_i}(v(\boldsymbol{x}) + q(\boldsymbol{x})) \, d\boldsymbol{x} \right)^2 + c_{\mathrm{P}} |v|_{H^2(\Omega)}^2.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\|v+q\|_{H^2(\Omega)}^2 \leq c_{\mathrm{P}} &\left(\int_\Omega (v(\boldsymbol{x}) + q(\boldsymbol{x})) \, d\boldsymbol{x} \right)^2 \\
&+ (1 + c_{\mathrm{P}}) c_{\mathrm{P}} \sum_{i=1}^{d} \left(\int_\Omega \frac{\partial}{\partial x_i}(v(\boldsymbol{x}) + q(\boldsymbol{x})) \, d\boldsymbol{x} \right)^2 \\
&+ (1 + (1 + c_{\mathrm{P}}) c_{\mathrm{P}}) |v|_{H^2(\Omega)}^2
\end{aligned}
$$

We now want to choose $q \in \mathcal{P}^1(\Omega)$ such that:

- $\displaystyle\int_\Omega (v(\boldsymbol{x}) + q(\boldsymbol{x})) \, d\boldsymbol{x} = 0$
- $\displaystyle\int_\Omega \frac{\partial}{\partial x_i}(v(\boldsymbol{x}) + q(\boldsymbol{x})) \, d\boldsymbol{x} = 0 \qquad \forall i = 1, ..., d$

This is, imposing restrictions over the coefficients $\{a_0, ..., a_d\}$ such that $q(\boldsymbol{x}) = a_0 + \sum_{i=1}^{d} a_i x_i$ meet the previous conditions. From the second condition, and as $\frac{\partial q}{\partial x_i} = a_i$, we ask for

$$
\int_\Omega \frac{\partial}{\partial x_i}(v(\boldsymbol{x}) + q(\boldsymbol{x})) \, d\boldsymbol{x} = \int_\Omega \frac{\partial}{\partial x_i} v(\boldsymbol{x}) \, d\boldsymbol{x} + a_i \int_\Omega \, d\boldsymbol{x} = 0
$$

i.e.

$$
a_i = -\frac{1}{|\Omega|} \int_\Omega \frac{\partial v}{\partial x_i}(\boldsymbol{x}) \, d\boldsymbol{x} \qquad \forall i = 1, ..., d.
$$

Meanwhile, from the first condition we need that

$$\int_\Omega (v(\boldsymbol{x}) + q(\boldsymbol{x}))\, d\boldsymbol{x} = \int_\Omega \left(v(\boldsymbol{x}) + a_0 + \sum_{i=1}^d a_i x_i \right)\, d\boldsymbol{x} = 0$$

i.e.

$$a_0 = -\frac{1}{|\Omega|} \int_\Omega \left(v(\boldsymbol{x}) + \sum_{i=1}^d a_i x_i \right)\, d\boldsymbol{x}.$$

So choosing $q \in \mathcal{P}^1(\Omega)$ in this fashion leads to,

$$\|v + q\|^2_{H^2(\Omega)} \leq (1 + (1 + c_\mathrm{P})c_\mathrm{P})|v|^2_{H^2(\Omega)},$$

but recall that $|f(v)| \leq c_f \|v + q\|_{H^2(\Omega)}$. So we conclude

$$|f(v)| \leq c_f \sqrt{1 + (1 + c_\mathrm{P})c_\mathrm{P}}|v|_{H^2(\Omega)},$$

as desired. Observe that this proof can be extended using similar arguments to arbitrary $k \in \mathbb{N}$. ■

1.8. **Traces.** In order to handle traditional boundary conditions in the context of partial differential equations, we need to introduce appropriate space of functions (Sobolev spaces) for functions defined over the boundary $\partial\Omega$ of a domain Ω.

Definition 1.23. Let $k \in \mathbb{N}_0$, $\eta \in (0, 1]$, $s \in [0, k + \eta]$, and Ω be a domain in $\mathbb{R}^d$ with $\mathcal{C}^{k,\eta}$-Class boundary $\Gamma := \partial\Omega$, as introduced in 1.1. Then, $\Gamma \subset \mathbb{R}^{d-1}$ is a compact set[a], and so, admits an open cover $\{D_i\}_{i=1}^I$ with $I < \infty$ (as exemplified in figure 5),

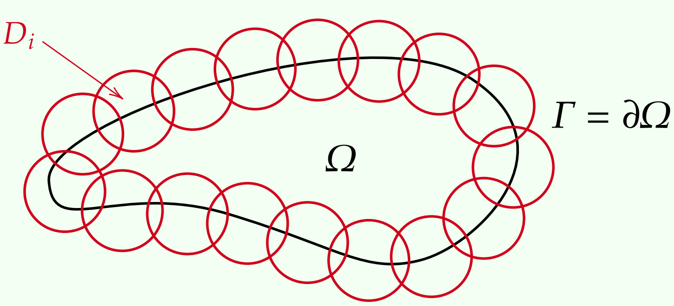


FIGURE 5. Example of a *patch system*.

such that,

$$\Gamma \cap D_i = \{x \in D_i \mid x_d = g(x_1, ..., x_{d-1})\},$$

with $g_i \in \mathcal{C}^{k,\eta}(D_i)$, for every $i = 1, ..., I$. A decomposition for Γ of this kind, i.e.

$$\Gamma = \bigcup_{i=1}^{I} g_i(D_i)$$

is called a *patch system*. In this settings, for $s \leq k + \eta$, we define the Sobolev space $W_p^s(\Gamma)$ as,

$$W_p^s(\Gamma) := \{v \in L^p(\Gamma) \mid v \circ g_i \in W_p^s(D_i), \quad i = 1, ..., I\}.$$

We endow this space with the norm,

$$\|v\|_{W_p^s(\Gamma)} := \max_{1 \leq i \leq I} \|v \circ g_i\|_{W_p^s(D_i)} \qquad \forall v \in W_p^s(\Gamma).$$

[a]Closed and bounded subset of a finite dimensional vector space.

Remark 1.23.1. When $p = 2$, we have that $H^s(\Gamma) := W_2^s(\Gamma)$ is a Hilbert space.

Before, we have introduced the Sobolev spaces as functions in $L^p(\Omega)$ with additional $L^p(\Omega)$-regularity imposed to the weak derivatives. Recall that L^p spaces are intrinsically sets of equivalence classes of functions induced by the Lebesgue measure, so, this functions are unique up to a set of measure zero.

In this sense, if $\Omega \subset \mathbb{R}^d$ we can see that $\partial\Omega \subset \mathbb{R}^{d-1}$ is a set of measure zero, i.e., $\mu(\partial\Omega) = 0$. Then, it is clear that we face a problem when restricting functions in $W_p^s(\Omega)$ to the boundary, as $\partial\Omega$ having measure zero implies that functions with different values over the boundary are indistinguishable for $\equiv_\mu$.

To handle this problem, we need to introduce an operator that gives sense to $u|_{\partial\Omega}$ for $u \in W_p^s(\Omega)$, this operator, to be introduced next, is known in literature as the *trace operator*.

Theorem 1.24 (Trace Operator). *Assume that Ω is a Lipschitz class domain in $\mathbb{R}^d$ and that $1 \leq p \leq \infty$. Then, there exists a linear and continuous operator,*

$$\gamma_0 : W_p^1(\Omega) \longmapsto L^p(\partial\Omega)$$

with the following properties,

(1) $\gamma_0 v = v|_{\partial\Omega}$ if $v \in W_p^1(\Omega) \cap \mathcal{C}(\Omega)$.

(2) There exists a constant $c > 0$ such that,

$$\|\gamma_0 v\|_{L^p(\partial\Omega)} \leq c\|v\|_{W_p^1(\Omega)}, \qquad \forall v \in W_p^1(\Omega).$$

(3) $\gamma_0 : W_p^1(\Omega) \longmapsto L^p(\partial\Omega)$ is compact, in other words, for all sequence $\{v_n\}_{n\in\mathbb{N}} \subset W_p^1(\Omega)$ there exists a subsequence $\{v_{\prime n}\}_{n'\in\mathbb{N}} \subset \{v_n\}_{n\in\mathbb{N}}$ such that $\{\gamma_0 v_{\prime n}\}_{n'\in\mathbb{N}} \subset L^p(\partial\Omega)$ is convergent in $L^p(\partial\Omega)$.

Meeting these properties, the operator γ_0 is then called trace operator.

Proof. Refer to [Bra01]. ■

Remark 1.24.1. $\gamma_0 : W_p^1(\Omega) \longmapsto L^p(\partial\Omega)$ is neither injective nor surjective. In fact

$$\gamma_0(H^1(\Omega)) = H^{1/2}(\partial\Omega)$$

moreover,

$$\|g\|_{H^{1/2}(\partial\Omega)} := \inf_{\substack{v\in H^1(\Omega) \\ \gamma_0 v = g}} \|v\|_{H^1(\Omega)}$$

defines an equivalent norm in $H^{1/2}(\partial\Omega)$, often called *minimal extension norm.* In general, we have that

$$\gamma_0(W_p^1(\Omega)) = W_p^{1-1/p}(\partial\Omega) \subset L^p(\partial\Omega).$$

2. Elliptic Variational Formulations.

2.1. **Symmetric case.** We want to study existence and uniqueness of solutions of elliptic problems in Sobolev spaces. For this so, we use as a model problem the Poisson equation

$$(P)\begin{cases} -\Delta u = f, & \text{in } \Omega \\ u = 0, & \text{on } \partial\Omega. \end{cases}$$

Here, we assume that $\Omega \subset \mathbb{R}^d$ is a Lipschitz domain with boundary $\partial\Omega$. Let us suppose now that $u \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\overline{\Omega})$. Then, if $v \in \mathcal{C}_0^\infty(\Omega)$, using integration by parts, we have that:

$$-\int_\Omega \Delta u v \, d\boldsymbol{x} = -\int_{\partial\Omega} \frac{\partial u}{\partial \hat{\boldsymbol{n}}} v \, dS + \int_\Omega \nabla u \cdot \nabla v \, d\boldsymbol{x} = \int_\Omega f v \, d\boldsymbol{x}$$

But as $v \in \mathcal{C}_0^\infty(\Omega)$ is compactly supported we see that,

$$\int_\Omega \nabla u \cdot \nabla v \, d\boldsymbol{x} = \int_\Omega f v \, d\boldsymbol{x} \qquad \forall v \in \mathcal{C}_0^\infty(\Omega).$$

Notice then that for this identity to make sense, it doesn't necessarily need that $u \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\overline{\Omega})$, $f \in \mathcal{C}(\overline{\Omega})$ and $v \in \mathcal{C}_0^\infty(\Omega)$. In fact, it suffices to ask for

$$u, v \in \mathring{H}^1(\Omega) = \overline{\mathcal{C}_0^\infty(\Omega)}^{\|\cdot\|_{H^1(\Omega)}} \quad \text{and} \quad f \in L^2(\Omega),$$

moreover, we can relax the second condition to $f \in (\mathring{H}^1(\Omega))' = H^{-1}(\Omega)$. This way, we define the *variational formulation* or *weak form* of the Poisson problem as:

Find $u \in \mathring{H}^1(\Omega)$ such that,

$$(PV) \qquad \int_\Omega \nabla u \cdot \nabla v \, d\boldsymbol{x} = \int_\Omega f v \, d\boldsymbol{x}, \qquad \forall v \in \mathring{H}^1(\Omega).$$

Observe that it is necessary that $u \in \mathring{H}^1(\Omega)$ so we guarantee $\gamma_0 u = 0$, where $\gamma_0 : \mathring{H}^1(\Omega) \longmapsto H^{1/2}(\partial\Omega)$ is the *trace operator*.

It is not hard to see that the solution of the strong problem (P), $u \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\overline{\Omega}) \subset \mathring{H}^1(\Omega)$, is also a solution of the weak problem. On the other hand, if $u \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\overline{\Omega})$ is solution of the weak problem and $f \in C(\overline{\Omega})$, we have that,

$$\int_\Omega \nabla u \cdot \nabla v \, d\boldsymbol{x} = \int_\Omega f v \, d\boldsymbol{x}, \qquad \forall v \in \mathring{H}^1(\Omega).$$

Integrating by parts again we obtain

$$\int_{\Omega}(-\Delta u - f)v \, d\boldsymbol{x} = 0, \qquad \forall v \in \mathring{H}^1(\Omega).$$

But as $C_0^\infty(\Omega) \subset \mathring{H}^1(\Omega)$, in particular we have that,

$$\int_{\Omega}(-\Delta u - f)v \, d\boldsymbol{x} = 0, \qquad \forall v \in \mathcal{C}_0^\infty(\Omega).$$

And as

$$-\Delta u - f \in \mathcal{C}(\overline{\Omega}) \subset L^1_{\text{loc}}(\Omega),$$

we conclude that $-\Delta u - f = 0$ almost everywhere in Ω as shown in Lemma 1.7. Now let $V = \mathring{H}^1(\Omega)$, then $a(\cdot,\cdot) : V \times V \mapsto \mathbb{R}$, defined by

$$a(u,v) = \int_{\Omega} \nabla u \cdot \nabla v \, d\boldsymbol{x} \qquad \forall u, v \in V$$

is a *bilinear form*. On the other hand, $\ell : V \mapsto \mathbb{R}$, defined by

$$\ell(v) = \int_{\Omega} f v \, d\boldsymbol{x} \qquad \forall v \in V,$$

is a linear functional. With these definitions, we can rewrite the *variational formulation* of (P) as:

Find $u \in V$ such that,

$$(PV) \qquad a(u,v) = \ell(v) \qquad \forall v \in V.$$

Definition 2.1 (Continuous Bilinear forms)**.** Let H be a normed linear space and let $a : H \times H \mapsto \mathbb{R}$ be a bilinear form. We say that $a(\cdot,\cdot)$ is *continuous* (or *bounded*) if there exists $C_a > 0$ such that,

$$|a(u,v)| \leq C_a \|u\|_H \|v\|_H \qquad \forall u, v \in H.$$

Definition 2.2 (Coercive Bilinear forms)**.** Let H be a normed linear space and let $a : H \times H \mapsto \mathbb{R}$ be a bilinear form. We say that $a(\cdot,\cdot)$ is *coercive* (or *elliptic*) in a subspace $V \subset H$ if there exists $c_a > 0$ such that,

$$a(v,v) \geq c_a \|v\|_H^2 \qquad \forall v \in V.$$

Proposition 2.3. *Let $(H, \|\cdot\|_H)$ be a Banach space, let V be a closed subspace of H and suppose that*

$$a : H \times H \longmapsto \mathbb{R}$$

is a symmetric bilinear form, that is also continuous over H with continuity constant $C_a > 0$, and coercive over $V \subset H$ with continuity constant $c_a > 0$. Then, $a(\cdot,\cdot)$ defines an inner product and $(V, a(\cdot,\cdot))$ is a Hilbert space.

Proof. We will first prove that $a(\cdot,\cdot)$ is an inner product in V, observe that symmetry and bilinearity are already guaranteed, so we only need to show positivity: Indeed, let $v = 0$, it is clear that $a(v,v) = 0$. For the converse, let $a(v,v) = 0$ and observe that by coercivity,

$$0 = a(v,v) \geq c_a \|v\|_H^2 \geq 0,$$

thus $v = 0$. So $(V, a(\cdot,\cdot))$ is an inner product space with induced norm

$$\|v\|_a := \sqrt{a(v,v)} \qquad \forall v \in V.$$

Now we show that $(V, a(\cdot,\cdot))$ is a Hilbert space, for this so, let $\{v_n\}_{n\in\mathbb{N}} \subset V$ be a Cauchy sequence in the topology induced by $\|\cdot\|_V$. Then, observe that

$$c_a \|v_n - v_m\|_H^2 \leq a(v_n - v_m, v_n - v_m) \leq C_a \|v_n - v_m\|_H^2 \qquad \forall m, n \in \mathbb{N},$$

so $\{v_n\}_{n\in\mathbb{N}}$ is also a Cauchy sequence in the topology induced by[4] $\|\cdot\|_H$. Furthermore, as H is a complete space, there exists $v \in H$ such that

$$v_n \xrightarrow[n\to\infty]{} v$$

in H. This way,

$$a(v_n - v_m, v_n - v_m) = \|v_n - v_m\|_a^2 \leq C_a \|v_n - v_m\|_H^2 \xrightarrow[n\to\infty]{} 0,$$

and as V is closed in H, we have that $v \in V$ and

$$v_n \xrightarrow[n\to\infty]{} v$$

in V. Thus, $\{v_n\}_{n\in\mathbb{N}}$ is convergent in V and so $(V, a(\cdot,\cdot))$ is a Hilbert space. ■

The symmetry concept referred in 2.3 defines a very specific family of variational formulations. In general, for a *symmetric bilinear form* $a : H \times H \mapsto \mathbb{R}$, a linear continuous functional $\ell \in H'$, and V a closed subspace of H, the problem:

[4]Notice that, in particular, $\|\cdot\|_V$ and $\|\cdot\|_H$ are equivalent norms.

Find $u \in V$ such that

$$(PVS) \qquad a(u,v) = \ell(v) \qquad \forall v \in V.$$

Is called *symmetric variational formulation.* A key result to address this kind of problems is presented next.

Theorem 2.4 (Riesz representation theorem)**.** *Let $(X, (\cdot,\cdot)_X)$ be a Hilbert space over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. Then, given any continuous linear functional $\ell \in X'$, there exists one and only one vector $y_\ell \in X$ such that*

$$\ell(x) = (x, y_\ell)_X \qquad \forall x \in X.$$

Besides,

$$\|\ell\|_{X'} = \|y_\ell\|_X,$$

and the Riesz isometry,

$$\tau : \ell \in X' \longmapsto \tau(\ell) := y_\ell \in X,$$

defined in this fashion is a bijection, which is linear if $\mathbb{K} = \mathbb{R}$ and semilinear if $\mathbb{K} = \mathbb{C}$. Consequently, any Hilbert space X can be identified with its dual space X' by means of the Riesz isometry $\tau : X' \mapsto X$. Furthermore, the dual space X' becomes a Hilbert space when it is equipped with the inner product $(\cdot,\cdot)_{X'} : X' \times X' \mapsto \mathbb{K}$ defined by

$$(x', y')_{X'} := \overline{(\tau x', \tau y')_X} \qquad \forall x', y' \in X'.$$

Proof. See [Cia13]. ∎

Now we are in conditions to state an existence and uniqueness result for symmetric variational problems.

Theorem 2.5 (Existence and uniqueness of solutions for symmetric variational formulations)**.** *Let $(H, \|\cdot\|_H)$ be a Banach space and $V \subset H$ be a closed subspace. Suppose furthermore that $\ell : H \mapsto \mathbb{R}$ is a linear and continuous functional under the $\|\cdot\|_H$-topology, and that*

$$a : H \times H \longmapsto \mathbb{R},$$

is a symmetric bilinear form that is also bounded over H and elliptic over $V \subset H$ with coercivity constant $c_a > 0$. Then there exists a unique solution to the symmetric variational problem (PVS).

Proof. The (PVS) reads: Find $u \in V$ such that

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

Observe, from proposition 2.3, that $(V, a(\cdot,\cdot))$ is a Hilbert space. Next, as $\ell \in H'$, there exists $c > 0$ such that $|\ell(v)| \leq c\|v\|_H$ for every $v \in H$, then, from the c_a-ellipticity of $a(\cdot,\cdot)$ over V, we can see that

$$|\ell(v)| \leq c\|v\|_H = \frac{c}{\sqrt{c_a}}\sqrt{c_a}\|v\|_H \leq \frac{c}{\sqrt{c_a}}\sqrt{a(v,v)} = \frac{c}{\sqrt{c_a}}\|v\|_a \qquad \forall v \in V.$$

So, $\ell \in V'$. It then follows, from the *Riesz representation theorem* 2.4 that there exists a unique $u_\ell \in V$ such that

$$\ell(v) = a(u_\ell, v) \qquad \forall v \in V.$$

■

We now can use theorem 2.5 to show existence and uniqueness of the model problem (PV), observe that in this case, we have that $H = H^1(\Omega)$ and $V = \mathring{H}^1(\Omega)$, the reader can verify that, in this settings, V is a closed subspace of H with V and H being Hilbert spaces. We now want to verify that

$$a(u,v) = \int_\Omega \nabla u \cdot \nabla v \, d\boldsymbol{x} \qquad \forall u, v \in \mathring{H}^1(\Omega)$$

is continuous over $H^1(\Omega)$ and coercive over $\mathring{H}^1(\Omega)$. We start with the continuity, notice that,

$$\begin{aligned} |a(u,v)| &= \left| \int_\Omega \nabla u \cdot \nabla v \, d\boldsymbol{x} \right| \\ &\leq \int_\Omega |\nabla u \cdot \nabla v| \, d\boldsymbol{x} \\ &\leq \|\nabla u\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)} \\ &\leq \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)} \end{aligned}$$

for every $u, v \in H^1(\Omega)$, so $a(\cdot,\cdot)$ is 1-continuous over $H^1(\Omega)$. Next, recall that,

$$\|v\|_{H^1(\Omega),\partial\Omega} = \left(\left(\int_{\partial\Omega} v(\boldsymbol{x}) \, dS_{\boldsymbol{x}} \right)^2 + \|\nabla v\|^2_{L^2(\Omega)} \right)^{1/2} \qquad \forall v \in H^1(\Omega)$$

is an equivalent norm in $H^1(\Omega)$, and so, there exists $c_{\text{PF}} > 0$ such that

$$\|v\|_{H^1(\Omega)} \leq c_{\text{PF}} \|v\|_{H^1(\Omega),\partial\Omega} \qquad \forall v \in H^1(\Omega).$$

This way,

$$\begin{aligned}\|v\|^2_{H^1(\Omega)} &\leq c^2_{\mathrm{PF}}\|v\|^2_{H^1(\Omega),\partial\Omega}\\ &\leq c^2_{\mathrm{PF}}\|\nabla v\|^2_{L^2(\Omega)}\\ &= c^2_{\mathrm{PF}}\int_\Omega \nabla v\cdot\nabla v\; d\boldsymbol{x}\\ &= c^2_{\mathrm{PF}}\; a(v,v)\end{aligned}$$

for each $v \in H^1(\Omega)$, and so, $a(\cdot,\cdot)$ is coercive in V with constant c^{-2}_{PF}. Finally, by Cauchy-Schwarz inequality, we obtain that $\ell \in H'$, indeed,

$$\begin{aligned}|\ell(v)| &= \left|\int_\Omega fv\; d\boldsymbol{x}\right|\\ &\leq \int_\Omega |fv|\; d\boldsymbol{x}\\ &\leq \|f\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)}\\ &\leq \|f\|_{L^2(\Omega)}\|v\|_{H^1(\Omega)},\end{aligned}$$

so ℓ is continuous over $H^1(\Omega)$ with continuity constant $\|f\|_{L^2(\Omega)}$. It then follows from theorem 2.5 that (PV) has one and only one solution $u \in \mathring{H}^1(\Omega)$. We now address a wider variety of symmetric problems.

2.1.1. *Dirichlet boundary conditions.* Consider now the Cauchy problem

$$\begin{cases} -\Delta u = f, & \text{in } \Omega\\ \quad\;\; u = g, & \text{on } \partial\Omega. \end{cases}$$

To study its variational formulation, we have seen that it is enough to ask for $f \in L^2(\Omega)$, we want to deliver now sufficient conditions on g for this problem to be *well-posed*.

Definition 2.6 (Well-posedness)**.** We will say that an arbitrary variational formulation is well-posed if it has a unique solution, that also depends continuously on the problem parameters.

When $g \neq 0$ we will say that the problem has *non-homogeneous Dirichlet boundary conditions*. To handle this type of problems, one may be tempted to search for solutions in the space

$$V_g := \{v \in H^1(\Omega) \mid \gamma v = g\},$$

where $\gamma : H^1(\Omega) \mapsto L^2(\partial\Omega)$ is the trace operator. An immediate problem the reader will identify is that V_g is not a Hilbert space when $g \neq 0$, furthermore, it is not even a subspace of $H^1(\Omega)$. Instead, we let $g \in H^{1/2}(\Omega) \subset L^2(\partial\Omega)$, and as,

$$\gamma H^1(\Omega) = H^{1/2}(\Omega)$$

we know that there exists $G \in H^1(\Omega)$ such that $\gamma G = g$. With this, we propose the following change of variables:

$$w := u - G \quad \text{s.t.} \quad \gamma G = g,$$

in other words,

$$\begin{cases} \Delta u = \Delta w + \Delta G, & \text{in } \Omega \\ \quad u = w + g, & \text{on } \partial\Omega. \end{cases}$$

But, $\gamma u = g$, $\gamma G = g$ on $\partial\Omega$ and $\Delta u = -f$ in Ω. So we obtain an equivalent boundary value problem

$$\begin{cases} -\Delta w = f + \Delta G, & \text{in } \Omega \\ \quad w = 0, & \text{on } \partial\Omega, \end{cases}$$

which is now a BVP with *homogeneous Dirichlet boundary conditions* in terms of the variable w. Observe that in fact one does not need to write the problem on its strong form, we do this only for illustrative purposes, moreover, we are not even asking G to have enough regularity for the term ΔG to make sense, instead, we can just apply the change of variables to the variational formulation obtained from (P), then we can see that,

$$\int_\Omega \nabla u \cdot \nabla v \, d\boldsymbol{x} = \int_\Omega \nabla(w + G) \cdot \nabla v \, d\boldsymbol{x} = \int_\Omega f v \, d\boldsymbol{x}$$

for all $v \in \mathring{H}^1(\Omega)$. So we are interested in solving the weak problem: Find $w \in \mathring{H}^1(\Omega)$ such that

$$\int_\Omega \nabla w \cdot \nabla v \, d\boldsymbol{x} = \int_\Omega f v \, d\boldsymbol{x} - \int_\Omega \nabla G \cdot \nabla v \, d\boldsymbol{x} \qquad \forall v \in \mathring{H}^1(\Omega).$$

If we define again $H = H^1(\Omega)$, $V = \mathring{H}^1(\Omega)$ and $a : H^1(\Omega) \times H^1(\Omega) \mapsto \mathbb{R}$ as

$$a(w, v) = \int_\Omega \nabla w \cdot \nabla v \, d\boldsymbol{x},$$

then $a(\cdot,\cdot)$ is a symmetric, linear, and bounded bilinear form over H and furthermore coercive in $V \subset H$, with V being a closed subspace of H, as we shown before. Then, let $\ell : H^1(\Omega) \mapsto \mathbb{R}$ be such that,

$$\ell(v) := \int_\Omega fv \, d\boldsymbol{x} - \int_\Omega \nabla G \cdot \nabla v \, d\boldsymbol{x}.$$

It is clear that ℓ is linear over H, and moreover,

$$\begin{aligned}
|\ell(v)| &\leq \|f\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)} + \|\nabla G\|_{L^2(\Omega)}\|\nabla v\|_{L^2(\Omega)} \\
&\leq \|f\|_{L^2(\Omega)}\|v\|_{H^1(\Omega)} + \|\nabla G\|_{L^2(\Omega)}\|v\|_{H^1(\Omega)} \\
&= \left(\|f\|_{L^2(\Omega)} + \|\nabla G\|_{L^2(\Omega)}\right)\|v\|_{H^1(\Omega)}
\end{aligned}$$

so ℓ is also bounded over H with continuity constant $\|f\|_{L^2(\Omega)} + |G|_{H^1(\Omega)} > 0$. Thus, given theorem 2.5, we have that there exists a unique $w \in \mathring{H}^1(\Omega)$ that is solution to our problem, for each choice of $G \in H^1(\Omega)$ with the desired extension $\gamma G = g$. The non-homogeneous Dirichlet boundary value problem then has a unique solution $u = w + G$.

2.1.2. *Neumann boundary conditions.* Consider this time the Cauchy problem

$$\begin{cases} -\Delta u + u = f, & \text{in } \Omega \\ \dfrac{\partial u}{\partial \hat{\boldsymbol{n}}} = g, & \text{on } \partial\Omega. \end{cases}$$

We call this class of boundary condition *non-homogeneous Neumann boundary conditions*. By means of integration by parts, the weak form of this problem will look like

$$\int_\Omega (\nabla u \cdot \nabla v + uv) \, d\boldsymbol{x} = \int_\Omega fv \, d\boldsymbol{x} + \int_{\partial\Omega} \frac{\partial u}{\partial \hat{\boldsymbol{n}}} \gamma v \, dS \qquad \forall v \in H^1(\Omega)$$

with $\gamma : H^1(\Omega) \mapsto L^2(\partial\Omega)$ the trace operator. This motivates the definition of the bilinear form $a : H^1(\Omega) \times H^1(\Omega) \mapsto \mathbb{R}$ as

$$a(u,v) = \int_\Omega (\nabla u \cdot \nabla v + uv) \, d\boldsymbol{x} \qquad \forall u, v \in H^1(\Omega),$$

and the linear form $\ell : H^1(\Omega) \mapsto \mathbb{R}$ as

$$\ell(v) = \int_\Omega fv \, d\boldsymbol{x} + \int_{\partial\Omega} \frac{\partial u}{\partial \hat{\boldsymbol{n}}} \gamma v \, dS \qquad \forall v \in H^1(\Omega).$$

Then, the variational formulation of this problem reads: Find $u \in H^1(\Omega)$ such that

$$a(u, v) = \ell(v) \qquad \forall v \in H^1(\Omega).$$

Define now, $H = V = H^1(\Omega)$ and observe that $a(\cdot, \cdot)$ is clearly symmetric and linear, also, using the Cauchy-Schwarz inequality we see that

$$\begin{aligned} |a(u, v)| &\leq \|\nabla u\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)} + \|u\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} \\ &\leq 2\|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)} \end{aligned}$$

for all $u, v \in H^1(\Omega)$ so $a(\cdot, \cdot)$ is bounded in $H^1(\Omega)$ with continuity constant 2. Furthermore,

$$a(v, v) = \int_\Omega |\nabla v|^2 + |v|^2 \, d\boldsymbol{x} = \|v\|^2_{H^1(\Omega)} \qquad \forall v \in H^1(\Omega),$$

so $a(\cdot, \cdot)$ is also elliptic over $H^1(\Omega)$ with coercivity constant 1. Finally, notice that

$$\begin{aligned} |\ell(v)| &\leq \|f\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} + \|g\|_{L^2(\partial\Omega)} \|\gamma v\|_{L^2(\partial\Omega)} \\ &\leq \|f\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} + \|g\|_{L^2(\partial\Omega)} \|\gamma\| \|v\|_{L^2(\Omega)} \\ &\leq \left(\|f\|_{L^2(\Omega)} + \|g\|_{L^2(\partial\Omega)} \|\gamma\| \right) \|v\|_{H^1(\Omega)}. \end{aligned}$$

This implies that $\ell \in H^{-1}(\Omega)$ with continuity constant $\|f\|_{L^2(\Omega)} + \|g\|_{L^2(\partial\Omega)} \|\gamma\| > 0$. All together, by theorem 2.5, we obtain that there exists one and only one solution $u \in H^1(\Omega)$ for the non-homogeneous Neumann boundary value problem.

Remark 2.6.1. Observe that the variational formulation holds for every $v \in H^1(\Omega)$, in particular, letting the $v = 1$ we see that

$$\int_\Omega u \, d\boldsymbol{x} = \int_\Omega f \, d\boldsymbol{x} + \int_{\partial\Omega} g \, dS$$

we call this *compatibility condition*, and is an additional requirement over the mean value of $u \in H^1(\Omega)$ for the problem to be well-posed.

We now examine an example of a Cacuchy problem with *homogeneous Neumann boundary conditions*

$$\begin{cases} -\Delta u = f, & \text{in } \Omega \\ \dfrac{\partial u}{\partial \hat{\boldsymbol{n}}} = 0, & \text{on } \partial\Omega. \end{cases}$$

Observe that after integrating against a test function $v \in H^1(\Omega)$ one obtain that

$$\int_\Omega \nabla u \cdot \nabla v \; d\boldsymbol{x} + \int_{\partial\Omega} \underbrace{\frac{\partial u}{\partial \hat{\boldsymbol{n}}}}_{0} \cdot \nabla v \; d\boldsymbol{x} = \int_\Omega fv \; d\boldsymbol{x},$$

but $v \in H^1(\Omega)$ is arbitrary, and so for the particular case $v = 1$ we see that

$$\int_\Omega f \; d\boldsymbol{x} = 0$$

which is the *compatibility condition* for the weak formulation. Observe that this breaks the uniqueness of solutions if we are not careful enough to define an appropriate space of solutions, indeed, let

$$a(u, v) = \int_\Omega \nabla u \cdot \nabla v \; d\boldsymbol{x} \qquad \forall u, v \in H^1(\Omega)$$

and

$$\ell(v) = \int_\Omega fv \; d\boldsymbol{x} \qquad \forall v \in H^1(\Omega)$$

and consider the variational formulation: Find $u \in H^1(\Omega)$

$$a(u, v) = \ell(v) \qquad \forall v \in H^1(\Omega).$$

We then observe that if $u \in H^1(\Omega)$ solves this problem, then $u + a$ is also a solution for every $a \in \mathbb{R}$. To see this clearly, we define

$$A[\cdot] : H^1(\Omega) \longmapsto H^{-1}(\Omega)$$

as

$$A[u] := a(u, \cdot).$$

Then the variational formulation can be equivalently formulated in the dual space $H^{-1}(\Omega)$ of $H^1(\Omega)$ as finding $u \in H^1(\Omega)$ such that

$$A[u] = \ell \;\text{ in }\; H^{-1}(\Omega).$$

Now, it is clear that $\ker(A) = \mathcal{P}^0(\Omega)$, so we need to exclude the subspace of constant functions from the space of solutions in order to invert A. For this so, we ask for any solution $u \in H^1(\Omega)$ to be orthogonal to $\mathcal{P}^0(\Omega)$, i.e.

$$0 = (u, a)_{H^1(\Omega)} = (u, a)_{L^2(\Omega)} \qquad \forall a \in \mathbb{R},$$

with $a \neq 0$, in other words, we need to impose the additional restriction

$$\int_\Omega ua\, d\boldsymbol{x} = a \int_\Omega u\, d\boldsymbol{x} = 0 \qquad \forall a \in \mathbb{R} \setminus \{0\}.$$

This way the Cauchy problem can be restated as

$$\begin{cases} -\Delta u = f, & \text{in } \Omega \\ \dfrac{\partial u}{\partial \hat{\boldsymbol{n}}} = 0, & \text{on } \partial\Omega \\ \displaystyle\int_\Omega u\, d\boldsymbol{x} = 0. \end{cases}$$

To provide a well-posed variational formulation, let us first introduce the subspace V of $H^1(\Omega)$ defined as

$$V = \left\{ v \in H^1(\Omega) \;\middle|\; \int_\Omega u\, d\boldsymbol{x} = 0 \right\} \subset H^1(\Omega).$$

Notice that this subspace is closed, as it is the pre-image of a singleton under a continuous map[5]. Next, we consider the following norm,

$$\|v\|_{H^1(\Omega),f} := \left(\left(\int_\Omega v\, d\boldsymbol{x} \right)^2 + \|\nabla v\|^2_{L^2(\Omega)} \right)^{1/2} \qquad \forall v \in H^1(\Omega),$$

introduced previously in theorem 1.21, in this case

$$f(v) = \int_\Omega v\, d\boldsymbol{x} \qquad \forall v \in H^1(\Omega)$$

and observe that

$$[f(v) = 0] \Leftrightarrow [v \equiv_\mu 0]$$

for all constant function $v \in \mathcal{P}^0(\Omega)$. So it follows that $\|\cdot\|_{H^1(\Omega),f}$ is an equivalent norm in $H^1(\Omega)$, i.e., there exists $c_{\mathrm{P}}, C_{\mathrm{P}} > 0$ such that

$$c_{\mathrm{P}} \|v\|_{H^1(\Omega),f} \leq \|v\|_{H^1(\Omega)} \leq C_{\mathrm{P}} \|v\|_{H^1(\Omega),f} \qquad \forall v \in H^1(\Omega).$$

In particular, $a(\cdot,\cdot)$ being elliptic over $H^1(\Omega)$ (as we saw previously) implies that it is also elliptic over $(V, \|\cdot\|_{H^1(\Omega),f})$, indeed:

$$\frac{c_{\mathrm{P}}^2}{c_{\mathrm{PF}}^2} \|v\|^2_{H^1(\Omega),f} \leq \frac{1}{c_{\mathrm{PF}}^2} \|v\|^2_{H^1(\Omega)} \leq a(v,v) \qquad \forall v \in V.$$

[5]Moreover, this shows that V is an hyperplane.

And furthermore, we have seen before that a is bilinear, symmetric and bounded over $H^1(\Omega)$, and that $\ell \in H^{-1}(\Omega)$. So, the variational formulation:

Find $u \in (V, \|\cdot\|_{H^1(\Omega),f})$ such that,

$$a(u,v) = \ell(v) \qquad \forall v \in (V, \|\cdot\|_{H^1(\Omega),f}).$$

has a unique solution by means of theorem 2.5, this solution satisfying the proposed homogeneous Neumann boundary value problem.

2.1.3. *Approximation of solutions for symmetric problems.* We will now take advantage of the structure of variational formulations in order to define suitable and optimal approximations of solutions to this problems.

Definition 2.7 (Ritz-Galerkin approximation)**.** Let V be a Hilbert space, and let $V_h \subset V$ be a finite dimensional subspace. Given a symmetric bilinear form $a : V \times V \mapsto \mathbb{R}$ and some $F \in V'$, the element $u_h \in V_h$ that satisfies:

$$a(u_h, v) = F(v) \qquad \forall v \in V_h,$$

is called the *Ritz-Galerkin approximation* of the solution to the infinite dimensional counterpart of this variational formulation, i.e., the element $u \in V$ that satisfies

$$a(u,v) = F(v) \qquad \forall v \in V.$$

Let $a : V \times V \mapsto \mathbb{R}$ be a symmetric bilinear form over some Hilbert space V, let $F \in V'$ and let $V_h \subset V$ be some finite dimensional subspace. Assume that the variational problem of finding $u \in V$ such that

$$a(u,v) = F(v) \qquad \forall v \in V,$$

is well-posed. We want to characterize the Ritz-Galerkin approximation $u_h \in V_h$ for this problem, for this so, notice first that as $\dim(V_h) < \infty$, there exists some finite collection $\{\psi_i\}_{i=1}^N \subset V_h$ such that

$$V_h = \text{span}\{\psi_1, ..., \psi_N\}$$

It then follows that $u_h \in V_h$ admits the representation

$$u_h = \sum_{i=1}^{N} c_i \psi_i$$

for some scalars $\{c_i\}_{i=1}^N$. Hence, when that the variational formulation is restricted to V_h, i.e. Find a collection of scalars $\{c_i\}_{i=1}^N$ such that

$$a\left(\sum_{i=1}^{N} c_i\psi_i, \psi_j\right) = \sum_{i=1}^{N} c_i a\left(\psi_i, \psi_j\right) = F(\psi_j) \qquad \forall j \in \{1, ..., N\},$$

we can see that it amounts to solve a linear system

$$A\boldsymbol{c} = \boldsymbol{b}$$

where

$$A_{ij} = a(\psi_i, \psi_j), \qquad \boldsymbol{c} = (c_1, ..., c_N)^T \in \mathbb{R}^N, \qquad \boldsymbol{b} = (F(\psi_1), ..., F(\psi_N))^T \in \mathbb{R}^N.$$

Furthermore, by proposition 2.3, $(V_h, a(\cdot,\cdot))$ is a Hilbert space, and the map

$$\|v\|_a := \sqrt{a(v,v)} \qquad \forall v \in V_h$$

is a norm. And so, for any $v = \sum_{i=1}^N c_i\psi_i \in V_h$ we have that

$$\boldsymbol{c}^T A\boldsymbol{c} = \sum_{i=1}^{M}\sum_{j=1}^{M} c_i a(\psi_i, \psi_j')c_j = a\left(\sum_{i=1}^{M} c_i\psi_i, \sum_{j=1}^{M} c_j\psi_j\right) = a(v,v) = \|v\|_a^2$$

with $\|v\|_a = 0$ if and only if $\boldsymbol{c} = \boldsymbol{0}$. Thus, A is a positive definite symmetric matrix, and then the linear system admits a unique solution.

Theorem 2.8 (Existence and uniqueness of the Ritz-Galerkin approximation)**.** *Under the same conditions of theorem 2.5, there exists a unique solution $u_h \in V_h$ to the Ritz-Galerkin approximation problem.*

Proof. As we saw before, $(V_h, a(\cdot,\cdot))$ is a Hilbert space, and furthermore, $F|_{V_h} \in V_h'$. Then, by the Riesz representation theorem we have that for any fixed $F \in V$ there exists a unique representative $u_h \in V_h$ such that

$$a(u_h, v) = F|_{V_h}(v)$$

for all $v \in V_h$. ■

Proposition 2.9 (Galerkin orthogonality)**.** *Let $u_h \in V_h$ be the Galerkin approximation to the solution $u \in V$ of a well-posed variational problem (PVS),*

then
$$a(u - u_h, v) = 0 \qquad \forall v \in V_h.$$

Proof. We have that $a(u, v) = F(v)$ for all $v \in V$, so in particular this also holds for all $v \in V_h$. We then can see that

$$\begin{aligned} a(u, v) &= F(v) \qquad \forall v \in V_h, \\ a(u_h, v) &= F(v) \qquad \forall v \in V_h. \end{aligned}$$

Subtracting these equations the result follows by linearity of a,

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

■

Corollary 2.9.1. *Under the same settings than proposition 2.9. Let*
$$\|v\|_a := \sqrt{a(v, v)} \qquad \forall v \in V,$$
then,
$$\|u - u_h\|_a = \min_{v \in V_h} \|u - v\|_a$$

Proof. Let $v \in V_h$ be arbitrary and observe that $u_h - v \in V_h$ as V_h is a subspace, this way,

$$\begin{aligned} \|u - v\|_a^2 &= \|u - u_h + u_h - v\|_a^2 \\ &= a(u - u_h, u - u_h) + 2a(u - u_h, u_h - v) + a(u_h - v, u_h - v), \end{aligned}$$

but by proposition 2.9 we have that $a(u - u_h, u_h - v) = 0$ for all $v \in V_h$, and so,

$$\|u - v\|_a^2 = \|u - u_h\|_a^2 + \|u_h - v\|_a^2 \geq \|u - u_h\|_a^2 \qquad \forall v \in V_h.$$

Hence, in particular,

$$\|u - u_h\|_a^2 \leq \inf_{v \in V_h} \|u - v\|_a^2,$$

and furthermore, as V_h is a closed subspace and $u_h \in V_h$ we infer that

$$\|u - u_h\|_a^2 = \min_{v \in V_h} \|u - v\|_a^2.$$

■

2.2. **Non-symmetric case.** Until now, he have exploited the fact that $a(\cdot,\cdot)$ is symmetric, in what follows, we will considering the following settings,

- $(H,(\cdot,\cdot))$ is a Hilbert sapace.
- V is a closed subspace of H
- $a(\cdot,\cdot)$ is a bilinear form in H, **not necessarily symmetric**.
- $a : H \times H \mapsto \mathbb{R}$ is continuous and $a : V \times V \mapsto \mathbb{R}$ is coercive.

Then, we want to study the following *non-symmetric variational formulation*,

$$\text{(PVNS)} \quad \text{Given } \ell \in V', \text{ find } u \in V \text{ such that } \ a(u,v) = \ell(v) \ \text{ for all } v \in V. \tag{9}$$

We will see now some examples of non-symmetric problems.

Example 2.9.1. Consider the following boundary value problem,

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

One then can see that the bilinear form $a : H^1(0,1) \times H^1(0,1) \mapsto \mathbb{R}$ associated to the weak formulation of this problem is given by

$$a(u,v) = \int_0^1 (u'v' + u'v + uv)\, dx,$$

which is clearly not symmetric. We can furthermore extend this example to higher dimensions, indeed, let $\Omega \subset \mathbb{R}^d$ with $d > 1$ and $\boldsymbol{c} \in \mathbb{R}^d$, the following boundary value problem

$$\begin{cases} -\Delta u + \nabla u \cdot \boldsymbol{c} + u = f, & \text{in } \Omega \\ u = 0, & \text{on } \partial\Omega \end{cases}$$

has an associated bilinear form $a : H^1 \times H^1 \mapsto \mathbb{R}$ given by

$$a(u,v) = \int_\Omega \left(\nabla u \cdot \nabla v + (\nabla u \cdot \boldsymbol{c})\, v + uv\right)\, dx,$$

which, again, is clearly non-symmetric.

To study existence and uniqueness of solutions for non-symmetric variational problems, the results we studied for the symmetric case become useless, in order to obtain new conditions for well-posedness we will first recall the following Real Analysis result.

Theorem 2.10 (Banach Fixed Point)**.** *Let (V, d) be a complete metric space and $T : V \mapsto V$ be a contractive function, i.e.,*

$$d(Tv_1, Tv_2) \leq Md(v_1, v_2) \qquad \forall v_1, v_2 \in V,$$

with $0 < M < 1$. Then T has one and only one fixed point $v \in V$, i.e., such that $Tv = v$. Besides, given any point $v_0 \in V$, the sequence $\{v_n\}_{n=0}^{\infty}$ defined by

$$v_{n+1} = T(v_n) \qquad n \geq 0,$$

converges to v as $n \to \infty$, and the following estimate holds:

$$\|v - v_n\| \leq CM^n, \; n \geq 0, \quad \text{with } \; C := \frac{d(Tv_0, v_0)}{1 - M}.$$

Proof. See [Cia13]. ■

Theorem 2.11 (Lax-Milgram)**.** *Let $(H, \|\cdot\|_H)$ be a Hilbert space and*

$$(V, \|\cdot\|_V) \subset (H, \|\cdot\|_H)$$

be a closed subspace. Let $a \in \mathcal{L}(H \times H; \mathbb{R})$ be a V-coercive bilinear form, i.e., there exists $c_a > 0$ such that

$$a(v, v) \geq c_a \|v\|_V^2 \qquad \forall v \in V.$$

And let $\ell : V'$. Then, the variational problem: Find $u \in V$ such that

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

has one and only one solution.

Proof. Let $u \in V$ be arbitrary, and define the linear form $Au(\cdot) : V \to \mathbb{R}$ as

$$Au(v) = a(u, v) \qquad \forall v \in V$$

observe that,

$$\frac{|Au(v)|}{\|v\|_V} = \frac{|a(u, v)|}{\|v\|_V} \leq \underbrace{\|a\|\|u\|_V}_{\text{Cont. constant of } Au} \qquad \forall v \in V - \{0\}$$

so $Au \in V'$ with

$$\|Au\|_{V'} \leq \|a\|\|u\|_V.$$

Furthermore, we can define the linear operator $A : V \longmapsto V'$ as

$$u \longmapsto Au(\cdot) \in V',$$

we can see that $A \in \mathcal{L}(V;V')$ as

$$\frac{\|Au\|_{V'}}{\|u\|_V} \leq \|a\| \qquad \forall u \in V - \{0\},$$

hence,

$$\|A\|_{\mathcal{L}(V,V')} \leq \|a\|.$$

We then can see the equivalence between the problems

$$[a(u,v) = \ell(v) \quad \forall v \in V] \quad \Leftrightarrow \quad [Au = \ell \quad \text{in } V'] \quad \Leftrightarrow \quad [\tau(Au - \ell) = 0 \quad \text{in } V]$$

here $\tau : V' \longmapsto V$ denotes the F. Riesz isometry. We now let $\omega > 0$, and deliberately define the mapping

$$f_\omega : V \longmapsto V$$

as,

$$f_\omega(v) = v - \omega\tau(Av - \ell) \qquad \forall v \in V.$$

One can show that, when choosing a suitable value for $\omega > 0$, the map f_ω is a contraction, indeed, let $v_1, v_2 \in V$ be arbitrary and $w := v_1 - v_2$, then

$$\begin{aligned}
\|f_\omega(v_1) - f_\omega(v_2)\|_V &= \|(v_1 - v_2) - \omega\tau A(v_1 - v_2)\|_V^2 \\
&= \|w\|_V^2 - 2\omega(w, \tau Aw)_V + \omega^2\|\tau Aw\|_V^2 \\
&= \|w\|_V^2 - 2\omega a(w,w) + \omega^2\|Aw\|_{V'}^2 \\
&\leq \|w\|_V^2 - 2\omega c_a\|w\|_V^2 + \omega^2\|a\|^2\|w\|_V^2 \\
&= \left(1 - 2\omega c_a + \omega^2\|a\|^2\right)\|v_1 - v_2\|_V^2
\end{aligned}$$

Observe that

$$[1 - 2\omega c_a + \omega^2\|a\|^2 < 1] \quad \Leftrightarrow \quad [\omega(\omega\|a\|^2 - 2c_a) < 0]$$

so f_ω is a contraction provided that

$$\omega \in \left(0, \frac{2c_a}{\|a\|^2}\right)$$

In this scenario, by the *Banach Fixed Point theorem*, f_ω has a unique fixed point $u \in V$ s.t.

$$u = f_\omega(u) = u - \omega\tau(Au - \ell),$$

therefore,

$$\tau(Au - \ell) = 0 \quad \text{in } V$$

and the existence and uniqueness follows. ■

Observe that, given the *Lax-Milgram* theorem, the only ingredient left to obtain sufficient conditions for well-posedness of non-symmetric variational formulations is the stability of solutions.

Remark 2.11.1. Notice that, in the same settings than theorem 2.11, the stability of the variational formulation follows from the V-coercivity of $a(\cdot,\cdot)$ and $\ell \in V'$, indeed, if $u \in V$ solves the weak problem, then

$$c_a \|u\|_V^2 \leq a(u,u) = \ell(u) \leq \|\ell\|_{V'} \|u\|_V$$

Thus,

$$\|u\|_V \leq \frac{\|\ell\|_{V'}}{c_a}.$$

This way, provided the conditions for the *Lax-Milgram* theorem to hold, we get not only existence and uniqueness but also stability of solutions.

2.2.1. *Approximation of solutions for non-symmetric problems.* We start by introducing the approximation concept for non-symmetric weak problems.

Definition 2.12 (Bubnov-Galerkin approximation)**.** In the settings of a non-symmetric variational formulation (9), let $V_h \subset V$ be a finite dimensional subspace. Then the element $u_h \in V_h$ that satisfies:

$$a(u_h, v) = \ell(v) \qquad \forall v \in V_h,$$

is called the *Bubnov-Galerkin approximation* to the solution of the infinite dimensional non-symmetric variational formulation.

We discuss now approximation results for non-symmetric variational formulations.

Theorem 2.13. *Under the same conditions than theorem 2.11, the Bubnov-Galerkin approximation problem has one and only one solution.*

Proof. Observe that $V_h \subset V \subset H$ is a finite dimensional closed subspace, and so, by inheritance it is a Hilbert space as well. Also, $a : V_h \times V_h \mapsto \mathbb{R}$ is bounded, coercive and $\ell|_{V_h}$ is linear and bounded. Hence, by *Lax-Milgram* theorem 2.11, there exists a unique $u \in V_h$ such that $a(u_h, v) = \ell(v)$ for all $v \in V_h$. ■

Now it results natural to ask for error estimates for the Bubnov-Galerkin approximation, to establish such a bound we intoduce the so called *Céa's lemma.*

Lemma 2.14 (Céa's lemma)**.** *Let $u \in V$ be the solution of the abstract non-symmetric variational problem* (9) *and $u_h \in V_h$ be the corresponding Bubnov-Galerkin approximation, then,*

$$\|u - u_h\|_V \leq \frac{\|a\|}{c_a} \min_{v \in V_h} \|u - v\|_V$$

where $\|a\|$ is the operator norm and c_a is the ellipticity constant of the bilinear form $a(\cdot, \cdot)$.

Proof. Observe that,

$$\begin{aligned} a(u, v) &= \ell(v) \qquad \forall v \in V \\ a(u_h, v) &= \ell(v) \qquad \forall v \in V_h \subset V. \end{aligned}$$

And so, by linearity of a,

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

It then follows from the ellipticity of a that for all $v \in V_h$:

$$\begin{aligned} c_a \|u - u_h\|_V^2 &\leq a(u - u_h, u - u_h) \\ &= a(u - u_h, u - v) + \underbrace{a(u - u_h, v - u_h)}_{=0, \text{ by } 2.9} \\ &= a(u - u_h, u - v) \\ &\leq \|a\| \|u - u_h\|_V \|u - v\|_V. \end{aligned}$$

Hence,

$$\|u - u_h\|_V \leq \frac{\|a\|}{c_a} \|u - v\|_V \qquad \forall v \in V_h,$$

and in particular,

$$\|u - u_h\|_V \leq \frac{\|a\|}{c_a} \min_{v \in V_h} \|u - v\|_V.$$

■

Remark 2.14.1. Recall that in the symmetric case we have, by corollary 2.9.1, that

$$\|u - u_h\|_a \leq \min_{v \in V_h} \|u - v\|_a$$

where $\|v\|_a = \sqrt{a(v,v)}$ for all $v \in V$. Notice now that from the V-coercivity and the boundedness of a it follows that

$$\sqrt{c_a}\|u - u_h\|_V \leq \|u - u_h\|_a \leq \min_{v \in V_h} \|u - v\|_a \leq \sqrt{\|a\|} \min_{v \in V_h} \|u - v\|_V,$$

this shows that the error bound for the symmetric case:

$$\|u - u_h\|_V \leq \sqrt{\frac{\|a\|}{c_a}} \min_{v \in V_h} \|u - v\|_V,$$

is better than the one obtained for the non-symmetric case by means of the *Cea's lemma*:

$$\|u - u_h\|_V \leq \frac{\|a\|}{c_a} \min_{v \in V_h} \|u - v\|_V.$$

To see this, observe that

$$c_a\|v\|_V^2 \leq a(v,v) \leq \|a\|\|v\|_V^2 \quad \forall v \in V \;\Rightarrow\; c_a \leq \|a\| \;\Rightarrow\; \sqrt{\frac{\|a\|}{c_a}} \leq \frac{\|a\|}{c_a}.$$

3. Finite element method

From the previous chapter we have seen that when replacing the Hilbert space V with a finite dimensional subspace $V_h \subset V$. We then look for a solution $u_h \in V_h$ such that

$$a(u_h, v) = \ell(v) \qquad \forall v \in V_h,$$

here $a \in \mathcal{L}(V \times V; \mathbb{R})$ is V-elliptic and $\ell \in V'$. As there exists a collection $\{\psi_i\}_{i=1}^N$ with $N < \infty$ such that

$$V_h = \text{span}\,\{\psi_1, ..., \psi_N\},$$

then both the Ritz-Galerkin and the Bubnov-Galerkin approximations are of the form $u_h = \sum_{i=1}^N c_i\psi_i$ such that

$$\sum_{i=1}^{N} c_i a(\psi_i, \psi_j) = \ell(\psi_j) \qquad \forall j \in \{1, ..., N\}$$

this way u_h is uniquely characterized by the coefficients $\{c_i\}_{i=1}^N$ that solve the linear system

$$A\boldsymbol{c} = \boldsymbol{b}$$

where $A_{ij} = a(\psi_i, \psi_j)$ and $b_i = \ell(\psi_i)$ for $i, j \in \{1, ..., N\}$, we have seen that for elliptic problems the matrix A is positive definite as $a(\cdot,\cdot)$ is V-coercive, indeed,

$$\boldsymbol{c}^T A\boldsymbol{c} = \sum_{i=1}^{M}\sum_{j=1}^{M} c_i a(\psi_i, \psi_j) c_j = a\left(\sum_{i=1}^{M} c_i\psi_i, \sum_{j=1}^{M} c_j\psi_j\right) = a(v,v) \geq \|v\|_V^2 > 0$$

if $v \neq 0$ and $\boldsymbol{c}^T A\boldsymbol{c} = a(v,v) = \|v\|_V = 0$ if $v = 0$. Then the existence and uniqueness of the Galerkin approximation $u_h \in V_h$ is ensured, in literature the matrix A is known as the *stiffness matrix*.

Proposition 3.1. *Let $V_h \subset V$ a finite dimensional subspace of V. Then, the solution of the Galerkin approximation problem $u_h \in V_h$ satisfies*

$$\|u_h\|_V \leq c_a^{-1}\|\ell\|_{V'}$$

where $c_a > 0$ is the ellipticity constant of $a(\cdot,\cdot)$. In other words, the Galerkin approximation is stable.

Proof. We have that $a(u_h, v) = \ell(v)$ for all $v \in V_h$. In particular,

$$a(u_h, u_h) = \ell(u_h).$$

It then follows from the coercivity of $a(\cdot,\cdot)$ and the continuity of ℓ that

$$c_a\|u_h\|_V^2 \leq a(u_h,u_h) \leq \ell(u_h) \leq \|\ell\|_{V'}\|u_h\|_V.$$

Thus,

$$\|u_h\|_V \leq c_a^{-1}\|\ell\|_{V'}.$$

■

The following example is considered the birth of the finite element method.

Example 3.1.1 (R. Courant, 1943). Suppose that we want to solve the Poisson equation in the unit square,

$$\begin{cases} -\Delta u = f, & \text{in } \Omega = (0,1)^2 \\ \quad u = 0, & \text{on } \partial\Omega \end{cases}$$

We consider an uniform partition of $\overline{\Omega}$ in triangles of uniform size $h > 0$ as shown in figure 6,

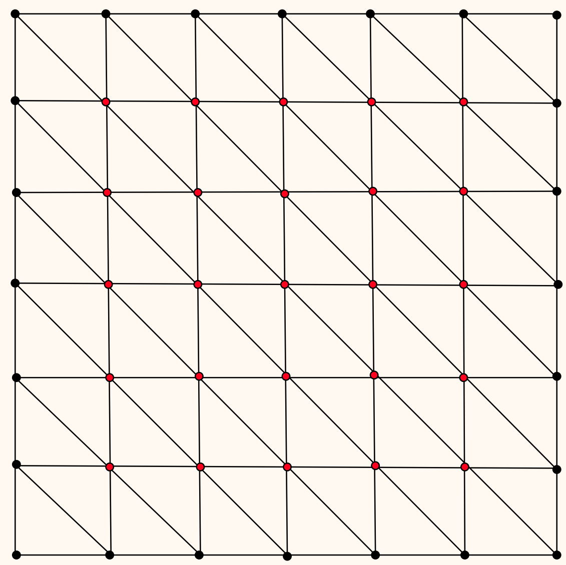


FIGURE 6. Uniform partition of $\overline{\Omega}$

where $h = 1/\tilde{N}$ and $\tilde{N} = 6$ in the figure. Our subspace V_h in this case would be

$$V_h = \{v \in \mathcal{C}(\overline{\Omega}) \mid v \text{ is linear over every triangle } T \text{ and } v = 0 \text{ over } \partial\Omega\}$$

So $v|_T \in \mathcal{P}^1(T)$, this way,

$$v|_T = a + bx + cy$$

for some coefficients $\{a, b, c\}$. From chapter 1.7 we can see that $V_h \subset \mathring{H}^1(\Omega)$. Furthermore, observing figure 7

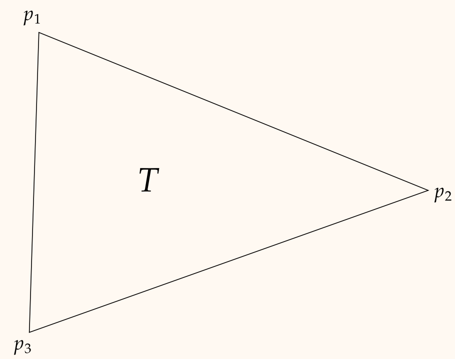


FIGURE 7. Arbitrary triangle T of the uniform partition of $\Omega = (0,1)$, with vertices $\{p_1, p_2, p_3\}$.

we notice that if $v|_T$ is known at the vertices $\{p_1, p_2, p_3\}$ then $\{a, b, c\}$ are uniquely determined. As $v|_{\partial\Omega}$ is known, then a function $v \in V_h$ is uniquely determined by its values at the vertices or "nodes" located at the interior of the domain, thus

$$\dim\,(V_h) = N = \#\text{ of nodes in the interior of }\Omega$$

Consider then the Galerkin approximation $u_h \in V_h$,

$$u_h(\boldsymbol{z}) = \sum_{j=1}^{N} c_i \psi_i(\boldsymbol{z})$$

where $\{\psi_i\}_{i=1}^N$ is a basis for V_h, if we denote the nodes of the mesh by $\{\boldsymbol{z}_i = (x_i, y_i)\}_{i=1}^N$ and we ask for

$$\psi_i(\boldsymbol{z}_j) = \delta_{i,j}$$

it then follows that

$$u_h(\boldsymbol{z}_i) = \sum_{j=1}^{N} c_j \psi_j(\boldsymbol{z}_i) = \sum_{j=1}^{N} c_j \delta_{i,j} = c_j.$$

Thus, for an arbitrary point $\boldsymbol{z} \in \overline{\Omega}$,

$$u_h(\boldsymbol{z}) = \sum_{i=1}^{N} u_h(\boldsymbol{z}_i)\psi_i(\boldsymbol{z}).$$

To determine this coefficients, we assemble the linear system

$$A\boldsymbol{c} = \boldsymbol{b},$$

here, $\boldsymbol{c} = (u_h(\boldsymbol{z}_1), ..., u_h(\boldsymbol{z}_N))^T$, $\boldsymbol{b} = (\ell(\psi_1)..., \ell(\psi_N))^T$ and

$$A_{ij} = \int_\Omega \nabla\psi_i \cdot \nabla\psi_j \, dx.$$

By construction, in an arbitrary interior node $\boldsymbol{c}$ with neighbors $\{N, S, E, W, NW, SE\}$, as illustrated in figure 8

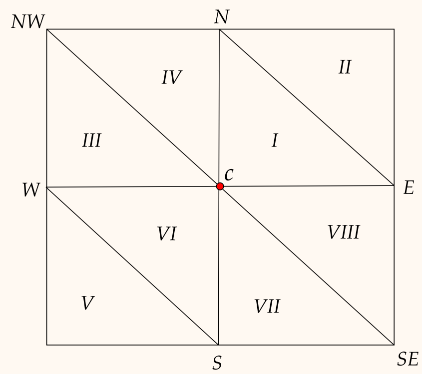


FIGURE 8. Arbitrary node $\boldsymbol{c}$ with labeled neighbors $\{N, S, E, W, NW, SE\}$.

we have that,

$$\psi_{\boldsymbol{c}}(\boldsymbol{c}) = 1$$

$$\psi_{\boldsymbol{c}}(N) = \psi_{\boldsymbol{c}}(E) = \psi_{\boldsymbol{c}}(W) = \psi_{\boldsymbol{c}}(S) = \psi_{\boldsymbol{c}}(NW) = \psi_{\boldsymbol{c}}(SE) = 0$$

$$\psi_{\boldsymbol{c}}|_T \in \mathcal{P}^1(T), \qquad T = I, ..., VIII.$$

Thus, $\psi_{\boldsymbol{c}}$ looks like in figure 9

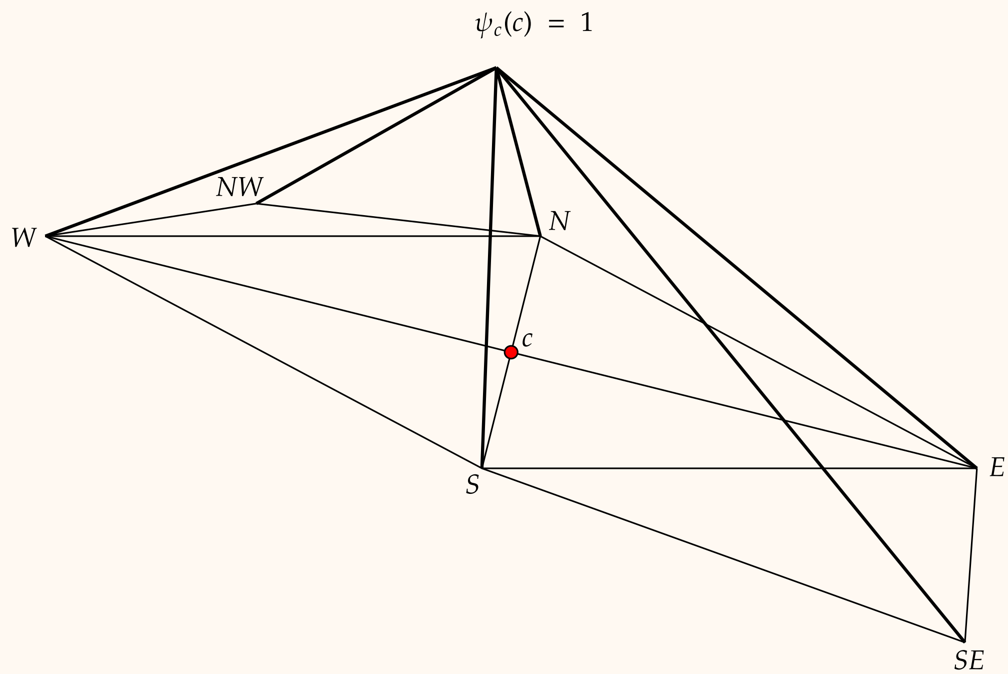


FIGURE 9. ψ_c representation.

and they receive the name of *hat functions* in literature. Observe now that as $\Omega = \bigcup_{i=1}^{M} T_i$ where M is the amount of triangles in the partition, then we have that

$$A_{ij} = \int_\Omega \nabla\psi_i \cdot \nabla\psi_j \; dx = \sum_{k=1}^{M} \int_{T_k} \nabla\psi_i \cdot \nabla\psi_j \; dx.$$

Given the shape of ψ_c we can see that

	I	II	III	IV	V	VI	VII	VIII
$\frac{\partial\psi_c}{\partial x}$	-1/h	0	1/h	0	0	1/h	0	-1/h
$\frac{\partial\psi_c}{\partial y}$	-1/h	0	0	-1/h	0	1/h	1/h	0

And so, denoting Area$(T) = h^2/2$ the area of the triangle $T = I, II, ..., VIII$, we can see that

$$a(\psi_{\boldsymbol{c}}, \psi_{\boldsymbol{c}}) = \int_{I+III+IV+VI+VII+VIII} |\nabla\psi_{\boldsymbol{c}}|^2 \; dxdy$$

$$
\begin{aligned}
&= 2\int_{I+III+IV} |\nabla\psi_{\boldsymbol{c}}|^2 \, dxdy \\
&= 2\int_{I+III+IV} \left(\frac{\partial\psi_{\boldsymbol{c}}}{\partial x}\right)^2 + \left(\frac{\partial\psi_{\boldsymbol{c}}}{\partial x}\right)^2 \, dxdy \\
&= 2\text{Area}(I)2h^{-2} + 2\text{Area}(III)h^{-2} + 2\text{Area}(IV)h^{-2} \\
&= 2\frac{h^2}{2}2h^{-2} + 2\frac{h^2}{2}h^{-2} + 2\frac{h^2}{2}h^{-2} \\
&= 2+1+1 \\
&= 4.
\end{aligned}
$$

Also,

$$
\begin{aligned}
a(\psi_c, \psi_N) &= \int_{I+IV} \nabla\psi_{\boldsymbol{c}} \cdot \nabla\psi_N \, dxdy \\
&= \int_I \left(\frac{-1}{h}\frac{\partial\psi_N}{\partial x} + \frac{-1}{h}\frac{\partial\psi_N}{\partial y}\right) \, dxdy + \int_{IV} \left(0\frac{\partial\psi_N}{\partial x} - \frac{1}{h}\frac{\partial\psi_N}{\partial y}\right) \, dxdy \\
&= \int_I \left(\frac{-1}{h}0 + \frac{-1}{h}h^{-1}\right) \, dxdy + \int_{IV} \left(0h^{-1} - \frac{1}{h}h^{-1}\right) \, dxdy \\
&= -\frac{h^2}{2}h^{-2} - \frac{h^2}{2}h^{-2} \\
&= -1,
\end{aligned}
$$

and by symmetry,

$$
a(\psi_{\boldsymbol{c}}, \psi_E) = a(\psi_{\boldsymbol{c}}, \psi_S) = a(\psi_{\boldsymbol{c}}, \psi_W) = a(\psi_{\boldsymbol{c}}, \psi_N) = -1.
$$

Finally,

$$
\begin{aligned}
a(\psi_{\boldsymbol{c}}, \psi_{NW}) &= \int_{III+IV} \nabla\psi_{\boldsymbol{c}} \cdot \nabla\psi_{NW} \, dxdy \\
&= \int_{III} \left(\frac{1}{h}\frac{\partial\psi_{NW}}{\partial x} + 0\frac{\partial\psi_{NW}}{\partial y}\right) \, dxdy \\
&\quad + \int_{IV} \left(0\frac{\partial\psi_{NW}}{\partial x} - \frac{1}{h}\frac{\partial\psi_{NW}}{\partial y}\right) \, dxdy \\
&= \int_{III} \left(\frac{1}{h}0 + 0\frac{1}{h}\right) \, dxdy + \int_{IV} \left(0\frac{-1}{h} - \frac{1}{h}0\right) \, dxdy \\
&= 0
\end{aligned}
$$

and again by symmetry

$$a(\psi_{\boldsymbol{c}}, \psi_{SE}) = \int_{III+IV} \nabla\psi_{\boldsymbol{c}} \cdot \nabla\psi_{NW} \, dxdy = 0.$$

Thus, the coefficients of A associated to the row corresponding to the node $\boldsymbol{c}$ can be represented as,

$$\begin{bmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{bmatrix}$$

according to the spatial location of the neighbor nodes. It is interesting that this coefficients are identical to the stencil of the second order cental finite difference discretization scheme for the Laplacian operator Δ.

The previous example introduces a uniform triangular partition for Ω, a generalization of this is the concept of *triangulation*.

Definition 3.2 (Triangulation)**.** A triangulation of the domain $\overline{\Omega} \subset \mathbb{R}^2$ is a partition $\mathcal{T}_h = \{K_i\}_{i=1}^M$ with $M < \infty$ into subsets K_i, which we call *elements*, that satisfy the following properties:

(1) $\overline{\Omega} = \bigcup_{K \in \mathcal{T}_h} K$.
(2) Each one of the *elements* $K \in \mathcal{T}_h$ is closed with non-empty interior and Lipschitz boundary.
(3) If $K_i \neq K_j$, $\{K_i, K_j\} \in \mathcal{T}_h$, then $\mathrm{int}(K_i) \cap \mathrm{int}(K_j) = \emptyset$, i.e. their topological interiors are disjoint.

For triangular elements (also for squared elements) we need the following regularity condition

(4) If $K_i \neq K_j$, $\{K_i, K_j\} \in \mathcal{T}_h$, then one of the following hold:
- $K_i \cap K_j = \emptyset$.
- K_i and K_j have a vertex in common.
- K_i and K_j have an edge in common.

We illustrate the last condition presented in the definition of triangulation in figure 10.

Furthermore, for polygonal triangulations, we need $\partial\Omega$ to be polygonal. Also, not every element of the triangulation will have the same "quality", this means that, under some metric, we can distinguish elements that are better than others, and so, they can lead to better approximations.

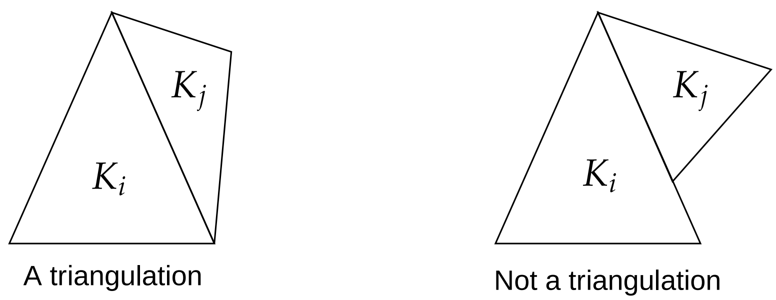


FIGURE 10. Regularity condition for triangular meshes.

Definition 3.3. For an arbitrary element K of a triangulation $\mathcal{T}_h$ we define,

$$h_K := \operatorname{diam}(K) = \max\{|\boldsymbol{x} - \boldsymbol{y}| \mid \boldsymbol{x}, \boldsymbol{y} \in K\}.$$

And

$$\rho_K := 2 \sup\{r \mid B_r(\boldsymbol{x}) \subset K, \boldsymbol{x} \in K\},$$

i.e. the diameter of the inscribed circle (of sphere) in the element K.

Remark 3.3.1. Notice that when K is a triangle, h_k corresponds to the longest edge.

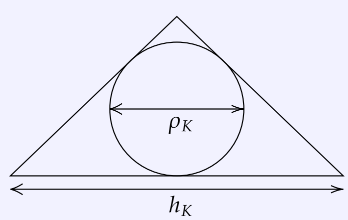


FIGURE 11. Illustration of h_K and ρ_K.

The previous definition formalizes the size h_K of an element K and the size $\max_K h_K$ of the triangulation $\mathcal{T}_h$. Also, the *condition number* of an element K is given by h_K/ρ_K, this is a measure for the quality of the element, if $h_K/\rho_K \to \infty$ the quality of the element gets deteriorated, meanwhile the quality of an element improves when $h_K/\rho_K \to 1$, as shown in picture 12

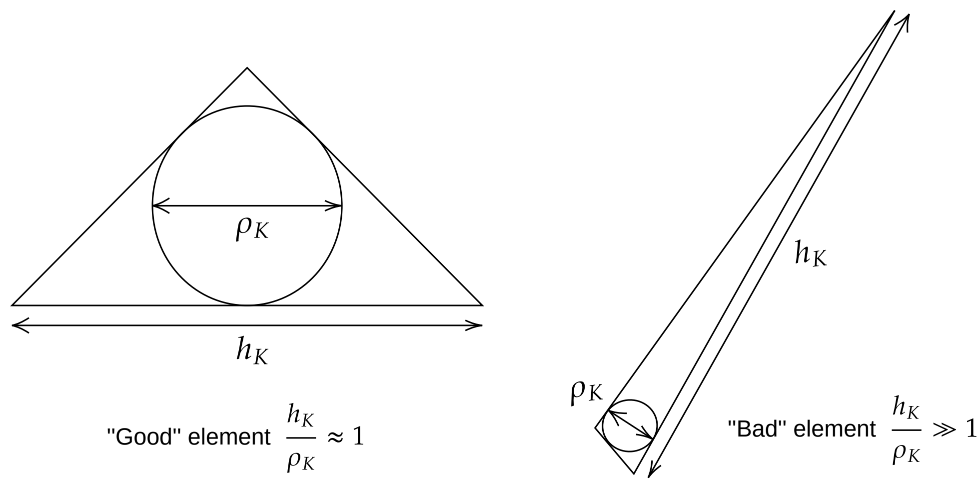


FIGURE 12. Element conditioning example.

Observe then that the ideal element is a equilateral triangle. From now on, we will focus our attention to the finite element method that considers linear and globally continuous basis functions (as introduced in the Courant example), in a domain $\Omega \subset \mathbb{R}^2$ and with triangulations based on triangles. To describe the implementation aspects of the finite element method, we will consider the following model problem

$$\begin{cases} -\Delta u + u = f, & \text{in } \Omega \\ \dfrac{\partial u}{\partial \hat{\boldsymbol{n}}} = g, & \text{on } \partial\Omega \end{cases}$$

We assume that Ω is Lipschitz and that $f \in L^2(\Omega), g \in L^2(\partial\Omega)$. Then, the variational formulation reads: Find $u \in H^1(\Omega)$ s.t.

$$a(u, v) = \ell(v) \qquad \forall v \in H^1(\Omega),$$

where,

$$a(u, v) = \int_\Omega \nabla u \cdot \nabla v + uv \; d\boldsymbol{x}$$

and

$$\ell(v) = \int_\Omega fv \; d\boldsymbol{x} + \int_{\partial\Omega} g\gamma v \; dS$$

As we have seen before, $a : H^1(\Omega) \times H^1(\Omega) \to \mathbb{R}$ is bilinear, continuous and coercive, meanwhile, $\ell \in H^{-1}(\Omega)$ and so, the variational formulation is well-posed. We will consider now the finite dimensional space inspired in the example 3.1.1, this space is well known in literature as the *Lagrange finite element space*,

$$V_h = \{v \in \mathcal{C}(\overline{\Omega}) \mid v|_K \in \mathcal{P}^1(\Omega),\ \forall K \in \mathcal{T}_h\},$$

here $\mathcal{T}_h$ is a triangulation for $\overline{\Omega}$ (based on triangles), analogously to figure 13

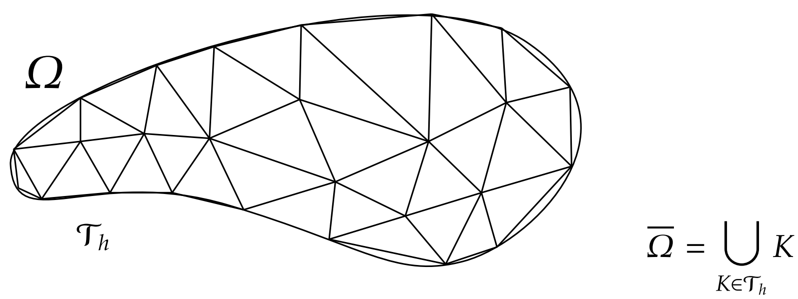


FIGURE 13. Example of triangulation based on triangles for a domain Ω.

For simplicity we assume that $\partial\Omega$ is a polygonal curve.

Proposition 3.4. *If $v|_K \in H^1(K)$ for all $K \in \mathcal{T}_h$ and $v \in \mathcal{C}(\overline{\Omega})$, then $v \in H^1(\Omega)$.*

Proof. See example 1.15.1. ■

Notice now that as $v \in V_h$, by definition $v \in \mathcal{C}(\overline{\Omega})$ and in particular

$$v|_{K_1}(K_1 \cap K_2) = v|_{K_2}(K_1 \cap K_2), \qquad \forall K_1, K_2 \in \mathcal{T}_h,\ \ K_1 \cap K_2 \neq \emptyset.$$

Let now $\{\boldsymbol{x}_i\}_{i=1}^N \subset \overline{\Omega}$ be the set of vertices associated with the collection $\{K_i\}_{i=1}^M = \mathcal{T}_h$. It follows, as we saw in example 3.1.1, that $\{\varphi_i\}_{i=1}^N$, where

$$\varphi_i \in V_h, \qquad \varphi_i(\boldsymbol{x}_j) = \delta_{ij}, \qquad 1 \leq i, j \leq N,$$

is an admissible basis for V_h. Observe also that for each element K such that $\boldsymbol{x}_i \in K$, we have that $\varphi_i|_K \in \mathcal{P}^1(K)$. In the case that $\boldsymbol{x}_i$ is not a vertex of K them $\varphi_i|_K = 0$.

This way, by construction we see that each $v \in V_h$ is uniquely determined by

$$v(\boldsymbol{x}) = \sum_{i=1}^{N} v_i \varphi_i(\boldsymbol{x})$$

where $v_i = v(\boldsymbol{x}_i)$ for all $i = 1, ..., N$. The Galerkin approximation problem that consists in finding $u_h \in V_h = \text{span}\ \{\varphi_i\}_{i=1}^N$ such that

$$\int_\Omega \nabla u_h \cdot \nabla v + u_h v \ d\boldsymbol{x} = \int_\Omega f v \ d\boldsymbol{x} + \int_{\partial\Omega} g v \ dS \qquad \forall v \in V_h.$$

is then equivalent to the linear system

$$\sum_{i=1}^{N} u_i \left(\int_\Omega \nabla\varphi_i \cdot \nabla\varphi_j + \varphi_i\varphi_j \ d\boldsymbol{x} \right) = \int_\Omega f\varphi_j \ d\boldsymbol{x} + \int_{\partial\Omega} g\varphi_j \ dS \qquad \forall j \in \{1, ..., N\}.$$

that uniquely determine the coefficients $\{u_i\}_{i=1}^N$. We rewrite this linear system as

$$A\boldsymbol{u} = \boldsymbol{b}$$

where $\boldsymbol{u} = (u_1, ..., u_N)^T$,

$$A_{ij} = \int_\Omega \nabla\varphi_i \cdot \nabla\varphi_j + \varphi_i\varphi_j \ d\boldsymbol{x} = \sum_{K\in\mathcal{T}_h} \int_K \nabla\varphi_i \cdot \nabla\varphi_j + \varphi_i\varphi_j \ d\boldsymbol{x}$$

and

$$b_j = \int_\Omega f\varphi_j \ d\boldsymbol{x} + \int_{\partial\Omega} g\varphi_j \ dS$$

for $i, j = 1, ..., N$. To calculate the coefficients of the matrix A for an arbitrary triangle $K \in \mathcal{T}_h$ we introduce the so called *barycentric coordinates*: Let $\boldsymbol{a}_j = (a_{1j}, a_{2j})^T$ for $j = 1, 2, 3$ be the vertices of the element K, then, the *barycentric coordinates* $\{\lambda_j\}_{j=1}^3$ associated with the vertices $\{\boldsymbol{a}_j\}_{j=1}^3$ are such that

$$\begin{aligned}
&\boldsymbol{x} \in K, \ \ \boldsymbol{x} = \lambda_1(\boldsymbol{x})\boldsymbol{a}_1 + \lambda_2(\boldsymbol{x})\boldsymbol{a}_2 + \lambda_3(\boldsymbol{x})\boldsymbol{a}_3, \\
&\sum_{i=1}^{3} \lambda_i(\boldsymbol{x}) = 1, \ \ \lambda_i(\boldsymbol{a}_j) = \delta_{ij}, \ \ i, j = 1, 2, 3.
\end{aligned}$$

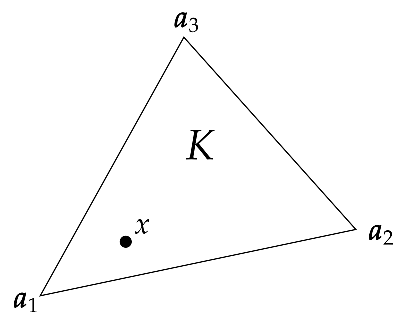


FIGURE 14. Example of vertices indexing for an element K.

The definition of the *barycentric coordinates* forces the following linear system to be satisfied

$$\begin{cases} a_{11}\lambda_1(\boldsymbol{x}) + a_{12}\lambda_2(\boldsymbol{x}) + a_{13}\lambda_3(\boldsymbol{x}) = x_1 \\ a_{21}\lambda_1(\boldsymbol{x}) + a_{22}\lambda_2(\boldsymbol{x}) + a_{23}\lambda_3(\boldsymbol{x}) = x_2 \\ \lambda_1(\boldsymbol{x}) + \lambda_2(\boldsymbol{x}) + \lambda_3(\boldsymbol{x}) = 1 \end{cases}$$

or equivalently,

$$\underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 1 & 1 & 1 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix}}_{\boldsymbol{\lambda}} = \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}.$$

The solution of this linear system is given by,

$$\lambda_1(\boldsymbol{x}) = \frac{1}{\det(A)} \begin{vmatrix} x_1 & a_{12} & a_{13} \\ x_2 & a_{22} & a_{23} \\ 1 & 1 & 1 \end{vmatrix}$$

$$\lambda_2(\boldsymbol{x}) = \frac{1}{\det(A)} \begin{vmatrix} a_{11} & x_1 & a_{13} \\ a_{21} & x_2 & a_{23} \\ 1 & 1 & 1 \end{vmatrix}$$

$$\lambda_3(\boldsymbol{x}) = \frac{1}{\det(A)} \begin{vmatrix} a_{11} & a_{12} & x_1 \\ a_{21} & a_{22} & x_2 \\ 1 & 1 & 1 \end{vmatrix}$$

observe that as $\det(A) = 2|K|$, then this solution can also be characterized as

$$\lambda_1(\boldsymbol{x}) = \frac{|k_1(\boldsymbol{x})|}{|K|}, \quad \lambda_2(\boldsymbol{x}) = \frac{|k_2(\boldsymbol{x})|}{|K|}, \quad \lambda_3(\boldsymbol{x}) = \frac{|k_3(\boldsymbol{x})|}{|K|}$$

where $\{k_1, k_2, k_3\}$ are the faces generated by the point $\boldsymbol{x} = (x_1, x_2) \in K$ and the vertices $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$ as shown in figure 15.

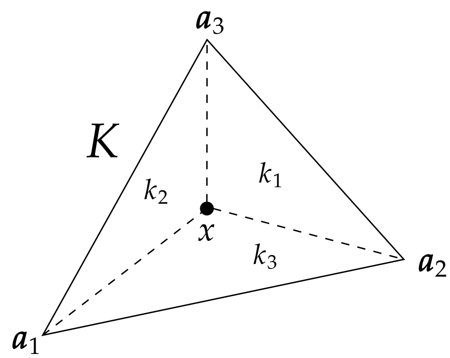


FIGURE 15. Faces $\{k_1, k_2, k_3\}$ generated by the point $\boldsymbol{x} \in K$.

From this construction we can see that

$$\varphi_i(\boldsymbol{x}) = \lambda_i(\boldsymbol{x}) \qquad i = 1, 2, 3,$$

as $\lambda_i(\boldsymbol{a}_j) = \delta_{ij}$. This holds for all $K \in \mathcal{T}_h$. Thus, over K, a function $v \in V_h$ is given by

$$v = \sum_{i=1}^{3} v(\boldsymbol{a}_i)\lambda_i(\boldsymbol{x}), \qquad \boldsymbol{x} \in K.$$

This allow us to calculate the integrals

$$\int_K \nabla\varphi_i \cdot \nabla\varphi_j \, d\boldsymbol{x} \quad \text{and} \quad \int_K \varphi_i\varphi_j \, d\boldsymbol{x},$$

that are necessary to obtain the coefficients of the matrix $A \in \mathbb{R}^{N\times N}$ of the linear system $A\boldsymbol{u} = \boldsymbol{b}$. On the other hand, the integrals

$$\int_\Omega f\varphi_j \, d\boldsymbol{x} \quad \text{and} \quad \int_{\partial\Omega} g\varphi_j \, dS, \qquad \text{for } f \in L^2(\Omega), g \in L^2(\partial\Omega),$$

needed to assemble the *load vector* $\boldsymbol{b}$, are usually calculated numerically using quadrature rules. For example, the quadrature rule

$$\int_K f(\boldsymbol{x})\, d\boldsymbol{x} = \frac{|K|}{3}(f(\boldsymbol{a}_1) + f(\boldsymbol{a}_2) + f(\boldsymbol{a}_3))$$

exactly integrates polynomia up to degree 2.

A different path to calculate the integrals needed to assemble the linear system $A\boldsymbol{u} = \boldsymbol{b}$ (and the most efficient) consist in using a *reference element* $\hat{K}$ that is easier to manipulate than the original element K. Then, using a linear transformation $F_K : \hat{K} \mapsto K$, one can recover the parameters of the element K. We fix the reference element $\hat{K}$ to be the triangle defined by the vertices $\hat{\boldsymbol{a}}_1 = (0, 0)$, $\hat{\boldsymbol{a}}_2 = (1, 0)$ and $\hat{\boldsymbol{a}}_3 = (0, 1)$, as illustrated in figure 16

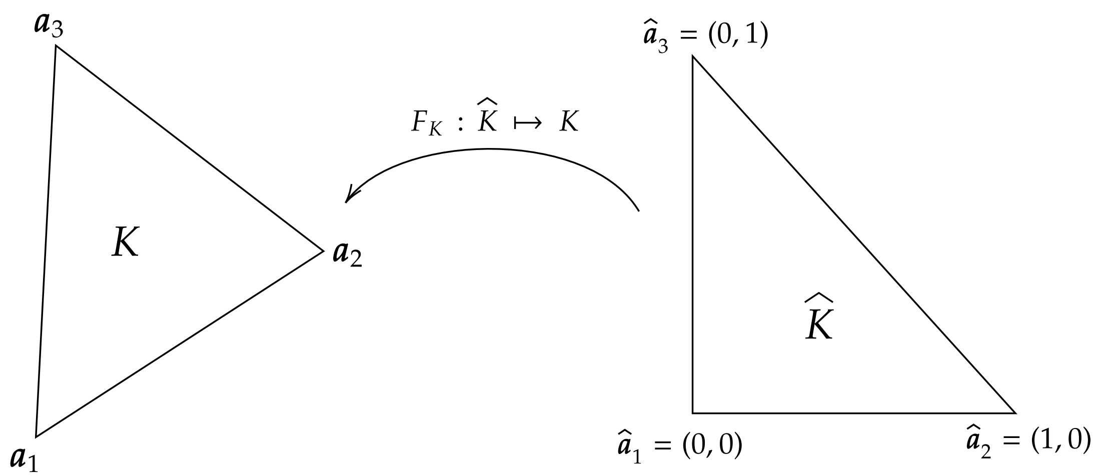


FIGURE 16. Reference element.

This way, the barycentric coordinates over $\hat{K}$ are given by

$$\hat{\lambda}_1 = 1 - \hat{x}_1 - \hat{x}_2, \qquad \hat{\lambda}_2 = \hat{x}_1, \qquad \hat{\lambda}_2 = \hat{x}_2,$$

where $\hat{\boldsymbol{x}} = (\hat{x}_1, \hat{x}_2)^T \in \hat{K}$. For an arbitrary triangle $K \in \mathcal{T}_h$ we then want to find an affine function $F_K : \hat{K} \mapsto K$ such that

$$\boldsymbol{a}_i = F_K(\hat{\boldsymbol{a}}_i), \qquad i = 1, 2, 3.$$

For $\hat{\boldsymbol{x}} \in \hat{K}$ this map looks like

$$\boldsymbol{x} = F_K(\hat{\boldsymbol{x}}) = \boldsymbol{b}_K + B_K\hat{\boldsymbol{x}},$$

for some $B_K \in \mathbb{R}^{2\times 2}$. As $\hat{\boldsymbol{a}}_1 = (0,0)$ we see that

$$\boldsymbol{a}_1 = F_K(\hat{\boldsymbol{a}}_1) = \boldsymbol{b}_K,$$

and so,

$$\boldsymbol{a}_2 = F_K(\hat{\boldsymbol{a}}_2) = \boldsymbol{a}_1 + B_K\hat{\boldsymbol{a}}_2 \quad \text{and} \quad \boldsymbol{a}_3 = F_K(\hat{\boldsymbol{a}}_3) = \boldsymbol{a}_1 + B_K\hat{\boldsymbol{a}}_3.$$

We put in matrix form this last two equations

$$B_K[\hat{\boldsymbol{a}}_2 \;\; \hat{\boldsymbol{a}}_3] = [\boldsymbol{a}_2 - \boldsymbol{a}_1 \;\; \boldsymbol{a}_3 - \boldsymbol{a}_1],$$

i.e.

$$B_K \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{12} - a_{11} & a_{13} - a_{11} \\ a_{22} - a_{21} & a_{23} - a_{21} \end{bmatrix}$$

and so,

$$\boldsymbol{x} = F_K(\hat{\boldsymbol{x}}) = \boldsymbol{a}_1 + [\boldsymbol{a}_2 - \boldsymbol{a}_1 \;\; \boldsymbol{a}_3 - \boldsymbol{a}_1]\hat{\boldsymbol{x}}, \qquad \hat{\boldsymbol{x}} \in \hat{K}, \boldsymbol{x} \in K.$$

One can verify that F_K is invertible, for this so, notice that

$$|\det(B_K)| = |(\boldsymbol{a}_3 - \boldsymbol{a}_1) \times (\boldsymbol{a}_2 - \boldsymbol{a}_1)| = 2|K|,$$

so B_K is invertible as long as $|K| = 0$, and so is F_K. Now we can use the affine transformation F_K to put in a bijection functions f defined over K with functions $\hat{f}$ defined over $\hat{K}$,

$$f(\boldsymbol{x}) = \hat{f}(\hat{\boldsymbol{x}}) = \left(\hat{f} \circ F_K^{-1}\right)(\boldsymbol{x}), \qquad \boldsymbol{x} \in K$$

or

$$\hat{f}(\hat{\boldsymbol{x}}) = f(F_K\hat{\boldsymbol{x}}) = (f \circ F_K)(\hat{\boldsymbol{x}}), \qquad \hat{\boldsymbol{x}} \in \hat{K}$$

Then,

$$\int_K f(\boldsymbol{x})\, d\boldsymbol{x} = \int_{\hat{K}} f(F_K\hat{\boldsymbol{x}})|\det(B_K)|\, d\hat{\boldsymbol{x}} = 2|K| \int_{\hat{K}} f(F_K\hat{\boldsymbol{x}})\, d\hat{\boldsymbol{x}}. \tag{10}$$

In the same fashion,

$$\begin{aligned} \nabla_{\hat{x}}\hat{f}(\hat{\boldsymbol{x}}) &= \nabla_{\hat{x}}[f(F_K\hat{\boldsymbol{x}})] \\ &= \left[\frac{\partial}{\partial \hat{x}_1} f\left((b_K)_1 + (B_K)_{11}\hat{x}_1 + (B_K)_{12}\hat{x}_2,\;\; (b_K)_2 + (B_K)_{21}\hat{x}_1 + (B_K)_{22}\hat{x}_2\right),\right. \end{aligned}$$

$$\left.\frac{\partial}{\partial \hat{x}_2} f\left((b_K)_1 + (B_K)_{11}\hat{x}_1 + (B_K)_{12}\hat{x}_2,\;\; (b_K)_2 + (B_K)_{21}\hat{x}_1 + (B_K)_{22}\hat{x}_2\right)\right]$$
$$= \nabla_{\boldsymbol{x}} f(F_K\hat{\boldsymbol{x}}) \cdot [(B_K)_{11} \;\; (B_K)_{21}] + \nabla_{\boldsymbol{x}} f(F_K\hat{\boldsymbol{x}}) \cdot [(B_K)_{12} \;\; (B_K)_{22}]$$
$$= \nabla_{\boldsymbol{x}} f(F_K\hat{\boldsymbol{x}}) B_K$$

and so can see that,

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \nabla_{\hat{\boldsymbol{x}}} f(\hat{\boldsymbol{x}}) B_K^{-1}. \tag{11}$$

Furthermore, as we know that the basis functions associated with the reference element are given by

$$\hat{\varphi}_i(\hat{\boldsymbol{x}}) = \hat{\lambda}_i(\hat{\boldsymbol{x}}), \qquad 1 = 1, 2, 3$$

we have that a linear function $\hat{v}$ defined over $\hat{K}$ can be represented as

$$\hat{v}(\boldsymbol{x}) = \sum_{i=1}^{N} \hat{v}(\hat{\boldsymbol{a}}_i)\hat{\lambda}_i(\hat{\boldsymbol{x}}),$$

this way, an arbitrary function $v \in V_h$ satisfies

$$v|_K(\boldsymbol{a}_i) = v|_K(F_K\hat{\boldsymbol{a}}_i) = \hat{v}(\hat{\boldsymbol{a}}_i)$$

for any $K \in \mathcal{T}_h$ with vertices $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$, hence

$$v|_K(\boldsymbol{x}) = \sum_{i=1}^{3} v|_K(\boldsymbol{a}_i)\underbrace{\varphi_i(\boldsymbol{x})}_{\lambda_i(\boldsymbol{x})} = \sum_{i=1}^{3} v|_K(\boldsymbol{a}_i)\underbrace{\varphi_i(F_K\hat{\boldsymbol{x}})}_{\lambda_i(F_K\hat{\boldsymbol{x}})} = \sum_{i=1}^{3} v|_K(\boldsymbol{a}_i)\underbrace{\hat{\varphi}_i(\hat{\boldsymbol{x}})}_{\hat{\lambda}_i(\hat{\boldsymbol{x}})}$$

for any $\boldsymbol{x} \in K$. Now, using equations (10) and (11) we can explicitly calculate the coefficients of the matrix $S^K \in \mathbb{R}^{3\times 3}$ defined by

$$\begin{aligned}
S_{ij}^K &= \int_K \nabla\varphi_i \cdot \nabla\varphi_j \; d\boldsymbol{x} \\
&= \int_{\hat{K}} \nabla\varphi_i(F_K\hat{\boldsymbol{x}}) \cdot \nabla\varphi_j(F_K\hat{\boldsymbol{x}}) \; |\det(B_K)| \; d\hat{\boldsymbol{x}} \\
&= |\det(B_K)| \int_{\hat{K}} \left(\nabla\hat{\lambda}_i(\hat{\boldsymbol{x}})B_K^{-1}\right) \cdot \left(\nabla\hat{\lambda}_j(\hat{\boldsymbol{x}})B_K^{-1}\right) \; d\hat{\boldsymbol{x}} \\
&= |\det(B_K)| \int_{\hat{K}} \nabla\hat{\lambda}_i(\hat{\boldsymbol{x}})^T \left(B_K^{-1}\right)^T \left(B_K^{-1}\right) \nabla\hat{\lambda}_j(\hat{\boldsymbol{x}}) \; d\hat{\boldsymbol{x}}.
\end{aligned}$$

For this so, we first notice that

$$B_K^{-1} = \frac{1}{\det(B_K)} \begin{bmatrix} a_{23} - a_{21} & a_{11} - a_{13} \\ a_{21} - a_{22} & a_{12} - a_{11} \end{bmatrix},$$

thus,

$$\begin{aligned} \left(B_K^{-1}\right)^T \left(B_K^{-1}\right) &= \frac{1}{|\det(B_K)|^2} \begin{bmatrix} a_{23} - a_{21} & a_{11} - a_{13} \\ a_{21} - a_{22} & a_{12} - a_{11} \end{bmatrix} \begin{bmatrix} a_{23} - a_{21} & a_{21} - a_{22} \\ a_{11} - a_{13} & a_{12} - a_{11} \end{bmatrix} \\ &= \frac{1}{|\det(B_K)|^2} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \end{aligned}$$

where

$$\begin{aligned} b_{11} &= (a_{23} - a_{21})^2 + (a_{11} - a_{13})^2 \\ b_{12} = b_{21} &= (a_{21} - a_{22})(a_{23} - a_{21}) + (a_{12} - a_{11})(a_{11} - a_{13}) \\ b_{22} &= (a_{21} - a_{22})^2 + (a_{12} - a_{11})^2. \end{aligned}$$

We use this to see that,

$$\begin{aligned} S_{11}^K &= |\det(B_K)| \int_{\hat{K}} \begin{bmatrix} 1 & 1 \end{bmatrix} \left(B_K^{-1}\right)^T \left(B_K^{-1}\right) \begin{bmatrix} 1 \\ 1 \end{bmatrix} \, d\hat{\boldsymbol{x}} \\ &= \frac{1}{2} \frac{|\det(B_K)|}{|\det(B_K)|^2} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \frac{1}{2|\det(B_K)|} \left[b_{11} + b_{12} + b_{21} + b_{22}\right] \\ &= \frac{1}{2|\det(B_K)|} \left[(a_{22} - a_{23})^2 + (a_{12} - a_{13})^2\right] \\ &= \frac{1}{2|\det(B_K)|} \|\boldsymbol{a}_3 - \boldsymbol{a}_2\|^2 \end{aligned}$$

by factoring a perfect square binomial, analogously one can verify,

$$S_{22}^K = \frac{1}{2|\det(B_K)|} \|\boldsymbol{a}_1 - \boldsymbol{a}_3\|^2, \qquad S_{33}^K = \frac{1}{2|\det(B_K)|} \|\boldsymbol{a}_1 - \boldsymbol{a}_2\|^2.$$

We now use the symmetry of $\left(B_K^{-1}\right)^T \left(B_K^{-1}\right)$ to compute

$$S_{12}^K = S_{21}^K = -|\det(B_K)| \int_{\hat{K}} \begin{bmatrix} 1 & 0 \end{bmatrix} \left(B_K^{-1}\right)^T \left(B_K^{-1}\right) \begin{bmatrix} 1 \\ 1 \end{bmatrix} \, d\hat{\boldsymbol{x}}$$

$$
\begin{aligned}
&= -\frac{1}{2}\frac{|\det(B_K)|}{|\det(B_K)|^2}\begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}b_{11} & b_{12}\\ b_{21} & b_{22}\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\\
&= -\frac{1}{2|\det(B_K)|}\left[b_{11}+b_{12}\right]\\
&= -\frac{1}{2|\det(B_K)|}\left[(a_{23}-a_{21})(a_{23}-a_{22})+(a_{11}-a_{13})(a_{12}-a_{13})\right]\\
&= -\frac{1}{2|\det(B_K)|}(\boldsymbol{a}_1-\boldsymbol{a}_3)\cdot(\boldsymbol{a}_2-\boldsymbol{a}_3).
\end{aligned}
$$

$$
\begin{aligned}
S^K_{13} = S^K_{31} &= -\frac{1}{2|\det(B_K)|}\begin{bmatrix}0 & 1\end{bmatrix}\begin{bmatrix}b_{11} & b_{12}\\ b_{21} & b_{22}\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\\
&= -\frac{1}{2|\det(B_K)|}\left[b_{21}+b_{22}\right]\\
&= -\frac{1}{2|\det(B_K)|}\left[(a_{21}-a_{22})(a_{23}-a_{22})+(a_{12}-a_{11})(a_{12}-a_{13})\right]\\
&= -\frac{1}{2|\det(B_K)|}(\boldsymbol{a}_1-\boldsymbol{a}_2)\cdot(\boldsymbol{a}_3-\boldsymbol{a}_2).
\end{aligned}
$$

$$
\begin{aligned}
S^K_{23} = S^K_{32} &= \frac{1}{2|\det(B_K)|}\begin{bmatrix}0 & 1\end{bmatrix}\begin{bmatrix}b_{11} & b_{12}\\ b_{21} & b_{22}\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}\\
&= \frac{1}{2|\det(B_K)|}b_{21}\\
&= \frac{1}{2|\det(B_K)|}\left[(a_{21}-a_{22})(a_{23}-a_{21})+(a_{12}-a_{11})(a_{11}-a_{13})\right]\\
&= \frac{1}{2|\det(B_K)|}(\boldsymbol{a}_3-\boldsymbol{a}_1)\cdot(\boldsymbol{a}_1-\boldsymbol{a}_2).
\end{aligned}
$$

All together, the matrix S^K is given by

$$
S^K = \frac{1}{4|K|}\begin{bmatrix}
\|\boldsymbol{a}_3-\boldsymbol{a}_2\|^2 & -(\boldsymbol{a}_1-\boldsymbol{a}_3)\cdot(\boldsymbol{a}_2-\boldsymbol{a}_3) & -(\boldsymbol{a}_1-\boldsymbol{a}_2)\cdot(\boldsymbol{a}_3-\boldsymbol{a}_2)\\
-(\boldsymbol{a}_1-\boldsymbol{a}_3)\cdot(\boldsymbol{a}_2-\boldsymbol{a}_3) & \|\boldsymbol{a}_1-\boldsymbol{a}_3\|^2 & (\boldsymbol{a}_3-\boldsymbol{a}_1)\cdot(\boldsymbol{a}_1-\boldsymbol{a}_2)\\
-(\boldsymbol{a}_1-\boldsymbol{a}_2)\cdot(\boldsymbol{a}_3-\boldsymbol{a}_2) & (\boldsymbol{a}_3-\boldsymbol{a}_1)\cdot(\boldsymbol{a}_1-\boldsymbol{a}_2) & \|\boldsymbol{a}_1-\boldsymbol{a}_2\|^2
\end{bmatrix}.
$$

Now we calculate the matrix $M^K \in \mathbb{R}^{3\times 3}$ such that

$$
M^K = \int_K \varphi_i(\boldsymbol{x})\varphi_j(\boldsymbol{x})\,d\boldsymbol{x}
$$

$$= |\det(B_K)| \int_{\hat{K}} \hat{\lambda}_i(\hat{\boldsymbol{x}})\hat{\lambda}_j(\hat{\boldsymbol{x}}) \, d\hat{\boldsymbol{x}}.$$

We first start with the diagonal,

$$\begin{aligned} M_{11}^K &= |\det(B_K)| \int_{\hat{K}} (1-\hat{x}_1-\hat{x}_2)(1-\hat{x}_1-\hat{x}_2) \, d\hat{\boldsymbol{x}} \\ &= |\det(B_K)| \int_0^1 \int_0^1 (1-\hat{x}_1-\hat{x}_2)(1-\hat{x}_1-\hat{x}_2) \, d\hat{x}_2 \, d\hat{x}_1 \\ &= \frac{|\det(B_K)|}{12} \\ &= \frac{|K|}{6} \end{aligned}$$

and notice that $M_{11}^K = M_{22}^K = M_{33}^K$. Now, observe that

$$\begin{aligned} M_{23}^K = M_{32}^K &= |\det(B_K)| \int_{\hat{K}} \hat{x}_1\hat{x}_2 \, d\hat{\boldsymbol{x}} \\ &= |\det(B_K)| \int_0^1 \int_0^1 \hat{x}_1\hat{x}_2 \, d\hat{x}_2 \, d\hat{x}_1 \\ &= \frac{|\det(B_K)|}{24} \\ &= \frac{|K|}{12}, \end{aligned}$$

and analogously to the diagonal case notice that

$$M_{23}^K = M_{32}^K = M_{12}^K = M_{21}^K = M_{13}^K = M_{31}^K,$$

so we obtain

$$M^K = \frac{|K|}{12} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

Finally, as

$$\int_K \nabla\varphi_i \cdot \nabla\varphi_j + \varphi_i\varphi_j \, d\boldsymbol{x} = S_{ij}^K + M_{ij}^K.$$

Thus, the finite element matrix A associated with our Galerkin problem $A\boldsymbol{u} = \boldsymbol{b}$ is given by,

$$A_{ij} = \int_\Omega \nabla\varphi_i \cdot \nabla\varphi_j + \varphi_i\varphi_j \, d\boldsymbol{x} = \sum_{K\in\mathcal{T}_h} \int_K \nabla\varphi_i \cdot \nabla\varphi_j + \varphi_i\varphi_j \, d\boldsymbol{x} = \sum_{K\in\mathcal{T}_h} S_{ij}^K + M_{ij}^K.$$

So we can see that in order to assemble the matrix $A \in \mathbb{R}^{N\times N}$ it suffices to iterate over the elements of the triangulation. An example of this is the following algorithm

Algorithm 1 Finite element matrix assembling

Require: N (# of nodes); $\mathcal{T}_h$ (triangulation)
Ensure: Matrix A
$A \leftarrow \text{zeros}(N, N)$
for $K \in \mathcal{T}_h$ **do**
    $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\} \leftarrow \text{vertices}(K)$
    $B_K \leftarrow \begin{bmatrix} \boldsymbol{a}_2 - \boldsymbol{a}_1 & \boldsymbol{a}_3 - \boldsymbol{a}_1 \end{bmatrix}$
    $|K| \leftarrow \frac{1}{2}|\det(B_K)|$
    Assemble M^K using $(\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}, |K|)$
    Assemble S^K using $(\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}, |K|)$
    for $i = 1, 2, 3$ **do**
        for $j = 1, 2, 3$ **do**
            $A_{ij} \leftarrow A_{ij} + S_{ij}^K + M_{ij}^K$
        end for
    end for
end for
return A

Globally, the matrix

$$S_{ij} = \sum_{K\in\mathcal{T}_h} S_{ij}^K$$

is known in literature as *stiffness* matrix, meanwhile, the matrix

$$M_{ij} = \sum_{K\in\mathcal{T}_h} S_{ij}^K$$

is known as *mass* matrix.

3.1. **Error analysis of the Finite Element Method.** We now look for error bounds for the solution approximations generated by the finite element method (we restrict ourselves to the Lagrange FEM case), we want this bounds to depend on the mesh size h. For this so we need some results that we present in what follows

Lemma 3.5. *Let $F_K : \hat{K} \mapsto K$ be an affine transformation. Then,*

$$\|B_K\| \leq \frac{h_K}{\hat{\rho}} \quad \text{and} \quad \|B_K^{-1}\| \leq \frac{\hat{h}}{\rho_K}$$

where $h_K = diam(K)$, $\hat{h} = diam(\hat{K})$, ρ_K is the diameter of the inscribed circle in the element K, and $\hat{\rho}$ is the diameter of the inscribed circle in the element $\hat{K}$.

Proof. Let $\hat{S}$ be the inscribed open ball (circle) in the element $\hat{K}$.

$$\begin{aligned}
\|B_K\| &= \sup_{\hat{\boldsymbol{x}} \neq 0} \frac{\|B_K \hat{\boldsymbol{x}}\|}{\|\hat{\boldsymbol{x}}\|} \\
&= \sup_{\|\hat{\boldsymbol{z}}\| = \hat{\rho}} \frac{\|B_K \hat{\boldsymbol{z}}\|}{\|\hat{\boldsymbol{z}}\|} \\
&= \hat{\rho}^{-1} \sup_{\|\hat{\boldsymbol{z}}\| = \hat{\rho}} \|B_K \hat{\boldsymbol{z}}\| \\
&= \hat{\rho}^{-1} \sup_{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}} \in \hat{S}} \|B_K (\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}})\| \\
&= \hat{\rho}^{-1} \sup_{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}} \in \hat{S}} \|(\underbrace{B_K \hat{\boldsymbol{x}} + b_K}_{\boldsymbol{x} \in K}) - (\underbrace{b_K + B_K \hat{\boldsymbol{y}}}_{\boldsymbol{y} \in K})\| \\
&= \hat{\rho}^{-1} \sup_{\boldsymbol{x}, \boldsymbol{y} \in K} \|\boldsymbol{x} - \boldsymbol{y}\| \\
&= \hat{\rho}^{-1} h_K
\end{aligned}$$

So we obtain,

$$\|B_K\| \leq \frac{h_K}{\hat{\rho}},$$

this bound being sharp. The second bound follows from a similar argument. ■

Definition 3.6. We define the *interpolation operator* associated with a reference element as, $\hat{\Pi} : \mathcal{C}(\hat{K}) \longmapsto \mathcal{P}^1(\hat{K})$, such that,

$$\hat{\Pi}\hat{v}(\boldsymbol{x}) = \sum_{i=1}^{3} \hat{v}(\hat{\boldsymbol{a}}_i)\hat{\varphi}_i(\hat{\boldsymbol{x}}).$$

We fix the reference element to be defined by the vertices $\hat{\boldsymbol{a}}_1 = (0,0)$, $\hat{\boldsymbol{a}}_2 = (1,0)$ and $\hat{\boldsymbol{a}}_3 = (0,1)$, with basis functions given by $\hat{\varphi}_1(\hat{\boldsymbol{x}}) = 1 - \hat{x}_1 - \hat{x}_2$, $\hat{\varphi}_1(\hat{\boldsymbol{x}}) = \hat{x}_1$ and $\hat{\varphi}_1(\hat{\boldsymbol{x}}) = \hat{x}_2$.
Then, the *interpolation operator* $\Pi : \mathcal{C}(K) \longmapsto \mathcal{P}^1(K)$ associated with an arbitrary element $K \in \mathcal{T}_h$, is defined by

$$\Pi_K v(\boldsymbol{x}) = \sum_{i=1}^{3} v(\boldsymbol{a}_i)\varphi_i^K(\boldsymbol{x}).$$

Here, $\{\varphi_i^K\}_{i=1}^3$ are the basis functions that correspond to the vertices $\{\boldsymbol{a}_i\}_{i=1}^3$, of the element $K \in \mathcal{T}_h$. By construction,

$$\varphi_i^K(\boldsymbol{x}) = \left(\hat{\varphi}_i \circ F_K^{-1}\right)(\boldsymbol{x}) \quad \text{and} \quad \hat{\varphi}_i^K(\hat{\boldsymbol{x}}) = \left(\varphi_i^K \circ F_K\right)(\boldsymbol{x}),$$

so we can see that

$$\hat{\Pi}\hat{v} = (\Pi_K v) \circ F_K, \quad \text{i.e.} \quad (\hat{\Pi}\hat{v})(\hat{\boldsymbol{x}}) = (\Pi_K v)(F_K \hat{\boldsymbol{x}}).$$

Theorem 3.7. *Let $0 \leq m \leq 2$, then there exists a constant $c > 0$ such that,*

$$|\hat{v} - \hat{\Pi}\hat{v}|_{H^m(\hat{K})} \leq c|\hat{v}|_{H^2(\hat{K})}, \qquad \forall \hat{v} \in H^2(\hat{K}).$$

Proof. By the Rellich-Kondrachov immersion theorem 1.20 we have that

$$H^2(\hat{K}) \underset{c}{\hookrightarrow} \mathcal{C}(\hat{K})$$

as $2 > d/2 = 1$, and so, the interpolation operator $\hat{\Pi} : \mathcal{C}(\hat{K}) \longmapsto \mathcal{P}^1(\hat{K})$ is defined for functions $\hat{v} \in H^2(\hat{K})$. Moreover,

$$\begin{aligned} \|\hat{\Pi}\hat{v}\|_{H^m(\hat{K})} &= \left\| \sum_{i=1}^{3} \hat{v}(\hat{\boldsymbol{a}}_i)\hat{\varphi}_i \right\|_{H^m(\hat{K})} \\ &\leq \sum_{i=1}^{3} |\hat{v}(\hat{\boldsymbol{a}}_i)| \|\hat{\varphi}_i\|_{H^m(\hat{K})} \end{aligned}$$

$$\leq \max_{\hat{\boldsymbol{x}}\in\hat{K}} |\hat{v}(\hat{\boldsymbol{x}})| \underbrace{\sum_{i=1}^{3} \|\hat{\varphi}_i\|_{H^m(\hat{K})}}_{\text{constant } \hat{c}}$$
$$= \hat{c}\|\hat{v}\|_{\mathcal{C}(\hat{K})}$$
$$\leq \tilde{c}\|\hat{v}\|_{H^2(\hat{K})},$$

for any fixed $m \in [0, 2]$ and for some $\tilde{c}, \hat{c} > 0$. This shows that $\hat{\Pi} : H^2(\hat{K}) \mapsto H^m(\hat{K})$ is not only linear but also continuous. Furthermore, Rellich-Kondrachov theorem also implies that

$$H^2(\hat{K}) \hookrightarrow\hookrightarrow H^m(\hat{K}),$$

in particular, the identity map

$$\iota : H^2(\hat{K}) \mapsto H^m(\hat{K})$$

is bounded by, lets say, $c_\iota > 0$. Thus,

$$\begin{aligned}|\hat{v} - \hat{\Pi}\hat{v}|_{H^m(\hat{K})} &\leq \|\hat{v} - \hat{\Pi}\hat{v}\|_{H^m(\hat{K})} \\ &\leq \|\hat{v}\|_{H^m(\hat{K})} + \|\hat{\Pi}\hat{v}\|_{H^m(\hat{K})} \\ &\leq c_\iota\|\hat{v}\|_{H^2(\hat{K})} + \tilde{c}\|\hat{v}\|_{H^m(\hat{K})} \\ &\leq c_f\|\hat{v}\|_{H^2(\hat{K})},\end{aligned}$$

for some $c_f > 0$. This shows that the function $f : H^m(\hat{K}) \mapsto \mathbb{R}$ defined by,

$$f(\hat{v}) := |\hat{v} - \hat{\Pi}\hat{v}|_{H^m(\hat{K})}$$

is not only linear, but continuous with constant $c_f > 0$. On the other hand, by definition, we can see that

$$\hat{\Pi}\hat{p} = \hat{p}, \qquad \forall \hat{p} \in \mathcal{P}^1(\hat{K}),$$

so $f(\hat{p}) = 0$ for all $\hat{p} \in \mathcal{P}^1(\hat{K})$. This way f meets the requirements of the Bramble-Hilbert lemma 1.22, this implies the existence of a positive constant $c > 0$ such that

$$|f(\hat{v})| = |\hat{v} - \hat{\Pi}\hat{v}|_{H^m(\hat{K})} \leq c|\hat{v}|_{H^2(\hat{K})}, \qquad \forall \hat{v} \in H^2(\hat{K}).$$

■

We now want to inspect what happens with the interpolation operator $\Pi_K : \mathcal{C}(K) \mapsto \mathcal{P}^1(K)$ over an arbitrary $K \in \mathcal{T}_h$.

Theorem 3.8. *Let $F_K : \hat{K} \mapsto K$ such that $F_K\hat{\boldsymbol{x}} = B_K\hat{\boldsymbol{x}} + b_K$ be a bijection from $\hat{K}$ onto K. Then,*

$$v \in H^m(K) \Leftrightarrow \hat{v} = v \circ F_K \in H^m(\hat{K}), \qquad \forall m \geq 0.$$

Furthermore, there exists $c > 0$ independent of $\hat{K}$ and K such that

$$|\hat{v}|_{H^m(\hat{K})} \leq c\|B_K\|^m |det(B_K)|^{-1/2} |v|_{H^m(K)}$$

and

$$|v|_{H^m(K)} \leq c\|B_K^{-1}\|^m |det(B_K)|^{1/2} |\hat{v}|_{H^m(\hat{K})}$$

Proof. Consider the following notation for an arbitrary multi-index $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ and $|\boldsymbol{\alpha}| = \alpha_1 + \alpha_2$,

$$\partial_{\hat{\boldsymbol{x}}}^{\boldsymbol{\alpha}} = \frac{\partial^{|\boldsymbol{\alpha}|}}{\partial \hat{x}_1^{\alpha_1} \partial \hat{x}_2^{\alpha_2}} \quad \text{and} \quad \partial_{\boldsymbol{x}}^{\boldsymbol{\alpha}} = \frac{\partial^{|\boldsymbol{\alpha}|}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2}}.$$

Now, by definition of the Sobolev semi-norm in multi-index notation,

$$|\hat{v}|^2_{H^m(\hat{K})} = \sum_{|\boldsymbol{\alpha}|=m} \int_{\hat{K}} \left|\partial_{\hat{x}}^{\boldsymbol{\alpha}} \hat{v}(\hat{\boldsymbol{x}})\right|^2 d\hat{\boldsymbol{x}},$$

We now perform the change of variables $\boldsymbol{x} = F_K(\hat{\boldsymbol{x}})$, for which

$$d\hat{\boldsymbol{x}} = |\det B_K|^{-1}\, d\boldsymbol{x}, \quad \hat{v}(\hat{\boldsymbol{x}}) = v(\boldsymbol{x}).$$

Thus,

$$|\hat{v}|^2_{H^m(\hat{K})} = \sum_{|\boldsymbol{\alpha}|=m} \int_K \left|\partial_{\hat{\boldsymbol{x}}}^{\boldsymbol{\alpha}} (\hat{v} \circ F_K^{-1})(\boldsymbol{x})\right|^2 |\det B_K|^{-1}\, d\boldsymbol{x}.$$

To compute $\partial_{\hat{\boldsymbol{x}}}^{\boldsymbol{\alpha}} \hat{v}$, note that for a single derivative ($m = 1$), the chain rule gives

$$\frac{\partial \hat{v}}{\partial \hat{x}_i}(\hat{\boldsymbol{x}}) = \sum_{j=1}^{2} (B_K)_{ji} \frac{\partial v}{\partial x_j}(F_K(\hat{\boldsymbol{x}})).$$

In vector form,

$$\nabla_{\hat{\boldsymbol{x}}} \hat{v}(\hat{\boldsymbol{x}}) = B_K \nabla_{\boldsymbol{x}} v(F_K(\hat{\boldsymbol{x}})).$$

For second derivatives ($m = 2$), applying the chain rule again yields

$$\frac{\partial^2 \hat{v}}{\partial \hat{x}_i \partial \hat{x}_k}(\hat{\boldsymbol{x}}) = \sum_{j,\ell=1}^{2} (B_K)_{ji} (B_K)_{\ell k} \frac{\partial^2 v}{\partial x_j \partial x_\ell}(F_K(\hat{\boldsymbol{x}})).$$

We see that each term is a product of m entries of B_K times a derivative of v of total order m. In general, for $|\boldsymbol{\alpha}| = m$, repeated application of the chain rule gives

$$\partial_{\hat{\boldsymbol{x}}}^{\boldsymbol{\alpha}} \hat{v}(\hat{\boldsymbol{x}}) = \sum_{|\boldsymbol{\beta}|=m} c_{\boldsymbol{\alpha},\boldsymbol{\beta}}(B_K)\, \partial_{\boldsymbol{x}}^{\boldsymbol{\beta}} v(F_K(\hat{\boldsymbol{x}})),$$

where each coefficient $c_{\boldsymbol{\alpha},\boldsymbol{\beta}}(B_K)$ is a finite sum of products of m entries of the matrix B_K. Since $|(B_K)_{ij}| \leq \|B_K\|$ for all i, j,

$$|c_{\boldsymbol{\alpha},\boldsymbol{\beta}}(B_K)| \leq C_{m,n} \|B_K\|^m,$$

where $C_{m,n}$ depends only on m and the spatial dimension $n = 2$. The formalization of this induction concept is known in literature as *Faà di Bruno formulae*. Now, substituting into the expression for $|\hat{v}|^2_{H^m(\hat{K})}$ and using the bound above:

$$\begin{aligned} |\hat{v}|^2_{H^m(\hat{K})} &\leq C_{m,n}\, \|B_K\|^{2m}\, |\det B_K|^{-1} \sum_{|\boldsymbol{\beta}|=m} \int_K |\partial_{\boldsymbol{x}}^{\boldsymbol{\beta}} v(\boldsymbol{x})|^2 d\boldsymbol{x} \\ &= C_{m,n}\, \|B_K\|^{2m}\, |\det B_K|^{-1}\, |v|^2_{H^m(K)}. \end{aligned}$$

Taking square roots gives

$$|\hat{v}|_{H^m(\hat{K})} \leq c\, \|B_K\|^m\, |\det B_K|^{-1/2}\, |v|_{H^m(K)},$$

with $c = \sqrt{C_{m,n}}$. The reverse inequality follows by applying the same reasoning to the inverse mapping $F_K^{-1} : K \to \hat{K}$, which replaces B_K by B_K^{-1} and swaps K with $\hat{K}$. ■

Theorem 3.9. *Let $\Pi_K : \mathcal{C}(K) \mapsto \mathcal{P}^1(K)$ be the interpolation operator associated with the element $K \in \mathcal{T}_h$. Then, there exists a constant $c > 0$, that depends only on $\hat{K}$ and $\hat{\Pi}$, such that*

$$|v - \Pi_K v|_{H^m(K)} \leq c \frac{h_K^2}{\rho_K^m} |v|_{H^2(K)} \qquad \forall v \in H^2(K),\ 0 \leq m \leq 2.$$

Proof. We first notice that

$$\hat{v} - \Pi\hat{v} = (v - \Pi v) \circ F_K,$$

then, by Theorem 3.8, it follows that,

$$|v - \Pi v|_{H^m(K)} \leq c\|B_K^{-1}\|^m |\det(B_K)|^{1/2} |\hat{v} - \hat{\Pi}\hat{v}|_{H^m(\hat{K})},$$

furthermore, Theorem 3.7 gives us,

$$|v - \Pi v|_{H^m(K)} \leq c\|B_K^{-1}\|^m |\det(B_K)|^{1/2} |\hat{v}|_{H^2(\hat{K})}.$$

But can bound the right hand side term $|\hat{v}|_{H^2(\hat{K})}$ by using again Theorem 3.8:

$$|\hat{v}|_{H^2(\hat{K})} \leq c\|B_K\|^2 |\det(B_K)|^{-1/2} |v|_{H^2(K)}.$$

Thus,

$$|v - \Pi v|_{H^m(K)} \leq c\|B_K^{-1}\|^m \|B_K\|^2 |v|_{H^2(K)},$$

and finally an application of Lemma 3.5 yields

$$|v - \Pi v|_{H^m(K)} \leq c\frac{h_K^2}{\rho_K^m} |v|_{H^2(K)}.$$

■

Now we want to start addressing global error bounds for the finite element method, for this so, in what follows, we will restrict ourselves to the case in which the related triangulation $\mathcal{T}_h$ is *regular*.

Definition 3.10. A family if meshes $\{\mathcal{T}_h\}_{h>0}$ is said to be regular if,
- There exists $\sigma > 0$ such that $h_K/\rho_K \leq \sigma$, $\forall K \in \mathcal{T}_h$, $\forall h > 0$.
- $\{\mathcal{T}_h\}_{h>0}$ is well defined in the limit $h \to 0$.

So, under the conditions of Theorem 3.9, we have seen that,

$$|v - \Pi_K v|_{H^m(K)} \leq c\frac{h_K^2}{\rho_K^m} |v|_{H^2(K)} \qquad \forall v \in H^2(K),\ 0 \leq m \leq 2.$$

But observe that we can construct the norm $\|\cdot\|_{H^m(K)}$ by just adding up semi-norms of the form $|\cdot|_{H^m(K)}$, for example, if $m = 3/2$,

$$\|\cdot\|_{H^{3/2}(K)}^2 = |\cdot|_{H^0(K)}^2 + |\cdot|_{H^1(K)}^2 + |\cdot|_{H^{1/2}(K)}^2.$$

So we can extend the bound from Theorem 3.9 to,

$$\|v - \Pi_K v\|_{H^m(K)} \leq c\frac{h_K^2}{\rho_K^m} |v|_{H^2(K)} \qquad \forall v \in H^2(K),\ 0 \leq m \leq 2,$$

where $c > 0$ remains depending only on $\hat{K}$ and $\hat{\Pi}$. Then, for a family of regular triangulations,

$$\begin{aligned}\|v - \Pi_K v\|_{H^m(K)} &\leq c h_K^{2-m} \frac{h_K^m}{\rho_K^m} |v|_{H^2(K)} \\ &\leq c h_K^{2-m} \sigma^m |v|_{H^2(K)} \\ &= c h_K^{2-m} |v|_{H^2(K)}.\end{aligned}$$

We now define the global interpolation in an element-wise fashion, as

$$(\Pi_h v)|_K := \Pi_K v \qquad \forall K \in \mathcal{T}_h,\ v \in V_h.$$

And so if $\{\boldsymbol{x}_i\}_{i=1}^N$ are the triangulation nodes as introduced in section 3, then,

$$\Pi_h v(\boldsymbol{x}) = \sum_{i=1}^{N} v(\boldsymbol{x}_i) \varphi_i(\boldsymbol{x}) \qquad (\varphi_i(\boldsymbol{x}_j) = \delta_{ij}),$$

where $V_h = \operatorname{span}\{\varphi_i\}_{i=1}^N$. In this notation,

$$\varphi_i = \varphi_i^K = \hat{\varphi}_i \circ F_K \qquad \forall K \in \mathcal{T}_h.$$

Recall now Cea's lemma 2.14,

$$\|u - u_h\|_{H^1(\Omega)} \leq c \inf_{v_h \in V_h} \|u - v_h\|_{H^1(\Omega)},$$

where $c > 0$, $u \in V$ is solution of

$$a(u, v) = \ell(v) \qquad \forall v \in V \subset H^1(\Omega)$$

and $u_h \in V_h$ solves

$$a(u_h, v) = \ell(v) \qquad \forall v \in V_h \subset H^1(\Omega).$$

Then we can see that

$$\|u - u_h\|_{H^1(\Omega)} \leq c \inf_{v_h \in V_h} \|u - v_h\|_{H^1(\Omega)} \leq c \|u - \Pi_h u\|_{H^1(\Omega)},$$

since $\Pi_h u \in V_h$, as $(\Pi_h u)(\boldsymbol{x}) = \sum_{j=1}^N u(\boldsymbol{x}_j) \varphi_j(\boldsymbol{x})$ and $V_h = \operatorname{span}\{\varphi_j\}_{j=1}^N$. This way we notice that if we want to find a bound for the global FEM error $\|u - u_h\|_{H^1(\Omega)}$ in terms of h, it suffices to estimate an upper bound in terms of h for the global interpolation error $\|u - \Pi_h u\|_{H^1(\Omega)}$.

Theorem 3.11. *Let $\{\mathcal{T}_h\}_{h>0}$ be a family of regular triangulations for a domain $\overline{\Omega}$. Then there exists a constant $c > 0$ independent of $h > 0$ such that,*

$$\|v - \Pi_h v\|_{H^m(\Omega)} \leq ch^{2-m}|v|_{H^2(\Omega)} \qquad \forall v \in H^2(\Omega),\ 0 \leq m \leq 2.$$

Proof. We have seen that

$$\|v - \Pi_K v\|_{H^m(K)} \leq ch_K^{2-m}|v|_{H^2(K)}$$

Then,

$$\begin{aligned}
\|v - \Pi_h v\|^2_{H^m(\Omega)} &= \sum_{K\in\mathcal{T}_h} \|v - \Pi_K v\|^2_{H^m(K)} \\
&\leq c^2 \sum_{K\in\mathcal{T}_h} h_K^{2(2-m)}|v|^2_{H^2(K)} \\
&\leq c^2 h^{2(2-m)} \sum_{K\in\mathcal{T}_h} |v|^2_{H^2(K)} \\
&= c^2 h^{2(2-m)}|v|^2_{H^2(\Omega)}
\end{aligned}$$

■

We use this result to show the following theorem.

Theorem 3.12. *Let $\{V_h\}_{h>0} \subset H^1(\Omega)$ be a family of finite element spaces that correspond to a family of triangulations $\{\mathcal{T}_h\}_{h>0}$ of $\overline{\Omega}$. Then, the finite element method converges, i.e.,*

$$\|u - u_h\|_{H^1(\Omega)} \xrightarrow[h\to 0]{} 0$$

Moreover, if $u \in H^2(\Omega)$, then there exists $c > 0$ such that,

$$\|u - u_h\|_{H^1(\Omega)} \leq ch|u|_{H^2(\Omega)}.$$

Proof. As we have seen before, by Cea's lemma 2.14, there exists $\tilde{c} > 0$

$$\|u - u_h\|_{H^1(\Omega)} \leq \tilde{c}\|u - \Pi_h u\|_{H^1(\Omega)}.$$

Also, if $u \in H^2(\Omega)$, by Theorem 3.11,

$$\|u - \Pi_h u\|_{H^1(\Omega)} \leq \hat{c}h|u|_{H^2(\Omega)}$$

for some $\hat{c} > 0$. Thus,

$$\|u - u_h\|_{H^1(\Omega)} \leq ch|u|_{H^2(\Omega)},$$

with $c := \tilde{c}\hat{c}$, and so, the finite element method converges. Now, assume that $u \in H^1(\Omega)$ but $u \notin H^2(\Omega)$. We recall that $\mathcal{C}^\infty(\Omega)$ is dense in $H^1(\Omega)$ in the H^1-topology, and so, given $\varepsilon > 0$, there exists $u_\varepsilon \in \mathcal{C}^\infty(\Omega)$ such that,

$$\|u - u_\varepsilon\|_{H^1(\Omega)} < \frac{\varepsilon}{2}$$

Now, observe that as $\mathcal{C}^\infty(\Omega) \subset H^2(\Omega)$, then $u_\varepsilon \in H^2(\Omega)$, and so, defining $v_\varepsilon := \Pi_h u_\varepsilon$, by Theorem 3.11, we see that

$$\|u_\varepsilon - v_\varepsilon\|_{H^1(\Omega)} \leq ch|u_\varepsilon|_{H^2(\Omega)},$$

hence, letting

$$h < \frac{\varepsilon}{2c|u_\varepsilon|_{H^2(\Omega)}},$$

we obtain

$$\|u_\varepsilon - v_\varepsilon\|_{H^1(\Omega)} \leq \frac{\varepsilon}{2}.$$

Thus,

$$\|u - v_\varepsilon\|_{H^1(\Omega)} \leq \|u - u_\varepsilon\|_{H^1(\Omega)} + \|u_\varepsilon - v_\varepsilon\|_{H^1(\Omega)} < \varepsilon.$$

Finally, as $v_\varepsilon \in V_h$, a new application of Cea's lemma 2.14 yields,

$$\|u - u_h\|_{H^1(\Omega)} \leq c \inf_{v_h \in V_h} \|u - v_h\|_{H^1(\Omega)} \leq c\|u - v_\varepsilon\|_{H^1(\Omega)} < c\varepsilon,$$

this implies that,

$$\|u - u_h\|_{H^1(\Omega)} \xrightarrow[h\to 0]{} 0.$$

■

To end this chapter, we will want to put together an error bound of the form,

$$\|u - u_h\|_{L^2(\Omega)} \leq ch^2,$$

for this so, we will use the following result.

Lemma 3.13 (Aubin-Nitsche Lemma). *Under the same conditions than Theorem 3.12 we have that,*

$$\|u-u_h\|_{L^2(\Omega)} \leq M\|u-u_h\|_{H^1(\Omega)} \sup_{g\in L^2(\Omega)} \left\{ \frac{1}{\|g\|_{L^2(\Omega)}} \inf_{v_h\in V_h} \|\varphi_g - v_h\|_{H^1(\Omega)} \right\}$$

where φ_g solves

$$a(v,\varphi_g) = (g,v)_{L^2(\Omega)} \qquad \forall v\in V \tag{12}$$

for $g \in L^2(\Omega)$.

Proof. As in the whole chapter, we assume that $a : V\times V \mapsto \mathbb{R}$ satisfies the Lax-Milgram hypothesis, this way, given any $g\in L^2(\Omega)$, φ_g not only solves (12) but also is unique. Furthermore, defining $e := u-u_h$, the Galerkin orthogonality (proposition 2.9) implies that

$$a(e,v_h) = 0 \qquad \forall v_h\in V_h.$$

Now, we consider the variational problem of finding $\varphi_e \in V$ such that

$$a(v,\varphi_e) = (e,v)_{L^2(\Omega)} \qquad \forall v\in V,$$

in this settings, $g = e \in V_h \subset L^2(\Omega)$ and so φ_ε exists and is unique as stated before. Notice that if we now fix the test function $v = e\in V$ in the previous formulation we obtain,

$$\|e\|^2_{L^2(\Omega)} = a(e,\varphi_\varepsilon) = a(e,\varphi_\varepsilon - v_h),$$

for all $v_h\in V_h$, as $a(\cdot,\cdot)$ is bilinear. Let also $c_a > 0$ be the continuity constant of $a(\cdot,\cdot)$, then it follows that

$$\|e\|^2_{L^2(\Omega)} \leq c_a\|e\|_{H^1(\Omega)}\|\varphi_\varepsilon - v_h\|_{H^1(\Omega)},$$

in particular,

$$\|e\|^2_{L^2(\Omega)} \leq c_a\|e\|_{H^1(\Omega)} \inf_{v_h\in V_h}\|\varphi_\varepsilon - v_h\|_{H^1(\Omega)}.$$

Observe now that if $e = 0$ the result trivially follows. This way, assume $e\neq 0$ and notice that,

$$\|e\|_{L^2(\Omega)} \leq c_a\frac{\|e\|_{H^1(\Omega)}}{\|e\|_{L^2(\Omega)}} \inf_{v_h\in V_h}\|\varphi_\varepsilon - v_h\|_{H^1(\Omega)}.$$

Hence, as

$$\frac{1}{\|e\|_{L^2(\Omega)}} \inf_{v_h\in V_h}\|\varphi_\varepsilon - v_h\|_{H^1(\Omega)} \leq \sup_{g\in L^2(\Omega)} \left\{ \frac{1}{\|g\|_{L^2(\Omega)}} \inf_{v_h\in V_h} \|\varphi_g - v_h\|_{H^1(\Omega)} \right\},$$

we obtain the desired result[6]. ■

Remark 3.13.1. Observe that in the case of symmetric elliptic variational problems, the element φ_g introduced in Lemma 3.13 is the Riesz representative of $g \in V'$ in the Hilbert space setting $(V, a(\cdot,\cdot))$.

Corollary 3.13.1. *Assume that the solution $\varphi_g \in V$ of (12) is such that $\varphi_g \in H^2(\Omega)$ and also that it is stable, i.e., there exists $c > 0$ such that*

$$\|\varphi_g\|_{H^2(\Omega)} \leq c\|g\|_{L^2(\Omega)} \qquad \forall g \in L^2(\Omega).$$

Then, we have that,

$$\|u - u_h\|_{L^2(\Omega)} \leq ch\|u - u_h\|_{H^1(\Omega)}.$$

Proof. As $\varphi_g \in H^2(\Omega)$, we get,

$$\inf_{v_h \in V_h} \|\varphi_g - v_h\|_{H^1(\Omega)} \leq c\|\varphi_g - \Pi_h \varphi_g\|_{H^1(\Omega)} \leq ch|\varphi_g|_{H^2(\Omega)},$$

by means of Cea's lemma 2.14 and Theorem 3.11 with $m = 1$. It then follows, using the stability of the solution φ_g and the Aubin-Nitsche Lemma 3.13, that

$$\begin{aligned}
\|u - u_h\|_{L^2(\Omega)} &\leq M\|u - u_h\|_{H^1(\Omega)} \sup_{g \in L^2(\Omega)} \left\{ \frac{1}{\|g\|_{L^2(\Omega)}} \inf_{v_h \in V_h} \|\varphi_g - v_h\|_{H^1(\Omega)} \right\} \\
&\leq M\|u - u_h\|_{H^1(\Omega)} \sup_{g \in L^2(\Omega)} \left\{ \frac{ch|\varphi_g|_{H^2(\Omega)}}{\|g\|_{L^2(\Omega)}} \right\} \\
&\leq M\|u - u_h\|_{H^1(\Omega)} \sup_{g \in L^2(\Omega)} \left\{ \frac{ch\|\varphi_g\|_{H^2(\Omega)}}{\|g\|_{L^2(\Omega)}} \right\} \\
&\leq ch\|u - u_h\|_{H^1(\Omega)},
\end{aligned}$$

here, we just relabeled c to absorb all other related constants, including M. ■

Now we can see that under the conditions of Corollary 3.13.1, we obtain,

$$\|u - u_h\|_{L^2(\Omega)} \leq ch^2|u|_{H^2(\Omega)},$$

provided $u \in H^2(\Omega)$.

[6]Observe that the $\sup_{g \in L^2(\Omega)}$ operation is well defined, as $g = 0$ forces $\varphi_g = 0$ by Lax-Milgram, and so $e = 0$ by construction.

References

[Bra01] D. Braess. Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics. Cambridge University Press, 2001.

[Cia13] Philippe G. Ciarlet. Linear and Nonlinear Functional Analysis with Applications. Society for Industrial and Applied Mathematics, USA, 2013.

[Eva98] Lawrence C. Evans. Partial Differential Equations, volume 19 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 1998.

[Joh94] C. Johnson. Numerical solution of partial differential equations by the finite element method. Studentlitteratur, 1994.

[Rel30] F. Rellich. Ein satz über mittlere konvergenz. Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, 1930:30–35, 1930.

[Sal09] Sandro Salsa. Partial Differential Equations in Action: From Modelling to Theory. Universitext. Springer, 1 edition, 2009.