Advanced Finite Element Methods

Eric Sonnendrücker, Ahmed Ratnani
Max-Planck-Institut für Plasmaphysik
und
Zentrum Mathematik, TU München

LECTURE NOTES
WINTERSEMESTER 2015/2016

February 5, 2016
Contents

1 The classical finite element method (FEM) 2
  1.1 Principle of the method ........................................... 2
  1.2 Basic properties of the Sobolev spaces $H^m(\Omega)$, $H(\text{div}, \Omega)$ and $H(\text{div}, \Omega)$ 5
    1.2.1 The Sobolev space $H^m$ ........................................ 6
    1.2.2 The Sobolev space $H(\text{div}, \Omega)$ ........................ 6
    1.2.3 The Sobolev space $H(\text{curl}, \Omega)$ ..................... 6
  1.3 The variational (or weak) form of a boundary value problem 6
    1.3.1 Case of homogeneous Dirichlet boundary conditions .......... 7
    1.3.2 Case of non homogeneous Dirichlet boundary conditions ..... 8
    1.3.3 Case of Neumann boundary conditions ........................ 8
  1.4 Solution of the the variational formulation ......................... 9
  1.5 Description of a Finite Element .................................. 10
    1.5.1 Formal definition of a Finite Element ....................... 10
    1.5.2 $C^0$ Lagrange Finite Elements .............................. 11
    1.5.3 $C^1$ Hermite Finite Elements .............................. 13
    1.5.4 $H(\text{div}, \Omega)$ conforming Finite Elements .......... 15
    1.5.5 $H(\text{curl}, \Omega)$ conforming Finite Elements .......... 18
  1.6 Change of local basis ............................................. 19
  1.7 Construction of the Finite Element approximation space .......... 21
    1.7.1 The mapping ..................................................... 21
    1.7.2 Bases of tangent and cotangent spaces in curvilinear coordinates 22
    1.7.3 $H^1$ conforming Finite Elements ........................... 23
    1.7.4 $H(\text{curl}, \Omega)$ conforming Finite Elements .......... 25
    1.7.5 $H(\text{div}, \Omega)$ conforming Finite Elements .......... 29
  1.8 Convergence of the Finite Element method ........................ 33

2 Non conforming and mixed Finite Element methods 36
  2.1 Variational crimes ............................................... 36
  2.2 Non conforming $P_k$ finite elements on triangles ................ 38
  2.3 Weak enforcement of Dirichlet boundary conditions ............... 40
  2.4 Discontinuous Galerkin methods .................................. 42
  2.5 Saddle-point problems ............................................ 43
    2.5.1 The theoretical framework ................................. 43
    2.5.2 Mixed formulation of the Poisson problem .................. 44
Chapter 1

The classical finite element method (FEM)

1.1 Principle of the method

For solving a problem on a computer that can only store a finite amount of information a discrete form of the problem is needed. In the Finite Difference method one simply computes an approximation of the solution at a finite number of grid points. In the Finite Element method, which is mathematically more involved, the idea is to look for the solution in a finite dimensional vector space, i.e. for some well chosen vector space $V_h$, with basis $(\varphi_i)_{0 \leq i \leq N-1}$, the approximate solution has the form

$$u_h(x) = \sum_{i=0}^{N-1} u_i \varphi_i(x).$$

The basis being given, the approximate solution $u_h$ is fully determined by its coefficients $u_i$ in this basis, which need not be values of $u_h$ at some points in the computational domain, but can in some cases.

The question now becomes how to choose $V_h$ and determine the coefficients $u_i$ such that $u_h$ is a good approximation of the solution $u$ of the original problem. Let us consider as a first example the Poisson problem with homogeneous Dirichlet boundary conditions:

$$-\Delta u = f, \quad \text{in } \Omega, \quad u = 0, \quad \text{on } \partial\Omega. \quad (1.1)$$

The first idea, introduced by Ritz in his thesis in Göttingen in 1902, was to transform the boundary problem into an equivalent minimisation problem. Indeed, via the Dirichlet principle (1.1) is equivalent to the minimisation problem

$$\min_{u \in H^1_0(\Omega)} \left( \frac{1}{2} \int_{\Omega} |\nabla u(x)|^2 \, dx - \int_{\Omega} f(x)u(x) \, dx \right). \quad (1.2)$$

We shall need the following Hilbert spaces, defined for a domain $\Omega \in \mathbb{R}^d$

$$H^1(\Omega) = \{ u \in L^2(\Omega), \ \nabla u \in (L^2(\Omega))^d \}, \quad H^1_0(\Omega) = \{ u \in H^1(\Omega), \ u = 0 \text{ on } \partial\Omega \}.$$ 

The scalar product associated to these Hilbert spaces is

$$\langle u, v \rangle_1 = \int_{\Omega} \nabla u(x) \cdot \nabla v(x) \, dx + \int_{\Omega} u(x)v(x) \, dx.$$
Moreover, we denote by
\[ ||u||_{1,\Omega} = \int_{\Omega} |\nabla u(x)|^2 \, dx + \int_{\Omega} u(x)^2 \, dx, \quad |u|_{1,\Omega} = \int_{\Omega} |\nabla u(x)|^2 \, dx, \quad ||u||_{0,\Omega} = \int_{\Omega} u(x)^2 \, dx, \]
the norm in \( H^1(\Omega) \), the semi-norm in \( H^1(\Omega) \) and the norm in \( L^2(\Omega) \). Then, the original problem being transformed into a minimisation problem, it becomes quite natural to look for an approximation in a finite dimensional subspace of the function space in which the minimisation problem is posed (in our case \( H^1_0(\Omega) \)), which means that the minimisation is performed by considering only minima in a finite dimensional subspace. Then if the form of finite dimensional space is chosen such that any function of the original space can be approximated to any given tolerance, by a function of the approximation space, we should be able to get a good approximation. Ritz who was actually looking at solutions for the bilaplacian equation, chose as basis functions for \( V_k \) a finite number of eigenfunctions of his operator.

The standard method to solve a minimisation problem with a cost functional \( J \) defined on a Hilbert space \( V \), of the form
\[ \min_{u \in V} J[u], \]
is to solve the associated Euler equation \( J'[u] = 0 \) obtained by computing the Fréchet derivative of the functional that we want to minimise. Note that the Fréchet derivative gives a rigorous definition of the functional derivative used in physics for functions that are in a Banach (including Hilbert) space. Consider a functional \( J \) from a Hilbert space \( V \) into \( \mathbb{R} \). Its Fréchet derivative \( J' \), assuming it exists, is a linear form on \( V \), which means that it maps any function from \( V \) to a scalar. It can be computed using the Gâteaux formula:
\[ J'[u](v) = \lim_{\varepsilon \to 0} \frac{J[u + \varepsilon v] - J[u]}{\varepsilon}. \tag{1.3} \]
Let us apply this formula to our problem for which
\[ J[u] = \frac{1}{2} \int_{\Omega} |
abla u(x)|^2 \, dx - \int_{\Omega} f(x)u(x) \, dx. \]
We have for any \( v \in V = H^1_0(\Omega) \)
\[ J[u + \varepsilon v] = \frac{1}{2} \int_{\Omega} |
abla u(x) + \varepsilon v(x)|^2 \, dx - \int_{\Omega} f(x)(u(x) + \varepsilon v(x)) \, dx \]
\[ = \frac{1}{2} \left( \int_{\Omega} |
abla u(x)|^2 \, dx + 2\varepsilon \int_{\Omega} \nabla u(x) \cdot \nabla v(x) \, dx + \varepsilon^2 \int_{\Omega} |
abla v(x)|^2 \, dx \right) \]
\[ - \int_{\Omega} f(x)u(x) \, dx - \varepsilon \int_{\Omega} f(x)v(x) \, dx. \]
From which we deduce, using the Gâteaux formula (1.3) that
\[ J'[u](v) = \int_{\Omega} \nabla u(x) \cdot \nabla v(x) \, dx - \int_{\Omega} f(x)v(x) \, dx. \]
Note that \( J'[u] \) being a linear form on \( V \) is defined by applying it to some vector \( v \in V \). Finally the solution of our minimisation problem (1.2), is a solution of the Euler equation \( J'[u] = 0 \) or equivalently \( J'[u](v) = 0 \) for all \( v \in V \), which reads
\[ \int_{\Omega} \nabla u(x) \cdot \nabla v(x) \, dx = \int_{\Omega} f(x)v(x) \, dx \quad \forall v \in H^1_0(\Omega). \tag{1.4} \]
This is also what is called the variational formulation, or the weak formulation of the original boundary value problem (1.1). Note that this variational formulation expresses in some sense the orthogonality of the residual to the space in which the solution is sought. This is more general than Euler equations of minimisation problems as noticed by Galerkin and has a wide range of applications. One can even extend this concept by making the residual orthogonal to a different function space, than the one in which the solution lives. Such methods are called Petrov-Galerkin methods and are beyond the scope of this lecture.

So the principle of the Galerkin (or Ritz-Galerkin) method is to look for a solution in a finite dimensional subspace $V_h \subset V$ of the original space and to use the same variational formulation (1.4) as the one defining the exact solution, with test functions also in $V_h$ to characterise the solution. What remains to be done now is to choose $V_h$ with good approximation properties. As we will see later, the stability of the Galerkin method follows directly from the well-posedness of the variational problem (1.4).

The finite dimensional space $V_h$ is in general defined by its basis functions. For those, Ritz used eigenfunctions of the problem. But those are in general cumbersome to compute. Galerkin proposed to use general classes of simple functions, trigonometric functions or polynomials, that are know to be able to approximate any continuous function with a finite number of basis functions. Trigonometric polynomials which are linked to Fourier series are very good in periodic domains, with a few simple extensions. Polynomials enjoy more widespread applications, however to get a good conditioning of the linear system that is obtained at the end, care needs to be taken in the choice of the basis functions. The monomial basis $(1, x, x^2, \ldots)$ has very bad properties. Best approximations are provided by the orthogonal Legendre polynomials or by the Chebyshev polynomials which are used in practice. Note that all the basis functions we have mentioned up to now have a global support in the computational domain and thus lead to full matrices in the linear system, which can be computationally expensive. Methods using such bases are actually not known as Finite Element methods but rather as spectral methods.

Another ingredient is needed to define what is known as Finite Element methods. This was introduced by Courant in 1943 [4] and consists in using basis functions with a small support in the computational domain, so that its product with other basis functions vanishes for most of the other basis functions leading to a very sparse matrix, i.e. with very few non zero terms, in the linear system, which can be solved very efficiently on a computer. To this aim, the computational domain is decomposed into small elements, in general triangles or quads in 2D and the basis functions are chosen to be relatively low order polynomials, on each of these elements. Convergence being achieved by taking smaller elements like the cells in the Finite Difference method. In 1D a finite element mesh will look like a finite difference mesh. An example of an unstructured Finite Element mesh in 2D is displayed in Figure 1.1, which shows the great flexibility in particular to handle complicated boundaries with finite elements, which finite differences do not provide. This is a key to its very wide usage.

The article by Gander and Wanner [7] provides a clear and well documented overview of the historical developments of the Finite Element method. For more technical historical developments of the Finite Difference and Finite Element methods one can also consult [10].

In summary, the finite element method consists in looking for a solution of a variational problem like (1.4), in a finite dimensional subspace $V_h$ of the space $V$ where
the exact solution is defined. The space $V_h$ is characterised by a basis $(\varphi_1, \ldots, \varphi_N)$ so that finding the solution of the variational problem amounts to solving a linear system. Indeed, express the trial function $u_h$ and the test function $\tilde{u}_h$ on this basis:

$$ u_h(x) = \sum_{j=1}^{N} u_j \varphi_j(x), \quad \tilde{u}_h(x) = \sum_{i=1}^{N} \tilde{u}_i \varphi_i(x), $$

and plug these expressions in the variational problem (1.4). This yields

$$ \sum_{i=1}^{N} \sum_{j=1}^{N} u_j \tilde{u}_i \int_{\Omega} \nabla \varphi_i(x) \cdot \nabla \varphi_j(x) \, dx = \sum_{i=1}^{N} \tilde{u}_i \int_{\Omega} f(x) \varphi_i(x) \, dx. $$

This can be expressed in matrix form, $\tilde{U}_h^\top A_h U_h = \tilde{U}_h^\top b_h$, which is equivalent to the linear system $A_h U_h = b_h$ as the previous equality is true for all $\tilde{U}_h$, where

$$ U_h = (u_1, \ldots, u_N)^\top, \quad \tilde{U}_h = (\tilde{u}_1, \ldots, \tilde{u}_N)^\top, \quad b_h = (\int_{\Omega} f(x) \varphi_1(x) \, dx, \ldots, \int_{\Omega} f(x) \varphi_N(x) \, dx)^\top $$

and the matrix $A_h$ is defined by its entries which are

$$ (\int_{\Omega} \nabla \varphi_i(x) \cdot \nabla \varphi_j(x) \, dx)_{1 \leq i, j \leq N}. $$

### 1.2 Basic properties of the Sobolev spaces $H^m(\Omega)$, $H(\text{div}, \Omega)$ and $H(\text{div}, \Omega)$

In the course of the lecture we shall work with the Sobolev spaces $H^m(\Omega)$, $H(\text{div}, \Omega)$ and $H(\text{div}, \Omega)$ and recall here their basic properties without proof. For a more detailed presentation with proofs we refer to Section 2.1. of [3].
1.2.1 The Sobolev space $H^m$

For any integer $m \geq 1$, one can define

\[ H^m(\Omega) = \{ v \in L^2(\Omega) | D^\alpha v \in L^2(\Omega), |\alpha| \leq m \} \tag{1.5} \]

The most classical second order operator is the Laplace operator, which reads in an arbitrary dimension $d$ (generally $d = 1, 2$ or $3$),

\[ \Delta u = \sum_{i=1}^{d} \frac{\partial^2 u}{\partial x_i^2} . \]

The classical Green formula for the Laplace operator reads

For essential boundary conditions related with this Green formula we shall define the space

\[ H^1_0(\Omega) = \{ v \in H^1(\Omega) | v_{\partial\Omega} = 0 \} . \]

And other classical operator which comes from elasticity is the bilaplacian operator $\Delta^2 = \Delta \Delta$, which is a fourth order operator. The Green formula needed for variational formulations of PDEs based on the bilaplacian reads

\[ \int_{\Omega} \Delta^2 u \, v \, dx = \int_{\Omega} u \, \Delta^2 v \, dx + \int_{\partial\Omega} \left( u \frac{\partial \Delta v}{\partial n} - v \frac{\partial \Delta u}{\partial n} + \Delta u \frac{\partial v}{\partial n} - \Delta v \frac{\partial u}{\partial n} \right) \, d\sigma \tag{1.6} \]

\[ H^2_0(\Omega) = \{ v \in H^2(\Omega) | v_{\partial\Omega} = 0, \frac{\partial v}{\partial n}_{\partial\Omega} = 0 \} . \]

1.2.2 The Sobolev space $H(\text{div}, \Omega)$

\[ \int_{\Omega} \nabla \cdot \mathbf{u} \, v \, dx = - \int_{\Omega} \mathbf{u} \cdot \nabla v \, dx + \int_{\partial\Omega} \mathbf{u} \cdot \mathbf{n} \, v \, d\sigma \tag{1.7} \]

1.2.3 The Sobolev space $H(\text{curl}, \Omega)$

The Green’s formula that will be useful for variational problems involving the curl operator reads

\[ \int_{\Omega} \mathbf{u} \cdot \nabla \times \mathbf{v} = \int_{\Omega} \nabla \times \mathbf{u} \cdot \mathbf{v} + \int_{\partial\Omega} (\mathbf{u} \times \mathbf{n}) \cdot \mathbf{v} \, d\sigma \tag{1.8} \]

1.3 The variational (or weak) form of a boundary value problem

The variational form of a boundary value problem contains all its elements, which are the partial differential equation in the interior of the domain and the boundary conditions. There are two very distinct ways to handle the boundary conditions depending on how they appear when deriving the variational formulation. If they appear on the test function they are called essential boundary conditions and need to be included in the space where the solution is looked for. If they appear on the trial function, which will be the approximate solution, they can be handled in a natural way in the variational formulation. Such boundary conditions are called natural boundary conditions. We will
see on the examples of Dirichlet and Neumann boundary conditions how this works in practice.

In order to define the variational formulation, we will need the following Green formula: For \( u \in H^2(\Omega) \) and \( v \in H^1(\Omega) \)

\[
- \int_{\Omega} \Delta u v \, dx = \int_{\Omega} \nabla u \cdot \nabla v \, dx - \int_{\partial\Omega} \frac{\partial u}{\partial n} v \, d\sigma. \tag{1.9}
\]

Here \( H^2(\Omega) \) denotes the Hilbert space of the functions whose partial derivatives up to second order are in \( L^2(\Omega) \) and \( \frac{\partial u}{\partial n} = \nabla u \cdot \nu \), where \( \nu \) is the outbound normal at any point of the boundary.

**Remark 1** Functions in Sobolev spaces are not necessarily continuous and hence their value on the boundary is not naturally defined. In fact, one can define, what is called a trace on the boundary, by first defining it for sufficiently smooth functions and then extending it by continuity, typically using a Green's formula like (1.9). So the Green's formula also tells us what kind of trace can be defined. For a more detailed description one can consult a classical functional analysis or a Finite Element textbook like for example [3].

### 1.3.1 Case of homogeneous Dirichlet boundary conditions

Let \( f \in L^2(\Omega) \). Consider the boundary value problem

\[
\begin{align*}
-\Delta u &= f & \text{in } \Omega, \\
u &= 0 & \text{on } \partial\Omega. \tag{1.10}
\end{align*}
\]

Assume that \( u \in H^2(\Omega) \), multiply (1.10) by \( v \in H^1(\Omega) \) and integrate using the Green formula (1.9), which yields

\[
\int_{\Omega} \nabla u \cdot \nabla v \, dx - \int_{\partial\Omega} \frac{\partial u}{\partial n} v \, d\sigma = \int_{\Omega} f v \, dx.
\]

Here \( u \) does not appear in the boundary integral, so we cannot apply the boundary condition directly. But in the end \( u \) will be in the same function space as the test function \( v \), which appears directly in the boundary integral. This is the case of an essential boundary condition. So we take test functions \( v \) vanishing on the boundary. We then get the following variational formulation:

Find \( u \in H^1_0(\Omega) \) such that

\[
\int_{\Omega} \nabla u \cdot \nabla v \, dx = \int_{\Omega} f v \, dx \quad \forall v \in H^1_0(\Omega). \tag{1.12}
\]

The solutions of this variational formulation are called weak solutions of the original boundary value problem. The solutions which are also in \( H^2(\Omega) \) are called strong solutions. Indeed we can prove that such a solution is also a solution of the initial boundary value problem (1.10)-(1.11). If \( u \in H^2(\Omega) \), the Green formula (1.9) can be used, and as \( \phi \) vanishes on the boundary it yields

\[
- \int_{\Omega} \Delta u \phi \, dx = \int_{\Omega} f \phi \, dx \quad \forall v \in H^1_0(\Omega).
\]

This implies, as \( H^1_0(\Omega) \) is dense in \( L^2(\Omega) \), that \( -\Delta u = f \) in \( L^2(\Omega) \) and so almost everywhere. On the other hand as \( u \in H^1_0(\Omega) \), \( u = 0 \) on \( \partial\Omega \). So \( u \) is a strong solution of (1.10)-(1.11).
1.3.2 Case of non homogeneous Dirichlet boundary conditions

Let \( f \in L^2(\Omega) \) and \( u_0 \in H^1(\Omega) \). We consider the problem

\[
-\Delta u = f \quad \text{in } \Omega, \quad u = u_0 \quad \text{on } \partial \Omega. \quad (1.13)
\]

As the value of \( u \) on the boundary cannot be directly put in the function space if it is not zero, as else the function space would not be stable by linear combinations, we need to bring the problem back to the homogeneous case. To this aim let \( \tilde{u} = u - u_0 \). We then show as previously that \( \tilde{u} \) is a solution of the variational problem Find \( \tilde{u} \in H^1_0(\Omega) \) such that

\[
\int_\Omega \nabla \tilde{u} \cdot \nabla v \, dx = \int_\Omega f v \, dx - \int_\Omega \nabla u_0 \cdot \nabla v \, dx \quad \forall v \in H^1_0(\Omega). \quad (1.15)
\]

This is the variational problem that needs to be solved for non homogeneous Dirichlet boundary conditions. In practice \( u_0 \) is only known on the boundary and needs to be extended to the whole domain. For standard finite elements one can pick the degrees of freedom on the boundary to be the projection (or finite element interpolation) of \( u_0 \) on the boundary and let the inner degrees of freedom be zero.

1.3.3 Case of Neumann boundary conditions

Let \( f \in L^2(\Omega) \) and \( g \in H^1(\Omega) \). We consider the problem

\[
\begin{align*}
-\Delta u + u &= f \quad \text{in } \Omega, \\
\frac{\partial u}{\partial n} &= g \quad \text{on } \partial \Omega.
\end{align*} \quad (1.16) \quad (1.17)
\]

Assuming that \( u \in H^2(\Omega) \), we multiply by a test function \( v \in H^1(\Omega) \) and integrate using the Green formula (1.9), which yields

\[
\int_\Omega \nabla u \cdot \nabla v \, dx - \int_{\partial \Omega} \frac{\partial u}{\partial n} v \, d\sigma + \int_\Omega uv \, dx = \int_\Omega fv \, dx.
\]

Replacing \( \frac{\partial u}{\partial n} \) by its value \( g \) on the boundary, we obtain the variational formulation Find \( u \in H^1(\Omega) \) such that

\[
\int_\Omega \nabla u \cdot \nabla v \, dx + \int_\Omega uv \, dx = \int_\Omega fv \, dx + \int_{\partial \Omega} gv \, d\sigma \quad \forall v \in H^1(\Omega). \quad (1.18)
\]

Let us now show that \( u \) is a strong solution of the boundary value problem, provided it is in \( H^2(\Omega) \). As \( H^1_0(\Omega) \subset H^1(\Omega) \) one can first take only test functions in \( H^1_0(\Omega) \). Then as in the case of homogeneous Dirichlet conditions it follows from the Green formula (1.9) that

\[
\int_\Omega (-\Delta u + u) \varphi \, dx = \int_\Omega f \varphi \, dx \quad \forall \varphi \in H^1_0(\Omega).
\]

This implies, as \( H^1_0(\Omega) \) is dense in \( L^2(\Omega) \), that \( -\Delta u + u = f \) in \( L^2(\Omega) \) and so almost everywhere.
It now remains to verify that we have the boundary condition
\[ \frac{\partial u}{\partial n} = g \text{ on } \partial \Omega. \]

For that we start from (1.18) and apply the Green formula (1.9), which yields
\[
- \int_{\Omega} \Delta u v \, dx + \int_{\partial \Omega} \frac{\partial u}{\partial n} v \, d\sigma + \int_{\Omega} uv \, dx = \int_{\Omega} f v \, dx + \int_{\partial \Omega} g v \, d\sigma \quad \forall v \in H^1(\Omega),
\]
and as \(-\Delta u + u = f\), it remains
\[
\int_{\partial \Omega} \frac{\partial u}{\partial n} v \, d\sigma = \int_{\partial \Omega} g v \, d\sigma \quad \forall v \in H^1(\Omega),
\]
which yields that \( \frac{\partial u}{\partial n} = g \text{ on } \partial \Omega. \)

### 1.4 Solution of the the variational formulation

The variational problems we consider in this chapter, as the examples from the previous section, can be written in the following abstract form

**Find** \( u \in V \) **such that**

\[
a(u, v) = l(v) \quad \forall v \in V,
\]

where \( V \) is a Hilbert space, \( a \) is a symmetric and coercive bilinear form and \( l \) a continuous linear form.

The most convenient tool for proving existence and uniqueness of the solution of a variational problem is the Lax-Milgram theorem that we recall here:

**Theorem 1 (Lax-Milgram)** Let \( V \) a Hilbert space with the norm \( \| \cdot \|_V \). Let \( a(\cdot, \cdot) \) be a continuous, symmetric and coercive bilinear form on \( V \times V \), i.e.

1. (Continuity): there exists \( C \) such that for all \( u, v \in V \)
\[
|a(u, v)| \leq C \| u \|_V \| v \|_V.
\]

2. (Coercivity): there exists a constant \( \alpha > 0 \) such that for all \( u \in V \)
\[
a(u, u) > \alpha \| u \|_V^2.
\]

Let \( l(\cdot) \) a continuous linear form on \( V \), i.e. there exists \( C \) such that for all \( v \in V \)
\[
|l(v)| \leq C \| v \|_V.
\]

*Then there exists a unique \( u \in V \) such that
\[
a(u, v) = l(v) \quad \forall v \in V.
\]*

To illustrate the use of the Lax-Milgram theorem, let us apply it to the examples of the previous section. The case of the problem with homogeneous Dirichlet boundary conditions, yields to the variational formulation (1.12), which fits into the framework of the Lax-Milgram theorem with \( V = H^1_0(\Omega) \), \( a(u, v) = \int_{\Omega} \nabla u \cdot \nabla v \, dx \), \( l(v) = \int_{\Omega} f v \, dx \).
The continuity of the bilinear form $a$ and the linear form $l$ follows from the Cauchy-
Schwarz inequality. $a$ is clearly symmetric and the coercivity is a consequence of the
Poincaré inequality:
\[
\|u\|_{0,\Omega} \leq C|u|_{1,\Omega}, \quad \forall u \in H^1_0(\Omega),
\]
which implies that
\[
a(u, u) = |u|_{1,\Omega} \geq \frac{1}{2}|u|_{1,\Omega} + \frac{1}{2C}|u|_{0,\Omega} \geq \min\left(\frac{1}{2}, \frac{1}{2C}\right)\|u\|_{1,\Omega}.
\]

The case of Neumann boundary conditions yields the variational formulation (1.18),
in which case $V = H^1(\Omega)$, $a$ is the $H^1$ scalar product which is clearly continuous,
symmetric and coercive, and $l(v) = \int_{\Omega} fv\,dx + \int_{\partial\Omega} gv\,d\sigma$, which is a continuous linear
form thanks to the Cauchy-Schwarz inequality and the continuity of the trace.

1.5 Description of a Finite Element

1.5.1 Formal definition of a Finite Element

Let $(K, P, \Sigma)$ be a triple such that
(i) $K$ is a closed subset of $\mathbb{R}^n$ of non empty interior,
(ii) $P$ is a finite dimensional vector space of functions defined on $K$,
(iii) $\Sigma$ is a set of linear forms on $P$ of finite cardinal $N$, $\Sigma = \{\sigma_1, \ldots, \sigma_N\}$.

Definition 1 The elements of $\Sigma$ are called degrees of freedom of the Finite Element.

The degrees of freedom characterise the basis of $P$ associated to the Finite Element.

Definition 2 $\Sigma$ is said to be $P$-unisolvent if for any $N$-tuple $(\alpha_1, \ldots, \alpha_N)$, there exists
a unique element $p \in P$ such that $\sigma_i(p) = \alpha_i$ pour $i = 1, \ldots, N$.

Definition 3 The triple $(K, P, \Sigma)$ of $\mathbb{R}^n$ is called Finite Element of $\mathbb{R}^n$ if it satisfies
(i), (ii) and (iii) and if $\Sigma$ is $P$-unisolvent.

In the definition of a Finite Element, $K$ is the domain on which the Finite Element is
defined, $P$ is the finite dimensional approximation space, and $\Sigma$ uniquely defines a basis
of $P$, which is needed to build the Finite Element matrices associated to the variational
formulation applied to functions in the Finite Dimensional approximation space. This
basis can be either defined through the degrees of freedom or directly by exhibiting an
explicit formula for the basis functions. In this last case the degrees of freedom are not
explicitly needed.

The unisolvence property is needed to establish that the elements of $P$ characterised
by the degrees of freedom actually form a basis of $P$. In practice, this is done by first
verifying that the dimension is right, i.e. that the number of degrees of freedom is equal
to $\dim P = N$ and then by checking either the injectivity or the surjectivity, which both
imply the bijectivity if the dimension is right, of the mapping
\[
P \rightarrow \mathbb{R}^N
\]
\[
p \mapsto (\sigma_1(p), \ldots, \sigma_N(p))
\]
This can be formalised in the two following lemmas:
Lemma 1 The set $\Sigma$ is $P$-unisolvent if and only if the two following properties are satisfied

(i) $\dim P = |\Sigma|$ (where $|\Sigma|$ denotes the number of elements in the set $\Sigma$).

(ii) $\sigma_j(p) = 0$ for $j = 1, \ldots, N \Rightarrow p = 0$.

Lemma 2 The set $\Sigma$ is $P$-unisolvent if and only if the two following properties are satisfied

(i) $\dim P = |\Sigma| = N$.

(ii) There exist $N$ linearly independent functions $p_i \in P, i = 1, \ldots, N$ such that $\sigma_j(p_i) = \delta_{ij}$.

In addition to its local definition, independent of its neighbours, the degrees of freedom of a Finite Element need to be chosen to allow a natural embedding into the function spaces in which the variational formulation is defined. In practice for Finite Element spaces embedded in $L^2$ no continuity is required, for Finite Element spaces embedded in $H^1 C^0$ continuity is required, for Finite Element spaces embedded in $H(\text{curl}, \Omega)$ $C^0$ continuity of the tangential component is required and for Finite Element spaces embedded in $H(\text{div}, \Omega)$ $C^0$ continuity of the normal component is required.

We will thus present examples of Finite Elements according to their conformity, i.e. to the space in which they can be naturally embedded. Note that the continuity requirement is enforced by sharing the degrees of freedom on the interface between two elements and by making sure that these define uniquely the trace of $P$ on the interface.

Remark 2 Note that not all degrees of freedom on the interface need to be shared between the elements sharing the interface. This needs to be decided when constructing the global Finite Element space from the reference element.

1.5.2 $C^0$ Lagrange Finite Elements

These are continuous Finite Elements, and the continuity will be enforced by sharing the degrees of freedom on the interface. Thanks to this continuity property the Finite Element constructed from those will be included in $H^1$. These elements are called $H^1$ conforming.

1D Lagrange $P_k$ Element Let $a, b \in \mathbb{R}, a < b$. Let $K = [a, b], P = \mathbb{P}_k$ the set of polynomials of degree $k$ on $[a, b], \Sigma = \{\sigma_0, \ldots, \sigma_k\}$, where $a = x_0 < x_1 < \cdots < x_k = b$ are distinct points and

$$\sigma_k : P \rightarrow \mathbb{R},$$

$$p \mapsto p(x_i).$$

Moreover $\Sigma$ is $P$-unisolvent, using Lemma 2. Indeed the Lagrange polynomials at the interpolation points $x_0, x_1, \ldots, x_k$ which read

$$l_{k,i}(x) = \prod_{0 \leq j \leq k}^{j \neq i} \frac{(x - x_j)}{(x_i - x_j)}$$

(1.21)
satisfy \( l_{k,i}(x_j) = \delta_{i,j}, \) 0 \( \leq \) \( i, j \leq k \).

The fact that the points are all distinct makes the Lagrange interpolation problem well posed. The first and last point being on the boundary, will correspond to shared degrees of freedom. The corresponding basis function will be shared with the neighbouring element.

For low degree polynomials, typically up to three, the degrees of freedom can be chosen uniformly space. For higher degree, the resulting matrices have better properties in particular conditioning if Gauss-Lobatto points are chosen. Note that high order Lagrange Finite Elements based on Gauss-Lobatto points are often called spectral elements.

The extension from one to multiple dimensions is generally done in two ways. Either one builds finite elements based on simplices, or on tensor products of 1D elements. There are other possibilities but they are less classical.

Simplex based \( \mathbb{P}_k \) Finite Elements. A simplex in \( n \) dimensions is the convex hull of \( n + 1 \) affine independent points. In 2D this is a non degenerate triangle in 3D a tetrahedron. In simplices basis functions are most easily expressed using the barycentric coordinates associated to the vertices of the simplex, that we shall denote by \( \mathbf{a}_0, \ldots, \mathbf{a}_n \in \mathbb{R}^n \). The barycentric coordinate, classically denoted by \( (\lambda_i)_{0 \leq i \leq n} \), associated to the vertex \( \mathbf{a}_i \) is the affine function with value 1 at \( \mathbf{a}_i \) and 0 at the \( n \) other vertices. More precisely, for any point \( \mathbf{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n \),

\[
\lambda_i(\mathbf{x}) = \alpha_{i,0} + \sum_{i=1}^n \alpha_{i,i} x_i, \quad i = 0, \ldots, n. \tag{1.22}
\]

The coefficients \( (\alpha_{i,j})_{0 \leq j \leq n} \) being determined by the relations \( \lambda_i(\mathbf{a}_j) = \delta_{i,j} \).

Example: \( \mathbb{P}_k \) in 2D. For 3 non aligned points \( \mathbf{a}_1, \mathbf{a}_2, \) and \( \mathbf{a}_3 \) in \( \mathbb{R}^2 \) let \( K \) be the triangle defined by \( \mathbf{a}_1, \mathbf{a}_2 \) and \( \mathbf{a}_3 \). Let \( P = \mathbb{P}_k \) be the vector space of polynomials of degree \( k \) in 2D

\[
\mathbb{P}_k = \{ \text{Span}(x^\alpha y^\beta), \quad (\alpha, \beta) \in \mathbb{N}^2, \quad 0 \leq \alpha + \beta \leq k \}.
\]

The dimension of \( \mathbb{P}_k \) is \( \frac{(k+1)(k+2)}{2} \). Putting the Lagrange interpolation points on a lattice with \( k + 1 \) points on the edges and then one point less at each level starting from one edge, we get \( (k + 1) + k + (k - 1) + \cdots + 1 = (k + 1)(k + 2)/2 \) interpolation points and so also degrees of freedom, which is the same as the dimension of the space (see Figure 1.2 for an example with \( k = 3 \)). The basis functions can be expressed using the barycentric coordinates. In our example with \( k = 3 \). The three basis functions associated to the three vertices are \( (9/2) \lambda_i(\lambda_i - 1/3)(\lambda_i - 2/3) \), the basis function associated associated to the point closest to \( \mathbf{a}_i \) on the edge \( [\mathbf{a}_i, \mathbf{a}_j] \) is \( (27/2) \lambda_i \lambda_j (\lambda_i - 1/3) \). There are six basis functions of this form. Finally the basis function associated to the center point is \( 27 \lambda_1 \lambda_2 \lambda_3 \). We thus find 10 basis functions for \( \mathbb{P}_3 \), whose dimension is 10.
To sum up, we have 10 unknowns so we need a total of 10 node points on the triangle element to determine the coefficients of all 6 basis functions in the quadratic case by inverting the matrix. In the case of cubic finite elements, the basis functions are cubic polynomials of degree 3 with the designation $P_3$.

First, consider the triangular element with four nodal points and obtain a set of Lagrange-type shape functions, namely $c = 0$. Do not specify the minimum number of symmetric terms to be dropped from the full polynomial.

Similarly, here are the interpolation functions of the nine-node Lagrange element. The nodes in the $x$-direction are placed at 0, and the nine basis functions are at

$$l_{k,i}(x)l_{k,j}(y), \quad 0 \leq i, j \leq 2.$$

### 1.5.3 $C^1$ Hermite Finite Elements

Such elements are $H^2$ conforming and are needed when second order derivatives appear in the variational formulation.

**1D Hermite $P_{2k+1}$ Finite Element.** Hermite interpolation on an interval $[a, b]$ consists in finding a polynomial $p$ of degree $2k + 1$ for $k \geq 1$, such that $p^{(l)}(a)$ and $p^{(l)}(b)$ are given for $0 \leq l \leq k$. This will enforce $C^1$ continuity if all the interface degrees of freedom are shared. To get $C^1$ continuity $k = 1$ is enough:

*Example $P_3$ Hermite element in 1D.* This is the simplest possible Hermite Finite Element. Let $a, b \in \mathbb{R}$, $a < b$. Let $K = [a, b]$, $P = P_3$ the linear space of cubic polynomials on $[a, b]$, $\Sigma = \{ \sigma_1, \sigma_2, \sigma_3, \sigma_4 \}$, with

$$
\begin{align*}
\sigma_1 & : P \to \mathbb{R}, & \sigma_2 & : P \to \mathbb{R}, & \sigma_3 & : P \to \mathbb{R}, & \sigma_4 & : P \to \mathbb{R}, \\
p & \mapsto p(a), & p & \mapsto p(b), & p & \mapsto p'(a), & p & \mapsto p'(b).
\end{align*}
$$

$C^1$ Finite Elements are much more involved in higher dimensions at least for triangles and higher dimensional simplices than $C^0$ finite elements. Tensor product constructions can be obtained from 1D Hermite Finite Elements, however they often have the constraint that the mapping also needs to be $C^1$, which is not the case for general affine mappings as are standardly used for $C^0$ Lagrange Finite Elements. Let us introduce the simplex examples on simplest triangle based and quad based $C^1$ finite elements.

Figure 1.2: (Left) $P_3$ Finite Element, (Right) $Q_2$ Finite Element
**Triangle $\mathbb{P}_3$ Hermite element.**  $K$ is a non degenerate triangle of vertices $(a_1, a_2, a_3)$, $P = \mathbb{P}_3$. So $\dim P = 10$ and so ten degrees of freedom are needed. To enforce $C^1$ continuity, the value and the two components of the gradient need to be fixed on each of the three vertices. This amounts to nine degrees of freedom, the tenth can be taken to be the value at the barycenter of the triangle. This is represented on the left of Figure 1.3.

**The Bogner-Fox-Schmitt rectangle $Q_3$ Finite element.**  Taking the tensor product of two 1D $P_3$ Hermite Finite element. One obtains the Bogner-Fox-Schmitt element, which in addition to the point and two derivative values at each of the four vertices has also the cross derivative $\frac{\partial^2 p}{\partial x \partial y}$ as a degree of freedom. The corresponding basis consists of the products of all possible 1D cubic Hermite basis functions. Its dimension is $4 \times 4 = 16$. This is represented on the right of Figure 1.3.

**Discontinuous finite elements**

Functions that are only in $L^2$ can be defined piecewise on each cell without any continuity requirements. Then there is no restriction on the degrees of freedom.

Classical discontinuous Finite Elements are *nodal* Lagrange elements based on Gauss points as the quadrature points. Even though their approximation order is lower for the same number of points it is sometimes convenient to place the degrees of freedom at Gauss-Lobatto points, where one point on each side is on the edge. This makes easier the computation of edge integrals. However in this case, as the Finite Element is discontinuous, degrees of freedom on the interface are not shared by neighbouring elements.

Another class of classical discontinuous Finite Elements are *modal* elements where the basis functions are generally taken to be the orthonormal Legendre polynomials $(L_i)_{0 \leq i \leq k}$, which have the advantage of yielding the identity as a mass matrix. The corresponding degrees of freedom are $\sigma_i(p) = \int p(x)L_i(x)\,dx$. Such constructions, where each element of the basis has a deferent degree, are called hierarchical finite elements and facilitate what is called the $p$ refinement consisting in obtaining a more accurate approximation by taking a higher order polynomial rather that refining the grid. Indeed in this case when refining only one basis function needs to be added to the existing one rather than replacing all the basis functions as would be necessary for the previously seem Finite Elements where all basis functions have the same degree.
1.5.4 $H(\text{div}, \Omega)$ conforming Finite Elements

The $H(\text{div}, \Omega)$ space consists of vector fields. Each element as $n$ components in $n$ dimensions. Moreover, as we have seen before $H(\text{div}, \Omega)$ vector fields have a continuous normal component and a discontinuous tangential component. Let us explain how such elements can be constructed in two dimensions. This can be generalised to higher dimensions.

**Tensor product construction.** Due to the given continuity requirements a $H(\text{div}, \Omega)$ conforming Finite Element can be constructed, by defining separately an approximation of the tangential component and of the normal component. In this case $K = [-1,1] \times [-1,1]$,

$$P = \{ \mathbf{p} = (p_x, p_y)^\top | p_x \in \mathbb{Q}_{k-1,k}, p_y \in \mathbb{Q}_{k,k-1} \},$$

where

$$\mathbb{Q}_{m,n} = \text{Span}(x^\alpha y^\beta, 0 \leq \alpha \leq m, 0 \leq \beta \leq n).$$

For $p_x$ the degrees of freedom are the Gauss points in $x$ and the Gauss-Lobatto points in $y$, such that $p_x$ is discontinuous in $x$ and continuous in $y$. It is the other way for $p_y$. This enables arbitrarily high order $H(\text{div}, \Omega)$ conforming elements.

**The Raviart-Thomas (RT) Finite Element.** $K$ is a non degenerate triangle of vertices $(a_1, a_2, a_3)$. For the element of order $k + 1$, $k \geq 0$, denoting by $\mathbb{P}_k$ the space of homogeneous polynomials of degree $k$ (i.e. all the monomials are exactly of degree $k$)

$$P = RT_k = (\mathbb{P}_k)^2 + \left( \begin{array}{c} x \\ y \end{array} \right) \mathbb{P}_k.$$  

In two dimensions $\dim \mathbb{P}_k = (k + 1)(k + 2)/2$ and $\dim \mathbb{P}_k = k + 1$, so that $\dim RT_k = (k + 1)(k + 3)$. This yields in particular $\dim RT_0 = 3$, $\dim RT_1 = 8$, $\dim RT_2 = 15$. The degrees of freedom are constructed starting from the edges in order to enforce the continuity requirements. This is our first example of moment based degrees of freedom, defined by 1D moments along the edges to enforce continuity and then 2D moments in the triangle to complete the missing degrees of freedom.

$$\Sigma = \left\{ \mathbf{p} \mapsto \int_{e_i} \mathbf{p} \cdot \mathbf{\nu} l^l \, ds, 0 \leq l \leq k, \quad \mathbf{p} \mapsto \int_K \mathbf{p}(x_1, x_2) \cdot x_i^l \, dx_i, \quad 0 \leq l \leq k - 1, \quad i = 1, 2 \right\}.$$  

We count here $k + 1$ degrees of freedom for each of the three edges, and $2k$ inner degrees of freedom, which are in total $3(k + 1) + 2k(k + 1)/2 = (k + 1)(k + 3)$ which is precisely the dimension of $RT_k$. To prove the unisolvence it is thus enough to prove that a polynomial which vanishes on all degrees of freedom is necessarily zero. Note that because the monomials in our degrees of freedom are a basis of the polynomial spaces $\mathbb{P}_k(e_i)$, $i = 1, 2, 3$ and $\mathbb{P}_{k-1}(K)$ respectively this is given by the following lemma:

**Lemma 3** Let $k \geq 0$, and $\mathbf{p} \in RT_k(K)$ then

$$\int_{e_i} \mathbf{p} \cdot \mathbf{\nu} \, r \, ds = 0, \quad \forall r \in \mathbb{P}_k(e_i), \quad (1.23)$$

$$\int_K \mathbf{p} \cdot \mathbf{q} \, dx = 0, \quad \forall \mathbf{q} \in (\mathbb{P}_{k-1})^2 \quad (1.24)$$

implies that $\mathbf{p} = 0.$
Remark 3 Note that instead of the monomials one can use (1.23) and (1.24) applied to arbitrary bases of $\mathbb{P}_k(e_i)$ and $(\mathbb{P}_{k-1})^2$ respectively.

Proof. Let $\mathbf{p} \in RT_k$. Then $\mathbf{p} = \mathbf{q} + \begin{pmatrix} x \\ y \end{pmatrix} \psi$ with $\mathbf{q} \in (\mathbb{P}_k)^2$ and $\psi \in \mathbb{P}_k$. Let us first observe that on any of the three edges $e_i$ we have $\mathbf{p} \cdot \mathbf{v} \in \mathbb{P}_k(e_i)$, this is obviously the case for $\mathbf{p} \cdot \mathbf{v}$. On the other hand if $(x(s), y(s))$ defines a parametrisation of $e_i$, then by definition of the normal $n_x(x(s) - x(s_0)) + n_y(y(s) - y(s_0)) = 0$ so that $n_x x(s) + n_y y(s)$ is constant on $e_i$.

It follows that on the edge $e_i$

$$\mathbf{ν} \begin{pmatrix} x \\ y \end{pmatrix} \psi = (n_x x + n_y y) \psi.$$

Then as $(n_x x + n_y y)$ is a constant, $n_x x + n_y y) \psi$ is a polynomial of degree $k$ like $\psi$.

Now computing the divergence of $\mathbf{p}$ we find

$$\nabla \cdot \mathbf{p} = \nabla \cdot \mathbf{q} + 2\psi + \begin{pmatrix} x \\ y \end{pmatrix} \cdot \nabla \psi = \nabla \cdot \mathbf{q} + (k + 2)\psi.$$

Hence $\nabla \cdot \mathbf{p}$ is a polynomial of degree $k$ orthogonal to all polynomials of degree $k$. Thus it is 0. Then because $\nabla \cdot \mathbf{q} \in \mathbb{P}_{k-1}$ it also follows that $\psi = 0$ and hence $\mathbf{p} \in (\mathbb{P}_k)^2$.

Let us take $\varphi \in \mathbb{P}_k$. Then $\nabla \varphi \in (\mathbb{P}_{k-1})^2$ and hence using (1.24) we get using a Green’s formula

$$\int_K \mathbf{p} \cdot \nabla \varphi \, dx = 0 = \int_{\partial K} \mathbf{p} \cdot \mathbf{ν} \varphi \, ds + \int_K \nabla \cdot \mathbf{p} \varphi \, dx.$$

As the boundary term vanishes due to (1.23), it follows that $\int_K \nabla \cdot \mathbf{p} \varphi \, dx = 0$. This for all $\varphi \in \mathbb{P}_k$.

Now in order to conclude we assume that the reference triangle has one edge on $x = 0$ and one edge on $y = 0$ and is in the positive quarter plane. The the condition $\mathbf{p} \cdot \mathbf{v} = 0$ yields that $p_x = 0$ for $x = 0$, hence $p_x = x \psi_1$ with $\psi_1 \in \mathbb{P}_{k-1}$ and $p_y = 0$ for $y = 0$, hence $p_y = y \psi_2$ with $\psi_2 \in \mathbb{P}_{k-1}$. Then taking $\mathbf{q} = (\psi_1, \psi_2)^\top$ in (1.24), we obtain

$$\int_K (x \psi_1^2 + y \psi_2^2) \, dx = 0$$

from which it follows as both terms are positive that they are both zero. So $\mathbf{p} = 0$ which was what we needed to prove. \[\blacksquare\]

Remark 4 Obviously, the procedure used here to construct the Finite Element, by first setting the boundary degrees of freedom to ensure the required continuity and then complete with the needed inner degrees of freedom can also be applied to quads.

The Brezzi-Douglas-Marini (BDM) Finite Element. $K$ is a non degenerate triangle of vertices $(a_1, a_2, a_3)$. Here $P$ is the full polynomial space in each direction $P = BDM_k = (\mathbb{P}_k)^2$. As continuity of the normal component is needed on each of the three faces, these requires at least three degrees of freedom. So and $H(\text{div}, \Omega)$ conforming $BDM_k$ element can only be constructed for $k \geq 1$ as dim $BDM_0 = 2$. In order to define the degrees of freedom, we need the following space

$$N_k = (\mathbb{P}_k)^2 + \left( \frac{y}{-x} \right) \widehat{\mathbb{P}}_k,$$
This space has obviously the same dimension as the corresponding $RT_k$: $\dim N_k = \dim RT_k = (k + 1)(k + 3)$. We can now define the degrees of freedom for the $BDM_k$ element: introducing $(\psi_j)_{0 \leq j \leq k}$ a basis of $P_k$ on each edge (this could be the monomials $s^j$ as we used for Raviart-Thomas or any other basis), and $(\varphi_j)_{0 \leq (k - 1)(k + 1)}$ a basis of $N_{k-2}$

$$\Sigma = \left\{ p \mapsto \int_{e_i} p \cdot \nu \psi_j \, ds, 0 \leq j \leq k, \ i = 1, 2, 3, \right.$$ \begin{align*}
p &\mapsto \int_K p \cdot \varphi_j \, dx, \ 0 \leq j \leq (k - 1)(k + 1) \left.\right\}.
\end{align*}

We first notice, that the number of degrees of freedom is $3(k + 1) + (k - 1)(k + 1) = (k + 1)(k + 2) = \dim P$. So that proving that $p \in P$ must vanish if all these degrees of freedom are zero is enough to prove unisolvence. Similarly as for the Raviart-Thomas element this follows from the following:

**Lemma 4** Let $k \geq 1$, and $p \in P_k(K)^2$ then

$$\int_{e_i} p \cdot \nu \psi \, ds = 0, \ \forall \psi \in P_k(e_i), \quad (1.25)$$

$$\int_K p \cdot q \, dx = 0 \ \forall q \in N_{k-2} \quad (1.26)$$

implies that $p = 0$.

**Proof.** Let $p \in P_k(K)^2$ verifying (1.25) and (1.26). For any $\varphi \in P_{k-1}$, we have that $\nabla \varphi \in (P_{k-2})^2 \subset N_{k-2}$, hence from (1.26) it follows that

$$\int_K p \cdot \nabla \varphi \, dx = 0 = - \int_K \nabla \cdot p \, \varphi + \int_{\partial K} p \cdot \nu \, \varphi \, ds.$$

The last term is zero because of (1.25). Then $\nabla \cdot p$ is a polynomial of degree $k - 1$ orthogonal to all polynomials of degree $k - 1$. Hence it is zero. Thus $p = (\partial_y \psi, -\partial_x \psi)$ for a $\psi \in P_{k+1}$. Moreover due to (1.25), $p \cdot \nu = 0$ on each edge, this implies that the tangential derivative of $\psi$ on each edge, and so $\psi$ is constant on all the edges. As it is defined up to a constant, on can choose $\psi$ such that it vanishes on the three edges. Then introducing the barycentric coordinates $\lambda_1, \lambda_2, \lambda_3$ of the triangle, it follows that each $\lambda_i$ divides $\psi$, so that $\psi = \lambda_1 \lambda_2 \lambda_3 \tilde{\psi}$ with $\tilde{\psi} \in P_{k-2}$.

Let us now take $q = \begin{pmatrix} y \\ -x \end{pmatrix}$ $\tilde{\psi} \in N_{k-2}$. Then from (1.26) it follows that

$$\int_K p \cdot q \, dx = \int_K \begin{pmatrix} \partial_y \psi \\ -\partial_x \psi \end{pmatrix} \cdot \begin{pmatrix} y \\ -x \end{pmatrix} \tilde{\psi} \, dx = \int_K k \lambda_1 \lambda_2 \lambda_3 \tilde{\psi}^2 \, dx = 0$$

which implies that $\tilde{\psi} = 0$ and thus $p = 0$, which is the desired result. ■

**Remark 5** The extension to 3D is straightforward, the degrees of freedom stay the same, the triangles being replaced by tetrahedra and the edges being replaced by faces.

17
1.5.5 \( H(\text{curl}, \Omega) \) conforming Finite Elements

Recall that for a 3D vector \( \mathbf{u} = (u_1, u_2, u_3)\)\(^T\), the curl is defined by the relation

\[
\nabla \times \mathbf{u} = \begin{pmatrix}
\frac{\partial u_3}{\partial y} - \frac{\partial u_2}{\partial z} \\
\frac{\partial u_1}{\partial z} - \frac{\partial u_3}{\partial x} \\
\frac{\partial u_2}{\partial x} - \frac{\partial u_1}{\partial y}
\end{pmatrix}.
\]

In 2D, there is no dependency on the third coordinate and the curl degenerates into the vector curl of a scalar, for the first two components and the scalar curl of a vector for the last component:

\[
\nabla \times \mathbf{u} = \begin{pmatrix}
\text{curl } u_3 \\
\text{curl } \mathbf{u}
\end{pmatrix}
\]

with \( \text{curl } u_3 = \begin{pmatrix}
\frac{\partial u_3}{\partial y} \\
-\frac{\partial u_3}{\partial x}
\end{pmatrix} \), \( \mathbf{u} = \begin{pmatrix}
u_1 \\
u_2
\end{pmatrix} \), \( \text{curl } \mathbf{u} = \partial_1 u_2 - \partial_2 u_1 \).

\( H(\text{curl}, \Omega) \) conforming elements have a continuous tangential component. They can thus be simply obtained in 2D by exchanging the two components of the vector in the tensor product case.

For triangles, the natural finite element space, built as the Raviart-Thomas element is the Nédélec Element defined by \( K = (a_1, a_2, a_3) \), a non degenerate triangle,

\[
P = N_k = (\mathbb{P}_k)^2 + \begin{pmatrix} y \\ -x \end{pmatrix} \mathbb{P}_k,
\]

and the degrees of freedom are obtained from the Raviart-Thomas degrees of freedom by replacing the normal \( \mathbf{\nu} = (\nu_1, \nu_2)\)\(^T\) by the tangent \( \mathbf{\tau} = (\nu_2, -\nu_1) \). Unisolvence is also easily adapted from the Raviart-Thomas case and is based on the following

**Lemma 5** Let \( k \geq 0 \), and \( \mathbf{p} \in RT_k(K) \) then

\[
\int_{e_i} \mathbf{p} \cdot \mathbf{\tau} \, ds = 0, \quad \forall r \in \mathbb{P}_k(e_i), \tag{1.27}
\]

\[
\int_K \mathbf{p} \cdot \mathbf{q} \, dx = 0 \quad \forall q \in (\mathbb{P}_{k-1})^2 \tag{1.28}
\]

implies that \( \mathbf{p} = 0 \).
Remark 6 The extension to 3D for tensor product element is also quite natural by considering the components separately and noticing that in 3D there are two tangential components and three normal components.

Simplex based $H(\text{curl}, \Omega)$ Finite Elements, which are constructed on tetrahedra in 3D are fundamentally different from their $H(\text{div}, \Omega)$ counterpart, unlike in 2D where one could go from one to the other by a simple rotation of the components. There are also two classes of finite elements, like Raviart-Thomas and Brezzi-Douglas-Marini in 2D, which are called respectively Nédélec elements of first type and of second type.

The 3D Nédélec space in which the elements of first type live and which is used to define the degrees of freedom of the Nédélec elements of second type reads

$$N_k = (P_k)^3 \oplus (x \times (P_k)^3)$$

(1.29)

and the degrees of freedom are defined by the following unisolvence lemma

Lemma 6 Let $K$ be a non degenerate tetrahedron, with faces $(f_i)_{0 \leq i \leq 4}$, and edges $(e_i)_{0 \leq i \leq 6}$ $k \geq 0$, and $p \in N_k(K)$ then

$$\int_{e_i} p \cdot \tau_i \cdot r \, ds = 0, \quad \forall r \in P_k(e_i),$$

(1.30)

$$\int_{f_i} (p \times \nu_i) \cdot s \, ds = 0, \quad \forall s \in (P_{k-1}(f_i))^2,$$

(1.31)

$$\int_K p \cdot q \, dx = 0 \quad \forall q \in (P_{k-2})^2$$

(1.32)

implies that $p = 0$.

We denote here by $\tau_i$ the unit vector along the edge $e_i$ and by $\nu_i$ the outward unit normal on face $f_i$.

1.6 Change of local basis

The local degrees of freedom $\Sigma$ of a Finite Element defined by $(K, P, \Sigma)$ can also be used to compute the element matrices with respect to the matrices corresponding to a simpler or easily computable basis of the same space $P$ of dimension $N$. This could for example
be an orthonormal basis. This allows to compute the matrices once for all in some basis and then to get the matrices in any other basis by just computing the corresponding generalised Vandermode matrix as follows.

Let us denote by \((\phi_l)_{0 \leq l \leq N}\) the reference basis in which the element matrices are known. Let now \((\hat{p}_j)_{0 \leq j \leq N}\) be another basis of the same linear space \(P\) associated to the degrees of freedom \((\sigma_i)_{0 \leq i \leq N}\). Then \(\hat{p}_j\) can be expressed in the basis \((\phi_l)_{0 \leq l \leq N}\):

\[
\hat{p}_j(x) = \sum_{l=1}^{N} \alpha_{j,l} \phi_l(x).
\]

(1.33)

Denoting by \(\hat{p} = (\hat{p}_1, \ldots, \hat{p}_N)^\top\), \(\phi = (\phi_1, \ldots, \phi_N)^\top\) and \(A = ((\alpha_{j,l}))_{1 \leq j,l \leq N}\), this relation can be written in matrix form

\[
\hat{p}(x) = A\phi(x).
\]

Applying the linear forms \(\sigma_i\) to relation (1.33) for \(1 \leq i \leq N\), we get

\[
I_N = VA^\top,
\]

(1.34)

where \(I_N\) is the identity matrix and \(V = ((\sigma_i(\phi_j)))_{1 \leq i,j \leq N}\). The matrix \(V\) is called generalised Vandermonde matrix, as in the case when \((\phi_j)_{j \leq N}\) is the monomial basis \((1, x, \ldots, x^{N-1})\) of \(P_k\) in 1D and for a Lagrange Finite Element where \(\sigma_i(\hat{p}_j) = \hat{p}_j(x_i)\), the matrix \(V\) is the classical Vandermonde matrix \(V = ((x_i^{j-1}))_{1 \leq i,j \leq N}\).

The generalised Vandermonde matrix \(V\) is explicitly computable when the basis \((\phi_j)_{j}\) and the degrees of freedom \((\sigma_i)_{i}\) are explicitly known. From relation (1.34), it follows immediately that \(A = V^{-1}\), where we denote by \(V^{-1}\) the inverse of the transpose of \(V\).

One can also take the derivative of formula (1.33). Thus

\[
\partial_x \hat{p}_j(x) = \sum_{l=1}^{N} \alpha_{j,l} \partial_x \phi_l(x),
\]

and the same for the other variables if needed. Applying again the linear form \(\sigma_i\) to this equation for \(1 \leq i \leq N\), we find

\[
D_x \hat{p} = D_x \phi A^\top = D_x \phi V^{-1},
\]

where the matrices \(D_x \hat{p}\) and \(D_x \phi\) are defined respectively by their generic element \(((\sigma_i(\partial_x \hat{p}_j)))_{1 \leq i,j \leq N}\) and \(((\sigma_i(\partial_x \phi_j)))_{1 \leq i,j \leq N_k}\). Note that if \((\phi_j)_{j}\) has been chosen such \(D_x \phi\) is explicitly computable, we can, thanks to this relation explicitly compute the matrix \(D_x \hat{p}\) and similarly the derivative matrices with respect to other variables.

We can express \(M\) the mass matrix on the reference element \(K\) using the Vandermonde matrix \(V\). Using formula (1.33),

\[
\hat{p}_j(x) = \sum_{l=1}^{N} \alpha_{j,l} \phi_l(x).
\]

Hence

\[
\int_K \hat{p}_i(x) \hat{p}_j(x) \, dx = \sum_{l=1}^{N} \sum_{m=1}^{N} \alpha_{i,l} \alpha_{j,m} \int_K \phi_l(x) \phi_m(x) \, dx = \sum_{l=1}^{N} \alpha_{i,l} \alpha_{j,m},
\]

20
if the basis functions \((\phi_j)_j\) are orthonormal on \(\hat{K}\). It follows that \(\hat{M} = AA^\top = V^{-\top}V^{-1}\).

If the space \(P\) is stable by derivation, as is the case for the polynomial space \(P_k\), we can also express the derivative \(\partial_x G_h\) or other derivatives in the basis \((p_i)_i\) and this derivative can also be characterised by the degrees of freedom \((\sigma_i(\partial_x p_j))\) \(1 \leq i \leq N_k\). We denote by \(D_x G\) the vector containing these degrees of freedom, which can be expressed using \(G\) as follows. We have

\[
\partial_x G_h(x) = \sum_{j=1}^{N_k} \sigma_j(G_h) \partial_x p_j(x),
\]

and so for \(1 \leq i \leq N_k\),

\[
\sigma_i(\partial_x G_h) = \sum_{j=1}^{N_k} \sigma_j(G_h) \sigma_i(\partial_x p_j).
\]

This can also be written \(D_x G = (D_x p)G\), denoting by \(D_x p\) the matrix with generic term \((\sigma_i(\partial_x p_j))\) \(1 \leq i, j \leq N\).

## 1.7 Construction of the Finite Element approximation space

### 1.7.1 The mapping

A conforming Finite Element space \(V_h\), is constructed from a reference element, by mapping the reference Finite Element defined previously to a collection of physical elements that define an admissible mesh of the computational domain and ensuring that \(V_h \subset V\) by equating degrees of freedom on the interface.

The coefficients of Finite Element matrices are defined as integrals involving the basis functions. The standard practice as well for convergence theory as for practical implementation is to make a change of variables from one unique reference element to each of the other elements. For Lagrange Finite Elements affine transformation can be used for simplices and multi-linear transformation for tensor product elements. For higher order elements, when one wants to get a corresponding high order approximation of the boundary, a polynomial transformation of the same order as the Finite Element approximation of the elements touching the boundary is used. This is called isoparametric Finite Elements.

Let us call \(\hat{K} \subset \mathbb{R}^d\) the reference element and \(K \subset \mathbb{R}^d\) any element of the physical mesh. We then define the mapping

\[
\mathbf{F} : \hat{K} \rightarrow K,
\]

\[
\hat{x} = (\hat{x}_1, \ldots, \hat{x}_d) \mapsto x = \mathbf{F}(\hat{x}) = (F_1(\hat{x}), \ldots, F_d(\hat{x}))^\top.
\]

As differential operators are involved we will need the Jacobian matrix of the mapping, denoted by

\[
DF(\hat{x}) = \left( \frac{\partial F_i}{\partial \hat{x}_j} \right)_{1 \leq i,j \leq d}.
\]

Let us denote by \(J(\hat{x}) = \det(DF(\hat{x}))\), for any scalar function \(v(x)\), \(\hat{v}(\hat{x}) = v(\mathbf{F}(\hat{x}))\), by \(\nabla = (\frac{\partial}{\partial \hat{x}_1}, \ldots, \frac{\partial}{\partial \hat{x}_d})^\top\) and finally by \(\hat{\nabla} = (\frac{\partial}{\partial \hat{x}_1}, \ldots, \frac{\partial}{\partial \hat{x}_d})^\top\). We then find, using the chain rule that

\[
\frac{\partial \hat{v}}{\partial \hat{x}_j} = \sum_{i=1}^d \frac{\partial F_i}{\partial \hat{x}_j} \frac{\partial v}{\partial x_i}.
\]
So that
\[ \hat{\nabla} \hat{v}(\hat{x}) = (DF)^\top \nabla v(x) \Rightarrow \nabla v(x) = (DF)^{-\top} \hat{\nabla} \hat{v}(\hat{x}). \] (1.35)

An important special case is when \( F \) is an affine map. This is typically used for simplicial Finite Elements (except on the boundary for high order elements as it brings back the method to second order). The classical Finite Element theory is also based on affine mappings. In this case \( F(\hat{x}) = A\hat{x} + b \), where \( A \) is a \( d \times d \) matrix and \( b \) a \( d \)-component column vector. For a simplicial mesh, the affine transformation, that is the components of \( A \) and \( b \), is completely determined by the imposing that \( F(\hat{a}_i) = a_i \) \( 0 \leq i \leq d \), i.e. that all vertices of the reference element \( \hat{K} \) are mapped to the vertices of the current element \( K \). We then have, for \( \hat{a}_0 = (0, \ldots, 0) \) and the components of \( a_{i,j} \) \( (1 \leq j \leq d) \) of \( \hat{a}_i \) are all 0 except the \( i^{th} \) component which is 1,

\[
A = \begin{pmatrix}
  a_{1,1} - a_{0,1} & a_{2,1} - a_{0,1} & \cdots & a_{d,1} - a_{0,1} \\
a_{1,2} - a_{0,2} & a_{2,2} - a_{0,2} & \cdots & a_{d,2} - a_{0,2} \\
  \vdots & \vdots & \ddots & \vdots \\
a_{1,d} - a_{0,d} & a_{2,d} - a_{0,d} & \cdots & a_{d,d} - a_{0,d}
\end{pmatrix}, \quad b = a_0.
\]

The jacobian matrix of the affine transformation is then the matrix \( A \).  

1.7.2 Bases of tangent and cotangent spaces in curvilinear coordinates

A mapping from \( F : \mathbb{R}^d \to \mathbb{R}^d \) defines a curvilinear coordinate system, which induces a natural vector space called tangent space a each point. The column vectors of the jacobian matrix \( DF(\hat{x}) \), \( t_i = \frac{\partial F}{\partial x_i} \), \( 1 \leq i \leq d \) define a basis of the tangent space, which is the image of the canonical basis under the tangent map defined by the Jacobian matrix \( DF(\hat{x}) \). In the case of an affine map \( F(\hat{x}) = A\hat{x} + b \), mapping the reference \( d \)-dimensional simplex into the simplex \((a_0, \ldots, a_d)\), we have \( DF(\hat{x}) = A \), independently of \( \hat{x} \) and then \( t_i = a_i - a_0 \), so that the tangent basis consists of the vectors going from \( a_0 \) to each of the other vertices of the simplex. See Figure 1.6 for a sketch in a triangle.

Note that the direction and length of the tangent basis vectors \( t_i \) are completely determined by the mapping \( F \). They form a basis of \( \mathbb{R}^d \) provided \( DF(\hat{x}) \) is invertible at all points, in other words, provided the mapping \( F \) is a \( C^1 \) diffeomorphism. However this basis is in general neither orthogonal nor normed. In an orthonormal basis the coefficients of any vector in this basis \( v = v_1e_1 + \cdots + v_de_d \) can be expressed in the euclidian space \( \mathbb{R}^d \), i.e. endowed with a scalar product, by \( v_i = v \cdot e_i \). This does not work anymore in a non-orthogonal basis. This is why in this case it is useful to use the

![Figure 1.6: Left: Tangent basis, Right: Tangent and cotangent basis vectors](image-url)
dual basis, which consists of elements of $\mathcal{L}(\mathbb{R}^d)$, that is linear forms on $\mathbb{R}^d$, which map any vector to a real number. The dual basis is by definition the unique basis of the dual space such that $d_i(t_j) = \delta_{i,j}$. Thanks to the scalar product, in a euclidian space, each element $d_i$ of the dual basis can be uniquely associated to a vector $n_i$ of $\mathbb{R}^d$ by defining $n_i$ as the unique vector verifying

\[ n_i \cdot v = d_i(v) \quad \forall v \in \mathbb{R}^d. \]

Considering now the basis $(t_1, \ldots, t_d)$, the dual basis $n_1, \ldots, n_d \in \mathbb{R}^d$ is defined by $n_i \cdot t_j = \delta_{i,j}$, which means that $n_i$ is locally orthogonal to the hyperplane defined by $(t_j)_{j \neq i}$ and normed by $n_i \cdot t_i = 1$.

For an affine mapping defined by a simplex, as introduced above, $n_i$ is normal to the face defined by all the vertices but $t_i$. See Figure 1.6 for a drawing in $\mathbb{R}^2$. An orthonormal basis is a special case: if the primal basis $(t_1, \ldots, t_d)$ is orthonormal, then we have, following the definition of the dual basis, that $n_i = t_i$ for $i = 1, \ldots, d$.

**Lemma 7** The matrix $N(\hat{x}) = [n_1, \ldots, n_d]$, whose columns are the elements of the dual basis of $(t_1, \ldots, t_d)$ verifies

\[ N(\hat{x}) = (DF(\hat{x}))^{-\top}. \]  

**Proof.** Let us denote $DF(\hat{x}) = [t_1, \ldots, t_d]$ to highlight that the tangent vectors are the columns of the Jacobian matrix. Then, due to the definition of the dual basis, we notice that

\[ N^\top DF = ((n_i \cdot t_j))_{1 \leq i, j \leq d} = ((\delta_{i,j}))_{1 \leq i, j \leq d} = I_d. \]

This shows that $N^\top = (DF)^{-1}$, which is equivalent to the result we are looking for.

**Proposition 1** In a three dimensional space ($d = 3$), the normal vectors are obtained from the tangent vector through the following relation:

\[ n_1 = \frac{1}{J} t_2 \times t_3, \quad n_2 = \frac{1}{J} t_3 \times t_1, \quad n_3 = \frac{1}{J} t_1 \times t_2. \]

**Proof.** As $n_1$ is by definition orthogonal to $t_2$ and $t_3$ is is colinear to $t_2 \times t_3$ and can be written $n_1 = \alpha t_2 \times t_3$ for some scalar $\alpha$. Now as $t_1 \cdot n_1 = 1$ and $t_1 \cdot t_2 \times t_3 = J$, we find $\alpha = 1/J$. The same reasoning applies to the two other components.

**Remark 7** In the physics literature, the tangent basis is generally called covariant basis and the basis of the dual cotangent space is called contravariant basis.

### 1.7.3 $H^1$ conforming Finite Elements

Let us build the conforming Finite Element space $V^0_h$ based on the reference element ($\hat{K}, \hat{P}, \hat{\Sigma}$) as a subspace of $H^1(\Omega)$. For this we need a mesh denoted by $T = \bigcup_{1 \leq e \leq N_{el}} K_i$ consisting of $N_{el}$ disjoint conforming elements denoted each by $K_i$. On this mesh we can then define $V^0_h \subset H^1(\Omega)$ by

\[ V^0_h = \{ v_h \in C^0(\Omega) | v_h|_{K_e} = \hat{v}_e \circ F_e^{-1} \in \hat{P}, \ 1 \leq e \leq N_{el} \}. \]

This means that the space $V^0_h$ is defined element by element and that the finite dimensional space on each element is defined as composition of the reference element space $\hat{P}$.
and the local element mapping $F_e$. In the case when $F_e$ is an affine mapping and $\hat{P}$ is a polynomial space, the mapped space is still a polynomial space of the same degree, but not in the general case. Let us denote by $N = \dim V_h^0$ the dimension of the full Finite Element approximation space, and $\hat{N} = \dim \hat{P}$ the dimension of the local approximation space on each Element. If there are degrees of freedom shared between several elements we have $N < N_e \hat{N}$, and in all cases $N \leq N_e \hat{N}$.

The Finite Element space $V_h^0$ being defined, the next step is to restrict the variational formulation of the problem being considered to functions in $V_h^0$. For the Poisson problem with Neumann boundary conditions (1.18) this leads to the following discrete variational formulation:

Find $u_h \in V_h^0$ such that

$$
\int_\Omega \nabla u_h \cdot \nabla v_h \, dx + \int_\Omega u_h v_h \, dx = \int_\Omega f v_h \, dx + \int_{\partial \Omega} g v_h \, d\sigma \quad \forall v_h \in V_h^0.
$$

(1.37)

This discrete variational formulation now this to be transformed to a linear system for practical implementation on a computer. To this aim, the integrals need to be decomposed on single elements and then mapped back to the reference elements so that the basis functions of the reference Finite Element space $\hat{P}$ can be plugged in.

First splitting the integrals over the elements (1.37) becomes

$$
\sum_{e=1}^{N_{el}} \int_{K_e} \nabla u_h \cdot \nabla v_h \, dx + \sum_{e=1}^{N_{el}} \int_{K_e} u_h v_h \, dx = \sum_{e=1}^{N_{el}} \int_{K_e} f v_h \, dx + \sum_{e=1}^{N_{el}} \int_{\Gamma_e} g v_h \, d\sigma \quad \forall v_h \in V_h^0.
$$

The last sum is on the element touching the boundary of $\Omega$. The next step is to make the change of variables $x = F_e(\hat{x})$ in the integral over $K_e$. Then $dx = J(\hat{x}) \, d\hat{x}$, where as introduced above, $J = \det \hat{D}$. In the integrals not involving derivatives we then simply get

$$
\int_{K_e} u_h \, dx = \int_{\hat{K}} \hat{u}_e \hat{v}_e J \, d\hat{x}, \quad \int_{K_e} f \, dx = \int_{\hat{K}} f \hat{v}_e J \, d\hat{x}.
$$

The boundary integral is dealt with the same way, using the restriction of the element mappings to the boundary and the appropriate Jacobian in the change of variables. For the integral involving the gradients, we need to make use of the chain rule (1.35)

$$
\int_{K_e} \nabla u_h \cdot \nabla v_h \, dx = \int_{\hat{K}} (DF)^{-T} \nabla \hat{u}_e \cdot (DF)^{-T} \nabla \hat{v}_e J \, d\hat{x}.
$$

Finally, we can express $\hat{u}_e, \hat{v}_e \in \hat{P}$, on the basis of $\hat{P}$ corresponding to the degrees of freedom $\hat{\Sigma}$. Let us denote by $(\hat{\varphi}_i)_{1 \leq i \leq \hat{N}}$ this basis. Then

$$
\hat{u}_e(\hat{x}) = \sum_{j=1}^{\hat{N}} u_j \hat{\varphi}_j(\hat{x}), \quad \hat{v}_e(\hat{x}) = \sum_{i=1}^{\hat{N}} v_i \hat{\varphi}_i(\hat{x}).
$$

By plugging these expressions in the element integrals defining the variational formulation we obtain a linear system. First

$$
\int_{\hat{K}} \hat{u}_e \hat{v}_e J \, d\hat{x} = \sum_{1 \leq i,j \leq \hat{N}} u_{ij} v_i \int_{\hat{K}} \hat{\varphi}_i \hat{\varphi}_j J \, d\hat{x} = V^T \hat{M}_e U, \quad \text{with} \quad \hat{M}_e = (\int_{\hat{K}} \hat{\varphi}_i \hat{\varphi}_j J \, d\hat{x})_{1 \leq i,j \leq \hat{N}}
$$
where $\hat{M}_e$ is called the element mass matrix, $V = (v_1, \ldots, v_N)^\top$, $U = (u_1, \ldots, u_N)^\top$. In the same way

$$
\int_{K} (DF)^{-\top} \hat{\nabla} \hat{v}_e (DF)^{-\top} \hat{\nabla} \hat{v}_e J \, d\hat{x} = V^T \hat{A}_e U, \quad \text{with } \hat{A}_e = \int_{K} (DF)^{-\top} \hat{\nabla} \hat{\varphi}_i (DF)^{-\top} \hat{\nabla} \hat{\varphi}_j J \, d\hat{x}.
$$

$\hat{A}_e$ is called the element stiffness matrix. Assuming for simplicity that the boundary term $g$ vanishes, it remains to assemble the right-hand-side depending on $f$. The contribution on the element $K_e$ is obtained by

$$
\int_{K} f \hat{v}_e J \, d\hat{x} = \sum_{i=1}^{N} v_i \int_{K} f \hat{\varphi}_i J \, d\hat{x} = \hat{V}^T b, \quad \text{with } b = (\int_{K} f \hat{\varphi}_i J \, d\hat{x})_{1 \leq i \leq N}.
$$

**From local to global.** To go from the element matrices to the global matrices, the sum over the elements need to be performed and contributions of degrees of freedom shared between different elements need to be added. For this a data structure associating a global number for each degree of freedom from the local number in each element needs to be created. This can define by a function $\Psi$ which to a couple, (element, local degree of freedom) $(e, i)$ associates the global number of the degree of freedom $I$ ($1 \leq I \leq \dim(V_0^h)$) of the basis function $\varphi_i$ of $V_0^h$ such that $(\varphi_I)_{|K_e} = \hat{\varphi}_i$. Then the global matrix $A$ is obtained by the algorithm:

\[
A = 0; \\
\text{for } e = 1, N_{el} \text{ do} \\
\quad \text{for } i, j = 1, \hat{N} \text{ do} \\
\quad \quad A(\Psi(e, i), \Psi(e, j)) = A(\Psi(e, i), \Psi(e, j)) + \hat{A}_e(i, j), \\
\quad \text{end} \\
\text{end}
\]

where $\hat{A}_e$ is the element stiffness matrix on $K_e$. The same procedure applies to all matrices involved in the variational formulation and also to the right-hand-side.

### 1.7.4 $H(\text{curl}, \Omega)$ conforming Finite Elements

Let us consider here the 3D case in which the curl operator is naturally defined. The 2D case can be obtained as a special case. The $H(\text{curl}, \Omega)$ function space and its conforming approximation consists of vectors. Hence on a change of basis, one needs to decide how the components of the vectors are transformed. In order to have a natural map of the tangents to a face used in the definition of the degrees of freedom and that guarantee conformity, the vectors defining the edges of the reference element needed to be mapped to either the covariant or the contravariant basis associated to the mapping. In the $H(\text{curl}, \Omega)$ case it turns out that the right basis is the contravariant basis ($\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3$).

A vector $\mathbf{v}$ on the image element is defined from a vector $\hat{\mathbf{v}}$ on the reference element by the transformation

$$
\mathbf{v}(\mathbf{x}) = (DF(\mathbf{x}))^{-\top} \hat{\mathbf{v}}(\mathbf{x}), \quad \text{with } \mathbf{x} = F(\mathbf{x}).
$$

This transformation is called the covariant Piola transform. In the context of differential forms it is the natural pullback transformation of a 1-form (in a 3D space).
As the columns of \((DF(\hat{x}))^{-\top}\) are \((n_1, n_2, n_3)\) as observed in (1.36), we see with this transformation formula that

\[ v(x) = \hat{v}_1(\hat{x})n_1(\hat{x}) + \hat{v}_2(\hat{x})n_2(\hat{x}) + \hat{v}_3(\hat{x})n_3(\hat{x}), \tag{1.39} \]

and in particular that the canonical basis \((\hat{e}_1, \hat{e}_2, \hat{e}_3)\) on the reference element is mapped into \((n_1, n_2, n_3)\). We will also need the images of the outward normals to the faces of the reference elements by this mapping.

**Lemma 8** The image of any outward normal \(\hat{v}\) of the reference element by the mapping (1.38) is \(\lambda v\), with \(\lambda = \|n_i\|\) in case the face of the reference element is orthogonal to \(\hat{e}_i\) and for the face opposite to the origin in the case of a tetrahedron \(\lambda = \|n_1 + n_2 + n_3\|/\sqrt{3}\).

**Proof.** The unit outward normal is constant on each face of the reference element.

For a tetrahedral reference element the unit normal to the face opposite to \(a_1\) is \(\hat{v}_1 = \hat{e}_1 = (1,0,0)^\top\), to the face opposite to \(a_2\) is \(\hat{v}_2 = \hat{e}_2 = (0,1,0)^\top\) and to the face opposite to \(a_3\) is \(\hat{v}_3 = \hat{e}_3 = (0,0,1)^\top\). Hence \(\hat{v}_i\) is mapped into \(n_i\), the corresponding vector of the contravariant basis. Because the outward unit vector on each face is orthogonal to the face, \(n_i\) and \(\nu\) are aligned and in the same direction. So the scaling factor with respect to the unit normal \(\tau\), whose norm is one, is given by \(\|n_i\|\). On the face opposite to \(a_0\), the outward unit normal is \(\nu = \frac{1}{\sqrt{3}}(\hat{e}_1 + \hat{e}_2 + \hat{e}_3)\), which is mapped by the transformation to \(n = \frac{1}{\sqrt{3}}(n_1 + n_2 + n_3)\). On the other hand, this \(n\) is orthogonal to the face \((a_1, a_2, a_3)\) of the element as \(n \cdot (t_2 - t_1) = n \cdot (t_3 - t_1) = 0\). Here the scaling factor compared to the unit normal is \(\|n\|\).

In the case of a cubic reference element, all the normals are parallel to one of the vector \(\hat{e}_i\) and the first part of the tetrahedron case applies. \(\blacksquare\)

In order to get an expression of a variational formulation defined in \(H(\text{curl}, \Omega)\), we now need to compute the curl of a vector in this basis.

**Proposition 2** We have the following properties:

1. For any vector field \(v\), defining \(Dv\) the Jacobian matrix of \(v\), we have

\[ (\nabla \times v) \times w = ((Dv)^{\top} - Dv) w \quad \forall w \in \mathbb{R}^d. \]

2. The transformation rule of the Jacobian matrix under the transformation (1.38) reads

\[ Dv(x) = S(\hat{x}) + (DF(\hat{x}))^{-\top}D\hat{v}(\hat{x})(DF(\hat{x}))^{-1} \]

where \(S\) is a symmetric matrix.

3. The transformation rule for the curl under the transformation (1.38) reads

\[ \nabla \times v(x) = \frac{1}{J(\hat{x})}DF(\hat{x})\hat{\nabla} \times \hat{v}(\hat{x}). \tag{1.40} \]

**Proof.** 1. Can be checked by comparing the expressions on the left and right hand sides.

2. We can express (1.38) as

\[ \hat{v}(\hat{x}) = (DF)^{\top}(\hat{x})v(F(\hat{x})), \]
so that
\[ \hat{v}_i(\hat{x}) = \sum_{k=1}^{3} \frac{\partial F_k(\hat{x})}{\partial \hat{x}_i} v_k(F(\hat{x})). \]

Let us now compute the coefficient on line \( i \) and column \( j \) of \( D\hat{\nu}(\hat{x}) \).
\[ \frac{\partial \hat{v}_i}{\partial \hat{x}_j} = \sum_{k=1}^{3} \left( \frac{\partial^2 F_k(\hat{x})}{\partial \hat{x}_i \partial \hat{x}_j} v_k(F(\hat{x})) \right) + \sum_{k=1}^{3} \left( \frac{\partial F_k(\hat{x})}{\partial \hat{x}_i} \sum_{l=1}^{3} \frac{\partial F_l}{\partial \hat{x}_j} \frac{\partial v_k}{\partial \hat{x}_l} \right), \]

so that we recognise that
\[ \hat{D}\hat{\nu} = \hat{S} + (DF)\top (D\nu(\hat{x}))(DF), \]

where \( \hat{S} = \sum_{k=1}^{3} \left( \frac{\partial^2 F_k(\hat{x})}{\partial \hat{x}_i \partial \hat{x}_j} v_k(F(\hat{x})) \right) \) is obviously symmetric. Multiplying this expression by \( (DF)^{-\top} \) on the left and by \( (DF)^{-1} \) on the right yields the result.

3. Using 1. and 2. we get, for an arbitrary vector \( w \in \mathbb{R}^3 \) we find
\[ (\nabla \times v) \times w = (DF)^{-\top} \left( (\hat{D}\hat{\nu}) - \hat{D}\hat{\nu} \right) (DF)^{-1} w = (DF)^{-\top} \left( (\hat{\nabla} \times \hat{v}) \times (DF)^{-1} w \right). \]

Not denoting by \( c_1, c_2, c_3 \) the columns of the matrix \( (DF)^{-1} \) and taking \( w = e_1 = (1,0,0)\top \) so that \( (DF)^{-1} w = c_1 \), we notice that
\[ (\nabla \times v) \times e_1 = \begin{pmatrix} 0 \\ (\nabla \times v)_3 \\ -(\nabla \times v)_2 \end{pmatrix} = \begin{pmatrix} c_1 \cdot (\hat{\nabla} \times \hat{v}) \times c_2 \\ c_2 \cdot (\hat{\nabla} \times \hat{v}) \times c_1 \\ c_3 \cdot (\hat{\nabla} \times \hat{v}) \times c_1 \end{pmatrix} = \begin{pmatrix} 0 \\ c_1 \cdot c_2 \cdot (\hat{\nabla} \times \hat{v}) \\ -c_3 \cdot c_1 \cdot (\hat{\nabla} \times \hat{v}) \end{pmatrix}. \]

And in the same way taking \( w = e_2 \) the second vector of the canonical basis
\[ (\nabla \times v) \times e_2 = \begin{pmatrix} -(\nabla \times v)_3 \\ 0 \\ (\nabla \times v)_1 \end{pmatrix} = \begin{pmatrix} c_1 \cdot (\hat{\nabla} \times \hat{v}) \times c_2 \\ c_2 \cdot (\hat{\nabla} \times \hat{v}) \times c_2 \\ c_3 \cdot (\hat{\nabla} \times \hat{v}) \times c_2 \end{pmatrix} = \begin{pmatrix} -c_1 \cdot c_2 \cdot (\hat{\nabla} \times \hat{v}) \\ 0 \\ c_2 \cdot c_3 \cdot (\hat{\nabla} \times \hat{v}) \end{pmatrix}. \]

It now remains to check that the inverse \( DF \) of \( (DF)^{-1} \) is defined by its three lines \( (c_2 \times c_3)/J, (c_3 \times c_1)/J, (c_1 \times c_2)/J \) to realise that the three components of the curl indeed satisfy (1.40).

Now using transformation (1.38) one can check by direct computation that all terms appearing in the curl Green’s formula (1.8) including the normal component are exactly preserved.

**Proposition 3** We have the following properties:

(i) \[ \int_K \nabla \times u \cdot v \, dx = \int_K \hat{\nabla} \times \hat{u} \cdot \hat{v} \, d\hat{x} \quad (1.42) \]

(ii) For any face \( f_i \) on the boundary of \( K \) image of the face \( \hat{f}_i \) of the reference element \[ \int_{f_i} u \times v \cdot d\sigma = \int_{\hat{f}_i} \hat{u} \times \hat{v} \cdot d\hat{\sigma} \quad (1.43) \]
We observe in particular, that the term involving the curl is completely independent of the metric and has the same expression under any mapping. This is true also for the boundary term.

**Proof.** (i) Let us observe, as the columns of $DF$ are the vectors $t_1, t_2, t_3$ that due to (1.40)

$$\nabla \times u(x) = \frac{1}{J(x)} (c_1 t_1 + c_2 t_2 + c_3 t_3),$$

where $c_1 = \partial_{x_2} \hat{u}_3 - \partial_{x_3} \hat{u}_2$, $c_2 = \partial_{x_3} \hat{u}_1 - \partial_{x_1} \hat{u}_3$, $c_3 = \partial_{x_1} \hat{u}_2 - \partial_{x_2} \hat{u}_1$. Then, making the change of variables $x = F(\hat{x})$, $dx = J(\hat{x}) d\hat{x}$ and plugging this expression into (1.42), we find

$$\int_K \nabla \times u \cdot v \, dx = \int_K \frac{1}{J(\hat{x})} (c_1 t_1 + c_2 t_2 + c_3 t_3) \cdot (\hat{v}_1 n_1 + \hat{v}_2 n_2 + \hat{v}_3 n_3) J(\hat{x}) \, d\hat{x} = \int_K (c_1 \hat{v}_1 + c_2 \hat{v}_2 + c_3 \hat{v}_3) \, d\hat{x} = \int_K \nabla \times \hat{u} \cdot \hat{v} \, d\hat{x}.$$

(ii) Lemma 8 gives us the image of the reference unit normal by the mapping. Let us first consider the case of a face of the reference element parallel to a coordinate plane. Let us express $u$ and $v$ in the contravariant basis using formula (1.39) and plug it into the expression of the left-hand side of (1.43), considering the face orthogonal to $n_1$, the others being dealt with in the same way. Note also that the boundary is defined as a parametric surface for which the area measure is defined by $d\sigma = ||t_2 \times t_3|| \, d\hat{\sigma}$, where $d\hat{\sigma} = d\hat{x}_2 \, d\hat{x}_3$. We use also that by definition of $n_1$ we have $||t_2 \times t_3|| = J ||n_1||$. We introduce the parameter $s$ to indicate whether $n_1$ and $\nu$ are in the same direction or not.

$$\int_{f_i} u \times v \cdot d\sigma = \int_{f_i} (\hat{u}_1(\hat{x}) n_1(\hat{x}) + \hat{u}_2(\hat{x}) n_2(\hat{x}) + \hat{u}_3(\hat{x}) n_3(\hat{x})) \times (sn_1/||n_1||) \cdot (\hat{v}_1 n_1(\hat{x}) + \hat{v}_2 n_2(\hat{x}) + \hat{v}_3 n_3(\hat{x})) J ||n_1|| \, d\hat{x}_2 \, d\hat{x}_3$$

$$= s \int_{f_i} (\hat{u}_3 \hat{v}_2 - \hat{u}_2 \hat{v}_3)(n_3 \times n_1 \cdot n_2) J \, d\hat{x}_2 \, d\hat{x}_3 = s \int_{f_i} (\hat{u}_3 \hat{v}_2 - \hat{u}_2 \hat{v}_3) J \, d\hat{x}_2 \, d\hat{x}_3$$

$$= \int_{f_i} \hat{u} \times \hat{v} \cdot d\hat{\sigma}$$

as $n_3 \times n_1 \cdot n_2 = 1/J$.

We can now assemble the matrices involved in a variational formulation posed on $H(\text{curl}, \Omega)$. Let us consider for example the variational formulation $Find \ u \in H(\text{curl}, \Omega)$ such that

$$\int_{\Omega} u \cdot v \, dx + \int_{\partial \Omega} \nabla \times u \cdot \nabla \times v \, dx = \int_{\Omega} f \cdot v \, dx \quad \forall v \in H(\text{curl}, \Omega). \quad (1.44)$$

Let us build the conforming Finite Element space $V_h^1 \subset H(\text{curl}, \Omega)$ based on the reference element $(K, \bar{P}, \bar{E})$. For this we need a mesh denoted by $T = \bigcup_{1 \leq e \leq N_{el}} K_e$ consisting of $N_{el}$ disjoint conforming elements denoted each by $K_e$. On this mesh we can then define $V_h^1 \subset H(\text{curl}, \Omega)$ by

$$V_h^1 = \{ v_h \in H(\text{curl}, \Omega) \mid v_h|_{K_e} = (DF)^{-T} \hat{v}_e \circ F_e^{-1} \in \bar{P}, \ 1 \leq e \leq N_{el} \}.$$
Now we can in the variational formulation (1.44) replace \( H(\text{curl}, \Omega) \) by \( V_h^1 \) and \( u \) and \( v \) by their discrete counterparts
\[
\mathbf{u}_h(x) = \sum_{j=1}^{N_1} u_j \varphi_j^1(x), \quad \mathbf{v}_h(x) = \sum_{i=1}^{N_1} v_i \varphi_i^1(x),
\]
where \((\varphi_i^1)_{1 \leq i \leq N_1}\) stands for the basis of \( V_h^1 \) associated to the Finite Element. As for the \( H^1(\Omega) \) conforming case, plugging these expressions into the variational formulation leads to the matrix expression:
\[
V^\top M_1 U + V^\top A_1 U = V^\top b,
\]
where \( M_1 = ((\int_{\Omega} \varphi_i^1 \cdot \varphi_j^1 \, dx))_{i,j} \) is the mass matrix in \( V_h^1 \), \( A_1 = ((\int_{\Omega} \nabla \times \varphi_i^1 \cdot \nabla \times \varphi_j^1 \, dx))_{i,j} \), and \( b = (\int_{\Omega} f \cdot \varphi_i^1)_{i} \). There now only remains to compute these integrals. Let us start with the mass matrix \( M_1 \).
\[
\int_{\Omega} \varphi_i^1 \cdot \varphi_j^1 \, dx = \sum_{e=1}^{N_e} \int_{K_e} \varphi_i^1 \cdot \varphi_j^1 \, dx.
\]
Then on each element \( K_e \) we make the change of variables \( x = F_e(\hat{x}) \), \( dx = J_e(\hat{x}) \, d\hat{x} \) to get
\[
\int_{K_e} \varphi_i^1 \cdot \varphi_j^1 \, d\hat{x} = \int_{K} (DF_e)^{-\top} \tilde{\varphi}_i^1 \cdot (DF_e)^{-\top} \tilde{\varphi}_j^1 \, J_e \, d\hat{x},
\]
\[
= \int_{K} (\tilde{\varphi}_{i,1}^1 n_1 + \tilde{\varphi}_{i,2}^1 n_2 + \tilde{\varphi}_{i,3}^1 n_3) \cdot (\tilde{\varphi}_{j,1}^1 n_1 + \tilde{\varphi}_{j,2}^1 n_2 + \tilde{\varphi}_{j,3}^1 n_3) J_e \, d\hat{x},
\]
\[
= \int_{K} (\tilde{\varphi}_i^1 \cdot G_e^c \tilde{\varphi}_j^1) J_e \, d\hat{x},
\]
where \( G_e^c = ((n_i \cdot n_j))_{1 \leq i,j \leq 3} \) is the contravariant metric tensor. This is the element mass matrix on the element \( K_e \) as for the \( H^1(\Omega) \), the full mass matrix is obtained by summing up the contributions from the different elements.

The procedure is the same for the \( A_1 \) matrix. Let us here simply compute the expression of the element matrix, using the expression of the curl given in (1.40):
\[
\int_{K_e} \nabla \times \varphi_i^1 \cdot \nabla \times \varphi_j^1 \, dx = \int_{K} \frac{1}{J_e} \tilde{\nabla} \times \tilde{\varphi}_i^1 \cdot \frac{1}{J_e} \tilde{D}(DF_e) \tilde{\nabla} \times \tilde{\varphi}_j^1 \, J_e \, d\hat{x},
\]
\[
= \int_{K} (c_{i,1} t_1 + c_{i,2} t_2 + c_{i,3} t_3) \cdot (c_{j,1} t_1 + c_{j,2} t_2 + c_{j,3} t_3) \frac{1}{J_e} \, d\hat{x},
\]
\[
= \int_{K} (\tilde{\nabla} \times \tilde{\varphi}_i^1)^\top G_e^c \tilde{\nabla} \times \tilde{\varphi}_j^1 \frac{1}{J_e} \, d\hat{x},
\]
where we have denoted \( \tilde{\nabla} \times \tilde{\varphi}_i = (c_{i,1}, c_{i,2}, c_{i,3})^\top \) and \( G_e^c = ((t_i \cdot t_j))_{1 \leq i,j \leq 3} \) is the covariant metric tensor.

### 1.7.5 \( H(\text{div}, \Omega) \) conforming Finite Elements

The \( H(\text{div}, \Omega) \) function space and its conforming approximation also consist of vectors. Hence on a change of basis, one needs to decide how the components of the vectors
are transformed. In order to have a natural map of the normal to a face used in the definition of the degrees of freedom and that guarantee conformity, the vectors defining the edges of the reference element needed to be mapped to either the covariant or the contravariant basis associated to the mapping. In the \( H(\text{div}, \Omega) \) case it turns out that the right basis is the covariant basis \((t_1, t_2, t_3)\). From the expression of the transformation rule for the curl (1.40) we see that components are exactly conserved if in addition we scale the length by \( 1/J(\hat{x}) \). This then leads to the following transformation rule

\[
v(\hat{x}) = \frac{1}{J(\hat{x})} DF(\hat{x}) \hat{v}(\hat{x}), \quad \text{with} \ x = F(\hat{x}). \quad (1.45)
\]

This transformation is called the contravariant Piola transform. In the context of differential forms it is the natural pullback transformation of a 2-form (in a 3D space).

In order to get an expression of a variational formulation defined in \( H(\text{div}, \Omega) \), we now need to compute the curl of a vector in this basis.

**Proposition 4** We have the following properties

(i) The jacobian matrix of a vector field transforms in the following way

\[
Dv = T + \frac{1}{J(\hat{x})} (DF(\hat{x}))^{-1} D\hat{v}(\hat{x})(DF(\hat{x})),
\]

where \( T \) is a matrix with vanishing trace.

(ii) The divergence transforms under the transformation rule (1.45)

\[
\nabla \cdot v(\hat{x}) = \frac{1}{J(\hat{x})} \nabla \cdot \hat{v}(\hat{x}).
\]

**Proof.** (i) The vector field \( v \) given by transform (1.45) can also be expressed on the covariant basis by

\[
v = \frac{1}{J}(\hat{v}_1 t_1 + \hat{v}_2 t_2 + \hat{v}_3 t_3).
\]

Taking the dot product with \( n_i \) yields \( \hat{v}_i = J n_i \cdot v \), so that the derivative becomes

\[
\frac{\partial \hat{v}_i}{\partial \hat{x}_j} = \frac{\partial J n_i}{\partial \hat{x}_j} \cdot v + J n_i \frac{\partial v}{\partial \hat{x}_j} = \frac{\partial J n_i}{\partial \hat{x}_j} \cdot v + J \sum_{k=1}^{3} \sum_{l=1}^{3} \frac{\partial \hat{v}_i}{\partial \hat{x}_k} \frac{\partial v_k}{\partial x_l} \frac{\partial x_l}{\partial \hat{x}_j},
\]

as \( n_i = \left( \frac{\partial \hat{v}_i}{\partial \hat{x}_1}, \frac{\partial \hat{v}_i}{\partial \hat{x}_2}, \frac{\partial \hat{v}_i}{\partial \hat{x}_3} \right)^\top \). Now denoting by \( \tilde{T} = ((\frac{\partial J n_i}{\partial \hat{x}_j} \cdot v))_{1 \leq i, j \leq 3} \), this implies that

\[
\tilde{D}v = \tilde{T} + J(DF)Dv(DF)^{-1}.
\]

Let us not verify that the trace of \( \tilde{T} \) vanishes. To this aim let us first recall that

\[
J n_1 = t_2 \times t_3, \quad J n_2 = t_3 \times t_1, \quad J n_3 = t_1 \times t_2.
\]

Moreover from the expression of \( t_j \) we find that \( \frac{\partial n_i}{\partial \hat{x}_j} = \frac{\partial t_i}{\partial \hat{x}_j} \), so that

\[
\text{tr} \tilde{T} = \left( \frac{\partial t_2}{\partial \hat{x}_1} \times t_3 + \frac{\partial t_3}{\partial \hat{x}_1} \times t_2 + \frac{\partial t_3}{\partial \hat{x}_2} \times t_1 + \frac{\partial t_1}{\partial \hat{x}_2} \times t_3 + \frac{\partial t_1}{\partial \hat{x}_3} \times t_2 + \frac{\partial t_2}{\partial \hat{x}_3} \times t_1 \right) \cdot v,
\]

\[
= \left( \frac{\partial t_2}{\partial \hat{x}_1} \times t_3 + \frac{\partial t_3}{\partial \hat{x}_1} \times t_2 + \frac{\partial t_3}{\partial \hat{x}_2} \times t_1 + \frac{\partial t_1}{\partial \hat{x}_2} \times t_3 + \frac{\partial t_1}{\partial \hat{x}_3} \times t_2 + \frac{\partial t_2}{\partial \hat{x}_3} \times t_1 \right) \cdot v,
\]

\[
= 0.
\]
We get the desired result by dividing by \( j \), multiplying by \((DF)^{-1}\) on the left and by \((DF)\). Indeed the matrix \( T \) obtained from \( \hat{T} \) in this process is similar to \( \hat{T} \) up to the scalar scaling factor \( J \) and as similar matrices have the same trace it follows that \( \text{tr} T = 0 \).

(ii) Relation (1.47) is obtained by taking the trace of (1.46) and using again that similar matrices have the same trace and that \( \text{tr} T = 0 \)

Now using transformation (1.45) one can check by direct computation that all terms appearing in the divergence Green’s formula (1.7) including the normal component are exactly preserved.

**Proposition 5** We have the following properties:

(i) 
\[
\int_K u \cdot \nabla \varphi \, dx = \int_{\hat{K}} \hat{u} \cdot \hat{\nabla} \hat{\varphi} \, d\hat{x} \tag{1.48}
\]

(ii) 
\[
\int_K \nabla \cdot u \, dx = \int_{\hat{K}} \hat{\nabla} \cdot \hat{u} \, d\hat{x} \tag{1.49}
\]

(iii) For any face \( f_i \) on the boundary of \( K \) image of the face \( \hat{f}_i \) of the reference element
\[
\int_{f_i} u \cdot \nu \varphi \, d\sigma = \int_{\hat{f}_i} \hat{u} \cdot \hat{\nu} \hat{\varphi} \, d\hat{\sigma} \tag{1.50}
\]

We observe in particular, that all these terms are completely independent of the metric and have the same expression under any mapping.

**Proof.** (i) Let us as usual perform the change of variables \( x = F(\hat{x}) \, dx = J(\hat{x}) \, d\hat{x} \), using the transformation rule (1.45) for \( u \) in the integral
\[
\int_K u \cdot \nabla \varphi \, dx = \int_{\hat{K}} \frac{1}{J} (DF) \hat{u} \cdot (DF)^{-\top} \hat{\nabla} \hat{\varphi} J \, d\hat{x} = \int_{\hat{K}} \hat{u} \cdot \hat{\nabla} \hat{\varphi} \, d\hat{x}.
\]

(ii) With the same change of variable and transformation rule for \( u \), using (1.47) we find
\[
\int_K \nabla \cdot u \, dx = \int_{\hat{K}} \frac{1}{J} \hat{\nabla} \cdot \hat{u} \hat{\varphi} J \, d\hat{x} = \int_{\hat{K}} \hat{\nabla} \cdot \hat{u} \hat{\varphi} \, d\hat{x}
\]

(iii) Here the face \( f_i \) is defined as a parametrized surface, the parametrization being given by one of the faces of the reference element. Let us do the computation for the face on the \((\hat{x}_1, \hat{x}_2)\) plane for with the normal vector is \( \hat{\nu} = -\hat{e}_3 \). The parametrization is then defined by
\[
(\hat{x}_1, \hat{x}_2) \mapsto F(\hat{x}_1, \hat{x}_2, 0)
\]
and the measure is
\[
d\sigma = \|t_1 \times t_2\| \, d\hat{x}_1 \, d\hat{x}_2 = J \|n_3\| \, d\hat{x}_1 \, d\hat{x}_2.
\]
Moreover the outward unit normal vector is related to the contravariant vector \( n_3 \), which is by definition orthogonal to the face, by \( \nu = -n_3/\|n_3\| \). Finally, applying the transformation rule (1.45) for \( u \) we get

\[
\int_{f_i} \mathbf{u} \cdot \nu \, d\sigma = -\int_{f_i} \frac{1}{J} (DF) \hat{u} \cdot \frac{n_3}{\|n_3\|} \hat{\nu} \|n_3\| \, d\hat{x}_1 \, d\hat{x}_2
\]

\[
= -\int_{f_i} (\hat{u}_1 t_1 + \hat{u}_2 t_2 + \hat{u}_3 t_3) \cdot n_3 \hat{\nu} \, d\hat{x}_1 \, d\hat{x}_2 = -\int_{f_i} \hat{u}_3 \hat{\nu} \, d\hat{x}_1 \, d\hat{x}_2 = \int_{f_i} \hat{u} \cdot \hat{\nu} \, d\hat{x}_1 \, d\hat{x}_2
\]

To conclude, we assemble the matrices involved in a variational formulation posed on \( H(\text{div}, \Omega) \). Let us consider for example the variational formulation

\[
\text{Find} \ u \in H(\text{div}, \Omega) \text{ such that }
\int_{\Omega} \mathbf{u} \cdot \mathbf{v} \, d\mathbf{x} + \int_{\Omega} \nabla \cdot \mathbf{u} \cdot \nabla \cdot \mathbf{v} \, d\mathbf{x} = \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} \quad \forall \mathbf{v} \in H(\text{div}, \Omega),
\]

(1.51)

We need to build the conforming Finite Element space \( V_h^2 \subset H(\text{div}, \Omega) \) based on the \( H(\text{div}, \Omega) \) reference element \( (K, \bar{P}, \bar{\Sigma}) \). For this we consider as previously a mesh denoted by \( T = \bigcup_{1 \leq e \leq N_{el}} K_e \) consisting of \( N_{el} \) disjoint conforming elements denoted each by \( K_e \).

On this mesh we can then define \( V_h^2 \subset H(\text{div}, \Omega) \) by

\[
V_h^2 = \{ \mathbf{v}_h \in H(\text{div}, \Omega) \mid \mathbf{v}_h|_{K_e} = \frac{1}{J} (DF) \hat{v}_e \circ F^{-1}_e \in \bar{P}, \ 1 \leq e \leq N_{el} \}.
\]

Now we can in the variational formulation (1.44) replace \( H(\text{div}, \Omega) \) by \( V_h^2 \) and \( u \) and \( v \) by their discrete counterparts

\[
\mathbf{u}_h(x) = \sum_{j=1}^{N_2} u_j \varphi_j^2(x), \quad \mathbf{v}_h(x) = \sum_{i=1}^{N_2} v_i \varphi_i^2(x),
\]

where \( (\varphi_i^2)_{1 \leq i \leq N_2} \) stands for the basis of \( V_h^2 \) associated to the Finite Element. As for the \( H^1(\Omega) \) and \( H(\text{curl}, \Omega) \) conforming case, plugging these expressions into the variational formulation leads to the matrix expression:

\[
V^\top M_2 U + V^\top A_2 U = V^\top b,
\]

where \( M_2 = ((\int_{\Omega} \varphi_i^2 \cdot \varphi_j^2 \, dx))_{i,j} \) is the mass matrix in \( V_h^2 \), \( A_2 = ((\int_{\Omega} \nabla \cdot \varphi_i^2 \cdot \nabla \cdot \varphi_j^2 \, dx))_{i,j} \), and \( b = ((\int \mathbf{f} \cdot \varphi_i^2)) \). There now only remains to compute these integrals. Let us start with the mass matrix \( M_2 \)

\[
\int_{\Omega} \varphi_i^2 \cdot \varphi_j^2 \, dx = \sum_{e=1}^{N_{el}} \int_{K_e} \varphi_i^2 \cdot \varphi_j^2 \, dx.
\]

Then on each element \( K_e \) we make the change of variables \( x = F_e(\hat{x}), \ dx = J_e(x) \, d\hat{x} \), using the contravariant Piola transform (1.45) for the vectors to get

\[
\int_{K_e} \varphi_i^2 \cdot \varphi_j^2 \, dx = \int_{K_e} \frac{1}{J_e} (DF_e) \hat{\varphi}_i \cdot \frac{1}{J_e} (DF_e) \hat{\varphi}_j \, d\hat{x},
\]

\[
= \int_{K_e} \frac{1}{J_e} (\hat{\varphi}_i^2 t_1 + \hat{\varphi}_i^2 t_2 + \hat{\varphi}_i^2 t_3) \cdot (\hat{\varphi}_j^2 t_1 + \hat{\varphi}_j^2 t_2 + \hat{\varphi}_j^2 t_3) \, d\hat{x},
\]

\[
= \int_{K_e} (\hat{\varphi}_i^2)^\top G bizarre \frac{1}{J_e} \, d\hat{x},
\]

32
where \( G_t = ((t_i \cdot t_j))_{1 \leq i,j \leq 3} \) is the covariant metric tensor. This is the element mass matrix on the element \( K_e \) as for the \( H^1(\Omega) \), the full mass matrix is obtained by summing up the contributions from the different elements.

The procedure is the same for the \( A_2 \) matrix. Let us here compute the expression of the element matrix, using the expression of the divergence given in (1.47):

\[
\int_{K_e} \nabla \cdot \varphi^2_i \nabla \cdot \varphi^2_j \, \mathrm{d}x = \int_{\hat{K}_e} \frac{1}{J_e} \hat{\nabla} \cdot \hat{\varphi}^2_i \hat{\nabla} \cdot \hat{\varphi}^2_j \, \mathrm{d}\hat{x},
\]

Note that here the divergence is a scalar. There is no metric tensor in the matrix. Only the inverse Jacobian is present because of the mapping.

### 1.8 Convergence of the Finite Element method

The Ritz-Galerkin method consists in finding an approximate solution \( u_h \) in a finite dimensional subspace of \( V \). For convergence studies one needs to consider a sequence of subspaces of \( V \) of larger and larger dimension so that they get closer to \( V \). One then defines a sequence of problems parametrised by \( h \) that read:

Find \( u_h \in V_h \) such that

\[
a(u_h, v_h) = l(v_h) \quad \forall v_h \in V_h,
\]

where \( V_h \subset V \) is a vector space of dimension \( N \). Let \( (\varphi_1, \ldots, \varphi_N) \) a basis of \( V_h \). An element \( u_h \in V_h \) can then be expanded as \( u_h(x) = \sum_{j=1}^{N} u_j \varphi_j(x) \). Taking \( v_h = \varphi_i \) the equation (1.52) becomes using the linearity

\[
\sum_{j=1}^{N} u_j a(\varphi_j, \varphi_i) = l(\varphi_i).
\]

Then using the symmetry of \( a \), we notice that the discrete variational formulation (1.52) is equivalent to the linear system

\[
AU_h = L,
\]

where \( A = (a(\varphi_i, \varphi_j))_{1 \leq i,j \leq N} \), \( L \) is the column vector with components \( l(\varphi_i) \) and \( U \) is the column vector with the unknowns \( u_i \) that are the coefficients of \( u_h \) in the basis \( (\varphi_1, \ldots, \varphi_N) \).

**Theorem 2** Assume that \( a \) is a symmetric continuous and coercive bilinear form on a Hilbert space \( V \) and \( l \) a continuous linear form on \( V \). Then the system (1.53) is equivalent to the discrete variational form (1.52) and admits a unique solution

**Proof.** For \( v_h \in V_h \), we denote by \( \tilde{V} \) the vector of its components in the basis \( (\varphi_1, \ldots, \varphi_N) \).

- Thanks to the bilinearity of \( a \) and the linearity of \( l \) the relation (1.52) can be written equivalently

\[
^t \tilde{V} AU_h = ^t \tilde{V} L \quad \forall \tilde{V} \in \mathbb{R}^N,
\]

which means that the vector \( AU_h - L \in \mathbb{R}^N \) is orthogonal to all the vectors of \( \mathbb{R}^N \), and so is the zero vector. Conversely it is clear that (1.53) implies (1.54) and so (1.52).
Let $v_h \in V_h$. Then, as $a$ is coercive, there exists $\alpha > 0$ such that

$$t \tilde{V} A \tilde{V} = a(v_h, v_h) \geq \alpha \|v_h\|^2 \geq 0,$$

and $t \tilde{V} A \tilde{V} = 0 = a(v_h, v_h) \Rightarrow \|v_h\| = 0$, which implies that $v_h = 0$ and so $\tilde{V} = 0$. So $A$ is symmetric, positive definite and therefore invertible.

After making sure the approximate solution exists for some given space $V_h$, one needs to make sure the approximation converges towards the exact solution. This results from two properties: 1) The Galerkin orthogonality, which comes from the conforming Gakerkin approximation, 2) The approximability property, which makes sure that for any $v \in V$ there exist $v_h$ in some finite dimensional space of the family which is close enough to $v$.

**Lemma 9 (Céa)** Let $u \in V$ the solution of (1.19) and $u_h \in V_h$ the solution of (1.52), with $V_h \subset V$. Then

$$\|u - u_h\| \leq C \inf_{v_h \in V_h} \|u - v_h\|.$$

**Proof.** We have

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

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

as $V_h \subset V$, we can take $v = v_h$ in the first equality and take the difference which yields

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

It results that $a(u - u_h, u - u_h) = a(u - u_h, u - v_h + v_h - u_h) = a(u - u_h, u - v_h)$, as $v_h - u_h \in V_h$ and so $a(u - u_h, v_h - u_h) = 0$. Then there exists $\alpha > 0$ and $\beta$ such that

$$\alpha \|u - u_h\|^2 \leq a(u - u_h, u - u_h) \quad \text{as } a \text{ is coercive},$$

$$\leq a(u - u_h, u - v_h) \quad \forall v_h \in V_h,$$

$$\leq \beta \|u - u_h\| \|u - v_h\| \quad \text{as } a \text{ is continuous}. $$

Whence $\|u - u_h\| \leq \frac{\beta}{\alpha} \|u - v_h\|$ for all $v_h \in V_h$. We get the desired results taking the infimum in $V_h$.

For the global error estimates, we make the following hypotheses on the triangulation $T_h$, where we denote by $h_K$ the diameter of the circumscribed circle and $\rho_K$ the diameter of the inscribed circle of triangle $K$:

(H1) We assume that the family of triangulations is regular in the following sense:

(i) There exists a constant $\sigma$ such that

$$\forall K \in \bigcup_{h} T_h \quad \frac{h_K}{\rho_K} \leq \sigma.$$

(ii) The quantity $h = \max_{K \in h} h_K$ tend to 0.
(H2) All finite elements \((K, P, \Sigma), K \in \bigcup h T_h\) are affine equivalent to a unique reference element \((\hat{K}, \hat{P}, \hat{\Sigma})\).

(H3) All finite elements \((K, P, \Sigma), K \in \bigcup h T_h\) are of class \(C^0\).

**Theorem 3** We assume the hypotheses (H1), (H2) and (H3) are verified. Moreover we assume that there exists an integer \(k \geq 1\) such that

\[
P_k \subset \hat{P} \subset H^1(\hat{K}),
\]

\[
H^{k+1}(\hat{K}) \subset C^0(\hat{K}) \quad \text{(true if } k + 1 > \frac{d}{2}, \hat{K} \subset \mathbb{R}^d).\]

Then there exists a constant \(C\) independent of \(h\) such that for any function \(v \in H^{k+1}(\Omega)\) we have

\[
\|v - \pi_h v\|_{0, \Omega} \leq C h^{k+1} |v|_{k+1, \Omega}, \\
|v - \pi_h v|_1 \leq C h^k |v|_{k+1, \Omega},
\]

where \(\pi_h\) is the finite element interpolation operator of the finite element \((K, P, \Sigma)\) defined by

\[
\pi_h v = \sum_{i=1}^{N} \sigma_i (v) p_i,
\]

where \(\sigma_i \in \Sigma\) and \(p_i\) the corresponding basis function of \(P\).

We consider a variational problem posed in \(V \subset H^1(\Omega)\).

**Theorem 4** We assume that (H1), (H2) and (H3) are verified. Moreover we assume that there exists an integer \(k \geq 1\) such that \(k + 1 > \frac{d}{2}\) with \(P_k(\hat{K}) \subset P \subset H^1(\hat{K})\) and that the exact solution of the variational problem is in \(H^{k+1}(\Omega)\), then

\[
\|u - u_h\|_{1, \Omega} \leq C h^k |u|_{k+1, \Omega},
\]

where \(u_h \in V_h\) is the discrete solution.

**Proof.** We have because of the polynomial approximation theorem

\[
\|u - \pi_h u\|_{1, \Omega} \leq C h^k |u|_{k+1, \Omega}.
\]

On the other hand Céa’s lemma gives us

\[
\|u - u_h\|_{1, \Omega} \leq C \inf_{v_h \in V_h} \|u - v_h\|_{1, \Omega} \leq C \|u - \pi_h u\|_{1, \Omega}.
\]

The result follows. \[\Box\]
Chapter 2

Non conforming and mixed Finite Element methods

2.1 Variational crimes

We introduce here a framework that can be used when the hypothesis of the conforming finite element methods breakdown, either because the discretisation space $V_h$ is not included in the space $V$ where the variational formulation to be approximated is defined, or when the discrete bilinear form and the discrete linear form are not exactly the same the continuous ones.

We consider the variational formulation: Find $u \in V$ such that

$$a(u,v) = l(v) \quad \forall v \in V.$$ (2.1)

In some cases it can be convenient to relax the constraints of the approximation space being a subspace of $V$ or to slightly modify the bilinear and linear form, using for example quadrature rules to compute the integrals or adding penalty terms. Let us then consider a mesh dependent family of Hilbert spaces $H_h$ equipped with a mesh dependent norm $\| \cdot \|_h$ such that $V \subset H_h$ and also $V_h \subset H_h$ for all $h$. One could choose $H_h = V + V_h$, but often some additional regularity of the exact solution is needed. Then one can choose $H_h = V \cap H^m(\Omega) + V_h$, where $H^m$ is a Sobolev space of the needed regularity. We then define the bilinear form $a_h$ on $H_h$ as an approximation of the bilinear form $a$ on $V$, and $l_h$ a linear form on $H_h$ approximating $l$ the linear form on $V$. We assume that $a_h$ and $l_h$ are continuous and that $a_h$ is uniformly coercive on $H_h$, meaning that the coercivity constant does not depend on $h$, even though the norm itself can depend on $h$.

As a typical example, given a subdivision of the domain $\Omega$ into elements $\Omega = \cup K_e$ we can define

$$H_h = \{ v \in L^2(\Omega) \mid v|_{K_e} \in H^1(K_e) \}, \quad \text{with } \| v \|_h^2 = \sum_e \int_{K_e} |\nabla v(x)|^2 \, dx.$$ 

This is a broken $H^1$ space. Its elements are locally $H^1$ but not globally. Obviously $V = H^1(\Omega) \subset H_h$. For $a(u,v) = \int_\Omega \nabla u(x) \cdot \nabla v(x) \, dx$ and for a given $f \in L^2(\Omega)$, $l(v) = \int_\Omega f(x)v(x) \, dx$, we define the approximations on $H_h$ by

$$a_h(u,v) = \sum_e \int_{K_e} \nabla u(x) \cdot \nabla v(x) \, dx.$$
On the other hand $l$ is well defined on $H_h$ so that we can take $l_h = l$. Obviously for $u, v \in V$, we have $a_h(u, v) = a(u, v)$ and $l_h(v) = l(v)$.

In this setting, we consider a discrete variational formulation of the following form: Find $u_h \in V_h$ such that

\[ a_h(u_h, v_h) = l_h(v_h) \quad \forall v_h \in V_h. \tag{2.2} \]

The error estimation of such an approximate variational formulation depends on the following lemma, known as Strang’s second lemma [9], which replaces Céa’s lemma:

**Lemma 10 (Strang)** Assume $u \in V$ is the solution of (2.1) and $u_h \in V_h$ its approximation, solution of (2.2), then

\[ \| u - u_h \|_h \leq C \left( \inf_{v_h \in V_h} \| u - v_h \|_h + \sup_{v_h \in V_h} \left| a_h(u, v_h) - l_h(v_h) \right| \right). \tag{2.3} \]

Note that as $u_h$ verifies $a_h(u_h, v_h) = l_h(v_h)$, we also could write $a_h(u, v_h) - l_h(v_h) = a_h(u - u_h, v_h)$.

**Remark 8** The difference between Céa’s Lemma that is used for the conforming approximation and the above Strang Lemma used for non conforming approximation is the presence of the second term on the right-hand-side, which is a consistency term, measuring how accurately the continuous solution verifies the discrete equation. This term vanishes when $V_h \subset V$ and $a_h = a$ because of the Galerkin orthogonality.

**Proof.** By bilinearity of $a_h$, we have

\[ a_h(v_h - u_h, v_h - u_h) = a_h(v_h - u + u - u_h, v_h - u_h) = a_h(v_h - u, v_h - u_h) + a_h(u - u_h, v_h - u_h) \]

and also

\[ a_h(u - u_h, v_h - u_h) = a_h(u, v_h - u_h) - a_h(u_h, v_h - u_h), \]

\[ = a_h(u, v_h - u_h) - l_h(v_h - u_h), \]

as $u_h$ verifies (2.2).

Using the uniform coercivity and continuity of $a_h$ in $H_h$, there exists $\alpha > 0$ and $\beta$ such that for any $v_h \in V_h$

\[ \alpha \|v_h - u_h\|^2 \leq a_h(v_h - u_h, v_h - u_h) \leq a_h(v_h - u, v_h - u_h) + a_h(u - u_h, v_h - u_h) \]

\[ \leq a_h(v_h - u, v_h - u_h) + a_h(u, v_h - u_h) - l_h(v_h - u_h), \]

\[ \leq \beta \|v_h - u\|_h \|v_h - u_h\|_h + \frac{\left| a_h(u, v_h - u_h) - l_h(v_h - u_h) \right|}{\|v_h - u_h\|_h} \|v_h - u_h\|_h \]

as $a_h$ is continuous. Whence

\[ \|v_h - u_h\|_h \leq \frac{\beta}{\alpha} \|u - v_h\|_h + \frac{1}{\alpha} \sup_{v_h \in V_h} \left| a_h(u, v_h) - l_h(v_h) \right| \]

for all $v_h \in V_h$. This combined with the triangle inequality yields

\[ \|u - u_h\|_h \leq \|u - v_h\|_h + \|v_h - u_h\|_h \leq \left( 1 + \frac{\beta}{\alpha} \right) \|u - v_h\|_h + \frac{1}{\alpha} \sup_{v_h \in V_h} \left| a_h(u, v_h) - l_h(v_h) \right| \|v_h\|_h. \]

We then get the desired results taking the infimum in $V_h$. \qed
2.2 Non conforming $\mathbb{P}_k$ finite elements on triangles

Let us consider the Poisson equation with Dirichlet boundary conditions given $f \in L^2(\Omega)$, for which the variational formulation reads: Find $u \in H^1_0(\Omega)$ such that

$$\int_{\Omega} \nabla u \cdot \nabla v \, dx = \int_{\Omega} f v \, dx \quad \forall v \in H^1_0(\Omega). \quad (2.4)$$

Standard conforming $\mathbb{P}_k$ finite elements on triangles have $k+1$ points on each edge, which are shared between elements to enforce continuity. Let us now consider the case, when only $k$ points are put on each side. The corresponding non conforming $\mathbb{P}_1$ and $\mathbb{P}_3$ elements are displayed on figure 2.1. They are also known as Crouzeix-Raviart Finite Elements. This is not enough to get continuity across the element and thus conformity, but is it still possible to define an optimal approximation? This is the case if the error coming from the second term in the second Strang Lemma is of the same order as the polynomial approximation error with $\mathbb{P}_k$ elements, i.e. $\|u - u_h\| \leq C h^k$. This can be obtained by putting the degrees of freedom on each edge at the Gauss points. Note however that this works only for $k$ odd, as for $k$ even the degrees of freedom obtained in this way are not linearly independent.

Consider the Lagrange Finite Element $(K_e, P_e, \Sigma_e)$, and $(a_i)_{i=1,N}$ the set of all degrees of freedom on all elements. As usual these degrees of freedom are shared by elements having an edge in common, but in the non-conforming case we are considering this is not enough to ensure continuity. The approximation space is then naturally defined as follows:

$$V_h = \{ v_h \in L^2(\Omega) \mid v_{h|K_e} \in \mathbb{P}_k, v_{h|K_{e1}}(a_i) = v_{h|K_{e2}}(a_i) \text{ if } a_i \in K_{e1} \cap K_{e2}, \quad v_h(a_i) = 0 \text{ if } a_i \text{ on } \partial\Omega\}. $$

As $V_h \not\subset H^1_0(\Omega)$, we cannot directly use the variational formulation (2.5). Instead we use its "broken" version: Find $u_h \in V_h$ such that

$$\sum_{e} \int_{K_e} \nabla u_h \cdot \nabla v_h \, dx = \int_{\Omega} f v_h \, dx \quad \forall v_h \in V_h. \quad (2.5)$$

We then are in the abstract framework defined above with

$$a_h(u_h, v_h) = \sum_{e} \int_{K_e} \nabla u_h \cdot \nabla v_h \, dx, \quad l_h(v_h) = l(v_h) = \int_{\Omega} f v_h \, dx.$$
Let us now express the consistency error in this non-conforming approximation. Let $u \in H^1_0(\Omega) \cap H^2(\Omega)$ be the exact solution of (2.5) and $v_h \in V_h$, then

$$a_h(u, v_h) - l_h(v_h) = \sum_{e} \int_{K_e} \nabla u \cdot \nabla v_h \, dx - \int_{\Omega} f v_h \, dx$$

(2.6)

$$= - \sum_{e} \int_{K_e} (\Delta u + f) \cdot v_h \, dx - \int_{\partial K_e} \frac{\partial u}{\partial n} v_h \, d\sigma$$

(2.7)

$$= \sum_{e} \int_{\partial K_e} \frac{\partial u}{\partial n} v_h \, d\sigma$$

(2.8)

as $u$ verifies $-\Delta u = f$.

For two adjacent elements $K_{e_1}$ and $K_{e_2}$ let us denote by $\Gamma_s = K_{e_1} \cap K_{e_2}$ their shared edge. Let $v_s$ be the interpolation of polynomial of degree $k - 1$ of $v_{h|K_{e_1}}$ and $v_{h|K_{e_2}}$ at the degrees of freedom on the edges of the two elements. As the degrees of freedom are shared by elements sharing the same edge, $v_s$ is the same on the two elements $K_{e_1}$ and $K_{e_2}$. Then as the unit outward normal for these two elements has opposite signs on $\Gamma_s$, it follows that

$$\int_{\Gamma_s} \frac{\partial u}{\partial n} v_s \, d\sigma + \int_{\Gamma_s} \frac{\partial u}{\partial n} v_s \, d\sigma = 0.$$

On the other hand if $\Gamma_s$ corresponds to a boundary of $\Omega$ the degrees of freedom there are set to zero, so that $v_s = 0$ on such a boundary. Hence adding the contributions over the whole mesh

$$a_h(u, v_h) - l_h(v_h) = \sum_{e} \int_{\partial K_e} \frac{\partial u}{\partial n} (v_h - v_s) \, d\sigma.$$

Let us now introduce $\pi_h u$, the Finite Element interpolant of $u$, which is by definition a polynomial of degree $k$ on each element. Then $\frac{\partial \pi_h u}{\partial n}$ is a polynomial of degree $k - 1$ on the edges of the mesh. So if, the degrees of freedom are chosen to be Gauss points on the edges, $v_h - v_s$ on each edge is a polynomial of degree $k$ having its zeros at the Gauss points. Hence it is, up to a multiplicative constant, the Legendre polynomial of degree $k$, which is orthogonal to all polynomials of degree $k - 1$. Thus for all edges

$$\int_{\Gamma_s} \frac{\partial \pi_h u}{\partial n} (v_h - v_s) \, d\sigma = 0.$$

Adding the contributions over the whole mesh it follows that

$$a_h(u, v_h) - l_h(v_h) = \sum_{e} \int_{\partial K_e} \left( \frac{\partial u}{\partial n} - \frac{\partial \pi_h u}{\partial n} \right) (v_h - v_s) \, d\sigma$$

$$\leq C \sum_{e} \| \nabla (u - \pi_h u) \cdot \mathbf{n} \|_{L^2(\partial K_e)} \| v_h - v_s \|_{\partial K_e},$$

using a Cauchy-Schwartz inequality. Then we can use the polynomial approximation properties on mesh faces as proved in [5] (Lemma 1.59, p. 32), assuming $u \in H^k(\Omega)$ with $k \geq 1$

$$\| \nabla (u - \pi_h u) \cdot \mathbf{n} \|_{L^2(\partial K_e)} \leq Ch^{k-1/2} |u|_{k+1, K_e}, \quad \| v_h - v_s \|_{\partial K_e} \leq C h^{k-1/2} |v_h|_{k+1, K_e},$$

(2.9)

$\pi_h u$ being a polynomial of degree $k$ approximating $u$ and $v_s$ a polynomial of degree $k - 1$ approximating $v_h$. Then it follows

$$\left| \frac{a_h(u, v_h) - l_h(v_h)}{\| v_h \|_h} \right| \leq C \sum_{e} h^{k-\frac{1}{2}} |u|_{k+1, K_e} h^{k-\frac{1}{2}} \leq C h^k |u|_{k+1, \Omega}.$$

39
2.3 Weak enforcement of Dirichlet boundary conditions

Another useful application of non conforming finite element is the weak enforcement of Dirichlet conditions following an idea of Nitsche [8]. We consider here a standard Finite Element conforming in $H^1$ (for the Poisson problem) but not in $H^0_0$.

Let us explain the method for the Poisson problem with Dirichlet boundary conditions. Here because the vector space where the solution lives is $H^1(\Omega)$ anyway, as for natural boundary conditions, non homogeneous boundary conditions are as easy to enforce as homogeneous boundary conditions. Our problem then reads

$$-\Delta u = f \quad \text{in } \Omega, \quad u = g \quad \text{on } \partial \Omega.$$  

In order to get a consistent variational formulation, as usual in the derivation of the Finite Element method we multiply the first equation by a test function $v$ and integrate on the whole domain

$$-\int_{\Omega} \Delta u v \, dx = \int_{\Omega} f v \, dx = \int_{\Omega} \nabla u \cdot \nabla v \, dx - \int_{\partial \Omega} \frac{\partial u}{\partial n} v \, d\sigma.$$  

In the standard conforming Finite Element method the boundary integral vanishes by taking test functions vanishing on the boundary. Instead here we keep it, moreover so that the problem remains symmetric, we add $\int_{\partial \Omega} u \frac{\partial v}{\partial n} \, d\sigma$, which is known as $u$ the exact solution is equal to a known function $g$ on the boundary. Then the exact solution of our problem satisfies the following relation

$$\int_{\Omega} \nabla u \cdot \nabla v \, dx - \int_{\partial \Omega} \frac{\partial u}{\partial n} v \, d\sigma - \int_{\partial \Omega} u \frac{\partial v}{\partial n} \, d\sigma = \int_{\Omega} f v \, dx - \int_{\partial \Omega} g \frac{\partial v}{\partial n} \, d\sigma \quad \forall v \in H^1(\Omega). \quad (2.10)$$  

The left-hand-side defines a symmetric bilinear form, but it is not coercive and thus does not lead to a well posed variational problem.

However, we can use this relation to define a consistent discrete variational problem. In order to make it coercive, we need to add an additional term $\mu_h \int_{\partial \Omega} u_h v_h \, d\sigma$, called penalty term which will be grid dependent, that vanishes for the exact solution and $\mu_h$ will be chosen such that the resulting bilinear form is coercive in some norm. For this we define the standard Lagrange approximation space for $H^1(\Omega)$

$$V_h = \{v_h \in C^0(\Omega) \mid v_h|_{K_e} \in P_k(K_e)\},$$

and then the discrete variational problem reads: Find $u_h \in V_h$ such that

$$\int_{\Omega} \nabla u_h \cdot \nabla v_h \, dx + \mu_h \int_{\partial \Omega} u_h v_h \, d\sigma - \int_{\partial \Omega} \frac{\partial u_h}{\partial n} v_h \, d\sigma - \int_{\partial \Omega} u_h \frac{\partial v_h}{\partial n} \, d\sigma$$

$$= \int_{\Omega} f v_h \, dx - \int_{\partial \Omega} g \frac{\partial v_h}{\partial n} \, d\sigma + \mu_h \int_{\partial \Omega} g v_h \, d\sigma \quad \forall v_h \in V_h.$$  

40
Proof. Let us first observe that for any abstract mesh depend variational problem of the form: 

\[ a_h(u_h, v_h) = \int_\Omega \nabla u_h \cdot \nabla v_h \, dx - \int_{\partial \Omega} u_h v_h \, d\sigma - \int_{\partial \Omega} u_h \frac{\partial v_h}{\partial n} \, d\sigma \]

\[ l_h(v_h) = \int_{\partial \Omega} g v_h \, d\sigma \]

Then, using the Cauchy-Schwarz inequality, the fact that the components of the normal vector are less than one, and the assumption of the Lemma \( \alpha > 0 \), one gets the optimal order of convergence.

\[ \alpha > h \lambda \]

Because \( V_h \subset H^1(\Omega) \), we can take in equation (2.10) \( u \) to be the exact solution and \( v = v_h \in V_h \). Then as in addition the exact solution \( u \) is equal to \( g \) on the boundary, the consistency error appearing in the second Strang Lemma (Lemma 10) vanishes completely:

\[ a_h(u, v_h) - l_h(v_h) = 0 \quad \forall v_h \in V_h. \]

We can now prove a lemma giving a condition on the penalisation parameter \( \mu_h \) such that the bilinear form \( a_h \) is coercive in the mesh depend norm

\[ \|v\|_h^2 = \|\nabla v\|_{0, \Omega}^2 + \frac{1}{h} \|v\|_{0, \partial \Omega}^2 \quad \forall v \in H^1(\Omega). \]

**Lemma 11** Assume that there exists \( C > 0 \) such that \( \|\nabla v\|_{0, \partial \Omega} \leq C h^{-1/2} \|\nabla v\|_{0, \Omega} \), then the bilinear form \( a_h \) is uniformly coercive in the mesh dependent norm \( \|\cdot\|_h \) if \( \mu_h = \alpha/h \) with \( \alpha > 2C^2 \).

**Proof.** Let us first observe that for any \( v \in H^1(\Omega) \)

\[ a_h(v, v) = \|\nabla v\|_{0, \Omega}^2 + \mu_h \|v\|_{0, \partial \Omega}^2 - 2 \int_{\partial \Omega} \frac{\partial v}{\partial n} v \, d\sigma = \|\nabla v\|_{0, \Omega}^2 + \frac{\alpha}{h} \|v\|_{0, \partial \Omega}^2 - 2 \int_{\partial \Omega} \frac{\partial v}{\partial n} v \, d\sigma. \]

Then, using the Cauchy-Schwarz inequality, the fact that the components of the normal vector are less than one, and the assumption of the Lemma

\[ 2 \int_{\partial \Omega} \frac{\partial v}{\partial n} v \, d\sigma \leq 2 \|\nabla v\|_{0, \partial \Omega} \|v\|_{0, \partial \Omega} \leq 2C h^{-1/2} \|\nabla v\|_{0, \Omega} \|v\|_{0, \partial \Omega} \leq \frac{1}{2} \|\nabla v\|_{0, \Omega}^2 + \frac{2C^2}{h} \|v\|_{0, \partial \Omega}^2 \]

using also for the last inequality the property \( ab \leq a^2/2 + b^2/2 \) for any real numbers \( a, b \). Plugging the opposite inequality into the expression of \( a_h(v, v) \) we get

\[ a_h(v, v) \geq \frac{1}{2} \|\nabla v\|_{0, \Omega}^2 + \frac{\alpha - 2C^2}{h} \|v\|_{0, \partial \Omega}^2 \geq \min\left\{ \frac{1}{2}, \frac{\alpha - 2C^2}{h} \right\} \|\nabla v\|_{0, \Omega}^2 + \frac{1}{h} \|v\|_{0, \partial \Omega}^2 \geq \frac{1}{2} \|\nabla v\|_{0, \Omega}^2 + \frac{1}{h} \|v\|_{0, \partial \Omega}^2. \]

This yields the uniform coercivity of \( a_h \) in the mesh dependent norm provided \( \alpha > 2C^2 \).

**Remark 9** In practical applications, it is best to choose \( \mu_h \) not too much larger than the minimal value needed for coercivity. This can be achieved for a given mesh, by solving the generalized eigenvalue problem

\[ \lambda_h u_h \cdot \nabla v_h \, dx - \int_{\partial \Omega} \frac{\partial u_h}{\partial n} v_h \, d\sigma - \int_{\partial \Omega} u_h \frac{\partial v_h}{\partial n} \, d\sigma = 0 \quad \forall v_h \in V_h \]

which can be written in matrix form, once the Finite Element matrices are assembled \( A = \lambda_h B \). The matrices being symmetric, all eigenvalues are real, and the smallest one denoted by \( \lambda_{h,0} \) should be negative because of the lack of coercivity. In order to get coercivity then one needs to choose \( \mu_h \geq \lambda_{h,0} \). If one chooses \( \mu_h = \alpha \lambda_{h,0} \) with \( \alpha \) a coefficient independent of \( h \) typically between 1 and 2, one gets the optimal order of convergence.
2.4 Discontinuous Galerkin methods

The Nitsche method to weakly enforce Dirichlet boundary conditions also inspired researchers for getting rid of the conformity at element interfaces using the idea of writing a consistent approximation involving the interface terms, then symmetrizing and penalizing to get a symmetric and coercive mesh dependent form. This leads to the so-called SIP (symmetric interior penalty) Discontinuous Galerkin (DG) method that was introduced, adapting previous works, and analyzed by Arnold [1]. See also [2] for a historical and more general discussion on Discontinuous Galerkin methods. A mathematical textbook presentation of the discontinuous Galerkin method can be found for example in [5].

Let us explain the method on the Poisson equation with homogeneous Dirichlet boundary conditions.

\[-\Delta u = f \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial \Omega.\]

Instead of multiplying by a smooth function over the whole domain as in the classical finite element method, we here introduce the mesh first \( T_h = \bigcup_e K_e \). Let \( v_h \in L^2(\Omega) \) such that \( v_h \in H^1(K_e) \). In practice \( v_h \) will be polynomial or in a local Finite Element space. Let us denote by \( n_e \) the outward unit normal for element \( K_e \) and denoting by \( (\Gamma_s)_s \) the edges of the grid, \( \Gamma_s = K_{e1} \cap K_{e2} \), by \( n_s \) a unit normal to the edge \( \Gamma_s \) (the sign being arbitrarily chosen so that \( n_s = n_{e1} \)). As usual in Discontinuous Galerkin formulations we denote by \( \{ v \} = (v|_{K_{e1}} + v|_{K_{e2}})/2 \) the average of \( v \) on the edge \( \Gamma_s \) and \( [ v ] = v|_{K_{e1}} - v|_{K_{e2}} \) the jump of \( v \) on the edge \( \Gamma_s \). For edges on the boundary of the domain, we assume that the outside value is zero so that \( \{ v \} = [ v ] = v \). Then, we multiply by a smooth function \( v_h|_{K_e} \) on one element, integrate over this element and sum over all the elements

\[
\sum_e \int_{K_e} (-\Delta u)v_h \, dx = \sum_e \left[ \int_{K_e} \nabla u \cdot \nabla v_h \, dx - \int_{\partial K_e} \frac{\partial u}{\partial n_e} v_h \, d\sigma \right]
= \sum_e \int_{K_e} \nabla u \cdot \nabla v_h \, dx - \sum_s \int_{\Gamma_s} \frac{\partial u}{\partial n_s} [v_h] \, d\sigma
= \sum_e \int_{K_e} \nabla u \cdot \nabla v_h \, dx - \sum_s \int_{\Gamma_s} \{ \frac{\partial u}{\partial n_s} \} [v_h] \, d\sigma \quad (2.11)
\]

as the exact solution, which is in \( H^2 \), has a continuous normal derivative across the edges, this is equal to its average.

This formula can be used to get a consistent approximation. As for the Nitsche method we now symmetrize and penalize the interface jumps. We then get the mesh dependent bilinear form

\[
a_h(u_h, v_h) = \sum_e \int_{K_e} (\nabla u_h \cdot \nabla v_h) \, dx - \sum_s \int_{\Gamma_s} \{ \frac{\partial u_h}{\partial n_s} \} [v_h] \, d\sigma - \sum_s \int_{\Gamma_s} [u_h] \{ \frac{\partial v_h}{\partial n_s} \} \, d\sigma + \mu_h \int_{\Gamma_s} [u_h] [v_h] \, d\sigma. \quad (2.12)
\]

Then, defining the linear form

\[
l_h(v_h) = \int_{\Omega} f v_h \, dx = \sum_e \int_{K_e} f v_h \, dx
\]
and the discontinuous polynomial space

\[ V_h = \{ v_h \in L^2(\Omega) \mid v_h|_{K_e} \in P^k(K_e) \} \],

the discrete variational formulation for our problem reads: Find \( u_h \in V_h \) such that

\[ a_h(u_h, v_h) = l_h(v_h) \quad \forall v_h \in V_h. \quad (2.13) \]

As the exact solution \( u \) is continuous, the jump of \( u \) at the element interfaces vanishes, i.e. \( [u] = 0 \), hence the two last terms in (2.12), which are respectively the symmetry term and the penalty term vanish for \( u_h = u \). Hence using the calculation (2.11) we can verify that the consistency error in the DG-SIP method (2.13) is exactly vanishing:

\[ a_h(u, v_h) - l_h(v_h) = \sum_e \int_{K_e} (-\Delta u)v_h \, dx - \int_{\Omega} fv_h \, dx = - \int_{\Omega} (\Delta u + f)v_h \, dx = 0 \]

as \( -\Delta u = f \).

Even though they are not conforming as \( V_h \not\subset V \), DG methods are built to be exactly consistent, so that the consistency error in the second Strang lemma 10 exactly vanishes. The error analysis described in the book by Di Pietro and Ern [5] relies on this property. See in particular Section 4.2, which provides a convergence analysis of the SIP-DG method in an ad hoc mesh dependent norm, following the same pattern as for the Nitsche method.

### 2.5 Saddle-point problems

#### 2.5.1 The theoretical framework

The variational problems we consider here fit into the following framework. Consider two Hilbert spaces \( V \) and \( W \), two continuous bilinear forms \( a \in L(V \times V, \mathbb{R}) \) and \( b \in L(V \times W, \mathbb{R}) \) and two continuous linear forms \( l_V \in L(V, \mathbb{R}) \) and \( l_W \in L(W, \mathbb{R}) \). Then we defined the abstract mixed variational problem as Find \( (u, p) \in V \times W \) such that

\[ a(u, v) + b(v, p) = l_V(v) \quad \forall v \in V; \quad (2.14) \]

\[ b(u, q) = l_W(q) \quad \forall q \in W. \quad (2.15) \]

For saddle point problems the Lax-Milgram framework cannot be applied. In this case the appropriate theoretical tool, called Banach-Nečas-Babuška (BNB) theorem in Ern and Guermond, which is more general than Lax-Milgram. We will specialise it to our type of mixed problems assuming that the bilinear form \( a \) is symmetric and coercive on \( K = \{ v \in V \mid b(q, v) = 0, \forall q \in W \} \). This corresponds to Theorem 2.34 p. 100 of [6], modified according to Remark 2.35.

**Theorem 5** Let \( V \) and \( W \) be Hilbert space. Assume \( a \) is a continuous bilinear form on \( V \times V \) and that \( b \) is a continuous linear form on \( V \times W \), that \( l_V \) and \( l_W \) are continuous linear forms on \( V \) and \( W \) respectively and that the following two hypotheses are verified

1) \( a \) is coercive on \( K = \{ v \in V \mid b(q, v) = 0, \forall q \in W \} \), i.e. there exists \( \alpha > 0 \) such that

\[ a(v, v) \geq \alpha \| v \|^2_V \quad \forall v \in K. \]
2) The Babuska-Brezzi, or inf-sup condition, is verified: there exists $\beta > 0$ such that

$$\inf_{q \in W} \sup_{v \in V} \frac{b(v, q)}{\|q\|_W \|v\|_V} \geq \beta.$$ 

Then the variational problem admits a unique solution and the solution satisfies the a priori estimate

$$\|u\|_V \leq \frac{1}{\alpha} \|f\|_V + \frac{1}{\beta} (1 + \frac{C}{\alpha}) \|g\|_W,$$

$$\|p\|_W \leq \frac{1}{\beta} (1 + \frac{C}{\alpha}) \|f\|_V + \frac{C}{\beta^2} (1 + \frac{C}{\alpha}) \|g\|_W,$$

where $C$ is the continuity constant of $a$.

The inf-sup conditions play an essential role, as it is only satisfied if the spaces $V$ and $W$ are compatible in some sense. This condition being satisfied at the discrete level with a constant $\beta$ that does not depend on the mesh size being essential for a well behaved Finite Element method. It can be written equivalently

$$\beta \|q\|_W \leq \sup_{v \in V} \frac{b(v, q)}{\|v\|_V} \forall q \in W. \quad (2.16)$$

And often, a simple way to verify it is, given any $q \in W$, to find a specific $v = v(q)$ depending on $q$ such that

$$\beta \|q\|_W \leq \frac{b(v(q), q)}{\|v(q)\|_V} \leq \sup_{v \in V} \frac{b(v, q)}{\|v\|_V}$$

with a constant $\beta$ independent of $w$.

Let us now come to the Galerkin discretisation. The principle is simply to construct finite dimensional subspaces $W_h \subset W$ and $V_h \subset V$ and to write the variational formulation (2.14)-(2.15) replacing $W$ by $W_h$ and $V$ by $V_h$. The variational formulations are the same as in the continuous case, like for conforming finite elements. This automatically yields the consistency of the discrete formulation. In order to get the stability property needed for convergence, we need that the coercivity constant $\alpha$ and the inf-sup constant $\beta$ are independent of $h$.

Because $V_h \subset V$ the coercivity property is automatically verified in the discrete case, with a coercivity constant that is the same as in the continuous case and hence does not depend on the discretisation parameter $h$.

Here, however, there is an additional difficulty, linked to the inf-sup conditions, which is completely dependent on the two spaces $V_h$ and $W_h$. By far not any conforming approximation of the two spaces will verify the discrete inf-sup condition with a constant $\beta$ that is independent on $h$. Finding compatible discrete spaces for a given mixed variational formulation, has been an active area of research.

### 2.5.2 Mixed formulation of the Poisson problem

Consider the Poisson problem

$$-\Delta p = f \quad (2.17)$$

$$p = 0 \quad (2.18)$$
Using that $\Delta p = \nabla \cdot \nabla p$, we set $u = \nabla p$, then the Poisson equation (2.17) can be written equivalently

$$u = -\nabla p, \quad \nabla \cdot u = f.$$ 

Instead of having one unknown, we now have two, along with the above two equations. In order to get a mixed variational formulation, we first take the dot product of the first one by $v$ and integrate by parts

$$\int u \cdot v \, dx - \int p \nabla \cdot v \, dx + \int_{\partial \Omega} p v \cdot n \, d\sigma = \int u \cdot v \, dx - \int p \nabla \cdot v \, dx = 0,$$

using $p = 0$ as a natural boundary condition. Then multiplying the second equation by $q$ and integrating yields

$$\int \nabla \cdot u q \, dx = \int f q \, dx.$$ 

No integration by parts is necessary here. And we thus get the following mixed variational formulation: Find $(u, p) \in H(\text{div}, \Omega) \times L^2(\Omega)$ such that

$$\int u \cdot v \, dx - \int p \nabla \cdot v \, dx = 0 \quad \forall v \in H(\text{div}, \Omega) \quad (2.19)$$

$$\int \nabla \cdot u q \, dx = \int f q \, dx, \quad \forall q \in L^2(\Omega). \quad (2.20)$$

Here also, we get an alternative formulation by not integrating by parts, the mixed term in the first formulation but in the second. The first formulation simply becomes

$$\int u \cdot v \, dx + \int \nabla p \cdot v \, dx = 0,$$

and the second, removing immediately the boundary term due to the essential boundary condition $q = 0$

$$\int \nabla \cdot u q \, dx = - \int u \cdot \nabla q \, dx = \int f q \, dx,$$

which leads to the variational formulation Find $(u, p) \in L^2(\Omega)^3 \times H^1_0(\Omega)$ such that

$$\int u \cdot v \, dx + \int \nabla p \cdot v \, dx = 0 \quad \forall v \in L^2(\Omega)^3 \quad (2.21)$$

$$\int u \cdot \nabla q \, dx = - \int f q \, dx, \quad \forall q \in H^1_0(\Omega). \quad (2.22)$$

Note that this formulation actually contains the classical variational formulation for the Poisson equation. Indeed for $q \in H^1_0(\Omega)$, $\nabla q \in L^2(\Omega)^3$ can be used as a test function in the first equation. And plugging this into the second we get

$$\int \nabla p \cdot \nabla q \, dx = \int f q \, dx, \quad \forall q \in H^1_0(\Omega).$$
Bibliography


