CHAPTER 4 : INTRODUCTION TO THE FINITE ELEMENT METHOD

1. Introduction

In this section, we consider static problems, and more precisely elliptic problems. The finite difference method consists **in approximating the differential operators** by finite difference, i.e., on formula of the form

$$f'(x) \approx \frac{f(x+dx) - f(x-dx)}{dx}.$$

In the Galerkin's method, on follows another strategy. In this approach, on consider actually the exact differential operators, but **approximate the space of solutions**.

2. Galerkin's method

Given $p \in \mathbb{N}$, consider an open domain $\Omega \in \mathbb{R}^p$. To detail the Galerkin's method, we focus on its application to the classical Laplace problem with, e.g., homogeneous Dirichlet's boundary conditions :

$$\begin{aligned} -\Delta u =& f \text{ in } \Omega \\ u =& 0 \text{ on } \partial\Omega. \end{aligned}$$

The first step consists in obtaining a weak formulation of the previous equations. This is done by multiplying the first equation by a function v and integrate by part. This results in the weak problem: *Find $u \in H$, such that*

$$a(u,v) = f(v), \text{ for all } v \in H. \tag{1}$$

From the theoretical point of view, the most standard tool to obtain existence and uniqueness of a solution of the latter problem is the Lax-Milgram's theorem, that requires H to be an Hilbert space, a is a continuous coercive bilinear form, and f is a continuous linear form. Note that a **does not** need to be symmetric.

2.1. Galerkin's approximation.

The principle of the method consists in approximating H by a **finite dimensional** space H_h, where h stands for a parameter controlling the discretization (for example $h = \delta x$ a step of a space discretization). Instead of considering the problem associated with (1), one rather considers *Find $u_h \in H_h$, such that*

$$a(u_h, v_h) = f(v_h), \text{ for all } v \in H_h. \tag{2}$$

One can check that as soon as Lax-Milgram applies to (1), then it also applies to (2). The advantage of this approach is that it directly gives rise to a numerical method, namely, a linear system to solve. Let us check this. Consider a (finite) basis of H_h denoted by $(e_i)_{i=1,\ldots,N_h}$. Solving (2) consists in looking for the coefficient

$x_1, \dots, x_{N_h}$ of $u_h = \sum_1^{N_h} x_j e_j$ such that (2) holds, meaning that $x_1, \dots, x_{N_h}$ must satisfy

$$\sum_{j=1}^{N_h} x_j a(e_j, v_h) = f(v_h), \text{ for all } v \in H_h. \tag{3}$$

We then replace successively v_h by e_1, e_2,..., e_{N_h} in (3) to get the system

$$\sum_1^{N_h} x_j a(e_j, e_1) = \quad f(e_1) \tag{4}$$

$$\sum_1^{N_h} x_j a(e_j, e_2) = \quad f(e_2) \tag{5}$$

$$\vdots \tag{6}$$

$$\sum_1^{N_h} x_j a(e_j, e_{N_h}) = \quad f(e_{N_h}), \tag{7}$$

so that we end up with the linear system

$$AX = b,$$

where A is a matrix whose coefficients are given by $A_{i,j} = [a(e_j, e_i)]$, $i, j = 1, \dots, N_h$, and b is the vector $[f(e_1), \dots, f(e_{N_h})]$.

Remarque 1. *Translate in the finite dimensional setting the assumptions of the Lax-Milgram theorem. What does it means for the matrix A ?*

2.2. **Convergence of the approximation.** A first *a priori* estimate of the error between the Galerkin's approximation and the true solution can be obtained thanks to the famous Céa's Lemma.

Lemme 1. *We keep the assumptions of Lax-Milgram theorem. There exists $c > 0$, independent of the choice of Galerkin approximation H_h, such that*

$$\|u - u_h\| \leq c \inf_{v_h \in H_h} \|u - v_h\|,$$

where $\| \cdot \|$ denotes the (hilbertian) norm of H.

Proof. On has

$$a(u, v_h - u_h) = f(v_h - u_h)$$
$$a(u_h, v_h - u_h) = f(v_h - u_h),$$

where the former equation is obtained by putting $v = v_h - u_h$ in (1), and the latter by putting $v = v_h - u_h$ in (2). This gives

$$a(u - u_h, v_h - u_h) = 0,$$

that holds for all $v_h \in H_h$[1]. Adding and substracting u in the term $v_h - u_h$ in the last equation, we obtain

$$a(u - u_h, u - v_h) = a(u - u_h, u - u_h).$$

[1]This identity is often called Galerkin's orthogonality.

The left-hand side term is bounded by $\gamma\|u-u_h\|\|u-v_h\|$ with γ the continuity constant of a, whereas the right-hand is lower bounded by $\alpha\|u-u_h\|^2$, with α is the coercivity constant of a. Simplifying by $\|u-u_h\|$, we get

$$\|u-u_h\| \leq \frac{\gamma}{\alpha}\|u-v_h\|.$$

Hence the result, obtained by considering the inf over $v_h \in H_h$. □

In practice,we complete this result by a more concrete on

Theorème 1. *Assume W to be a dense subspace of H, and that for each h, there exists $r_h : W \to H_h$ such that for all $v \in W$*

$$\lim_{h\to 0} r_h v = v. \tag{8}$$

Then

$$\lim_{h\to 0} u_h = u.$$

Proof. Let $\varepsilon > 0$. Since W is dense in H, there exists $w \in W$ such that

$$\|u-w\|_H \leq \frac{\varepsilon}{2C},$$

where C is the Céa constant. According to (8), there exists $h(\varepsilon) > 0$ such that, for all $h \leq h(\varepsilon)$:

$$\|u-w\|_H \leq \frac{\varepsilon}{2C}.$$

Finally, Céa's lemma gives, for $h \leq h(\varepsilon)$

$$\|u-u_h\|_H \leq C(\|u-w\|_H + \|w-r_h w\|_H) \leq \varepsilon,$$

the result follows. □

In practice, H is often a Sobolev space, in which $\mathcal{C}^m(\overline{\Omega})$ is dense. In this case, r_h corresponds to a truncation, so that the term $\|w-r_h w\|_H$ is easier to study than the bound in Céa's Lemma.

3. The Finite Element Method

We now focus on a particular Galerkin method, namely, the finite element method.

3.1. **Principle.**

The finite element method applies when Ω is bounded. In view of the previous description of the Galerkin's method, the method is completely defined as soon as H_h is fixed. In the finite element method, this space is built as follows. Up to an additional preliminary approximation, Ω is assumed to be polygonal. We then consider a *mesh*, i.e., a set of elementary polyhedrons (generally triangles in 2D and tetrahedrons in 3D) $(T_\ell)_{\ell=1,\ldots,L}$, such that

- $\cup_{\ell=1}^{L} T_\ell = \Omega$,
- $\overset{\circ}{T}_\ell \cap \overset{\circ}{T}_{\ell'} = \emptyset$, if $\ell \neq \ell'$,
- all edge of an element T_ℓ is either the edge of another element $T_{\ell'}$, or a part of the boundary $\partial\Omega$.

Then, H_h is of the form:

$$H_h := \{v_h \in \mathcal{C}^k(\Omega), \forall \ell = 1, \dots, L, v_{h|T_\ell} \text{ polynomial of degre } \leq m\}.$$

It follows that a finite element method is parametrized by to integers k and m. Note that a basis is given by piecewise polynomial and compactly supported functions, which can be defined by their values in some points of the elements.

3.2. **General process.** In order to define a finite element method, one needs to define:

- a set K, often polygonal,
- a set $\Sigma = (M_i^K)_{i=1,\dots,n_K}$ of points of K, usually called *degre of freedom*,
- a set P of polynomials containing, for some k, $\mathbb{P}^k(K)$, the set of polynomials of maximum order k on K.

Of course, K represents the elementary domain of a mesh, in the sense that an element ℓ in the mesh is obtained by an affine transformation φ_ℓ applied to K. Since composition by affine transformation preserves regularity of smooth functions and polynomials of a given degree, one consider local basis on K and deduce from it a global basis.

Note also that to be well-defined, the mapping $p \in P \mapsto (p(M_i^K))_{i=1,\dots,N_K}$ needs to be bijective. If this property holds, P is said to be $\Sigma-$*unisolvent*, and it exists in particular n_K polynomials $(\tau_i^K)_{i=1,\dots,N_K}$ in P such that

$$\tau_i^K(M_j^K) = \delta_{i,j}. \tag{9}$$

Given a mesh of L elements, we can then assemble this elementary functions to get a basis $(w_{\ell_i})_{Ln_K}$ of the whole mesh, which satisfies a property similar to (9), but for every degree of freedom of the mesh. A simple way to define it consists in defining it by its restriction on $\varphi_\ell(K)$, the element ℓ

$$(w_{\ell_i})_{|\varphi_\ell(K)} = \tau_i^K \circ \varphi_\ell^{-1}.$$

Note that this basis gives rise to sparse matrices, which one of the main advantage of the finite element method.

Remarque 2. *Two matrices are often considered in finite element (and more generally in Galerkin's method):*

- *The mass matrix, associated with the bilinear form*

$$a_M(u,v) = \int_\Omega uv,$$

- *The stiffness matrix, associated with the bilinear form*

$$a_S(u,v) = \int_\Omega \nabla u \cdot \nabla v.$$

These matrices appear, for example when considering the problem

$$-\Delta u + u = f.$$

3.3. **Convergence.** Let us consider an elliptic problem set in $H^1(\Omega)$, and an interpolation operator

$$\pi_h : \mathcal{C}^0(\overline{\Omega}) \to H_h$$
$$v \to \pi_h(v) = \sum_{\ell=1,\dots,L; i=1,\dots,n_K} v(\varphi_\ell(M_i)) w_{\ell_i}$$

Thanks to Céa's Lemma, we have, for all $w \in \mathcal{C}^0(\overline{\Omega})$

$$\|u - u_h\|_{H^1(\Omega)} \leq \|u - \pi_h w\|_{H^1(\Omega)}.$$

Assuming u to be regular enough to belong to $\mathcal{C}^0(\overline{\Omega})$, we can set $w = u$ in the previous inequality, so that it remains to estimate

$$\|u - \pi_h u\|_{H^1(\Omega)} = \left(\sum_{\ell=1,\dots,L} \|u - \pi_h u\|^2_{H^1(\varphi_\ell(K))} \right)^{\frac{1}{2}}.$$

We then need the following result, where we consider the semi-norm $|v|_{m,D}$ defined as sum of the L^2 norms on D of the derivatives of order m of v.

Theorème 2. *Assume the method under consideration is of order k, meaning that*

$$\mathbb{P}^k(K) \subset P \subset H^{k+1}(K).$$

Then, there exists c depending only on (K, Σ, P), such that

$$|v - \pi_h v|_{m,\varphi_\ell(K)} \leq C \frac{(h_{\varphi_\ell(K)})^{k+1}}{(\rho_{\varphi_\ell(K)})^m} |v|_{k+1,\varphi_\ell(K)}, \tag{10}$$

where $h_{\varphi_\ell(K)}$ and $\rho_{\varphi_\ell(K)}$ are the radii of excircle and incircles associated with $\varphi_\ell(K)$, respectively.

Formula (10) should be compared with Taylor's expansions in 1D. The proof of this result is very technical[2]. Using the previous result with $m = 0$ and $m = 1$, we get:

$$\|v - \pi_h v\|_{L^2(\varphi_\ell(K))} \leq C (h_{\varphi_\ell(K)})^{k+1} |v|_{k+1,\varphi_\ell(K)}, \tag{11}$$

$$|v - \pi_h v|_{1,\varphi_\ell(K)} \leq C \frac{(h_{\varphi_\ell(K)})^{k+1}}{\rho_{\varphi_\ell(K)}} |v|_{k+1,\varphi_\ell(K)}. \tag{12}$$

We are now in the position to get a convergence result.

Theorème 3. *Keep the assumptions of Theorem 2 and suppose that the solution u belongs to $H^{k+1}(\Omega)$. Assume also that there exists $\sigma > 0$ such that the mesh elements satisfy for $\ell = 1, \dots, L$*

$$\frac{(h_{\varphi_\ell(K)})}{(\rho_{\varphi_\ell(K)})} \leq \sigma.$$

---

[2]This proof can be found in, e.g.,

- Philippe G. Ciarlet, *Basic error estimates for elliptic problems, Dans le Hand- book of numerical analysis*, Vol. II, Eds. P.G. Ciarlet and J.-L. Lions, North Holland, pp. 17–351 (1991).
- P.-A. Raviart, J.-M. Thomas, *Introduction à l'analyse numérique des équations aux dérivées partielles*, Masson (1983).

Then there exists $C(\sigma)$ such that

$$\|u - u_h\|_{H^1(\Omega)} \leq C(\sigma) h^k |u|_{k+1,\Omega}.$$

This result can be (easily) deduced from (11) and (12).

4. To go further

This chapter is in its greater part inspired from the book

"LA METHODE DES ELEMENTS FINIS, Part. 1",

from P. Ciarlet & E. Lunéville.