Chapter 3

Numerical Analysis of the FEM for Elliptic Problems

The Finite Element Method is a method because it defines a systematic way to approximate solutions to boundary value problems to any desired degree of accuracy. Up to now we have largely focused on defining how finite element solutions are constructed, but we have not established that such constructions in fact provide a systematic way to approximate the exact solution of the boundary value problem. This is what we do in this chapter.

In building the tools to present the analysis, we will also answer questions that have likely lingered in the reader's mind. First, we will clearly state what we mean by a function being *smooth* in specifying the trial and test spaces. To this end, we will (lightly) introduce the concept of a Hilbert space, which can be regarded as the generalization of finite dimensional Euclidean spaces ($\mathbb{R}^n$) to infinite-dimensional sets of functions. Second, we will show why we can expect the stiffness matrix that we obtain from the Finite Element Method to be invertible. Third, a question that may or may not have materialized in the reader's mind: If a finite element solution approximates the values of a function, does its derivative also approximate the derivative of the function? How about higher-order derivatives, like the second derivative?

3.1 Basic Ideas

Let's illustrate some basic ideas we use in approaching the question of convergence of the finite element method. To this end, consider a scalar-valued function $F(v)$ for $v \in \mathcal{V} = \mathbb{R}$, as in Fig. 3.1. We seek u such that

$$F(u) = 0. \tag{3.1}$$

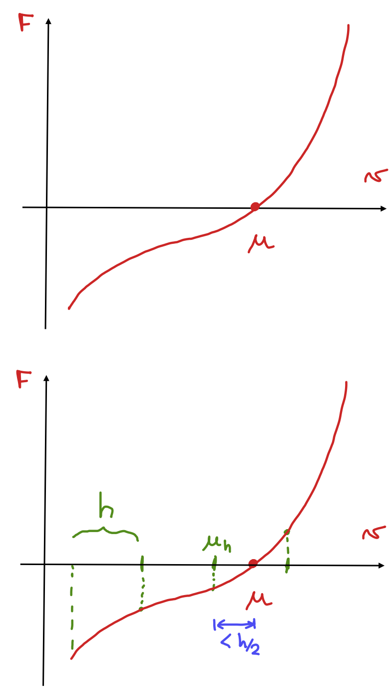

Figure 3.1

This is a generally nonlinear root-finding problem among an infinite set of potential values of u: any real number. To draw parallels between the way finite approximations to the solution approximate it, we may propose the following (not optimal) way of approximating the solution of this problem. Consider the set $\mathcal{V}_h = \{ih \mid i \in \mathbb{Z}\} = \{\dots, -2h, -h, 0, h, 2h, \dots\}$ for $h > 0$, and the approximate problem: Find $u_h \in \mathcal{V}_h$ such that

$$|F(u_h)| = \inf_{v \in \mathcal{V}_h} |F(v)|. \tag{3.2}$$

In other words, we seek u_h for which $F(u_h)$ is closest to zero. We examine next some conditions on F to guarantee that $u_h \to u$ as $h \to 0$.

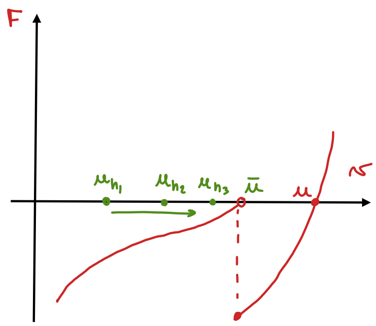

Figure 3.2

Approximability: An essential condition to approximate solutions of the problem is that as we consider sets $\mathcal{V}_h$ with larger number of elements, any potential solution should have an element of such sets at progressively smaller distances. More precisely, for any potential solution u, we should be able to identify $v_h \in \mathcal{V}_h$ for each h such that $|v_h - u| \to 0$ as $h \searrow 0$. This is true in this case. In fact, for any $u \in \mathbb{R}$, there is at least one element of $u_h \in \mathcal{V}_h$ that is at a distance smaller than $h/2$ of u, $|u_h - u| \le h/2$, for any h; see Fig. 3.1. Hence our sequence of spaces $\mathcal{V}_h$ is able to approximate any u to any desired accuracy.

Instead, the method could fail to approximate the exact solution if we defined $\tilde{\mathcal{V}}_h = \{u_h \in \mathcal{V}_h \mid |u_h| > 1\}$ and sought infimizers of $|F(v)|$ therein. For example, if $u = 0$ is the solution, we would not be able to approximate it with the sequence of spaces $\{\tilde{\mathcal{V}}_h\}_h$; see Fig. 3.3.

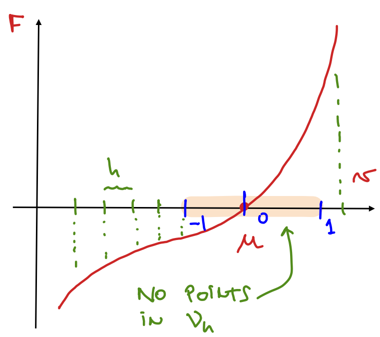

Figure 3.3

Continuity: If the function F is continuous, then we can say that if $u_h \to \bar{u}$, then $F(u_h) \to F(\bar{u})$ as $h \searrow 0$. Hence, if $F(u_h) \to 0$ as $h \searrow 0$, then $F(\bar{u}) = 0$ and $\bar{u}$ is a solution of the problem. In contrast, lack of continuity of F may lead to $F(\bar{u}) \neq 0$, or the sequence $\{u_h\}_h$ to converge to a value that is *not* the solution of the problem, see Fig. 3.2.

Approximability guarantees that we will be able to approximate any potential solution, and continuity guarantees that if the sequence of values we find converges, then the limit will be a solution. What they do not guarantee, however, is that the exact problem (3.1) or the discrete problem (3.2) has a solution, or if they do, that this solution is unique. For example, the solutions of the discrete problem could "jump around" different solutions of the exact problem, and never show convergence,[1] see Fig. 3.4. Sufficient solutions for the solution to exists and be unique follow from the following two properties (see [5]).

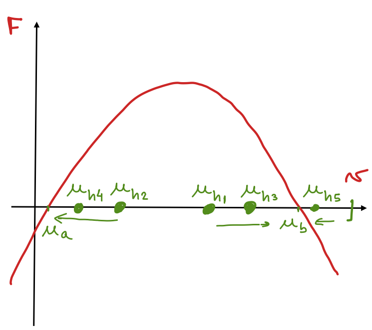

Figure 3.4

Coercivity: Vector-valued function $f: \mathbb{R}^n \to \mathbb{R}^n$, $n \in \mathbb{N}$, is **coercive** if

$$\lim_{\|x\| \to \infty} \frac{f(x) \cdot x}{\|x\|} = +\infty. \tag{3.3}$$

In particular, the scalar-valued function F in our problem ($n = 1$) is coercive if

$$\lim_{x \to +\infty} F(x) = +\infty \qquad \text{and} \qquad \lim_{x \to -\infty} F(x) = -\infty. \tag{3.4}$$

[1] A subsequence may converge.

Strict Monotonicity: Vector-valued function $f : \mathbb{R}^n \to \mathbb{R}^n$, $n \in \mathbb{N}$, is **strictly monotone** if

$$(f(x) - f(y)) \cdot (x - y) > 0 \text{ for all } x, y \in \mathbb{R}^n, x \neq y. \tag{3.5}$$

In particular, for the scalar-valued function F in our problem this implies that

$$F(x) > F(y) \Longleftrightarrow x > y.$$

Strict monotonicity implies that F^{-1} exists, and is continuous if F is.

Clearly, if F is continuous and coercive, we can conclude that its range is the entire real line. In other words, $F(u)$ takes all real values as u varies over the real line. Strict monotonicity then implies that it takes each real value only once, see Fig. 3.5. Therefore, the equation $F(u) = a$ for any $a \in \mathbb{R}$ (c.f. (3.1)) has exactly one solution. This is a particular case of a more general result that says that given a continuous, strictly monotone and coercive vector field $f : \mathbb{R}^n \to \mathbb{R}^n$, $n \in \mathbb{N}$, then for every $a \in \mathbb{R}^n$ there exists a unique $u \in \mathbb{R}^n$ such that $f(u) = a$ [5, Corollary 10.42].

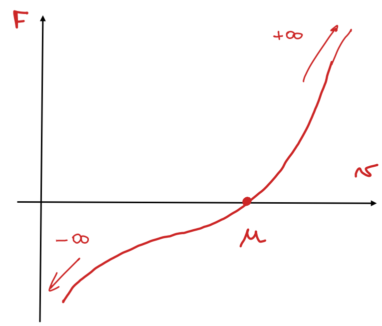

Figure 3.5

Basic convergence result: Assume that F is continuous, coercive and strictly monotone, and that for any $u \in \mathbb{R}$, there exists $v_h \in \mathcal{V}_h$ such that $|u - v_h| < h/2$ for each h. Then, there exists a unique u such that

$$F(u) = 0,$$

for each h there exists $u_h \in \mathcal{V}_h$ such that

$$|F(u_h)| = \inf_{v \in \mathcal{V}_h} |F(v)|,$$

and $u_h \to u$.

Therefore, under these conditions, the sequence of discrete solutions exists and is guaranteed to converge to the sole solution of the exact problem.

Basic convergence result

The existence and uniqueness of u follows from the earlier discussion. The existence of the discrete solution u_h is seen as follows. Because F is invertible, for any $M > 0$ we can define $v_l = F^{-1}(-M)$ and $v_r = F^{-1}(M)$, and let $I = [v_l, v_r]$ (see Fig. 3.6). From the strict monotonicity of F, $|F(v)| \leq M$ for $v \in I$, and $|F(v)| > M$ for $v \notin I$. Therefore, if the set $\mathcal{V}_h \cap I$ is not empty, we can write

$$\inf_{v \in \mathcal{V}_h} |F(v)| = \min_{v \in \mathcal{V}_h \cap I} |F(v)|.$$

The set $\mathcal{V}_h \cap I$ contains a finite number of elements, which allowed us to change the infimum for a minimum. So, if we show that $\mathcal{V}_h \cap I$ is not empty, then one of the elements of the set will be the minimizer u_h we seek, and we would have proved its existence. To this end, notice that because F is coercive, we can select M large enough so that $v_r - v_l > h$ (as we let $M \to \infty$, $v_r \to \infty$ and $v_l \to -\infty$). Then, the set I contains at least one element $v_h \in \mathcal{V}_h$, or $\mathcal{V}_h \cap I \neq \emptyset$. This follows because the approximation property of $\mathcal{V}_h$ implies that there should be an element of $\mathcal{V}_h$ that is at a distance at most $h/2$ from $(v_l + v_r)/2$. This proves that the discrete solution exists.

To see that $u_h \to u$, we select a sequence that approximates the exact solution $v_h \to u$ as $h \searrow 0$ with $v_h \in \mathcal{V}_h$. Such a sequence exists because of the approximation properties of $\mathcal{V}_h$. Since $v_h \in \mathcal{V}_h$ and u_h is the solution of the discrete problem (3.2) (they may be different), we can write that

$$0 \le |F(u_h)| \le |F(v_h)|.$$

This is illustrated in Fig. 3.7. Computing the limit as $h \searrow 0$, and using the continuity of F, we have that

$$0 \le \lim_{h\searrow 0} |F(u_h)| \le \lim_{h\searrow 0} |F(v_h)| = |F(u)| = 0,$$

and hence that $\lim_{h\searrow 0} F(u_h) = 0$. But F^{-1} exists and is continuous, so $u_h = F^{-1}(F(u_h)) \to F^{-1}(0) = u$, or $u_h \to u$.

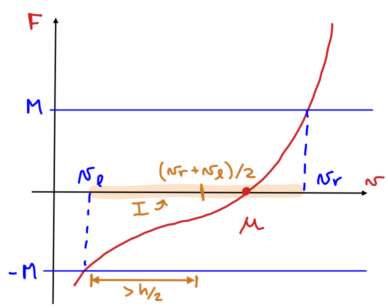

Figure 3.6

As a final remark, in the important case in which $F = ax + b$, then F is coercive if and only if $a > 0$, and as a results it is strictly monotone as well. In the following we will see how these basic ideas play a role in understanding the convergence of solutions obtained with the Finite Element method.

3.2 Abstract Error Estimate for Galerkin Method

The beauty of having formulated and abstract weak form and a Galerkin Method to approximate its solutions is that we can now obtain an error estimate that is applicable to any problem formulated in this way. Specifically, the abstract weak form, Problem 1.4, reads

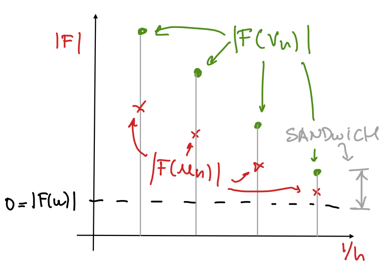

Figure 3.7

Problem 3.1 (Abstract Weak Form)**.** *Let $\mathcal{W}$ be a vector space, $a\colon \mathcal{W} \times \mathcal{W} \to \mathbb{R}$ be a bilinear form, and $\ell\colon \mathcal{W} \to \mathbb{R}$ be a linear functional. Let the trial space $\mathcal{S}$ be an affine subspace of $\mathcal{W}$, and let the test space $\mathcal{V} \subset \mathcal{W}$ be the direction of $\mathcal{S}$.*

$$\text{Find } u \in \mathcal{S} \text{ such that } a(u,v) = \ell(v) \text{ for all } v \in \mathcal{V}. \tag{3.6}$$

and the Galerkin Method for it, Problem 1.5, reads

Problem 3.2 (Abstract Galerkin Method for Problem 3.1)**.** *Let $\mathcal{W}_h \subset \mathcal{W}$ be a vector space, and let $\mathcal{S}_h = \mathcal{S} \cap \mathcal{W}_h$ and $\mathcal{V}_h = \mathcal{V} \cap \mathcal{W}_h$.*

$$\text{Find } u_h \in \mathcal{S}_h \text{ such that } a(u_h, v_h) = \ell(v_h) \text{ for all } v_h \in \mathcal{V}_h. \tag{3.7}$$

The question we ask in this part is what we can say about the error $e_h = u - u_h$. In particular, we want to find a function $f_u(h)$ that may depend on u such that

$$\|u - u_h\| \le f_u(h).$$

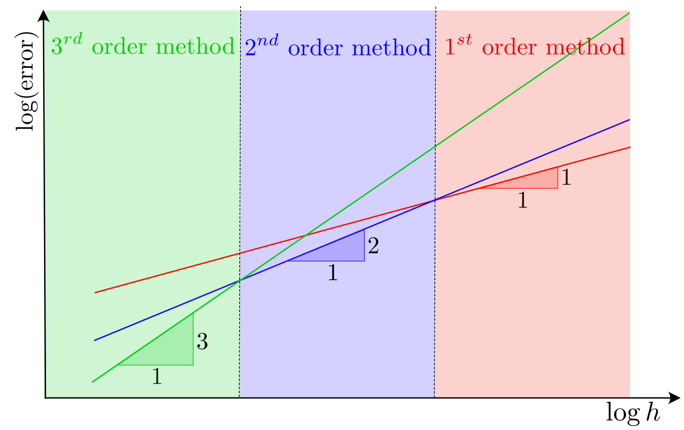

Figure 3.8 Methods with a higher-order of convergence are always more accurate than lower-order ones for a small enough value of h. In this figure we indicate the slope of each line, and hence the order of the method, with the triangles drawn under each line (the two axes have different scales). While a first order method may be convenient at large values of h (red region), higher-order methods will eventually have a smaller error as h decreases.

This is what we call an **error estimate**. Although here h can be any parameter identifying the space $\mathscr{W}_h$, we will study convergence of Galerkin Method as $h \searrow 0$, following the interpretation of h as the mesh size found in the context of Finite Element methods. Thus, if $f_u(h) \to 0$ as $h \searrow 0$, we conclude that solutions to Galerkin Method converge, namely, $u \to u_h$ as $h \searrow 0$. Typically,

$$f_u(h) = C_u h^r,$$

where $C_u > 0$ is a value that depends only on u or its derivatives, h is the diameter of the largest element in the mesh, and $r > 0$ is called the **order of convergence** of the method. The order of convergence of the method defines how fast the method converges to the exact solution. A method with a higher-order of convergence will always be more accurate for a fine enough mesh, as Fig. 3.8 illustrates. Since $\log(f_u) = \log C + r \log h$, the slope of each line in the figure is the order of convergence of the method.

To draw a parallel to the definitions introduced in §3.1 in the case of Galerkin Method, it is convenient to identify F with a vector-valued function as follows. First, for $w_h \in \mathscr{S}_h$, we write $w_h = \overline{u}_h + \hat{w}_h$, where $\hat{w}_h \in \mathscr{V}_h$. We denote by $\hat{W}$ the vector of components of $\hat{w}_h$ in $\mathscr{V}_h$. Then, the vector-valued function F is defined as

$$\begin{aligned} F_b(\hat{U}) &= a(\overline{u}_h + \hat{u}_h, N_b) - \ell(N_b) \qquad \text{for } b \in \eta_a, \\ &= a(\hat{u}_h, N_b) + a(\overline{u}_h, N_b) - \ell(N_b). \end{aligned} \tag{3.8}$$

and the equations for Galerkin Method follow as

$$F_b(\hat{U}) = 0 \qquad \text{for } b \in \eta_a. \tag{3.9}$$

A similar identification can be made for the abstract variational form, but because $\mathcal{V}$ is an infinite-dimensional space, additional concepts are required which are better left to a more advanced stage.

To perform the analysis outlined here, we will introduce the definitions of a normed space, continuity and coercivity in the setting we are interested in, which will enable us to establish the convergence results later on.

3.2.1 Normed Spaces

To be able to assess convergence, or how close our approximations are to the exact solution, we need to define a way to measure distances in a vector space. The most common way to do this is through a *norm*.

Definition 3.1 (Norm)**.** *Let V be a vector space. A norm is a function $\|\cdot\|: V \to \mathbb{R}$ such that for $v, u \in V$ and $\alpha \in \mathbb{R}$:*

1. **N.1.** $\|v\| \geq 0$*, and* $\|v\| = 0$ *if and only* $v = 0$.
2. **N.2.** $\|\alpha v\| = |\alpha| \|v\|$.
3. **N.3.** $\|v + u\| \leq \|v\| + \|u\|$ *(triangle inequality).*

The typical norm that you are familiar with is the "Euclidean norm" in $\mathbb{R}^3$. For example, if $x = (x_1, x_2, x_3)$, then $\|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2}$. Clearly if $\|x\| \geq 0$, and if $\|x\| = 0$, then $x = 0$. The second condition, N.2, is also simple to verify, and the triangle inequality is the common statement that the sum of the lengths of two sides of a triangle is always greater or equal than the length of the third. These three conditions are intuitive to understand in the case of $\mathbb{R}^n$, and the fact that the Euclidean norm satisfies them is easy to see. Defining a norm for vector spaces of functions is more delicate, and less intuitive. Let's look at some examples.

Examples:

3.1 For $x = (x_1, x_2, x_3) \in \mathbb{R}^3$, the function $g(x) = \sqrt{2x_1^2 + 3x_2^2 + 4x_3^2}$ is also a norm in $\mathbb{R}^3$. We will not prove this.

The importance of this example is to illustrate that we can endow the same set of vectors with different norms. A simple way to think about this example is that we are using different units to measure distances in each coordinate direction.

3.2 For $v \in V_1 = \{f: [a,b] \to \mathbb{R} \text{ smooth}\}$, we define the L^∞*-norm*

$$\|v\|_{0,\infty} = \max_{x \in [a,b]} |v(x)|. \tag{3.10}$$

Let's check the conditions for this to be a norm, since it is simple in this case. For N.1, since $|v(x)| \geq 0$ for all $x \in [a,b]$, then $\|v\|_{0,\infty} \geq 0$. Also, if

$0 = \|v\|_{0,\infty} = \max_{x\in[a,b]} |v(x)| \geq |v(x)| \geq 0$ for any $x \in [a,b]$, then $v = 0$.
For N.2,

$$\|\alpha v\|_{0,\infty} = \max_{x\in[a,b]} |\alpha v(x)| = \max_{x\in[a,b]} |\alpha||v(x)| = |\alpha| \max_{x\in[a,b]} |v(x)| = |\alpha|\|v\|_{0,\infty}.$$

Finally, for N.3,

$$\begin{aligned}\|u+v\|_{0,\infty} &= \max_{x\in[a,b]} |u(x)+v(x)| \leq \max_{x\in[a,b]} |u(x)|+|v(x)| \\ &\leq \max_{x\in[a,b]} |u(x)| + \max_{x\in[a,b]} |v(x)| = \|u\|_{0,\infty} + \|v\|_{0,\infty}.\end{aligned}$$

For instance, let $[a,b] = [0,\pi]$, then:

i. If $v(x) = \cos(x)$, then $\|v\|_{0,\infty} = 1$.
ii. If $v(x) = x(x-\pi)$, then $\|v\|_{0,\infty} = -v(\pi/2) = \pi^2/4$.

3.3 For $v \in V_1$ from Example 3.2, we define the L^2-*norm*

$$\|v\|_{0,2} = \left[\int_a^b v(x)^2\, dx\right]^{1/2}. \tag{3.11}$$

We will not check that this is a norm, but just state it. For $[a,b] = [0,\pi]$:

i. If $v(x) = \cos(x)$, then $\|v\|_{0,2} = \left[\int_0^\pi \cos(x)^2\, dx\right]^{1/2} = \sqrt{\pi/2}$.
ii. If $v(x) = x(x-\pi)$, then $\|v\|_{0,2} = \left[\int_0^\pi x^2(x-\pi)^2\, dx\right]^{1/2} = \pi^{5/2}/\sqrt{30}$.

3.4 For $v \in V_2 = \{w \in V_1 \mid w(a) = w(b) = 0\}$, we define the H^1-*seminorm*

$$|v|_{1,2} = \left[\int_a^b v'(x)^2\, dx\right]^{1/2}. \tag{3.12}$$

The fact that this is a norm requires a discussion of condition N.1: Why does it hold? To answer this, notice that if $|v|_{1,2} = 0$, we can conclude that $v'(x) = 0$ for all $x \in [0,1]$, since the integrand $v'(x)^2 \geq 0$ everywhere. Therefore, $v(x)$ is a constant function. Since $v(a) = 0$, then $v(x) = 0$ for $x \in [a,b]$.
Because of this discussion, $|\cdot|_{1,2}$ is not a norm in the space V_1 in Example 3.2, since functions therein need not be zero at the boundaries, and hence condition N.1 is not satisfied. All we would be able to say if $|v|_{1,2} = 0$ is that v is a constant function. For example, let $[a,b] = [0,1]$,

i. If $v(x) = \sin(\pi x)$, $v \in V_2$, then $|v|_{1,2} = \left[\int_0^1 (\pi\cos(\pi x))^2\, dx\right]^{1/2} = \frac{\pi}{\sqrt{2}}$.
ii. If $v(x) = 3$, $v \notin V_2$, then $|v|_{1,2} = \left[\int_0^1 0\, dx\right]^{1/2} = 0$.

3.5 For $v \in V_1$ from Example 3.2, we define the H^1-*norm*

$$\begin{aligned}\|v\|_{1,2} &= \left[\int_a^b v(x)^2\, dx + \int_a^b v'(x)^2\, dx\right]^{1/2} \\ &= \left[\|v\|_{0,2}^2 + |v|_{1,2}^2\right]^{1/2}.\end{aligned} \tag{3.13}$$

In contrast to what happens with $|v|_{1,2}$ in Example 3.4, condition N.1 is satisfied in this case, since it is satisfied for $\|v\|_{0,2}$.

Notice that we talked about three different norms for space V_1 above: We defined the L^∞-norm, the L^2-norm and the H^1-norm. The three norms measure distance differently, emphasizing different aspects of the functions.

We can now define the notion of a normed space.

Definition 3.2 (Normed Space.)**.** *A vector space V with a norm defined over it $\|\cdot\|: V \to \mathbb{R}$ is called a* **normed space***, and denoted by* $(V, \|\cdot\|)$.

Examples:

3.6 The space $\mathbb{R}^n$, $n \in \mathbb{N}$, with the Euclidean norm $\|\cdot\|$ is a normed space $(\mathbb{R}^n, \|\cdot\|)$, since the norm is defined for every element of $\mathbb{R}^n$.

3.7 Consider the space $V_1 = \{f: [a,b] \to \mathbb{R} \text{ smooth}\}$ with the L^∞-norm $\|\cdot\|_{0,\infty}$. Since all smooth functions are bounded in $[a,b]$, the L^∞-norm is well defined for every function in V_1, and hence $(V_1, \|\cdot\|_{0,\infty})$ is a normed space.

3.8 Consider again the space $V_1 = \{f: [a,b] \to \mathbb{R} \text{ smooth}\}$ with the L^2-norm $\|\cdot\|_{0,2}$. Since all smooth functions are bounded in $[a,b]$, the integrals needed to compute the L^2-norm are well defined for every function in V_1, and hence $(V_1, \|\cdot\|_{0,2})$ is a normed space. It is, however, a different normed space than $(V_1, \|\cdot\|_{0,\infty})$, since functions that are close in one may not be close in the other, as we shall see.

3.9 Consider the space $V_2 = \{f: (a,b) \to \mathbb{R} \text{ smooth}\}$ (notice the open interval) with the L^2-norm $\|\cdot\|_{0,2}$. The function $f(x) = 1/(x-a)$ is in V_2, since it is smooth in (a,b), but

$$\|f\|_{0,2}^2 = \int_a^b \frac{1}{(x-a)^2}\, dx = +\infty,$$

so the norm is not defined for f. Therefore, $(V_2, \|\cdot\|_{0,2})$ is *not* a normed space.

3.10 Let $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$. For such domain Ω, the norm $\|v\|_{0,2}$ of $v: \Omega \to \mathbb{R}$ is defined as

$$\|v\|_{0,2} = \left[\int_\Omega v(x)^2\, d\Omega\right]^{1/2}. \tag{3.14}$$

The set

$$L^2(\Omega) = \{v: \Omega \to \mathbb{R} \mid \|v\|_{0,2} < \infty\} \tag{3.15}$$

is called the $L^2(\Omega)$ **space**, and $(L^2(\Omega), \|\cdot\|_{0,2})$ is a normed space. The space $L^2(\Omega)$ is said to contain all *square-integrable* functions, and these need not be smooth. For example, if $\Omega = [-1,1]$, it contains the Heaviside function

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

In contrast, $H(x) \notin L^2(\mathbb{R})$, since $\|H\|_{0,2} = \infty$ in this case.

3.11 Let $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$. For such domain Ω, we define the H^1-*norm* as

$$\|v\|_{1,2} = \left[\|v\|_{0,2}^2 + \sum_{i=1}^{n} \left\| \frac{\partial v}{\partial x_i} \right\|_{0,2}^2 \right]^{1/2}.$$

With it, we can define the $H^1(\Omega)$**-space** as

$$H^1(\Omega) = \left\{ v\colon \Omega \to \mathbb{R} \mid \|v\|_{1,2} < \infty \right\}, \tag{3.16}$$

and $(H^1(\Omega), \|\cdot\|_{1,2})$ is normed space. Functions in $H^1(\Omega)$ contain all functions in which both the function and each one of its partial derivatives is square integrable. Alternatively, the function and each one of its partial derivatives is in $L^2(\Omega)$. Therefore, if a funtion $v \in H^1(\Omega)$, then $v \in L^2(\Omega)$. For example: Let $\Omega = [-1,1] \times [-1,1]$, then

i. The function $v(x_1, x_2) = x_1^2 + x_2^3 \in H^1(\Omega)$, since

$$\begin{aligned} \|v\|_{1,2}^2 &= \int_{-1}^{1}\int_{-1}^{1} (x_1^2 + x_2^3)^2 \, dx_1 dx_2 + \int_{-1}^{1}\int_{-1}^{1} (2x_1)^2 \, dx_1 dx_2 \\ &\quad + \int_{-1}^{1}\int_{-1}^{1} (3x_2^2)^2 \, dx_1 dx_2 = \frac{292}{21} < \infty. \end{aligned}$$

ii. The function $v(x_1, x_2) = \ln(1 + x_1) + \ln(1 + x_2) \notin H^1(\Omega)$, but $v \in L^2(\Omega)$, since

$$\begin{aligned} \|v\|_{0,2}^2 &= \int_{-1}^{1}\int_{-1}^{1} (\ln(1+x_1) + \ln(1+x_2))^2 \, dx_1 dx_2 \\ &= 24 + 8\ln(4)(\ln(2) - 2) < \infty. \\ \|v\|_{1,2}^2 &= \|v\|_{0,2}^2 \\ &\quad + \int_{-1}^{1} \frac{1}{(1+x_1)^2} \, dx_1 dx_2 + \int_{-1}^{1} \frac{1}{(1+x_2)^2} \, dx_1 dx_2 = \infty. \end{aligned}$$

☞ The definition of H^1 here is incomplete. We will have an opportunity to complete the definition later.

3.12 Let $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$. For $v \in H^1(\Omega)$ we can define the H^1-*seminorm* as

$$|v|_{1,2} = \left[\int_{\Omega} \|\nabla v\|^2 \, d\Omega \right]^{1/2}. \tag{3.17}$$

Notice that

$$\begin{aligned} |v|_{1,2}^2 &= \int_{\Omega} \|\nabla v\|^2 \, d\Omega = \int_{\Omega} \sum_{i=1}^{n} \left| \frac{\partial v}{\partial x_i} \right|^2 \, d\Omega = \sum_{i=1}^{n} \int_{\Omega} \left| \frac{\partial v}{\partial x_i} \right|^2 \, d\Omega \\ &= \sum_{i=1}^{n} \left\| \frac{\partial v}{\partial x_i} \right\|_{0,2}^2. \end{aligned}$$

This allows us to write the H^1-norm as

$$\|v\|_{1,2}^2 = \|v\|_{0,2}^2 + |v|_{1,2}^2. \tag{3.18}$$

3.13 Let $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$ and $m \in \mathbb{N}_0 = \{0\} \cup \mathbb{N}$. The H^m-seminorm of a function $u\colon \Omega \to \mathbb{R}$ is defined as

$$|u|_{m,2}^2 = \sum_{\substack{i_1,\ldots,i_n \geq 0 \\ i_1+\ldots+i_n=m}} \left\| \frac{\partial^m u}{\partial x_1^{i_1} \ldots \partial x_n^{i_n}} \right\|_{0,2}^2, \tag{3.19}$$

where $\partial^m u / \partial x_k^0 = u$.
For example, the H^2-seminorm in $\mathbb{R}^2$ is

$$|u|_{2,2}^2 = \left\| \frac{\partial^2 u}{\partial x_1^2} \right\|_{0,2}^2 + 2 \left\| \frac{\partial^2 u}{\partial x_1 \partial x_2} \right\|_{0,2}^2 + \left\| \frac{\partial^2 u}{\partial x_2^2} \right\|_{0,2}^2,$$

and the H^0-seminorm is directly the L^2-norm.
The H^m-norm is then defined as

$$\|u\|_{m,2}^2 = \sum_{i=0}^{m} |u|_{i,2}^2 \tag{3.20}$$

With it, we can define the $H^m(\Omega)$**-space** as

$$H^m(\Omega) = \left\{ v\colon \Omega \to \mathbb{R} \mid \|u\|_{m,2} < \infty \right\}. \tag{3.21}$$

3.2.2 Continuity

We begin by introducing the definition of continuity of a bilinear form and a linear functional, using the $\varepsilon - \delta$ description of continuity from any introductory calculus course.

Definition 3.3 (Continuous Linear Functional)**.** *Let $(V, \|\cdot\|)$ be a normed space. A linear functional $\ell\colon V \to \mathbb{R}$ is continuous at $v \in V$ if for any $\varepsilon > 0$ there exists $\delta > 0$ such that, for any $w \in V$, if $\|w - v\| < \delta$ then $|\ell(v) - \ell(w)| < \varepsilon$.*

Definition 3.4 (Continuous Bilinear Form)**.** *Let $(V, \|\cdot\|)$ be a normed space. A bilinear form $a\colon V \times V \to \mathbb{R}$ is continuous at $(v, w) \in V \times V$ if for any $\varepsilon > 0$ there exists $\delta > 0$ such that, for any $(\bar{v}, \bar{w}) \in V \times V$, if $\|v - \bar{v}\| < \delta$ and $\|w - \bar{w}\| < \delta$, then $|a(v, w) - a(\bar{v}, \bar{w}))| < \varepsilon$.*

A useful characterization of continuity is provided by the following lemma.

Lemma 3.1 (Continuity)**.** *Let $(V, \|\cdot\|)$ be a normed space. Then:*

1. *A bilinear form $a\colon V \times V \to \mathbb{R}$ is continuous if and only if $\exists M > 0$ such that for all $u, v \in V$*

$$|a(u, v)| \leq M \|u\| \|v\|. \tag{3.22}$$

2. *A linear functional $\ell : V \to \mathbb{R}$ is continuous if and only if* $\exists M > 0$ *such that for all* $v \in V$
$$|\ell(v)| \leq M\|v\|. \tag{3.23}$$

A linear functional that satisfies (3.23) or a bilinear form that satisfies (3.22) is said to be **bounded**.

Proof. We first prove that (3.23)$\Longrightarrow$ ℓ is continuous. To show that ℓ is continuous at $v \in V$, for any $\varepsilon > 0$ we need to find $\delta > 0$ such that $|\ell(v)-\ell(w)| < \varepsilon$ if $\|v-w\| < \delta$ for $w \in V$. Let $\delta = \varepsilon/M$, so from (3.23) we conclude that

$$|\ell(v) - \ell(w)| = |\ell(v-w)| \leq M\|v-w\| < M\delta = \varepsilon.$$

For the converse, ℓ is continuous$\Longrightarrow$(3.23), choose $\varepsilon > 0$. There exists $\delta > 0$ such that if $\|w\| < \delta$ then $|\ell(w)| < \varepsilon$ for $w \in V$ (continuity at 0). Now, given $v \in V$, define $w = \delta v/(\|v\|2)$, so that

$$\|w\| = \frac{\delta}{2} < \delta \Longrightarrow \varepsilon > |\ell(w)| = \frac{\delta}{2\|v\|}|\ell(v)|,$$

or
$$|\ell(v)| < \frac{2\varepsilon}{\delta}\|v\|.$$
Hence, (3.23) follows with $M = 2\varepsilon/\delta$.

The proof for the bilinear form is similar, but more cumbersome because of the two arguments. Let's see first that (3.22)$\Longrightarrow$ a is continuous. To show that a is continuous at $(v,w) \in V \times V$, for any $\varepsilon > 0$ we need to find $\delta > 0$ such that if $\|v-\bar{v}\| < \delta$ and $\|w-\bar{w}\| < \delta$ then $|a(v,w) - a(\bar{v},\bar{w})| < \varepsilon$ for $(\bar{v},\bar{w}) \in V \times V$. In this case, let $\delta_0 = \|v\| + \|w\|$ and set $\delta = \delta_0 \min\{1, \varepsilon/(2M\delta_0)\}$, so from (3.22) we conclude that

$$\begin{aligned}
&|a(v,w) - a(\bar{v},\bar{w})| \\
&\qquad = |a(v,w) - a(v,\bar{w}) + a(v,\bar{w}) - a(\bar{v},\bar{w}) - a(v-\bar{v},w) + a(v-\bar{v},w)| \\
&\qquad \leq |a(v,w) - a(v,\bar{w})| + |a(v,\bar{w}) - a(\bar{v},\bar{w}) - a(v-\bar{v},w)| + |a(v-\bar{v},w)| \\
&\qquad \leq |a(v,w-\bar{w})| + |a(v-\bar{v},\bar{w}-w)| + |a(v-\bar{v},w)| \\
&\qquad \leq M\|v\|\|w-\bar{w}\| + M\|v-\bar{v}\|\|w-\bar{w}\| + M\|v-\bar{v}\|\|w\| \\
&\qquad \leq M(\|v\| + \|w\|)\delta + M\delta^2 \\
&\qquad = M\delta_0^2\left(\frac{\delta}{\delta_0} + \left(\frac{\delta}{\delta_0}\right)^2\right)
\end{aligned}$$

From our choice δ, we know that $0 < \delta/\delta_0 \leq 1$. For $x \in [0,1]$, $x(x-1) \leq 0$, and hence $x^2 + x < 2x$ therein. Using $x = \delta/\delta_0$, we conclude that

$$|a(v,w) - a(\bar{v},\bar{w})| \leq M\delta_0^2 2\frac{\delta}{\delta_0} = 2M\delta_0\frac{\varepsilon}{2M\delta_0} = \varepsilon.$$

To prove the converse, a is continuous $\Longrightarrow$ (3.22), choose $\varepsilon > 0$. There exists $\delta > 0$ such that if $\|\bar{v}\| < \delta$ and $\|\bar{w}\| < \delta$ then $|a(\bar{v}, \bar{w})| < \varepsilon$ for $\bar{v}, \bar{w} \in V$. Given $v, w \in V$, define $\bar{v} = \delta v/(\|v\|2)$ and $\bar{w} = \delta w/(\|w\|2)$, so that

$$\|\bar{v}\| = \|\bar{w}\| = \frac{\delta}{2} \Longrightarrow \varepsilon > |a(\bar{v}, \bar{w})| = \frac{\delta^2}{4\|v\|\|w\|}|a(v, w)|,$$

and (3.22) follows with $M = 4\varepsilon/\delta^2$. □

For the examples that follow, we will need to apply Cauchy-Schwartz inequality, in two alternative forms.

Theorem 3.1 (Cauchy-Schwartz inequality for vectors in $\mathbb{R}^n$)**.** *Let $u, v \in \mathbb{R}^n$, $n \in \mathbb{N}$, then*

$$|u \cdot v| \leq \|u\|\|v\|, \tag{3.24}$$

where $u \cdot v$ is the scalar product of u and v in $\mathbb{R}^n$ and $\|u\|$ is the Euclidean norm.

In components of an orthonormal basis $\{e_1, \ldots, e_n\}$, $u = u_1 e_1 + \ldots + u_n e_n$ and $v = v_1 e_1 + \ldots + v_n e_n$, this inequality reads

$$\left|\sum_{i=1}^{n} u_i v_i\right| \leq \left(\sum_{i=1}^{n} u_i^2\right)^{1/2} \left(\sum_{i=1}^{n} v_i^2\right)^{1/2}. \tag{3.25}$$

Theorem 3.2 (Cauchy-Schwartz inequality for functions in $L^2(\Omega)$)**.** *Let $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{R}^n$. Let $u, v \in L^2(\Omega)$, then*

$$\left|\int_\Omega uv \, d\Omega\right| \leq \|u\|_{0,2}\|v\|_{0,2}. \tag{3.26}$$

Examples:

3.14 Consider first the weak form for the general second-order problem in one dimension, Problem 1.3, with $\max_{x\in[0,L]}\{|k(x)|, |b(x)|, |c(x)|\} = M/3 < \infty$ for all x. The bilinear form for this problem is

$$a(u, v) = \int_0^L k(x)u'(x)v'(x) + b(x)u'(x)v(x) + c(x)u(x)v(x)\,dx.$$

This bilinear form is well-defined and continuous for any $u, v \in H^1([0, L])$, since

$$
\begin{aligned}
|a(u,v)| &= \left|\int_0^L k(x)u'(x)v'(x) + b(x)u'(x)v(x) + c(x)u(x)v(x)\,dx\right| \\
&\le \int_0^L |k(x)u'(x)v'(x) + b(x)u'(x)v(x) + c(x)u(x)v(x)|\,dx && \text{Triangle inequality for integrals} \\
&\le \int_0^L |k(x)||u'(x)||v'(x)| + |b(x)||u'(x)||v(x)| \\
&\qquad + |c(x)||u(x)||v(x)|\,dx && \text{Triangle inequality} \\
&\le \frac{M}{3}\Big(\int_0^L |u'(x)||v'(x)|\,dx + \int_0^L |u'(x)||v(x)|\,dx \\
&\qquad + \int_0^L |u(x)||v(x)|\,dx\Big) \\
&\le \frac{M}{3}\big(|u|_{1,2}|v|_{1,2} + |u|_{1,2}\|v\|_{0,2} + \|u\|_{0,2}\|v\|_{0,2}\big) && \text{Cauchy-Schwartz for functions, (3.26)} \\
&\le M\|u\|_{1,2}\|v\|_{1,2}. && \|u\|_{0,2}, |u|_{1,2} \le \|u\|_{1,2}
\end{aligned}
$$

3.15 Consider the weak form of the 2D diffusion problem, Problem 2.2 with $H = 0$ and $K = k\mathbf{I}_{2\times 2}$ being a constant and isotropic matrix with $k > 0$. In this case, the bilinear form $a\colon \mathscr{W} \times \mathscr{W} \to \mathbb{R}$ and linear functional $\ell\colon \mathscr{W} \to \mathbb{R}$ are

$$
\begin{aligned}
a(u,v) &= \int_\Omega k\nabla u \cdot \nabla u \, d\Omega \\
\ell(v) &= \int_\Omega f v \, d\Omega.
\end{aligned}
$$

To obtain continuity of both the bilinear form and the linear functional, we need to select the normed space to which u and v belong. For example, if we set $\mathscr{W} = H^1(\Omega)$, and $u, v \in \mathscr{W}$, then continuity of a is readily obtained:

$$
\begin{aligned}
|a(u,v)| &= \left|\int_\Omega k\nabla u \cdot \nabla v \, d\Omega\right| \\
&= k\left|\int_\Omega \|\nabla u\|\|\nabla v\| \, d\Omega\right| && \text{Cauchy-Schwartz in } \mathbb{R}^2, (3.25) \\
&= k|u|_{1,2}|v|_{1,2} && \text{Cauchy-Schwartz in } L^2(\Omega), (3.26)
\end{aligned}
$$

To apply the Cauchy-Schwartz inequality in the last line, we used the fact that if $w \in H^1(\Omega)$, then the function $\|\nabla w\| \in L^2(\Omega)$, since from (3.18),

$$
\int_\Omega \|\nabla w\|^2 \, d\Omega = |w|_{1,2}^2 \le \|w\|_{1,2}^2 < \infty.
$$

So $a(\cdot,\cdot)$ is continuous in $\mathscr{W}$ with $M = k$ in Lemma 3.1.

The continuity of ℓ is simpler, but some information about f is needed. For example, we can request $f \in L^2(\Omega)$. In this case, for $v \in H^1(\Omega)$,

$$
\begin{aligned}
|\ell(v)| &= \left|\int_\Omega f v \, d\Omega\right| \\
&\leq \int_\Omega |f v| \, d\Omega && \text{Triangle inequality for integrals} \\
&\leq \|f\|_{0,2}\|v\|_{0,2} && \text{Cauchy-Schwartz in } L^2(\Omega),\ (3.26) \\
&\leq \|f\|_{0,2}\|v\|_{1,2}
\end{aligned}
$$

In the last line we used that $\|v\|_{0,2} \leq \|v\|_{0,2} + |v|_{1,2} = \|v\|_{0,2}$.

3.16 The choice of $\mathscr{W} = H^1(\Omega)$ in Example 3.15 is important. For example, $a(\cdot,\cdot)$ is not continuous in $\mathscr{W} = L^2(\Omega)$. To see this, we first note that continuity in this case means finding $M > 0$ such that for all $u, v \in L^2(\Omega)$,

$$|a(u,v)| \leq M\|u\|_{0,2}\|v\|_{0,2}. \tag{3.27}$$

If we find a function $\bar{u} \in L^2(\Omega)$, and hence for which $\|\bar{u}\|_{0,2} < \infty$, such that $|a(\bar{u},\bar{u})| = \infty$, we establish that there is no such $M > 0$, and hence that $a(\cdot,\cdot)$ is not continuous in $L^2(\Omega)$.

Consider then an example in which $\Omega = \{(x_1, x_2) \in \mathbb{R} \mid x_1^2 + x_2^2 < 1\}$, or the unit circle. Let $r = \sqrt{x_1^2 + x_2^2}$, and for $\alpha > -1$ define the function

$$v_\alpha(r) = \sqrt{\alpha+1}\, r^\alpha.$$

Function $v_\alpha \in L^2(\Omega)$, since

$$
\begin{aligned}
\|v_\alpha\|_{0,2}^2 = 2\pi(\alpha+1)\int_0^1 (r^\alpha)^2 r \, dr \\
= 2\pi(\alpha+1)\int_0^1 r^{2\alpha+1} \, dr = 2\pi(\alpha+1)\frac{1}{2\alpha+2} = \pi.
\end{aligned}
$$

Hence, v_α has the same L^2-norm for any α. However, the value of $a(v_\alpha, v_\alpha)$ is infinite for $\alpha < 0$. To see this, the components of ∇v_α in a polar basis are $\partial v_\alpha/\partial r$ in the radial direction and 0 in the azimuthal one. Hence, for $\alpha < 0$

$$
\begin{aligned}
a(v_\alpha, v_\alpha) = k2\pi(\alpha+1)\int_0^1 (\alpha r^{\alpha-1})^2 r \, dr = k2\pi(\alpha+1)\alpha^2 \int_0^1 r^{2\alpha-1} \, dr \\
= k\frac{2\pi(\alpha+1)\alpha^2}{2\alpha}\left(1 - \lim_{r\to 0^+} r^{2\alpha}\right) = \infty.
\end{aligned}
$$

This shows that $a(\cdot,\cdot)$ is *not* continuous in $L^2(\Omega)$.

What has happened in this example? How is it that $a(\cdot,\cdot)$ is continuous in $H^1(\Omega)$ but not in $L^2(\Omega)$? The intuitive answer to these questions is that the L^2-norm only measures the size of a function, and not of its derivative. As the functions v_α show, it is possible to have functions that have a bounded L^2-norm but an infinite H^1-norm. Since the gradients are involved in computing the value of $a(\cdot,\cdot)$, it is necessary to have a norm that measures the size of the gradient on the right hand side of the inequality. In terms of continuity, this means that we could have a sequence of function $\{u_i\}_{i\in\mathbb{N}} \subset L^2(\Omega)$ that converges to $u \in L^2(\Omega)$, so that $\lim_{i\to\infty} \|u_i - u\|_{0,2} = 0$, but for which $a(u_i, v) \not\to a(u, v)$ for some $v \in \mathscr{W}$. In contrast, because of the continuity, this cannot happen for a sequence $\{u_i\}_{i\in\mathbb{N}} \subset H^1(\Omega)$ that converges to $u \in H^1(\Omega)$. In this case, $a(u_i, v) \to a(u, v)$ for any $v \in \mathscr{W}$.

3.2.3 Coercivity

We introduce now the notion of coercivity of a bilinear form, which is inextricably linked to the uniqueness of solutions of the weak form of a problem and of the Galerkin method, and hence, to the invertibility of the stiffness matrix.

Definition 3.5 (Coercivity of a Bilinear Form)**.** *Let $(W, \|\cdot\|)$ be a normed space, and let $V \subseteq W$ be a vector space. A bilinear form $a\colon W \times W \to \mathbb{R}$ is coercive on V if there exists $c_V > 0$ such that for all $v \in V$*

$$c_V \|v\|_W^2 \leq a(v, v). \tag{3.28}$$

The constant c_V is called the **coercivity constant**. Notice the use of a "containing" space W in which the bilinear form is defined, and the satisfaction of the coercivity condition (3.28) only on a subspace V. This mimics the setup we have for the abstract weak form in Problem 1.4. In many problems, we will only be able to assert coercivity on a subspace of W only, as the examples below will show.

Let's see next how the coercivity of a bilinear form is related to the uniqueness of the solution of a weak form.

Theorem 3.3 (Uniqueness of solutions of a weak form)**.** *Consider the abstract weak form in Problem 1.4. Assume that the space $(\mathscr{W}, \|\cdot\|_{\mathscr{W}})$ is a normed space, and that the bilinear form $a\colon \mathscr{W} \times \mathscr{W} \to \mathbb{R}$ is coercive on $\mathscr{V}$. Then, if a solution $u \in \mathscr{S}$ to the problem exists, it is unique.*

Proof. To prove this result, we assume that there exist $u_1, u_2 \in \mathscr{S}$ that are solutions of Problem 1.4, and conclude that $u_1 = u_2$. As solutions of the problem, u_1 and u_2 satisfy that

$$a(u_1, v) = \ell(v) \tag{3.29a}$$

$$a(u_2, v) = \ell(v) \tag{3.29b}$$

for all $v \in \mathscr{V}$. Subtracting the two equations, we get

$$a(u_1 - u_2, v) = 0 \tag{3.29c}$$

for all $v \in \mathcal{V}$. Since $u_1, u_2 \in \mathcal{S}$, $u_1 - u_2 \in \mathcal{V}$, given that $\mathcal{V}$ is the direction of $\mathcal{S}$. Therefore, in (3.29c) we can choose $v = u_1 - u_2$, and use the coercivity of the bilinear form to get

$$0 = a(u_1 - u_2, u_1 - u_2) \geq c_{\mathcal{V}} \|u_1 - u_2\|^2_{\mathcal{W}} \geq 0,$$

from where we conclude that $\|u_1 - u_2\|_{\mathcal{W}} = 0$, and hence that $u_1 = u_2$. □

The weak form for Galerkin Method, Problem 1.5, is also an abstract weak form as in Problem 1.4, in which $\mathcal{W}_h$, $\mathcal{S}_h$ and $\mathcal{V}_h$ play the role of $\mathcal{W}$, $\mathcal{S}$, and $\mathcal{V}$, respectively. Therefore, Theorem 3.3 guarantees that if $a(\cdot,\cdot)$ is coercive on $\mathcal{V}_h$, then the solution of Galerkin Method is unique. Below we see that coercivity also implies that the solution exists.

Let's take a look at an example.

Examples:

3.17 Consider first the weak form for the general second-order problem in one dimension, Problem 1.3, with $b(x) = 0$, $k(x) \geq k_0 > 0$ and $c(x) \geq c_0 > 0$ for all x. The bilinear form for this problem is

$$a(u, v) = \int_0^L k(x)u'(x)v'(x) + c(x)u(x)v(x)\,dx.$$

From Example 3.14, this bilinear form is well-defined and continuous for any $u, v \in \mathcal{W} = H^1([0, L])$. It is also coercive on $\mathcal{W}$:

$$\begin{aligned} a(u, u) &= \int_0^L k(x)u'(x)^2 + c(x)u(x)^2\,dx \\ &\geq \min\{k_0, c_0\} \int_0^L u'(x)^2 + u(x)^2\,dx \\ &\geq \min\{k_0, c_0\}\|u\|^2_{1,2}. \end{aligned}$$

The coercivity constant is $c_{\mathcal{V}} = \min\{k_0, c_0\} > 0$. This implies that if the problem has a solution, it is unique, as already stated in Theorem 1.1.

For the forthcoming examples, we will need the following theorem which under appropriate conditions allows us to estimate the size of the function with the size of its derivative.

Theorem 3.4 (Poincaré's inequality)**.** *Let $\Omega \subset \mathbb{R}^n$ be a smooth-enough, connected and bounded domain. Let $\partial\Omega_D \subset \partial\Omega$ with $|\partial\Omega_D| > 0$. There exists a constant $C_p(\Omega)$ such that*

$$\|u\|_{0,2} \leq C_p(\Omega)|u|_{1,2} \tag{3.30}$$

for any $u \in \{v \in H^1(\Omega) \mid v = 0 \text{ on } \partial\Omega_D\}$.

Here $|\partial\Omega_D|$ means the *measure* of $\partial\Omega_D$, which is the number of points in the boundary of one-dimensional domains, the length for boundary curves of two-dimensional domains, and the area for boundary surfaces of three-dimensional domains. The notation $C_p(\Omega)$ is used to indicate that the constant depends on the domain, and it becomes progressively large as the size of the domain grows. Let's see this in an example.

Examples:

3.18 Let $\Omega = [-L, L]$, and $v_L(x) = \sin(\pi x/L)$. Then, $v_L = 0$ on $\partial\Omega = \{-L, L\}$ and $v_L \in H^1(\Omega)$, since

$$\|v_L\|_{0,2} = \left[\int_{-L}^{L} \sin\left(\pi \frac{x}{L}\right)^2 \, dx\right]^{1/2} = \sqrt{L},$$
$$|v_L|_{1,2} = \frac{\pi}{L}\left[\int_{-L}^{L} \cos\left(\pi \frac{x}{L}\right)^2 \, dx\right]^{1/2} = \frac{\pi}{\sqrt{L}},$$

and hence $\|v_L\|_{1,2} = \sqrt{L + \pi^2/L} < \infty$. The constant in Poincaré's inequality should then satisfy that

$$C_p(\Omega) \geq \frac{\|v_L\|_{0,2}}{|v_L|_{1,2}} = \frac{L}{\pi},$$

and hence it grows as the size of the domain does. Intuitively, what is happening is that as L grows, the derivative can become smaller since the function has to grow (or decrease) from -1 to 1 over a longer interval, as reflected in the $1/L$ in front of v_L'. At the same time, the L^2-norm of v_L grows simply because the domain does. So for the size of the derivative to bound the size of the function, the constant has to become larger as the domain grows.

3.19 The fact that the function needs to be equal to zero on some part of the boundary plays a crucial role in Poincaré's inequality. Otherwise, we can consider the constant function $v(x) = 1$ in Ω, whose gradient is equal to zero everywhere. Hence, $\|v\|_{0,2} \neq 0$ and $|v|_{1,2} = 0$, and in this case there does not exist any constant C_p to satisfy (3.30).

Intuitively, the requirement for the function to be zero on some part of the boundary anchors the value of the function, so for the function to grow or decrease in value away from $\partial\Omega_D$ it needs to have a non-zero derivative, which can be accounted for on the right hand size of the inequality.

We are now ready to examine the coercivity for some other examples.

Examples:

3.20 Let's revisit Example 3.17 and set $b(x) = c(x) = 0$ and $k(x) \geq k_0 > 0$, for all x, so the bilinear form is

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

In this case, the proof of coercivity in Example 3.17 breaks down, since, following the argument, $c_0 = 0$ and hence the coercivity constant would be zero as well. In fact, the bilinear form is no longer coercive in $\mathscr{W} = H^1([0, L])$. For example, a constant function u different than zero is in $\mathscr{W}$, but $a(u, u) = 0$.

However, what matters for Problem 1.3 is coercivity on $\mathscr{V}$, not on $\mathscr{W}$; specifically, on

$$\mathscr{V} = \{v \in \mathscr{W} \mid w(0) = 0\}.$$

Coercivity on $\mathscr{V}$ follows because the boundary condition at $x = 0$ enables the of use Poincaré's inequality. To wit, for $u \in \mathscr{V}$,

$$\begin{aligned}
a(u, u) &= \int_0^L k(x)u'(x)^2\, dx \\
&\geq k_0 \int_0^L u'(x)^2\, dx \\
&= k_0 |u|_{1,2}^2 \\
&\geq \frac{k_0}{2}|u|_{1,2}^2 + \frac{1}{C_p^2}\frac{k_0}{2}\|u\|_{0,2}^2 \qquad \text{Poincaré's inequality, (3.30)} \\
&\geq \frac{k_0}{2}\min\{1, C_p^{-2}\}\|u\|_{1,2}^2.
\end{aligned}$$

3.21 We revisit now Example 3.15, which involved a particular case of the the diffusion problem in two-dimension, Problem 2.2. Example 3.15 showed that the bilinear form and the linear functional are continuous if $\mathscr{W} = H^1(\Omega)$. However, a constant function $u \neq 0$ is in $\mathscr{W}$, and since its gradient is zero, $a(u, u) = 0$. This shows that the bilinear form is not coercive on $\mathscr{W}$.

Instead, consider the test space in Problem 2.2,

$$\mathscr{V} = \{v \in \mathscr{W} \mid v(x) = 0 \text{ for all } x \in \partial\Omega_D\},$$

where the length of $\partial\Omega_D$ is assumed to be positive (Hypothesis **H1** in Problem 2.1). In this case, we can apply Poincaré's inequality, and obtain coercivity on $\mathscr{V}$. We will also use that K is positive definite with a smallest eigenvalue greater or equal than $\kappa_0 > 0$ (Hypothesis **H2** in Problem 2.1), to get

$$(K\nabla u)\cdot\nabla u = \sum_{i,j=1}^{2} K_{ij}\frac{\partial u}{\partial x_j}\frac{\partial u}{\partial u_i} \geq \kappa_0\|\nabla u\|^2. \tag{3.31}$$