Université de Rennes 1
U.F.R. Mathématiques

Master Calcul Scientifique et Modélisation
Support de cours

Introduction to the Finite Element Method

Éric DARRIGRAND

UNIVERSITÉ DE RENNES 1
UFR Mathématiques

Contents

Introduction

Modelization and numerical simulation are nowadays essential tools of engineering sciences. Beside experimental approaches, the mathematical models and their resolution by numerical methods are widely used in many domains of physics, chemistry, biology or economy. Numerical methods aim to reduce the mathematical equations to finite dimensional problems such that the solutions to those equations are approximated by the solutions to finite dimensional systems. The numerical approaches are mandatory when the configuration is complicated such that there is no analytical solution of the mathematical model. The finite element method is one of these numerical techniques and was introduced in the 1950'. Firstly proposed by the physicists, the finite element method is now well known, carefully explained in a mathematical point of view, and implemented in many codes of numerical simulation. Several finite element libraries are also proposed on the web. In particular, the library MÉLINA++ ([11]) was developed at the university of Rennes 1, France, initially by Daniel Martin, former faculty member at the university of Rennes 1, now retired. MÉLINA++ is a generic finite element library which contains a wide range of finite elements and can solve numerous problems. The library is based on the consideration of the variational formulation and the involved integrands. A significant part of this document is inherited from a course by Daniel Martin on finite elements oriented towards the use of an initial version of MÉLINA++. Now, the library is maintained by Yvon Lafranche and Eric Darrigrand but is not anymore developed. The project of a new finite element library XLiFE++ ([5]) started a few years ago to extend the library MÉLINA++ to more complicated configurations requiring other families of finite elements.

The first chapter defines the general two-dimensional framework that we aim to consider in the document and introduces a fundamental ingredient for finite elements, the variational formulation. This presentation includes mathematical tools which are required for the justification of the variational formulation. The second chapter introduces the ingredients of the finite element methods in the framework of a two-dimensional model problem. This approach enables the reader to understand the main tools of the finite element methods in a rather simple configuration. Chapter 3 is devoted to a presentation of different families of finite elements. In Chapter 4, we detail the presentation of some technical ingredients of the finite element method: the elementary objects are carefully described; the assembling procedure is considered with storage concerns. These aspects are presented according to the choices done for the development of MÉLINA++. In the last chapter, some extensions to more complicated configurations are evoked such as three dimensional finite elements, time dependent problems, integral equations, Nédélec finite elements, or linear-system resolution issues.

Chapter 1

Motivation of Finite Elements

1.1 Elliptic problems – partial differential equations and eigenvalues problems

The generic configuration which we deal with in this document consists of elliptic partial differential equations or related eigenvalue problems. In this section, we first introduce very briefly these families of mathematical problems. Then, we define the generic boundary value problem which is considered in this document.

1.1.1 Elliptic boundary value problems – partial differential equations and eigenvalue problems

Elliptic boundary value problems are widely used for modelling physical phenomena. Depending on the problem, they can take several forms such as *partial differential equations* (PDE) or *eigenvalue problems*. In this document, we consider boundary value problems defined on a bounded domain Ω with boundary $\partial\Omega = \Gamma_D \cup \Gamma_N$.

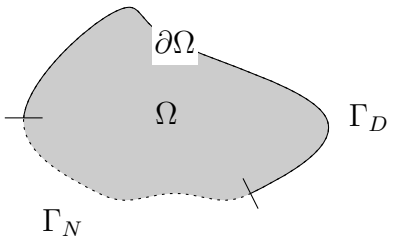

Figure 1.1: (*generated with Fig4TeX* [7]) A bounded 2D domain Ω

The *second-order elliptic partial differential equations* are defined thanks to an equation involving a second-order differential operator D. Combined with boundary conditions, they define a boundary value problem of the form:

$$\begin{cases} D\,u = f & \text{in } \Omega\,, \\ \text{boundary conditions on } \Gamma_N \text{ and } \Gamma_D\,. \end{cases}$$

The finite element method aims to approximate the solution to such a problem with a finite dimensional solution u_h assimilated to the vector solution U to a linear system of the form

$$\mathrm{D}\,U = \mathrm{F}\,,$$

The system includes the consideration of the boundary conditions.

The *second-order elliptic eigenvalue problems* are defined thanks to an eigenvalue equation involving a second-order differential operator D. Combined with boundary conditions, they define a boundary value problem of the form:

$$\begin{cases} D\,u = \lambda\,u & \text{in } \Omega\,, \\ \text{boundary conditions on } \Gamma_N \text{ and } \Gamma_D\,. \end{cases}$$

where the unknown is the eigen pair (λ, u). The finite element method aims to approximate the solution (λ, u) to such a problem with a finite dimensional solution (λ_h, u_h) where u_h is assimilated to the eigenvector U of a linear generalized eigenvalue problem

$$\mathrm{D}\,U = \lambda_h\,\mathrm{M}\,U\,,$$

The eigenvalue system includes the consideration of the boundary conditions and the matrix M is the finite element approximation of the identity operator, known as the *mass matrix*.

1.1.2 A generic 2D elliptic partial differential equation

We consider an open, bounded, polygonal, convex domain $\Omega \subset \mathbb{R}^2$, of boundary $\partial\Omega = \Gamma_D \cup \Gamma_N$ such that $\Gamma_D \cap \Gamma_N$ is empty or reduced to a finite number of points of $\partial\Omega$. We denote by $x = (x_1, x_2)$ the generic point in Ω, by γ the generic point on $\partial\Omega$, and by $\nu(\gamma) = (\nu_1(\gamma), \nu_2(\gamma))$ the outgoing unit normal vector at point γ of $\partial\Omega$.

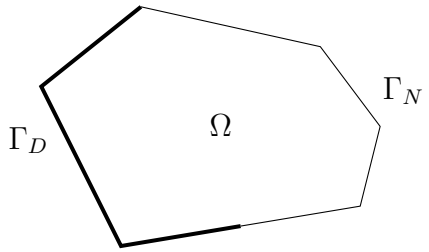

Figure 1.2: (*generated with Fig4TeX* [7]) The 2D domain Ω

We consider the functions

- $a_{\alpha\beta}$ (for $\alpha, \beta = 1, 2$), a_{00}, f_Ω defined on Ω,
- b_N, g_N defined on Γ_N,
- u_D defined on Γ_D.

The generic PDE that we are going to deal with in this document is based on the differential operator defined by:

$$\begin{aligned} \mathcal{A}u(x) &= -\sum_{\alpha,\beta=1}^{2} \frac{\partial}{\partial x_\beta}\Big(a_{\alpha\beta}(x)\frac{\partial u}{\partial x_\alpha}(x)\Big) + a_{00}(x)\,u(x)\,,\ x\in\Omega \\ \frac{\partial u}{\partial \nu_{\mathcal{A}}}(\gamma) &= \sum_{\alpha,\beta=1}^{2} a_{\alpha\beta}(\gamma)\frac{\partial u}{\partial x_\alpha}(\gamma)\nu_\beta(\gamma)\,,\ \gamma\in\partial\Omega\,. \end{aligned} \tag{1.1}$$

We then aim to explain the application of the finite element method in the context of the generic boundary value problem: find u solution to

$$\left\{\begin{array}{rcll} \mathcal{A}u = & f_\Omega & \text{in } \Omega & \\ \dfrac{\partial u}{\partial \nu_{\mathcal{A}}} + b_N u = & g_N & \text{on } \Gamma_N & \textit{Neumann/Fourier condition} \\ u = & u_D & \text{on } \Gamma_D & \textit{Dirichlet condition} \end{array}\right. \tag{1.2}$$

A very simple case corresponds to the functions $a_{\alpha\beta} = \delta^\alpha_\beta\,(\alpha,\beta = 1,2)$, where δ^α_β is the Kronecker delta, equal to 1 if $\alpha=\beta$ and 0 otherwise:

$$\mathcal{A}u = -\Big(\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2}\Big) + a_{00}\,u = -\Delta u + a_{00}\,u \quad \text{and} \quad \frac{\partial u}{\partial \nu_{\mathcal{A}}} = \operatorname{grad} u\cdot\nu = \frac{\partial u}{\partial \nu}.$$

The consideration of these equations requires the introduction of suitable spaces, the so-called Sobolev spaces (see Section 1.2).

1.2 From PDE to variational formulation – Green's formulae

A first step of the finite element method is the definition of a corresponding variational formulation. For mathematical justification, this task requires important tools of functional analysis and distribution theory such as Hilbert spaces, Sobolev spaces, trace operators, Green's formulae. Numerous results are mentioned in this section with no demonstration. For more details or demonstrations, we refer the reader to literature (e.g. [13, 8, 10]).

1.2.1 Introduction to the functional spaces

The well known $\mathcal{C}^\alpha$ spaces are not sufficient for the mathematical justification of the different steps of the finite element method. We introduce here a specific family of Hilbert spaces which are crucial in the analysis of variational formulations. Some hypotheses are also required on the regularity of the domain. To this aim, we first give a definition for the regularity of Ω. Let $\mathcal{C}^{0,\alpha}(\Omega)$ be the set of functions f satisfying for all $(x,y)\in\Omega\times\Omega$:

$$\|f(x)-f(y)\| \le \|x-y\|^\alpha$$

Definition 1.1. *$\Omega\subset\mathbb{R}^d$ is said of class $\mathcal{C}^k$ (resp. $\mathcal{C}^{k,1}$) if for any point m_0 of the boundary $\partial\Omega$ of Ω, there exists a ball $B(m_0,R)$, with $R>0$, a system of axes $\{O_0;\ x_1^0,...,x_d^0\}$ and a function $a_0\in\mathcal{C}^k(\mathbb{R}^{d-1})$ (resp. $\mathcal{C}^{k,1}$) such that*

$$\begin{aligned} \partial\Omega\cap B(m_0,R) &= \{x\in B(m_0,R)/\ x_d^0 = a_0(x_1^0,...,x_{d-1}^0)\} \\ \Omega\cap B(m_0,R) &= \{x\in B(m_0,R)/\ x_d^0 > a_0(x_1^0,...,x_{d-1}^0)\} \end{aligned}$$

Notions of Hilbert spaces

Definition 1.2 (norm, inner product)**.** *Let H be a real space.*

- ⋆ *A norm on H is a function $\mathcal{N}$ defined from H to $\mathbb{R}^+$ satisfying:*
  - ◇ $\forall u \in H, \forall \lambda \in \mathbb{R}, \quad \mathcal{N}(\lambda u) = |\lambda| \mathcal{N}(u)$,
  - ◇ $\mathcal{N}(u) = 0 \quad \Longrightarrow \quad u = 0$,
  - ◇ $\forall u, v \in H, \quad \mathcal{N}(u+v) \leq \mathcal{N}(u) + \mathcal{N}(v)$ *(triangular inequality).*
- ⋆ *We call inner product on H a function $\mathcal{P}$ defined from $H \times H$ to $\mathbb{R}$ satisfying:*
  - ◇ *$\mathcal{P}$ is bilinear symmetric,*
  - ◇ $\forall u \in H^*, \quad \mathcal{P}(u,u) > 0$.
- ⋆ *If $\mathcal{P}$ is an inner product on H, the function $\mathcal{N} : u \mapsto (\mathcal{P}(u,u))^{1/2}$ is a norm on H.*

Definition 1.3 (Cauchy sequence)**.** *Let H be a normed space, with norm $\|\cdot\|$ and $(u_n)_{n\in\mathbb{N}}$ be a sequence of points of H.*

- ⋆ *$(u_n)_{n\in\mathbb{N}}$ is said convergent in $(H, \|\cdot\|)$ if there exists $u \in H$ / $\lim_{n\to\infty} \|u_n - u\| = 0$. We say that $(u_n)_{n\in\mathbb{N}}$ converges towards u in $(H, \|\cdot\|)$.*
- ⋆ *$(u_n)_{n\in\mathbb{N}}$ is said a Cauchy sequence if, for all $p \in \mathbb{N}$, $\lim_{n\to\infty} \|u_n - u_{n+p}\| = 0$.*
- ⋆ *A convergent sequence is a Cauchy sequence. The reciprocal is generally wrong.*
- ⋆ *If all Cauchy sequences are convergent, the space is said complete.*

Definition 1.4 (Hilbert space)**.** *Let H be a real space.*

- ⋆ *H is said pre-Hilbert space if H is equipped with an inner product $\langle \cdot, \cdot \rangle$ and the corresponding norm $\|\cdot\|$.*
  Property (Cauchy-Schwarz inequality): $\forall u, v \in H, \quad |\langle u, v \rangle| \leq \|u\| \, \|v\|$.
- ⋆ *H is said Hilbert space if H is a complete pre-Hilbert space.*

The spaces of integrable functions L^1, L^p, equipped with their natural norms, are Hilbert spaces. The spaces $\mathcal{C}^\alpha$ equipped with their natural norms are not Hilbert spaces. They suffer from a lack of completeness. Indeed, the Sobolev spaces that we are going to introduce can be interpreted as the completion of the corresponding spaces $\mathcal{C}^\alpha$ equipped with the ad hoc norm.

Notions of distributions and dual product

In order to introduce the Sobolev spaces, we first briefly present the theory of distributions. A *distribution* is a concept which generalizes the notion of function and makes possible the operation of derivation in a particular sense so-called *distributional sense* or *weak sense*. In this section, we use the following notation: for $\alpha = (\alpha_1, \alpha_2, ..., \alpha_d) \in \mathbb{R}^d$ a multi-index,

$$\partial^\alpha \varphi = \frac{\partial^{|\alpha|} \varphi}{\partial_{x_1}^{\alpha_1} \partial_{x_2}^{\alpha_2} \cdots \partial_{x_d}^{\alpha_d}} \qquad \text{with } |\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_d \, .$$

Definition 1.5 (test function space – distribution)**.** *Let Ω be an open subset of $\mathbb{R}^d$.*

- ⋆ *A subset K of Ω is said compact if it is a closed and bounded subset of Ω.*

- ⋆ *We denote by $\mathcal{D}(\Omega)$ the set of functions which are $\mathcal{C}^\infty$ on $\mathbb{R}^d$ and the support of which is compact in Ω.*

- ⋆ *u is said a distribution on Ω if u is a linear form on $\mathcal{D}(\Omega)$*

$$u : \varphi \in \mathcal{D}(\Omega) \mapsto \langle u, \varphi \rangle$$

and satisfy the continuity property: for all compact K of Ω, there exists an integer k and a constant C_K such that

$$\forall \varphi \in \mathcal{D}(\Omega) \text{ such that } \operatorname{supp}(\varphi) \subset K, \qquad |\langle u, \varphi \rangle| \leq C_K \max_{|\alpha| \leq k} \left\{ \|\partial^\alpha \varphi\|_{L^\infty(\Omega)} \right\} .$$

- ⋆ *We denote by $\mathcal{D}'(\Omega)$ the space of distributions on Ω.*

In this definition, $\langle \cdot, \cdot \rangle$ is the duality product on $\mathcal{D}'(\Omega) \times \mathcal{D}(\Omega)$. For a function $f \in L^1_{\text{loc}}(\Omega) = \{ f \,/ f \in L^1(K)\,,\ \forall K \text{ compact } \subset \Omega \}$, the linear form

$$\varphi \mapsto \int_\Omega f\, \varphi \, dx$$

defines a distribution. The continuity property is satisfied with $k = 0$ and $C_K = \int_K |f(x)|\, dx$. The distribution is identified to the function and we also denote by f the distribution associated with f.

The space of distributions is equipped with notions of convergence and derivation thanks to the following definition.

Definition 1.6 (distributional convergence and derivation)**.** *Let Ω be an open subset of $\mathbb{R}^d$.*

- ⋆ *The sequence of distributions $(u_n)_{n \in \mathbb{N}}$ is said convergent towards $u \in \mathcal{D}'(\Omega)$ if and only if*

$$\forall \varphi \in \mathcal{D}(\Omega), \qquad \langle u_n, \varphi \rangle \longrightarrow \langle u, \varphi \rangle \text{ in } \mathbb{R}\,.$$

- ⋆ *For $u \in \mathcal{D}'(\Omega)$, $\Omega \subset \mathbb{R}^d$, we denote by $\frac{\partial u}{\partial x_i}$ ($1 \leq i \leq d$) the distribution defined by*

$$\forall \varphi \in \mathcal{D}(\Omega), \qquad \langle \frac{\partial u}{\partial x_i}, \varphi \rangle = -\langle u, \frac{\partial \varphi}{\partial x_i} \rangle \ .$$

If α is a multi-index, we denote $\partial^\alpha u$ the distribution defined by

$$\forall \varphi \in \mathcal{D}(\Omega), \qquad \langle \partial^\alpha u, \varphi \rangle = (-1)^{|\alpha|} \langle u, \partial^\alpha \varphi \rangle \ .$$

Thanks to this definition, any function L^1, L^p, which can be identified to a distribution, is derivable in the distributional sense.

Notions of Sobolev spaces

The previous definitions introduce a strong ingredient of functional analysis: the distributional derivation. When a function is classically derivable, its classical derivative coincides with the distributional derivative, but when the function is not derivable we consider the distributional derivative. In the sequel, we will consider either the classical or the distributional derivative without specifying.

Definition 1.7 (Sobolev space $H^1(\Omega)$)**.** *Let Ω be an open subset of $\mathbb{R}^d$. We say that u is in $H^1(\Omega)$ if $u \in L^2(\Omega)$ and $\frac{\partial u}{\partial x_i} \in L^2(\Omega)$ for $i = 1, ..., d$ (distributional derivatives ... of course). This space is equipped with the inner product*

$$\langle u, v \rangle_{H^1} = \int_\Omega u\, v\, dx + \sum_{i=1}^{d} \int_\Omega \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_i}\, dx$$

and the corresponding norm

$$\|u\|_{H^1} = \left(\|u\|_{L^2}^2 + \sum_{i=1}^{d} \left\| \frac{\partial u}{\partial x_i} \right\|_{L^2}^2 \right)^{1/2}$$

Theorem 1.8. *The space $H^1(\Omega)$ equipped with this inner product $\langle \cdot, \cdot \rangle_{H^1}$ is a Hilbert space.*

Remark 1.9. *This definition can be extended:*

- ⋆ *The spaces $H^m(\Omega)$, for $m \in \mathbb{N}$, defined by*

  $$H^m(\Omega) = \left\{ u \in L^2(\Omega) \,/\, \forall \alpha \text{ multi-index such that } |\alpha| \leq m\,,\ \partial^\alpha u \in L^2(\Omega) \right\}\ ,$$

  equipped with the inner product

  $$(u, v)_{H^m} = \sum_{|\alpha| \leq m} \int_\Omega \partial^\alpha u\, \partial^\alpha v\, dx\ ,$$

  and the corresponding norm

  $$\|u\|_{H^m} = \left(\sum_{|\alpha| \leq m} \|\partial^\alpha u\|_{L^2}^2 \right)^{1/2}\ ,$$

  are Hilbert spaces.

- ⋆ *When $\Omega = \mathbb{R}^d$, we define the spaces $H^s(\mathbb{R}^d)$ with $s \in \mathbb{R}^+$:*

  $$H^s(\mathbb{R}^d) = \left\{ u \in L^2(\mathbb{R}^d) \,/\, (1 + |\xi|^2)^{s/2} \hat{u} \in L^2(\mathbb{R}^d) \right\}\ ,$$

where $\hat{u}$ is the Fourier transform of the function u. When s is an integer, the last definition is equivalent to the previous. The demonstration is based on Fourier transform of functions of L^2 and its distributional derivatives.

All these spaces satisfy properties such as density results or compactness injection. A crucial result concerns the completion $H_0^1(\Omega)$ of $\mathcal{D}(\Omega)$ in $(H^1(\Omega), \|\cdot\|_{H^1})$. The result is done in Theorem 1.12. The consideration of the space $H_0^1(\Omega)$ enables us to introduce $H^{-1}(\Omega)$ as the dual of $H_0^1(\Omega)$ (that is the set of linear forms defined on $H_0^1(\Omega)$) which is characterized by:

$$H^{-1}(\Omega) \overset{\text{def}}{=} \left(H_0^1(\Omega)\right)' = \left\{ u \in \mathcal{D}'(\Omega) \,/\, u = f_0 + \sum_{i=1}^{d} \frac{\partial f_i}{\partial x_i}, f_0 \in L^2(\Omega), f_i \in L^2(\Omega), 1 \leq i \leq d \right\}$$

Remark 1.10. *In $\mathbb{R}^d$, the Sobolev spaces can be defined thanks to the Fourier transform: for $s \in \mathbb{R}$,*

$$H^s(\mathbb{R}^d) = \left\{ u \in S'(\mathbb{R}^d) / (1 + |\xi|^2)^{s/2} \hat{u} \in L^2(\mathbb{R}^d) \right\}$$

This definition coincides with the previous ones and uses the concept of Fourier transform on the space of tempered distributions $S'(\mathbb{R}^d)$ which is the dual of the so-called Schwartz space $S(\mathbb{R}^d)$. The Schwartz space is the space of the functions of $\mathcal{C}^\infty$ which are rapidly decreasing at infinity.

Notions of trace operators and trace spaces

A boundary value problem requires the consideration of the restriction of functions to the boundary $\partial\Omega$ of the computational domain Ω. The mathematical justification of such consideration is given thanks to *trace operators*. The restrictions of functions which can be restricted to the boundary belong to *trace spaces*.

Theorem 1.11. *Let Ω be a bounded open subset of $\mathbb{R}^d$, of class $\mathcal{C}^{0,1}$.*

i) $H_0^1(\Omega) = \{u \in H^1(\Omega) \,/\, \widetilde{u} \in H^1(\mathbb{R}^d)\}$ where $\widetilde{u}$ is the extension of u by 0 outside of Ω.

ii) The mapping $\gamma_0 : u \mapsto u|_{\partial\Omega}$ is linear and continuous from $H^1(\Omega)$ to $L^2(\partial\Omega)$, and is called "trace operator".

Theorem 1.12. *Let Ω be bounded and of class $\mathcal{C}^{0,1}$. Then,*

$$H_0^1(\Omega) = \{u \in H^1(\Omega) \,/\, \gamma_0 u = 0\} = \ker \gamma_0$$

Two trace spaces will be considered in the justification of the finite element method:

$$H^{1/2}(\partial\Omega) = \{\underline{v} \in L^2(\partial\Omega) / \ \exists v \in H^1(\Omega) \text{ such that } \underline{v} = \gamma_0 v\} \,,$$
$$H^{3/2}(\partial\Omega) = \{\underline{v} \in L^2(\partial\Omega) / \ \exists v \in H^2(\Omega) \text{ such that } \underline{v} = \gamma_0 v\} \,.$$

1.2.2 Green's formulae

The well known one dimensional integration by parts is extended to higher dimensions by the so-called *Green's formula*:

Theorem 1.13 (Green's formula)**.** *If $\Omega \subset \mathbb{R}^d$ is $\mathcal{C}^{0,1}$, then $\forall u, v \in H^1(\Omega)$, we have*

$$\int_\Omega \frac{\partial u}{\partial x_i} v = -\int_\Omega u \frac{\partial v}{\partial x_i} + \int_{\partial\Omega} \gamma_0 u \, \gamma_0 v \, \nu_i \,, \qquad 1 \le i \le d \,, \tag{1.3}$$

where ν_i is the i-th component of the outgoing unit normal vector ν on $\partial\Omega$.

The demonstration of the formula is easy when $u, v \in \mathcal{C}^1(\Omega)$. The difficulty occurs in the consideration of Sobolev spaces. The result is extended to those spaces thanks to density theorems of functional analysis. In the sequel, we ignore the trace operator for readability and write

$$\int_\Omega \frac{\partial u}{\partial x_i}(x)\, v(x)\, dx = -\int_\Omega u(x)\, \frac{\partial v}{\partial x_i}(x)\, dx + \int_{\partial\Omega} u(\gamma)\, v(\gamma)\, \nu_i(\gamma)\, d\gamma \,, \qquad 1 \le i \le d \,,$$

but the integral over $\partial\Omega$ of a function defined on Ω would have to be considered in terms of trace operator.

A first consequence of Green's formula is the following corollary that will be used for the first derivation of the 2D finite element method in this document:

Corollary 1.14. *If $\Omega \subset \mathbb{R}^d$ is $\mathcal{C}^{0,1}$, then $\forall u \in H^2(\Omega), \forall v \in H^1(\Omega)$, we have*

$$\int_\Omega \Delta u(x)\, v(x)\, dx = -\int_\Omega \nabla u(x) \cdot \nabla v(x)\, dx + \int_{\partial\Omega} \frac{\partial u}{\partial \nu}(\gamma)\, v(\gamma)\, d\gamma\,. \tag{1.4}$$

The generic formula associated with the generic problem defined in 1.1.2 is given by the corollary:

Corollary 1.15. *If $\Omega \subset \mathbb{R}^2$ is $\mathcal{C}^{0,1}$, then $\forall u \in H^2(\Omega), \forall v \in H^1(\Omega)$, we have*

$$\begin{aligned} &\int_\Omega \sum_{\alpha,\beta=1}^{2} \frac{\partial}{\partial x_\beta}\Big(a_{\alpha\beta}\frac{\partial u}{\partial x_\alpha}\Big)(x)\, v(x)\, dx \\ = \; &-\int_\Omega \sum_{\alpha,\beta=1}^{2} a_{\alpha\beta}(x)\frac{\partial u}{\partial x_\alpha}(x)\frac{\partial v}{\partial x_\beta}(x)\, dx + \int_{\partial\Omega} \frac{\partial u}{\partial \nu_{\mathcal{A}}}(\gamma)\, v(\gamma)\, d\gamma\,. \end{aligned} \tag{1.5}$$

1.2.3 From the PDE to the variational formulation

To derive the generic variational formulation, we multiply the generic PDE by a test function and integrate over the domain

$$-\int_\Omega \sum_{\alpha,\beta=1}^{2} \frac{\partial}{\partial x_\beta}\Big(a_{\alpha\beta}(x)\frac{\partial u}{\partial x_\alpha}(x)\Big)\; v(x)\, dx + \int_\Omega a_{00}(x)\, u(x)\, v(x)\, dx = \int_\Omega f_\Omega(x)\, v(x)\, dx$$

Then, the global order of the derivatives is reduced thanks to the Green's formula. The application of the formula (1.5) leads to

$$\begin{aligned} &\int_\Omega \sum_{\alpha,\beta=1}^{2} a_{\alpha\beta}(x)\frac{\partial u}{\partial x_\alpha}(x)\frac{\partial v}{\partial x_\beta}(x)\, dx - \int_{\partial\Omega} \frac{\partial u}{\partial \nu_{\mathcal{A}}}(\gamma)\, v(\gamma)\, d\gamma + \int_\Omega a_{00}(x)\, u(x)\, v(x)\, dx \\ = \; &\int_\Omega f_\Omega(x)\, v(x)\, dx \end{aligned} \tag{1.6}$$

Let us introduce some notations:

- ⋆ the unknown belongs to the space $W = H^1(\Omega)$;
- ⋆ the test function of the variational formulation is taken in W, satisfying $v|_{\Gamma_D} = 0$, $v \in V$ with
$$V = \{v \in W / v|_{\Gamma_D} = 0\};$$

⋆ for $u \in W$ and $v \in V$, we consider the bilinear form

$$\begin{aligned} a(u,v) = & \int_\Omega \Big(\sum_{\alpha,\beta=1}^{2} a_{\alpha\beta}(x) \frac{\partial u}{\partial x_\alpha}(x) \frac{\partial v}{\partial x_\beta}(x) \; + \; a_{00}(x) u(x) v(x) \Big) \, dx \\ & + \int_{\Gamma_N} b_N(\gamma) u(\gamma) v(\gamma) \, d\gamma \, ; \end{aligned} \tag{1.7}$$

and the linear form

$$f(v) = \int_\Omega f_\Omega(x) v(x) \, dx + \int_{\Gamma_N} g_N(\gamma) v(\gamma) \, d\gamma \, . \tag{1.8}$$

The combination of the boundary conditions, the definition of the spaces W and V, the definition of the forms a and f, together with the expression (1.6), leads to the variational form of the generic problem (1.2):

$$\left| \begin{array}{ll} \text{Find } u \in W \;\; \text{such that} & u = u_D \;\; \text{on } \Gamma_D \\ & a(u,v) = f(v) \, , \; \forall v \in V \, . \end{array} \right. \tag{1.9}$$

Due to the regularity hypotheses on the functions, we can ensure existence and uniqueness of the solution to (1.9). A fundamental lemma is done in Section 1.3 regarding this question.

1.2.4 The Dirichlet condition

The Dirichlet condition is not considered in the expression of the equation involved in the variational form. This condition only appears in the definition of the functional space. For the establishment of theoretical properties, it is usual to consider first the case of a homogeneous Dirichlet condition (i.e. $u_D = 0$), then to deal with the general Dirichlet condition by a change of unknown which links to the case of a homogeneous condition. Even in a practical point of view, it may be easier to implement the homogeneous Dirichlet condition ($u_D = 0$).

Let us consider the problem

$$\left| \begin{array}{ll} \text{Find } u \in W \;\; \text{such that} & u = u_D \text{ on } \Gamma_D \, , \\ & a(u,v) = f(v) \, , \; \forall v \in V \end{array} \right. \tag{1.10}$$

where we suppose that there exists a function $U_D \in H^1(\Omega)$ such that $U_D|_{\Gamma_D} = u_D$. We then introduce a new unknown $U = u - U_D$ and the problem (1.10) becomes

$$\left| \begin{array}{ll} \text{Find } U \in V \;\; \text{such that} & \forall v \in V \\ & a(U,v) = f(v) - a(U_D, v) \, , \end{array} \right. \tag{1.11}$$

Remark 1.16. *The function U_D is not unique. This means that the solution U depends on U_D but the solution $u = U + U_D$ does not depend on U_D.*

1.3 Wellposedness of variational formulations – Lax-Milgram Lemma

A fundamental result for the justification of the wellposedness of variational formulations is Lax-Milgram Lemma:

Lemma 1.17 (Lax-Milgram Lemma)**.** *Let us consider a variational formulation*

$$\left| \begin{array}{ll} \text{Find } u \in V \text{ such that} & \forall v \in V \\ & a(u,v) = f(v)\,, \end{array} \right. \tag{1.12}$$

where

- $\star$ V *is a Hilbert space,*
- $\star$ a *is a bilinear continuous form on* $V \times V$*,*
- $\star$ f *is a linear continuous form on* V*.*

If the bilinear form a *is coercive, that is*

$$\exists \alpha > 0 \,/\, \forall u \in V,\ a(u,u) \geq \alpha \left\|u\right\|_V^2\ ,$$

then, the variational formulation (1.12) has a unique solution.

The most difficult aspect of the theorem is the satisfaction of the coercivity of the bilinear form. For some problems, the Poincaré inequality may help:

Theorem 1.18 (Poincaré inequality)**.** *Let* Ω *be a bounded subset of* $\mathbb{R}^d$*, of class* $\mathcal{C}^{0,1}$*. There exists a constant* $C_\Omega > 0$ *depending only on* Ω *such that*

$$\forall u \in H_0^1(\Omega)\,, \qquad \left\|u\right\|_{L^2} \leq C_\Omega \left\|\nabla u\right\|_{L^2}\ .$$

Remark 1.19. *An equivalent result holds even if we force* u *to be 0 only on a part* Γ_D *of* $\partial\Omega$ *as soon as* $\mathrm{meas}(\Gamma_D) \neq 0$*.*

Such results will lead to the following statements, for suitable hypotheses on Ω, f_Ω, g_N and u_D:

- $\star$ The following problem and its corresponding variational formulation have a unique solution:

$$\text{Find } u \in H^2(\Omega) \text{ such that } \left\{ \begin{array}{rcll} -\Delta u + u & = & f_\Omega & \text{in } \Omega\,, \\ \dfrac{\partial u}{\partial \nu} & = & g_N & \text{on } \partial\Omega\,. \end{array} \right.$$

  In this case, $V = H^1(\Omega)$, and the bilinear form is the inner product on $H^1(\Omega)$.

- $\star$ The following problem and its corresponding variational formulation do not have a unique solution:

$$\text{Find } u \in H^2(\Omega) \text{ such that } \left\{ \begin{array}{rcll} -\Delta u & = & f_\Omega & \text{in } \Omega\,, \\ \dfrac{\partial u}{\partial \nu} & = & g_N & \text{on } \partial\Omega\,. \end{array} \right.$$

  If we get a solution u, for any constant α, $u + \alpha$ is also a solution. The bilinear form of the variational formulation is not coercive.

- $\star$ The following problem and its corresponding variational formulation have a unique solution:

$$\text{Find } u \in H^2(\Omega) \text{ such that } \left\{ \begin{array}{rcll} -\Delta u & = & f_\Omega & \text{in } \Omega\,, \\ u & = & u_D & \text{on } \partial\Omega\,. \end{array} \right.$$

  The bilinear form of the variational formulation is coercive thanks to Poincaré inequality.

1.4 Galerkin method: discretization of variational formulations

In this section, we consider a general 2D configuration with homogeneous Dirichlet condition, based on a generic variational formulation

$$\left\{ \begin{array}{ll} \text{Find } u \in V \text{ such that} & \forall v \in V \\ & a(u,v) = f(v)\,. \end{array} \right. \tag{1.13}$$

The functional space V, and the bilinear form a and the linear form f are supposed to satisfy some properties which ensure existence and uniqueness of the solution. For example, they satisfy the hypotheses of Lemma (1.17). Moreover, we consider that the space V contains the homogeneous Dirichlet condition on Γ_D: $\forall v \in V$, $v|_{\Gamma_D} = 0$. We derive formally a finite element approximation and the different aspects of its implementation. This section introduces the general concept of finite elements and indicates the usual strategy of application and implementation.

From the variational formulation, the first step consists in the choice of a partition of the computational domain Ω named a *mesh* or a *triangulation* when it consists of triangles, for the discretization of the geometry. For the sake of simplicity, we consider and describe here a triangulation $\mathcal{T}_h$ of Ω, a set of triangles K satisfying:

i) $\overline{\Omega}_h = \bigcup_{K\in\mathcal{T}_h} K$ is a polygonal approximation of $\overline{\Omega}$, of boundary $\Gamma_{N,h} \cup \Gamma_{D,h}$ with $\Gamma_{N,h}$ (resp. $\Gamma_{D,h}$) approximation of Γ_N (resp. Γ_D);

ii) $\forall K \in \mathcal{T}_h\,, \overset{\circ}{K} \neq \emptyset\,;$

iii) $\forall K_1 \in \mathcal{T}_h\,, \forall K_2 \in \mathcal{T}_h,\ \overset{\circ}{K}_1 \cap \overset{\circ}{K}_2 = \emptyset$ or $\overset{\circ}{K}_1 = \overset{\circ}{K}_2$;

iv) $\forall K'$ edge of $K_1 \in \mathcal{T}_h$, K' is either an edge of another triangle $K_2 \in \mathcal{T}_h$, or a part of $\Gamma_{D,h}$ or a part of $\Gamma_{N,h}$.

The parameter $h = \max\limits_{K\in\mathcal{T}_h} \mathrm{diam}(K)$ measures how fine the triangulation is. "diam" stands for "diameter". For a triangle, the diameter is the length of the longest edge. The triangulation defines an approximation of the geometry when the domain Ω is not polygonal, otherwise it is possible to choose the triangulation such that $\Omega_h = \Omega$.

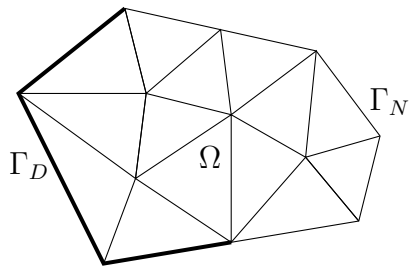

Figure 1.3: (*generated with Fig4TeX* [7]) Example of a triangulation of a polygonal domain Ω

The initial variational formulation (1.13) is now replaced with

$$\left\{ \begin{array}{ll} \text{Find } u \in \widetilde{V} \text{ such that} & \forall v \in \widetilde{V} \\ & a_h(u,v) = f_h(v)\,, \end{array} \right. \tag{1.14}$$

where $\widetilde{V}$, a_h and f_h are defined on Ω_h like V, a and f are defined on Ω (e.g., if $V = H^1(\Omega)$, then $\widetilde{V} = H^1(\Omega_h)$).

A second step consists in the approximation of the space $\widetilde{V}$ by a finite dimensional space V_h. Let us consider V_h defined by a basis $(\varphi_i)_{i=1,...,n}$ where n denotes the dimension of V_h. The functions φ_i, $i = 1, ..., n$ are the so-called *basis functions*.

Replacing $\widetilde{V}$ by V_h, the unknown u of (1.14) is approximated by $u_h = \sum_{j=1}^n u_j \varphi_j$, with $\{\varphi_i, i = 1, ..., n\}$ a basis of V_h, and the equation in (1.14) requires only to be written for v in $\{\varphi_i, i = 1, ..., n\}$. This leads to the system

$$\forall i = 1, ..., n\,, \qquad \sum_{j=1}^{n} u_j\, a_h(\varphi_j, \varphi_i) = f_h(\varphi_i) \tag{1.15}$$

where the unknown is the vector $(u_1, ..., u_n) \in \mathbb{R}^n$. The system (1.15) can be written in matricial form:

$$AX = B \tag{1.16}$$

with the matrix $A \in \mathbb{R}^{n\times n}$ and the right hand side $B \in \mathbb{R}^n$ defined by

$$A_{ij} = a_h(\varphi_j, \varphi_i)\,, \qquad B_i = f_h(\varphi_i) \tag{1.17}$$

and the unknown in $\mathbb{R}^n$: $X = (u_1, ..., u_n)^{\mathrm{T}}$. This corresponds to a Galerkin formulation of the initial problem (1.13), a finite dimensional variational formulation.

Next step consists in specifying the choice of V_h or more exactly the choice of a set of basis functions $\{\varphi_i, i = 1, ..., n\}$. Indeed, the finite element method is such a Galerkin method which answers this question with remarquable properties:

- ⋆ ease of calculation of the expressions $a_h(\varphi_j, \varphi_i)$ and $f_h(\varphi_i)$ which usually involve evaluations of integrals;
- ⋆ explicit expression of the basis $\{\varphi_i, i = 1, ..., n\}$ of the functional space V_h;
- ⋆ convenient properties of the matrix A for the resolution of the system;
- ⋆ convergence of the solution u_h to the linear system towards the solution to the variational formulation on Ω.

These aspects are carefully discussed in next chapter for a toy 2D problem and more generally considered in the following chapters.

Chapter 2

A complete 2D Lagrange Finite Element application

In this chapter, we propose to derive carefully the finite element approximation in the case of a 2D classical problem where the required mathematical properties are rather easy to check.

2.1 A toy 2D problem

We consider the following problem

$$\left\{\begin{array}{rcll} \text{Find } u & \in & H^2(\Omega) & \text{such that} \\ -\Delta u + u & = & f_\Omega & \text{in } \Omega\,, \\ \dfrac{\partial u}{\partial \nu} & = & g_N & \text{on } \Gamma_N \quad \textit{(Neumann condition)}\,, \\ u & = & u_D & \text{on } \Gamma_D \quad \textit{(Dirichlet condition)}\,, \end{array}\right. \tag{2.1}$$

where $\Omega \in \mathbb{R}^2$, Γ_N and Γ_D are defined like in Section 1.1, see Figure 1.2. If Ω is regular enough, considering $f_\Omega \in L^2(\Omega)$, $u_D \in H^{3/2}(\Gamma_D)$ and $g_N \in H^{1/2}(\Gamma_N)$, the problem is wellposed: the solution exists and is unique.

2.2 Variational formulation and wellposedness

To derive the variational form of (2.1), we first multiply the equation on Ω by a test function v satisfying $v|_{\Gamma_D} = 0$ and integrate over Ω:

$$-\int_\Omega \Delta u(x)v(x)\,dx + \int_\Omega u(x)v(x)\,dx = \int_\Omega f_\Omega(x)v(x)\,dx\,.$$

By considering the Green's formula (1.4), we obtain

$$\int_\Omega \nabla u(x)\cdot\nabla v(x)\,dx + \int_\Omega u(x)v(x)\,dx = \int_\Omega f_\Omega(x)v(x)\,dx + \int_{\partial\Omega} \frac{\partial u}{\partial \nu}(\gamma)v(\gamma)\,d\gamma\,.$$

The consideration of the boundary conditions (vanishing the test function on the Dirichlet boundary, using the Neumann condition) gives

$$\int_\Omega \nabla u(x)\cdot\nabla v(x)\,dx + \int_\Omega u(x)v(x)\,dx = \int_\Omega f_\Omega(x)v(x)\,dx + \int_{\Gamma_N} g_N(\gamma)v(\gamma)\,d\gamma\,.$$

To ensure the wellposedness of all the terms in the previous expression, the suitable functional space is $H^1(\Omega)$. Then, the variational form of Problem (2.1) is

$$\left| \begin{array}{ll} \text{Find } u \in W \text{ such that} & u|_{\Gamma_D} = u_D \,, \\ & a(u,v) = f(v) \,, \ \forall v \in V \,, \end{array} \right. \tag{2.2}$$

with

$$\begin{aligned} & W = H^1(\Omega) \,, \qquad V = \{v \in W \,/\, v|_{\Gamma_D} = 0\} \,, \\ & a(u,v) = \int_\Omega \nabla u(x) \cdot \nabla v(x)\, dx + \int_\Omega u(x) v(x)\, dx \,, \\ & f(v) = \int_\Omega f_\Omega(x) v(x)\, dx + \int_{\Gamma_N} g_N(\gamma) v(\gamma)\, d\gamma \,. \end{aligned} \tag{2.3}$$

The bilinear form a is indeed the usual inner product on $H^1(\Omega)$, and the linear form f is continuous due to the hypothesis on f_Ω and the trace theorems of functional analysis. The variational formulation satisfies the hypotheses of Lax-Milgram Lemma 1.17, the existence and uniqueness of the solution is ensured and the solution to (2.2) is the solution to (2.1). Existence and uniqueness of the solution to (2.2) can be checked with weaker hypotheses, $f_\Omega \in H^{-1}(\Omega)$, $u_D \in H^{1/2}(\Gamma_D)$ and $g_N \in H^{-1/2}(\Gamma_N)$, but the solution to (2.2) would not be a solution to (2.1) in a standard sense anymore. Such a solution would belong to $H^1(\Omega) \setminus H^2(\Omega)$ and would be said *weak solution to* (2.1) or *distributional solution*.

2.3 $\mathbb{P}_1$ Lagrange triangulation

For the mesh of the domain Ω, we choose a triangulation $\mathcal{T}_h$ which consists of $\mathbb{P}_1$ triangles. Each triangle is defined by its three vertices. We suppose that $\mathcal{T}_h$ satisfies the usual properties:

i) $\overline{\Omega}_h = \bigcup_{K \in \mathcal{T}_h} K$ is a polygonal approximation of $\overline{\Omega}$, of boundary $\Gamma_{N,h} \cup \Gamma_{D,h}$ with $\Gamma_{N,h}$ (resp. $\Gamma_{D,h}$) approximation of Γ_N (resp. Γ_D);

ii) $\forall K \in \mathcal{T}_h \,, \overset{\circ}{K} \neq \emptyset$;

iii) $\forall K_1 \in \mathcal{T}_h \,, \forall K_2 \in \mathcal{T}_h, \ \overset{\circ}{K}_1 \cap \overset{\circ}{K}_2 = \emptyset$ or $\overset{\circ}{K}_1 = \overset{\circ}{K}_2$;

iv) $\forall K'$ edge of $K_1 \in \mathcal{T}_h$, K' is either an edge of another triangle $K_2 \in \mathcal{T}_h$, or a part of $\Gamma_{D,h}$ or a part of $\Gamma_{N,h}$.

The parameter $h = \max\limits_{K \in \mathcal{T}_h} \operatorname{diam}(K)$ measures how fine the triangulation is. We refer to Figure 1.3 of Section 1.4 for an example of such a triangulation.

For the definition of the finite dimensional functional space, we introduce a $\mathbb{P}_1$-piecewise interpolation of the unknown function based on the knowlegde of the function at the vertices of the triangles. This corresponds to the so-called $\mathbb{P}_1$ *Lagrange finite elements*. The finite element related to a given triangle K is defined by :

⋆ the geometrical element, entirely described by the vertices a_1^K, a_2^K, a_3^K of the triangle K;

⋆ the functional space on K: $\mathbb{P}_1(K)$ = space of polynomials of degree at most 1 with respect to all its variables considered all together;

- ⋆ the basis functions of $\mathbb{P}_1(K)$ associated with the nodes a_1^K, a_2^K, a_3^K: λ_1^K, λ_2^K, λ_3^K of $\mathbb{P}_1(K)$ defined by $\lambda_n^K(a_\ell^K) = \delta_n^\ell$ for $n, \ell = 1, 2, 3$.

The basis functions satisfy the following properties

- ⋆ $(\lambda_1^K, \lambda_2^K, \lambda_3^K)$ is a basis of $\mathbb{P}_1(K)$;
- ⋆ If $p \in \mathbb{P}_1(K)$ satisfies $p(a_\ell^K) = \alpha_\ell$ for $\ell = 1, 2, 3$, then p is uniquely determined and $p = \sum_{\ell=1}^3 \alpha_\ell \lambda_\ell^K$;
- ⋆ $\lambda_1^K + \lambda_2^K + \lambda_3^K = 1$;
- ⋆ $x = \sum_{\ell=1}^3 a_\ell^K \lambda_\ell^K(x) \quad \forall x \in \mathbb{R}^2$.

The functions λ_1^K, λ_2^K, λ_3^K are called barycentric coordinates with respect to the vertices a_1^K, a_2^K, a_3^K of the triangle K.

The definition of the basis functions is based on the association of each basis function with one point of K. These points which are associated with the basis functions are called the *nodes* of the finite element. In the case of $\mathbb{P}_1$ Lagrange finite elements, the nodes are the vertices but this is not the case for more general finite elements (see next chapter). The $\mathbb{P}_1$ Lagrange finite element is illustrated in Figure 2.1 where the points n_ℓ, $\ell = 1, ..., 3$, are the nodes of the finite element.

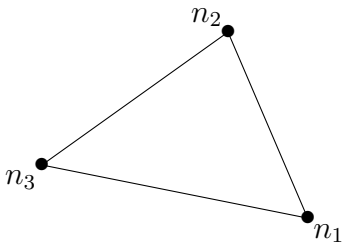

Figure 2.1: (*generated with Fig4TeX* [7]) Set of nodes of the $\mathbb{P}_1$ Lagrange finite element

2.4 Discrete space, numbering, basis functions

Let us denote by $\mathcal{N}^K$ the set of nodes of the element $K \in \mathcal{T}_h$, and $N^K = \mathrm{card}(\mathcal{N}^K)$. In the current configuration of $\mathbb{P}_1$ finite elements, $N^K = 3$, and the following properties are satisfied: for any triangles K_1 and K_2 sharing an edge,

- ⋆ $\mathcal{N}^{K_1} \cap K_2 = \mathcal{N}^{K_2} \cap K_1$;
- ⋆ $\mathbb{P}_1(K_1)|_{K_1 \cap K_2} = \mathbb{P}_1(K_2)|_{K_1 \cap K_2}$.

We now introduce $\mathcal{M}_h$ the set of nodes of the triangulation $\mathcal{T}_h$ as the union of the nodes of the elements:

$$\mathcal{M}_h = \bigcup_{K \in \mathcal{T}_h} \mathcal{N}^K$$

and $M_h = \mathrm{card}(\mathcal{M}_h)$. We introduce in the same way $\mathcal{N}_h$ the set of nodes of $\mathcal{M}_h$ which are not contained in Γ_D, and $N_h = \mathrm{card}(\mathcal{N}_h)$. The set of elements which contain the node $n \in \mathcal{M}_h$ is

$$\mathcal{L}_h(n) = \{K \in \mathcal{T}_h \,/\, n \in \mathcal{N}^K\}.$$

These notations enable us to define the finite dimensional spaces

$$\begin{aligned} W_h &= \left\{v = (v_K)_{K\in\mathcal{T}_h} \in \textstyle\prod_{K\in\mathcal{T}_h} \mathbb{P}_1(K) \,/\, \forall n \in \mathcal{M}_h,\ \forall K_1, K_2 \in \mathcal{L}_h(n),\ v_{K_1}(n) = v_{K_2}(n)\right\} \\ V_h &= \{v \in W_h \,/\, \forall n \in \mathcal{M}_h \cap \Gamma_D,\ \forall K \in \mathcal{L}_h(n),\ v_K(n) = 0\} \end{aligned}$$

For two adjacent elements K_1 and K_2, the functions v_{K_1} and v_{K_2} coincide at the nodes of $K' = K_1 \cap K_2$. In the case of $\mathbb{P}_1$-Lagrange finite elements, these objects coincide at the interfaces between two elements such that an object $v \in W_h$ is indeed a function in $\mathcal{C}^0(\Omega)$ and for any $K \in \mathcal{T}_h$, $v_K = v|_K$:

$$W_h = \{v \in \mathcal{C}^0(\Omega) \,/\, \forall K \in \mathcal{T}_h,\ v|_K \in \mathbb{P}_1(K)\}\,,$$

However, it is important to mention that in a general configuration, the objects of W_h are not necessarily functions even if it is the case for any $\mathbb{P}_k$-Lagrange finite elements, $k > 0$ (defined in next chapter).

A function $v \in W_h$ is then fully and uniquely determined by the knowledge of its values $\{v(n)\}_{n\in\mathcal{M}_h}$ at the nodes of $\mathcal{M}_h$, associated with $\mathcal{T}_h$, and called the *degrees of freedom* (d.o.f.) of W_h.

A crucial ingredient in the implementation of the finite element method is the transition from local numbering to global numbering. A function g^K is introduced to define this transition: the node n_i^K of K, $i = 1, ..., N^K$, is the node n_I of $\mathcal{M}_h$, $I \in \{1, ..., M_h\}$, identified by the relation $I = g^K(i)$. Moreover, the functions g^K, $K \in \mathcal{T}_h$, satisfy the property: $\forall n_I \in \mathcal{M}_h, \forall K_1 \in \mathcal{L}_h(n_I), \forall K_2 \in \mathcal{L}_h(n_I), \{(g^{K_1}(i) = g^{K_2}(j) = I) \Longleftrightarrow (n_i^{K_1} = n_j^{K_2} = n_I)\}$.

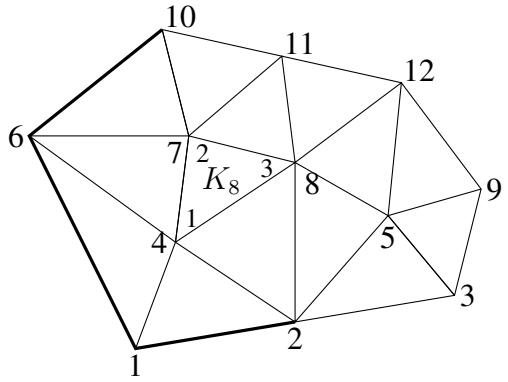

Figure 2.2: (*generated with Fig4TeX* [7]) Example of local and global numbering, $g^{K_8}(\{1,2,3\}) = \{4, 7, 8\}$

We now define the basis functions of the finite dimensional spaces. Let n_J be a node of $\mathcal{M}_h$, $J \in \{1, ..., M_h\}$, we denote by w_J the basis function related to this node:

$$w_J \in W_h \quad \text{and} \quad w_J(n_I) = \delta_I^J\,, I = 1, ..., M_h\,,$$

where δ_I^J is the Kronecker delta, equal to 1 if $I = J$ and 0 otherwise. One can check that

- ⋆ $\{w_J\}_{J=1,...,M_h}$ is a basis of W_h, and $\dim W_h = M_h = \mathrm{card}(\mathcal{M}_h)$;
- ⋆ $\mathrm{supp} w_J = \mathcal{L}_h(n_J)$;
- ⋆ $w_J|_K = w_j^K$ if there is $n_j^K \in \mathcal{N}^K$ such that $J = g^K(j)$, where w_j^K is the basis function associated with n_j^K on the element K. In our configuration of $\mathbb{P}_1$ Lagrange elements, $w_j^K = \lambda_j^K$.

2.5 Discrete problem and derivation of the linear system

The variational formulation (2.2) is replaced with the discrete variational form

$$\left| \begin{array}{ll} \text{Find } u_h \in W_h \text{ such that} & u_h|_{\Gamma_{D,h}} = u_{D_h}\,, \\ & a_h(u_h, v_h) = f_h(v_h)\,,\ \forall v_h \in V_h\,. \end{array} \right. \tag{2.4}$$

where

$$\begin{aligned} a_h(u_h, v_h) &= \int_{\Omega_h} \nabla u_h(x) \cdot \nabla v_h(x)\, dx + \int_{\Omega_h} u_h(x) v_h(x)\, dx\,, \\ f_h(v_h) &= \int_{\Omega_h} f_\Omega(x) v_h(x)\, dx + \int_{\Gamma_{N,h}} g_N(\gamma) v_h(\gamma)\, d\gamma\,, \\ & u_{D_h} \in W_h \text{ satisfying } u_{D_h}(n) = 0 \text{ for } n \in \mathcal{N}_h, \\ & \text{and } u_{D_h}(n) = u_D(n) \text{ for } n \in \mathcal{M}_h \setminus \mathcal{N}_h = \mathcal{M}_h \cap \Gamma_{D,h}\,. \end{aligned} \tag{2.5}$$

In order to simplify the notation, we consider that the nodes of $\mathcal{N}_h$ are numbered from 1 to N_h such that the nodes of $\mathcal{M}_h \cap \Gamma_D$ are numbered from $N_h + 1$ to M_h. Introducing the notation $u_J = u_h(n_J)$, $J = 1, ..., M_h$, we can define u_h thanks to the basis of W_h:

$$u_h = \sum_{J=1}^{M_h} u_J w_J = \sum_{J=1}^{N_h} u_J w_J + \sum_{J=N_h+1}^{M_h} u_D(n_J) w_J,$$

The equation of (2.5) written for all $v_h \in V_h$ can equivalently be considered for the basis $\{w_I\}_{I=1,...,N_h}$ of V_h. The finite dimensional variational form is then equivalent to the problem

$$\left| \begin{array}{ll} \text{Find } \{u_J\}_{J=1,...,N_h} \in \mathbb{R}^{N_h} & \text{such that } \forall I = 1, ..., N_h, \\ & \displaystyle\sum_{J=1}^{N_h} u_J a_h(w_J, w_I) = f_h(w_I) - \sum_{J=N_h+1}^{M_h} u_D(n_J) a_h(w_J, w_I) \end{array} \right. \tag{2.6}$$

which corresponds to the linear system

$$A_h X_h = F_h \tag{2.7}$$

with the matrix $A_h \in \mathbb{R}^{N_h \times N_h}$ and the right hand side $F_h \in \mathbb{R}^{N_h}$ defined by: $\forall I, J = 1, ..., N_h$

$$(A_h)_{IJ} = a_h(w_J, w_I)\,, \qquad (F_h)_I = f_h(w_I) - \sum_{\widetilde{I}=N_h+1}^{M_h} u_D(n_{\widetilde{I}}) a_h(w_{\widetilde{I}}, w_I) \tag{2.8}$$

and the unknown in $\mathbb{R}^{N_h}$: $X_h = (u_1, ..., u_{N_h})^{\mathrm{T}}$.

Due to the properties of the bilinear forms a and a_h and of the linear forms f and f_h of the variational formulations (2.2-2.3) and (2.4-2.5),

- $\star$ A_h is symmetric;
- $\star$ A_h is sparse:

  $$\mathrm{supp} w_I \cap \mathrm{supp} w_J = \emptyset \Longrightarrow (A_h)_{IJ} = 0\,,$$

  which means

  $$(A_h)_{IJ} \neq 0 \Longrightarrow \mathcal{L}_h(n_I) \cap \mathcal{L}_h(n_J) \neq \emptyset\,.$$

  The statement $(\mathcal{L}_h(n_I) \cap \mathcal{L}_h(n_J) \neq \emptyset)$ is equivalent to $(\exists K \in \mathcal{T}_h \ / \ n_I \in \mathcal{N}^K, n_J \in \mathcal{N}^K)$. Moreover, the matrix has a bandwidth the length of which can be determined by the formula

  $$L = \max_{K \in \mathcal{T}_h} \max_{n_I, n_J \in K} |I - J|\,.$$

  This indicates that the bandwidth of the matrix can be optimized by an appropriate numbering of the nodes of $\mathcal{T}_h$.

The first term of A_h, $\left(\int_{\Omega_h} \nabla w_J(x) \cdot \nabla w_I(x)\, dx \right)_{IJ}$, is known as the *stiffness matrix* and its second term, $\left(\int_{\Omega_h} w_J(x) w_I(x)\, dx \right)_{IJ}$, is known as the *mass matrix*.

2.6 Elementary matrices and right-hand-sides

In the finite element method, the matrix and right hand side of the linear system are computed through the consideration of *elementary matrices and right hand sides*. The elementary matrices and right hand sides are defined on the elements of the triangulation.

For $u_h, v_h \in W_h$, we introduce the notations: $\forall K \in \mathcal{T}_h$,

$$\begin{aligned} a_K(u_h, v_h) &= \int_K \nabla u_h(x) \cdot \nabla v_h(x)\, dx + \int_K u_h(x) v_h(x)\, dx\,, \\ f_K(v_h) &= \int_K f_\Omega(x) v_h(x)\, dx + \int_{K \cap \Gamma_{N,h}} g_N(\gamma) v_h(\gamma)\, d\gamma\,, \end{aligned} \tag{2.9}$$

Then, the matrix A_h and the right hand side F_h satisfy: $\forall I, J = 1, ..., N_h$,

$$(A_h)_{I,J} = \sum_{K \in \mathcal{L}_h(n_I) \cap \mathcal{L}_h(n_J)} a_K(w_J, w_I)\,, \qquad (F_h)_I = \sum_{K \in \mathcal{L}_h(n_I)} f_K(w_I)\,.$$

Combined with the definition of the function g^K, we get

$$\begin{aligned} (A_h)_{I,J} &= \sum_{K \in \mathcal{L}_h(n_I) \cap \mathcal{L}_h(n_J)} \sum_{i=1}^{N^K} \sum_{j=1}^{N^K} \delta^I_{g^K(i)} \delta^J_{g^K(j)} a_K(w_j^K, w_i^K)\,, \\ (F_h)_I &= \sum_{K \in \mathcal{L}_h(n_I)} \sum_{i=1}^{N^K} \delta^I_{g^K(i)} f_K(w_i^K) \\ &\quad - \sum_{\widetilde{I}=N_h+1}^{M_h} \sum_{K \in \mathcal{L}_h(n_I) \cap \mathcal{L}_h(n_{\widetilde{I}})} \sum_{i=1}^{N^K} \sum_{j=1}^{N^K} \delta^I_{g^K(i)} \delta^{\widetilde{I}}_{g^K(j)} u_D(n_{\widetilde{I}}) a_K(w_j^K, w_i^K)\,. \end{aligned} \tag{2.10}$$

where $\delta^I_{g^K(i)}$ is the Kronecker delta, equal to 1 if $I = g^K(i)$ and 0 otherwise. In this relation, indeed the summations "$\sum_{i=1}^{N^K} \sum_{j=1}^{N^K}$" involve exactly one term.

The relation (2.10) indicates that the matrix and the right hand side of the linear system to be solved are entirely determined by the quantities $a_K(w_j^K, w_i^K)$, $f_K(w_i^K)$, for $K \in \mathcal{T}_h$, $i, j = 1, ..., N^K$. These quantities correspond to the so-called *elementary matrices and right hand sides* associated with the triangulation $\mathcal{T}_h$:

Definition 2.1. *Let K be an element of the triangulation $\mathcal{T}_h$, the elementary matrix and right hand side associated with K are the matrix A^K and the vector F^K defined by: $\forall i, j = 1, ..., N^K$*

$$A^K_{ij} = a_K(w_j^K, w_i^K)\,, \qquad F^K_i = f_K(w_i^K)$$

Thanks to this definition, the calculation of A_h and F_h reduces to:

- ⋆ the calculation of the elementary matrices and right hand sides;
- ⋆ the assembling of A_h and F_h from the elementary matrices and right hand sides.

The calculation of the elementary objects is done by a generic strategy which is defined through the consideration of a reference element. Essentially, the calculation of A^K and F^K consists of the explicit definition of the basis functions on K and quadrature rules for the numerical evaluation of the integrals. These objects are given on a reference element and the numerical approximation of A^K and F^K is obtained from this reference element by a change of variable. The assembling of A_h and F_h is done carefully in a manner which minimizes the algorithm cost of the operation.

2.7 The reference element

The reference triangle $\widehat{K}$ is defined by the vertices $\widehat{a}_1 = (1, 0)$, $\widehat{a}_2 = (0, 1)$ and $\widehat{a}_3 = (0, 0)$. The barycentric coordinates of $\widehat{K}$ are defined by

$$\begin{cases} \widehat{\lambda}_1(\widehat{x}) &= \widehat{x}_1 \\ \widehat{\lambda}_2(\widehat{x}) &= \widehat{x}_2 \\ \widehat{\lambda}_3(\widehat{x}) &= 1 - \widehat{x}_1 - \widehat{x}_2 \end{cases}$$

Any triangle K of $\mathcal{T}_h$ can be defined from the reference element. If we denote by a^K_ℓ, $\ell = 1, 2, 3$, the vertices of K, there exists a unique $\mathbb{P}_1$ function F_K defined on $\widehat{K}$ such that $a^K_\ell = F_K(\widehat{a}_\ell)$. F_K satisfies the following properties:

- ⋆ F_K is invertible for any non degenerate triangle K;
- ⋆ $F_K(\widehat{x}) = x = \sum_{\ell=1}^3 a^K_\ell \widehat{\lambda}_\ell(\widehat{x})$;
- ⋆ $\exists B_K \in \mathbb{R}^{2\times 2}, b_K \in \mathbb{R}^2$ such that

$$F_K \begin{pmatrix} \widehat{x}_1 \\ \widehat{x}_2 \end{pmatrix} = B_K \begin{pmatrix} \widehat{x}_1 \\ \widehat{x}_2 \end{pmatrix} + b_K \tag{2.11}$$

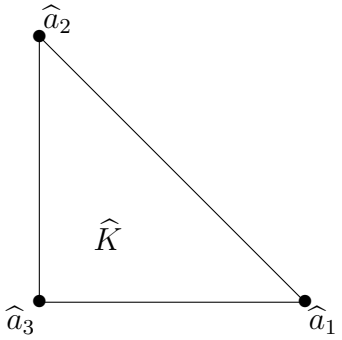

Figure 2.3: (*generated with Fig4TeX* [7]) The reference triangle

⋆ B_K and b_K are explicitly defined by

$$B_K = \begin{pmatrix} a_{1,1}^K - a_{3,1}^K & a_{2,1}^K - a_{3,1}^K \\ a_{1,2}^K - a_{3,2}^K & a_{2,2}^K - a_{3,2}^K \end{pmatrix} \qquad \text{and} \qquad b_K = \begin{pmatrix} a_{3,1}^K \\ a_{3,2}^K \end{pmatrix}.$$

where $\{a_{\ell,j}^K\}_{j=1,2}$, are the components of a_ℓ^K.

In the same way, we define the reference segment $\widehat{K}' = [0,1]$ which is used for the calculation of the integrals on edges of the triangles which interact with the Neumann condition. A generic edge K' is obtained from the reference element thanks to the function

$$F_{K'}(\widehat{\gamma}\,) = \gamma = (1-\widehat{\gamma}\,)a_1^{K'} + \widehat{\gamma} a_2^{K'}$$

where $\{a_\ell^{K'}\}_{\ell=1,2}$ are the extremities of the edge K'.

2.8 Elementary objects and reference element

Thanks to the reference element, the calculation of the elementary matrices and right hand sides is done through the change of variable defined by the function F_K:

$$\begin{aligned} a_K(w_j^K, w_i^K) &= \int_{\widehat{K}} \nabla w_j^K(F_K(\widehat{x})) \cdot \nabla w_i^K(F_K(\widehat{x}))\, |\mathrm{det} JF_K(\widehat{x})|\; d\widehat{x} \\ &\quad + \int_{\widehat{K}} w_j^K(F_K(\widehat{x})) w_i^K(F_K(\widehat{x}))\, |\mathrm{det} JF_K(\widehat{x})|\; d\widehat{x}\,, \\ f_K(w_i^K) &= \int_{\widehat{K}} f_\Omega(F_K(\widehat{x})) w_i^K(F_K(\widehat{x}))\, |\mathrm{det} JF_K(\widehat{x})|\; d\widehat{x} \\ &\quad + \sum_{K' \subset \partial K \cap \Gamma_{N,h}} \int_{\widehat{K}'} g_N(F_{K'}(\widehat{\gamma})) w_i^K(F_{K'}(\widehat{\gamma}))\, |\mathrm{det} JF_{K'}(\widehat{\gamma})|\; d\widehat{\gamma}\,, \end{aligned} \tag{2.12}$$

where $\mathrm{det} JF_K$ and $\mathrm{det} JF_{K'}$ are the determinants of the Jacobian matrices of F_K and $F_{K'}$.

We can check the remarquable properties:

⋆ For $x \in K$ and $\widehat{x} \in \widehat{K}$,

$$\widehat{w}_i(\widehat{x}) = w_i^K \circ F_K(\widehat{x})\,, \quad w_i^K(x) = \widehat{w}_i \circ F_K^{-1}(x)$$

⋆ For $\alpha = 1, 2$ and $x \in K$,

$$\begin{aligned}\frac{\partial w_i^K}{\partial x_\alpha}(x) &= \frac{\partial \widehat{w}_i}{\partial \widehat{x}_1} \circ F_K^{-1}(x)\, \frac{\partial (F_K^{-1})_1}{\partial x_\alpha}(x) + \frac{\partial \widehat{w}_i}{\partial \widehat{x}_2} \circ F_K^{-1}(x)\, \frac{\partial (F_K^{-1})_2}{\partial x_\alpha}(x) \\ &= \big(\widehat{\mathrm{grad}}\, \widehat{w}_i \circ F_K^{-1}(x)\big)^{\mathrm{T}} \big(JF_K^{-1}(x)\big)_{\alpha}\,.\end{aligned}$$

⋆ In the current case of $\mathbb{P}_1$ Lagrange finite elements, with $\mathbb{P}_1$ triangles, the Jacobian matrix JF_K is constant with respect to $\widehat{x}$ and equal to B_K.

These statements show that the elementary matrices and right hand sides are entirely determined by the knowledge of:

⋆ F_K for each element K (i.e. the vertices of K and the basis functions $\{\widehat{\lambda}_\ell\}_{\ell=1,...,3}$, on $\widehat{K}$);

⋆ the finite element basis functions on $\widehat{K}$, $\widehat{w}_\ell$, $\ell = 1, ..., 3$;

⋆ an integration strategy on the reference element $\widehat{K}$.

In the current configuration of $\mathbb{P}_1$ Lagrange finite elements combined with a $\mathbb{P}_1$ Lagrange interpolation of the geometry (F_K affine), we have $\widehat{\lambda}_\ell = \widehat{w}_\ell$ but this is not necessary.

2.9 Quadrature rules

It is usually impossible to evaluate exactly some of the integrals involved in the definition of the elementary objects $a_K(w_j^K, w_i^K)$ or $f_K(w_i^K)$. Numerically, the evaluation is operated using quadrature rules. The following nodal rules are standard rules for the reference triangle or reference segment:

⋆ Quadrature rule on triangle, exact for $\mathbb{P}_1$ polynomials:

$$\int_{\widehat{K}} \varphi(\widehat{x})\, d\widehat{x} \approx \frac{S_{\widehat{K}}}{3}\big(\varphi(\widehat{a}_1) + \varphi(\widehat{a}_2) + \varphi(\widehat{a}_3)\big).$$

⋆ Quadrature rule on triangle, exact for $\mathbb{P}_2$ polynomials:

$$\int_{\widehat{K}} \varphi(\widehat{x})\, d\widehat{x} \approx \frac{S_{\widehat{K}}}{3}\big(\varphi(\widehat{a}_4) + \varphi(\widehat{a}_5) + \varphi(\widehat{a}_6)\big).$$

⋆ Quadrature rule on segment, exact for $\mathbb{P}_1$ polynomials:

$$\int_{\widehat{K}'} \varphi(\widehat{\gamma})\, d\widehat{\gamma} \approx \frac{L_{\widehat{K}'}}{2}\big(\varphi(\widehat{a}'_1) + \varphi(\widehat{a}'_2)\big).$$

⋆ Quadrature rule on segment, exact for $\mathbb{P}_2$ polynomials:

$$\int_{\widehat{K}'} \varphi(\widehat{\gamma})\, d\widehat{\gamma} \approx \frac{L_{\widehat{K}'}}{6}\big(\varphi(\widehat{a}'_1) + 4\varphi(\frac{\widehat{a}'_1 + \widehat{a}'_2}{2}) + \varphi(\widehat{a}'_2)\big).$$

where $S_{\widehat{K}}$ and $L_{\widehat{K}'}$ are respectively the surface of $\widehat{K}$ ($1/2$ for the reference triangle) and length of $\widehat{K}'$ ($= 1$ for the reference segment).

2.10 Assembling procedure

The efficient procedure consists in considering the elements of $\mathcal{T}_h$ one by one: for each element, the contributions of its elementary objects are distributed to the corresponding matrix or right hand side element. The correspondence is specified by the function g^K.

A simple version of the procedure is written below, with no detail on the storage consideration. Indeed, due to the sparsity of the matrix, storage preoccupation is crucial and carefully answered in next chapter.

- Initialization of A_h and F_h to 0.
- For each element K:
  For each couple of nodes (i, j) of K:
    If $n_i^K \notin \Gamma_D$, $n_j^K \notin \Gamma_D$;
      $I = g^K(i)$; $J = g^K(j)$;
      $(A_h)_{IJ} \leftarrow (A_h)_{IJ} + A_{ij}^K$;
    End if;
    If $n_i^K \notin \Gamma_D$, $n_j^K \in \Gamma_D$;
      $I = g^K(i)$; $J = g^K(j)$;
      $(F_h)_I \leftarrow (F_h)_I - u_D(n_J) A_{ij}^K$;
    End if;
  End for (i, j);
  For each i, node of K such that $n_i^K \notin \Gamma_D$:
    $I = g^K(i)$;
    $(F_h)_I \leftarrow (F_h)_I + F_i^K$;
  End for i ;
End for K ;

Chapter 3

Some usual Finite Elements

3.1 Straight triangular Lagrange Finite Element

3.1.1 Definition of a finite element

Let us consider a convex closed subset K of $\mathbb{R}^2$ with non empty interior, a finite dimensional space P of functions $p : K \mapsto \mathbb{R}$, and Σ a set of N linear forms Φ_i defined on P.

Definition 3.1. *Σ is said P-unisolvent if Σ and P satisfy: for any set of N real numbers $\{\alpha_i\}_{i=1,...,N}$, there exists a unique $p \in P$ such that $\Phi_i(p) = \alpha_i$, for $i = 1, ..., N$.*

Proposition 3.2. *There is equivalence between the following assertions*

(i) *Σ is P-unisolvent.*

(ii) *"$\Phi_i(p) = 0$ for $i = 1, ..., N$" $\Longrightarrow$ "$p = 0$", and* $\dim P = N$.

(iii) *There exist N functions p_i, $i = 1, ..., N$, such that $\Phi_i(p_j) = \delta^i_j$ for $i, j = 1, ..., N$, and* $\dim P = N$.

Definition 3.3. *With the previous definition of K, P and Σ, the triple (K, P, Σ) is said a finite element if Σ is P-unisolvent. The linear forms Φ_i of the previous definition are called the degrees of freedom and the functions p_i satisfying (iii) are called the basis functions of the finite element.*

Usually, in 2D, K is the image by an invertible mapping F_K of the reference triangle or the reference square $]0, 1[\times]0, 1[$; P is a space of polynomials; the linear forms Φ_i are of one of these types:

1. $\Phi_i : p \mapsto \Phi_i(p) = p(n_i)$ where $n_i \in K$, case of Lagrange degrees of freedom or *nodal-value* degrees of freedom.

2. $\Phi_i : p \mapsto \Phi_i(p) = Dp(n_i)v_i$ where $n_i \in K$, $Dp(n_i)$ is the first derivative of p at n_i and v_i is a vector of $\mathbb{R}^2$, case of Hermite degrees of freedom of order 1.

3. $\Phi_i : p \mapsto \Phi_i(p) = D^k p(n_i)(v_i^1, v_i^2, ..., v_i^k)$ where $n_i \in K$, $D^k p(n_i)$ is the derivative of order k of p at n_i, and v_i^j, $j = 1, ..., k$, are vectors of $\mathbb{R}^2$, case of Hermite degrees of freedom of order k.

When all the linear forms of the family Σ are in the form 1, we say that (K, P, Σ) is a Lagrange finite element. For any of the forms 1, 2 or 3, the points n_i related to the degrees of freedom are called the *nodes* of the interpolation. All these finite elements are called nodal finite elements. In this chapter, we consider only Lagrange finite elements.

3.1.2 $\mathbb{P}_k$ Lagrange triangular finite element

The most common finite element is the *Lagrange triangular finite element* where the space P is the space $\mathbb{P}_k$ of polynomials of degree at most k with respect to all its variables considered all together.

We introduce the *straight* $\mathbb{P}_k$ *Lagrange triangle* which is characterized by the nodes given in Figures 3.1-3.2 for $k = 0, 1, 2, 3$. For a triangle $K$of vertices a_1^K, a_2^K, a_3^K, the set of nodes is:

$\star$ for $k = 0$:

$$\mathcal{L}_0^K = \left\{ n_1^{0,K} = \sum_{\ell=1}^{3} \lambda_\ell a_\ell^K ; \lambda_\ell = \frac{1}{3}, \ell = 1, 2, 3 \right\},$$

$\star$ for $k > 0$:

$$\mathcal{L}_k^K = \left\{ n_i^{k,K} \in K ; n_i^{k,K} = \sum_{\ell=1}^{3} \lambda_\ell a_\ell^K ; \lambda_\ell \in \{0, \frac{1}{k}, ..., \frac{k-1}{k}, 1\} ; \sum_{\ell=1}^{3} \lambda_\ell = 1 \right\}.$$

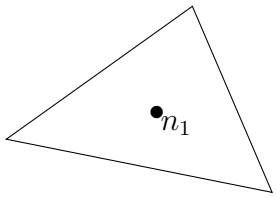

Figure 3.1: (*generated with Fig4TeX* [7]) Node for $k = 0$

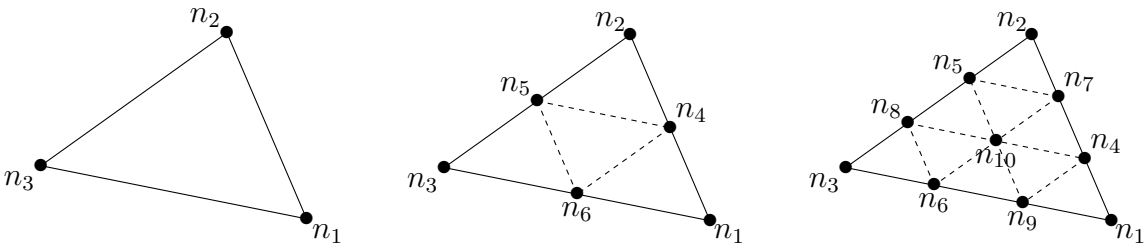

Figure 3.2: (*generated with Fig4TeX* [7]) Nodes repartition $k = 1, 2, 3$

These sets satisfy $\mathrm{card}(\mathcal{L}_k^K) = \binom{k+2}{k} = \dim \mathbb{P}_k$, moreover a polynomial $p \in \mathbb{P}_k$ is uniquely determined by its values $p(n_i^{k,K})$ at the nodes $n_i^{k,K} \in \mathcal{L}_k^K$. This means that Σ is $\mathbb{P}_k$-unisolvent with

$$\Sigma = \{\Phi_i : p \mapsto p(n_i^{k,K}), \text{ for } n_i^{k,K} \in \mathcal{L}_k^K\}.$$

In the case $k = 1$, the basis functions w_i^K are the barycentric coordinates

$$w_i^K = \lambda_i^K, \quad i = 1, 2, 3\,.$$

In the case $k = 2$, for $i = 1, 2, 3$,

$$\begin{aligned} w_i^K &= \lambda_i^K(2\lambda_i^K - 1)\,, \\ w_{i+3}^K &= 4\lambda_i^K \lambda_{1+i \bmod 3}^K\,. \end{aligned}$$

In the case $k = 3$, they are given by

$$\left.\begin{aligned} w_i^K &= \tfrac{1}{2}\,\lambda_i^K(3\lambda_i^K - 1)(3\lambda_i^K - 2) \\ w_{i+3}^K &= \tfrac{9}{2}\,\lambda_i^K \lambda_{1+i \bmod 3}^K(3\lambda_i^K - 1) \\ w_{i+6}^K &= \tfrac{9}{2}\,\lambda_i^K \lambda_{1+i \bmod 3}^K(3\lambda_i^K - 2) \end{aligned}\right\} \qquad i = 1, 2, 3\,,$$
$$w_{10}^K = 27\,\lambda_1^K \lambda_2^K \lambda_3^K\,.$$

3.1.3 Non conforming $\mathbb{P}_1$ finite element

The *non conforming* $\mathbb{P}_1$ *finite element* is a nodal-value finite element like the $\mathbb{P}_1$ Lagrange finite element with the difference that the set of nodes consists of the centers of the edges of the triangle K of vertices a_1^K, a_2^K, a_3^K:

$$\mathcal{L}_1^K = \left\{ n_i^{1,K} \in K; n_i^{1,K} = \frac{1}{2}(a_{1+i}^K + a_{2+i}^K),\ i = 1, 2, 3 \right\} \quad \text{with } a_4^K = a_1^K,\ a_5^K = a_2^K.$$

Due to this choice of the nodes, if we consider such a finite element for all the geometrical elements of a triangulation, the interpolation of a given function is not continuous at the interface between two adjacent elements. This justifies the terminology of *non conforming* finite element. In the opposite, the $\mathbb{P}_1$ finite element introduced in Section 3.1.2 is said H^1*-conforming finite element*.

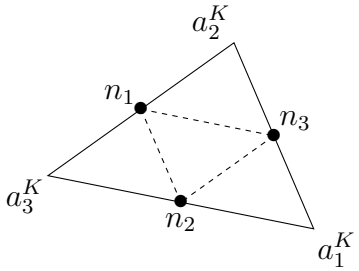

Figure 3.3: (*generated with Fig4TeX* [7]) Non conforming triangle of order 1

Similarly to the conforming $\mathbb{P}_1$ Lagrange finite element, we define the non conforming $\mathbb{P}_1$ finite element by choosing $P = \mathbb{P}_1$ and $\Sigma = \{\Phi_i(p) = p(n_i^K) \text{ for } n_i^K \in \mathcal{L}_1^K\}$. Its basis functions w_i^K are given by

$$w_i^K = 1 - 2\lambda_i^K\,, \quad i = 1, 2, 3\,.$$

3.2 Usual Finite Elements

3.2.1 Reference finite element

In a mesh, each element K is built from a reference element $\widehat{K}$ by application of an invertible function F_K. A finite element on K can be derived from a finite element on $\widehat{K}$. We introduce here the relation between a generic finite element and the reference finite element which simplifies the presentation of more elaborated finite elements such as curved finite elements.

Let us consider N points $\widehat{a}_i$ in $\mathbb{R}^2$ and denote by $\widehat{K}$ the closed convex envelope of $\{\widehat{a}_i\}_{i=1,...,N}$. We also consider $\widehat{P}$ a space of dimension N and $\widehat{\Sigma} = \{\widehat{\Phi}_i \,/\, \widehat{\Phi}_i(v) = v(\widehat{a}_i)\}_{i=1,...,N}$ such that $\widehat{\Sigma}$ is $\widehat{P}$-unisolvent. Then the following result holds: for any set of N distinct points a_i in $\mathbb{R}^2$, there is a unique mapping F

$$\widehat{x} \in \widehat{K} \mapsto F(\widehat{x}) = (F_1(\widehat{x}), F_2(\widehat{x})) \in \mathbb{R}^2$$

such that $F_\ell \in \widehat{P}$, for $\ell = 1, 2$ and $F(\widehat{a}_i) = a_i$, for $i = 1, ..., N$. This mapping is given by

$$F(\widehat{x}) = \sum_{i=1}^{N} a_i \, \widehat{\tau}_i(x)$$

where $\widehat{\tau}_i$ is the unique function of $\widehat{P}$ such that $\widehat{\tau}_i(\widehat{a}_j) = \delta^i_j$ (a basis function).

Considering an element K built from a reference element $\widehat{K}$ thanks to the invertible mapping F_K, a finite element on K can be derived from the finite element $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ with

$$P_K = \{p \,/\, p \circ F_K \in \widehat{P}\}, \quad \text{and} \quad \Sigma_K = \{\Phi \,/\, \exists\, \widehat{\Phi} \in \widehat{\Sigma} \ \text{ such that } \Phi(\widehat{p} \circ F_K^{-1}) = \widehat{\Phi}(\widehat{p}) \,, \forall \widehat{p} \in \widehat{P}\} \,.$$

(K, P_K, Σ_K) is a finite element. Its basis functions are $\{w_i^K\}$ deduced from the functions $\{\widehat{w}_i\}$:

$$\widehat{w}_i = w_i^K \circ F_K.$$

Thanks to this statement, any nodal-value polynomial finite element is entirely defined by the sole consideration of the nodes on the reference element. In the case of Lagrange finite elements, F_K is polynomial, and if F_K is affine, $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ and (K, P_K, Σ_K) are said affine equivalent.

3.2.2 Curved triangular Lagrange finite element

When the boundary of the geometrical domain Ω is curved, the consideration of curved finite elements can be useful. The *curved triangular Lagrange finite element* is derived from the definition of a suitable set of nodes on the reference element. In the case of $\mathbb{P}_2$-curved Lagrange finite elements, the sets of nodes on $\widehat{K}$ and K are given in Figure 3.4.

And the corresponding function F_K is defined by

$$F_K(\widehat{x}) = \sum_{i=1}^{6} a_i^K \, \widehat{\tau}_i(\widehat{x})$$

where $\{\widehat{\tau}_i\}_{i=1,...,6}$ are the basis functions of the reference triangular $\mathbb{P}_2$ Lagrange element.

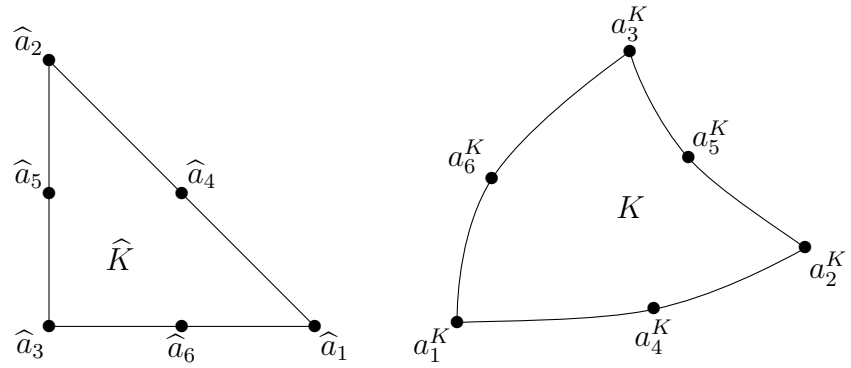

Figure 3.4: (*generated with Fig4TeX* [7]) Reference and generic triangular $\mathbb{P}_2$ Lagrange finite elements

3.2.3 Quadrangular finite elements

The quadrangular finite elements are built for quadrangles and based on $\mathbb{Q}_k$ the space of polynomials of degree at most k for each of the variables. We present here the quadrangular $\mathbb{Q}_k$ Lagrange finite element based on a lattice of the unit square $\widehat{K}$ of vertices $\widehat{a}_1 = (1,0)$, $\widehat{a}_2 = (1,1)$, $\widehat{a}_3 = (0,1)$, $\widehat{a}_4 = (0,0)$. The corresponding lattice is defined by:

$$\mathcal{L}_0^{\widehat{K}} = \left\{ n_1^{0,\widehat{K}} \, ; \, n_{1,1}^{0,\widehat{K}} = n_{1,2}^{0,\widehat{K}} = \frac{1}{2} \right\} \quad \text{for} \;\; k = 0,$$

$$\mathcal{L}_k^{\widehat{K}} = \left\{ n_i^{k,\widehat{K}} \, ; \, n_{i,j}^{k,\widehat{K}} \in \{0, \frac{1}{k}, ..., \frac{k-1}{k}, 1\}; j = 1,2 \right\} \quad \text{for} \;\; k > 0.$$

where $n_{i,j}^{k,\widehat{K}}$ is the j-th component of the node $n_i^{k,\widehat{K}}$.

The reference $\mathbb{Q}_1$ Lagrange finite element is defined by the four vertices of $\widehat{K}$. Its basis functions are $\widehat{\tau}_i \in \mathbb{Q}_1(\widehat{K})$:

$$\begin{aligned} \widehat{\tau}_1(\widehat{x}) &= \widehat{x}_1(1-\widehat{x}_2) \\ \widehat{\tau}_2(\widehat{x}) &= \widehat{x}_1\widehat{x}_2 \\ \widehat{\tau}_3(\widehat{x}) &= (1-\widehat{x}_1)\widehat{x}_2 \\ \widehat{\tau}_4(\widehat{x}) &= (1-\widehat{x}_1)(1-\widehat{x}_2) \end{aligned}$$

A quadrangular $\mathbb{Q}_1$ element K of vertices $a_1^K, a_2^K, a_3^K, a_4^K$ can be curved if F_K is not affine. F_K is defined by

$$F_K \begin{pmatrix} \widehat{x}_1 \\ \widehat{x}_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \sum_{\ell=1}^{4} a_\ell^K \widehat{\tau}_\ell(\widehat{x})$$

and the nodes are given in Figure 3.5.

The case of $\mathbb{Q}_2$ elements is illustrated in Figure 3.6

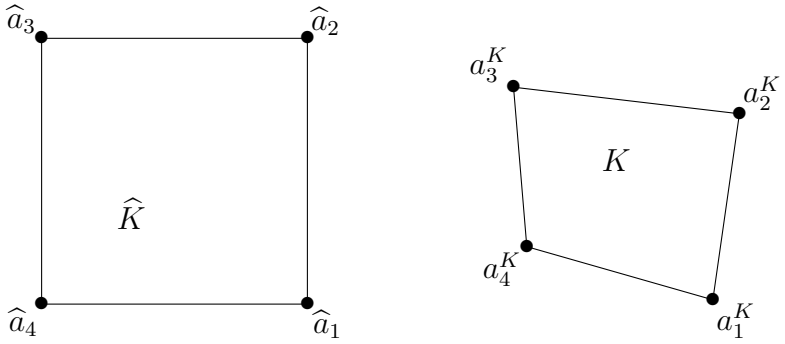

Figure 3.5: (*generated with Fig4TeX* [7]) Reference and generic quadrangular $\mathbb{Q}_1$ Lagrange finite elements

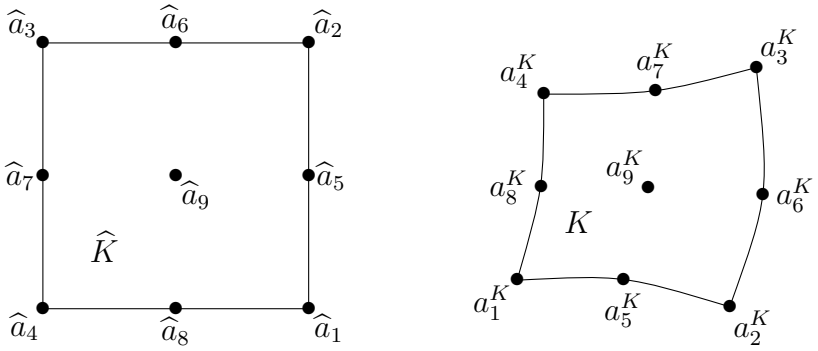

Figure 3.6: (*generated with Fig4TeX* [7]) Reference and generic quadrangular $\mathbb{Q}_2$ Lagrange finite elements

3.3 Geometrical Finite Element and interpolation Finite Element

This chapter is dedicated to the definition of classical finite elements. In the previous section, we saw that the finite element on any element K of a mesh is entirely determined by its definition on the reference element and the transformation F_K. This transformation itself is defined thanks to the basis functions of a geometrical finite element. When the finite element chosen for the geometrical approximation (definition of F_K) and the finite element used for the interpolation of the unknow are identical, the finite-element approximation is said isoparametric. This definition will be more explicitely expressed in Subsection 4.2.3.

Chapter 4

Generic configuration and FE details

4.1 Integrands of the generic 2D configuration

We first remind that the generic problem is posed on an open bounded polygonal convex domain $\Omega \subset \mathbb{R}^2$, of boundary $\partial\Omega = \Gamma_D \cup \Gamma_N$ such that $\Gamma_D \cap \Gamma_N$ is empty or reduced to a finite number of points of $\partial\Omega$ (see Fig. 1.2).

The generic PDE is defined by the relations (1.1) and Equation (1.2). We remind that the generic Green's formula associated with the generic configuration is given by Equation (1.5) which results from the initial Green's formula (1.3).

Using these tools, the variational formulation has been written in Subsection 1.2.3 and is given by the relations (1.7-1.8) and Equation (1.9). The problem has a unique solution $u \in H^2(\Omega)$ if the following assumptions are satisfied

- $a_{\alpha\beta} \in H^1(\Omega)$; $a_{00} \in L^2(\Omega)$; $f_\Omega \in L^2(\Omega)$; $b_N, g_N \in H^{\frac{1}{2}}(\Gamma_N)$; $u_D \in H^{\frac{3}{2}}(\Gamma_D)$;
- ellipticity of the operator of order 2:

$$\exists c > 0 \text{ such that } \forall \xi \in \mathbb{R}^2\,,\ \forall x \in \Omega\,,\ \sum_{\alpha,\beta=1}^{2} a_{\alpha\beta}(x)\xi_\alpha\xi_\beta \geq c \sum_{\alpha=1}^{2} \xi_\alpha^2\,;$$

- $\Gamma_D \neq \emptyset$, or $\exists c > 0$ such that $a_{00}(x) \geq c > 0$ for all $x \in \Omega$.

The variational formulation (1.9) involves the following integrands

$$\int_\Omega a_{\alpha\beta}(x)\frac{\partial u}{\partial x_\alpha}(x)\frac{\partial v}{\partial x_\beta}(x)\,dx\,,\quad \int_\Omega a_{00}(x)u(x)v(x)\,dx\,,\quad \int_{\Gamma_N} b_N(\gamma)u(\gamma)v(\gamma)\,d\gamma\,,$$
$$\int_\Omega f_\Omega(x)v(x)\,dx\,,\qquad \int_{\Gamma_N} g_N(\gamma)v(\gamma)\,d\gamma\,. \tag{4.1}$$

After decomposition over the elements of the mesh, these integrands are calculated thanks to a change of variable which expresses them as an assembling of integrals over a reference element. These integrals are evaluated by numerical integration schemes such as nodal quadrature rules.

We refer to Chapter 3 for a detailed presentation of 2D finite elements. Section 4.2 is devoted to the decomposition into elementary integrands, the change of variable to the reference element, and quadrature rules.

4.2 Elementary matrices and right hand sides

In this section, we consider carefully the derivation and calculation of the elementary matrices and right hand sides related to the following integrands introduced in Section 4.1

$$\int_\Omega a_{\alpha\beta}(x)\frac{\partial u}{\partial x_\alpha}(x)\frac{\partial v}{\partial x_\beta}(x)\,dx\ , \qquad \int_\Omega a_{00}(x)u(x)v(x)\,dx\ , \qquad \int_{\Gamma_N} b_N(\gamma)u(\gamma)v(\gamma)\,d\gamma\ ,$$

$$\int_\Omega f_\Omega(x)v(x)\,dx\ , \qquad \int_{\Gamma_N} g_N(\gamma)v(\gamma)\,d\gamma\ .$$

A nodal-value finite element approximation of basis functions $\{w_I\}_{I=1,...,M_h}$ leads to the consideration of the following terms

$$\int_{\Omega_h} a_{\alpha\beta}(x)\frac{\partial w_J}{\partial x_\alpha}(x)\frac{\partial w_I}{\partial x_\beta}(x)\,dx\ , \quad \int_{\Omega_h} a_{00}(x)w_J(x)w_I(x)\,dx\ , \quad \int_{\Gamma_{N,h}} b_N(\gamma)w_J(\gamma)w_I(\gamma)\,d\gamma\ ,$$

$$\int_{\Omega_h} f_\Omega(x)w_I(x)\,dx\ , \qquad \int_{\Gamma_{N,h}} g_N(\gamma)w_I(\gamma)\,d\gamma\ .$$

4.2.1 Definition of the elementary matrices and right hand sides

The consideration of the elementary matrices and right hand sides requires that we remind some notations:

- ⋆ g^K the mapping from local numbering on K to global numbering on $\mathcal{T}_h$;
- ⋆ $\mathcal{L}_h(n_I)$ is the set of elements which contributes to the support of the basis function w_I, when n_I is a node of $\mathcal{T}_h$;
- ⋆ $\{w_i^K,\ i = 1, ..., N^K\}$ the set of local basis functions on K, where N^K is the number of nodes on K.

We then have the following relations

$$\begin{aligned}
\int_{\Omega_h} a_{\alpha\beta}(x)\frac{\partial w_J}{\partial x_\alpha}(x)\frac{\partial w_I}{\partial x_\beta}(x)\,dx &= \sum_{K\in\mathcal{L}_h(n_J)\cap\mathcal{L}_h(n_I)} \int_K a_{\alpha\beta}(x)\frac{\partial w_j^K}{\partial x_\alpha}(x)\frac{\partial w_i^K}{\partial x_\beta}(x)\,dx\,,\\
\int_{\Omega_h} a_{00}(x)w_J(x)w_I(x)\,dx &= \sum_{K\in\mathcal{L}_h(n_J)\cap\mathcal{L}_h(n_I)} \int_K a_{00}(x)w_j^K(x)w_i^K(x)\,dx\,,\\
\int_{\Gamma_{N,h}} b_N(\gamma)w_J(\gamma)w_I(\gamma)\,d\gamma &= \sum_{K\in\mathcal{L}_h(n_J)\cap\mathcal{L}_h(n_I)}\ \sum_{K'\in\partial K\cap\Gamma_{N,h}} \int_{K'} b_N(\gamma)w_j^K(\gamma)w_i^K(\gamma)\,d\gamma\,,\\
\int_{\Omega_h} f_\Omega(x)w_I(x)\,dx &= \sum_{K\in\mathcal{L}_h(n_J)\cap\mathcal{L}_h(n_I)} \int_K f_\Omega(x)w_i^K(x)\,dx\,,\\
\int_{\Gamma_{N,h}} g_N(\gamma)w_I(\gamma)\,d\gamma &= \sum_{K\in\mathcal{L}_h(n_J)\cap\mathcal{L}_h(n_I)}\ \sum_{K'\in\partial K\cap\Gamma_{N,h}} \int_{K'} g_N(\gamma)w_i^K(\gamma)\,d\gamma\,.
\end{aligned} \tag{4.2}$$

where i and j satisfy $I = g^K(i)$ and $J = g^K(j)$. This indicates that the matrix and the right hand side of the system obtained by application of the finite element method to the problem

(1.2) introduced in Section 1.1.2 are entirely determined by the assembling of the elementary matrices $A^{K,1}$, $A^{K,2}$, $A^{K',N}$ given by

$$\begin{aligned} A_{ij}^{K,1} &= \int_K a_{\alpha\beta}(x) \frac{\partial w_j^K}{\partial x_\alpha}(x) \frac{\partial w_i^K}{\partial x_\beta}(x)\, dx\,, \\ A_{ij}^{K,2} &= \int_K a_{00}(x) w_j^K(x) w_i^K(x)\, dx\,, \\ A_{ij}^{K',N} &= \int_{K'} b_N(\gamma) w_j^K(\gamma) w_i^K(\gamma)\, d\gamma\,, \end{aligned} \tag{4.3}$$

and the elementary right hand sides $B^{K,\Omega}$, $B^{K',N}$ defined by

$$\begin{aligned} B_i^{K,\Omega} &= \int_K f_\Omega(x) w_i^K(x)\, dx\,, \\ B_i^{K',N} &= \int_{K'} g_N(\gamma) w_i^K(\gamma)\, d\gamma\,. \end{aligned} \tag{4.4}$$

We refer to Section 4.3 for details on the assembling procedure.

4.2.2 Integrals over the reference element

The calculation of the integrals in 4.2.1 is done using a change of variable defined by the mappings F_K from $\widehat{K}$ to K and $F_{K'}$ from $\widehat{K}'$ to an edge K' of K. Since we consider here the case of nodal finite elements, the reference element $\widehat{K}$ is characterized by its nodes $\widehat{a}_\ell$, $\ell = 1, ..., \widetilde{N}^K$. We first remind the definition of F_K and $F_{K'}$, and significant properties:

- F_K (resp. $F_{K'}$) is invertible for any non degenerate element K (resp. K');
- $F_K(\widehat{x}) = x = \sum_{\ell=1}^{\widetilde{N}^K} a_\ell^K \widehat{\tau}_\ell(\widehat{x})$;
- $\widehat{w}_i(\widehat{x}) = w_i^K \circ F_K(\widehat{x})$; $w_i^K(x) = \widehat{w}_i \circ F_K^{-1}(x)$;
- for $\alpha = 1, 2$,

$$\begin{aligned} \frac{\partial w_i^K}{\partial x_\alpha}(x) &= \frac{\partial \widehat{w}_i}{\partial \widehat{x}_1} \circ F_K^{-1}(x) \frac{\partial \widehat{x}_1}{\partial x_\alpha}(x) + \frac{\partial \widehat{w}_i}{\partial \widehat{x}_2} \circ F_K^{-1}(x) \frac{\partial \widehat{x}_2}{\partial x_\alpha}(x) \\ &= \big(\widehat{\mathrm{grad}}\, \widehat{w}_i \circ F_K^{-1}(x)\big)^{\mathrm{T}} \big(JF_K^{-1}(x)\big)_\alpha\,. \end{aligned}$$

The change of variable leads to

$$\begin{aligned} A_{ij}^{K,1} &= \int_{\widehat{K}} a_{\alpha\beta}(F_K(\widehat{x})) \frac{\partial w_j^K}{\partial x_\alpha}(F_K(\widehat{x})) \frac{\partial w_i^K}{\partial x_\beta}(F_K(\widehat{x}))\ |\mathrm{det} JF_K(\widehat{x})|\ d\widehat{x}\,, \\ A_{ij}^{K,2} &= \int_{\widehat{K}} a_{00}(F_K(\widehat{x}))\, w_j^K(F_K(\widehat{x}))\, w_i^K(F_K(\widehat{x}))\ |\mathrm{det} JF_K(\widehat{x})|\ d\widehat{x}\,, \\ A_{ij}^{K',N} &= \int_{\widehat{K}'} b_N(F_{K'}(\widehat{\gamma}))\, w_j^K(F_{K'}(\widehat{\gamma}))\, w_i^K(F_{K'}(\widehat{\gamma}))\ |\mathrm{det} JF_{K'}(\widehat{\gamma})|\ d\widehat{\gamma}\,, \end{aligned} \tag{4.5}$$

and the elementary right hand sides $B^{K,\Omega}$, $B^{K',N}$ defined by

$$\begin{aligned} B_i^{K,\Omega} &= \int_{\widehat{K}} f_\Omega(F_K(\widehat{x}))\, w_i^K(F_K(\widehat{x}))\ |\mathrm{det} JF_K(\widehat{x})|\ d\widehat{x}\,, \\ B_i^{K',N} &= \int_{\widehat{K'}} g_N(F_{K'}(\widehat{\gamma}))\, w_i^K(F_{K'}(\widehat{\gamma}))\ |\mathrm{det} JF_{K'}(\widehat{\gamma})|\ d\widehat{\gamma}\,. \end{aligned} \tag{4.6}$$

4.2.3 Organization of the calculation

The calculation is then reduced to the determination of an integration scheme, and for each point $\widehat{x} \in \widehat{K}$ or $\widehat{\gamma} \in \widehat{K'}$ involved in the integration scheme, the evaluation of the quantities:

- ⋆ for $i = 1, ..., N^K$, $\alpha = 1, 2$, $w_i^K(F_K(\widehat{x}))$, $\dfrac{\partial w_i^K}{\partial x_\alpha}(F_K(\widehat{x}))$;
- ⋆ $a_{\alpha\beta}(F_K(\widehat{x}))$, $a_{00}(F_K(\widehat{x}))$, $b_N(F_{K'}(\widehat{\gamma}))$, $f_\Omega(F_K(\widehat{x}))$, $g_N(F_{K'}(\widehat{\gamma}))$;
- ⋆ $JF_K(\widehat{x})$, $JF_{K'}(\widehat{\gamma})$;

We can find two classes of calculus:

- ⋆ Some of them are related to the interpolation of the unknown, the finite dimensional spaces W_h and V_h: $\widehat{\mathrm{grad}\, w_i}$ and $\widehat{w}_i$, for $i = 1, ..., N^K$.
- ⋆ The other class concerns the calculus related to the interpolation of the geometry involving the function F_K and requiring:
  - ◇ the shape functions $\{\widehat{\tau}_i\}_{i=1,...,\widetilde{N}^K}$ on $\widehat{K}$, and the corresponding shape functions on the one dimensional reference element: $\{\widehat{\tau}'_i\}_{i=1,...,\widetilde{N}^{K'}}$;
  - ◇ the nodes $\{a_i^K\}_{i=1,...,\widetilde{N}^K}$ of K and the nodes $\{a_i^{K'}\}_{i=1,...,\widetilde{N}^{K'}}$ of the edge K';
  - ◇ the mapping $F_K : \widehat{K} \mapsto K \subset \mathbb{R}^2$ defined by

$$F_K(\widehat{x}) = \sum_{i=1}^{\widetilde{N}^K} a_i^K \widehat{\tau}_i(\widehat{x})\,,$$

    which leads to

$$\frac{\partial F_K}{\partial \widehat{x}_\alpha}(\widehat{x}) = \sum_{i=1}^{\widetilde{N}^K} a_i^K \frac{\partial \widehat{\tau}_i}{\partial \widehat{x}_\alpha}(\widehat{x})\,,$$

    for the definition of the Jacobian $JF_K(\widehat{x})$;
  - ◇ the 1D mapping $F_{K'} : [0, 1] \mapsto K' \subset \mathbb{R}^2$,

$$F_{K'}(\widehat{\gamma}) = \sum_{i=1}^{\widetilde{N}^{K'}} a_i^{K'} \widehat{\tau}'_i(\widehat{\gamma})\,.$$

    and the corresponding Jacobian

$$JF_{K'}(\widehat{\gamma}) = \Big(\Big(\frac{dF_{K',1}}{d\widehat{\gamma}}(\widehat{\gamma})\Big)^2 + \Big(\frac{dF_{K',2}}{d\widehat{\gamma}}(\widehat{\gamma})\Big)^2\Big)^{\frac{1}{2}}\,.$$

To summarize, the calculation of the elementary matrix and right hand side on a given element K requires the knowledge of

- ⋆ an integration scheme;
- ⋆ a procedure which defines the geometrical finite element;
- ⋆ a procedure which defines the interpolation finite element;
- ⋆ the coordinates of the points $\{a_i^K\}_{1\leq i\leq \widetilde{N}^K}$;
- ⋆ the table which indicates if the edges are subset of Γ_N or Γ_D or not;
- ⋆ when required, the list of nodes and Dirichlet values for the nodes on Γ_D.

Usually the basis functions $\{\widehat{w}_i\}_{i=1,\ldots,N^K}$ and the shape functions $\{\widehat{\tau}_i\}_{i=1,\ldots,\widetilde{N}^K}$ coincide. In this case, the finite element approximation is said *isoparametric*.

4.2.4 Quadrature rules

The evaluation of the integrals requires numerical strategies, either for the exact calculation of integrals involving only polynomials of low degree, or for the evaluation of integrals where the primitive of the integrand is not known. We list here some quadrature rules which are commonly used in finite element libraries (for other examples, see [15, 16]):

$$\int_{\widehat{K}} \varphi(\widehat{x})\, d\widehat{x} \approx \frac{S_{\widehat{K}}}{3}(\varphi(\widehat{a}_1)+\varphi(\widehat{a}_2)+\varphi(\widehat{a}_3)) \quad \text{for the triangle, exact for } \mathbb{P}_1,$$

$$\int_{\widehat{K}} \varphi(\widehat{x})\, d\widehat{x} \approx \frac{S_{\widehat{K}}}{3}(\varphi(\widehat{a}_4)+\varphi(\widehat{a}_5)+\varphi(\widehat{a}_6)) \quad \text{for the triangle, exact for } \mathbb{P}_2,$$

$$\int_{\widehat{K}} \varphi(\widehat{x})\, d\widehat{x} \approx \frac{S_{\widehat{K}}}{4}(\varphi(\widehat{a}_1)+\varphi(\widehat{a}_2)+\varphi(\widehat{a}_3)+\varphi(\widehat{a}_4)) \quad \text{for the square, exact for } \mathbb{Q}_1,$$

$$\int_{\widehat{K}} \varphi(\widehat{x})\, d\widehat{x} \approx \frac{S_{\widehat{K}}}{36}((\varphi(\widehat{a}_1)+\varphi(\widehat{a}_2)+\varphi(\widehat{a}_3)+\varphi(\widehat{a}_4)) \quad \text{for the square, exact for } \mathbb{Q}_2$$
$$+4(\varphi(\widehat{a}_5)+\varphi(\widehat{a}_6)+\varphi(\widehat{a}_7)+\varphi(\widehat{a}_8)) + 16\varphi(\widehat{a}_9)),$$

$$\int_{\widehat{K}'} \varphi(\widehat{\gamma})\, d\widehat{\gamma} \approx \frac{L_{\widehat{K}'}}{2}(\varphi(\widehat{a}'_1)+\varphi(\widehat{a}'_2)) \quad \text{for the segment, exact for } \mathbb{P}_1, (\widehat{a}'_1=0, \widehat{a}'_2=1),$$

$$\int_{\widehat{K}'} \varphi(\widehat{\gamma})\, d\widehat{\gamma} \approx \frac{L_{\widehat{K}'}}{6}(\varphi(\widehat{a}'_1)+4\varphi(\tfrac{\widehat{a}'_1+\widehat{a}'_2}{2})+\varphi(\widehat{a}'_2)) \quad \text{for the segment, exact for } \mathbb{P}_2.$$

where $S_{\widehat{K}}$ and $L_{\widehat{K}'}$ are respectively the surface of $\widehat{K}$ ($1/2$ for the reference triangle, 1 for the reference square) and length of $\widehat{K}'$ (1 for the reference segment).

Equivalent rules could be used for the integration over a generic element K. This can be done for validation.

4.3 Assembling and storages

The assembling procedure in Section 2.10 is optimized in terms of calculation time. In order to make the algorithm optimal in terms of memory requirements, the storage of the matrix should be done carefully. As already observed, the matrix defined by (2.8) is sparse. The storage of the matrix should be adapted to sparse structures. Several storages were developed for sparse matrices such as the *compressed sparse row storage* (CSR) or the *compressed sparse column storage* (CSC). Essentially, a sparse storage contains the non-zero values of the matrix in a vector table and indicates the corresponding row and column numbers for each stored value thanks to suitable tables. The memory requirement is then linearly proportional to the number of non-zero entries of the matrix. In this section, we present sparse storages used in the library MÉLINA++ [11]:

- ⋆ The *ordered sparse storage* consists in the storage of the non-zero entries of the matrix in a vector table, ordered with respect to their locations in the matrix considered row by row. This storage is suitable for some algebraic operations but not accessible directly by the assembling procedure.

- ⋆ The *non-ordered sparse storage* consists in the storage of the non-zero entries of the matrix in a vector table, ordered with respect to their first occurrence in the assembling procedure. This storage is accessible by the assembling procedure.

In the definition of these storages, we call "non-zero entry" any entry $(A_h)_{IJ}$ satisfying $\mathcal{L}_h(n_I) \cap \mathcal{L}_h(n_J) \neq \emptyset$. Such entries are generally non-zero valued but some of them may numerically vanish for specific geometrical configurations. The storages in MÉLINA++ take also in consideration the symmetry of the matrix. For the sake of simplicity, we consider here only the case of symmetric matrices.

4.3.1 Ordered sparse storage

The ordered sparse storage as implemented in MÉLINA++ consists in the consideration of the following tables:

- ⋆ `Diag`(ℓ), $\ell = 1, ..., N_h$: values of the diagonal entries of A_h.

- ⋆ `LMat`(ℓ), $\ell = 1, ..., N_{L_h} + 1$: values of non-zero entries of L_h (= strict lower part of A_h), where N_{L_h} is the number of non-zero entries of L_h. `LMat`(ℓ) is the value of the ℓ-th non-zero entry in terms of its location in L_h ordered row by row. The entry numbered $N_{L_h} + 1$ is a fictitious entry considered to be the only entry in row N_h of L_h (justified just below).

- ⋆ `ColInd`(ℓ), $\ell = 1, ..., N_{L_h}$: column number of the ℓ-th non-zero entry of `LMat`.

- ⋆ `FirstAdLi`(i), $i = 1, ..., N_h$: location (or "address") in `LMat` of the first non-zero entry of the i-th row of L_h, part of the $(i+1)$-th row of A_h. If the i-th row of L_h does not have non-zero entry, we set `FirstAdLi`(i) = `FirstAdLi`($i+1$). The last value `FirstAdLi`(N_h) is for the fictitious entry.

Thanks to this definition, and the fictitious entry, the number of non-zero entries of the i-th row of L_h (which is the $(i+1)$-th row of A_h) is precisely `FirstAdLi`($i+1$) - `FirstAdLi`(i).

Example 1

The matrix

$$A_h = \begin{pmatrix} X_1 & & & & \\ A & X_2 & & \text{Sym} & \\ B & 0 & X_3 & & \\ 0 & C & D & X_4 & \\ 0 & E & 0 & F & X_5 \end{pmatrix}$$

leads to the following storage:

$$\begin{aligned} &\texttt{Diag} = [X_1\ X_2\ X_3\ X_4\ X_5] \\ &\texttt{LMat} = [\overset{1}{A}\ \overset{2}{B}\ \overset{3}{C}\ \overset{4}{D}\ \overset{5}{E}\ \overset{6}{F}\ \overset{7}{\Diamond}] \\ &\texttt{FirstAdLi} = [1\ 2\ 3\ 5\ 7] \\ &\texttt{ColInd} = [1\ 1\ 2\ 3\ 2\ 4\ ] \end{aligned}$$

where $\Diamond$ represents the fictitious entry.

Example 2

The matrix

$$A_h = \begin{pmatrix} X_1 & & & & \\ 0 & X_2 & & \text{Sym} & \\ A & 0 & X_3 & & \\ 0 & 0 & 0 & X_4 & \\ B & 0 & C & 0 & X_5 \end{pmatrix}$$

leads to the following storage:

$$\begin{aligned} &\texttt{Diag} = [X_1\ X_2\ X_3\ X_4\ X_5] \\ &\texttt{LMat} = [\overset{1}{A}\ \overset{2}{B}\ \overset{3}{C}\ \overset{4}{\Diamond}] \\ &\texttt{FirstAdLi} = [1\ 1\ 2\ 2\ 4\ ] \\ &\texttt{ColInd} = [1\ 1\ 3] \end{aligned}$$

4.3.2 Non-ordered sparse storage

The non-ordered sparse storage as implemented in MÉLINA++ consists in the consideration of the following tables:

- ⋆ `Diag`(ℓ), $\ell = 1, ..., N_h$: values of the diagonal entries of A_h.
- ⋆ `LMat`(ℓ), $\ell = 1, ..., N_{L_h}$: values of non-zero entries of L_h (= strict lower part of A_h), where N_{L_h} is the number of non-zero entries of L_h. `LMat`(ℓ) is the value of the ℓ-th non-zero entry of L_h ordered with respect to its first occurrence in the assembling procedure.

- ⋆ `FollowingAdLi`(ℓ), $\ell = 1, ..., N_{L_h}$: location (or "address") in `LMat` of the next non-zero entry on the same row of L_h after the ℓ-th entry of `LMat`, ordered with respect to their first occurrences in the assembling procedure. This table is initialized to 0 at the beginning of the assembling procedure.
- ⋆ `ColInd`(ℓ), $\ell = 1, ..., N_{L_h}$: column number of the ℓ-th non-zero entry of `LMat`.
- ⋆ `FirstAdLi`(i), $i = 1, ..., N_h - 1$: location (or "address") in `LMat` of the first non-zero entry of the i-th row of L_h, part of the $(i+1)$-th row of A_h. This table is initialized to 0 at the beginning of the assembling procedure.

The non-zero entries of a row of L_h are ordered with respect to their first occurrences in the assembling algorithm. The transcription to the ordered storage is done row by row in two steps per row: for each row,

- ⋆ get all the entries of the row (corresponding parts of `LMat` and `ColInd`) thanks to tables `FirstAdLi` and `FollowingAdLi`.
- ⋆ sort the entries of the row thanks to the corresponding part of table `ColInd`.

Warning: In MÉLINA++, this storage transcription includes a treatment of the Dirichlet condition because the elimination of the Dirichlet degrees of freedom in the interpolation of the unknown is not done like we considered in this document.

4.4 Resolution

The application of a finite element method in the context of a variational formulation satisfying the hypotheses of Lax-Milgram theorem and symmetric property of the bilinear form induces a linear system the matrix of which is sparse, symmetric and positive-definite. The Cholesky method is then one of the most suitable resolution procedures. The method factorizes the matrix of the system into triangular matrices which makes very easy the resolution. Moreover, the Cholesky method preserves the profile of the matrix. In this section, we explain a strategy for the application of Cholesky method in the context of sparse matrices, including a useful factorization procedure. To this aim, we define the skyline storage which is adapted for preserving the profile of a matrix. Let us first remind that we call "non-zero entry" of the finite element matrix A_h any entry $(A_h)_{IJ}$ satisfying $\mathcal{L}_h(n_I) \cap \mathcal{L}_h(n_J) \neq \emptyset$.

4.4.1 Profile of a matrix and skyline storage

The profile of the matrix is defined by the consideration of the entries of each row from the first non-zero entry to the last non-zero entry. We present here a *row-oriented skyline storage* for symmetric matrices as used in MÉLINA++: for each row, only the entries from the first non-zero subdiagonal to the diagonal entry are stored:

- ⋆ `DiagP`(ℓ), $\ell = 1, ..., N_h$: values of the diagonal entries of A_h.
- ⋆ `LMatP`(ℓ), $\ell = 1, ..., N_{L_h} + 1$: values of the entries in the profile of L_h (= strict lower part of A_h), where N_{L_h} is the number of entries in the profile of L_h. `LMatP`(ℓ) is the value of the ℓ-th entry in terms of its location in the profile of L_h ordered row by row. The entry numbered $N_{L_h} + 1$ is a fictitious entry considered to be the only entry in row N_h of L_h (justified just below).

⋆ `Profile`(i), $i = 1, ..., N_h$: location in `LMatP` of the first non-zero entry of the i-th row of L_h, part of the $(i+1)$-th row of A_h. If the i-th row of L_h does not have non-zero entry, we set `Profile`(i) = `Profile`($i+1$).

Thanks to this definition and the fictitious entry, the number of stored entries of the i-th row of L_h (the $(i+1)$-th row of A_h) is precisely `Profile`($i+1$) - `Profile`(i), and

$$\begin{aligned} & \texttt{LMatP}[\texttt{Profile}(i) : \texttt{Profile}(i+1) - 1] \\ = \; & A_h\,(i+1\,,\; i+1-\texttt{Profile}(i+1)+\texttt{Profile}(i) : i)\;. \end{aligned}$$

The transcription from the ordered sparse storage to the skyline storage is done by insertion, for each row, of the zero entries located between the first non-zero entry of L_h and the diagonal entry of A_h of the row.

4.4.2 Cholesky factorization and resolution

The Cholesky method is justified by the theorem

Theorem 4.1 (Cholesky factorization)**.** *If A is a real symmetric positive-definite matrix, there exists a unique real lower triangular matrix L satisfying*

$$\begin{aligned} & A = LL^{\mathrm{T}}\,, \\ & \forall i\,,\; L_{ii} > 0\,. \end{aligned}$$

Using this theorem and the definition of L, the resolution of the system $AX = B$ is given by the resolution of systems

$$LY = B \quad \text{and} \quad L^{\mathrm{T}}X = Y\,.$$

The triangular systems are solved efficiently by usual upper/lower-triangular system-resolution procedures.

Last but not least, we describe the factorization as implemented in MÉLINA++. The matrix L is calculated row by row by identification in the relation $A = LL^{\mathrm{T}}$:

⋆ The first row is explicitly defined by: $L_{11}\,L_{11} = A_{11}$.

⋆ The second row satisfies:

$$\begin{cases} L_{21}L_{11} = A_{21}\,, \\ L_{22}L_{22} + L_{21}L_{21} = A_{22}\,, \end{cases}$$

which gives:

$$\begin{cases} L_{21} = A_{21}L_{11}^{-1}\,, \\ L_{22} = \sqrt{A_{22} - L_{21}L_{21}}\,. \end{cases}$$

⋆ An identification of the i-th row, $i \geq 2$, of the relation $A = LL^{\mathrm{T}}$ leads to:

$$\begin{cases} \begin{pmatrix} L_{1,1} & 0 & \cdots & 0 \\ L_{2,1} & L_{2,2} & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ L_{i-1,1} & L_{i-1,2} & \cdots & L_{i-1,i-1} \end{pmatrix} \begin{pmatrix} L_{i,1} \\ L_{i,2} \\ \vdots \\ L_{i,i-1} \end{pmatrix} = \begin{pmatrix} A_{i,1} \\ A_{i,2} \\ \vdots \\ A_{i,i-1} \end{pmatrix}\,, \\ \\ L_{ii} = \sqrt{A_{ii} - \displaystyle\sum_{k=1}^{i-1} L_{i,k}L_{i,k}}\,. \end{cases}$$

which indicates that the calculation of the subdiagonal part of the i-th row of L can be done by the application of an usual lower-triangular system-resolution procedure to a system the matrix of which is the already calculated part of L.

The hypotheses of Theorem 4.1 are sufficient and necessary to ensure the positivity of the arguments of the square-root in this algorithm. Indeed, there is equivalence between the existence of the Cholesky factorization and the realizability of the factorization according to this algorithm.

4.5 Validation

The error analysis of the finite element method requires the consideration of

- ⋆ the error induced by the geometrical approximation (choice of the triangulation),
- ⋆ the error induced by the interpolation of the unknown (choice of V_h).

Both the geometrical and the unknown interpolation error estimations are based on standard interpolation error estimations. The convergence analysis of the finite element method is given by the consideration of Céa's Lemma which derives from the properties of the variational formulation:

Lemma 4.2 (Céa's Lemma)**.** *If u and u_h are respectively the solutions to the variational formulation (2.2-2.3) and the finite dimensional formulation (2.4-2.5), and if the hypotheses of Lax-Milgram Lemma are satisfied, then*

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

where M is the constant of continuity of the bilinear form a and α is its constant of coercivity.

The finite element method constructs V_h according to a size parameter h which makes the distance between V_h (resp. Ω_h) and V (resp. Ω) converging towards 0 when h tends to 0 itself, and approximates u by its interpolate u_h in V_h in a sense that the distance between u and u_h is the distance between u and V_h. For polynomial finite elements, this gives error estimates which are derived from interpolation theory combined with Céa's Lemma. In our configuration of $\mathbb{P}_1$ Lagrange finite elements, under some regularity properties, if $u \in H^2(\Omega)$, we have for example:

$$\|u - u_h\|_{H^1(\Omega)} \leq Ch\,|u|_{H^2} \qquad \text{and} \qquad \|u - u_h\|_{L^2(\Omega)} \leq Ch^2\,|u|_{H^2} \tag{4.7}$$

where $|\cdot|_{H^2}$ is the semi-norm on $H^2(\Omega)$:

$$|u|_{H^m} = \left(\sum_{|\alpha|=m} \|\partial^\alpha u\|_{L^2}^2 \right)^{1/2}.$$

The validation of a finite element code is done by considering error estimates like the ones given by relations (4.7). We choose a problem such that the exact solution is known analytically. Then, the plot of the relative error between the exact solution and the finite element solution with respect to the characteristic size h of the triangulation $\mathcal{T}_h$ should illustrate the theoretical error behavior. To observe this numerical result, we plot the error in a log-log scale such that the

curve should be linear with a slope corresponding to the order of the convergence with respect to h.

Some simple test-cases are defined by problems of the form:

$$\text{find } u \text{ such that } \begin{cases} -\Delta u + u &= f_\Omega \quad \text{in } \Omega\,, \\ \dfrac{\partial u}{\partial \nu} &= g_N \quad \text{on } \partial\Omega\,, \end{cases}$$

or

$$\text{find } u \text{ such that } \begin{cases} -\Delta u &= f_\Omega \quad \text{in } \Omega\,, \\ u &= u_D \quad \text{on } \partial\Omega\,, \end{cases}$$

where Ω, f_Ω, g_N and u_D are determined such that the problem is satisfied by a chosen function u. For example, considering $\Omega =]0,1[\times]0,1[$, one can determine easily f_Ω and g_N

- in the first problem, such that $u(x) = \cos(\pi x_1)\cos(\pi x_2)$ is solution,
- in the second problem, such that $u(x) = \sin(\pi x_1)\sin(\pi x_2)$ is solution.

Chapter 5

Extensions – to go further

5.1 Three dimensional finite elements

The three dimensional finite element method is mathematically derived in the same way as the two dimensional method. The method is then described by the three dimensional reference finite elements. The usual reference elements are the tetrahedron, the hexahedron, the prism, the pyramid. For example, Figure 5.1 shows nodal-value three dimensional reference elements.

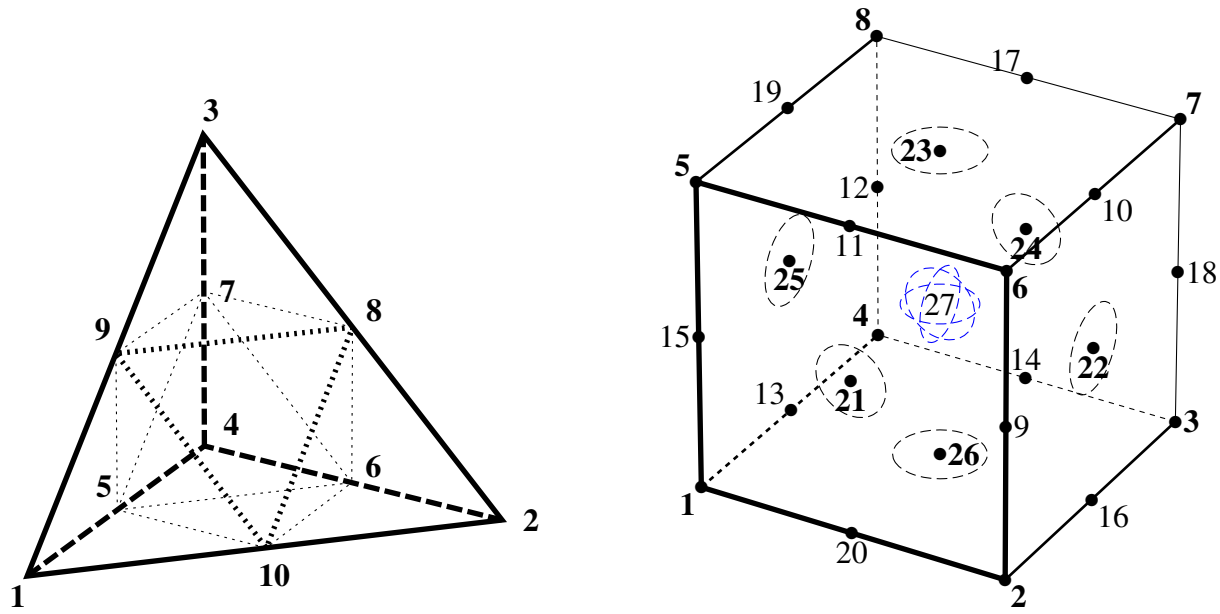

Figure 5.1: (*generated with Fig4TeX* [7]) Reference tetrahedron and hexahedron of degree 2

The three dimensional case is usually rather difficult to deal with due to computational cost issues:

- ⋆ For real-life geometries, the construction of the mesh is often a real challenge and requires expertise in the use of mesh generators.
- ⋆ The optimization of the bandwidth of the matrix with respect to the numbering of the nodes is limited compared to the two dimensional case.

⋆ The size of the finite element system can be very large according to the problem to be solved (several millions).

5.2 Time-dependent problems

The finite element method is also considered for the resolution of evolution problems such as parabolic or hyperbolic differential equations. These equations involve the time variable. For some configurations, one can try to solve the problem by application of the finite element in the space+time variables. However, the mathematical justification usually imposes another strategy: the finite element is used only for the discretization of the space variable and the time variable is discretized thanks to a finite difference scheme.

Let us consider the generic problem

$$\left\{\begin{array}{l} D_t\,u(x,t) + D_x\,u(x,t) = f(x,t) \qquad \text{in } \Omega\,, \\ \text{boundary conditions on } \Gamma_N \text{ and } \Gamma_D \text{ at any time } t\,, \\ \text{initial conditions at time } t = t_0\,. \end{array}\right. \tag{5.1}$$

where D_t (resp. D_x) is a differential operator with respect to the time variable (resp. space variable). A finite difference scheme in the time variable consists of the choice of a time-step size and a discrete approximation of D_t. For example, $\frac{\partial u}{\partial t}$ can be approximated by $\frac{u^n - u^{n-1}}{\Delta t}$ where Δt is the time-step size and u^{n-1}, u^n are the iterates of the approximation of u with respect to t at two consecutive discrete times t^{n-1} and t^n. The generic operator $D_t\,u(x,t)$ is approximated by a function of $p+1$ consecutive iterates

$$D_t\,u(x,t_n) \approx F(u^{n-p}, \cdots, u^{n-1}, u^n)(x)$$

for the determination of u^n, considering u^j already known for $j = n-p, ..., n-1$. At each time step, the problem (5.1) is reduced to the problem:

$$\text{Find } u^n \text{ such that } \left\{\begin{array}{l} F(u^{n-p}, \cdots, u^{n-1}, u^n)(x) + D_x\,u^n(x,t_n) = f(x,t_n) \qquad \text{in } \Omega\,, \\ \text{boundary conditions on } \Gamma_N \text{ and } \Gamma_D \text{ at time } t_n\,. \end{array}\right. \tag{5.2}$$

Under some mathematical hypotheses on the initial problem and the chosen finite difference scheme, the problem 5.2 can be solved by application of the finite element method at each time step. For more details and specificity, we refer the reader to [8].

5.3 Case of integral equations

The document is devoted to the application of the finite element method to partial differential equations but there is another family of equations which also uses finite element procedures for the numerical resolution, the integral equations. The essential difference is the operator which is not a differential operator but an integral operator.

An example is the so-called *trace of the single-layer potential* in scattering theory:

$$u \mapsto \left(x \mapsto \int_{\partial\Omega} G_k(x,y) u(y) d\gamma(y) \right)\,, \tag{5.3}$$

where k is a physical parameter (the wavenumber in wave propagation problems), $G_k(x, y)$ is a so-called *Green's kernel* which is singular at $x = y$. Such operators are often used in scattering theory where the computational domain is infinite, as the exterior of a bounded obstacle. The integral equations are a strategy which consists in replacing the exterior problem by the determination of a representative quantity defined on the boundary of the obstacle which is solution to a problem defined on the boundary of the obstacle and involving integral operators.

These operators are mathematical tools significantly different from partial differential operators in terms of

- ⋆ mathematical studies: the study of integral operators is rather difficult and the mathematical justification of the wellposedness of integral equations involves mathematical ingredients which are rather different from the ones used for partial differential equations. Some reference books on integral equations are [6], [3], [12]. In three dimensions, the operator (5.3) is a surface integral operator and its definition requires the consideration of trace Hilbert spaces.

- ⋆ mathematical properties: the integral operators are global and singular.

The mathematical properties impact drastically the application of finite element method:

- ⋆ The discretization of (5.3) leads to the matrix defined by

$$A_{ij} = \int_{\partial\Omega}\int_{\partial\Omega} G_k(x, y)w_j(y)w_i(x)d\gamma(y)d\gamma(x)\,, \tag{5.4}$$

  which is dense since the basis function and the test function are not considered at the same integration variable. To deal with this specificity, several strategies are developed: preconditioning and fast methods. The fast methods aim to somehow sparsify the matrix. An algebraic sparsifying strategy would be unefficient. These problems require the consideration of analytical properties of the Green's kernel and lot of efforts for the generation of efficient preconditioners or fast methods.

- ⋆ The singularity of the Green's kernel induces a singularity of the integral. According to the strength of the singularity, various strategies are developed for the evaluation of the singular integrals. Numerical strategies are explained in [14]. A rather recent analytical strategy is presented in [9].

5.4 Edge finite elements

This document presented polynomial nodal-value finite elements including the H^1-conforming $\mathbb{P}_1$ Lagrange finite element in the case of scalar problems. In some applications, the modelization leads to a mixed problem which involves a scalar unknown and a vectorial unknown and motivated the development of mixed finite elements. Such problems are usually affected by a constraint on the vectorial unknown which has to be of constrained divergence. To preserve the divergence and avoid unexpected numerical effects like porious modes, the nodal finite elements are not sufficient and other classes of finite elements were introduced in literature. To ensure divergence preserving, the mixed finite elements are based on the consideration of the so-called *edge finite elements* for the discretization of the vectorial unknown. The edge elements are known as the *Raviart-Thomas / Nédélec* elements and are based on moments calculation

instead of nodal evaluation. The edge elements are associated to the edges of the mesh. For example, in the case of 3D Maxwell equations, the degrees of freedom of the edge element on a tetrahedron K are the moments M_{e} defined by

$$M_{\mathrm{e}}(v) = \int_{\mathrm{e}} v \cdot t$$

for any edge e of K, where t is an oriented tangential vector on e, and for v in the space

$$\mathcal{R}_1 = \{v \in (\mathbb{P}_1(K))^3 \,/\, v(x) = a + bx,\ a, b \in \mathbb{R}^3\}\,.$$

This finite element is considered for the discretization of the functional space

$$V_h = \{v \in H(\mathrm{curl}, \Omega) \,/\, v\,|_K \in \mathcal{R}_1(K), \forall K \in \mathcal{T}_h\}$$

with $H(\mathrm{curl}, \Omega) = \{v \in (L^2(\Omega))^3 \,/\, \mathrm{curl}\, v \in (L^2(\Omega))^3\}$. The finite element is said $H(curl)$-*conforming element*.

Edge elements were initially introduced for Navier-Stokes equations with the so-called $H(\mathrm{div}, \Omega)$-*conforming elements*. For more details on these elements, we refer for example to [2, 4, 1].

5.5 Resolution

In terms of resolution of the linear system, the Cholesky method was introduced as the most suitable one in the case of the application of the finite element method to coercive symmetric linear variational formulation related to a partial differential equation but the finite element method is also applied to configurations which do not satisfy these properties all together:

- ⋆ In the case of a non-symmetric configuration, the coercive property of the bilinear form ensures the invertibility of the matrix of the system and the Gauss method works in any case.

- ⋆ When the variational formulation is derived from an integral equation, the matrix of the corresponding discrete system is dense and a direct method is often unusable due to the large size of the matrix. In such cases where the storage of the full matrix is impossible, the matrix is described by algebraic operations such as matrix-vector products. Theses operations are defined thanks to fast methods which reduces the computational cost in terms of time and memory requirements. With the described algebraic operations, the methods which can be used for the resolution of the system are not numerous. For example, in the case of a symmetric variational formulation obtained from an integral equation, an isoparametric finite element method leads to a symmetric positive definite matrix which permits the use of a Conjugate Gradient method. In the case of integral equations, the matrix is also known to be badly conditioned. The iterative method should be implemented with a preconditioner which is generally derived from the analytical properties of the corresponding integral operators.

- ⋆ In the case of large matrices where the direct solvers are not possible, if the symmetric property is not ensured, one can consider the GMRES method which is a generalization of the methods based on Krylov spaces such as the Conjugate Gradient method.

⋆ In the case of eigenvalue problems, the choice of the solver is also a crucial question. The library ARPACK++[1] answers the question offering an implementation of the "Implicit Restarted Arnoldi Method" (IRAM), which combines Arnoldi factorizations with an implicitly shifted QR method. The library considers carefully the hypotheses of the matrices involved in the eigenvalue problem, usually a generalized eigenvalue problem.

[1] http://www.ime.unicamp.br/∼chico/arpack++/

Bibliography

[1] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*, volume 15. Springer-Verlag, 1991.

[2] Philippe G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.

[3] David Colton and Rainer Kress. *Inverse acoustic and electromagnetic scattering theory*, volume 93 of *Applied Mathematical Sciences*. Springer-Verlag, Berlin, second edition, 1998.

[4] Vivette Girault and Pierre-Arnaud Raviart. *Finite element methods for Navier-Stokes equations*, volume 5 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1986. Theory and algorithms.

[5] Nicolas Kielbasiewicz and Eric Lunéville. http://uma.ensta-paristech.fr/soft/XLiFE++/. XLiFE++ documentation, ENSTA, in progress.

[6] Rainer Kress. *Linear integral equations*, volume 82 of *Applied Mathematical Sciences*. Springer, New York, third edition, 2014.

[7] Yvon Lafranche. https://perso.univ-rennes1.fr/yvon.lafranche/fig4tex/index.html. Fig4TeX documentation, Université Rennes 1, last update: november 2011.

[8] Stig Larsson and Vidar Thomée. *Partial differential equations with numerical methods*, volume 45 of *Texts in Applied Mathematics*. Springer-Verlag, Berlin, 2009. Paperback reprint of the 2003 edition.

[9] Marc Lenoir and Nicolas Salles. Evaluation of 3-D singular and nearly singular integrals in Galerkin BEM for thin layers. *SIAM J. Sci. Comput.*, 34(6):A3057–A3078, 2012.

[10] Brigitte Lucquin and Olivier Pironneau. *Introduction to scientific computing*. John Wiley & Sons, Ltd., Chichester, 1998. Translated from the French by Michel Kern.

[11] Daniel Martin, Eric Darrigrand, and Yvon Lafranche. https://anum-maths.univ-rennes1.fr/melina/melina++_distrib/. MÉLINA++ documentation, Université Rennes 1, last update: april 2015.

[12] Jean-Claude Nédélec. *Acoustic and electromagnetic equations*, volume 144 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2001.

[13] Walter Rudin. *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, second edition, 1991.

[14] Stefan A. Sauter and Christoph Schwab. *Boundary element methods*, volume 39 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2011. Translated and expanded from the 2004 German original.

[15] A. H. Stroud. *Approximate calculation of multiple integrals*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1971. Prentice-Hall Series in Automatic Computation.

[16] A. H. Stroud and Don Secrest. *Gaussian quadrature formulas*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1966.